1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

AN EXTENSION OF THE ALLAN VARIANCE WITH INCREASED CONFIDENCE AT LONG TERM

D. A. Howe
Time and Frequency Division, National Institute of Standards and Technology
Boulder, CO 80303

ABSTRACT - I build a new statistic for the characterization of frequency stability of clocks and oscillators. It improves on the traditional Allan variance statistic especially at long averaging times.

INTRODUCTION AND SUMMARY

The impulse response of the commonly used technique of differencing of independent random variables overlaps (correlates) "odd" and "even" deviates except the first and last one if the data are not wrapped. This implies there are about twice as many deviates as there are actual degrees of freedom in the mean-square, and the resulting sample Allan variance is appropriately divided by 2(M-1). We obtain

$$\hat{\sigma}^2_{\bar{y}} = \frac{1}{2(M-1)} \sum_{k=1}^{M-1} (\bar{y}_{k+1} - \bar{y}_k)^2 = \frac{1}{2(M-1)} \sum_{k=1}^{M-1} (\nabla \bar{y}_{k+1})^2,$$

where $\{\bar{y}_{k'}\}$, k' = 1,2,3,...,N-1 are fractional frequency differences averaged over τ_0 and $\{\bar{y}_k\}$, k = 1,2,3,...,M are fractional frequency differences averaged over interval $m\tau_0$. Hence $\hat{\sigma}^2_{\bar{y}}$ is implicitly dependent on dimensionless quantity m, a scale parameter which for efficiency can be limited to rational powers of 2, i.e., 2^i=m, i=0,1,2,3,...

The sample Allan variance is useful as a power-law (octave band) spectral estimator but is time-shift (phase) sensitive and depends on where we start the calculation with respect to data in process. For large data sets and small scale values of m, the odd and even deviates average together in the equation for a fairly accurate estimation of a broadband spectral distribution or variance of first differences. The division by 2(M-1) is arguably due to overlapping two sets of deviates but ought to approach M-1 for an accurate estimate in the statistic as $m\tau_0 \rightarrow T/2$ since the first and last deviates do not overlap (unless the data are wrapped). Under most conditions, wrapping the data improves the estimate. However, at the largest scale, the estimate degrades for another reason, namely the sample variance has no response to an evenly symmetric function at this scale. These reasons cause an estimation error or what can be misinterpreted as a "bias" at long intervals in virtually all cases even though the estimation is actually unbiased.

Realizing this, we can re-express deviates in terms of "in-phase" and "phase-shifted" versions of sample variances. Examining this result reveals that the argument in the sample Allan variance above is essentially twice the in-phase variance only. Shifting the observation or sampling window by τ_0 and wrapping the data with an overall frequency difference removed will conveniently yield N-1 sample variances which can be averaged with the usual in-phase sample Allan variance. For $\{\bar{y}_{k'}\} = \bar{y}_1,\ldots,\bar{y}_{N-1}$, this statistic is given by

$$\hat{\sigma}^2_{total}(\tau) = \frac{1}{N-1} \sum_{j}^{N-1} \left[\frac{1}{2(M-1)} \sum_{k=1}^{M-1} (\nabla \bar{y}_{k+1,j})^2 \right]$$

where $\{\bar{y}_{k',j}\} = \bar{y}_{j+1}, \bar{y}_{j+2},\ldots, \bar{y}_{N-1}, \bar{y}_1, \bar{y}_2,\ldots,\bar{y}_j$ are spaced by τ_0 and $\{\bar{y}_{k'}\}$ is therefore wrapped and reindexed by j. $\{\bar{y}_{k,j}\}$ are averages implied over $\tau = m\tau_0$. This variance reduces estimation errors universally seen in previous treatments, thereby providing a much better estimate of frequency stability for measurement times longer than say 20% of the data length.

DISCUSSION

This paper presents a very simple example which shows a common and heretofore undiscussed pitfall in the calculation of the sample version of the Allan variance for long averaging times τ. The sample Allan variance is preferred over the modified variance for long-term frequency stability estimation since τ can be as long as T/2 whereas 50% more data length is required overall for the modified version

using the same τ. The concepts of this paper can easily be extended to the modified variance however.

In frequency-standards metrology, measurements of average relative frequencies are passed through a first-order high-pass filter which is used for removing nonstationary behavior [Box and Jenkins, 1970]. A digital (discrete) version of such a filter (known as finite impulse response or FIR) is routine for creating stationary variates from original nonstationary variables with a "red" PSD having a power-law noise process no steeper than $\propto f^{-2}$ [Rabiner and Gold, 1975]. Furthermore it necessarily correlates otherwise random uncorrelated measurements assumed to be independent [Beran, 1992] [Box, Hunter, Hunter, 1978]. The sample variance of overlapping (not necessarily maximally overlapped) first-differences quantifies spectral features of the data using a smooth broadband, constant-Q, equivalent frequency-domain response [Allan, Weiss, Jespersen, 1991] [Howe and Percival, 1994]. This filter unfortunately has deep nulls at the reciprocal of the averaging time and its harmonics. For broadband noise, a sequence of many overlapping deviates is averaged; hence there is an average of as many arbitrary phases associated with an equivalent frequency-domain filter. However, for long scales, the nulls in the filter response can dramatically affect the result because averaging cannot be claimed as an advantage since only one or two deviates may be involved. The problem is that particular stochastic processes having even functional symmetry over the finite observation can put virtually all of the noise power precisely in the filter's nullpoints at the longest scale.

