Stochastic Models and Statistical Analysis for Clock Noise

Donald B. Percival
Applied Physics Laboratory
Box 355640
University of Washington
Seattle, WA 98195–5640
USA
e-mail: dbp@apl.washington.edu

Abstract

In this primer we first give an overview of stochastic models that can be used to interpret clock noise. Because of their statistical tractability, we concentrate on fractionally differenced (FD) processes, which we relate to fractional Brownian motion, discrete fractional Brownian motion, fractional Gaussian noise and discrete pure power-law processes. We discuss several useful extensions to FD processes, namely, composite FD processes, autoregressive fractionally integrated moving average (ARFIMA) processes and time-varying FD processes. We then consider the statistical analysis of clock noise in terms of how these models are manifested via the spectral density function (SDF) and the wavelet variance (WV), the latter being a generalization of the well-known Allan variance. Both the SDF and the WV decompose the process variance with respect to an independent variable (Fourier frequency for the SDF, and scale for the WV); similarly, judiciously chosen estimators of the SDF and WV decompose the sample variance of clock noise. We give an overview of the statistical properties of estimators of the SDF and the WV and show how these estimators can be used to deduce the parameters of stochastic models for clock noise.

1 Introduction

The quality of a time scale formed from the intercomparison of a collection of high performance clocks is limited by the inherent instability ('noise') in the individual clocks. Quantification of the characteristics of clock noise is possible using the notion of a stochastic process, which by definition is a set of random variables (RVs) indexed by t (a variable representing the nominal time at which a measurement of clock noise is taken). We let $\{X_t\}$ denote the process itself, while X_t represents the clock noise at time t. We will concentrate on models for sampled data, for which we assume for convenience that $t \in \mathbb{Z} \equiv \{\ldots, -1, 0, 1, \ldots\}$ (i.e., the collection of all integers); however, we will find it handy in some cases to restrict t to just the nonnegative integers, i.e., $t \in \mathbb{Z}^* \equiv \{0, 1, 2, \ldots\}$.

Because the notion of a stochastic process is quite general, we need to consider carefully what classes of stochastic processes can serve as useful models for clock noise. We concentrate here on fractionally differenced (FD) processes, which we define and explore in §2. FD processes depend upon two parameters. Given a set of clock measurements, we can estimate these parameters and determine the adequacy of the FD process as a model by using two complementary analysis of variance ('power') techniques, namely, spectral analysis and wavelet analysis. These are discussed in, respectively, §3 and §4. We conclude with a brief summary in §5.

2 Fractionally Differenced and Related Processes

2.1 Background: Stationary Processes

Let us begin by reviewing the notion of a stationary process. Let $E\{X_t\}$ denote the expected value of the RV X_t, and let

$$\operatorname{cov}\{X_s, X_t\} \equiv E\{(X_s - E\{X_s\})(X_t - E\{X_t\})\}$$

denote the covariance between the RVs X_s and X_t. A stochastic process $\{X_t\}$ is said to stationary if

1. $E\{X_t\} = \mu_X$ for all t, where μ_X is a constant that does not depend on t, and
2. $\operatorname{cov}\{X_t, X_{t+\tau}\} = s_{X,\tau}$ for all possible t and $t + \tau$, where $s_{X,\tau}$ is a constant depending on the lag (time difference) between the RVs, but not on t.

The sequence $\{s_{X,\tau}\}$ indexed by the lag $\tau \in \mathbb{Z}$ is known as the autocovariance sequence (ACVS). Because

$$s_{X,0} = \operatorname{cov}\{X_t, X_t\} = E\{(X_t - \mu_X)^2\} \equiv \operatorname{var}\{X_t\},$$

where $\operatorname{var}\{X_t\}$ denotes the process variance, one consequence of stationarity is that the process variance is constant across time.

The stationary processes of interest here possess a spectral density function (SDF) that can be expressed as the Fourier transform of the ACVS; i.e.,

$$S_X(f) = \sum_{\tau=-\infty}^{\infty} s_{X,\tau} e^{-i2\pi f\tau}, \quad |f| \leq 1/2. \tag{1}$$

The SDF is a function of a variable f known as frequency (or, more precisely, Fourier frequency to avoid confusion with notions such as fractional frequency that arise in the clock literature). We take f to have units of cycles per some basic time interval (this is the interval in time between when we observe the value for, say, X_t and the one for X_{t+1}).

Although formally the SDF is defined over both positive and negative frequencies, in fact $S_X(-f) = S_X(f)$ when $\{X_t\}$ is a real-valued process (the negative frequencies are a convenient 'fiction' that is introduced to simplify certain mathematical expressions). If a stationary process $\{X_t\}$ does possess an SDF, then we can recover the ACVS from $S_X(\cdot)$ via an inverse Fourier transform, namely,

$$\int_{-1/2}^{1/2} S_X(f) e^{i2\pi f\tau}\, df = s_{X,\tau}, \quad \tau \in \mathbb{Z}.$$

If we set $\tau = 0$ in the above, we obtain the fundamental result that

$$\int_{-1/2}^{1/2} S_X(f)\, df = s_{X,0} = \text{var}\,\{X_t\}; \tag{2}$$

i.e., the SDF decomposes the process variance across frequencies f.

We can use the theory of linear time invariant filters to give us a complementary interpretation for the SDF. Let $\{a_u, u \in \mathbb{Z}\}$ denote a filter, and let

$$Y_t \equiv \sum_{u=-\infty}^{\infty} a_u X_{t-u}, \quad t \in \mathbb{Z}, \tag{3}$$

represent the output from this filter when its input is the stationary process $\{X_t\}$. Under a mild 'matching condition' on $\{a_u\}$ and $\{X_t\}$ to ensure that the output process has finite variance, the process $\{Y_t\}$ is stationary with an SDF given by

$$S_Y(f) = \mathcal{A}(f) S_X(f), \text{ where } \mathcal{A}(f) \equiv \left| \sum_{u=-\infty}^{\infty} a_u e^{-i2\pi f u} \right|^2. \tag{4}$$

The function $\mathcal{A}(\cdot)$ is known as the squared gain function for $\{a_u\}$. Suppose now that $\{a_u\}$ is a narrow-band filter of bandwidth Δf centered on some fixed frequency f; i.e.,

$$\mathcal{A}(f') = \begin{cases} 1/2\Delta f, & f - \Delta f/2 \le |f'| \le f + \Delta f/2 \\ 0, & \text{otherwise,} \end{cases}$$

We then have then have following interpretation for $S_X(f)$:

$$\text{var}\,\{Y_t\} = \int_{-1/2}^{1/2} S_Y(f')\, df' = \int_{-1/2}^{1/2} \mathcal{A}(f') S_X(f')\, df' \approx S_X(f); \tag{5}$$

i.e., the SDF $S_X(\cdot)$ at frequency f is approximately equal to the variance of the process formed by subjecting $\{X_t\}$ to a narrow-band filter centered at that frequency.

2.2 White Noise and Integer Differences of White Noise

The simplest example of a stationary process is white noise. By definition, the stochastic process $\{\epsilon_t, t \in \mathbb{Z}\}$ is said to be white noise if it satisfies the following three conditions:

1. $E\{\epsilon_t\} = \mu_\epsilon$ for all t, where μ_ϵ is some real-valued constant (often taken to be zero);

2. $\operatorname{var}\{\epsilon_t\} = \sigma_\epsilon^2$ for all t, where σ_ϵ^2 is some positive constant; and

3. $\operatorname{cov}\{\epsilon_s, \epsilon_t\} = 0$ for all $s \neq t$.

White noise thus satisfies the conditions of a stationary process, and its ACVS takes a very simple form, namely,

$$s_{\epsilon,\tau} = \operatorname{cov}\{\epsilon_t, \epsilon_{t+\tau}\} = \begin{cases} \sigma_\epsilon^2, & \tau = 0; \\ 0, & \text{otherwise.} \end{cases}$$

It follows from (1) that the SDF for white noise is given by

$$S_\epsilon(f) = \sum_{\tau=-\infty}^{\infty} s_{X,\tau} e^{-i2\pi f\tau} = \sigma_\epsilon^2;$$

i.e., the SDF is constant across all frequencies, with the constant being equal to the variance of the process.

Despite its simplicity, by suitably manipulating white noise, we can create entire classes of potentially interesting processes. As a first step in this direction, let us consider the first order backward difference of white noise, namely,

$$X_t = \epsilon_t - \epsilon_{t-1}.$$

The above can be expressed in terms of a linear filter by noting that

$$X_t = \sum_{u=-\infty}^{\infty} a_u \epsilon_{t-u} \text{ if we set } a_u \equiv \begin{cases} 1, & u = 0; \\ -1, & u = 1; \\ 0, & \text{otherwise} \end{cases}$$

(cf. (3)). It follows from (4) that the squared gain function for a backward difference filter is

$$\mathcal{A}(f) = \left|1 - e^{-i2\pi f}\right|^2 = \left|e^{-i\pi f}(e^{i\pi f} - e^{-i\pi f})\right|^2 = \left|(e^{i\pi f} - e^{-i\pi f})\right|^2 = |2\sin(\pi f)|^2,$$

where we have made use of the relationship $\sin(x) = (e^{ix} - e^{-ix})/2i$. Equation (4) says that the SDF for $\{X_t\}$ is given by

$$S_X(f) = \mathcal{A}(f)S_\epsilon(f) = |2\sin(\pi f)|^2\sigma_\epsilon^2. \tag{6}$$

Using the approximation $\sin(x) \approx x$ (valid for for small x), we see that $S_X(f) \approx |2\pi f|^2\sigma_\epsilon^2$ at low frequencies.

Define B to be the backward shift operator; i.e., $B\epsilon_t = \epsilon_{t-1}$, $B^2\epsilon_t = \epsilon_{t-2}$, $B^k\epsilon_t = \epsilon_{t-k}$ for any positive integer, and $(1 - B)\epsilon_t = \epsilon_t - \epsilon_{t-1}$. We can use B to help us define the dth order backward difference of white noise, where d is a positive integer:

$$\begin{aligned} X_t = (1-B)^d\epsilon_t &= \sum_{k=0}^{d} \binom{d}{k} (-B)^k \epsilon_t \\ &= \sum_{k=0}^{d} \frac{d!}{k!(d-k)!}(-1)^k\epsilon_{t-k} \\ &= \sum_{k=0}^{d} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}(-1)^k\epsilon_{t-k}, \end{aligned}$$

where $\Gamma(\cdot)$ is the gamma function, one of whose properties is that $\Gamma(k) = (k-1)!$ for all $k \in \mathbb{Z}^+ \equiv \{1, 2, \ldots\}$. Now the gamma function has poles at all the nonpositive integers, i.e., $\Gamma(k) = \infty$ for $k = 0, -1, -2, \ldots$. Hence

$$\frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}(-1)^k = 0 \text{ when } d-k+1 \leq 0, \text{ i.e., when } k \geq d+1.$$

This fact allows us to write

$$X_t = \sum_{k=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}(-1)^k \epsilon_{t-k}.$$

To be consistent with a convention in the literature, let us define $\delta \equiv -d$ (i.e., δ is a negative integer) so that the above becomes

$$X_t = \sum_{k=0}^{\infty} \frac{\Gamma(1-\delta)}{\Gamma(k+1)\Gamma(1-\delta-k)}(-1)^k \epsilon_{t-k}. \tag{7}$$

Since $(1-B)^d = (1-B)\cdots(1-B)$, a dth order backward difference can be expressed as d successive applications of a first order backward difference. It follows from (6) that the SDF for $\{X_t\}$ is given by

$$S_X(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^{2\delta}}. \tag{8}$$

At low frequencies we have $S_X(f) \approx \sigma_\epsilon^2/|2\pi f|^{2\delta}$.

2.3 Stationary Fractionally Differenced Processes

Granger and Joyeux [12] and Hosking [14] both show that (7) makes sense for all $\delta < 1/2$, i.e., not just for δ equal to a negative integer. When δ is not an integer, we can write

$$X_t = \sum_{k=0}^{\infty} a_k(\delta)\epsilon_{t-k} \text{ with } a_k(\delta) \equiv \frac{\Gamma(1-\delta)}{\Gamma(k+1)\Gamma(1-\delta-k)}(-1)^k \neq 0, \tag{9}$$

which means that the summation above truly involves an infinite number of RVs from the white noise process. The process $\{X_t\}$ is stationary and is called a fractionally differenced (FD) process. Its SDF has exactly the same form as before, namely, that of (8). We note that an FD process reduces to white noise when $\delta = 0$.

When $\delta < 1/2$ and $\delta \neq 0, -1, \ldots$, the ACVS for an FD process is given by

$$s_{X,\tau} = \sigma_\epsilon^2 \frac{\sin(\pi\delta)\Gamma(1-2\delta)\Gamma(\tau+\delta)}{\pi\Gamma(1+\tau-\delta)};$$

when $\delta = 0, -1, \ldots$, we have

$$s_{X,\tau} = \begin{cases} \sigma_\epsilon^2 \frac{(-1)^\tau \Gamma(1-2\delta)}{\Gamma(1+\tau-\delta)\Gamma(1-\tau-\delta)}, & 0 \leq |\tau| \leq -\delta \text{ and} \\ 0, & |\tau| > -\delta. \end{cases}$$

For all $\delta < 1/2$, we have

$$s_{X,0} = \operatorname{var}\{X_t\} = \sigma_\epsilon^2 \frac{\Gamma(1-2\delta)}{\Gamma^2(1-\delta)}, \tag{10}$$

and the remaining values in the ACVS can be computed easily via the recursion

$$s_{X,\tau} = s_{X,\tau-1}\frac{\tau+\delta-1}{\tau-\delta}, \quad \tau \in \mathbb{Z}^{+} \tag{11}$$

(for negative lags τ, recall that $s_{X,\tau} = s_{X,-\tau}$). Also, for all $\delta < 1/2$, the partial autocorrelation sequence (PACS) is given by

$$\phi_{t,t} \equiv \frac{\delta}{t-\delta}, \quad t \in \mathbb{Z}^{+}$$

(this sequence is of importance in, e.g., constructing best linear predictors of X_t based upon the t prior values X_{t-1}, X_{t-2}, $\ldots, X_0$, which in turn play a key role in exact maximum likelihood estimation of the FD parameters δ and σ_ϵ^2).

In summary, stationary FD processes have simple and easily computed expressions for the SDF, ACVS and PACS.

