

# Should the Classical Variance Be Used As a Basic Measure in Standards Metrology?

DAVID W. ALLAN

**Abstract**—Since a measurement is no better than its uncertainty, specifying the uncertainty is a very important part of metrology. One is inclined to believe that the fundamental constants in physics are invariant with time and that they are the foundation upon which to build international system (SI) standards and metrology. Therefore clearly specifying uncertainties for these physical invariants at state-of-the-art levels should be one of the principal goals of metrology. However, by the very act of observing some physical quantity we may perturb the standard, thus introducing uncertainties. The random deviations in a series of observations may be caused by the measurement system, by environmental coupling or by intrinsic deviations in the standard. For these reasons and because correlated random noise is as commonly occurring in nature as uncorrelated random noise, the universal use of the classical variance, and the standard deviation of the mean may cloud rather than clarify questions regarding uncertainties; i.e., these measures are well behaved only for random uncorrelated deviations (white noise), and white noise is typically a subset of the spectrum of observed deviations. The assumption that each measurement in a series is independent because the measurements are taken at different times should be called into question if, in fact, the series is not random and uncorrelated, i.e., does not have a white spectrum. In this paper, studies of frequency standards, standard-volt cells, and gauge blocks provide examples of long-term random-correlated time series which indicate behavior that is not “white” (not random and uncorrelated). This paper outlines and illustrates a straightforward time-domain statistical approach, which for power-law spectra yields an alternative estimation method for most of the important random power-law processes encountered. Knowing the spectrum provides for clearer uncertainty assessment in the presence of correlated random deviations, the statistical approach outlined also provides a simple test for a white spectrum, thus allowing a metrologist to know whether use of the classical variance is suitable or whether to incorporate better uncertainty assessment procedures, e.g., as outlined in the paper.

## I. INTRODUCTION

**T**IME AND FREQUENCY metrology provides some of the most accurate measurements known to man. Over the last three decades a large number of papers dealing with the statistics of atomic time and frequency standards have been published. The statistical procedures used in this discipline have been developed to a high level, driven by the need to deal with random correlated time series. The lessons derived from this work clearly require a negative answer to the question posed in the title of this paper. In essentially all atomic clocks the standard deviation of the long-term frequency fluctuations diverges without limit as the observation time increases. There are

appropriate places to use classical statistics within this discipline, however, they are limited.

Since the statistical methods developed for time and frequency metrology are generally applicable for any equispaced time series, opportunity was taken to apply these methods in some other areas of metrology, namely, standard voltage cells. Gauge block data were also studied but these data were not equispaced yielding some limitations to the procedure. These will be discussed in the text.

It is worth noting that if one has a viable model for a time series, a spectral density model proportional to  $f^\alpha$ , and if  $\alpha = -1$ , then the classical variance and standard deviation are divergent. Because of the ubiquitous nature of  $1/f$  noise for low-frequency components, perhaps it is reasonable to ask the question posed in the title. Aside from frequency standards  $1/f$  noise has been observed in important systems: transistor junctions, semiconductor diodes, resistors, thermistors, carbon microphones, thin films, light sources, RF propagation fluctuations, and in a surprising number of other processes [1], [2]. If noise with  $\alpha \leq -1$  is found to be a reasonable model for measurement deviations in basic standards in general, then the classical variance and standard deviation may have limited usefulness. The problem becomes significant when very long-term averaging is used. This is precisely what is required for maintenance of fixed standards which form the “invariant” building blocks of our measurement system.

In maintaining a set of standards and deriving calibrations of other standards from the set, several questions arise. Two important ones are: 1) to what degree does continued averaging of repeated measurements improve upon the quality of the result? and 2) what physical processes influence the long-term behavior of a standard? The first question is directly answerable within the context of the analysis presented here and, once answered, should show the conditions under which the classical variance can be used safely. One might note here that, for certain spectral models (e.g.,  $1/f^2$ ), the sample variance gets larger with increased averaging or with increased number of readings. The second question cannot be addressed directly, but experience in time and frequency metrology suggests that time-series analysis can provide a measure of the “health” of a standard. With high-quality oscillators, we have found that certain types of behavior in the

Manuscript received June 23, 1987.

The author is with the National Bureau of Standards, Time and Frequency Division, Boulder, CO 80303.

IEEE Log Number 8613595.

noise spectrum can be associated with certain physical processes, so that the noise analysis provides information on the performance limiting phenomena. This is especially valuable for considerations of long-term performance. At this point one can only speculate that the techniques outlined herein will be as useful in understanding and in improving uncertainties in other areas of standards metrology.

In time and frequency metrology it has been found that random variations are often well modeled by integer-power-law spectra proportional to  $f^\alpha$ , with  $-2 \leq \alpha \leq +2$ . Power-law-spectrum estimation procedures have been developed and these allow one to develop appropriate power-spectral-density models. The nonclassical statistical methods outlined herein are simple to use and are convergent for the power-spectral-noise models encountered in many physical systems.

To show the divergent behavior of the classical variance for some correlated time series note that it can be written [3] (Parseval's Theorem):

$$\sigma^2 = \frac{1}{\pi} \int_0^\infty S(\omega) d\omega$$

where  $S(\omega)$  is a symmetrical two-sided spectral density and  $\omega = 2\pi f$ . If in metrology the variable being measured has random variations that have a power spectrum proportional to  $f^\alpha$ , then for  $\alpha = -1$  the classical variance is infinite at both limits of the integral, i.e., unless there are both high- and low-frequency cutoffs to the process being analyzed, the classical variance is unbounded. A high-frequency cutoff  $f_h$  always exists and is typically determined by the measurement system. However, a low-frequency cutoff  $f_l$  is needed for  $\alpha \leq -1$ , and  $f_l$  seems to be elusive, inaccessible, and, sometimes, impossible to determine. If  $f_l$  is not reasonably measurable, the classical variance is not useful in characterizing  $f^\alpha$  processes with  $\alpha \leq -1$ .

An assumption often employed in metrology is that the random uncertainties associated with the measurement of a given physical quantity are uncorrelated in time and, hence, they are often not treated as a time series. If the deviations in a series of measurements are random and uncorrelated (white spectrum), then the calculation of the mean, the standard deviation, and the standard deviation of the mean are meaningful and well understood. If the random deviations are correlated when taken as a time series, then a knowledge of their spectrum and/or autocorrelation function is necessary to properly interpret the results, define the uncertainties, and predict future values. It is often the case that, as one moves from short-term measurements to long-term measurements of the fundamental constants of physics and of nature, one is forced to abandon the simple assumption of random uncorrelated (white) measurement noise. Short and long can take on very different values depending on the area of metrology. At state-of-the-art levels there are often processes which are correlated in time which influence the measurements,

e.g., there may be a coupling to the environment, the physical device may suffer internal changes with time, or changes may be induced by the measurement process.

