Spectral Analysis of Clock Noise: A Primer

Donald B Percival

Applied Physics Laboratory

University of Washington

Seattle, WA 98195–5640, USA

Abstract. The statistical characterization of clock noise is important for understanding how well a clock can perform in applications where timekeeping is important. The usual frequency domain characterization of clock noise is the power spectrum. We present a primer on how to estimate the power spectrum of clock noise given a finite sequence of measurements of time (or phase) differences between two clocks. The simplest estimator of the spectrum is the periodogram. Unfortunately this estimator is often problematic when applied to clock noise. Three estimators that overcome the deficiencies of the periodogram are the sinusoidal multitaper spectral estimator, Welch's overlapped segment averaging estimator and Burg's autoregressive estimator. We give complete details on how to calculate these three estimators. We apply them to two examples of clock noise and find that they all improve upon the periodogram and give comparable results. We also discuss some of the uses for the spectrum and its estimates in the statistical characterization of clock noise.

1. Introduction

Spectral analysis is one of the most commonly used techniques for studying measurements collected at regularly spaced intervals of time. Its use is ubiquitous in, e.g., atmospheric science, oceanography, geophysics, astronomy, physics and engineering. The subject has been covered in a number of textbooks over the last fifty years [4, 6, 7, 8, 9, 13, 15, 18, 19, 23, 25, 26, 27, 35, 37, 41, 42, 49]. Given

a real-valued time series X_t, $t = 0, 1, \ldots, N-1$, the basic idea behind spectral analysis to decompose the sample variance

$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{t=0}^{N-1} (X_t - \overline{X})^2, \text{ where } \overline{X} = \frac{1}{N} \sum_{t=0}^{N-1} X_t, \tag{1}$$

into components that can be associated with various Fourier frequencies f. The decomposition is expressed as an empirical power (or variance) spectrum $\hat{S}_X(\cdot)$, which is a nonnegative function of Fourier frequency. If we let Δt be the amount of time that elapses between observing X_t and X_{t+1}, then we have the basic constraint that

$$\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \hat{S}_X(f)\,\mathrm{d}f = 2\int_0^{f_{\mathcal{N}}} \hat{S}_X(f)\,\mathrm{d}f = \hat{\sigma}_X^2, \tag{2}$$

where $f_{\mathcal{N}} = 1/(2\,\Delta t)$ is known as the Nyquist frequency. The first equality above follows because, by construction, $\hat{S}_X(-f) = \hat{S}_X(f)$. The physically meaningful frequencies are those in the interval $[0, f_{\mathcal{N}}]$ (a 'two-sided' spectrum with negative frequencies is introduced to simplify mathematical developments and to allow generalization to complex-valued time series). The interpretation of equation (2) follows from a consideration of the Fourier representation for X_t, which reexpresses a time series as a sum of sines and cosines (collectively we refer to these as sinusoids). The Fourier representation essentially takes the form

$$X_t = \sum_{0 \le f \le f_{\mathcal{N}}} A(f)\cos(2\pi f t) + B(f)\sin(2\pi f t), \tag{3}$$

where $\hat{S}_X(f)$ is associated with $A^2(f) + B^2(f)$ (the expression above ignores the exact form of the summation so that we can focus our discussion on the ideas rather than the mathematical details). Thus, when $\hat{S}_X(f)$ is large, the amplitudes associated with the sinusoids of frequency f are important contributors to the Fourier representation for X_t. Studying the spectrum $\hat{S}_X(f)$ can give us insight into how a time series is structured and how it compares to other time series.

Given the prominent role that the spectrum plays in the physical sciences and engineering, it comes as no surprise that it is one of the major characterizations

of clock noise. In this context the difference in time as kept by two clocks (or, equivalently, the phase difference between the oscillators that are driving the clocks) is measured, and X_t represents the tth such measured time difference. Measurements are commonly taken at regularly spaced intervals, in part because this sampling design greatly simplifies the estimation of spectra and of other measures of clock performance (in practice, instrumentation problems can result in missing measurements, but the proportion of missing data in metrology is usually small, and lost data can be filled in using a variety of interpolation schemes). Examples of measured time differences are shown in figure 1. If both clocks being compared were perfect timekeepers, then X_t would be identically zero at all measured times, its sample variance would be zero, and its spectrum would be zero at all frequencies. In practice X_t measures imperfections (hence justifying the pejorative name 'clock noise'), and the spectrum gives insight into the nature of the imperfections and what impact these might have when a particular clock is used in a particular application. As one example, if the spectrum for one clock has a lower level over a certain band of Fourier frequencies as compared to a second clock, the first clock will be more suitable for applications in which good timekeeping is required over time spans related to the inverses of the Fourier frequencies (i.e., periods). As a second example, a sharp peak in a spectrum implies that there is some structure in the clock noise that needs to be investigated and possibly exploited to improve the performance of the clock. Finally, one approach for forming a time scale from an ensemble of clocks is to weight each clock according to how predictable its clock noise is. The best linear predictor for clock noise can be deduced once its spectrum is known.

The goal of this article is to provide a self-contained primer on spectral analysis for clock noise measurements. The primer starts with a review of the definition of the spectrum $S_X(f)$ for intrinsically stationary stochastic processes (section 2.).

This class of processes provides realistic models for detrended clock noise. Given a set of clock noise measurements that can be regarded as a realization of a portion $X_0, X_1, \ldots, X_{N-1}$ of an intrinsically stationary process, our task is to estimate $S_X(f)$. The spectrum estimation problem in general is quite complicated. In the case of clock noise, the simplest estimator (the periodogram) is often inadequate (section 3.1.), but, in its place, we can recommend either a sinusoidal multitaper estimator (section 3.2.), Welch's overlapped segment averaging estimator (3.3.) or Burg's autoregressive spectral estimator (3.4.). We provide all the formulae needed to compute these estimators. We compare these spectral estimates on the two examples of clock noise shown in figure 1 and find that the estimates are consistent with one another (3.6.). We then describe some of the many uses for the spectrum in characterizing clock noise (4.). We conclude with a brief summary in section 5.

2. Power Spectra for Intrinsically Stationary Processes

Once any sort of deterministic pattern in the time differences between two clocks has been removed (e.g., subtraction of a systematic linear drift between the time kept by the clocks), the resulting clock noise series is inherently stochastic in nature. Accordingly we regard clock noise as a realization of a stochastic process X_t, i.e., a collection of random variables indexed by a time index t. In metrology great care is taken to maintain a stable environment for high precision clocks. As a result, an appropriate class of stochastic processes for modeling clock noise is the class of intrinsically stationary processes. The notion of stationarity posits that certain statistical properties of a process are invariant with regard to shifts in time. Basically, stationarity just requires that the underlying stochastic nature of the process not be changing over time. Even though individual realizations from an intrinsically stationary process certainly vary over time and differ from each

other in their specific patterns, their variations are similar in that they have the same 'look and feel' with the passage of time.

The class of intrinsically stationary processes is a generalization of the class of stationary processes. The latter admits a spectral representation of the form

$$X_t = \mu + \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} e^{\mathrm{i}2\pi f t\,\Delta t}\,\mathrm{d}Z(f), \quad t = \ldots, -2, -1, 0, 1, 2, \ldots, \tag{4}$$

where μ is a constant representing the expected value of X_t; as before, $f_{\mathcal{N}}$ is the Nyquist frequency, which is dictated by the sampling interval Δt; $e^{\mathrm{i}2\pi f t\,\Delta t} = \cos(2\pi f t\,\Delta t) + \mathrm{i}\sin(2\pi f t\,\Delta t)$, with $\mathrm{i} = \sqrt{-1}$; and $Z(f)$ is a complex-valued stochastic process (indexed by the Fourier frequency f) with the very special property of having orthogonal increments. At first glance the above representation looks formidable, but the basic idea is that X_t is a linear combination of sinusoids over a continuum of frequencies. The amplitude that is assigned to the cosine and sine with frequency f is generated by the increment $\mathrm{d}Z(f) = Z(f+\mathrm{d}f) - Z(f)$ from the stochastic process $Z(f)$. Because $\mathrm{d}Z(f)$ is stochastic, different realizations will have different amplitudes assigned to the sinusoids with frequency f. The spectrum tells us how large these amplitudes are likely to be. Formally, for the stationary processes of interest in modeling clock noise, the expected squared magnitude of $\mathrm{d}Z(f)$ is given by $S_X(f)\,\mathrm{d}f$, where $S_X(f)$ is called the spectrum (or spectral density function). The fact that the increments of $\mathrm{d}Z(f)$ are orthogonal (in the sense that the correlation between $\mathrm{d}Z(f)$ and $\mathrm{d}Z(f')$ is zero when $f \neq f'$) leads to the basic result that

$$\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} S_X(f)\,\mathrm{d}f = \sigma_X^2, \tag{5}$$

where σ_X^2 is the variance of the stationary process (one of the consequences of stationarity is that the variance of X_t does not depend upon t). The above can be regarded as the theoretical analog of equation (2), which involves quantities that can be computed from actual measurements. In words, the spectrum breaks up

the process variance for X_t into components, each of which is associated with a different Fourier frequency f. By examining $S_X(f)$ as a function of frequency, we can determine which sinusoids are important in constructing X_t via its spectral representation.

In preparation for defining an intrinsically stationary process, we need the following result, which follows from the theory of linear time invariant filters (see, e.g., [30, 31, 35, 37] for details). Suppose X_t is a stationary process with spectrum $S_X(f)$, and let $X_t^{(1)} = X_t - X_{t-1}$ be its first-order backward difference. Then $X_t^{(1)}$ is also a stationary process, and its spectrum is given by

$$S_{X^{(1)}}(f) = |2\sin(\pi f\,\Delta t)|^2\, S_X(f). \tag{6}$$

If we take the first-order backward difference of $X_t^{(1)}$, we obtain the second-order backward difference of X_t, namely, $X_t^{(2)} = X_t^{(1)} - X_{t-1}^{(1)} = X_t - 2X_{t-1} + X_{t-2}$. The process $X_t^{(2)}$ is stationary with spectrum

$$S_{X^{(2)}}(f) = |2\sin(\pi f\,\Delta t)|^2\, S_{X^{(1)}}(f) = |2\sin(\pi f\,\Delta t)|^4\, S_X(f). \tag{7}$$