2.4 Simulation of Stationary Fractionally Differenced Processes

Models for clock noise are often used to generate simulated noise sequences on a digital computer for use in, e.g., evaluation of different time scale algorithms. We note here that it is possible to obtain 'exact' simulations of an FD process via a so-called 'circulant embedding' scheme (this is sometimes referred to in the statistical literature as the Davies–Harte algorithm) [7, 8, 25]. Suppose that we wish to generate a realization of a portion $X_0, X_1, \ldots, X_{N-1}$ of an FD process with parameters $-1 \le \delta < 1/2$ and $\sigma_\epsilon^2 > 0$. We begin by using (10) and (11) to compute $s_{X,0}, \ldots, s_{X,N}$. We then use the discrete Fourier transform (DFT) to compute

$$S_k \equiv \sum_{\tau=0}^{N} s_{X,\tau} e^{-i2\pi k\tau/2N} + \sum_{\tau=N+1}^{2N-1} s_{X,2N-\tau} e^{-i2\pi k\tau/2N}, \quad k = 0, \ldots, N; \tag{12}$$

i.e., the desired S_k values are the first $N+1$ values in the DFT of the following sequence of length $2N$:

$$s_{X,0}, \ldots, s_{X,N-2}, s_{X,N-1}, s_{X,N}, s_{X,N-1}, s_{X,N-2}, \ldots, s_{X,1}.$$

Given $2N$ independent Gaussian deviates ε_t with mean zero and variance σ_ϵ^2 (these can be readily simulated on a digital computer), we form the complex-valued sequence

$$\mathcal{Y}_k \equiv \begin{cases} \varepsilon_0\sqrt{2NS_0}, & k = 0; \\ (\varepsilon_{2k-1} + i\varepsilon_{2k})\sqrt{NS_k}, & 1 \le k < N; \\ \varepsilon_{2N-1}\sqrt{2NS_N}, & k = N; \text{ and} \\ \mathcal{Y}^*_{2N-k}, & N < k \le 2N-1 \end{cases}$$

(in the above, the asterisk denotes a complex conjugate). Finally we use the inverse DFT to construct the real-valued sequence

$$Y_t = \frac{1}{2N}\sum_{k=0}^{2N-1} \mathcal{Y}_k e^{i2\pi kt/2N}.$$

The first N values of this sequence, i.e., $Y_0, Y_1, \ldots, Y_{N-1}$, constitute an exact simulation of X_0, X_1, $\ldots$, X_{N-1} in the sense that the multivariate Gaussian distributions of both sequences are identically the same.

The circulant embedding scheme works for a wide variety of – but not all – stationary processes. The scheme is guaranteed to work as long as all the S_k in (12) are nonnegative. Results in [5, 11, 19] prove that (i) $S_k \geq 0$ when $-1 \leq \delta < 1/2$ and (ii) this nonnegativity condition can fail when $\delta < -1$ for certain sample sizes. We note that, since an FD process with $\delta < -1$ can be expressed as a certain number of backward differences of an FD process with $-1/2 \leq \delta < 1/2$, we can in fact easily generate any stationary FD process.

Finally, an interesting implication of the circulant embedding scheme is that we can represent $X_0, X_1, \ldots, X_{N-1}$ as

$$X_t = \sum_{k=0}^{2N-1} c_{t,k}(\delta)\varepsilon_k \text{ rather than } X_t = \sum_{k=0}^{\infty} a_k(\delta)\epsilon_{t-k}; \tag{13}$$

i.e., even though an FD process is formally defined in terms of an infinite weighted moving average of a white noise sequence, in fact we can represent a finite portion of length N in terms of a weighted sum of $2N$ deviates from another white noise sequence.

2.5 Nonstationary Fractionally Differenced Processes

Here we extend the definition of an FD process to include all $\delta \geq 1/2$, leading to an interesting class on nonstationary processes that can generate realistic realizations of various kinds of clock noise. To make this extension, suppose that $X_t^{(1)}$ is an FD process with parameter $\delta^{(s)}$ such that $-1/2 \leq \delta^{(s)} < 1/2$. Let us define X_t as the cumulative sum of $X_t^{(1)}$; i.e.,

$$X_t \equiv \sum_{l=0}^{t} X_l^{(1)}, \;\; t \in \mathbb{Z}^*$$

(for $l < 0$, let $X_t \equiv 0$). It can be argued that the process $\{X_t\}$ is nonstationary because its variance is not invariant over time, but rather increases unboundedly with time. Since, for $t \in \mathbb{Z}^*$,

$$X_t^{(1)} = X_t - X_{t-1} \text{ and } S_{X^{(1)}}(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^{2\delta^{(s)}}},$$

filtering theory suggests using the relationship

$$S_{X^{(1)}}(f) = |2\sin(\pi f)|^2 S_X(f)$$

to *define* an SDF $S_X(\cdot)$ for $\{X_t\}$; i.e.,

$$S_X(f) \equiv \frac{S_{X^{(1)}}(f)}{|2\sin(\pi f)|^2} = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^{2\delta}}, \text{ where } \delta \equiv \delta^{(s)} + 1.$$

Note that this SDF has the same functional form as the SDF for a stationary FD process (see (8)). Yaglom [26] shows that, although the integral of $S_X(\cdot)$ from $-1/2$ to $1/2$ is infinite, the function $S_X(\cdot)$ has many of the properties of SDFs associated with stationary processes, including the bandpass interpretation of (5).

By cumulatively summing a stationary FD process with parameter $\delta^{(s)}$ satisfying $-1/2 \leq \delta^{(s)} < 1/2$, we define a nonstationary FD process with parameter δ satisfying $1/2 \leq \delta < 3/2$. If we now take this nonstationary process and in turn cumulatively sum it, we obtain another nonstationary FD process, this time with a δ satisfying $3/2 \leq \delta < 5/2$. By continuing in this manner, we can define a nonstationary FD process for any $\delta \geq 1/2$.

Let us consider four special cases of nonstationary FD processes whose backward differences of a certain order are stationary FD processes. Suppose first that $X_t^{(1)}$ is white noise (i.e., $\delta^{(s)} = 0$) so $S_{X^{(1)}}(f) = \sigma_\epsilon^2$. If we cumulatively sum it, the resulting process $\{X_t\}$ is called a random walk, which is associated with $\delta = 1$. Its SDF is given by

$$S_X(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^2} \approx \frac{\sigma_\epsilon^2}{|2\pi f|^2} \text{ at low frequencies.}$$

If $X_t^{(2)}$ is white noise and if

$$X_t^{(1)} \equiv \sum_{l=0}^{t} X_l^{(2)} \text{ and } X_t \equiv \sum_{l=0}^{t} X_l^{(1)}, \quad t \in \mathbb{Z}^*, \tag{14}$$

then X_t is called a random run. A random run is associated with $\delta = 2$ and has an SDF given by

$$S_X(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^4} \approx \frac{\sigma_\epsilon^2}{|2\pi f|^4} \text{ at low frequencies.}$$

Now suppose that $X_t^{(1)}$ is an FD process with $\delta^{(s)} = -1/2$. If we cumulatively sum it, we obtain an FD process that is associated with $\delta = 1/2$ and that has an SDF given by

$$S_X(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|} \approx \frac{\sigma_\epsilon^2}{|2\pi f|} \text{ at low frequencies.}$$

Clock noise with an SDF that varies as $1/|f|$ is often called flicker phase noise. Similarly, if $X_t^{(2)}$ is an FD process with $\delta^{(s)} = -1/2$ and if X_t is formed as per (14), then X_t is an FD process with $\delta = 3/2$ and has an SDF given by

$$S_X(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^3} \approx \frac{\sigma_\epsilon^2}{|2\pi f|^3} \text{ at low frequencies.}$$

Clock noise with an SDF varying as $1/|f|^3$ is known as flicker frequency noise.

Since a nonstationary FD process can be formed by cumulatively summing a stationary FD process a certain number of times, we can easily simulate the former via cummulative summations of a simulation of the latter generated using the technique described in §2.4.

2.6 Alternatives to Fractionally Differenced Processes

One of the key properties of FD processes is that their SDFs obey a power law at low frequencies. Formally we say that a process $\{X_t\}$ obeys a power law at low frequencies if it possesses an SDF such that

$$\lim_{f \to 0} \frac{S_X(f)}{C|f|^\alpha} = 1$$

for some constant $C > 0$; i.e., $S_X(f) \approx C|f|^\alpha$ at low frequencies. It follows from (8) that FD processes obey the above with $\alpha = -2\delta$. Several other processes have been considered in the literature that obey power laws at low frequencies. It is instructive to compare and contrast these to FD processes.

Let us first consider fractional Brownian motion (FBM), which we denote by $B_H(t), 0 \le t < \infty$, where here t ranges over all nonnegative numbers (rather than just the integers). The SDF for an FBM is given by

$$S_{B_H(t)}(f) = \frac{\sigma_X^2 C_H}{|f|^{2H+1}}, \quad -\infty < f < \infty,$$

where $\sigma_X^2 > 0$, $C_H > 0$ and $0 < H < 1$. The parameter H is called the Hurst parameter (or Hurst coefficient), while $\sigma_X^2 > 0$ and $C_H > 0$ are, respectively, a variance for an associated fractional Gaussian noise (see below) and a constant that depends upon H. An FBM obeys a power law over all frequencies with $\alpha = -(2H+1)$, which means that $-3 < \alpha < -1$. If we sample an FBM at the nonnegative integers, we obtain a discrete fractional Brownian motion (DFBM); i.e., $\{B_t, t \in \mathbb{Z}^*\}$ is DFBM if $B_t = B_H(t)$. A DFBM has an SDF given by

$$S_{B_t}(f) = \sigma_X^2 C_H \sum_{j=-\infty}^{\infty} \frac{1}{|f+j|^{2H+1}}, \quad |f| \le 1/2$$

(see [23]). We can argue that a DFBM obeys a power law at low frequencies also with $-3 < \alpha < -1$, so it is similar to an FD process with $1/2 < \delta < 3/2$. We note that a DFBM reduces to a random walk process if $H = 1/2$.

A process closely related to DFBM is fractional Gaussian noise (FGN) [15]. An FGN $\{X_t, t \in \mathbb{Z}^*\}$ is defined to be the first forward difference of DFBM; i.e., $X_t = B_{t+1} - B_t$. The SDF for an FGN is given by

$$S_X(f) = 4\,\sigma_X^2 C_H \sin^2(\pi f) \sum_{j=-\infty}^{\infty} \frac{1}{|f+j|^{2H+1}}, \quad |f| \le 1/2.$$

The above obeys a power law at low frequencies with $\alpha = -(2H-1)$, which means that $-1 < \alpha < 1$. An FGN $\{X_t\}$ is in fact a stationary process, with an ACVS given by

$$s_{X,\tau} = \frac{\sigma_X^2}{2}\left(|\tau+1|^{2H} - 2|\tau|^{2H} + |\tau-1|^{2H}\right), \quad \tau \in \mathbb{Z},$$

where we see that $s_{X,0} = \sigma_X^2 = \mathrm{var}\,\{X_t\}$. When $H = 1/2$, an FGN reduces to white noise.

The final process of interest to us is a discrete pure power law (PPL) process, which is defined to have an SDF given by

$$S_X(f) = C_S |f|^{\alpha}, \;\; |f| \le 1/2,$$

for some constant $C_S > 0$. This SDF obviously obeys a power law at low frequencies. If $\alpha > -1$, a PPL process is stationary, but its ACVS takes some effort to compute; if $\alpha = 0$, the process reduces to white noise; and if $\alpha \le -1$, the process is nonstationary, but its backward differences of a certain order are stationary.

Let us now note the following five ways in which FD processes compare and contrast with FBMs, DFBMs, FGNs and PPL processes.

1. All five processes obey power laws at low frequencies, but, whereas FD and PPL processes cover the full range of possible power laws (i.e., $-\infty < \alpha < \infty$), FBMs, DFBMs and FGNs cover limited ranges (FBMs and DFBMs carry the restriction $-3 < \alpha < -1$, while FGNs are such that $-1 < \alpha < 1$).

2. If we difference an FD process, we obtain another FD process. For the other four processes, differencing yields a new type of process. Because differencing is such an important operation in time series analysis, the fact that differencing an FD process yields another FD process is quite convenient in describing the effect of differencing.

3. An FD process has a simple SDF and, if stationary, a simple ACVS and PACS. By contrast, FBM has a simple SDF, but DFBM has a complicated SDF involving an infinite summation (since both of these processes are nonstationary, formally they do not possess an ACVS or PACS). The form of the ACVS for an FGN is simple, but its SDF and PACS have complicated expressions. Finally, we note that a PPL process has a simple SDF, but, when stationary, its ACVS and PACS are complicated.

4. FD processes, DFBMs, FGNs and PPL processes all model sampled noise, which can be problematic if the sampling rate at which the process is observed is subject to change. Only FBM models unsampled noise.

Although FGNs, PPL processes and FD processes are formally different models, they are actually quite close to each other when their parameters are set such that they obey the same power law at low frequencies. Figures 1 and 2 demonstrate two aspects of this closeness. In Figure 1, we show log/log plots of the SDFs for two specific examples of the three processes. The left-hand, middle and right-hand plots show SDFs for, respectively, FGNs, PPL processes and FD processes. The two examples for FGNs have Hurst parameters of $H = 0.6$ (thick curve) and 0.9 (thin), which, using the relationship $\alpha = -(2H - 1)$, yield power laws at low frequencies with, respectively, $\alpha = -0.2$ and -0.8. The statement $S_X(f) \approx C|f|^\alpha$ implies that $\log(S_X(f)) \approx \log(C) + \alpha \log(f)$ at positive low frequencies, so the exponent α dictates the slope of the approximate linear behavior that appears in the plots of $S_X(f)$ versus f on log/log axes. The middle plot shows the two PPL processes, with α set to the two values previously mentioned. The right-hand plot gives corresponding examples for FD processes, with δ set via the relationship $\delta = -\alpha/2$. From these plots, we see that the power law approximation is evidently quite good when $|f| < 1/8$ or so for FGNs and FD processes. We can also see that the SDFs for corresponding models differ only by a small amount at high frequencies.

Figure 2 shows plots of realizations for the six processes whose SDFs are shown in Figure 1. The top three plots show, respectively, an FGN with $H = 0.6$, a PPL process with $\alpha = -0.2$ and an FD process with $\delta = 0.1$, while the bottom three plots show the cases $H = 0.9$, $\alpha = -0.8$ and $\delta = 0.4$. Each realization is of length $N = 512$ and was created using the circulant embedding method described in §2.4. This method forms a particular realization by transforming a set of $2N = 1024$ independent deviates $\varepsilon_0, \dots, \varepsilon_{2N-1}$ from a standard Gaussian distribution. For all six realizations plotted in Figure 2, we used the *same* set of 1024 deviates. The differences amongst comparable realizations (i.e., the top three plots or the bottom three) are thus due to the underlying processes rather than to the use of different sets of Gaussian deviates. Visually the three plots for realizations with comparable parameters are strikingly similar, emphasizing our contention that FGNs, PPL processes and FD processes are quite similar statistical models (estimates of the SDFs for the six realizations do in fact show high frequency differences consistent with those shown in Figure 1).

