INSTITUTE OF PHYSICS PUBLISHING METROLOGIA
Metrologia **43** (2006) S322–S331 doi:10.1088/0026-1394/43/4/S20

ThêoH: a hybrid, high-confidence statistic that improves on the Allan deviation

D A Howe

National Institute of Standards and Technology, 325 Broadway, Boulder, CO 80303, USA

Received 2 November 2005
Published 4 August 2006
Online at stacks.iop.org/Met/43/S322

Abstract

'Theo1' (short terminology for 'theoretical variance #1') is derived based on $\psi_y^2(\tau_s, \tau)$, a variance that separates the Allan variance's *averaging time* (τ) from its *sampling interval* (τ_s), also called 'stride', as applied to normalized fractional-frequency $y(t)$. The τ-length average of $y(t)$ is denoted ${}^{\tau}\overline{y}(t)$, and every permissible squared difference of ${}^{\tau}\overline{y}(t)$-values in $\psi_y^2(\tau_s, \tau)$ is averaged in $m\tau_0$ segments of $y(t)$ to compute Theo1(m, τ_0). Although biased relative to the Allan variance, Theo1(m, τ_0) mimics all statistical properties of the Allan variance and works to very long stride. Its estimator $\widehat{\text{Theo1}}(\tau_s)$ has substantially better confidence compared with the best Allan estimator $\hat{\sigma}_y^2(\tau_s)$, called 'Avar'. To make an Allan-compatible statistic, a novel 'bias-removed' version of $\widehat{\text{Theo1}}$ variance, called '$\widehat{\text{TheoBR}}$' variance, yields ThêoH variance ('H' to indicate high confidence and/or hybrid Allan and $\widehat{\text{TheoBR}}$ functions). ThêoH deviation combines the Allan deviation in short term and $\widehat{\text{TheoBR}}$ deviation in long term, reporting very long-term frequency stability 50% beyond that possible using the Allan deviation alone and with excellent confidence.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

For a given data run, it is often vital to reliably determine the level and type of noise associated with the longest-term frequency stability. Three examples of such a need are: (1) in an ensemble of clocks, steering each clock for combined best overall frequency stability relies on a weighting algorithm that depends on accurate values of long-term frequency stability and correct FM-noise identification; (2) in a primary frequency standard, an important test of claimed accuracy (the threshold) is that the flicker frequency modulation (FLFM) floor must be the same as or lower than any stated accuracy; and (3) oscillators that are mass-produced require standardized stability measurements out to the longest possible averaging time τ, given that they are tested for a relatively short time.

Frequency stability is characterized by the square root of the Allan variance [1]. Throughout this paper, 'Avar' is the overlapping Allan variance estimator and 'Adev' is its square root [2]. If long-term measurements of frequency stability are important, a 'total' Allan variance estimator, called Total Avar, is recommended [3]. Total Avar is simply Avar that is applied to a τ-length extension of the original data [4–6]. It is an example of a 'hybrid' statistic in the sense of improved estimation of frequency stability at long-term τ-values, where the extension is greatest, while essentially computing plain Avar at mid- and short-term τ-values, where the extension becomes progressively smaller and smaller [7–10]. Total Avar(τ, T) has dependence on τ and data-run length T. The spectral filter response associated with Total Avar's sampling for large τ-values is smoother than Avar for determining FM noise and is the main reason for using Total Avar in the first place [11]. Confidence, expressed as the equivalent degrees of freedom (edf), increases from Avar's 1 to Total Avar's 1.5, 2.1 and 3 in the presence of random-walk frequency modulation (RWFM), FLFM and white frequency modulation (WHFM), respectively, for computations of frequency stability at $\tau = T/2$. There is an easily-removed slight negative bias associated with RWFM and FLFM noises [7, 11].

Analysts of frequency stability are confident of Total Avar's properties [9] and take advantage of its benefits [12]. Easy-to-use 32-bit Windows software is commercially distributed which implements Total Avar on large data sets,

computes its confidence intervals and automatically adjusts for bias[1]. Interestingly, Total Avar can be computed for intervals beyond half the data run, but these values, not reported since Avar itself is undefined in this region [8], contain useful information [11].