FIR filter concepts are used in this paper because they relate directly to the nth-order differences used in the statistics quantifying frequency stability. They also help clarify the problem addressed in this paper in using the sample Allan variance.

To begin, the N-sample standard variance is not convergent for the often encountered case of red noise processes of the measured relative phase of an oscillator $\{x_{k'}\}$; it is highly variable with averaging time $\tau = m\tau_o$ or correspondingly half of the reciprocal sampling frequency and its aliases. We assume that differencing techniques allow probability theory to be applied to such non-convergent time series. Differencing creates the new series $\{x'_1,\ldots,x'_{N-1}\}$ formed from the original series $\{x_1,\ldots,x_N\}$ by the "∇" operator as [Brockwell and Davis, 1987]

$$x'_k = x_{k+1} - x_k = \nabla x_{k+1}. \tag{1}$$

Essentially, differences of high enough order "pre-whiten" measured data which is then subject to classical statistical treatment. A first-difference of average frequency has proven to be a simple and adequate whitening filter for oscillators [Barnes, et al., 1971]. Thus the two-sample (first difference) standard variance of frequency was one of the first suggested frequency stability measures [Allan, 1966]. Briefly, it is (using the notation in [Allan, 1966]):

$$\sigma_y^2(\tau) = \sigma_{std}^2(2, \tau = T). \tag{2}$$

All measurements are discrete, not continuous. Interval Δk is taken to be minimum (denoted in a time series by τ_0) with other longer intervals given by $m\Delta k$ (denoted by τ), where m = 1,2,3,....for efficiency however m is often limited to increments of rational powers of 2, i.e., $m=2^i$, i = 0,1,2,3,... More specifically, we start with assumed independent discrete variables $\{x_{k'}\}$, sampling (spatial or temporal) interval Δk (separating time marks $t_{k'}$ if a time series), and m which designates the shift or "stride" in an unprimed index k derived from k′ such that k is incremented once for every mth increment of k′. Hence k = k′/m, and $m\Delta k$ (i.e., τ) is the physical spacing or scaling of measurements within total observation M or T for k = 1,2,3,...,M. Recall that M and T are actually dimensions of space or time which are proportioned to dimensionless integers for mathematical convenience. Although confusing, averaging-time dimension "τ" is often used in the context of statistics having dimensionless Δk, k, and M such that $\frac{m\Delta k}{M} = \frac{m\tau_o}{T}$. Most commonly, the unprimed index k has an implied scale m and $\{x_{k,m}\}$ is a sequence with spacing $m\Delta k$ of measured phase differences between two oscillators; $\{\bar{y}_{k,m}\}$ are average frequency differences. That is

$$\bar{y}_{k,m} = \bar{y}_{k,m}(t_{k'} - m\tau_o, t_{k'}) = \frac{1}{m\tau_o} \int_{t_{k'}-m\tau_o}^{t_{k'}} y(t')dt' = \frac{x_{k'-m}(t_{k'} - m\tau_o) - x_{k'}(t_{k'})}{m\tau_o} = \frac{x_{k-1,m} - x_{k,m}}{m\tau_o}. \tag{3}$$

Additionally, an important procedure assumed in this paper is that overall frequency difference ($\Delta\bar{y}_{1\to N}$) is removed. This means that $x_N = x_1$ which without loss could be set to 0. From eq (1) and eq (3) it

follows that $\{\bar{y}_{k'}\} = \bar{y}_1,\ldots,\bar{y}_{N-1}$ is such that $\bar{y}_{k'} = \frac{1}{\tau_o} x'_{k'}$, and generally

$$\bar{y}_{k,m} = \frac{1}{m\tau_o} x'_{k,m}. \qquad (4)$$

By removing an initial time offset, we can construct $\{\bar{y}_{k'}\}$ from $\{x_{k'}\}$ and vice-versa.

If process $\{\bar{y}_{k,m}\}$ is white, eq (2) is equivalent to

$$\sigma_{\bar{y}}^2(m\tau_o) = \tfrac{1}{2}\langle(\bar{y}_{k+1,m} - \bar{y}_{k,m})^2\rangle, \qquad (5)$$

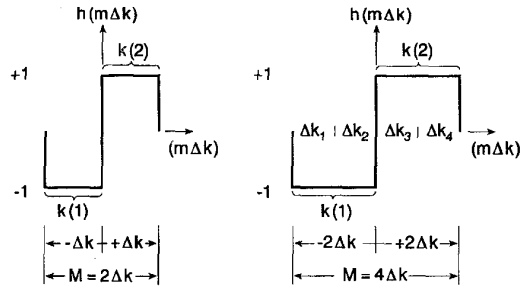
where $\{\bar{y}_{k,m}\}$ is an infinitely long series of nonoverlapping averaged discrete measurements of process y(t); the averaging time is given by $m\tau_o = \tau$; "$\langle\rangle$" denotes an infinite-time ensemble average; and k(origin) = t_0. We of course never can calculate this variance but can calculate a statistic which serves as an estimate. We often refer to a calculation of "the variance" when we usually mean "the statistic" or sample variance.