The obvious generalization to the dth order backward difference of X_t says that the spectrum for $X_t^{(d)}$ is given by

$$S_{X^{(d)}}(f) = |2\sin(\pi f\,\Delta t)|^{2d}\, S_X(f). \tag{8}$$

The class of intrinsically stationary process is constructed by considering what can be regarded as the inverse operation to differencing, namely, cumulative summation. For example, suppose we define

$$W_t = \begin{cases} C + \sum_{u=1}^{t} X_u, & t = 1, 2, \ldots; \\ C, & t = 0 \text{ and} \\ C - \sum_{u=0}^{|t|-1} X_{-u}, & t = -1, -2, \ldots, \end{cases} \tag{9}$$

where C is random variable with variance $\sigma_C^2 \geq 0$ (we assume C is uncorrelated with each X_t). If X_t is stationary with variance $\sigma_X^2 > 0$ and spectrum $S_X(f)$, then

the process W_t cannot be stationary since this would dictate that the variance of W_t is finite and the same for all t; however, the variances for W_0 and $W_{\pm 1}$ are, respectively, σ_C^2 and $\sigma_C^2 + \sigma_X^2$. By construction, the first-order backward difference $W_t^{(1)}$ of W_t is just X_t. Since the process $W_t^{(1)}$ is stationary with spectrum $S_X(f)$, equation (6) suggests a way of *defining* a spectrum for W_t. Formally, we replace X with W in that equation to obtain $S_{W^{(1)}}(f) = |2\sin(\pi f\,\Delta t)|^2\, S_W(f)$, which yields $S_X(f) = |2\sin(\pi f\,\Delta t)|^2\, S_W(f)$. The desired definition is

$$S_W(f) = \frac{S_X(f)}{|2\sin(\pi f\,\Delta t)|^2}. \tag{10}$$

More generally, if W_t is any nonstationary process whose dth order backward differences form a stationary process X_t with spectrum $S_X(f)$, then the spectrum for W_t is defined to be

$$S_W(f) = \frac{S_X(f)}{|2\sin(\pi f\,\Delta t)|^{2d}}. \tag{11}$$

In this case the process W_t is said to be a dth order intrinsically stationary process (stationary processes can be considered to be intrinsically stationary with order zero). Formally we have

$$\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} S_W(f)\,\mathrm{d}f = \infty, \tag{12}$$

which, when compared to equation (5), suggests that we assign an infinite variance to a nonstationary process with stationary backward differences (this viewpoint is consistent with the construction given in equation (9) if we take σ_C^2 to be infinite). The results of Yaglom [48] can be used to argue that the spectrum for W_t as defined above is fully consistent with the definition of the spectrum for a stationary process. In particular, if W_t is used as the input to an ideal bandpass filter of narrow bandwidth Δf centered at frequency f, then the variance of the output is given by $2S_W(f)\,\Delta f$ to an approximation that improves as $\Delta f \to 0$.

The above theory provides us with a well-defined spectrum for the class of intrinsically stationary processes, members of which have proven to be useful

models for clock noise. The task that now faces us is to take a set of clock measurements that we assume to be a realization of a portion $X_0, X_1, \ldots, X_{N-1}$ of an intrinsically stationary process with spectrum $S_X(f)$ and to use these measurements to estimate the spectrum. The efficacy of the methodology that we describe in the next section is supported by computer experiments [14, 28, 29] and some theoretical work that has appeared over the last decade [12, 21, 45, 46].

3. Estimation of Clock Noise Spectra

In the subsections that follow, we briefly describe some well-known estimators of the spectrum and state all the formulae that are needed to compute them (sections 3.1. to 3.5.), after which we comment upon how they perform on the clock noise series shown in figure 1 (section 3.6.). All the estimators are intended to be used on a time series X_t after it has been centered by subtracting off its sample mean. We denote the centered series by

$$\widetilde{X}_t = X_t - \overline{X}, \quad t = 0, 1, \ldots, N-1, \text{ where, as before, } \overline{X} = \frac{1}{N} \sum_{t=0}^{N-1} X_t. \tag{13}$$

Each estimator makes use of the discrete Fourier transform (DFT). By definition the DFT of the time series $W_0, W_1, \ldots, W_{N'-1}$ is given by

$$\Delta t \sum_{t=0}^{N'-1} W_t e^{-\mathrm{i}2\pi t j/N'}, \quad j = 0, 1, \ldots, N'-1, \tag{14}$$

where Δt is the sampling interval for W_t. The jth element of the DFT is associated with the Fourier frequency $f'_j = j/(N'\,\Delta t)$. The DFT can be efficiently computed using a fast Fourier transform (FFT) algorithm, for which Δt is often standardized to be either unity or $1/N'$. The most common form of this algorithm presumes that N' is a power of two; i.e., $N' = 2^J$ for some positive integer J. In the formulae that we give in the following subsections, we will *not* assume that the sample size N is a power of two, but we will restrict ourselves to using a 'power of two' FFT algorithm. We accomplish this by 'zero padding' the centered series

$\widetilde{X}_t$. Accordingly, let

$$\widetilde{X}'_t = \begin{cases} \widetilde{X}_t, & t = 0, 1, \ldots, N-1, \text{ and} \\ 0, & t = N, N+1, \ldots. \end{cases} \tag{15}$$

In what follows, if N is not a power of two, we let N' be any power of two such that $N' > N$; if N is a power of two, we can take N' to be the same as N. When $N' > N$, then $\widetilde{X}'_t = 0$ for $t = N, N+1, \ldots, N'-1$.

3.1. A Naive Estimator: The Periodogram

A basic estimator of the spectrum is the periodogram, which is defined as

$$\hat{S}_X^{(p)}(f) = \frac{\Delta t}{N} \left| \sum_{t=0}^{N-1} \widetilde{X}_t e^{-\mathrm{i}2\pi f t\, \Delta t} \right|^2, \quad |f| \le f_{\mathcal{N}}. \tag{16}$$

The above can be computed using

$$\hat{S}_X^{(p)}(f'_j) = \frac{\Delta t}{N} \left| \sum_{t=0}^{N'-1} \widetilde{X}'_t e^{-\mathrm{i}2\pi t j/N'} \right|^2, \quad j = 0, 1, \ldots, \frac{N'}{2}, \tag{17}$$

where the right-hand side involves the first $\frac{N'}{2}+1$ values in the DFT of the centered and zero padded series $\widetilde{X}'_t$. Note that $f'_0 = 0$ and $f'_{N'/2} = f_{\mathcal{N}}$ so that f'_j defines a grid of $\frac{N'}{2} + 1$ equally spaced frequencies that partition the interval $[0, f_{\mathcal{N}}]$ into $N'/2$ subintervals of equal size.

While the periodogram does satisfy the basic constraint of equation (2), it can suffer from a phenomenon known as leakage, in which portions of the estimated spectrum are artificially elevated. The frequencies at which leakage can occur are portions of $\hat{S}_X^{(p)}(f)$ that do not contribute very much to the sample variance. The periodogram is thus capable of describing what is going on at frequencies that are the dominant contributors to the sample variance, but it can be very misleading elsewhere. In addition, the periodogram is inherently quite noisy looking, making it necessary to smooth it somehow so that we can see what it might be telling us.

As examples, the left-hand column of figure 2 shows periodograms for the clock noise series $\widetilde{X}_{1,t}$ (upper plot) and $\widetilde{X}_{2,t}$ (lower). Since $N = 4000$ for both

series, we set $N' = 4096$ and plot $\hat{S}_X^{(p)}(f'_j)$ versus f'_j for $j = 1, \ldots, N'/2$ on log/log scales (we cannot display $\hat{S}_X^{(p)}(f'_0)$ versus f'_0 on a log/log scale for two reasons: $f'_0 = 0$ and $\hat{S}_X^{(p)}(f'_0) = 0$ due to centering). Both periodograms begin at $f'_1 = 1/(N'\,\Delta t)$ and end at $f'_{2048} = f_{\mathcal{N}} = 1/(2\,\Delta t)$, but the spanned frequencies are different because the sampling interval Δt is 1 second for the first series and 60 seconds for the other, leading to $f_{\mathcal{N}} = 0.5$ Hz (cycles per second) and 0.0083 Hz, respectively. Both periodograms look very noisy, so we show smoothed versions $\bar{S}_X^{(p)}(f'_j)$ in the right-hand column (thin jagged curves). Along with the smoothed periodograms are shown corresponding Burg autoregressive spectral estimates (thick smoother curves), which are described in more detail in section 3.4.. Note that the smoothed periodograms are elevated systematically above Burg's estimate over certain frequencies. For $\widetilde{X}_{1,t}$, the elevation is confined to frequencies close to the Nyquist frequency $f_{\mathcal{N}} = 0.5$ Hz, but, for for $\widetilde{X}_{2,t}$, it is spread out over a wider range of frequencies. The systematic elevation is greater than an order of magnitude at some frequencies. These discrepancies can be attributed to leakage in the periodogram; i.e., whereas Burg's estimate can be considered to be largely free of bias, the periodogram appears to be badly biased at certain frequencies.

The next two subsections consider two other estimators that, like the Burg estimator, can get around the limitations of the periodogram.

3.2. Multitaper Spectral Estimation

As we noted in the previous section, one problem with the periodogram is that it can suffer from leakage. One way to alleviate leakage is to use a data taper. figure 3 shows the Hanning data taper

$$h_{N,t} = \left(\frac{2}{3(N+1)}\right)^{1/2} \left[1 - \cos\left(\frac{2\pi(t+1)}{N+1}\right)\right], \quad t = 0, 1, \ldots, N-1, \qquad (18)$$

along with an example of a tapered series $h_{N,t}\widetilde{X}_{1,t}$ using the centered clock noise series $\widetilde{X}_{1,t}$ shown in the top of figure 1. Tapering essentially forces the beginning

and end of a time series to damp down smoothly to zero, thus forcing the extremes of the series to match up with each other. To obtain an estimator of the spectrum that does not suffer from leakage to the degree that the periodogram does, we define

$$h'_{N,t} = \begin{cases} h_{N,t}, & t = 0, 1, \ldots, N-1, \text{ and} \\ 0, & t = N, N+1, \ldots, \end{cases} \tag{19}$$

and use $h'_{N,t}\widetilde{X}'_t$ in place of $\widetilde{X}'_t$ in equation (17), with $\Delta t/N$ replaced by just Δt (the division by N has been eliminated because of the manner in which $h_{N,t}$ is normalized).

Why tapering helps alleviate leakage is based upon the following chain of thought. First, the periodogram for a time series is proportional to the squared amplitudes of its Fourier representation. Second, the Fourier extrapolation of a time series outside of its measured values is periodic with a period of N so that, e.g., $\widetilde{X}_{-1}$ and $\widetilde{X}_N$ are extrapolated to be, respectively, $\widetilde{X}_{N-1}$ and $\widetilde{X}_0$. Third, a discontinuity either in the time series or in its extrapolation can disrupt the amplitudes of its Fourier representation over a range of frequencies. This disruption is one explanation for leakage. Tapering essentially helps to eliminate discontinuities between a time series and its periodic extrapolation by forcing both ends of the centered series towards zero. Both clock noise series shown in figure 1 are such that $\widetilde{X}_0$ and $\widetilde{X}_{N-1}$ are not reasonable extrapolations for $\widetilde{X}_N$ and $\widetilde{X}_{-1}$.

While tapering helps eliminate leakage, it arguably does so by reducing the sample size of the time series (compare the top plot of figure 1 with the bottom plot of figure 3). The information that is lost by using a single taper can be recovered by using more than one taper, leading to a multitaper spectral estimator [44, 35, 38]. The idea behind multitapering is to use a set of K orthonormal data tapers $\{h_{k,N,t}\}$, $k = 0, \ldots, K-1$, where the orthonormality stipulation means that

$$\sum_{t=0}^{N-1} h_{k,N,t} h_{l,N,t} = \begin{cases} 1, & \text{if } k = l; \\ 0, & \text{if } k \neq l. \end{cases} \qquad 0 \le k, l \le K-1. \tag{20}$$

With $h'_{k,N,t}$ defined in terms of $h_{k,N,t}$ as per equation (19), the multitaper spectral estimator is given by

$$\hat{S}_X^{(mt)}(f'_j) = \frac{\Delta t}{K} \sum_{k=0}^{K-1} \left| \sum_{t=0}^{N'-1} h'_{k,N,t} \widetilde{X}'_t e^{-\mathrm{i}2\pi tj/N'} \right|^2, \quad j = 0, 1, \ldots, \frac{N'}{2}. \tag{21}$$

Note that the summation within the absolute value involves the first $\frac{N'}{2}+1$ values in the DFT of the centered, tapered and zero padded series $h'_{k,N,t}\widetilde{X}'_t$.