2.7 Extensions to Fractionally Differenced Processes

An FD process is completely characterized by just two parameters (δ and σ_ϵ^2) and hence is a quite simple statistical model. While some clock noise can be successfully modeled by this simple process, usually more complex models are needed to capture the salient features of clock noise. Here we consider three extensions to basic FD processes that can provide more realistic models.

1) **Composite fractionally differenced processes**

Suppose that, for $m = 0, \ldots, M-1$, $\{X_{m,t}\}$ is an FD process with parameters δ_m and σ_m^2. Suppose the processes $\{X_{m,t}\}$ and $\{X_{m',t}\}$ are independent of each other when $m \neq m'$ (this means that $X_{m,t}$ and $X_{m',t'}$ are independent for all t and t'). We define a composite FD process as

$$X_t = \sum_{m=0}^{M-1} X_{m,t}.$$

This process will have an SDF given by

$$S_X(f) = \sum_{m=1}^{M} \frac{\sigma_m^2}{|2\sin(\pi f)|^{2\delta_m}}.$$

A simple example of a composite FD process that has been used to model phase noise consists of $M = 2$ components, namely, a random run process ($\delta_1 = 2$) and a random walk process ($\delta_2 = 1$). A realization from such a process is shown in the upper right-hand plot of Figure 5, and its corresponding SDF is shown as the thick smooth curves in Figures 4 and 6 (for this example, we set $\sigma_1^2 = 0.0001$ and $\sigma_2^2 = 0.0064$).

2) **Autoregressive fractionally integrated moving average processes**

As is evident from the discussion in §2.3 and §2.5, all stationary and nonstationary FD processes are ultimately based upon (possibly infinite) linear combinations of a white noise process $\{\epsilon_t\}$. One simple way of extending an FD process is to replace the white noise process in this construction with a stationary autoregressive moving average (ARMA) process. By definition, an ARMA process takes the form

$$U_t = \sum_{k=1}^{p} \phi_k U_{t-k} + \epsilon_t - \sum_{k=1}^{q} \theta_k \epsilon_{t-k}$$

and is stationary if we suitably choose the ϕ_k parameters (see, e.g., [4]). With this replacement, the resulting process is called an autoregressive fractionally integrated moving average (ARFIMA) process (for details, see Beran [3], where these processes are referred to as FARIMA processes). An ARFIMA process has an SDF given by

$$S_X(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^{2\delta}} \cdot \frac{\left|1 - \sum_{k=1}^{q} \theta_k e^{-i2\pi f k}\right|^2}{\left|1 - \sum_{k=1}^{p} \phi_k e^{-i2\pi f k}\right|^2}.$$

The introduction of the ARMA part gives us more flexibility in modeling the high-frequency structure of clock noise.

3) **Time-varying fractionally differenced processes**

Since an FD process is either stationary or has stationary backward differences of a certain order, either all or some of its pertinent statistical properties are invariant over time. As a clock ages, however, its statistical properties can change. It is possible to model these changes by introducing the notion of a time-varying FD (TVFD) process. In the case of stationary FD processes, this can be done by letting δ in (9) depend upon t, thus yielding the TVFD process

$$X_t = \sum_{k=0}^{\infty} a_k(\delta_t)\epsilon_{t-k},$$

which is well-defined as long as $\delta_t < 1/2$ for all t. When dealing with a finite portion X_0, X_1, ..., X_{N-1}, we can also modify the alternative representation given in (13) to obtain

$$X_t = \sum_{k=0}^{2N-1} c_{t,k}(\delta_t)\varepsilon_k, \quad t = 0, 1, \ldots, N-1,$$

which is particularly useful for generating simulations of TVFDs. The above ideas can be extended to allow δ_t to vary in an arbitrary manner (i.e., we can remove the restriction $\delta_t < 1/2$) and to make σ_ϵ^2 a function of t also (for details, see [19]).

As an example, Figure 3 shows realizations of four different TVFDs, along with plots of δ_t versus t. In plot (a), δ_t starts off close to what is appropriate for white noise ($\delta_t = 0$) and ends up close to a process having a power law at low frequencies with $\alpha = -1$ ($\delta_t = 1/2$). The change in the spectral properties of the process is evident in the plotted realization. In plot (b), the parameter δ_t is constant and just below 1/2 for the first half of the time series, after which it switches to another constant – this time just above 1/2 – for the second half. Although the TVFD process is thus stationary over the first half and nonstationary over the second half (but with stationary first differences), there is no visible qualitative difference between the two halves of the series. In plot (c), we show an example of a TVFD process that starts off as a nonstationary FD process with δ close to 1/2 and then transitions linearly almost up to a random walk process ($\delta = 1$). In plot (d), the TVFD process starts off with $\delta_t = 0.4$ (a stationary FD process). It then linearly ramps up to a nonstationary FD process with $\delta = 3/4$, after which it linearly ramps down to a white noise process ($\delta = 0$) and then back up to a process close to the stationary/nonstationary boundary of $\delta = 1/2$. While this last realization does not resemble actual clock noise, it does demonstrate the potential for evolutionary changes in TVFDs.

2.8 Summary of Fractionally Differenced Processes

Let us briefly summarize some of the key results discussed above for FD processes. The stochastic process $\{X_t\}$ is said to be an FD process if its SDF is given by

$$S_X(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^{2\delta}} \approx \frac{\sigma_\epsilon^2}{|2\pi f|^{2\delta}} \text{ at low frequencies.}$$

An FD process depends upon two parameters, namely, δ (which can assume any real value) and σ_ϵ^2 (a positive value that is the variance of the white noise process used to construct the FD process). At low frequencies, an FD process obeys a power law with exponent $\alpha = -2\delta$. If $\delta < 1/2$, the FD process is stationary with an ACVS given by

$$s_{X,\tau} = \begin{cases} \sigma_\epsilon^2 \frac{\sin(\pi\delta)\Gamma(1-2\delta)\Gamma(\tau+\delta)}{\pi\Gamma(1+\tau-\delta)}, & \text{when } \delta \text{ is not an integer;} \\ \sigma_\epsilon^2 \frac{(-1)^\tau\Gamma(1-2\delta)}{\Gamma(1+\tau-\delta)\Gamma(1-\tau-\delta)}, & \text{when } \delta \text{ is an integer and } 0 \le |\tau| \le -\delta; \text{ and} \\ 0, & \text{when } \delta \text{ is an integer and } |\tau| > -\delta, \end{cases}$$

and a PACS given by

$$\phi_{t,t} \equiv \frac{\delta}{t-\delta}, \quad t \in \mathbb{Z}^+.$$

Expressions for the ACVS that are convenient for computational purposes and that are valid for both integer and noninteger δ are

$$s_{X,0} = \sigma_\epsilon^2 \frac{\Gamma(1-2\delta)}{\Gamma^2(1-\delta)} \text{ and } s_{X,\tau} = s_{X,\tau-1}\frac{\tau+\delta-1}{\tau-\delta}, \quad \tau \in \mathbb{Z}^+.$$

If $\delta \geq 1/2$, the FD process is nonstationary but its dth order backward difference is a stationary FD process with parameter $\delta^{(s)}$, where

$$d \equiv \lfloor \delta + 1/2 \rfloor \text{ and } \delta^{(s)} \equiv \delta - d$$

(here $\lfloor x \rfloor$ is the largest integer less than or equal to x).

Stationary FD processes can be readily simulated using the circulant embedding scheme described in §2.4, while realizations of nonstationary FD processes can be generated by suitable cumulative summations of realizations of stationary FD processes.

3 Spectral Analysis of Fractionally Differenced Processes

Suppose now that we have some clock noise and that we regard it as a realization of a portion $X_0, \ldots, X_{N-1}$ of an FD process with unknown parameters δ and σ_ϵ^2. Given the available data, how should we go about estimating these parameters? Many different estimators have been proposed for this task (see, e.g., [3]), and the development of new estimators and the statistical analysis of previously proposed estimators are areas of active research. In this section, we concentrate on estimators based upon spectral analysis (i.e., frequency-based estimators), while we devote the next section (§4) to estimators based upon wavelet analysis (i.e., scale-based estimators). These estimators have the following advantages over certain other techniques that have been proposed in the literature.

1. Both spectral and wavelet analysis are analysis of variance techniques, so they are useful for much more than just estimating δ and σ_ϵ^2.

2. Both techniques can help assess the need for models that are more complex than a simple FD process (e.g., a composite FD process).

3. Both techniques can provide preliminary estimates for more complicated schemes that depend upon numerical optimization (an example is exact maximum likelihood estimation).

4. The basic components of both techniques have appealing physical interpretations that directly relate to fundamental notions about the characterization of clock noise.

The use of spectral analysis to estimate δ and σ_ϵ^2 has its basis in what happens when we take the log of both sides of the expression for the SDF of an FD process given in (8). For $0 < f < 1/2$, we obtain

$$\log(S_X(f)) = \log(\sigma_\epsilon^2) - 2\delta \log(|2\sin(\pi f)|),$$

which says that a plot of $\log(S_X(f))$ versus $\log(|2\sin(\pi f)|)$ is linear with a slope of $\alpha = -2\delta$. As we have previously noted, the approximation $\sin(\pi f) \approx \pi f$ is quite good for $0 < f < 1/8$ or so. This fact implies that

$$\log(S_X(f)) \approx \log(\sigma_\epsilon^2) - 2\delta \log(2\pi f);$$

i.e., a plot of $\log(S_X(f))$ versus $\log(2\pi f)$ is approximately linear at low frequencies with the same slope as before.

Based upon the above log/log relationships, the basic scheme for estimating δ and σ_ϵ^2 via spectral analysis consists of the following four steps.

1. Using $X_0, \ldots, X_{N-1}$, we estimate $S_X(f)$ via some SDF estimator, say $\hat{S}_X(f)$.

2. We then fit a linear model to $\log(\hat{S}_X(f))$ versus $\log(2\pi f)$ over low frequencies to obtain an estimated slope and an estimated intercept. The simplest fitting procedures are variations on least squares.

3. We then use the estimated slope $\hat{\alpha}$ to estimate δ via $-\hat{\alpha}/2$.

4. Finally we use the antilog of the estimated intercept to estimate σ_ϵ^2.

In what follows, we consider the above scheme using the periodogram in §3.1 and multitaper SDF estimators in §3.2.

3.1 Spectral Analysis based upon the Periodogram

The most basic estimator of $S(f)$ is the periodogram, which is given by

$$\hat{S}^{(p)}(f) \equiv \frac{1}{N}\left|\sum_{t=0}^{N-1} X_t e^{-i2\pi f t}\right|^2, \quad |f| \le 1/2$$

(for convenience, we assume $E\{X_t\} = 0$; when this is not the case, as is true in most practical applications, we need to replace X_t in the above with $X_t - \overline{X}$, where $\overline{X}$ is the sample mean of $X_0, \ldots, X_{N-1}$). Just as the SDF gives a decomposition of the variance of a stationary process, so does the periodogram give a decomposition of the sample variance:

$$\int_{-1/2}^{1/2} \hat{S}^{(p)}(f)\, df = \frac{1}{N}\sum_{t=0}^{N-1} X_t^2. \tag{15}$$

For stationary processes and for large sample sizes N, theory suggests that

$$\hat{S}^{(p)}(f) \stackrel{\mathrm{d}}{=} S(f)\chi_2^2/2, \;\; 0 < f < 1/2, \tag{16}$$

approximately; in words, $\hat{S}^{(p)}(f)$ is approximately equal in distribution to a chi-square RV (denoted by χ_2^2) with two degrees of freedom after this RV has been multiplied by $S(f)/2$. Since $E\{\chi_2^2\} = 2$ and $\operatorname{var}\{\chi_2^2\} = 4$, it follows that $E\{\hat{S}^{(p)}(f)\} \approx E\{S(f)\chi_2^2/2\} = S(f)$ and $\operatorname{var}\{\hat{S}^{(p)}(f)\} \approx \operatorname{var}\{S(f)\chi_2^2/2\} = S^2(f)$. Additionally, theory suggests that, at Fourier frequencies $f_j \equiv j/N$ and $f_k \equiv k/N$ satisfying $0 < f_j < f_k < 1/2$, we have $\operatorname{cov}\{\hat{S}^{(p)}(f_j), \hat{S}^{(p)}(f_k)\} \approx 0$.

If we apply a log transform to both sides of (16), we obtain

$$\log(\hat{S}^{(p)}(f)) \stackrel{\mathrm{d}}{=} \log\left(S(f)\chi_2^2/2\right) = \log(S(f)) + \log\left(\chi_2^2/2\right).$$

Let χ_η^2 represent a chi-square RV with η degrees of freedom. Bartlett and Kendall [2] determined that the first moment and variance of a $\log(\chi_\eta^2/\eta)$ RV are given by

$$E\left\{\log\left(\chi_\eta^2/\eta\right)\right\} = \psi(\eta/2) - \log(\eta/2) \;\text{ and }\; \operatorname{var}\left\{\log\left(\chi_\eta^2/\eta\right)\right\} = \psi'(\eta/2),$$

where $\psi(\cdot)$ and $\psi'(\cdot)$ are the di– and trigamma functions. Their results imply that

$$E\{\log(\hat{S}^{(p)}(f))\} = \log(S(f)) + \psi(1) - \log(1) = \log(S(f)) - \gamma,$$

where $\gamma \doteq 0.57721$ is Euler's constant, and

$$\operatorname{var}\{\log(\hat{S}^{(p)}(f))\} = \psi'(1) = \pi^2/6.$$

In light of the above, let us define

$$Y^{(p)}(f_j) = \log(\hat{S}^{(p)}(f_j)) + \gamma \text{ and } \epsilon(f_j) = Y^{(p)}(f_j) - \log(S(f_j)).$$

It follows that

$$Y^{(p)}(f_j) = \log(S(f_j)) + \epsilon(f_j) \approx \log(\sigma_\epsilon^2) - 2\delta\log(2\pi f_j) + \epsilon(f_j) \tag{17}$$

over low frequencies indexed by, say, $1 \le j \le J$. The above is a simple linear regression model with dependent variables $Y^{(p)}(f_j)$, intercept $\log(\sigma_\epsilon^2)$, slope -2δ, independent variables $\log(2\pi f_j)$ and error terms $\epsilon(f_j)$ having the following approximate statistical properties:

1. $E\{\epsilon(f_j)\} = 0$;
2. $\mathrm{var}\{\epsilon(f_j)\} = \pi^2/6$, a known constant;
3. $\mathrm{cov}\{\epsilon(f_j), \epsilon(f_k)\} \approx 0$ when $j \ne k$ under the additional assumption that $\{X_t\}$ obeys a Gaussian distribution; and
4. $\epsilon(f_j) \stackrel{\mathrm{d}}{=} \log(\chi_2^2) + \gamma - \log(2)$.