If series $\{\bar{y}_k\}$ is derived from a continuous function y(t), the fractional frequency difference of two oscillators, then we must infer that each sampled random variable is an average over $\tau = m\tau_o$. The recommended sample variance of first-differences of $\{\bar{y}_k\}$ (denoted hereafter as AVAR) is

$$\hat{\sigma}_{\bar{y}}^2(\tau) = \frac{1}{2(M-1)} \sum_{k=1}^{M-1} (\bar{y}_{k+1} - \bar{y}_k)^2 = \frac{1}{2(M-1)} \sum_{k=1}^{M-1} (\nabla\bar{y}_{k+1})^2. \qquad (6)$$

Equation (3) can be combined with eq (6) for AVAR in terms of $\{x_{k'}\}$. The elements $\bar{y}_{k+1} - \bar{y}_k$ (called "deviates") corresponding to a particular value of τ derive from changes in the data having a corresponding frequency of $\frac{1}{2\tau}$ (and its aliases) and a corresponding phase given by the starting point at k=k(origin) [Howe, Allan, Barnes, 1981]. The sample variance is therefore not only dependent on τ but is additionally always phase-sensitive, i.e., dependent on where we start the calculation with respect to data in process.

Although this paper is not directly about an inconsistent statistic due to an improper number of degrees of freedom, the improper normalization or scaling by degrees of freedom for short data lengths also create an obvious inaccuracy or bias in its corresponding statistic. Briefly, inaccuracy or bias can be quantified using simulated data by an inconsistent trend (given by the slope) of the calculated statistic decomposed as a function of τ or equivalently what is called the power-law characterization (again the slope) of an estimated power spectral density (PSD). The division by 2(M-1) in eq (6) is justified by the argument that the two-sample variance in the infinite limit is the same as the standard variance (which divides by M-1 only) for a white Gaussian power-law process since the cross product of $(\bar{y}_{k+1} - \bar{y}_k)^2$ goes to zero yielding $(\bar{y}^2_{k+1} + \bar{y}^2_k)$, namely, the sum of two standard variances shifted by mΔk. For M = 2Δk and M = 4Δk, eq (6) has finite impulse response (FIR) or "convolving function" shown in figure 1(a) and 1(b).



Typical first-order high-pass impulse responses; mΔk=k.
Deviates ($x_{k+1} - x_k$) are from adjacent random variables.
The usual variance overlaps adjacent deviates.

Fig. 1

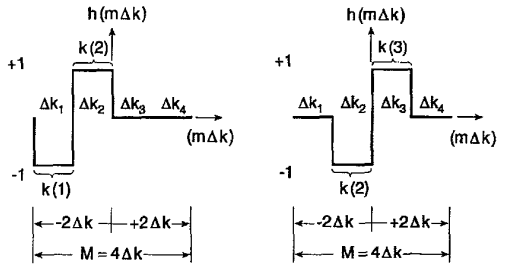
The sample variance of first-difference deviates is an average of these squared deviates at spacings of mΔk within the total observation M as shown in eq (6). Adjacent deviates are not independent. Each element $(\bar{y}_{k+1} - \bar{y}_k)$ is half of the preceding and following elements (except for the first and last). For example, element $(\bar{y}_3 - \bar{y}_2)$ is not independent of preceding element $(\bar{y}_2 - \bar{y}_1)$ since both elements equally include $\bar{y}_2$. Similarly, the following element $(\bar{y}_4 - \bar{y}_3)$ contains $\bar{y}_3$. As a result the variance will correlate each element even though the original data is uncorrelated (as is the case with a white Gaussian process) [Yoshimura, 1976]. This correlation vanishes at the longest scale because deviates no longer overlap, that is, as mΔk → M, there are no overlapped deviates; in this case therefore the division (or normalization) should be M-1 in eq (6) rather than 2(M-1). Wrapping the $\{x_{k'}\}$ data allows 2(M-1) to be used even at the longest scale and hence is a way to get around this particular long-scale bias. Wrapping means that $x_\varsigma = x_{\varsigma \mathrm{mod} N}$ for $\varsigma < 1$ and

$\varsigma > N$, i.e., $x_1 = x_{N+1}$. An assumption of stationarity implies that the wrap can be applied. The procedure also must include matching the endpoints (i.e., $x_N = x_1$) of $\{x_{k'}\}$ to avoid a step in the wrapped data. This, as mentioned, is easily done when an overall frequency difference is removed. We must eliminate the increment x_N to x_{N+1} in the wrap to avoid a potential bias since they are made the same value. Furthermore we must pay attention that N is even (N-1 is odd) because AVAR (in terms of $\{x_{k'}\}$) is a second difference of distinct phase values and could inadvertently respond with 0 at the longest possible scale if N is odd. Using $\{\bar{y}_{k'}\}$ derived from $\{x_{k'}\}$ eliminates the concern regarding whether the total should end up being odd or even.