A convenient set of multitapers to use is the family of sinusoidal tapers. These tapers are easily computed since they are given by

$$h_{k,N,t} = \left(\frac{2}{N+1}\right)^{1/2} \sin\left(\frac{(k+1)\pi(t+1)}{N+1}\right), \quad t = 0, \ldots, N-1. \tag{22}$$

The left-hand column of figure 4 shows these tapers for $k = 0, 1, \ldots, 5$. The right-hand column shows the product of $h_{k,N,t}$ and the clock noise series $\widetilde{X}_{1,t}$. The left-hand column of figure 5 shows the $K = 6$ sinusoidal multitaper estimates for $\widetilde{X}_{1,t}$ (upper plot) and $\widetilde{X}_{2,t}$ (lower).

Statistical theory suggests that a 95% confidence interval for the unknown $S_X(f)$ is given by

$$\left[\frac{\nu \hat{S}_X^{(mt)}(f)}{Q_\nu(0.975)}, \frac{\nu \hat{S}_X^{(mt)}(f)}{Q_\nu(0.025)}\right], \tag{23}$$

where here $\nu = 2K$, and $Q_\nu(p)$ is the percentage point such that the probability of a chi-square random variable with ν degrees of freedom being less than $Q_\nu(p)$ is p (when $K = 6$ so that $\nu = 2K = 12$, we have $Q_{12}(0.025) \doteq 4.404$ and $Q_{12}(0.975) \doteq 23.337$). The width of such a confidence interval on a log/log plot is independent of the actual value of $\hat{S}_X^{(mt)}(f)$. The confidence interval based upon a fictitious estimate $\hat{S}_X^{(mt)}(0.1) = 10^{-18}$ is shown in the upper left-hand plot of figure 5 and can be used to roughly gauge the variability we can expect the multitaper estimator to have. In addition, for $0 < f'_j < f'_{j'} < f_{\mathcal{N}}$, we can expect the estimators $\hat{S}_X^{(mt)}(f'_j)$ and $\hat{S}_X^{(mt)}(f'_{j'})$ to be approximately uncorrelated as long as the separation between frequencies $f'_{j'} - f'_j$ is at least as large as the bandwidth of the multitaper estimator, which is given by $\frac{K+1}{(N+1)\,\Delta t}$.

The choice of K dictates the statistical properties of $\hat{S}_X^{(mt)}(f)$. As K increases, Equation (21) says that more averaging is used to form the estimator, which results in a decrease in the variance of $\hat{S}_X^{(mt)}(f)$ and in a tighter confidence interval for the unknown $S_X(f)$. Since the variance of $\hat{S}_X^{(mt)}(f)$ is proportional to $1/K$, there is a rapid drop in variability as we initially increase K beyond unity. Setting K to be in the range of 5 to 10 yields an estimator with a reduction in variance that is a clear improvement over the variability inherent in the periodogram (compare the left-hand plots in figures 2 and 5). As K increases, the bandwidth of the estimator also increases, which unfortunately can offset some of the advantages of decreased variability. In particular, if the true spectrum has a sharp peak, increasing the bandwidth will broaden the peak in the estimated spectrum, thus increasing uncertainty in its location. If the bandwidth is too large, two peaks that are separated by less than the bandwidth can be smeared together by the estimator. Restricting K to be in the range of 5 to 10 has the additional benefit of keeping the bandwidth small enough so that the estimated spectrum reliably preserves sharp features in most applications.

3.3. Welch's Overlapped Segment Averaging Spectral Estimation

While multitapering is one strategy for obtaining a spectral estimator with better properties than the periodogram, another commonly used approach is Welch's overlapped segment averaging (WOSA) spectral estimator [47]. The idea is to break the centered time series $\widetilde{X}_t$ up into overlapping subseries of equal length, to apply a common data taper to each subseries and then to form an estimator based upon averaging the squared magnitudes of the DFTs of each tapered subseries. Let N_S represent the length of each subseries, and let K be the number of subseries, with the individual subseries being indexed by $k = 0, 1, \ldots, K-1$. Figure 6 illustrates the tapering of $K = 6$ subseries from $\widetilde{X}_{1,t}$, each of length

$N_S = 1024$. Here we take the common taper to be the Hanning data taper $h_{N_S,t}$ of equation (18). The kth subseries consists of $\widetilde{X}_{1,t_k}, \widetilde{X}_{1,t_k+1}, \ldots, \widetilde{X}_{1,t_k+N_S-1}$, where the starting indices are given by

$$t_k = \left\lfloor k \left(\frac{N - N_S}{K - 1} \right) \right\rfloor, \quad k = 0, 1, \ldots, K-1 \tag{24}$$

(in the above, $\lfloor x \rfloor$ is the greatest integer that is less than or equal to x). The left-hand column of plots shows the Hanning data taper aligned to $t_0 = 0$, $t_1 = 595$, $\ldots$, $t_5 = 2976$. The right-hand column shows the product of each aligned $h_{N_S,t}$ with $\widetilde{X}_{1,t}$. Note that, except at the extremes of the time series, values that happen to be severely tapered in one subseries are relatively unaltered in an adjacent subseries. For the Hanning data taper, it is recommended that, once N_S has been selected, the number of subseries K be set so that the overlap between adjacent subseries is about 50%. For the example, the overlap is $1 - \frac{N-N_S}{N_S(K-1)} = 41.9\%$.

Since we can select the subseries length N_S to be a power of two, it is easy to formulate the WOSA estimator to take advantage of an FFT algorithm that requires a power of two. For a bit more generality, we formulate the estimator in terms of N', which we take to be any power of two satisfying $N' \geq N_S$. With $h'_{N_S,t}$ defined in terms of $h_{N_S,t}$ as per equation (19), the WOSA spectral estimator is given by

$$\hat{S}_X^{(wosa)}(f'_j) = \frac{\Delta t}{K} \sum_{k=0}^{K-1} \left| \sum_{t=0}^{N'-1} h'_{N_S,t} \widetilde{X}'_{t+t_k} e^{-\mathrm{i}2\pi t j / N'} \right|^2, \quad j = 0, 1, \ldots, \frac{N'}{2}. \tag{25}$$

If N' is set to the same power of two in the above and in equation (21), the WOSA and multitaper estimates are computed over the same grid of frequencies.

The middle column of figure 5 shows the WOSA spectral estimates for the clock noise series $\widetilde{X}_{1,t}$ (top plot) and $\widetilde{X}_{2,t}$ (bottom). To get an idea of how much sampling variability there is in the WOSA estimator, the upper plot also shows a 95% confidence interval for the true spectrum at $f = 0.1$ based upon a fictitious estimate of $\hat{S}_X^{(wosa)}(0.1) = 10^{-18}$. This interval is computed as indicated

by equation (23), but we need to replace $\hat{S}_X^{(mt)}(f)$ with $\hat{S}_X^{(wosa)}(f)$ and to set ν using

$$\nu = \frac{2K}{1 + 2\sum_{k=1}^{K-1}\left(1 - \frac{k}{K}\right)\left|\sum_{t=0}^{N_S-1} h_{N_S,t} h_{N_S,t+t_k}\right|^2}. \tag{26}$$

For the present example, this yields $\nu \doteq 11.9$, which is virtually the same as for the multitaper estimates ($\nu = 12$). If the degree of overlap between adjacent subseries is close to 50% (as is the case in our example), then, to a good approximation, the above becomes

$$\nu \approx \frac{2K}{1 + \frac{2(K-1)}{K}\left|\sum_{t=0}^{N_S/2-1} h_{N_S,t} h_{N_S,t+N_S/2}\right|^2} \approx \frac{36K^2}{19K - 1}. \tag{27}$$

The second approximation above yields $\nu \doteq 11.5$. Finally, we note that the bandwidth for the Hanning-based WOSA estimator is approximately $\frac{2}{N_S\,\Delta t}$, which implies that $\hat{S}_X^{(wosa)}(f'_j)$ and $\hat{S}_X^{(wosa)}(f'_{j'})$ with $0 < f'_j < f'_{j'} < f_{\mathcal{N}}$ are uncorrelated to a reasonable approximation as along as $f'_{j'} - f'_j$ is at least as large as this bandwidth.

3.4. Autoregressive Spectral Estimation

The spectral estimators that we have discussed in the previous three subsections are formed based upon one or more DFTs of either part or all of a time series (after centering and possibly tapering). An alternative approach that is quite effective for typical clock noise is to use an autoregressive (AR) process to model X_t. Under this model, we presume that

$$X_t = \mu + \sum_{k=1}^{p} \phi_{p,k}(X_{t-k} - \mu) + \epsilon_t, \tag{28}$$

where μ is a constant representing the expected value of X_t; p is the order of the AR model; $\phi_{p,1}, \phi_{p,2}, \ldots, \phi_{p,p}$ are constants known as the AR coefficients; and ϵ_t is a set of uncorrelated random variables with mean zero and variance σ_p^2 (i.e., ϵ_t

is white noise). An AR process has an associated theoretical spectrum given by

$$S_X(f) = \frac{\sigma_p^2 \, \Delta t}{\left|1 - \sum_{k=1}^{p} \phi_{p,k} e^{-i2\pi f k \, \Delta t}\right|^2}, \quad |f| \le f_{\mathcal{N}}. \tag{29}$$

The scheme is to plug estimates $\hat{\phi}_{p,1}, \hat{\phi}_{p,2}, \ldots, \hat{\phi}_{p,p}$ and $\hat{\sigma}_p^2$ of the p+1 AR parameters into the above to create an AR spectral estimator $\hat{S}_X^{(ar)}(f)$. Note that we can use the DFT to evaluate this estimator over the same grid of frequencies f_k' as before. To do so, define

$$\hat{\phi}'_{p,k} = \begin{cases} 1, & k = 0; \\ -\hat{\phi}_{p,k}, & k = 1, \ldots, p; \text{ and} \\ 0, & k = p+1, p+2, \ldots. \end{cases} \tag{30}$$

Then

$$\hat{S}_X^{(ar)}(f_j') = \frac{\hat{\sigma}_p^2 \, \Delta t}{\left|\sum_{k=0}^{N'-1} \hat{\phi}'_{p,k} e^{-i2\pi k j / N'}\right|^2}, \quad j = 0, \ldots, N'/2. \tag{31}$$