ThêoH variance to be described here outperforms max-overlap estimators Avar and Total Avar, having respective edfs at $\tau = T/2$ of 2.1 for RWFM, 4.3 for FLFM, and 6 for WHFM. For a primary frequency standard properly exhibiting WHFM, ThêoH variance has a six-fold increase in edfs, hence, subsantially improved confidence, on the reported level of frequency stability at the maximum-τ value of Avar [13]. ThêoH variance also gives levels of frequency stability with good confidence beyond the maximum-τ of Avar [14, 15].

Section 2 defines a generic first-difference variance called 'psi-variance' ($\psi_y^2(\tau_s, \tau)$) in which measurement dead time is a free parameter given by sampling interval τ_s called 'stride,' minus averaging time or $\tau_s - \tau$. Nonzero dead time expressed as $\tau_s \neq \tau$, causes bias relative to Avar that depends on the ratio of dead-to-live time and so is generally thought to be detrimental compared to Allan's zero-dead-time approach [16–18]. However, simulation studies show that confidence is not substantially lower with dead time [19]. An average of psi-variance confined to a span of $m\tau_0$, where m is an integer, is defined representing a new class of statistics called 'Theo1 variance' with recognized benefits [12–15, 20]. Throughout this paper, the maximum-overlap estimator of Theo1 is designated '$\widehat{\text{Theo1}}$.' The properties of $\widehat{\text{Theo1}}$ variance are compared with Avar in section 3. Section 4 gives a technique for substantially reducing long-term bias relative to Avar, which leads to the ThêoH variance defined in section 5. Section 6 gives $\widehat{\text{Theo1}}$ edf formulae. Section 7 shows that by using ThêoH variance, the primary goal of distinguishing power-law noise types and levels is attained substantially better than by using Avar.

2. Processing fractional-frequency data

2.1. A generic two-sample frequency variance with dead time: $\psi_y^2(\tau_s, \tau)$

For comparison to the Allan variance, define a continuous-analogue two-sample frequency variance based on average frequency differences spaced by stride τ_s as follows:

$$\psi_y^2(\tau_s, \tau) \equiv \frac{\tau}{\tau_s}\langle[{}^{\tau}\overline{y}(t) - {}^{\tau}\overline{y}(t-\tau_s)]^2\rangle, \qquad (1)$$

where ${}^{\tau}\overline{y}(t)$ is an average frequency over duration $\tau = m\tau_0$ and $\langle\cdot\rangle$ denotes an ensemble average of realizations of sequential ${}^{\tau}\overline{y}(t)$ [21]. In terms of the time-error function $x(t)$ between two oscillators, ${}^{\tau}\overline{y}(t) = \frac{x(t)-x(t-\tau)}{\tau}$. Definition (1) is based on taking sequential mean frequency measurements spaced τ_s apart, differencing them, and computing the mean square [22]. The τ/τ_s makes the expected value the same for WHFM irrespective of τ and τ_s for a constant interval. Figure 1 shows the sampling-sequence function associated with $\psi_y^2(\tau_s, \tau)$ acting on $y(t)$ for different τ_s and τ in interval $m\tau_0$. The 'sampling sequence' is that function which is convolved with the frequency-data vector before computing the mean-square variance. τ_n in figure 1 is one discrete realization from the set of all those possible, and τ_{s_n} is its corresponding stride. As τ increases from τ_0 to $(m/2)\tau_0$ in figure 1(*a*)–(*d*), stride τ_s decreases and averages to $\overline{\tau_s} = (1/N)\sum_{n=1}^{N}\tau_{s_n} = 0.75m\tau_0$.