One of the main goals of this paper is to outline a simple statistical approach, for characterizing spectral models of the random deviations observed. Methods are developed to replace the nonconvergent classical variance and standard deviation with variance measures that are as easy to compute but often reveal much more information about the random processes involved. Another goal is to provide a simple test for whiteness of the spectrum of the random deviations of a series of measurements. For the above reasons, if the spectrum is not known it should be measured. If it is white, then one may proceed in the traditional way—using the simple mean, the standard deviation, and the standard deviation of the mean, etc. If it is not white, then one can proceed as outlined herein for spectrum estimation, or use some of the more traditional frequency-domain analysis procedures. Limitations of some of the methods outlined herein are associated with the fact that the development in these methods does not treat data unequally spaced in time. In these cases, the methods outlined in this paper have additional uncertainties and limitations. However, other methods shown allow some conclusions to still be obtained even in these cases, however, more research is necessary before clear spectral modeling with good confidence intervals is available for unequally spaced data. In this regard frequency-domain methods are much further developed.

## II. SOME STATISTICAL TOOLS

Because the usefulness of the classical variance, the standard deviation and the standard deviation of the mean depends upon the implicit assumption that a set of measurements are random uncorrelated (have a white spectrum), the emphasis of this section will be on the development of some alternate statistical tools which have been found to be useful in time and frequency metrology.

Over the years, the time and frequency community has published numerous papers and several books and has held conferences, seminars, and workshops which in whole or in part have addressed the problems of spectrum estimation, modeling, and systematic errors [7]–[12]. Though difficult to summarize in a few words, one of the conclusions of all of this work is that most of the random processes occurring in time and frequency metrology can be well modeled by a power-law spectrum,  $S_y(f) \sim f^\alpha$  [13]. As will be shown later this model may also be appropriate for standard volt cells. If power-law spectra are good models in other areas of metrology, then those areas may benefit from the statistical analysis procedures for modeling of the random deviations of a set of measurements taken as a time series. Characteristics desired in a measure of random deviations are that the measure:

- 1) is convergent;
- 2) has a solid theoretical basis;
- 3) is convenient to use;

- 4) is easy to interpret;
- 5) has a natural relationship to frequently occurring experiments and measurement methods;
- 6) has tractable transformations between the frequency domain and the time domain and vice-versa; and
- 7) is an efficient estimator.

After some years of dealing with the problems of non-white spectra in time and frequency metrology, the IEEE formed a subcommittee that published recommended measures of frequency stability [8]. The focus in this paper is only on their recommended measure for the time domain developed for analysis of random correlated time series that cannot be adequately handled with a classical statistical approach. The measure they recommended satisfies most of the above requirements.

The measure is simple to estimate for finite data sets and is defined as

$$\sigma_y^2(\tau) = \frac{1}{2} \langle (\Delta y)^2 \rangle \quad (1)$$

where the brackets  $\langle \rangle$  denote an infinite time average and the delta  $\Delta$  denotes a first finite difference of adjacent fractional frequency measurements  $y$ , each sampled over a sample time  $\tau$ . The squared first differences are averaged over a theoretical infinite series. The  $y$  can be considered as any metrological measure taken as a time series. This measure is sometimes called the "two-sample variance," the "pair variance," the "Allan variance" or one-half the "mean square successive difference," and the confidence of the estimate of this measure has been reported in several papers [4], [8], [14]–[19].

In practice, for a discrete series of measurements,  $y_k$  ( $k = 1$  to  $N$ ), each taken over a sample time  $\tau_0$ , one can compute an estimate of  $\sigma_y^2(\tau_0)$  as

$$\sigma_y^2(\tau_0) = \frac{1}{2(N-1)} \sum_{k=1}^{N-1} (y_{k+1} - y_k)^2. \quad (2)$$

For time and frequency measurements,  $y$  is taken as the normalized rate or frequency of a clock  $y = (\nu - \nu_0)/\nu_0$  where  $\nu$  is the actual average clock frequency and  $\nu_0$  is the nominal frequency. It is important to distinguish between the carrier frequency  $\nu$  and the Fourier frequency  $f$  of the random processes. This normalized frequency will have systematic deviations as well as random deviations. The spectral density of the random deviations is well modeled by  $S_y(f) = h_\alpha f^\alpha$  with  $\alpha = -2, -1, 0, +1, +2$ . The coefficient  $h_\alpha$  is the intensity of the particular type  $\alpha$  of power-law process. Which value of  $\alpha$  is appropriate is a function of the kind of clock being modeled and of the region of Fourier frequency  $f$  or of averaging time  $\tau$  of interest. Generally, one expects the spectral type to change as one goes from short-term to long-term averages; and  $\alpha$  tends to decrease as one goes to longer averages. Once the random deviations in a measurement are characterized, optimum procedures can be developed for measuring other biases and systematics such as the time offset, the frequency offset, and the frequency drift of a

clock [4]. Such a characterization also allows one to combine the measurements from an ensemble of clocks. These measurements can be statistically and optimally weighted such that, for example, the performance of the ensemble average is better than that of the best clock in the system. Such procedures are routinely used in the National Bureau of Standards (NBS) time scale. This scale which is quite smooth is calibrated periodically with the primary frequency standard, NBS-6, and the analysis of deviations in the scale provides guidance for the optimum interval between these calibrations [5]. Analogous approaches to the above could be utilized in others areas of metrology.

Some of the limitations in the use of  $\sigma_y^2(\tau)$  are now addressed in an effort to broaden its application in other areas of metrology. First, in contrast to adjacent frequency measurements, it is often the case in metrology that a significant amount of time will elapse between successive measurements of a quantity. This has been called "dead time," and its effects on correlated time series analysis have been studied, establishing relationships between the Allan variance and variances computed with dead time present. What has not been published is the effect of averaging such measurements so that dead time is distributed across the averages [20]. This will be discussed later in this paper. Second, the Allan variance has an ambiguity in that it cannot easily distinguish between white phase noise ( $\alpha = +2$ ) and flicker phase noise ( $\alpha = +1$ ). This limitation was resolved by developing a "modified Allan variance,"  $\text{Mod} \cdot \sigma_y^2(\tau)$ , which is slightly more difficult to compute, but does give additional insights into the spectral characteristics [21], [22]. This latter measure will also be useful as we proceed. And third, if the spectrum is known, the Allan variance can be computed, but the reverse is not generally true. Fortunately, if the spectral model is a power law, the reverse is true and the transformation is fairly straightforward. If the spectral model is an integer power law, then the transformation is easy (see Table II).