There are many estimators in the literature for the AR parameters. For typical clock noise spectra, we recommend the Burg estimator [11], which is remarkably effective and can be computed efficiently [10, 25, 27, 35, 42]. For $t = 0, \ldots, N-1$, define $f_{0,t} = \widetilde{X}_t$ and $b_{0,t} = \widetilde{X}_t$. Let $A_1 = 2N\hat{\sigma}_X^2 - \widetilde{X}_0^2 - \widetilde{X}_{N-1}^2$ and $\hat{\sigma}_0^2 = \hat{\sigma}_X^2$, where $\hat{\sigma}_X^2$ is the sample variance of X_t. We obtain the estimators $\hat{\phi}_{p,1}, \hat{\phi}_{p,2}, \ldots, \hat{\phi}_{p,p}$ and $\hat{\sigma}_p^2$ by recursively computing, for $l = 1, \ldots, p$,

$$B_l = 2 \sum_{t=l}^{N-1} f_{l-1,t} b_{l-1,t-l} \tag{32}$$

$$\hat{\phi}_{l,l} = B_l / A_l \tag{33}$$

$$\hat{\phi}_{l,k} = \hat{\phi}_{l-1,k} - \hat{\phi}_{l,l} \hat{\phi}_{l-1,l-k}, \;\; 1 \le k \le l-1 \;\; \text{(this is skipped if } l = 1) \tag{34}$$

$$\hat{\sigma}_l^2 = \hat{\sigma}_{l-1}^2 (1 - \hat{\phi}_{l,l}^2) \tag{35}$$

$$f_{l,t} = f_{l-1,t} - \hat{\phi}_{l,l} b_{l-1,t-l}, \;\; l \le t \le N-1 \tag{36}$$

$$b_{l,t-l} = b_{l-1,t-l} - \hat{\phi}_{l,l} f_{l-1,t}, \;\; l \le t \le N-1 \tag{37}$$

$$A_{l+1} = \left(1 - \hat{\phi}_{l,l}^2\right) A_l - f_{l,l}^2 - b_{l,N-l-1}^2 \tag{38}$$

(the last three equations can be skipped when $l = p$). The desired estimators of the $p+1$ AR parameters are available as output at the end of the pth recursion.

Equation (34) is also part of the Levinson–Durbin recursions that are used to form the well-known Yule–Walker estimator of the AR parameters [25, 27, 35, 42].

The remaining task is to set p. Three approaches are to use Akaike's final prediction error (FPE) [1], Akaike's information criterion (AIC) [2] or the Bayesian information criterion (BIC) [40]. All three criteria require that we initially set p to some maximum value, say p', and use the Burg method to fit a model of order p' to the data, keeping track of $\hat{\sigma}_l^2$, $l = 1, \ldots, p'$, as they are computed as part of the Burg recursive equations. The FPE approach consists of plotting

$$\mathrm{FPE}(l) = \frac{N + l + 1}{N - l - 1} \hat{\sigma}_l^2 \tag{39}$$

versus $l = 1, \ldots, p'$ and then setting p to the value of l for which $\mathrm{FPE}(l)$ is the smallest. The AIC approach is similar in spirit, with $\mathrm{FPE}(l)$ being replaced by

$$\mathrm{AIC}(l) = \log\left(\hat{\sigma}_l^2\right) + \frac{2l}{N} \tag{40}$$

(a 'bias corrected' version of the AIC has also been proposed [43, 22], which consists of replacing $2l/N$ with $2(l+1)/(N-l-2)$). The BIC approach makes use of

$$\mathrm{BIC}(l) = \log\left(\hat{\sigma}_l^2\right) + \frac{2\log(N)}{N} \tag{41}$$

(for the AIC and BIC, 'log' should interpreted as 'log base e'). In practice, caution must be exercised in using these criteria since they do not always yield satisfactory choices for p, but they are often good starting points.

The right-hand column of figure 5 shows the Burg AR spectral estimates for the two clock noise series. With p' set to 100, we used the FPE criterion to set $p = 5$ and 25 for, respectively, $\widetilde{X}_{1,t}$ and $\widetilde{X}_{2,t}$. While the two estimated AR spectra are much smoother looking than either the multitaper or WOSA estimates, it is not easy to assess the effect of sampling variability, and hence we cannot display a 95% confidence interval here as we could in the left-hand and middle columns. It is also not easy to give a measure of the bandwidth of $\hat{S}_X^{(ar)}(f)$. Despite these

limitations, the Burg AR spectral estimator is guaranteed to integrate to the sample variance as per equation (2).

3.5. Prewhitening and Postcoloring

A technique that can potentially improve the spectral estimators we have discussed so far is prewhitening. The basic idea is to filter a time series so that the output from the filter is closer to white noise than the unfiltered series. We then perform a spectral analysis on the filtered series and estimate the spectrum for the original (unfiltered) series using the estimate based on the filtered series after it has been adjusted for the effect of the prewhitening filter. The rationale for this scheme is that the spectrum for white noise is much easier to estimate for any other process since it is just a constant proportional to the variance of the process. When presented with white noise, all of the estimators we have considered so far are unbiased estimators of the true spectrum. In practice, prewhitening often works well even if the spectrum for the filtered series is still markedly different from white noise.

For processes whose spectra are dominated by low frequency components, a commonly used prewhitening filter is a first-order backward difference. Such a filter damps down low frequency components considerably and is useful for intrinsically stationary processes as long as the spectrum for the differenced process does not damp down to zero too fast as $f \to 0$. In the context of clock noise, the first-order backward difference $X_t^{(1)} = X_t - X_{t-1}$ is proportional to the fractional frequency deviation $\overline{Y}_t = (X_t - X_{t-1})/\Delta t$, so we will consider the latter instead of $X_t^{(1)}$. The spectra for $\overline{Y}_t$ and X_t are related by

$$S_{\overline{Y}}(f) = \frac{4\sin^2(\pi f\,\Delta t)}{(\Delta t)^2} S_X(f). \tag{42}$$

Once we have estimated $S_{\overline{Y}}(f)$ using, say, $\hat{S}_{\overline{Y}}(f)$, we can estimate $S_X(f)$ using

$$\hat{S}_X(f) = \frac{(\Delta t)^2}{4\sin^2(\pi f\,\Delta t)}\hat{S}_{\overline{Y}}(f). \tag{43}$$

Figure 7 shows multitaper, WOSA and AR spectral estimates based upon $\overline{Y}_t$ for the two clock noise series $\widetilde{X}_{1,t}$ and $\widetilde{X}_{2,t}$ (the plots are arranged in the same manner as for figure 5). These estimates were obtained by using $\overline{Y}_t$ (after centering and possible zero padding) in place of $\widetilde{X}_t$ or $\widetilde{X}'_t$ in all the equations stated previously in this section. For the AR estimates, the FPE criterion picked out $p = 4$ and 24 for the two fractional frequency series, in contrast to 5 and 25 for $\widetilde{X}_{1,t}$ and $\widetilde{X}_{2,t}$. Whereas the spectra in figure 5 have variations ranging over 7 or 8 decades, the ones in figure 7 range only over about 2 decades, so the first-order backward difference filter has indeed had a whitening effect. There is, however, some concern with the AR estimate corresponding to $\widetilde{X}_{1,t}$ (upper right-hand plot). The corresponding multitaper and WOSA estimates show the spectra to be decreasing as $f \to 0$, whereas the AR estimate flattens out. This points out a potential problem with AR spectral estimates, namely that, whereas they have been demonstrated to work well with processes whose spectra increase as $f \to 0$, they are also known to perform poorly in comparison to multitaper and WOSA estimators for processes whose spectra decrease rapidly as $f \to 0$. (In fact, prewhitening is rarely used with AR spectral estimation, but, rather, low-order AR models are often fit to a time series to identify a potential prewhitening filter; for more details, see, e.g., [35], section 9.10).

Figure 8 shows the multitaper, WOSA and AR spectral estimates for $S_X(f)$ based upon converting the estimates for $S_{\overline{Y}}(f)$ in figure 7 using equation (43). The agreement amongst the three spectra is quite good, with the exception that the AR estimate for $\widetilde{X}_{1,t}$ (upper right-hand plot) is elevated about half an order of magnitude above the multitaper and WOSA estimates at the very lowest frequency. This discrepancy can be attributed to the deficiency in AR spectral

estimation that we just pointed out. If we compare the multitaper and WOSA spectra in figures 8 with those in figure 5, again the agreement is quite good except at the very lowest frequencies. The ones in figure 5 tend to flatten out, whereas those in figure 8 do not. The tendency of the estimates in figure 5 to flatten out can be attributed to the bandwidth of these estimates, which begin to trap the value $f = 0$ at the very lowest frequencies. Because of the need to center the time series, the estimated spectra are biased toward zero at frequencies that are less than a bandwidth away from $f = 0$. The prewhitening scheme has evidently removed some of this bias, so the multitaper and WOSA spectra in figures 8 are to be preferred over those in figure 5.

3.6. Comparison of Spectral Estimators

Let us now take a closer look at the multitaper and WOSA spectral estimates of figure 8 and the AR spectral estimates of figure 5. Note that the AR spectral estimates are much smoother than the multitaper and WOSA estimates. In principle, the multitaper estimate could be smoothed out by increasing K, and the WOSA also, by increasing K along with a reduction in N_S (we could also decrease the variability in the estimates by smoothing across frequencies); however, there is no compelling reason for smoothing here since the variability in the multitaper and WOSA estimates is not so large that we cannot easily see how the spectra vary with frequency. The advantage of the relatively unsmoothed multitaper and WOSA estimates over the AR estimates is that their statistical properties are tractable, and hence they are suitable for future manipulations (see the discussion in the next section). While the AR spectral estimates are certainly quite good, their statistical properties are difficult to assess without some rather strong assumptions (in particular, that there is no error in the selected order p).

Visually the multitaper, WOSA and AR spectral estimates follow the same

overall patterns as functions of frequency. For a more sensitive comparison, the top row of plots in figure 9 shows, from left to right, the ratios $\hat{S}_X^{(mt)}(f)/\hat{S}_X^{(wosa)}(f)$, $\hat{S}_X^{(mt)}(f)/\hat{S}_X^{(ar)}(f)$ and $\hat{S}_X^{(wosa)}(f)/\hat{S}_X^{(ar)}(f)$ for $X_{1,t}$, while the bottom row is for $X_{2,t}$. Despite the inherently bumpy nature of the multitaper and WOSA estimates, the left-hand column of plots show that, except at a very few isolated frequencies, they differ by no more than a factor of two. The general conclusion that we can draw from figure 9 is that, although the three estimators we have discussed have quite different theoretical foundations, they are all telling us the same story about the spectra of $\widetilde{X}_{1,t}$ and $\widetilde{X}_{2,t}$.

4. Uses for the Spectrum in Assessing Clock Noise

In this section we briefly mention some of the uses for the spectrum $S_X(f)$ and its estimates $\hat{S}_X(f)$ in assessing the performance of clocks. Since X_t involves a comparison between two clocks, the estimated spectrum reflects the performance of both. If one of the clocks in the comparison is regarded as an approximation to a 'noise free' standard, then the spectrum will largely reflect the performance of a clock under test. A direct comparison of the levels of the estimated spectra for two clocks under test can reveal ranges of Fourier frequencies over which one clock is superior. While such comparisons are the most obvious use for estimated spectra, in fact spectra can contribute to the characterization of clock noise in several other ways. (In passing, we note that, when intercomparing three or more clocks, none of which can be regarded as a 'noise free' standard, it is in theory possible to obtain an estimate of the spectrum for an individual clock by examining the cross-spectrum between contemporaneous measurements of $X_{1,t}$ and $X_{2,t}$, both of which involve the individual clock of interest. To make this work, there is a critical assumption, namely, that all three clocks involved in the two sets of measurements are uncorrelated with one another.)