Yet another pathological analysis error can occur. We first separate the variance into two nonoverlapped versions with odd and even indexes to emphasize that it is a sum of variances [see, for example, Jenkins and Watts, 1968]:

$$\sigma_{\bar{y}}^2(\tau) = \tfrac{1}{2}\langle(\bar{y}_{k+1} - \bar{y}_k)^2\rangle$$
$$= \tfrac{1}{2}[\langle(\bar{y}_{2k} - \bar{y}_{2k-1})^2\rangle + \langle(\bar{y}_{2k+1} - \bar{y}_{2k})^2\rangle] \quad (7)$$

where the first term on the right is the odd-indexed deviates and the second term is the even-indexed deviates.



(a) Adjacent impulse response of odd marks only for m=1

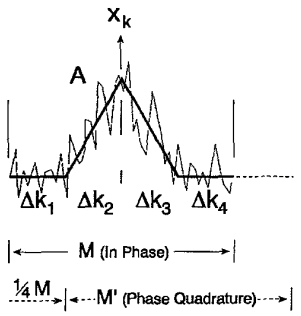
(b) Adjacent impulse response of even marks only for m=1

Fig. 2

To go further, note that deviate $\bar{y}_{k+1} - \bar{y}_k$ requires a total interval of $2m\Delta k$ and is centered at k. That is, $\bar{y}_k$ is derived from the interval spanning k-1 to k, and $\bar{y}_{k+1}$ is derived from span k to k+1. So the location of $\nabla\bar{y}_{k+1}$ is centered at k and its information content spans $2m\Delta k$ with odd indices at k = 1,3,5... and even indices at k = 2,4,6... as shown in figure 2 where m = 1. From this view, the localization of the odd and even marks of span $2\Delta k$ have a fixed phase relationship or phase difference of Δk corresponding to a half-period of a fundamental periodic change to which the deviate identifies and assigns an amplitude. Now consider AVAR as a power-law (constant-Q) spectral estimator [Allan, Weiss, Jespersen, 1991]. For a frequency component in $\{\bar{y}_k\}$ at $f = 1/(2\tau)$, the first term of the bottom expression of eq (7) detects an in-phase or 0° reference phase, and the second term detects an out-of-phase or 180° (relative to 0°) phase. By inspection, there is a fundamental change at the same frequency $f = 1/(2\tau)$ which can go undetected in the interval, namely, one-quarter and three-quarter period (or 90° and 270° phase-shifted) changes.

Deviations of a particular kind can go undetected in a $2m\Delta k$ (or 2τ) interval using AVAR. From a spectral point of view, the usual finite estimation of the original Allan definition is problematic because of this. From a time domain viewpoint the problem is that deviates are zero if the average frequency of the first interval equals the average frequency of the second which is true for an infinite set of even functions. The Allan statistic therefore has no response to any even function over the whole interval. As an example of the problem, hypothetical data are shown in figure 3 by the noisy plot over observation M. The statistic as described by eq (6) for $M = 4\Delta k$ will have deviates made from mean values $\bar{y}_k$ for segments k=1,...,4. A linear fit is indicated by the solid-line segments plotted with slopes 0, $+A/\Delta k$, $-A/\Delta k$, and 0 respectively. The resultant plot has even functional symmetry over M with

$$\bar{y}_1 = \bar{y}_4 = 0$$
$$\bar{y}_2 = \bar{y}_3 = A/2.$$



Example of even symmetric function over observation M

Fig. 3

According to AVAR, the sample statistic of eq (6), $\sigma^2(2\Delta k) = 0$. Obviously, the variance is not really 0 yet our estimate is 0. In this regard, determining and removing a polynomial (particularly a drift coefficient) over the interval has meaning only if the procedure is physically correct. Drift removal is model-dependent [Barnes, 1983]. A popular method involves quantifying drift as an overall second-difference of $\{x_{k'}\}$ which, when removed, results in the usual sample Allan variance being precisely 0 at the longest τ [Weiss, Allan, Howe, 1992] [Weiss and Hackman, 1992].