Since the eye has excellent data evaluation capability and is also a fair spectrum analyzer, it is always good to look at a measurement sequence as a first step. Fig. 1 shows plots of some power-law spectra. For time and frequency the time series illustrated in Fig. 1 are analogous to the time deviations in various precision clocks and oscillators and for different sampling or averaging times. Since time is the integral of frequency, one may write [8]

$$S_y(f) = (2\pi f)^2 S_x(f) \quad (3)$$

where  $y = dx/dt$  and  $x$  represents the time deviations.

It is useful to name the integer-power-law spectral values. If  $S_y(f) = h_\alpha f^\alpha$ , and  $\alpha = -2, -1, 0, 1$ , or  $2$ , then using (3) one may write  $S_x(f) = h_\alpha f^\beta / (2\pi)^2$ , with corresponding values of  $\beta = -4, -3, -2, -1$ , or  $0$ . Table I gives the names for the different values of  $\alpha$  and  $\beta$ , and Table II gives the Fourier transform relationship. Note that for time and frequency, the time deviations and phase deviations are proportional, that is,  $x = \phi / (2\pi\nu_0)$ , where

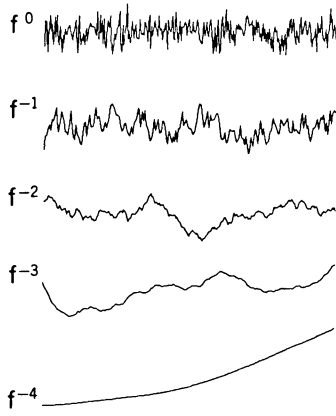


Fig. 1. Simulated power-law spectra where  $f$  is the Fourier frequency. (These are commonly occurring random processes in nature. They are named  $f^0$ , white noise;  $f^{-1}$ , flicker noise;  $f^{-2}$ , random walk;  $f^{-3}$ , flicker walk; and  $f^{-4}$ , random run.)

TABLE I  
NAME OF NOISE

$\alpha$	$\beta$	FREQUENCY MODULATION (FM)	PHASE MODULATION (PM)
2	0	SUPER WHITE	WHITE
1	-1	SUPER FLICKER	FLICKER
0	-2	WHITE	RANDOM WALK
-1	-3	FLICKER	FLICKER WALK
-2	-4	RANDOM WALK	RANDOM RUN

$\phi$  denotes the phase deviation of an oscillator. The power-law processes depicted in Fig. 1 are computer simulations and could be analogous to a time series from any metrology set. One notes from the above relationships, for example, that white noise on  $y$  ( $\alpha = 0$ ) is random walk noise on  $x$  ( $\beta = -2$ ). One can use either  $y$  or  $x$  as the deviations in a measurement series. Since we believe frequency is one of the invariants in physics, we may be inclined to use  $y$  as our generic measurement deviation variable for a general metrology time series, but the logical development here allows for the use of  $x$  or  $y$ . A few of the advantages and disadvantages of specific choices of  $x$  or  $y$  will be presented.

For discrete data the continuous derivative  $y = dx/dt$  becomes  $y_k = \Delta x_k / \tau_0$  where  $k$  is the counting index in the measurement time series and where  $\Delta x_k = x_{k+1} - x_k$  is the first finite difference on the time deviation series  $x_k$ , with  $x_k$  spaced  $\tau_0$  apart, i.e.,  $\tau_0$  is the average or nominal spacing between the beginning of the  $y_k$  and the beginning of the  $y_{k+1}$ . Given the discrete measurement time series  $y_k$ , (2) provides an estimate of  $\sigma_y^2(\tau_0)$ . To observe the

dependence of  $\sigma_y^2(\tau_0)$  on  $\tau$ ,  $\tau$  can be easily varied by averaging as many adjacent values on the  $y_k$ s as desired

$$\sigma_y^2(\tau) = \frac{1}{2(N - 2n + 1)} \sum_{k=1}^{N-2n+1} \left( \frac{1}{n} \sum_{j=k+n}^{k+2n-1} y_j - \frac{1}{n} \sum_{j=k}^{k+n-1} y_j \right)^2 \quad (4)$$

where  $\tau = n\tau_0$ . A mathematically equivalent relationship can be used to estimate  $\sigma_y^2(\tau)$  from the time deviation data

$$\sigma_y^2(\tau) = \frac{1}{2\tau^2(N - 2n + 1)} \sum_{k=1}^{N-2n+1} (x_{k+2n} - 2x_{k+n} + x_k)^2. \quad (5)$$

Varying  $n$  ( $\tau = n\tau_0$ ) in either (4) or (5) gives a time-domain power-law spectrum-estimation procedure. For power-law spectra  $\sigma_y^2(\tau)$  is proportional to  $\tau^\mu$ . Fig. 2 gives the relationship between  $\mu$  and  $\alpha$ . Notice, for the above types of power-law spectra, the Fourier transform relationship is simply  $\alpha = -\mu - 1$  ( $-3 < \alpha \leq +1$ ) and  $\mu = -2$  ( $\alpha \geq +1$ ). There is a one-to-one correspondence between  $h_\alpha$  and the proportionality constant for  $\sigma_y(\tau)$  (see Table II and [8]).

Since  $\log \sigma_y^2(\tau) \sim \mu \log \tau$ , it is convenient to plot  $\sigma_y(\tau)$  on a log-log plot with  $n = 2^i$ ,  $i = 0, 1, 2, \dots$ . The slope  $\mu/2$  and the intensity  $\sigma_y(\tau)$  from this log-log plot focus on two parameters (spectral intensity and spectral type) which may characterize a random process over several decades of Fourier frequency. Notice, that with a measurement time-series data stream  $y_k$  the sample time  $\tau = n\tau_0$  can be varied in the software as in (4) or (5). Of course, in principle, it can be changed in the hardware or as part of the measurement process, but that is usually much more cumbersome.