First, estimated spectra can be used to provide a nonparametric method for assessing the variability in time-domain measures of clock instability. A well-known such measure is the Allan variance [3, 5]. For moderate sample sizes N, the distribution of the maximal-overlap estimator of the Allan variance is, to a good approximation, normally distributed with a mean given by the population Allan variance and with a variance that is proportional to

$$\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \mathcal{A}_\tau(f) S_X^2(f)\,\mathrm{d}f, \tag{44}$$

where $\mathcal{A}_\tau(f)$ can be computed based upon knowledge of the filter whose squared output forms the basis for the maximal-overlap estimator at averaging time τ [33, 36]. A nonparametric method for assessing the variability in the estimated Allan variance is to use $\hat{S}_X(f)$ in place of $S_X(f)$ in the above. The current procedure for assessing variability is parametric and involves associating X_t with a canonical power-law process [17]. This nonparametric approach is simpler, is asymptotically correct, and avoids the canonical power-law assumptions behind the parametric approach. (The above remarks hold for time-domain characterizations of clock noise other than the Allan variance.)

Second, the estimated spectra can be used to assess the power-law assumption underlying the statistical assessment of most time-domain characterizations of clock noise. A glance at the estimated spectra for the two examples we have examined shows that, while the spectrum for $\widetilde{X}_{1,t}$ might be well-modeled as a piece-wise power-law process, the spectrum for $\widetilde{X}_{2,t}$ is not, particularly, for frequencies between $f_{\mathcal{N}}/2$ and $f_{\mathcal{N}}$.

Third, estimates of the spectrum can be used to assess the validity of the hypothesis that one of the canonical power-law spectra (i.e., $S_X(f) = |f|^\alpha$, with $\alpha = 0$, -1, -2, -3 or -4) is tenable as a spectral model over a certain range of frequencies for a given set of clock noise measurements. Here the fact that the multitaper and WOSA estimators have tractable statistical properties plays an

important role because it is possible to compute an estimate $\hat{\alpha}$ of α based upon a linear least squares fit of $\log(\hat{S}_X(f))$ versus $\log(f)$ and to assess the variability in $\hat{\alpha}$ [28, 29, 34]. Using statistical methodology to assess the validity of the canonical power-law hypothesis has generally been ignored in practice, but it is important to test as much as possible the assumptions underlying any statistical analysis to avoid incorrect inferences.

Fourth, the development of new statistical measures for characterizing clock noise in the time domain continues to be an area of active research, which is evidence that no existing time domain measure is fully satisfactory (see, e.g., [20]). To date, all such measures are similar to the Allan variance in the sense of being based on the squared outputs from non-ideal band-pass filters. Comparison between old and new measures would be easier if there were a 'gold standard.' In the frequency domain, there is an obvious choice, namely, an ideal band-pass variance, which by definition would be just the integral of the spectrum $S_X(f)$ over the relevant pass-band [39, 32]. New and existing time domain measures can be evaluated as estimators of this target variance in terms of, e.g., bias, variance and mean squared error. This approach would lead to a viable way of objectively determining whether or not one measure is superior to another when analyzing a certain type of clock noise.

5. Summary

We have demonstrated that three different spectral estimators (multitaper, WOSA and autoregressive) yield basically the same results when applied to two representative clock noise series. Clock noise spectra can thus be successfully tackled using any of these estimators, even though the estimation of power spectra is in general complicated due to the wide variety of spectra that are encountered in the engineering and physical sciences. The theoretical underpinnings of these three estimators are

quite different. The fact that the end results are in such good agreement reassures us that our estimates are on target. Given a routine that efficiently computes the discrete Fourier transform, we have given all the equations needed to fully implement all three estimators. Finally, we have noted some important uses for spectra and their estimates in the statistical characterization of clock noise, both in the frequency and time domains.

In passing we note that the spectral analyses we have presented demonstrate the effectiveness of autoregressive processes as models for clock noise. The problem of forming a time scale based upon an ensemble of clocks can be formulated as a Kalman filter (see, e.g., [16] and references therein), but this approach has been problematic in that such a filter cannot easily handle the canonical power-law processes with $\alpha = -1$ and $\alpha = -3$ (phase and frequency flicker noises). The Kalman filter, however, can readily accommodate autoregressive processes [24], so a promising avenue of research would be to use just autoregressive processes (with appropriate parameter estimation) as clock noise models.

6. Software and Acknowledgments

A collection of functions in the freeware statistical package R that implements the procedures described here is available upon request from the author. The R package itself can be downloaded from the Comprehensive R Archive Network at `http://cran.r-project.org/`.

The author would like to thank (i) the National Institute of Standards and Technology, Time and Frequency Division, for support and (ii) Demetrios Matsakis, Lara Schmidt and Sam Stein for the clock noise data.

References

1. Akaike H 1970 Statistical predictor identification *Annals of the Institute of Statistical Mathematics* **22** 203–17

2. Akaike H 1974 A new look at the statistical model identification *IEEE Transactions on Automatic Control* **19** 716–22

3. Allan D W 1966 Statistics of atomic frequency standards *Proceedings of the IEEE* **54** 221–30

4. Anderson T W 1971 *The Statistical Analysis of Time Series* (New York: John Wiley & Sons) 704 p.

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

6. Blackman R B and Tukey J W 1958 *The Measurement of Power Spectra* (New York: Dover Publications) 190 p.

7. Bloomfield P 2000 *Fourier Analysis of Time Series: An Introduction* 2nd edn (New York: John Wiley & Sons) 261 p.

8. Brillinger D R 1981 *Time Series: Data Analysis and Theory* expanded edn (San Francisco: Holden–Day) 540 p.

9. Brockwell P J and Davis R A 1991 *Time Series: Theory and Methods* 2nd edn (New York: Springer–Verlag) 577 p.

10. Brockwell P J and Davis R A 2002 *Introduction to Time Series and Forecasting* 2nd edn (New York: Springer) 434 p.

11. Burg J P 1968 A new analysis technique for time series data *NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics*; also in Childers D G 1978 *Modern Spectrum Analysis* (New York: IEEE Press) 334 p.

12. Chan G, Hall P and Poskitt D S 1995 Periodogram-based estimators of fractal properties *Annals of Statistics* **23** 1684–711

13. Chatfield C 2004 *The Analysis of Time Series: An Introduction* 6th edn (Boca Raton, LA: Chapman & Hall/CRC) 333 p.

14. Fougere P F 1985 On the accuracy of spectrum analysis of red noise processes using maximum entropy and periodogram methods: simulation studies and application to geophysical data *Journal of Geophysical Research* **90** 4355–66

15. Fuller W A 1996 *Introduction to Statistical Time Series* 2nd edn (New York: John Wiley & Sons) 698 p.

16. Galleani L and Tavella P 2003 On the use of the Kalman filter in timescales *Metrologia* **40** S326–34

17. Greenhall C A 1991 Recipes for degrees of freedom of frequency stability estimators *IEEE Transactions on Instrumentation and Measurement* **40** 994–9.

18. Grenander U and Rosenblatt M 1984 *Statistical Analysis of Stationary Time Series* 2nd edn (New York: Chelsea Publishing Company) 308 p.

19. Hamilton J D 1994 *Time Series Analysis* (Princeton, NJ: Princeton University Press) 799 p.

20. Howe D A 2006 TheoH: a hybrid, high-confidence statistic that improves on the Allan deviation *Metrologia* under revision

21. Hurvich C M and Ray B K 1995 Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes *Journal of Time Series Analysis* **16** 17–41

22. Hurvich C M and Tsai C–L 1989 Regression and time series model selection in small samples *Biometrika* **76** 297–301

23. Jenkins G M and Watts D G 1968 *Spectral Analysis and Its Applications* (San Francisco, CA: Holden–Day) 525 p.

24. Jones R H 1980 Maximum likelihood fitting of ARMA Models to time series with missing observations *Technometrics* **22** 389–95

25. Kay S M 1988 *Modern Spectral Estimation: Theory and Application* (Englewood Cliffs, NJ: Prentice–Hall) 543 p.

26. Koopmans L H 1995 *The Spectral Analysis of Time Series* 2nd edn (San Diego, CA: Academic Press) 366 p.

27. Marple S L Jr 1987 *Digital Spectral Analysis with Applications* (Englewood Cliffs, NJ: Prentice–Hall) 492 p.

28. McCoy E J, Walden A T and Percival D B 1998 Multitaper spectral estimation of power law processes *IEEE Transactions on Signal Processing* **46** 655–68.

29. Nichols–Pagel G A, Percival D B and Reinhall P G 2006 Should structure functions be used to estimate power laws in turbulence? A comparative study *Physica D* under revision

30. Oppenheim A V and Schafer R W 1989 *Discrete-Time Signal Processing* (Englewood Cliffs, NJ: Prentice–Hall) 879 p.

31. Papoulis A and Pillai S U 2002 *Probability, Random Variables, and Stochastic Processes* 4th edn (New York: McGraw–Hill) 304 p.

32. Percival D B 1991 Characterization of frequency stability: frequency domain estimation of stability measures *Proceedings of the IEEE* **79** 961–72

33. Percival D B 1995 On estimation of the wavelet variance *Biometrika* **82** 619–31

34. Percival D B 2003 Stochastic models and statistical analysis for clock noise *Metrologia* **40** S289–304

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

36. Percival D B and Walden A T 2000 *Wavelet Methods for Time Series Analysis* (Cambridge, UK: Cambridge University Press) 594 p.

37. Priestley M B 1981 *Spectral Analysis and Time Series* (London: Academic Press) 890 p.

38. Riedel K S and Sidorenko A 1995 Minimum bias multiple taper spectral estimation *IEEE Transactions on Signal Processing* **43** 188–95

39. Rutman J 1978 Characterization of phase and frequency instabilities in precision frequency sources: fifteen years of progress *Proceedings of the IEEE* **66** 1048–75

40. Schwarz G 1978 Estimating the dimension of a model *Annals of Statistics* **6** 461–4

41. Shumway R H and Stoffer D S 2006 *Time Series Analysis and Its Application* 2nd edn (New York: Springer) 592 p.

42. Stoica P and Moses R L 1997 *Introduction to Spectral Analysis* (Upper Saddle River, NJ: Prentice Hall) 319 p.

43. Sugiura N 1978 Further analysis of the data by Akaike's information criterion and the finite corrections *Communications in Statistics: Theory and Methods* **7** 13–26

44. Thomson D J 1982 Spectrum estimation and harmonic analysis *Proceedings of the IEEE* **70** 1055–96

45. Velasco C 1999 Gaussian semiparametric estimation of non-stationary time series *Journal of Time Series Analysis* **20** 87–127

46. Velasco C 1999 Non-stationary log-periodogram regression *Journal of Econometrics* **91** 325–71

47. Welch P D 1967 The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms *IEEE Transactions on Audio and Electroacoustics* **15** 70–3

48. Yaglom A M 1958 Correlation theory of processes with random stationary nth increments *American Mathematical Society Translations (Series 2)* **8** 87–141

49. Yaglom A M 1987 *Correlation Theory of Stationary and Related Random Functions, Volume I: Basic Results* (New York: Springer–Verlag) 526 p.