We replace the Allan two-sample function (which I will denote as "real" or "in-phase") shown in figure 1 and reproduced in figure 4(a) by a phase-shifted version of it to pick up variations in a phase-quadrature component of the data over the coarsest interval or longest scale of τ. This turns out to be a three-sample sequence over each interval which is shown in figures 4(b) and 4(c) and denoted as "imaginary" or "phase-quadrature" functions. The Allan two-sample in-phase function has odd functional symmetry in the interval 2τ as shown in figure 4(a) and hence is suitably an "odd sampling window function." The three-sample sequence has even symmetry in the same interval 2τ as shown in figures 4(b) and 4(c) given by sequence

$$\pm [\bar{y}_{k+3} - (\bar{y}_{k+2} + \bar{y}_{k+1}) + \bar{y}_k],$$

where the average frequency is taken over $\tau/2$ rather than τ.

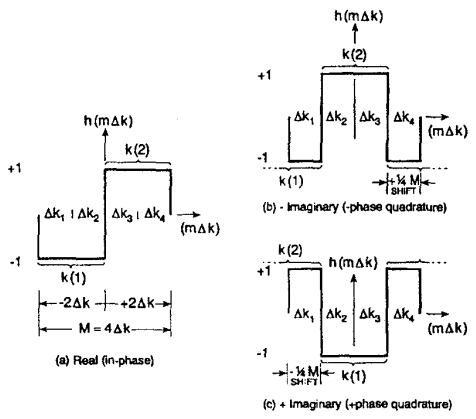


Fig. 4

Without loss eq (7) can be rewritten as

$$\sigma_{\bar{y}}^2(\tau) = \frac{1}{2}\left\langle \left[\frac{(\bar{y}_{k+3} + \bar{y}_{k+2})}{2} - \frac{(\bar{y}_{k+1} + \bar{y}_k)}{2} \right]^2 \right\rangle, \qquad (8)$$

$$\bar{y}_k = \bar{y}_k(t_k - \tau/2,\ t_k),\ k=1,3,5...$$

in which the variance is still for scale τ but where $\{\bar{y}_k\}$ are average frequencies over $\tau/2$ rather than τ.

Modified in the form of finite impulse responses shown in figure 4 and rewritten as a mean-square combination of an "in-phase" variance and "phase-quadrature" variance gives

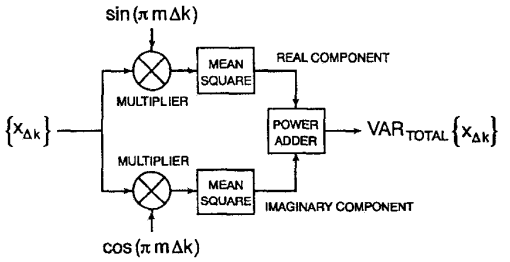
$$\sigma_{\bar{y}*}^2(\tau) = \frac{1}{4}\left[\left\langle \left[\frac{(\bar{y}_{k+3} + \bar{y}_{k+2})}{2} - \frac{(\bar{y}_{k+1} + \bar{y}_k)}{2} \right]^2 \right\rangle + \left\langle \left[\bar{y}_{k+3} - \frac{(\bar{y}_{k+2} + \bar{y}_{k+1})}{2} + \bar{y}_k \right]^2 \right\rangle \right], \qquad (9)$$

where the average frequency is taken over $\tau/2$.

Another variance can be defined from two variances of orthogonal in-phase and phase-quadrature elements. This formalism bears a similarity to the method of complex demodulation used in signal processing [Bingham, Godfrey, and Tukey, 1967]. In this method, time-series data are assumed to contain periodicities and viewed as consisting of separate real and imaginary parts; the variance is calculated for each part and the total is taken as

$$\sigma_{\Delta\, total}^2 = Re|\sigma^2| \oplus Im|\sigma^2| = [Re^2\sigma^2 + Im^2\sigma^2]^{1/2}, \qquad (10)$$

where "$\oplus$" means power-added together. Functionally, the method of complex demodulation is shown in figure 5. The real part is comprised of



Method of complex demodulation

Fig. 5

deviates of the actual data (the in-phase observation) and the $\pm$ imaginary part are the deviates created by shifting the observation window by $\pm$¼T (the phase-

quadrature observation). In the case of a first-order difference filter, we obtain impulse responses precisely as shown in figure 4(a) for the real part and figures 4(b) and 4(c) for the $\pm$ imaginary part. The sign of the imaginary part becomes unimportant as the deviates are squared. The imaginary part that is outside the actual data are wrapped; that is, h(M+k) = h(k). This is justified because one assumes the measurement is for periodic (sine and cosine) functions which are in process, that is, unwindowed and periodically extended.