A simple *white-noise-test* procedure can be readily developed. Given the data stream  $y_k$  one may also compute the standard deviation for  $N$  samples in the conventional way

$$\sigma(N) = \left[ \frac{1}{N-1} \sum_{k=1}^N (y_k - \bar{y})^2 \right]^{1/2} \quad (6)$$

where  $\bar{y}$  denotes the mean value of the sample. Taking the ratio of the standard deviation squared to the Allan variance illustrates the convergence or the lack of the convergence of standard deviation, for the various power-law spectra of interest (see Fig. 3). This ratio is also a measure of the amount of correlation in the data. As the data length increases without limit, the standard deviation diverges without limit for all  $\alpha \leq -1$ , e.g., flicker and random walk. In fact, one can use this ratio, i.e., the divergent nature of the standard deviation, to estimate the value of  $\mu$  from the following relationship:

$$\sigma^2(N) / \sigma_y^2(\tau_0) = \frac{N(N^\mu - 1)}{2(N-1)(2^\mu - 1)}. \quad (7)$$

TABLE II  
FOR  $S_y(f) = h_\alpha f^\alpha$

Noise Name	$\alpha$	$S_y(f) =$	$S_x(f) =$	$\sigma_y^2(\tau) =$
White PM	+2	$h_2 f^2$	$\frac{h_2}{(2\pi)^2}$	$\frac{3f_h}{(2\pi)^2} h_2 \tau^{-2}$
Flicker PM	+1	$h_1 f$	$\frac{h_1}{(2\pi)^2} f^{-1}$	$\frac{1.038 + 3 \ln(2\pi f_h \tau)}{(2\pi)^2} h_1 \tau^{-2}$
White FM or Random Walk PM	0	$h_0$	$\frac{h_0}{(2\pi)^2} f^{-2}$	$\frac{1}{2} h_0 \tau^{-1}$
Flicker FM	-1	$h_{-1} f^{-1}$	$\frac{h_{-1}}{(2\pi)^2} f^{-3}$	$2 \ln(2) h_{-1}$
Random Walk FM	-2	$h_{-2} f^{-2}$	$\frac{h_{-2}}{(2\pi)^2} f^{-4}$	$\frac{(2\pi)^2}{6} h_{-2} \tau$

These power-law processes have been very useful in modeling the random processes in precision clocks and oscillators.

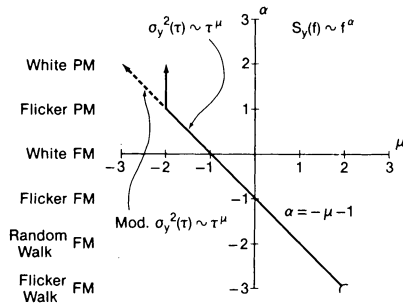


Fig. 2. Fourier transform mapping from time domain to frequency domain and vice versa. (With  $\sigma_y^2(\tau) \sim \tau^\mu$  ( $-2 \leq \mu < +2$ ) and with  $\text{Mod} \cdot \sigma_y^2(\tau) \sim \tau^\mu$  ( $-4 < \mu \leq -2$ ,  $\alpha = -\mu - 1$ ). Analyzing the data in the time domain and using the above mapping is like a super-fast Fourier transform in the case of power-law spectra.) See [22], [23] for definition of  $\text{Mod} \cdot \sigma_y^2(\tau)$ .

If the spectrum is white ( $\alpha = 0$ ,  $\mu = -1$ ), then the ratio is 1; i.e., the classical variance equals the Allan variance; hence, if the ratio is not 1, then the spectrum is not white. If  $\alpha < 0$  (correlated random deviations) then this ratio will be greater than 1. If the spectrum is not known, this simple ratio test should be conducted. See von Neumann and Hart for information on the distribution of this ratio [19], [20]. As an approximate rule, if the ratio is unity within  $100/\sqrt{N}$  percent, then one can probably safely proceed with the classical variance and the traditional approach. If not, one has reason to doubt the validity of the classical variance and a procedure such as that outlined here may be useful as a time-domain analysis tool and for estimating the spectrum of the measurement deviations. Tabular values for (7) are listed in [26] and [27] for various values of  $N$  and for data with or without dead time. The "ratio = 1" test is valid with or without dead time. Equation (7) is for the no dead-time case, and is the basis for the curves plotted in Fig. 3. Other interesting values of the ratio in (7) are:  $N \ln N / (2(N-1) \ln 2)$  for  $\alpha = -1$  ( $\mu = 0$ , flicker noise); and  $N/2$  for  $\alpha = -2$  ( $\mu = +1$ , random walk).

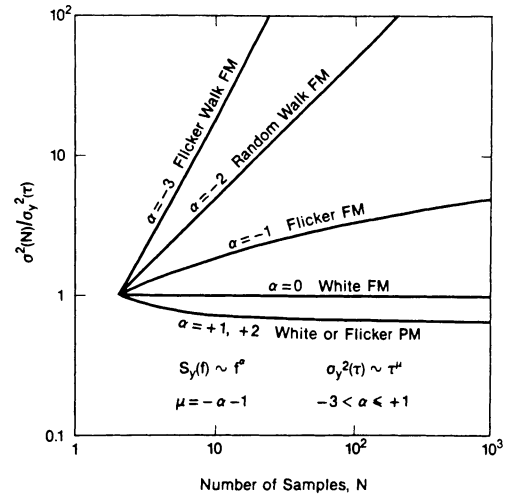


Fig. 3.  $\sigma_y^2(\tau)$  is a convergent stable measure in the presence of the random processes named in Fig. 1. (Plotted on the ordinate is the ratio of the standard deviation squared to  $\sigma_y^2(\tau)$  as a function of the number of samples  $N$ , where  $\tau_0$  is the nominal averaging time for each measurement in a series. For  $1/f$ ,  $1/f^2$  and  $1/f^3$  power-law processes the standard deviation increases with increasing  $N$  without bound.)

In contrast to time and frequency metrology where adjacent frequency measurements are easy to make as calculated from the time measurements, in other areas of metrology a significant amount of time may elapse between the measurements. The effects of this dead time have been studied and one can still determine power-law spectral density models, if they are appropriate for the random variations [21]–[28]. In the case of white noise FM, the dead time has no effect since the deviations are random and uncorrelated regardless of their spacing. For the dead-time case, the following more general variance is defined

$$\langle \sigma_y^2(N=2, T_0, \tau_0, f_h) \rangle = \frac{1}{2} \langle (\Delta'y)^2 \rangle \quad (8)$$

where the  $N = 2$  denotes a two-sample variance,  $T_0$  is the

time between the beginning of one data sample and the next,  $\tau_0$  is the nominal averaging time for each  $y$  ( $T_0 - \tau_0$  is the dead-time), and  $f_h$  is the measurement system bandwidth. Notice that (8) is the same on the right as (1) except for the  $\Delta'$ , where the prime denotes the presence of dead time between each measure of  $y$ . Equation (3) can be used to compute an estimate of  $\langle \sigma_y^2(N = 2, T_0, \tau_0, f_h) \rangle$ . If the random deviations have a white spectrum, then (1) and (8) give the same variance. If they are not white then the variances will be different. Again the differences have been studied and published, and it has been shown that even with dead time present, the underlying power-law spectrum can be estimated [26]–[28]. If the ratio  $T_0/\tau_0$  is held constant, then the Fourier transform relationship  $\alpha = -\mu - 1$  is still valid [28].