Received on xx xxxx 2006.

Figure captions

1. Two examples of clock noise. The top plot shows the time differences $\widetilde{X}_{1,t}$ between two atomic clocks taken once per second for 4000 seconds (data collected using a TSC 5110A analyzer and courtesy of Dr. Samuel R. Stein, Timing Solutions Corporation). The bottom plot shows $\widetilde{X}_{2,t}$, the time differences between two hydrogen masers sampled once per minute for 4000 minutes (data courtesy of Drs. Lara S. Schmidt and Demetrios Matsakis, US Naval Observatory). This series was extracted from a much longer segment that was adjusted for a linear trend. Both series have been centered so that their sample means are zero. The units for $\widetilde{X}_{1,t}$ and $\widetilde{X}_{2,t}$ are in nanoseconds.

2. Periodograms $\hat{S}_X^{(p)}(f)$ for clock noise series $\widetilde{X}_{1,t}$ and $\widetilde{X}_{2,t}$ (left-hand plots) and smoothed periodograms $\bar{S}_X^{(p)}(f)$ and Burg autoregressive spectral estimates $\hat{S}_X^{(ar)}(f)$ (right-hand). The spectral estimates for $\widetilde{X}_{1,t}$ are in the upper row, and those for $\widetilde{X}_{2,t}$, in the lower row. In the right-hand plots, the thicker (and smoother looking) curves are the Burg estimates.

3. Hanning data taper $h_{N,t}$ (top plot) and the result of multiplying it point by point with the centered clock noise series $\widetilde{X}_{1,t}$ to form the tapered series $h_{N,t}\widetilde{X}_{1,t}$ (bottom).

4. Sinusoidal tapers $h_{k,N,t}$ for $k = 0, 1, \ldots, 5$ (left-hand column, top to bottom) and corresponding tapered and centered series $h_{k,N,t}\widetilde{X}_{1,t}$ (right-hand) for $\widetilde{X}_{1,t}$.

5. Sinusoidal multitaper $\hat{S}_X^{(mt)}(f)$ (left-hand plots), WOSA $\hat{S}_X^{(wosa)}(f)$ (middle) and Burg autoregressive $\hat{S}_X^{(ar)}(f)$ (right) spectral estimates for the clock noise series $\widetilde{X}_{1,t}$ (top row) and $\widetilde{X}_{2,t}$ (bottom). The multitaper estimates are based upon $K = 6$ data tapers. The WOSA estimates are based upon

$K = 6$ subseries, each of length $N_S = 1024$. The model orders for the Burg estimates are $p = 5$ and $p = 25$ for, respectively, $\widetilde{X}_{1,t}$ and $\widetilde{X}_{2,t}$. The left-hand and middle plots also show what a typical 95% confidence interval for the true spectrum looks like when based upon the spectral estimates in the plots.

6. Aligned Hanning data tapers $h_{N_S,t}$ (left-hand column) and the product of the aligned tapers with $\widetilde{X}_{1,t}$ (right-hand).

7. Sinusoidal multitaper $\hat{S}_{\overline{Y}}^{(mt)}(f)$ (left-hand plot), WOSA $\hat{S}_{\overline{Y}}^{(wosa)}(f)$ (middle) and Burg autoregressive $\hat{S}_{\overline{Y}}^{(ar)}(f)$ (right) spectral estimates for fractional frequency deviations $\overline{Y}_t$ derived from the clock noise series $\widetilde{X}_{1,t}$ (top row) and $\widetilde{X}_{2,t}$ (bottom). As before, the multitaper estimates are based upon $K = 6$ data tapers, while the WOSA estimates are based upon $K = 6$ subseries, each of length $N_S = 1024$. The model orders for the Burg estimates are $p = 4$ and $p = 24$ for $\overline{Y}_t$ based upon, respectively, $\widetilde{X}_{1,t}$ and $\widetilde{X}_{2,t}$.

8. As in figure 5, but now for spectral estimates based upon postcoloring the estimates shown in figure 7.

9. Ratios of spectral estimates for $\widetilde{X}_{1,t}$ (top row) and $\widetilde{X}_{2,t}$ (bottom). The first column shows the ratio $\hat{S}_X^{(mt)}(f)/\hat{S}_X^{(wosa)}(f)$ versus f for the multitaper and WOSA estimates of figure 8; the second and third columns show, respectively, $\hat{S}_X^{(mt)}(f)/\hat{S}_X^{(ar)}(f)$ and $\hat{S}_X^{(wosa)}(f)/\hat{S}_X^{(ar)}(f)$ using the AR estimates of figure 5.

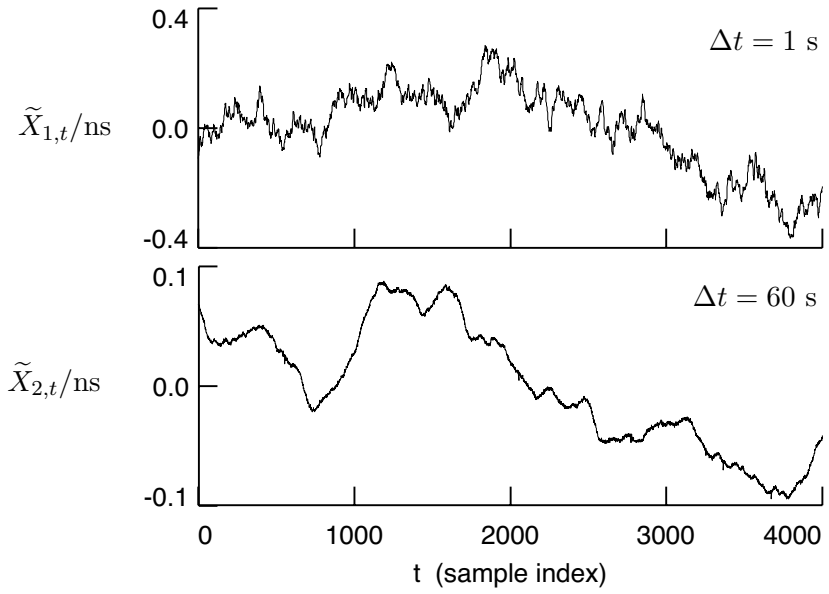

Figure 1: Two examples of clock noise. The top plot shows the time differences $\widetilde{X}_{1,t}$ between two atomic clocks taken once per second for 4000 seconds (data collected using a TSC 5110A analyzer and courtesy of Dr. Samuel R. Stein, Timing Solutions Corporation). The bottom plot shows $\widetilde{X}_{2,t}$, the time differences between two hydrogen masers sampled once per minute for 4000 minutes (data courtesy of Drs. Lara S. Schmidt and Demetrios Matsakis, US Naval Observatory). This series was extracted from a much longer segment that was adjusted for a linear trend. Both series have been centered so that their sample means are zero. The units for $\widetilde{X}_{1,t}$ and $\widetilde{X}_{2,t}$ are in nanoseconds.

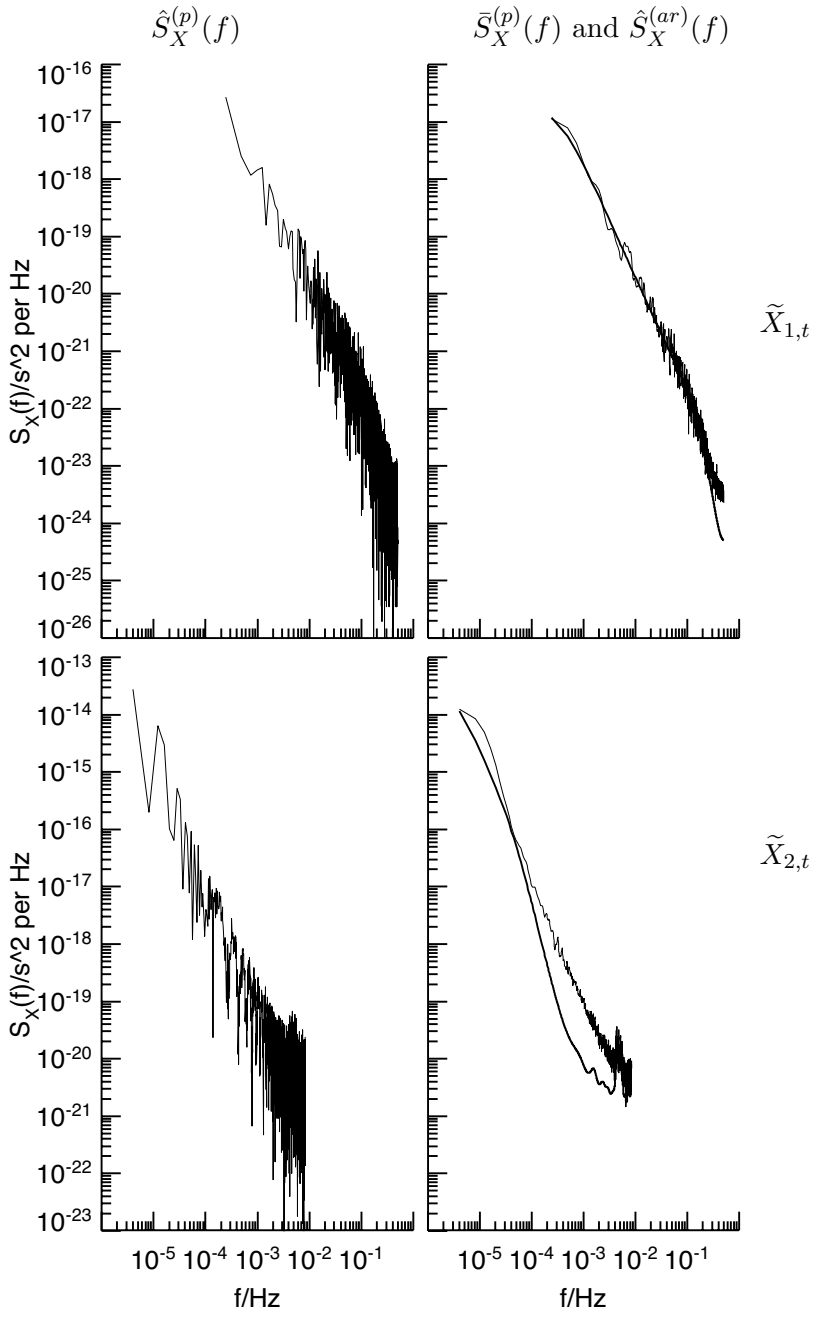

Figure 2: Periodograms $\hat{S}_X^{(p)}(f)$ for clock noise series $\widetilde{X}_{1,t}$ and $\widetilde{X}_{2,t}$ (left-hand plots) and smoothed periodograms $\bar{S}_X^{(p)}(f)$ and Burg autoregressive spectral estimates $\hat{S}_X^{(ar)}(f)$ (right-hand). The spectral estimates for $\widetilde{X}_{1,t}$ are in the upper row, and those for $\widetilde{X}_{2,t}$, in the lower row. In the right-hand plots, the thicker (and smoother looking) curves are the Burg estimates.