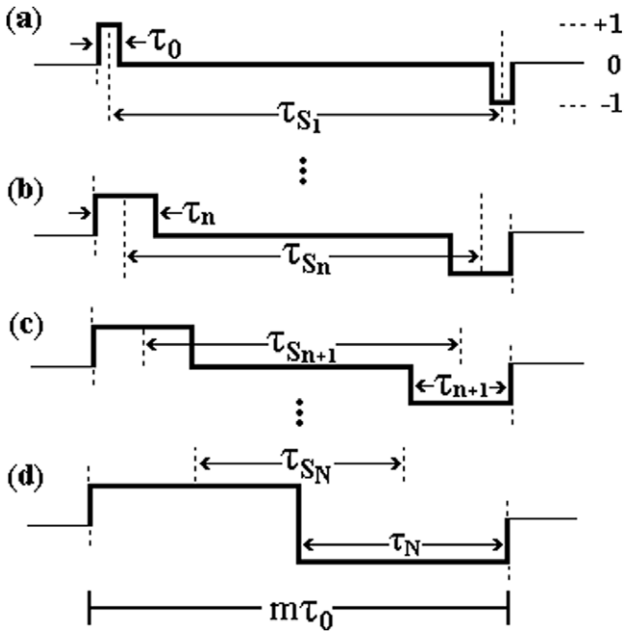

Figure 1. Theo1 variance averages $\psi_y^2(\tau_s, \tau)$ over all τs constrained to a span of length $m\tau_0$. (*a*) and (*d*) correspond to sampling at *max*-stride $\psi_y^2((m-1)\tau_0, \tau_0)$ and *min*-stride $\psi_y^2((m/2)\tau_0, (m/2)\tau_0)$, respectively. (*b*) and (*c*) are realizations between *max* and *min* separated by one m increment. Note that average stride $\overline{\tau_s} = 0.75m\tau_0$.

2.2. Definition of Theo1

A new class of long-stride statistics is introduced in this writing by 'Theo1' that averages $\psi_y^2(\tau_s, \tau)$ over combinations of τ_s and τ. τ_s, τ, $x(t)$ and $y(t)$ are continuous, but, in fact, actual measurements are represented as digital, equi-spaced discrete numbers in a time-series. It is usual practice to measure the time-error between two oscillators with a time interval counter [2]. Measurement samples of the time-error function $x(t)$ occur at a rate f_s having an interval $\tau_0 = 1/f_s$. Given a sequence of time errors $\{x_n : n = 1, \ldots, N_x\}$ with a sampling period between adjacent observations given by τ_0, we define the $\alpha\tau_0$-average fractional-frequency deviate as

$$\overline{y}_n(\alpha) \equiv \frac{1}{\alpha}\sum_{j=0}^{\alpha-1} y_{n-j},$$

where we obtain $\{y_n\}$ from $y_n = (1/\tau_0)(x_n - x_{n-1})$ with $n = t/\tau_0$ starting from a designated origin $t_0 = 0$. Index n, integers m and N_x and sample-rate period τ_0 are used in discrete-continuous versions of actual measurements [23].

Theo1 variance averages $\psi_y^2(\tau_s, \tau)$ in interval $m\tau_0$. Its expectation value, in terms of symmetric differences of

[1] Stable 32, Frequency Stability Analysis for Windows and NT, Hamilton Technical Services, 650 Distant Island Drive, Beaufort, SC 29907, USA, Phone: 843-525-6495, http://www.wriley.com. No endorsement is implied.

time-error values $\{x_n\}$, is

$$\text{Theo1}(m,\tau_0) \equiv \frac{1}{2}\left\langle \frac{2}{0.75m}\sum_{\delta=0}^{\frac{m}{2}-1}\frac{1}{m(\frac{m}{2}-\delta)} \times \frac{1}{\tau_0^2}\left[\left(x_{n+\frac{m}{2}}-x_{n+\delta}\right)-\left(x_{n-\delta}-x_{n-\frac{m}{2}}\right)\right]^2\right\rangle, \quad (2)$$

where τ_0 is the minimum sampling interval. $\text{Theo1}(m,\tau_0)$ characterizes frequency stability at stride $0.75(m-1)\tau_0$. For WHFM, Theo1 variance is made equal to the Allan variance when the strides in each are equal.