Because it is convenient to do averaging in the software, we may think of averaging adjacent values even when there is significant dead time between them. In other words, use (4) to ascertain the  $T$  or  $\tau$  dependence, where  $T = nT_0$  and  $\tau = n\tau_0$ . This has been studied and also shown to be useful in estimating the appropriate underlying power-law spectral density model. A  $\sigma_y(\tau)$  versus  $\tau$  plot may more meaningfully become a  $\sigma_{yd}^2(T)$  versus  $T$  plot with the definition

$$\begin{aligned} \langle \sigma_{yd}^2(N = 2, T, \tau, f_h) \rangle &= \sigma_{yd}^2(T) \\ &= \frac{1}{2} \langle (\Delta' y')^2 \rangle \end{aligned} \quad (9)$$

where the  $y$  is primed to denote it is an average of  $n$  sequential  $y_k$  values across dead time, and  $T$  is the nominal total time window over which each  $y'$  is averaged. This distributed dead-time approach is useful for  $-3 < \alpha \leq 0$ , which is an important range of power-law spectral-exponent values. The dependence of  $\sigma_{yd}^2(T)$  on  $T$  asymptotically approaches the same behavior as in the no dead-time case for useful values of  $\alpha$ , making it a convenient measure, that is,  $\sigma_{yd}^2(T) \sim T^\mu$  for  $T \gg T_0$  and  $\alpha = -\mu - 1$ .

A slight limitation for the distributed dead-time approach for determining  $\mu$  is that if  $\mu = -1$  (white noise), then  $\alpha$  can have a range of values  $\alpha \geq 0$ . This is typically not a problem, because in this case classical statistics can be used.

As an alternative, if dead time exists in a data stream, its effect can be bypassed by using the  $x_k$ s in (5) as analog to the measurement deviations. In this case the modified Allan variance,  $\text{Mod} \cdot \sigma_y^2(\tau) \sim \tau^{\mu'}$ , often yields more information from the  $\tau$  dependence. The range of usefulness is for  $-3 \leq \mu' \leq +3$ . If  $\mu' = -3$ , the  $x$ 's have a white spectrum; if  $\mu' = -2$ , the spectrum of the  $x$ 's is flicker or  $1/f$ . For  $\mu' > -2$  the asymptotic behavior for large  $T$  or  $\tau$  is the same as for the Allan variance, but there are some significant variations for small  $n$  ( $\tau = n\tau_0$ ) [22], [23]. In the case of  $\mu > -2$ , it is usually easier to use the Allan variance directly—with or without the dead time.

Another limitation of the above measures is that equispaced data are assumed. Unequally spaced data are a

common occurrence in metrology; hence, dealing with this limitation is a significant concern. Limited information is available for spectrum estimation for unequally spaced time series [29], [30]. Much more theoretical work is needed in this area. Responding to the question in the title of this paper, there are some things that can be done using the current analysis to recognize nonwhite noise conditions in an unequally spaced time series. Consider an unequally spaced sampled set extracted from a process with a white spectrum. First, the classical variance is equal to the Allan variance independent of the data spacing for such a set [31]. Hence, the test using (7) can be employed as a test for correlation in an unequally spaced time series. If the process is random and uncorrelated (white), then the ratio is 1. Hence, if the ratio is not 1, then the spectrum is not white even if the data are unequally spaced. Again, for power-law spectra, if the ratio is greater than 1, then correlations exist in the time series; however, the interpretation of the ratio is not the same for unequally spaced data. On the other hand, even for unequally spaced data, the ratio will be a flag as to whether the classical variance is safe to use. And second, a  $\sigma_y(\tau)$  or  $\text{Mod} \cdot \sigma_y(\tau)$  analysis of such a set will yield the conclusion that the spectrum is white under the assumption that the values taken from the set are equispaced even though they may not be as a further test of the spectrum being white.

We have not discussed the distribution of the random deviations. An erroneous assumption that is sometimes made is that if a noise process has a normal distribution, it is white. What is, in fact, the case is that the distribution is independent of the spectral type of noise; for example, flicker noise can and does have a normal distribution for those cases investigated in precision oscillators. On the other hand, a square distribution can, of course, have a white spectrum. It is the experience of the author that normal distributions are encountered in most of nature's random processes. The distribution becomes important when dealing with the confidence of an estimate and normality is assumed throughout this paper.

### III. EXPERIMENTAL FINDINGS

#### A. Atomic Clocks

Fig. 4 is a  $\sigma_y(\tau)$  plot of the atomic clocks in the NBS ensemble. The  $\tau^{-1/2}$  behavior ( $\mu = -1$ ,  $\alpha = 0$ ) is classical white noise FM and typically persists for sample times out to several days. For the longer sample times the behavior changes to  $\sigma_y(\tau) \sim \tau^{1/2}$  ( $\mu = +1$ ,  $\alpha = -2$ , random walk FM) for the appropriate model [31]. Using the measured levels of the white noise FM and of the random-walk FM for each clock, an algorithm is employed that assigns optimum weighting factors to each clock to generate an optimum software standard, which should be better than the best clock in the ensemble. Some simulated estimates are also plotted in Fig. 4. These frequency stability estimates have been verified through international comparisons. The frequency uncertainty of the NBS

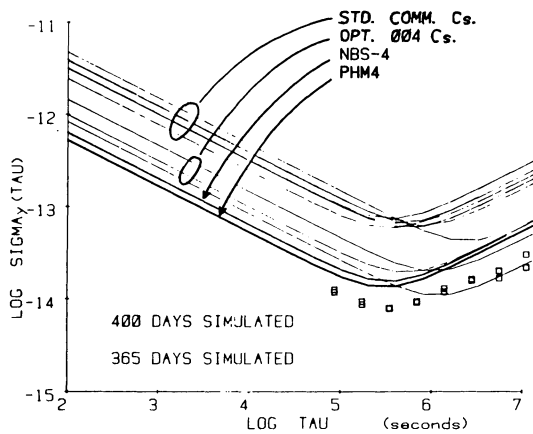


Fig. 4. A fractional frequency stability plot of the various members of the NBS clock ensemble as a function of the sample or averaging time  $\tau$ . (The clocks are modeled by white noise FM for sample times out to a few days, and then by random walk FM for sample times longer. The circles are the estimated stability from simulated data of the computed NBS weighted clock ensemble. The estimated stability has been verified through international comparisons.)

weighted clock ensemble slowly degrades to the inaccuracy of the NBS primary frequency standard (NBS-6 accuracy = 8 parts in  $10^{14}$ ) after about a year; the optimum calibration interval is, therefore, about one year.