The properties of the errors $\epsilon(f_j)$ are compatible with standard regression models with the exception that their distribution is markedly non-Gaussian. Assuming this model, we can use a least square procedure to obtain estimates $\hat{\delta}$ and $\hat{\sigma}_\epsilon^2$ for δ and σ_ϵ^2, along with estimates of the variability in $\hat{\delta}$ and $\hat{\sigma}_\epsilon^2$ [9, 10, 16, 17].

3.2 Spectral Analysis based upon Multitaper Spectral Estimators

While the periodogram-based procedure works reasonably well for stationary FD processes, the approximations to its statistical properties quoted above are of questionable validity for nonstationary FD processes. As an example of the potential problems that can arise, Figure 4 illustrates that this estimator can suffer from a phenomenon called 'leakage,' resulting in unacceptably large biases. The thick curve in plot (a) shows the SDF for the nonstationary composite FD process described in § 1). The thin curve in that plot shows the periodogram based upon a time series of length $N = 512$ generated from this composite process (the series itself is shown in the upper right-hand plot of Figure 5). There are substantial biases in the periodogram.

One well-known technique for alleviating leakage is tapering, which leads to an SDF estimator of the following form, known as a direct spectral estimator:

$$\hat{S}^{(d)}(f) \equiv \left| \sum_{t=0}^{N-1} a_t X_t e^{-i2\pi f t} \right|^2.$$

In the above, $\{a_t\}$ is called a data taper and usually resembles a bell-shaped curve. Plots (b) and (c) in Figure 4 show two direct spectral estimates based upon two different data tapers. The taper used in (b) leads to an SDF estimate that is less biased than the periodogram, but still suffers from substantial bias, particularly at high frequencies. The taper used in (c) yields a direct spectral estimate that is essentially bias free. The idea behind applying a data taper is to force the beginning and end of the tapered time series to be close to each other. One interpretation of leakage is that it is due to a mismatch between a time series and an implicit assumption in Fourier analysis, namely, that the finite sequence $X_0, \ldots, X_{N-1}$ is part of a periodic infinite sequence with a period of N. If $|X_0 - X_{N-1}|$ is large compared to typical values for $|X_t - X_{t-1}|$, $t = 1, \ldots, N-1$, then Fourier methods can have a hard time dealing with the discontinuity between X_0 and X_{N-1}, leading to leakage in SDF estimates. Tapering in effect forces a match between the beginning and end of the series.

A critique of tapering is that we are effectively decreasing the sample size N of the time series that we are dealing with; i.e., by tapering, we lose certain 'information' contained at the beginning and end of the series so that N is in effect shortened. Thomson [24] proposed a scheme for recovering this lost information called multitapering that, as its name implies, uses several different data tapers. The idea behind multitapering is to use a set of K orthonormal data tapers $\{a_{k,t}\}$, $k = 0, \ldots, K-1$, where the orthonormality condition states that

$$\sum_{t=0}^{N-1} a_{k,t} a_{l,t} = \begin{cases} 1, & \text{if } k = l; \\ 0, & \text{if } k \neq l. \end{cases} \quad 0 \leq k, l \leq K-1.$$

Each of these tapers is used to form a direct spectral estimator:

$$\hat{S}_k^{(d)}(f) \equiv \left| \sum_{t=0}^{N-1} a_{k,t} X_t e^{-i2\pi ft} \right|^2, \ \ k = 0, \ldots, K-1.$$

We then average all K such estimators together to obtain the simplest form of a multitaper SDF estimator, namely,

$$\hat{S}_K^{(mt)}(f) \equiv \frac{1}{K} \sum_{k=0}^{K-1} \hat{S}_k^{(d)}(f).$$

Thomson [24] proposed using tapers based upon discrete prolate spheroidal sequences (now commonly referred to as Slepian sequences). These require specialized software in order to compute (for details, see Chapter 8 of [20]). In subsequent work, Riedel and Sidorenko [22] proposed a 'minimum bias' multitapering scheme based upon an easily computed family of sinusoidal tapers. These tapers are given by

$$a_{k,t} = \left\{ \frac{2}{(N+1)} \right\}^{1/2} \sin \left\{ \frac{(k+1)\pi(t+1)}{N+1} \right\}, \quad t = 0, \ldots, N-1.$$

Figures 5 and 6 show an example of multitapering using the sinusoidal tapers. The left-hand column of Figure 5 shows tapers of orders $k = 0, 1, 2$ and 3 that are appropriate for a time series of length $N = 512$. The top plot in the right-hand column shows a series of this length simulated from the composite FD process with two components described in § 1). The plots below this one show the effect of multiplying the various sinusoidal tapers and the time series together to form the tapered series $a_{k,0}X_0, \ldots a_{k,N-1}X_{N-1}$. The jagged curves in the top row of Figure 6 show the direct spectral estimates $\hat{S}_k^{(d)}(\cdot)$ based upon the kth order sinusoidal tapers, $k = 0, 1, 2$ and 3 (left- to right-hand plots). The smooth thicker curve in each plot shows the true SDF for the composite FD process. The bottom row shows multitaper SDF estimates $\hat{S}_K^{(mt)}(\cdot)$ based upon an average of K direct spectral estimates, where $K = 1, 2, 3$ and 4 (left to right). Note that, as K increases, the variability in the resulting multitaper estimate decreases. This fact can be explained theoretically from the result that, if $S(\cdot)$ is slowly varying around $S(f)$ and if N is large, then

$$\hat{S}_K^{(mt)}(f) \stackrel{\mathrm{d}}{=} S(f)\chi_{2K}^2/2K$$

approximately for $0 < f < 1/2$. Since $\text{var}\{\chi^2_\eta\} = 2\eta$, we have

$$\text{var}\{\hat{S}^{(mt)}_K(f)\} \approx S^2(f)\,\text{var}\{\chi^2_{2K}\}/4K^2 = S^2(f)/K.$$

In light of the above, we can now set up a linear regression model that closely parallels the model of (17) but is based upon the multitaper estimator. Let us define

$$Y^{(mt)}(f_j) = \log(\hat{S}^{(mt)}_K(f_j)) - \psi(K) + \log(K) \text{ and } \zeta(f_j) = Y^{(mt)}(f_j) - \log(S(f_j)).$$

We can model $Y^{(mt)}(f_j)$ as

$$Y^{(mt)}(f_j) = \log(S(f_j)) + \zeta(f_j) \approx \log(\sigma^2_\epsilon) - 2\delta\log(2\pi f_j) + \zeta(f_j) \tag{18}$$

over low frequencies. As in the case of the periodogram, the above is a simple linear regression model, but now the errors $\zeta(f_j)$ have the following statistical properties:

1. $E\{\zeta(f_j)\} = 0$;
2. $\text{var}\{\zeta(f_j)\} = \psi'(K)$, a known constant (as before, $\psi'(\cdot)$ is the trigamma function);
3. the errors are now correlated, but the form of the correlation takes the following simple structure:

$$\text{cov}\{\zeta(f_j), \zeta(f_{j+\nu})\} \approx \begin{cases} \psi'(K)\left(1 - \frac{|\nu|}{K+1}\right), & \text{if } |\nu| \le K+1; \\ 0, & \text{otherwise; and} \end{cases}$$

4. $\zeta(f_j)$ approximately obeys a Gaussian distribution when $K \ge 5$.

We can compensate for the fact that the errors are now correlated by using a generalized least squares procedure [9]. This procedure yields estimates $\hat{\delta}$ and $\hat{\sigma}^2_\epsilon$ for δ and σ^2_ϵ, along with estimates of the variability in $\hat{\delta}$ and $\hat{\sigma}^2_\epsilon$. Monte Carlo studies indicate that the estimates of δ and σ^2_ϵ obtained from the multitaper approach are superior to those obtained from the periodogram approach (see [16] for details).

4 Wavelet Analysis of Fractionally Differenced Processes

In addition to spectral analysis, another important method for estimating the parameters of an FD processes and for assessing the characteristics of clock noise in general is based on the discrete wavelet transform (DWT). One form of wavelet-based analysis is closely related to the Allan variance [1], which is a standard tool in the characterization of clock noise. In what follows, we give a brief introduction to the DWT, following which we describe some of its uses in analyzing clock noise.

4.1 Background: the Discrete Wavelet Transform

Let $\mathbf{X} = [X_0, X_1, \ldots, X_{N-1}]^T$ denote a vector representing our observed time series. For convenience, we assume that the sample size N is an integer multiple of some power of two, which we denote as 2^{J_0} (in practice, wavelet methods can be readily adapted to handle arbitrary sample sizes). The DWT is an orthonormal transform of $\mathbf{X}$. The result of the transform is a new vector of length N, which we call the DWT coefficients and denote as $\mathbf{W}$. This vector is created by premultiplying $\mathbf{X}$ by an $N \times N$ real-valued matrix $\mathcal{W}$ that defines the DWT; i.e., $\mathbf{W} = \mathcal{W}\mathbf{X}$. The orthonormality conditions means that the inverse of $\mathcal{W}$ is just its transpose, which in turn means that $\mathbf{X} = \mathcal{W}^T\mathbf{W}$. Since we can exactly recover $\mathbf{X}$ from its DWT coefficients, the vector $\mathbf{W}$ can be regarded as an alternative (but fully equivalent) representation for the time series.

The structure of $\mathcal{W}$ is such that we can attach a physical meaning to each element of the DWT coefficient vector $\mathbf{W}$. Qualitatively we can partition $\mathbf{W}$ such that

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_{J_0} \\ \mathbf{V}_{J_0} \end{bmatrix}$$

and describe the subvectors as follows. The subvector $\mathbf{W}_j$ contains $N_j = N/2^j$ wavelet coefficients that are associated with changes in averages of segments of the time series. Each average spans a certain segment of data, and the amount that is spanned is referred to as the scale associated with the wavelet coefficient. The scale associated with all the coefficients in $\mathbf{W}_j$ is $\tau_j = 2^{j-1}$ time units, where 'time unit' refers to the time interval between adjacent measurements of the clock noise (because the scale is always a power of two, τ_j is sometimes referred to as the jth 'dyadic' scale). In addition, each wavelet coefficient in $\mathbf{W}$ is related to a particular segment of time spanning 2^j time units, and adjacent wavelet coefficients are related to time segments that are separated by 2^j time units. The final subvector of $\mathbf{W}$ is $\mathbf{V}_{J_0}$, which contains $N_{J_0} = N/2^{J_0}$ scaling coefficients. These coefficients are averages of the time series over a scale denoted by $\lambda_{J_0} = 2^{J_0}$.

As an example, Figure 7 depicts the matrix $\mathcal{W}$ for the Haar DWT with $N = 8$ and $J_0 = 3$. This matrix has eight rows, and the contents of each row are plotted in the figure. Let us consider the first row (indexed by $n = 0$), which consists of two nonzero values, followed by six zeros. When we form the inner product of this row and $\mathbf{X}$, we get a wavelet coefficient that is proportional to $X_1 - X_0$, i.e., the difference between two adjacent 'one point' averages. The first four rows collectively yield $\mathbf{W}_1$, with each coefficient in this vector being proportional to changes on a scale of $\tau_1 = 1$. The next two rows (indexed by $n = 4$ and 5) collectively yield $\mathbf{W}_2$, whose contents are proportional to changes on a scale of $\tau_2 = 2$. For example, if we take the inner product of row $n = 4$ and $\mathbf{X}$, the resulting coefficient is proportional to $\frac{1}{2}(X_3 + X_2) - \frac{1}{2}(X_1 + X_0)$, i.e., the difference between two adjacent 'two point' averages. The next row ($n = 6$) yields $\mathbf{W}_3$, with its single coefficient being proportional to a change on a scale of $\tau_3 = 4$. The final row yields $\mathbf{V}_3$. This single scaling coefficient is proportional to an average on a scale of $\lambda_3 = 8$, i.e., to the sample mean of all the data.

The wavelet and scaling coefficients that make up the DWT can also be expressed as outputs from filters applied to the time series $\mathbf{X}$. If we let $\{h_{j,l}, l = 0, \ldots, L_j - 1\}$ represent the wavelet filter associated with the jth scale, we can apply this filter to the time series to obtain the output

$$2^{j/2}\widetilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l \bmod N}, \quad t = 0, 1, \ldots, N-1 \tag{19}$$

(note that this filtering operation treats the data as if it were periodic with a period of N because, e.g., $t - l \bmod N = t - l + N$ when $-N \leq t - l \leq -1$ in the above summation). We obtain the tth element $W_{j,t}$ of $\mathbf{W}_j$ by subsampling this filter output; i.e.,

$$W_{j,t} = 2^{j/2}\widetilde{W}_{j,2^j(t+1)-1}, \quad t = 0, 1, \ldots, N_j - 1. \tag{20}$$

Figure 8 shows four sets of wavelet filters for scales τ_1 up to τ_7. The left-hand plot of Figure 8 shows the Haar wavelet filters, which, as we have already seen, yield wavelet coefficients that are proportional to the difference between adjacent sample averages. The remaining plots in this figure show, from left to right, the Daubechies D(4), C(6) and LA(8) wavelet filters (for details, see [6] and [21]). The D(4) filters are asymmetric, whereas, by design, the C(6) and LA(8) wavelet filters are much more symmetric and lead to an interpretation for the associated wavelet coefficients of differences between a centrally located weighted average and averages of values before and after it. In terms of their squared gain functions, the jth wavelet filter is approximately a band-pass filter with a pass-band given by $[1/2^{j+1}, 1/2^j]$. Thus the notion of scale is related to an *interval* of frequencies rather than to just one particular frequency (although wavelet analysis is sometimes referred to as a 'time/frequency' technique, a better descriptor for it is 'time/scale').

In a manner completely analogous to (19), we can obtain the scaling coefficients in $\mathbf{V}_{J_0}$ by using a scaling filter $\{g_{J_0,l}, l = 0, \ldots, L_{J_0} - 1\}$. Figure 9 shows four sets of scaling filters for $J_0 = 1, \ldots, 7$ corresponding to the wavelet filters shown in Figure 8. The Haar scaling filters yield scaling coefficients that are proportional to averages of the time series, whereas the C(6) and LA(8) scaling filters yield coefficients that are proportional to weighted averages (the D(4) filters can also be interpreted as such, but the asymmetric weights are not as appealing as the more symmetric looking weights for the C(6) and LA(8) filters). The J_0th scaling filter is approximately a low-pass filter with a pass-band given by $[0, 1/2^{J_0+1}]$. Collectively the wavelet filters of levels $j = 1, \ldots, J_0$ and the scaling filter of level J_0 partition the entire frequency interval $[0, 1/2]$.

The first level Haar wavelet filter has a width of two, whereas the corresponding D(4), C(6) and LA(8) filters have widths of, respectively, four, six and eight. As the width of the first level filter increases, the wavelet and scaling filters become better approximations to ideal band-pass and low-pass filters. Also, each wavelet filter has a certain number of backward difference filters contained within it: the Haar wavelet filter has one; the D(4) and C(6) filters each have two; and the LA(8) filter has four. For time series that are well modeled by an FD process, the number of embedded difference filters dictates the range of δ that can be handled by a particular wavelet filter (e.g., the Haar wavelet filter is limited to $\delta < 3/2$, whereas the LA(8) filter is less restricted in that it can in theory handle FD processes with $\delta < 9/2$).