The estimate to eq (9) is an equivalent description to an "in-phase" and "phase-quadrature" discrete functional component separation. Using this method, the variance of the even-symmetric function is computed by moving the observation window by m/2, for the "phase-quadrature" variance and adding this to the "in-phase" variance. This has a sample statistic given by

$$\hat{\sigma}^2_{\bar{y}}(m\tau_o)=\frac{1}{4(M-1)}\left[\sum_{k=1}^{M-1}(\nabla\bar{y}_{k+1,m})^2+\sum_{k=1}^{M-1}(\nabla\bar{y}^*_{k+1,m})^2\right], \quad (11)$$

$$\text{where } \bar{y}_{k,m}=\frac{x_{k'-m}(t_{k'}-m\tau_o)-x_{k'}(t_{k'})}{m\tau_o},$$

$$\text{and } \bar{y}^*_{k,m}=\frac{x_{k'-m/2}(t_{k'}-\frac{m\tau_o}{2})-x_{k'+m/2}(t_{k'}+\frac{m\tau_o}{2})}{m\tau_o},$$

$x_1, x_2, \ldots x_N$ with wrap such that $x_\varsigma = x_{\varsigma \bmod N}$ for $\varsigma < 1$ and $\varsigma > N$, i.e., $x_1 = x_{N+1}$ which reindexes to x_N as noted earlier. Shifting the data and using a wrap simplifies the form of the sample variance of eq (11) corresponding to eq (9) [for a discussion of wrapped data, see Howe and Percival, 1995; also briefly in Bloomfield, 1976]. Equation (11) is an average of variances whereas eq (10) implies an RSS for orthogonal (sin vs. cos) basis functions or an RMS for random functions. Using simulation of common power laws, the difference between an average and RMS of variances turns out to be negligible in practice. Returning to the previously mentioned hypothetical example and using eq (11), the result is not 0 but rather

$$\hat{\sigma}^2_{\bar{y}^*}(m=2) = \frac{1}{12}\left(\frac{A}{2}\right)^2.$$

HIGHER ORDER DIFFERENCES

Given continuous function f(k) with nth order derivative ($\frac{d^n f(k)}{dk}$ where "k" is the independent variable), there are n derivatives for n possible orders. Continuous function f(k) is an idealization; all physical measurements will sample f(k) in discrete, usually equispaced, increments Δk_k despite the fact that Δk_k may be infinitesimally small. In all cases, therefore, we are representing f(k) as discrete continuous f(mΔk) which describes a space- or time-ordered sampled function with nth order possible differences. The variance is like an average of these squared differences. In this case, there are not n, but always 2^n independent differences in the average for n orders. To preserve independence, the proper variance is therefore the square root of the sum of squares of 2^n variances. To obtain stationary deviates, the order of differencing is usually quite small, frequently 1 or 2. This is because the corresponding (high-pass) impulse response makes new deviates which have a PSD which goes as f^{+2n} for each order n. Hence an f^{-4} PSD would need to be differenced only twice (order 2) to realize white Gaussian noise. Similarly, many functions can be well approximated on an interval of finite length by a polynomial of low degree. For the simple high pass first difference (order 1), there are 2^1 or two independent sets of differences and the proper variance is just the RMS of the two variances from each set as discussed in this paper. For the second difference (order 2), there are 2^2 or four independent sets of differences, and so on.

AN IMPROVED STATISTIC AT LONG TERM

I construct a sample variance (eq (11)) that is an average of variances of separate in-phase and phase-quadrature deviates (time-shifted by m/2) and given by a procedure which wraps $\{x_{k'}\}$ or equivalently $\{\bar{y}_{k'}\}$. This is sufficient to avoid a potentially serious estimation error compared to traditional AVAR at long averaging times. Historically the Allan variance is the composite variance of interest. It formally derives from the standard variance (see eq (2)). We compute various things from finite data which are calculations of statistics (one being AVAR) as estimators of the Allan variance. For a set of noise processes, a good estimator (1) has the same mean as the variance and (2) itself has a low variance. A question is whether the three-sample sequence introduced in this paper when combined with the

original AVAR implies a new definition having better confidence or is a better estimate of the existing definition. Better statistical certainty usually derives from more degrees of freedom. An in-phase convolving function (first term of eq (9)) is being combined with another convolving function at phase quadrature (second term of eq (9)). This is either an extension of the historically used two-sample variance or is a different concept.

We can establish an even better long-term estimator which describes and hence predicts longer-term frequency stability from available measurements. We can average all possible N-1 sample variances, each time shifted by τ_o. For $\{\overline{y}_{k'}\} = \overline{y}_1,\ldots,\overline{y}_{N-1}$, this statistic is given by

$$\hat{\sigma}^2_{total}(\tau) = \frac{1}{N-1}\sum_{j}^{N-1}\left[\frac{1}{2(M-1)}\sum_{k=1}^{M-1}(\nabla\overline{y}_{k+1,j})^2\right], \qquad (12)$$

where $\{\overline{y}_{k',j}\} = \overline{y}_{j+1}, \overline{y}_{j+2},\ldots, \overline{y}_{N-1}, \overline{y}_1, \overline{y}_2,\ldots,\overline{y}_j$ are spaced by τ_o and $\{\overline{y}_{k'}\}$ is therefore wrapped and reindexed by j. $\{\overline{y}_{k,j}\}$ are averages implied over $\tau = m\tau_o$. Equation (3) can be combined into eq (12) for $\hat{\sigma}^2_{total}$ in terms of $\{x_{k'}\}$.