### B. Standard Voltage Cells

The preliminary investigations reported in this paper were performed on data provided by Marshall [33]. The simple test of calculating the ratio of the standard deviation squared to  $\sigma_y^2(\tau_0)$ , as outlined above, was performed on the first cell and gave  $\sigma^2(N = 22)/\sigma_y^2(\tau_0) = 2 \pm 0.2$ , which is consistent with  $\mu \approx 0$  ( $\alpha = -1$  or  $1/f$  noise). The statistical methods for frequency standards were employed to characterize the voltage fluctuations for four transfer cells relative to the average of the four reference cells. In the first case, the voltage deviations were treated as analogous to frequency deviations  $y$ . Each voltage difference was measured for a few minutes once a week. Fig. 5 is a plot of one of the comparisons as a function of time. Fig. 6 is the square root of the Allan variance  $\sigma_{yd}(T)$ , with the distributed dead time for the data plotted in Fig. 5. A modified-Allan-variance analysis taking the voltage to be analogous to the time deviation  $x$  (assuming again a once-per-week measurement) yields good agreement with the  $1/f$  spectral density model for the voltage deviations. It is apparent from Fig. 6 that, within the confidence of the estimate, the white noise model is excluded. Barnes has shown that the Allan variance estimate from a finite data set is an unbiased estimate, and that it has a chi-square distribution. The confidence of the estimates are as given in [15]. All of the other cell pair comparison combinations were consistent with a  $1/f$  noise model. Since the data for the standard volt cells were equispaced, the theoretical tools developed in the previous section have been employed with the only assumption that the data are well modeled by power-law spectra. Within the confidence of the estimates this appears to be a valid assumption.

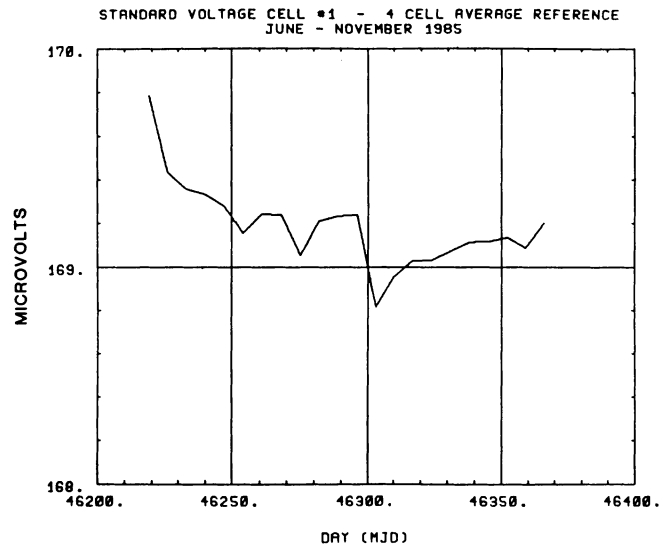


Fig. 5. A plot of the voltage stability of one of four standard volt cells with respect to a four-cell average reference. (Measurements were made on each of the four standard cells with respect to the reference for a few minutes once per week.)

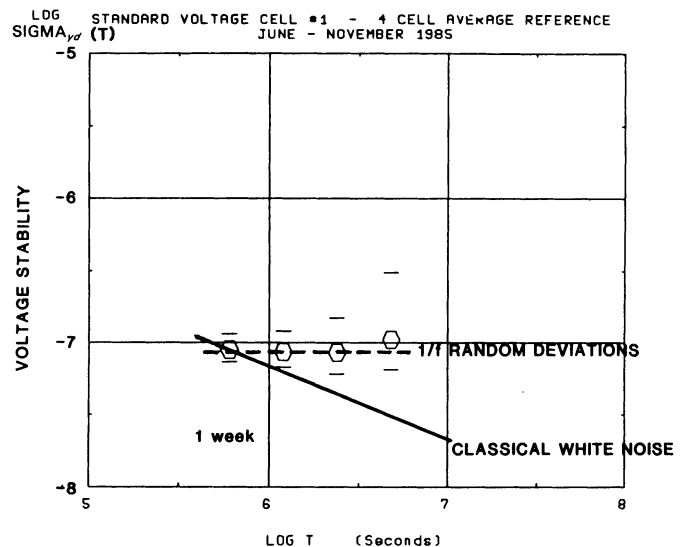


Fig. 6. A voltage stability plot  $\sigma_{yd}(T)$ , as a function of sample interval  $T$ . The confidence intervals shown are 90 percent (An  $1/f$  model is clearly a better fit than the white noise model. The standard deviation increases as the data length without bound for  $1/f$  voltage deviations. The modified Julian day (MJD) is a day count equal to the Julian day; 2 400 000.5 days. MJD 46431, for example, begins 0000 h universal time on January 1, 1986.)

Fig. 7 for the standard voltage cells is somewhat analogous to Fig. 4 for the atomic clocks. Fig. 7 is a plot of  $\sigma_{yd}(T)$  for each of the five versus a weighted ensemble. The power-law-spectral model for each is clearly not white noise, and, in general, a  $1/f$  behavior seems to be a better model, i.e.,  $\sigma_{yd}(T) \sim T^0$ .

### C. Gauge Blocks

The data analyzed in this section were provided by Dr. R. Hocken's staff of NBS. The data fell into two categories: "mechanical" measurements, which are the comparisons of two gauge blocks of nearly identical size and material; and "interferometric" measurements, which are

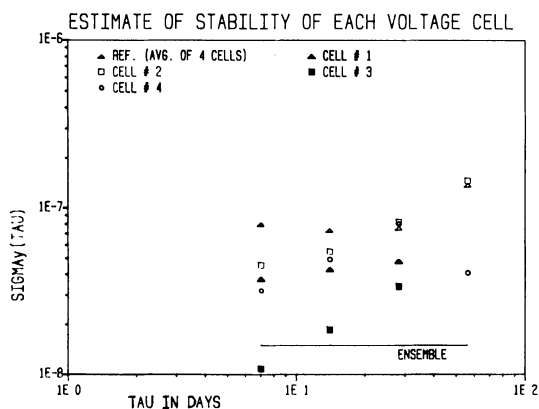


Fig. 7. A plot of the voltage stability  $\sigma_{yd}(T)$ , of the four-cell reference, and of each of the four standard cells with respect to a statistically weighted voltage ensemble. (The abscissa is the stability at sample times of one, two, four, and eight weeks. The voltage deviations appear to be reasonably well modeled by a power-law spectra in the range from  $1/f$  to  $1/f^2$ , with  $1/f$  being a somewhat better model on the whole. The weighted ensemble stability can be described by  $\sigma_{yd}(T) = 15$  nV in the range of sample intervals shown.)

the measurements of gauge blocks with a wavelength standard. The problem of unequally spaced data was a significant one in the case of the gauge-block data. For the mechanical measurements there were some sets that were nearly equally spaced and some not. For the interferometric measurements, most of the data sets were far from equispaced. Hence, the conclusions drawn from the gauge-block data are not as definitive as those for the standard voltage cells and for atomic clocks.