As an example, Figure 10 shows the $J_0 = 6$ D(4) DWT coefficients for a sample of $N = 512$ daily phase deviates for clock 55 (a commercial cesium beam atomic clock) as measured against a US Naval Observatory clock time scale (this series is measured in nanoseconds and is the last part of a longer series of 564 daily measurements, with the first 52 days being used to detrend the remaining measurements via extrapolation of a linear least squares fit). The time series of these clock noise measurements is shown in the bottom plot in the left-hand column. Above this plot, there are seven other plots, which depict (from bottom to top) $\mathbf{W}_1, \ldots, \mathbf{W}_6$ and $\mathbf{V}_6$ plotted versus the times associated with each coefficient (the wavelet coefficients are plotted as deviations from zero, whereas the scaling coefficients are plotted by connected lines). Note that, as we go from the jth scale to the $(j + 1)$st scale, the number of wavelet coefficients decreases by a factor of two, whereas the times between adjacent coefficients increases by a factor of two. Wavelet coefficients with large magnitudes are associated with portions of the phase record where the atomic clock performed relatively poorly at the associated scale τ_j (i.e., there was a large change in the phase from one time period of length τ_j to the next); conversely, wavelet coefficients that are close to zero indicate time periods during which the clock performance was quite stable over a scale of τ_j. Because the scaling coefficients (left-hand column, top row) are proportional to averages over 64 days, their plot resembles a highly smoothed version of the original data.

The right-hand column of Figure 10 illustrates an interesting feature of the DWT, namely, that it can approximately decorrelate time series that have a high degree of autocorrelation. The bottom plot in this column shows the sample autocorrelation sequence (ACS) for the phase deviates out to a lag of 32 days:

$$\hat{\rho}_\tau \equiv \frac{\sum_{t=0}^{N-\tau-1}(X_t - \overline{X})(X_{t+\tau} - \overline{X})}{\sum_{t=0}^{N-1}(X_t - \overline{X})^2}, \quad \tau = 0, \ldots, 32.$$

This plot shows that the ACS damps down to zero at a very slow rate, which is indicative of a nonstationary FD process (i.e., $\delta > 1/2$). The six plots about this one show the sample ACSs for $\mathbf{W}_1, \ldots, \mathbf{W}_6$, along with upper and lower curves depicting 95% confidence limits under the assumption that the wavelet coefficients in $\mathbf{W}_j$ are a sample of white noise of length N_j. We see that, when sampling variability is taken into account, each subvector of wavelet coefficients is consistent with a hypothesis of white noise. In addition, it can be argued that the wavelet coefficients between different scales are approximately uncorrelated. Thus the DWT acts as a 'whitening transform' for these highly correlated clock noise measurements.

4.2 Wavelet-Based Analysis of Variance

We noted in our discussion of the periodogram that it gives a decomposition of the sample variance across frequencies (see (15)). This result is related to the fact that the discrete Fourier transform is proportional to an orthonormal transform. Because the DWT is an orthonormal transform, we can formulate a wavelet-based decomposition of the sample variance across scales. To see this, let us consider the squared norm (or 'energy') in our time series $\mathbf{X}$:

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

It is easy to argue that this energy is preserved in the DWT coefficients because we have

$$\|\mathbf{W}\|^2 = \|\mathcal{W}\mathbf{X}\|^2 = \mathbf{X}^T\mathcal{W}^T\mathcal{W}\mathbf{X} = \mathbf{X}^T\mathbf{X} = \|\mathbf{X}\|^2$$

(recall that the inverse of the DWT matrix $\mathcal{W}$ is its transpose, and hence $\mathcal{W}^T\mathcal{W}$ is the identity matrix). Since the subvectors $\mathbf{W}_1, \ldots, \mathbf{W}_{J_0}$ and $\mathbf{V}_{J_0}$ completely partition $\mathbf{W}$, we have

$$\|\mathbf{W}\|^2 = \sum_{j=1}^{J_0} \|\mathbf{W}_j\|^2 + \|\mathbf{V}_{J_0}\|^2,$$

where $\|\mathbf{W}_j\|^2$ is the contribution to the energy in the series due to variations at scale τ_j. Combining the above, we obtain the following scale-based analysis of the sample variance:

$$\frac{1}{N}\left(\sum_{j=1}^{J_0} \|\mathbf{W}_j\|^2 + \|\mathbf{V}_{J_0}\|^2\right) = \frac{1}{N}\sum_{t=0}^{N-1} X_t^2.$$

The above is analogous to the frequency-based decomposition of (15).

4.3 The Maximal Overlap Discrete Wavelet Transform

As noted in (19) and (20), the DWT can be expressed in terms of a subsampling of the output from a filter applied to the time series $\mathbf{X}$. We can define an interesting and useful variation on the DWT that is known as the maximal overlap DWT (MODWT) by eliminating the subsampling. If we let $\widetilde{\mathbf{W}}_j$ represent a vector of length N containing the MODWT wavelet coefficients for scale τ_j, then the tth element of this vector is given by

$$\widetilde{W}_{j,t} \equiv \frac{1}{2^{j/2}} \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l \bmod N}, \quad t = 0, 1, \dots, N-1 \tag{21}$$

(cf. (19)). A similar equation involving the scaling filter $\{g_{j,l}\}$ rather than the wavelet filter $\{h_{j,l}\}$ can be used to define the elements of $\widetilde{\mathbf{V}}_{J_0}$, a vector of length N containing the MODWT scaling coefficients.

Unlike the DWT, the MODWT is not an orthonormal transform, but is in fact highly redundant. This is evident because, in the case of the DWT, there are exactly N DWT coefficients, whereas a level J_0 MODWT contains $N \cdot (J_0 + 1)$ coefficients in its constituent vectors $\widetilde{\mathbf{W}}_1, \dots, \widetilde{\mathbf{W}}_{J_0}$ and $\widetilde{\mathbf{V}}_{J_0}$. Like the DWT, we can use the MODWT to carry out an analysis of variance because it can be shown that

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

(for details, see Chapter 5 of [21]). Unlike the DWT, the MODWT is naturally defined for all samples sizes N (i.e., we need not make an assumption that N is related to a multiple of a power of two).

As an example, Figure 11 shows the D(4) MODWT coefficients out to level $J_0 = 6$ for the same series of phase measurements analyzed using the DWT in Figure 10. In comparison to that figure, we see that the elimination of subsampling in the MODWT yields a more detailed look at the scale-based fluctuations in $\mathbf{X}$ as a function of time. This type of analysis can help us determine if the fluctuations at a particular scale are homogeneous across time; if not, then the series might be a candidate for modeling as a time-varying FD process.

4.4 The Wavelet Variance

Let $\{X_t,\, t \in \mathbb{Z}\}$ be a stochastic process that we use as input to a jth level wavelet filter, obtaining as output (after rescaling)

$$\overline{W}_{j,t} \equiv \frac{1}{2^{j/2}} \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l}, \quad t \in \mathbb{Z}. \tag{22}$$

We define the wavelet variance (WV) as

$$\nu_{X,t}^2(\tau_j) \equiv \operatorname{var}\{\overline{W}_{j,t}\},$$

assuming that $\operatorname{var}\{\overline{W}_{j,t}\}$ exists and is finite (the quantity $\nu_{X,t}^2(\tau_j)$ is sometimes called the wavelet spectrum). In general, the WV depends on τ_j and t, but the statistical characteristics of typical clock noise appear to be invariant over long stretches of time, leading us to consider processes for which the WV is in fact independent of time; i.e.,

$$\nu_X^2(\tau_j) \equiv \operatorname{var}\{\overline{W}_{j,t}\}$$

is constant for all t. In practice, the statistical techniques that we outline below for the time-independent case can be readily adapted to certain processes with time-dependent WVs if the variations over time are relatively slow.

One rationale for considering the WV is that it decomposes the variance of a stationary process on a scale-by-scale basis. Recall that, if $\{X_t\}$ is a stationary process with an SDF $S_X(\cdot)$, then (2) states the fundamental result that the integral of the SDF is equal to the process variance; i.e., the SDF decomposes $\text{var}\,\{X_t\}$ across frequencies f. This decomposition involves an uncountable infinite number of frequencies f. As noted in (5), we have the interpretation that $S_X(f)\,\Delta f$ is the contribution to $\text{var}\,\{X_t\}$ due to frequencies in an interval of length Δf centered at f. If we are willing to interpret $\text{var}\,\{X_t\}$ as being infinite for nonstationary processes with stationary backward differences, then (2) and (5) hold for these process also,

For the WV the analog to (2) is

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

i.e., the WV decomposes the process variance across scales τ_j, and we can regard $\nu_X^2(\tau_j)$ as the contribution to $\text{var}\,\{X_t\}$ due to variations at scale τ_j. The fact that the WV decomposes the process variance mimics results that we have already noted, namely, that the DWT and MODWT of a time series $\mathbf{X}$ decompose its sample variance across scales. In a certain sense, the decomposition of the process variance given by the WV is simpler than the SDF-based decomposition because the former depends on a countably infinite number of scales, whereas the latter involves an uncountable infinite number of frequencies. In addition, the square root $\nu_X(\tau_j)$ of the WV has the same units as the process $\{X_t\}$ and hence is arguably easier to interpret.

A second rationale for considering the WV is that it is often a useful substitute or complement for the SDF. The fact that the wavelet filter $\{h_{j,l}\}$ is approximately a band-pass filter with a passband given by $[1/2^{j+1}, 1/2^j]$ implies that

$$\nu_X^2(\tau_j) \approx 2 \int_{1/2^{j+1}}^{1/2^j} S_X(f)\,df. \tag{23}$$

If the SDF $S_X(\cdot)$ is relatively featureless within every passband of the form $[1/2^{j+1}, 1/2^j]$, then, loosely speaking, the information in the WV over all scales is equivalent to the information in the SDF; however, the WV offers a more succinct presentation of this information since we only need one $\nu_X^2(\tau_j)$ for each octave band of frequencies.

The SDF $S_X(\cdot)$ for an FD process is a good example of one that is relatively featureless. Since $S_X(f) \approx \sigma_\epsilon^2/|2\pi f|^{2\delta}$ at low frequencies, (23) implies that we should have

$$\nu_X^2(\tau_j) \propto \tau_j^{2\delta-1} \tag{24}$$

approximately. Taking logs yields

$$\log\,(\nu_X^2(\tau_j)) \approx a + (2\delta - 1)\log\,(\tau_j),$$

where a is related to the constant of proportionality in (24). We can thus deduce δ from slope of plots of $\log\,(\nu_X^2(\tau_j))$ versus $\log\,(\tau_j)$. Once we have estimates of $\nu_X^2(\tau_j)$ over a range of moderate to large scales and once we have determined the statistical properties of these estimates, we can estimate δ and σ_ϵ^2 by applying regression analysis to the log of these estimates.

We can formulate a useful and efficient estimator of $\nu_X^2(\tau_j)$ based upon the MODWT wavelet coefficients of (21) (we could also formulate an estimator based upon the DWT wavelet coefficients, but it turns out that this estimator is less efficient than the corresponding MODWT estimator; i.e., the variance of the DWT-based estimator cannot be smaller than that of the MODWT-based estimator and generally is larger). Recalling that by definition $\nu_X^2(\tau_j) = \text{var}\{\overline{W}_{j,t}\}$, a comparison of (21) for $\widetilde{W}_{j,t}$ and (22) for $\overline{W}_{j,t}$ shows that $\widetilde{W}_{j,t} = \overline{W}_{j,t}$ as long as the 'modulo N' operation is not formally needed in the definition of the former. This happens when $L_j - 1 \leq t < N$, so, under the assumption that $N - L_j \geq 0$, we can obtain an unbiased estimator of $\nu_X^2(\tau_j)$ using

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

where $M_j \equiv N - L_j + 1$. By including all possible MODWT wavelet coefficients, we can also construct a biased estimator of $\nu_X^2(\tau_j)$ via

$$\tilde{\nu}_X^2(\tau_j) \equiv \frac{1}{N} \sum_{t=0}^{N-1} \widetilde{W}_{j,t}^2 = \frac{1}{N} \left(\sum_{t=0}^{L_j-2} \widetilde{W}_{j,t}^2 + \sum_{t=L_j-1}^{N-1} \overline{W}_{j,t}^2 \right),$$

where the first sum in the parentheses involves the MODWT wavelet coefficients that are influenced by the circularity assumption.

While the unbiased estimator $\hat{\nu}_X^2(\tau_j)$ might seem preferable, in fact there are reasons for considering the biased estimator $\tilde{\nu}_X^2(\tau_j)$. First, the biased estimator is actually unbiased if $\{X_t\}$ is a white noise process. Second, the biased estimator offers an exact decomposition of the sample variance, whereas the unbiased estimator need not. The final and most compelling reason for considering $\tilde{\nu}_X^2(\tau_j)$ is that its mean square error can be smaller than $\hat{\nu}_X^2(\tau_j)$; i.e., since the mean square error of an estimator is equal to the sum of its variance and its squared bias, the bias in $\tilde{\nu}_X^2(\tau_j)$ is compensated for by a decrease in its variability due to the use of the additional MODWT wavelet coefficients. For details, see Greenhall et al. [13], where it is noted that the improvement in mean square error comes about only if, instead of analyzing $\{X_t\}$ itself, we apply the MODWT to a series of length $2N$ formed by extending $\{X_t\}$ with a reflected (i.e., time reversed) version of itself; i.e., we analyze

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

rather than just $X_0, X_1, \ldots, X_{N-2}, X_{N-1}$.

Suppose now that $\{\overline{W}_{j,t}\}$ obeys a Gaussian distribution and is a stationary process with zero mean and with an SDF given by $S_j(\cdot)$ (this latter condition holds if, e.g., $\{X_t\}$ is an FD process and if the wavelet filter has enough embedded backward difference filters to reduce the FD process to a zero mean stationary process). Since $\{\overline{W}_{j,t}\}$ is the result of filtering $\{X_t\}$ with the wavelet filter $\{h_{j,l}\}$, (4) tells us that $S_j(\cdot)$ involves the product of the SDF for $\{X_t\}$ and the squared gain function for the filter. Suppose further that $S_j(f) > 0$ at almost all frequencies (as is true in all practical situations) and that the following square integrability condition holds:

$$A_j \equiv \int_{-1/2}^{1/2} S_j^2(f)\, df < \infty$$

(this condition is true for all FD processes as long as the wavelet filter has enough embedded backward differences filters). Under these conditions, it is possible to show that $\hat{\nu}_X^2(\tau_j)$ is asymptotically normally distributed with a mean of $\nu_X^2(\tau_j)$ and a large sample variance given by $2A_j/M_j$ [18]. In practice, we need to estimate A_j (see Chapter 8 of [21] for three methods for doing so), after which we can assess the variability in $\hat{\nu}_X^2(\tau_j)$ and hence in $\log(\hat{\nu}_X^2(\tau_j))$.