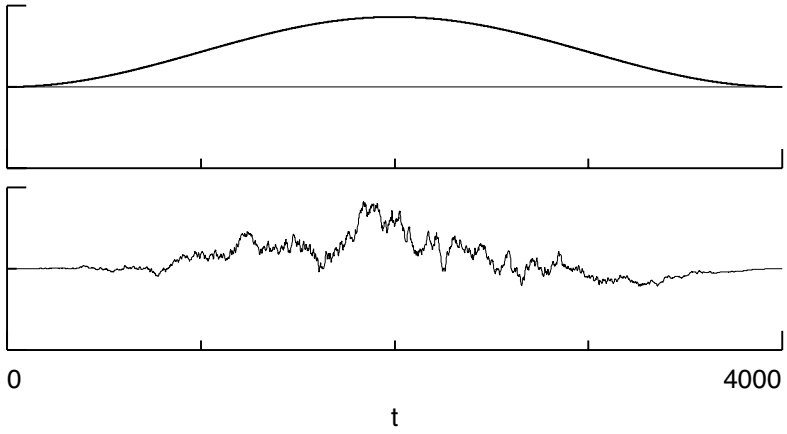

Figure 3: Hanning data taper $h_{N,t}$ (top plot) and the result of multiplying it point by point with the centered clock noise series $\widetilde{X}_{1,t}$ to form the tapered series $h_{N,t}\widetilde{X}_{1,t}$ (bottom).

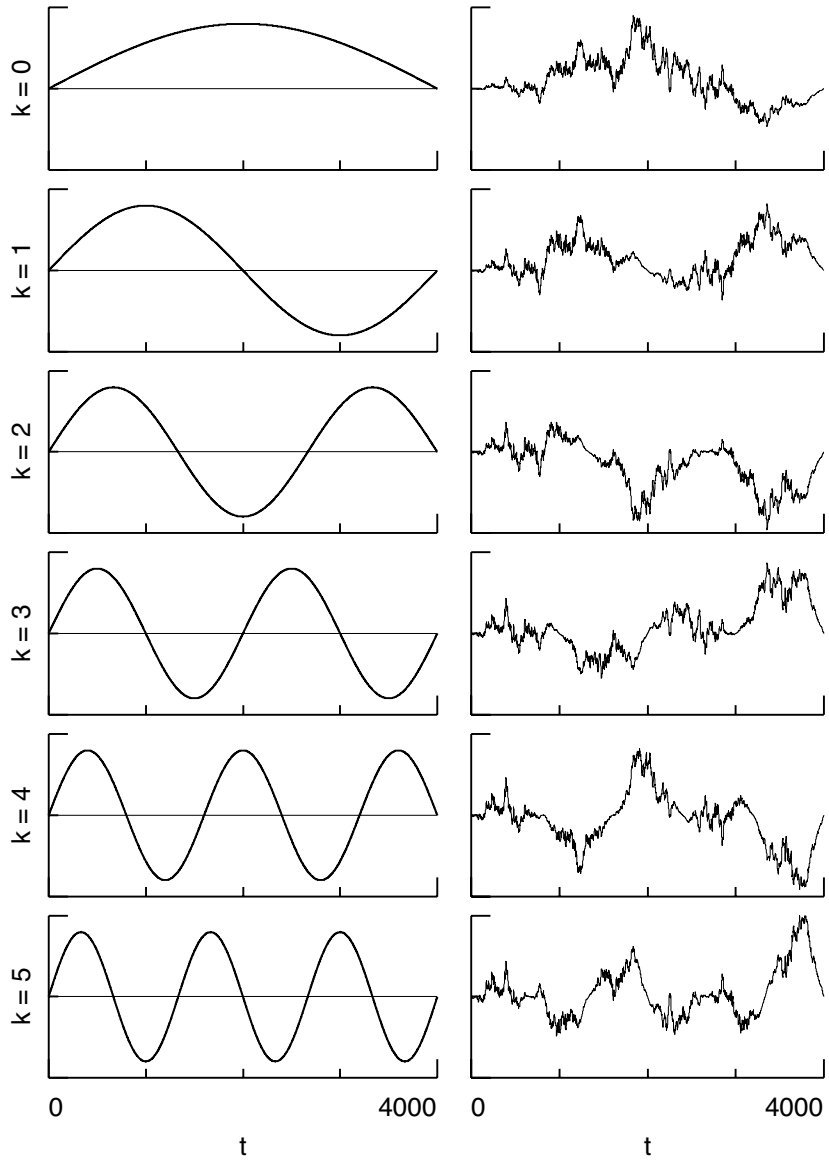

Figure 4: Sinusoidal tapers $h_{k,N,t}$ for $k = 0, 1, \ldots, 5$ (left-hand column, top to bottom) and corresponding tapered and centered series $h_{k,N,t}\widetilde{X}_{1,t}$ (right-hand) for $\widetilde{X}_{1,t}$.

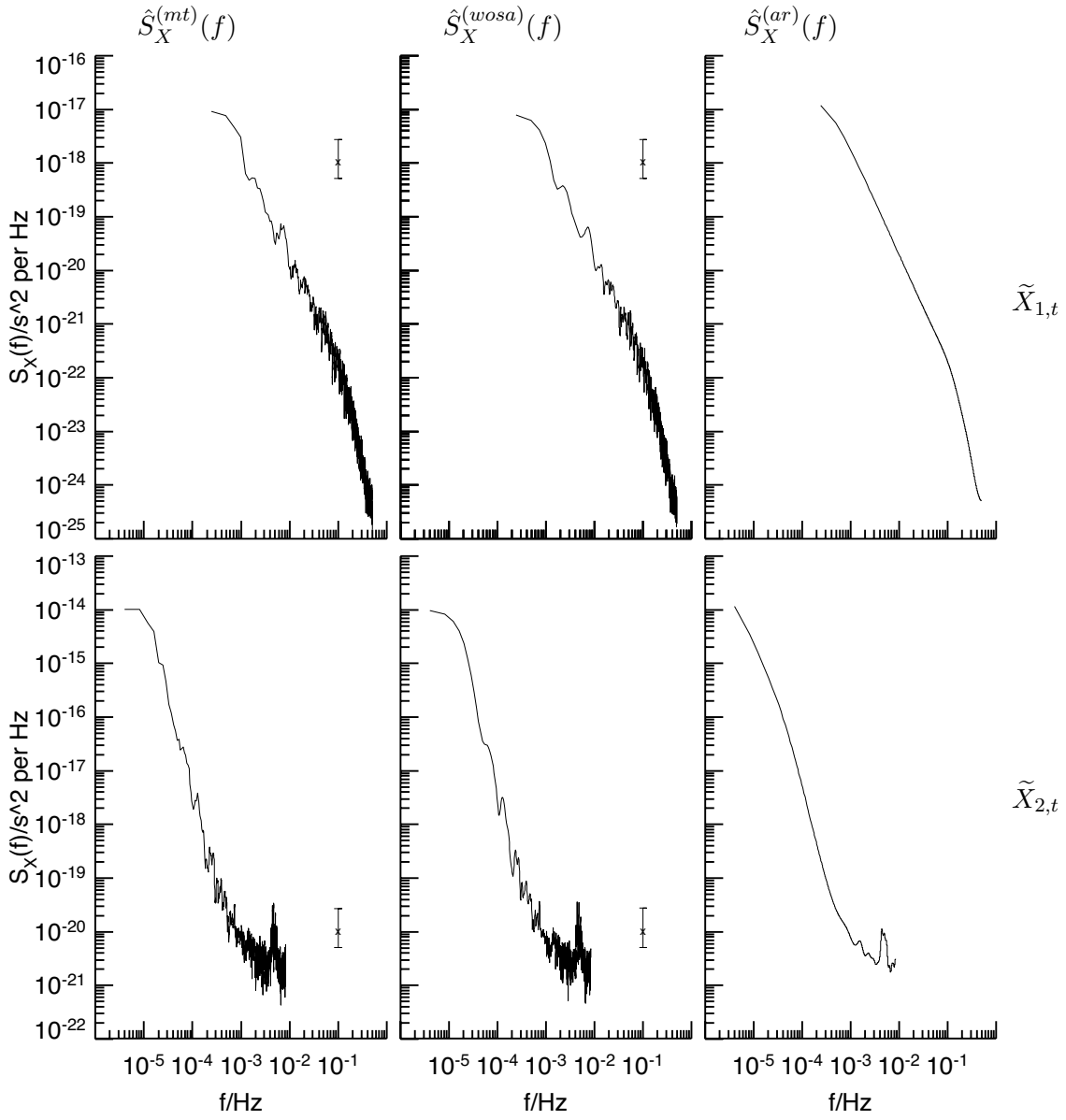

Figure 5: Sinusoidal multitaper $\hat{S}_X^{(mt)}(f)$ (left-hand plots), WOSA $\hat{S}_X^{(wosa)}(f)$ (middle) and Burg autoregressive $\hat{S}_X^{(ar)}(f)$ (right) spectral estimates for the clock noise series $\widetilde{X}_{1,t}$ (top row) and $\widetilde{X}_{2,t}$ (bottom). The multitaper estimates are based upon $K = 6$ data tapers. The WOSA estimates are based upon $K = 6$ subseries, each of length $N_S = 1024$. The model orders for the Burg estimates are $p = 5$ and $p = 25$ for, respectively, $\widetilde{X}_{1,t}$ and $\widetilde{X}_{2,t}$. The left-hand and middle plots also show what a typical 95% confidence interval for the true spectrum looks like when based upon the spectral estimates in the plots.