The average time between mechanical measurements was about 18 days, and for the interferometric about 385 days. Twenty mechanical and thirty interferometric time series were analyzed. One 25.4-cm (10-in) Cervit "stack" was also analyzed. Figs. 8, 9, and 10 are respective time series plots from each of these three types of measurements.

Most of the mechanical time series passed the ratio test for white noise; however, the  $\sigma_{yd}(T)$  and  $\text{Mod} \cdot \sigma_y(\tau)$  analyses both indicated some nonwhite (excess low-frequency deviations) noise behavior for  $T$  and  $\tau$  values of the order of three months and longer. Most of the interferometric time series did *not* pass the white noise ratio test, and the value of the ratio averaged over the thirty sets of data was consistent with a  $1/f$  power-law spectral density model. The white noise ratio test given in (7) fails for the Cervit "stack" data shown in Fig. 10. Since the data are unequally spaced and there are only 37 measurements, it is difficult to estimate the spectrum. The ratio is consistent with  $1/f$  noise but certainly not conclusive.

An indication that the long-term deviations are in the gauge blocks rather than in the measurement system is that the square root of the Allan variance at  $\sigma_{yd}(T = 10^8 s)$  was proportional to the lengths of the gauge blocks to within a few percent.

Because much of the gauge block data was unequally spaced, more work is necessary before any definitive conclusions can be drawn. The indications, however, are that the random processes involved are not well modeled by

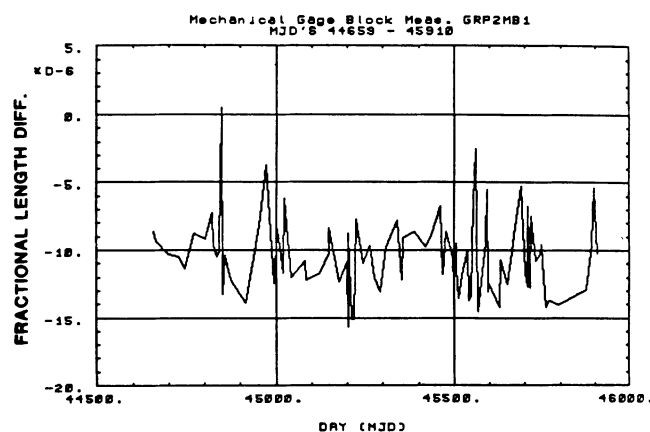


Fig. 8. A plot of the fractional length difference deviations between two 2.54-mm (0.1 in.) gauge blocks over the course of about 1,200 days. (A  $-10 \times 10^{-6}$  would correspond to  $-1 \mu$  in, for example. See Fig. 6 for explanation of MJD.)

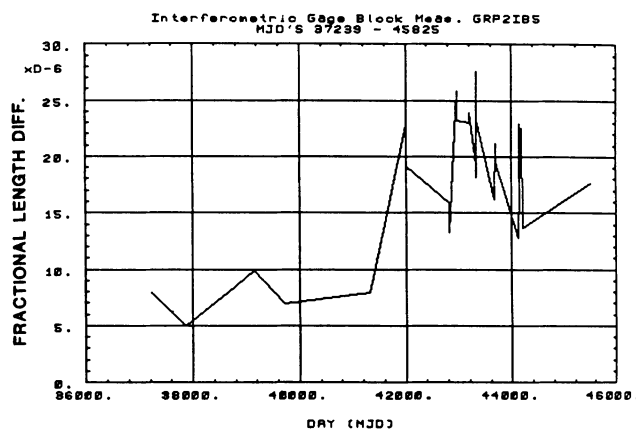


Fig. 9. Similar to Fig. 8, this is a plot of the fractional length difference of another gauge block versus a wavelength standard taken over a little more than 8,000 days. (In contrast to Fig. 8, the data plotted in this figure have more energy in the low Fourier frequencies. See Fig. 6 for explanation of MJD.)

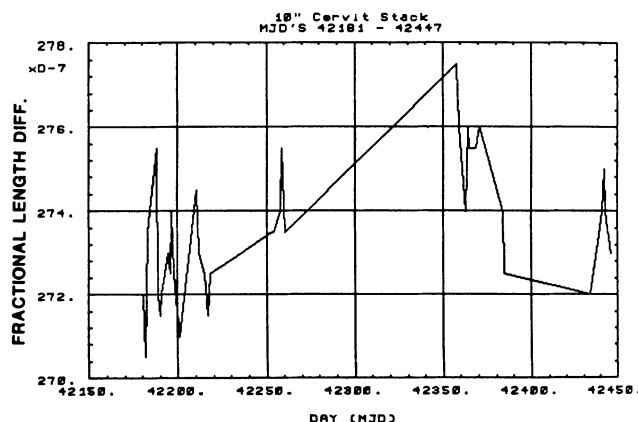


Fig. 10. A plot of the fractional length deviations of a 25.4-cm cervit gauge block stack measured over about 270 days. (The nonwhite character of the deviations is visually apparent.)

white noise. Hence, the use of the classical variance, the mean value, and the standard deviation of the mean should still be called into question for sampling periods of the order of a year and longer.



## IV. CONCLUSIONS

For the data analyzed, one concludes that the classical variance or the standard deviation can give misleading results for the uncertainties in measurements on standard voltage cells and for long-term variations in gauge blocks. This has also been shown to be the case for precision oscillators used in time and frequency metrology. One should study short-term deviations in standard volt cells to determine whether the spectrum is properly modeled by white noise. If the short-term behavior can be shown to be white, then the traditional classical statistical approach is appropriate for averages over that data region. For the long-term deviations in frequency standards, standard voltage cells, and gauge blocks, a white noise model appears not to be viable. The models for the power-law spectral behavior of the long-term deviations can be estimated by the time-domain analysis techniques outlined in the text, except in the cases where the data are significantly unequally spaced. In general, frequency-domain techniques can be employed for spectrum estimation. Knowing the spectral-density models allows estimation procedures to be employed for improved uncertainty assessment. The resulting uncertainties will typically be smaller than those obtained from the classical variance, will better characterize the current "health" of a standard, and will generally be time dependent. There is a simple test to determine if the spectrum of the measurement deviations is white, allowing one to know when and when not to use the classical variance and the standard deviation. Perhaps the real answer to the question posed in the title is a limited yes; safely limited to those cases where the spectrum is white (random uncorrelated measurement deviations). Otherwise, a sample variance approach, as outlined in this paper, should probably be employed. Since the occurrence of nonwhite spectra, giving rise to a divergent classical variance, appears in long-term measurements, the spectrum should be tested for whiteness. If it is nonwhite, then it should be characterized and modeled for proper uncertainty assessment.