As an example, Figure 12 shows the clock errors X_t for clock 55 (previously considered in Figures 8 and 13), along with plots of its first and second order backward differences $X_t^{(1)} \equiv$

$X_t - X_{t-1}$ and $X_t^{(2)} \equiv X_t^{(1)} - X_{t-1}^{(1)}$. The first order backward differences are proportional to the fractional frequency deviates $\overline{Y}_t$ (these are obtained by dividing $X_t^{(1)}$ by 8.64×10^{13} since (i) X_t is measured in nanoseconds and (ii) the spacing between X_{t-1} and X_t is one day). While X_t is too dispersive to be a reasonable realization from a Gaussian stationary process, the plots in this figure suggest that its first and second differences can be modeled as such; however, $X_t^{(1)}$ appears to have a sample mean that is inconsistent with the assumption of a zero mean, whereas $X_t^{(2)}$ appears to be a realization of a zero mean process.

Figure 13 shows the square root $\hat{\nu}_X(\tau_j)$ of the estimated WVs versus τ_j on a log/log scale for the clock errors X_t. The estimates $\hat{\nu}_X(\tau_j)$ in the left-hand plot are based on three different wavelet filters, namely, the Haar wavelet filter (`x`'s), the D(4) filter (circles) and the D(6) filter (pluses). While the D(4) and D(6) estimates are in good agreement, the Haar estimates are systematically larger. This anomalous result can be attributed to the fact that the Haar wavelet filter involves a single backward difference, whereas a second order backward difference is required to reduce $\{X_t\}$ to a series that can reasonably be interpreted as a realization of a zero mean stationary process. The fact that the Haar estimates are elevated can be regarded as a form of leakage, which can be handled easily in wavelet analysis by merely using a wavelet filter with more embedded backward differences (for the D(L) and LA(L) Daubechies wavelet filters, the number of such backward differences is $L/2$). The right-hand plot in Figure 13 shows the D(4) values for $\hat{\nu}_X(\tau_j)$ again, now along with 95% confidence intervals and a linear least squares fit for deducing the δ parameter in a possible simple FD model for X_t (the slope of the line in log/log space is 0.48, which translated to an estimate of 0.98 for δ – this is quite close to the value $\delta = 1$ for a random walk process).

Figure 14 shows a WV analysis of the daily average fractional frequency deviates $\{\overline{Y}_t\}$ (these are proportional to $\{X_t^{(1)}\}$). The left-hand plot shows $\hat{\nu}_{\overline{Y}}(\tau_j)$ versus τ_j on a log/log scale using the Haar wavelet filter (`x`'s) and the D(4) wavelet filter (circles). Because the first difference of $\{\overline{Y}_t\}$ can be reasonably regarded as a realization of a zero mean stationary process and because there is at least one backward difference filter embedded in both wavelet filters, the resulting estimates of $\nu_{\overline{Y}}(\tau_j)$ for the two filters are quite similar. When we use the Haar wavelet filter, the resulting WV $\nu^2_{\overline{Y}}(\tau_j)$ is proportional to the Allan variance $\sigma^2_{\overline{Y}}(2, \tau_j)$; in fact, we have

$$\nu^2_{\overline{Y}}(\tau_j) = \tfrac{1}{2}\sigma^2_{\overline{Y}}(2, \tau_j).$$

The right-hand plot shows the D(4) estimates again, but this time with 95% confidence intervals and a regression fit for deducing the δ parameter in a simple FD model for $\overline{Y}_t$ (the estimate slope is now -0.47, which translates to $\delta = 0.03$ and is quite close to the value $\delta = 0$ for a white noise process).

5 Summary

To briefly summarize our main points, fractionally differenced (FD) processes provide a flexible and tractable class of models that obey power laws at low frequencies with any arbitrary exponent. FD processes are easy to work with because the expressions for their SDFs, ACVSs and PACSs are all quite simple. In addition, these processes can be readily extended in several directions, including composite FD processes, ARFIMA processes and time-varying FD processes.

Given a sample of clock noise that we propose to model as a portion of a realization of an FD process, we can use either spectral analysis or wavelet analysis to obtain estimates of the parameters for FD processes. There are also numerous time-domain techniques for estimating these parameters, but, while some of these are competitive with spectral or wavelet methods, these latter techniques also offer a decomposition of the sample variance across frequencies (in the case of spectral analysis) or across scales (in the case of wavelet analysis). The two techniques provide complementary analyses, but there are some potential advantages in using wavelet analysis for clock noise. First, wavelet-based estimators of δ and σ_ϵ^2 tend to perform somewhat better than SDF-based estimators. Second, wavelet analysis is useful for characterizing and analyzing clock noise with time-varying statistical properties. Third, wavelet filters can handle clock noise observed in the presence of certain polynomial trends (for details, see Chapter 9 of [21]). Finally, estimators of the WV have the same units as X_t^2, which facilitates their interpretation.

6 Acknowledgments

The author would like to thank (i) the organizers of the IV International Time Scales Algorithms Symposium (Patrizia Tavella, Demetrios Matsakis and Felicitas Arias) for the invitation and encouragement to put together this primer; (ii) the IEN Galileo Ferraris, Torino, Italy, for funding that made attendance possible; (iii) Lara Schmidt, for providing the data for clock 55; and (iv) two anonymous reviewers and the editor for very helpful critiques.

References

[1] Allan, D. W., *Proceedings of the IEEE*, 1966, **54**, 221–230.

[2] Bartlett, M. S., Kendall, D. G., *Supplement to the Journal of the Royal Statistical Society*, 1946, **8**, 128—138.

[3] Beran, J., *Statistics for Long-Memory Processes*, New York, Chapman & Hall, 1994, 315 p.

[4] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., *Time Series Analysis: Forecasting and Control*, 3rd ed., Englewood Cliffs, New Jersey, Prentice Hall, 1994, 598 p.

[5] Craigmile, P. F., *Journal of Time Series Analysis*, 2003, to appear.

[6] Daubechies, I., *Ten Lectures on Wavelets*, Philadelphia, SIAM, 1992, 357 p.

[7] Davies, R. B., Harte, D. S., *Biometrika*, 1987, **74**, 95–102.

[8] Dietrich, C. R., Newsam, G. N., *SIAM Journal on Scientific Computing*, 1997, **18**, 1088–1107.

[9] Draper, N. R., Smith, H., *Applied Regression Analysis*, 3rd ed., New York, John Wiley & Sons, 1998, 706 p.

[10] Geweke, J., Porter–Hudak, S., *Journal of Time Series Analysis*, 1983, **4**, 221–238.

[11] Gneiting, T., *Statistics and Probability Letters*, 1998, **39**, 119–122.

[12] Granger, C. W. J., Joyeux, R., *Journal of Time Series Analysis*, 1980, **1**, 15–29.

[13] Greenhall, C. A., Howe, D. A., Percival, D. B., *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 1999, **46**, 1183–1191.

[14] Hosking. J. M. R., *Biometrika*, 1981, **68**, 165–176.

[15] Mandelbrot, B. B., van Ness, J. W., *SIAM Review*, 1968, **10**, 422–437.

[16] McCoy, E. J., Walden, A. T., Percival, D. B., *IEEE Transactions on Signal Processing*, 1998, **46**, 655–668.

[17] Percival, D. B., *Proceedings of the IEEE*, 1991, **79**, 961–972.

[18] Percival, D. B., *Biometrika*, 1995, **82**, 619–631.

[19] Percival, D. B., Constantine, W. L. B., *Journal of Computational and Graphical Statistics*, 2002, submitted.

[20] Percival, D. B., Walden, A. T., *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*, Cambridge, UK, Cambridge University Press, 1993, 583 p.

[21] Percival, D. B., Walden, A. T., *Wavelet Methods for Time Series Analysis*, Cambridge, UK, Cambridge University Press, 2000, 594 p.

[22] Riedel, K. S., Sidorenko, A., *IEEE Transactions on Signal Processing*, 1995, **43**, 188–195.

[23] Sinai, Y. G., *Theory of Probability and Its Applications*, 1976, **21**, 64–80.

[24] Thomson, D. J., *Proceedings of the IEEE*, 1982, **70**, 1055–1096.

[25] Wood, A. T. A., Chan, G., *Journal of Computational and Graphical Statistics*, 1994, **3**, 409–432.

[26] Yaglom, A. M., *American Mathematical Society Translations, Series 2*, 1958, **8**, 87–141.

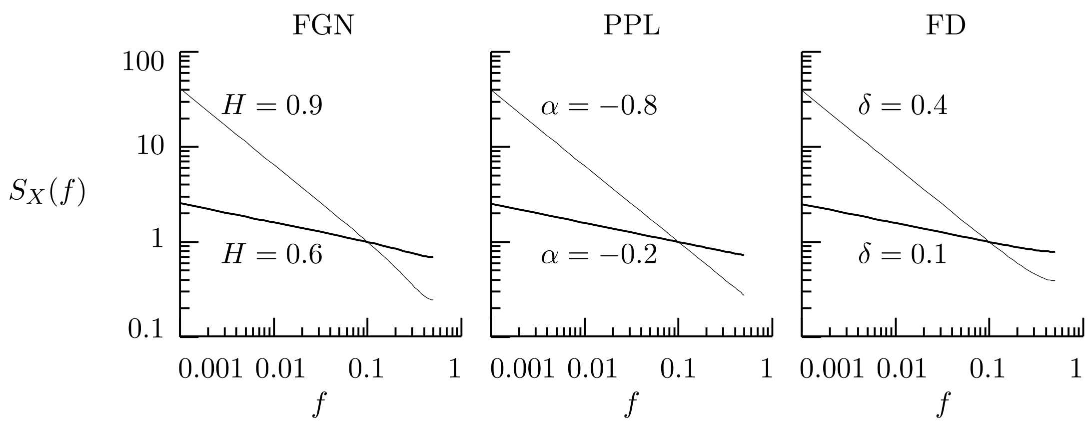

Figure 1: SDFs for FGN, PPL and FD processes on log/log axes. Each SDF $S_X(\cdot)$ is normalized such that $S_X(0.1) = 1$.

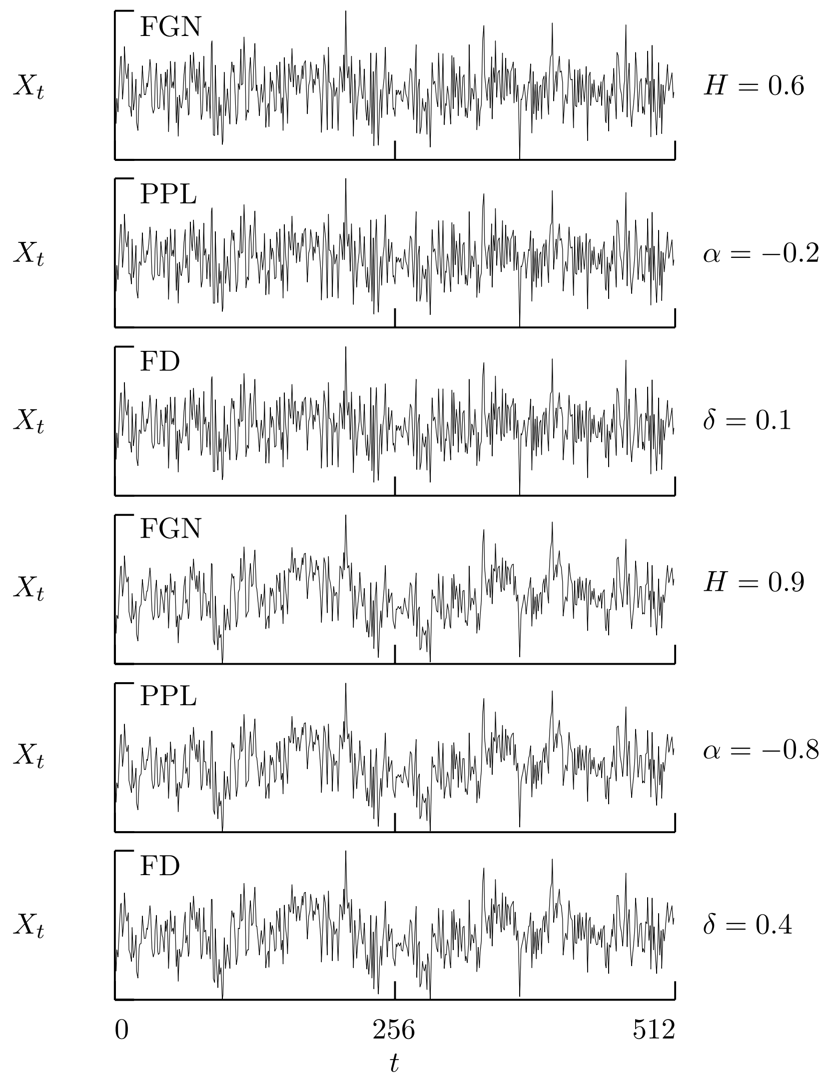

Figure 2: Simulated realizations X_t of FGN, PPL and FD processes (the SDFs for these processes are shown in Figure 1).

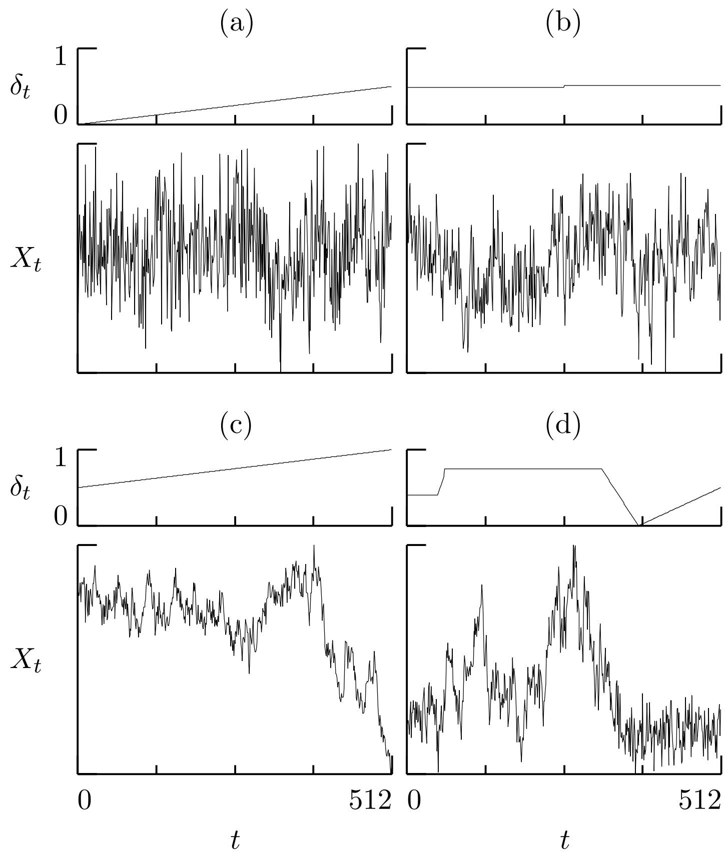

Figure 3: Four examples of time series simulated from TVFD processes. The upper panel of each plot shows the sequence δ_t, $t = 0, \ldots, 511$, used to generate the simulated time series in the lower panel.