This statistic and its associated impulse responses average the variances of all possible fractional frequency deviates. The new statistic properly picks up and normalizes all functional variations at all τ scales, and at the longest ones, namely as $\tau \rightarrow T/2$. Its use is recommended for at least the longest four scales.

SIMULATION WITH RANDOM WALK OF $\{x_{k'}\}$

Figure 6(a) shows 100 plots of calculations of the square root of $\hat{\sigma}^2_{total}(\tau)$ for 100 simulations of random walk of phase modulation (RWPM). Equation (12) is the expression used for these calculations and N=1024 for each simulation. Each of the simulation averages of two-sample variances is equal to one. The bottom plot is the 100-simulation total mean and shows excellent agreement with theory. Figure 6(b) is the same set of calculations using traditional square root of maximally-overlapped AVAR. We see that the spread in the estimates is greater using AVAR instead of the new statistic in figure 6(a) especially at $\tau = T/2$. The new statistic is preferred at long averaging times since it yields a distribution which is less skewed and which is less susceptible to optimistic (sometimes very optimistic) estimation errors.

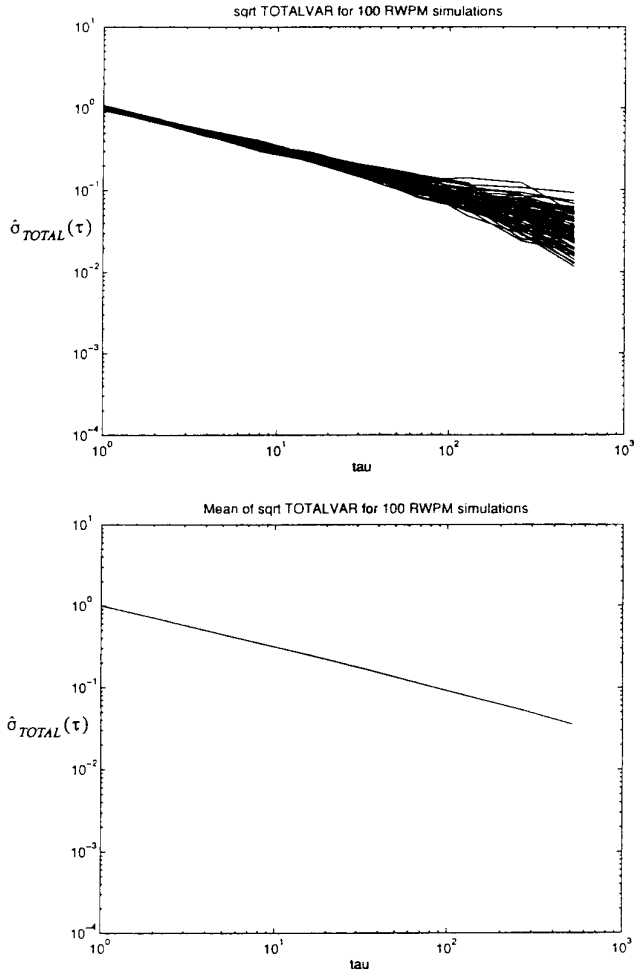


Fig. 6(a)

Top: New statistic (indicated as square root of TOTALVAR) calculated for 100 RWPM simulations with unit (two-sample) mean. Note the reduced skew and spread in the response for long term averaging times compared to traditional square root of maximally-overlapped AVAR shown at right at the top of figure 6(b).

Bottom: Mean of the 100 simulations using square root of TOTALVAR agrees satisfactorily with theory.

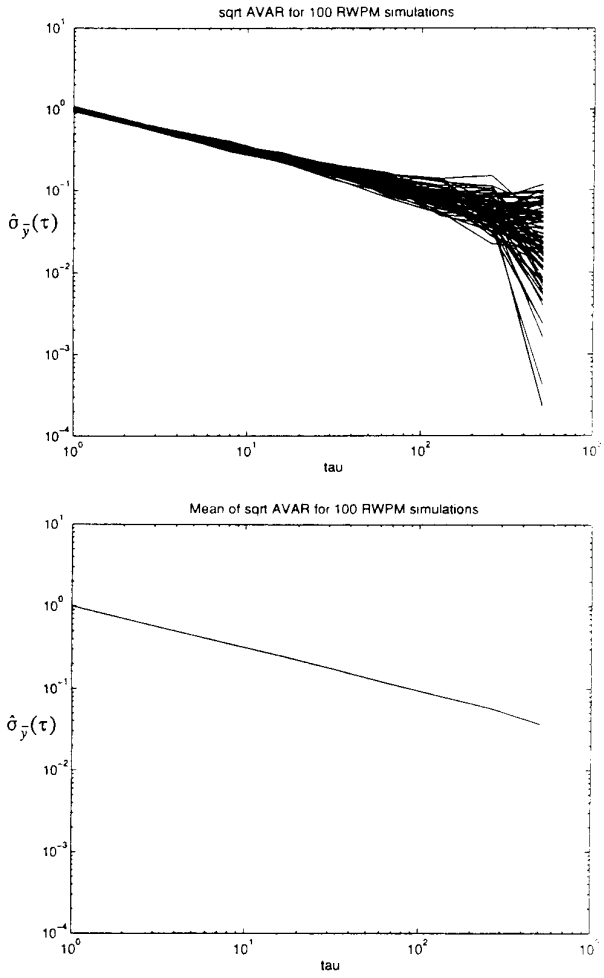


Fig. 6(b)

Top: For comparison, traditional square root of maximally-overlapped AVAR is calculated for the same 100 RWPM simulations as used at left for square root of TOTALVAR in figure 6(a). Note the wider skew and spread for long term averaging times and the tendency toward an optimistic (sometimes very optimistic) response.