## ACKNOWLEDGMENT

The author is indebted to Dr. R. Kamper, Dr. D. Sullivan, Dr. M. Weiss, Dr. J. Moulder, and Mr. D. Davis for extremely helpful comments as readers of this paper. The encouragement and support from Dr. K. Kessler, Dr. R. Kamper, and D. Sullivan, in the completion of this work, are greatly appreciated by the author. Thanks are due to the several people who provided data for the analysis, specifically, Mr. J. Marshall, Dr. T. Doiron, Mr. L. Carroll, Dr. J. Levine, and Dr. J. Beers were very helpful. Ms. T. Peppler has been untiring in assisting with the data analysis.

## REFERENCES

- [1] D. Halford, "A general mechanical model for  $|f|^\alpha$  spectral density random noise with special reference to flicker noise  $1/|f|$ ," *Proc. IEEE*, vol. 56, pp. 251-258, 1968.
- [2] M. S. Keshner, "1/f noise," *Proc. IEEE*, vol. 70, pp. 212-218, 1982.
- [3] L. S. Cutler and C. L. Searle, "Some aspects of the theory and measurement of frequency fluctuations in frequency standards," *Proc. IEEE*, vol. 54, pp. 136-154, 1966.
- [4] D. W. Allan, "Clock specifications, characterization and prediction," in *Proc. 17th Annu. PTTI Meet.*, 1985, pp. 45-68.
- [5] D. W. Allan, H. Hellwig, and D. J. Glaze, "An accuracy algorithm for an atomic time scale," *Metrologia*, vol. 11, pp. 133-138, 1975.
- [6] John Vig, *Proc. IEEE*, vol. 54, Feb. 1966.
- [7] John Vig, *Proc. IEEE Annu. Freq. Control Symp.*, May 1981.
- [8] J. A. Barnes *et al.*, "Characterization of frequency stability," *IEEE Trans. Instrum. Meas.*, vol. IM-20, pp. 105-120, 1971.
- [9] B. E. Blair, Ed., "Time and frequency: Theory and fundamental," NBS, Boulder, CO, NBS Monograph 140, 1974.
- [10] P. Kartaschoff, *Frequency and Time*. New York: Academic, 1978.
- [11] V. F. Kroupa, Ed., *Frequency Stability: Fundamentals and Measurement*. New York: IEEE 1983.
- [12] E. A. Gerber and A. Ballato, Eds., *Precision Frequency Control*, vol. 2. New York: Academic, 1985.
- [13] P. Lesage and C. Audoin, "Characterization and measurement of time and frequency stability," *Radio Sci.*, vol. 14, pp. 521-539, 1979.
- [14] P. Bloomfield, "The practice of spectrum analysis" (Short course), Princeton Univ., Princeton, NJ, Nov. 1979.
- [15] D. A. Howe, D. W. Allan, and J. A. Barnes, "Properties of signal sources and measurement methods," in *Proc. 35th Annu. Freq. Control Symp.*, (NBS, Gaithersburg, MD), May 1981.
- [16] W. C. Lindsay and C. M. Chie, "Theory of oscillator stability based upon structure functions," *Proc. IEEE*, vol. 64, pp. 1652-1666, Dec. 1976.
- [17] P. Lesage and C. Audoin, "Characterization of frequency stability: Uncertainty due to finite number of measurements," *IEEE Trans. Instrum. Meas.*, vol. IM-22, pp. 157-161, June 1973.
- [18] P. Lesage and C. Audoin, "Estimation of the two-sample variance with limited number of data," in *Proc. 31st Ann. Symp. Frequency Control*, (NBS, Gaithersburg, MD), 1977, pp. 311-318.
- [19] J. von Neumann, "Distribution of the ratio of the mean square successive difference to the variance," *Ann. Math. Stat.*, vol. 12, pp. 367-395, 1941.
- [20] B. I. Hart, "Tabulation of the probabilities for the ratio of the mean square successive difference of the variance," *Ann. Math. Stat.*, vol. 13, pp. 207-214, 1942.
- [21] "The relationships for distributed dead time," J. A. Barnes, Austron, Inc., unpublished.
- [22] D. W. Allan and J. A. Barnes, "A modified 'Allan variance' with increased oscillator characterization ability," in *Proc. 35th Annu. Symp. Freq. Control* (Fort Monmouth, NJ), 1981, p. 470.
- [23] P. Lesage, "Characterization of frequency stability: Analysis of the modified Allan variance and properties of its estimate," *IEEE Trans. Instrum. Meas.*, vol. IM-33, pp. 332-336, 1984.
- [24] D. A. Howe, D. W. Allan, and J. A. Barnes, "Properties of signal sources and measurement methods," in *Proc. 35th Annu. Freq. Control Symp.* (NBS, Gaithersburg, MD), May 1981.
- [25] M. J. Lighthill, *Introduction to Fourier Analysis and Generalized Functions* (Cambridge monograph on mechanics and applied mathematics), Cambridge Univ., Cambridge, MA, 1964.
- [26] J. A. Barnes, "Tables of bias functions,  $B_1$  and  $B_2$ , for variances based on finite samples of processes with power law spectral densities," NBS, Boulder, CO, NBS Tech. Note 375, 1974.
- [27] D. W. Allan, J. H. Shoaf, and D. Halford, "Statistics of time and frequency data analysis," NBS, Boulder, CO, NBS Monograph 140, pp. 153-190, 1974.
- [28] D. W. Allan, "Statistics of atomic frequency standards," in *Proc. IEEE*, vol. 54, pp. 221-230, 1966.
- [29] R. H. Jones, "Maximum likelihood fitting of ARMA models to time series with missing observations," *Tecnometrics*, vol. 22, no. 3, pp. 389-395, Aug. 1980.
- [30] R. H. Jones, "Continuous time series models for unequally spaced data applied to modeling atomic clocks," *SIAM J. Stat. Comput.*, vol. 7, 1986.
- [31] D. W. Allan, "Statistics of atomic frequency standards," *Proc. IEEE*, vol. 54, no. 2, pp. 221-230, Feb. 1966.
- [32] P. V. Tryon and R. H. Jones, "Estimation of parameters in models for cesium beam atomic clocks," *J. Res.*, vol. 88, pp. 3-16, 1983.
- [33] J. Marshall, Measurement Standards Lab., Hewlett Packard Co., Santa Clara, CA, unpublished.
- [34] J. S. Beers, "A gage block measurement process using single wavelength interferometry," NBS, Boulder, CO, NBS Monograph 152, Dec. 1975.