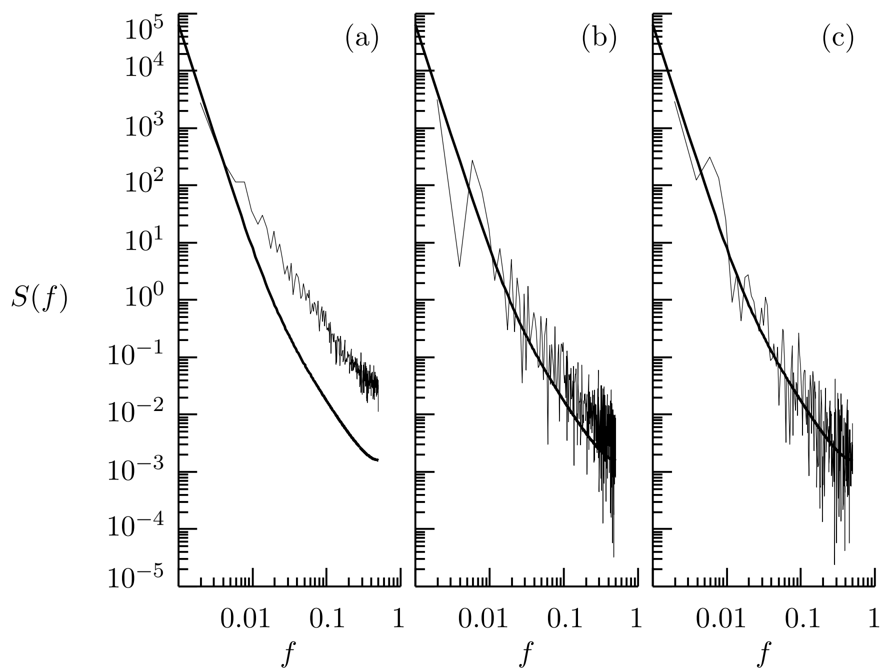

Figure 4: Periodogram (a) and direct spectral estimates (b, c) of clock noise simulated from a composite FD process with two components specified by $\delta_1 = 2$, $\sigma_1^2 = 0.0001$, $\delta_2 = 1$ and $\sigma_2^2 = 0.0064$. The true SDF is the thick curve in each plot, while the thin curves are the spectral estimates. The data tapers used in the direct spectral estimates are Slepian tapers (i.e., discrete prolate spheroidal sequences) with the resolution bandwidth W set via (b) $NW = 1$ and (c) $NW = 2$. The simulated clock noise is shown in the upper right-hand plot of Figure 5.

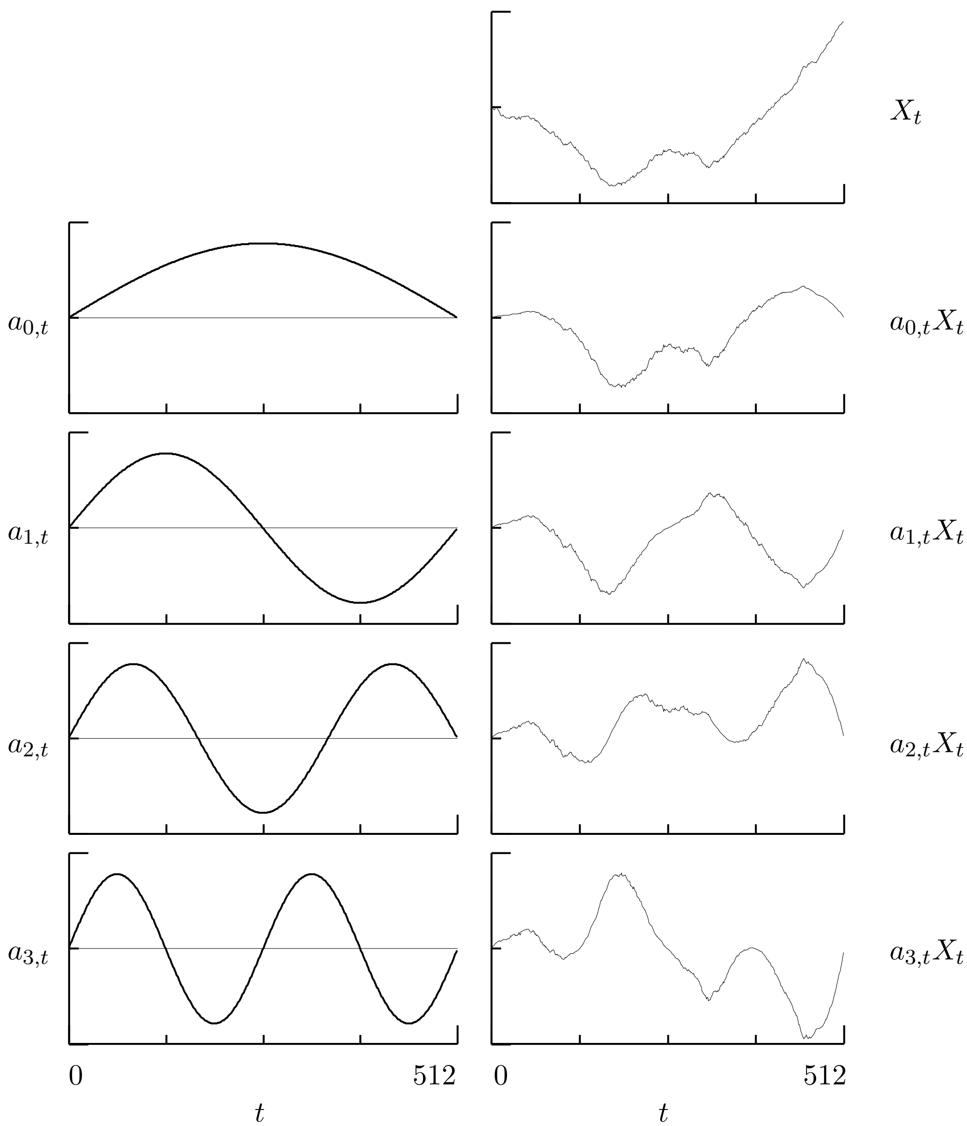

Figure 5: Sinusoidal tapers (left-hand column) as applied to a simulated composite FD time series (top plot, right-hand column), resulting in four tapered series (right-hand column, second to bottom rows).

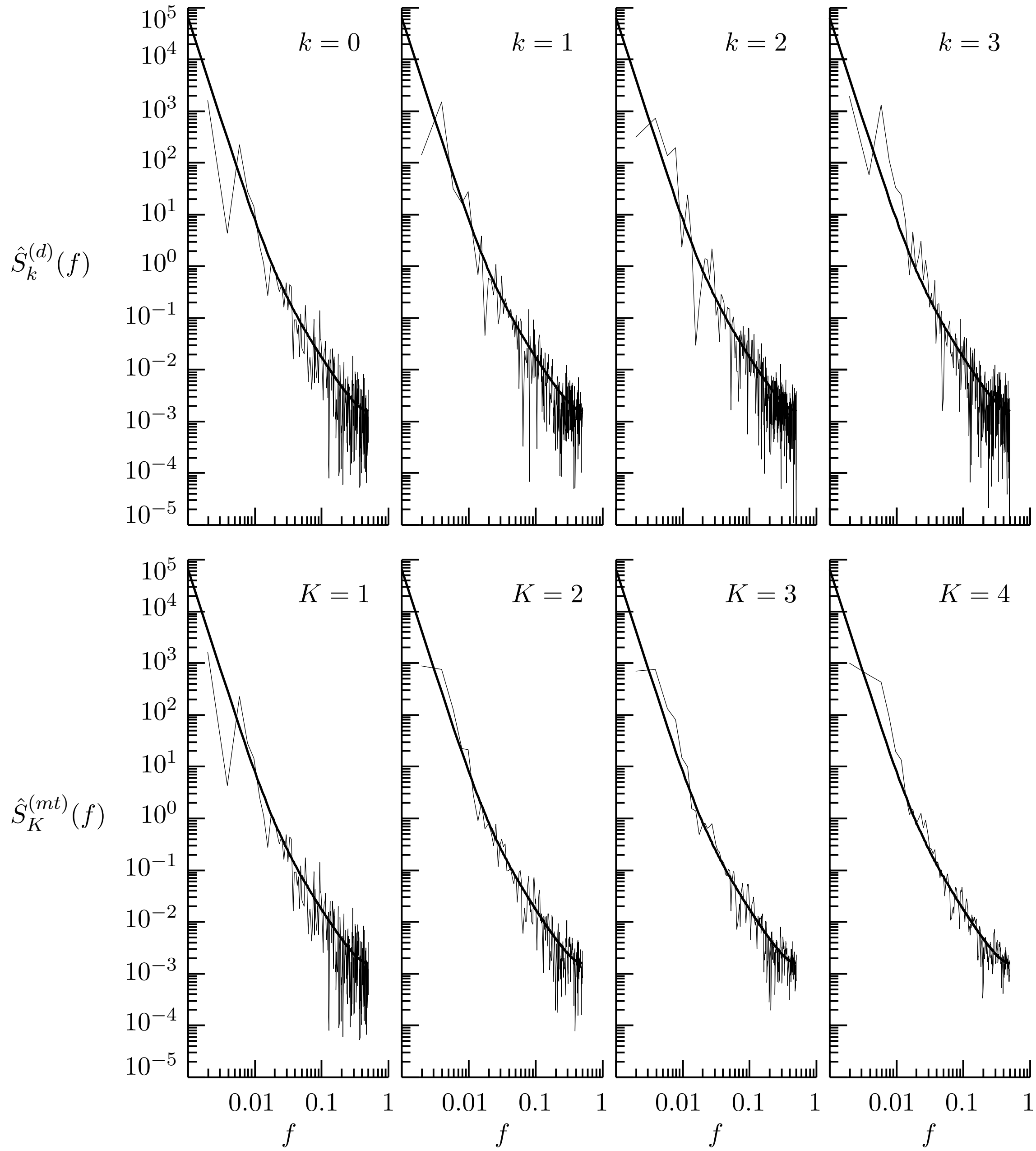

Figure 6: Direct spectral estimates $\hat{S}_k^{(d)}(\cdot)$ formed using the kth sinusoidal taper, $k = 0, \ldots, 3$ (top row, left- to right-hand plots), along with multitaper estimates $\hat{S}_K^{(mt)}(\cdot)$ formed by averaging $K = 1, \ldots, 4$ of these direct spectral estimates (bottom row).

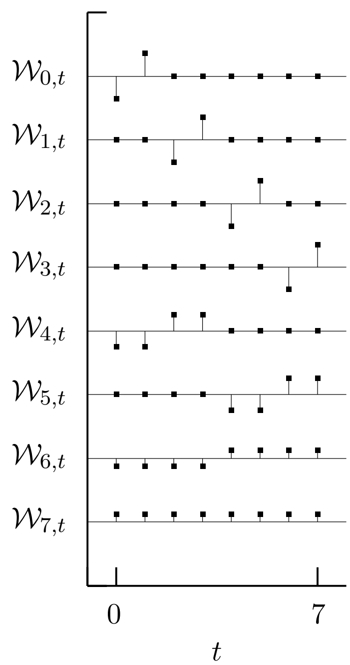

Figure 7: Haar discrete wavelet transform matrix $\mathcal{W}$ for the case $N = 8$ and $J_0 = 3$. The (n, t)th element of this 8×8 matrix is denoted as $\mathcal{W}_{n,t}$, where $n, t = 0, \dots, 7$. The eight values in each row n are plotted as deviations from zero (indicated by a horizontal line).

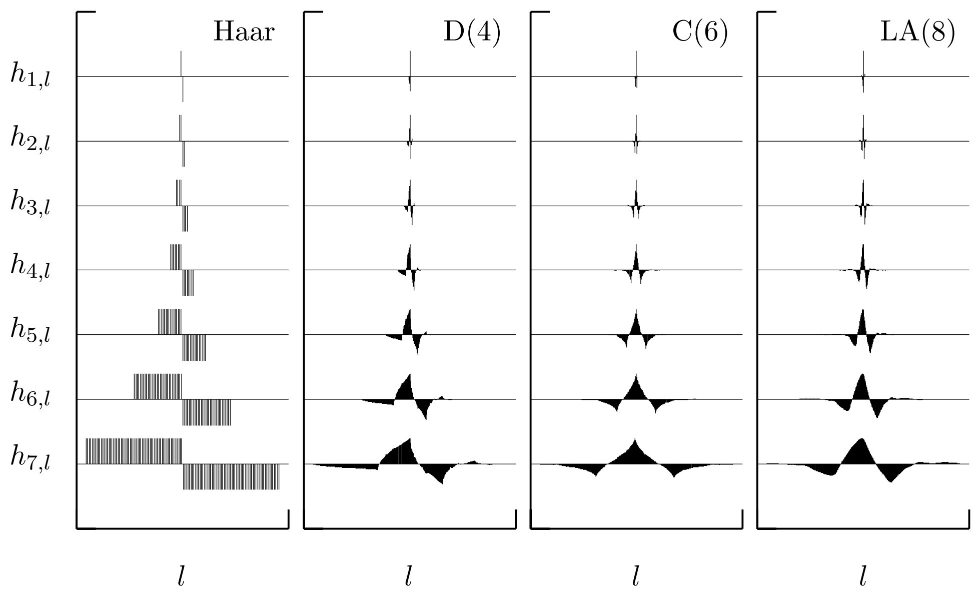

Figure 8: Haar, D(4), C(6) and LA(8) wavelet filters $h_{j,l}$ for scales $\tau_j = 2^{j-1}$, $j = 1, 2, \ldots, 7$.