Bottom: Mean of the 100 simulations using square root of maximally-overlapped AVAR is shown.

REFERENCES

Allan, D. W., Statistics of atomic frequency standards, Proc. IEEE 54, 221-230 (1966).

Allan, D.W. and Barnes, J.A., A modified "Allan Variance" with increased oscillator characterization ability, Proc. 35th Freq. Control Symp., 470-476 (1981).

Allan, D.W., Weiss, M.A., and Jespersen, J.L., A frequency-domain view of time-domain characterization of clocks and time and frequency distribution systems, Proc. 45th Frequency Control Symposium, 667-678 (1991).

Barnes, J.A., Atomic timekeeping and the statistics of precision signal generators, Proc. IEEE 54, 207-220 (1966).

Barnes, J. A., Chi, A. R., Cutler, L. S., Healy, D. J., Leeson, D. B., McGunigal, T. E., Mullen, J. A., Jr., Smith, W. L., Sydnor, R. L., Vessot, R. C. F., and Winker, G. M. R., Characterization of frequency stability, IEEE Trans. Instrum. Meas., IM-20, 105-120 (1971).

Barnes, J.A., The measurement of linear frequency drift in oscillators, Proc. 15th Ann. PTTI Meeting, 1983. Article is also contained in Natl. Inst. Stand. Technol. TN-1337, 1990.

Beran, J., Statistical methods for data with long-range dependence, Statistical Science, 7, 404-427 (1992).

Bingham, C., Godfrey, M. D., and Tukey, J. W., Modern techniques of power spectrum estimation, IEEE Trans. Audio Electroacoustics, AU-15, 56-66 (1967).

Bloomfield, P., Fourier analysis of time series: an introduction, Wiley, New York, 1976. Pages 55-60 are a useful discussion of the "circular definition" of a noise (or roughness) coefficient, namely the variance, using "wrapped" data.

Box, G. E. P., Hunter, W. G., and Hunter, J. S., Statistics for experimenters, Wiley, New York, 1978.

Box, G. E. P. and Jenkins, G. M., Time series analysis: forecasting and control, 1970, Holden-Day, San Francisco.

Brockwell, P. J. and Davis, R. A., Time Series: Theory and Methods, 1987, Springer-Verlag, New York.

Cutler, L.S. and Searle, C.L., Some aspects of the theory and measurement of frequency fluctuations in frequency standards, Proc. IEEE 54, 136-154 (1966).

Howe, D.A., Allan, D.W., and Barnes, J.A., Properties of signal sources and measurement mehtods, Proc. 35th Ann. Symp. Frequency Cont., Philadelphia, PA, May 27-29, 1981, A25-A35, 1981. Article is also contained in Natl. Inst. Stand. Technol. TN-1337, 1990.

Howe, D.A. and Percival, D.B., The wavelet variance, Allan variance, and leakage, Proc. IEEE Trans. on I & M, IM-44, no. 2, 94-97 (1995).

Howe, D.A. and Percival, D.B., The role of circularity in analyzing the variance of a time series, to be published. Contact D.A. Howe, NIST, 325 Broadway, Boulder, CO 80303.

Howe, D. A., Circular representation of infinitely extended sequences, to be published, Proc. 1995 IEEE Freq. Control Symp. (this issue).

Lindsey, W. C. and Simon, M. K., Telecommunication system engineering, 1973, Prentice-Hall, New Jersey.

Rabiner, L. R. and Gold, B., Theory and application of digital signal processing, 1975, Prentice-Hall, Inc., New Jersey.

Weiss, M.A., Allan, D.W., and Howe, D.A., Confidence on the second difference estimation of frequency drift, Proc. IEEE 46th Freq. Control Symp., 300-305, 1992.

Weiss, M.A. and Hackman, C., Confidence on the three-point estimator of frequency drift, Proc. 24th Ann. PTTI Meeting, 451-460, 1992.

Yoshimura, K., Characterization of frequency stability: Uncertainty due to the autocorrelation of the frequency fluctuations, IEEE Trans. Instrum. Meas., IM-27, 1-7 (1978). The number of degrees of freedom is about half the total number of deviates, so the scaling factor ends up being $\simeq 1/2(M-1)$ as shown in eq (6). It is exactly ½ in the infinite limit. This paper is concerned with the typical finite case and the special case of just a few deviates.