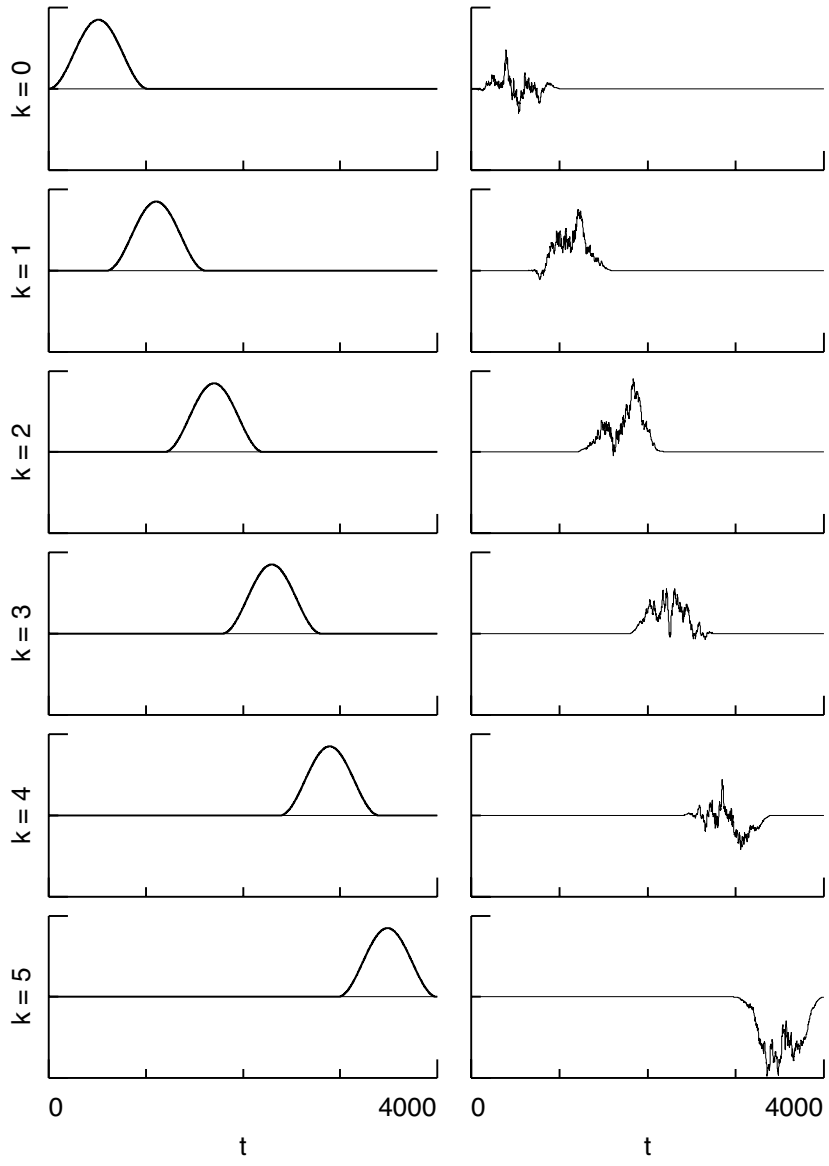

Figure 6: Aligned Hanning data tapers $h_{N_S,t}$ (left-hand column) and the product of the aligned tapers with $\widetilde{X}_{1,t}$ (right-hand).

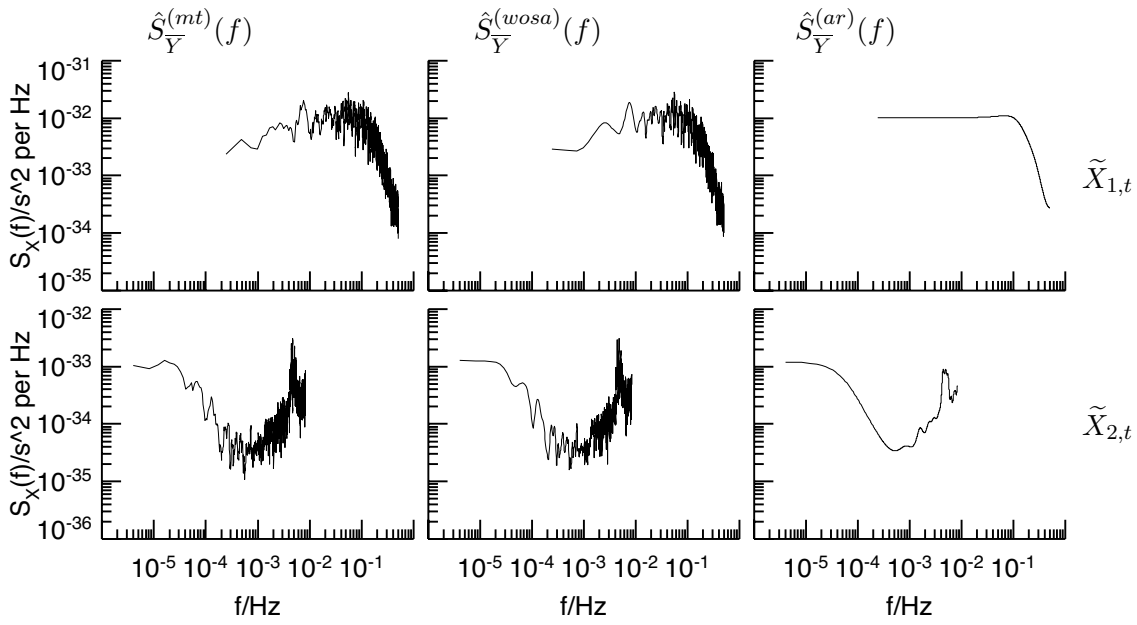

Figure 7: Sinusoidal multitaper $\hat{S}_{\overline{Y}}^{(mt)}(f)$ (left-hand plot), WOSA $\hat{S}_{\overline{Y}}^{(wosa)}(f)$ (middle) and Burg autoregressive $\hat{S}_{\overline{Y}}^{(ar)}(f)$ (right) spectral estimates for fractional frequency deviations $\overline{Y}_t$ derived from the clock noise series $\widetilde{X}_{1,t}$ (top row) and $\widetilde{X}_{2,t}$ (bottom). As before, the multitaper estimates are based upon $K = 6$ data tapers, while the WOSA estimates are based upon $K = 6$ subseries, each of length $N_S = 1024$. The model orders for the Burg estimates are $p = 4$ and $p = 24$ for $\overline{Y}_t$ based upon, respectively, $\widetilde{X}_{1,t}$ and $\widetilde{X}_{2,t}$.

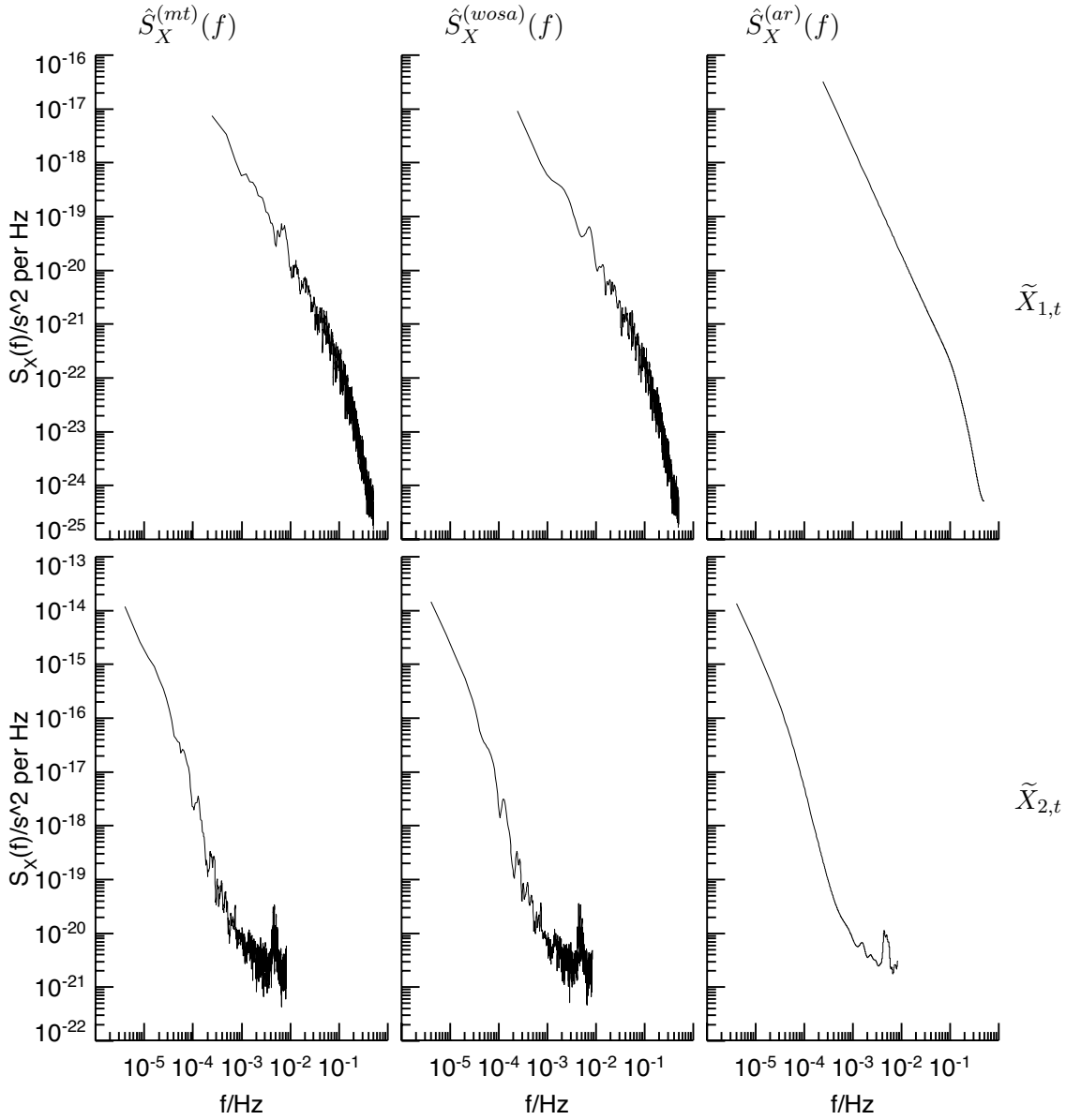

Figure 8: As in figure 5, but now for spectral estimates based upon postcoloring the estimates shown in figure 7.

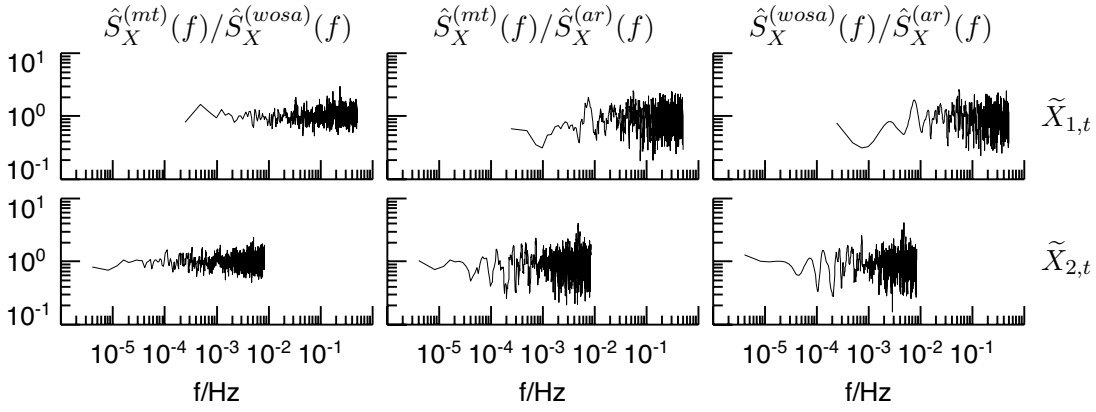

Figure 9: Ratios of spectral estimates for $\widetilde{X}_{1,t}$ (top row) and $\widetilde{X}_{2,t}$ (bottom). The first column shows the ratio $\hat{S}_X^{(mt)}(f)/\hat{S}_X^{(wosa)}(f)$ versus f for the multitaper and WOSA estimates of figure 8; the second and third columns show, respectively, $\hat{S}_X^{(mt)}(f)/\hat{S}_X^{(ar)}(f)$ and $\hat{S}_X^{(wosa)}(f)/\hat{S}_X^{(ar)}(f)$ using the AR estimates of figure 5.