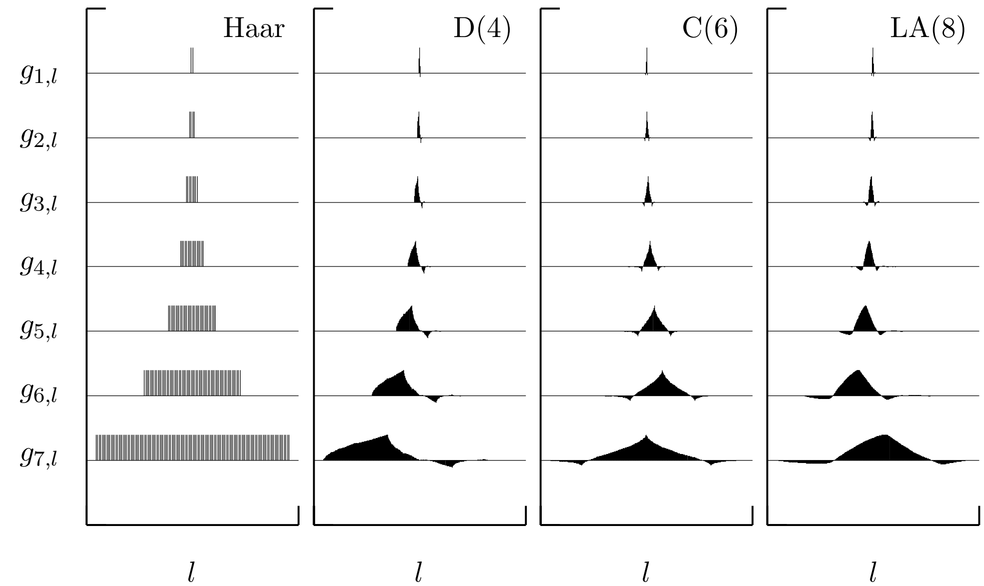

Figure 9: Haar, D(4), C(6) and LA(8) scaling filters $g_{J_0,l}$ for scales $\lambda_{J_0} = 2^{J_0}$, $J_0 = 1, 2, \ldots, 7$.

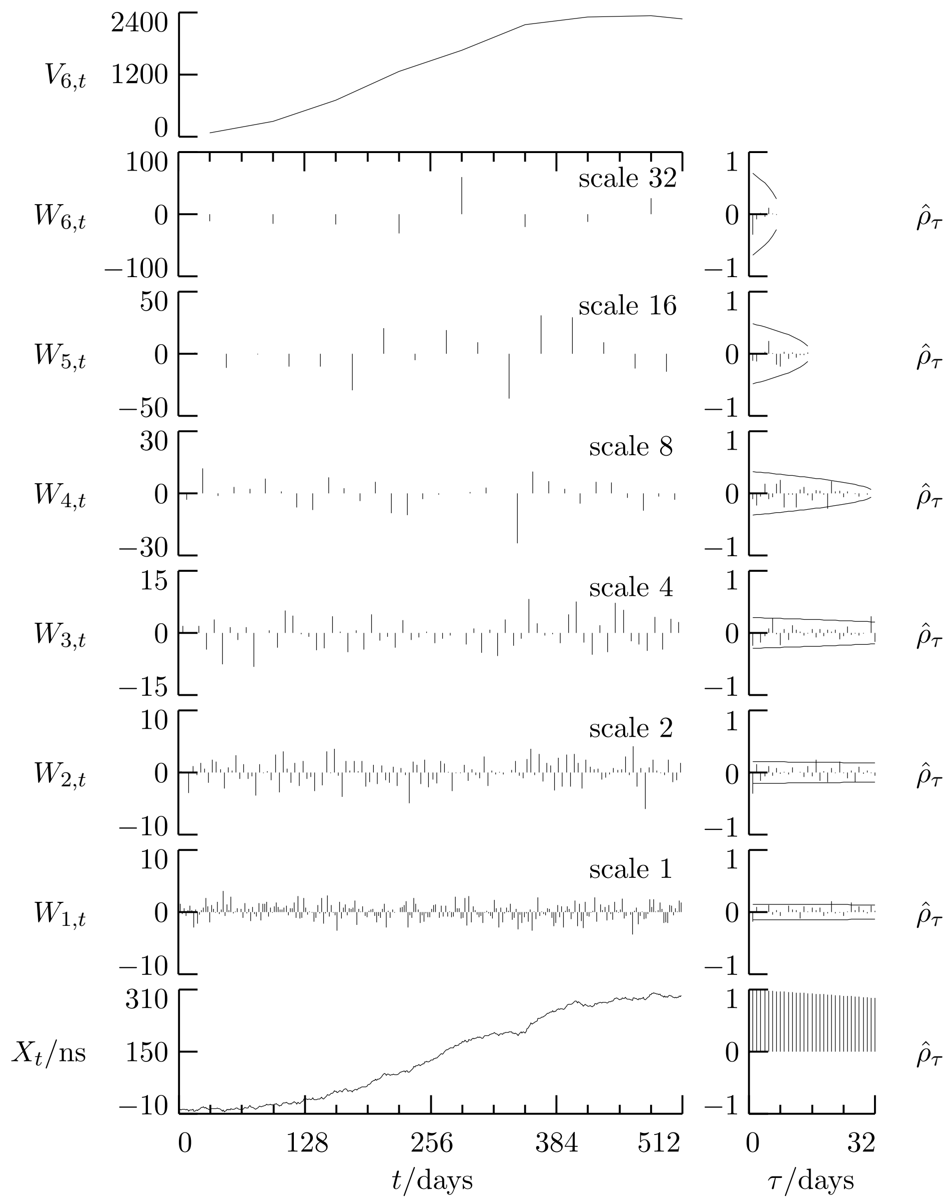

Figure 10: D(4) DWT wavelet and scaling coefficients ($W_{j,t}$ and $V_{6,t}$) for clock 55 and sample autocorrelation sequences ($\hat{\rho}_\tau$).

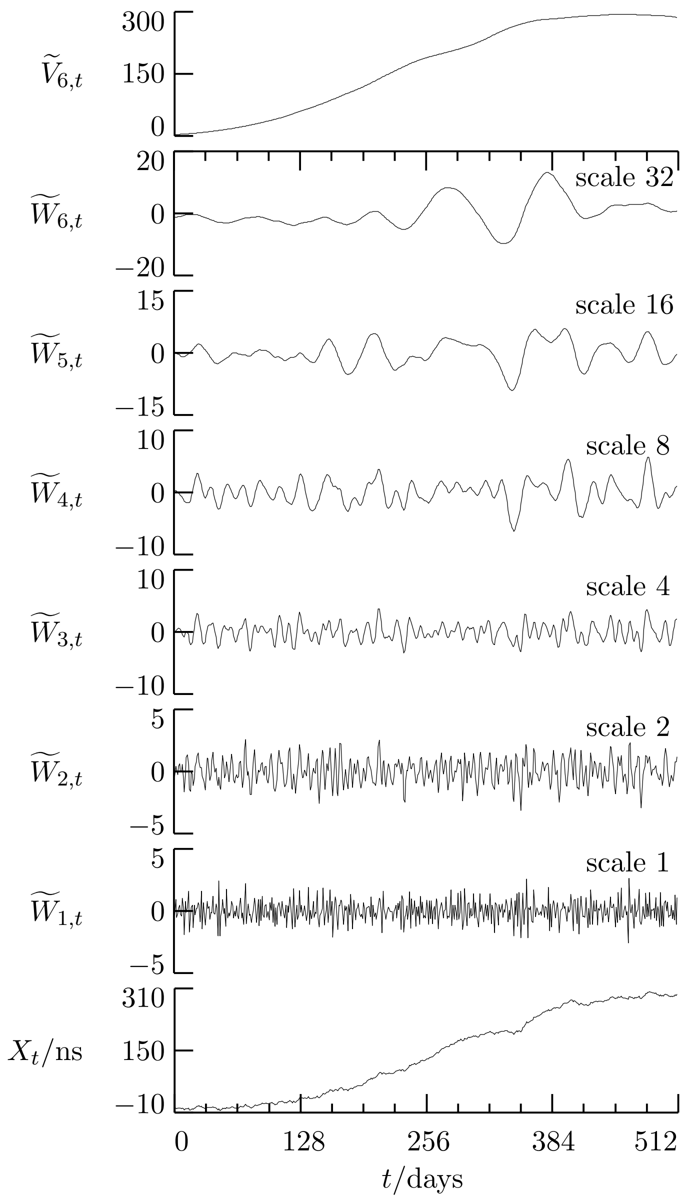

Figure 11: D(4) MODWT wavelet and scaling coefficients ($\widetilde{W}_{j,t}$ and $\widetilde{V}_{6,t}$) for clock 55.

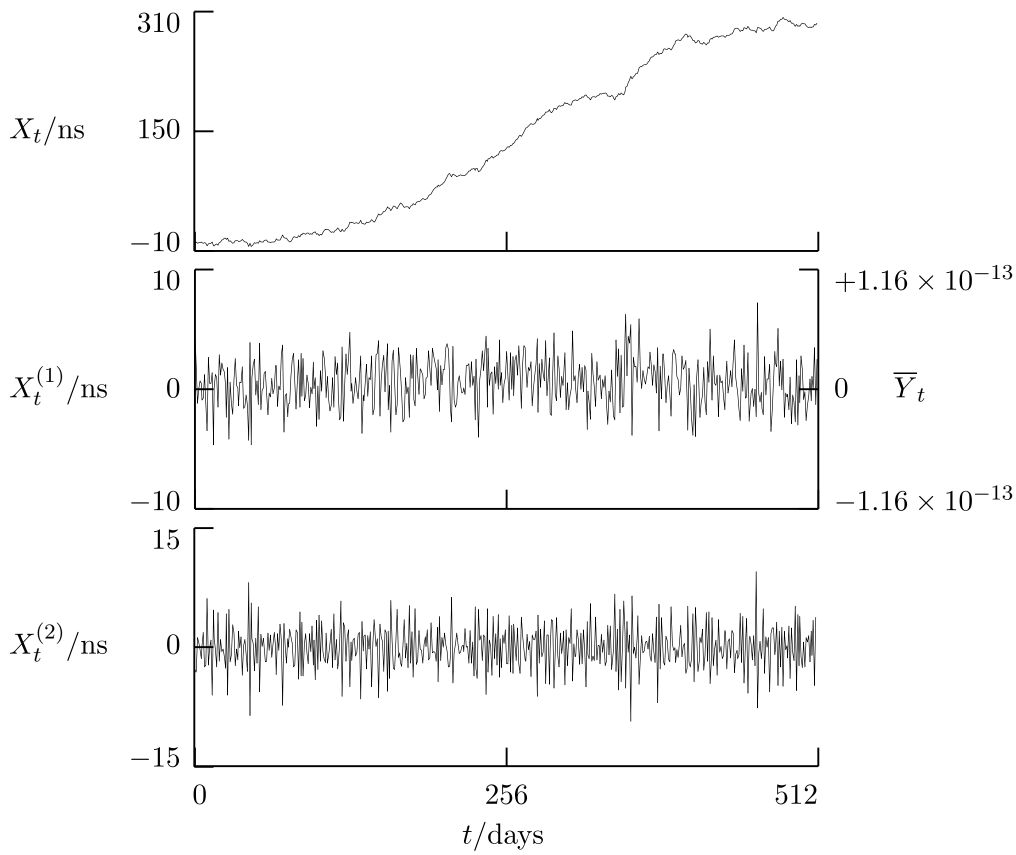

Figure 12: Plot of differences in time $\{X_t\}$ as kept by clock 55 and as kept by a time scale maintained by the US Naval Observatory, Washington, DC (top plot); its first backward difference $\{X_t^{(1)}\}$ (middle); and its second backward difference $\{X_t^{(2)}\}$ (bottom). In the middle plot, $\overline{Y}_t$ denotes the daily average fractional frequency deviates (plotted according to the right-hand vertical axis) – these are proportional to $X_t^{(1)}$.

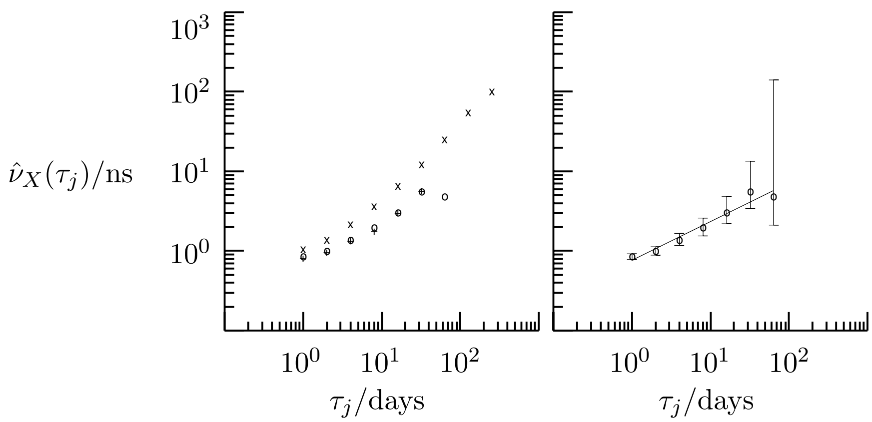

Figure 13: Square roots $\hat{\nu}_X(\tau_j)$ of wavelet variance estimates for clock 55's phase differences X_t based upon the unbiased MODWT estimator and the following wavelet filters: Haar (`x`'s in left-hand plot), D(4) (circles in left- and right-hand plots) and D(6) (pluses in left-hand plot). The right-hand plot also shows 95% confidence intervals for the unknown wavelet variances and a linear least squares fit to $\log_{10}(\hat{\nu}_X(\tau_j))$ versus $\log_{10}(\tau_j)$.

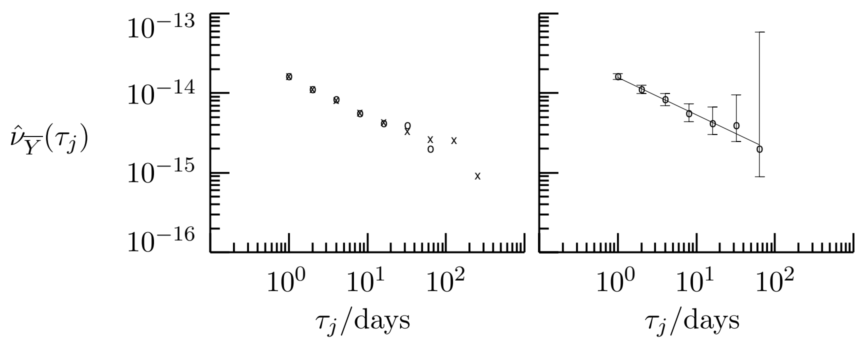

Figure 14: Square roots $\hat{\nu}_{\overline{Y}}(\tau_j)$ of wavelet variance estimates for clock 55's one day average fractional frequency deviates $\overline{Y}_t(\tau_1)$ based upon the unbiased MODWT estimator and the following wavelet filters: Haar (x's in left-hand plot) and D(4) (circles in left and right-hand plots). The right-hand plot also gives 95% confidence intervals for the $\nu_{\overline{Y}}(\tau_j)$ and a least squares fit to $\log_{10}(\hat{\nu}_{\overline{Y}}(\tau_j))$ versus $\log_{10}(\tau_j)$.