2.3. Estimator $\widehat{Theo1}$

The hat '^' over Theo1 variance denotes an estimator for length N_x consistent with nomenclature '$\hat{\sigma}_y^2(\tau)$,' or what has been called Avar, the maximum-overlap, an unbiased estimator of the Allan variance [2–4]. The maximum-overlap estimator of Theo1 is the main subject of all discussions hereafter and is given by

$$\widehat{\text{Theo1}}(m,\tau_0,N_x) = \frac{1}{0.75(N_x-m)(m\tau_0)^2} \times \sum_{i=1}^{N_x-m}\sum_{\delta=0}^{\frac{m}{2}-1}\frac{1}{(m/2)-\delta} \times[(x_{i+m}-x_{i+\delta+\frac{m}{2}})-(x_{i-\delta+\frac{m}{2}}-x_i)]^2, \quad (3)$$

for m even, $10 \leqslant m \leqslant N_x - 1$, where frequency stability is evaluated, in essence, at an average sampling interval $\overline{\tau_s} = 0.75m\tau_0$. $\widehat{\text{Theo1}}$ variance estimates mean $\psi_y^2(\tau_s,\tau)$ by summing over all τ_{s_n} and τ_n in span $m\tau_0$. This means that the longest $\overline{\tau_s} = 0.75(N_x-1)\tau_0$ or an averaging time corresponding to 3/4 of the duration of a data-run length. This is 50% longer than the longest τ-value of maximum-overlap estimator $\hat{\sigma}_y^2(\tau)$ in which m's range is only $1 \leqslant m \leqslant (N_x-1)/2$ [2,3].

In (3), the symmetry of $\widehat{\text{Theo1}}$'s sampling (see figure 1) means that m can take only even values, that is, $m = 2, 4, 6, \cdots$. Thus, $\widehat{\text{Theo1}}(\tau)$'s evaluations actually occur at $\tau = 1\tau_0, 3\tau_0, 5\tau_0, \cdots$. An evaluation of $\widehat{\text{Theo1}}(m-1,\tau_0,N_x)$ in a time-domain setting simply means $\widehat{\text{Theo1}}(\overline{\tau_s}-\tau_0)$. As a practical matter, since τ_0 and its effect can be made arbitrarily small, $\widehat{\text{Theo1}}(\overline{\tau_s}) = \widehat{\text{Theo1}}(\tau)$ just as $\hat{\sigma}_y^2(\tau_s) = \hat{\sigma}_y^2(\tau)$ with Avar.

3. Properties of $\widehat{Theo1}(\tau)$

3.1. Compatibility with Allan deviation

The square-root of $\widehat{\text{Theo1}}$-variance of (3), or $\widehat{\text{Theo1}}$-deviation is in convenient units of fractional frequency fluctuation, like square-root-of-Avar, or Adev. m, as always, is the number of units of data spacing τ_0 in seconds that defines time interval τ such that $\tau = 0.75m\tau_0$. Thus, notation $\widehat{\text{Theo1}}(\tau)$ means $\widehat{\text{Theo1}}(\tau = \overline{\tau_s} = 0.75m\tau_0)$.

Table 1 gives coefficients for transforms of $\widehat{\text{Theo1}}$-deviation level and slope on log–log plot to the square root of noise spectrum, $\sqrt{S_y(f)}$, that is, carrier-signal rms fractional frequency fluctuations $(\Delta\nu/\nu_o)_{\text{rms}}$ in a 1 Hz bandwidth evaluated at Fourier frequency f. 'QPM' is the quantized phase modulation, and 'QSFM(f_0)' means quasi-sinusoidal frequency modulation (QSFM) at f_0. Coefficients are consistent with straight-line (on log–log plots) Allan-deviation transforms to the Fourier-frequency domain [24].

Table 1. Transform coefficients of $\widehat{\text{Theo1}}$-deviation level and slope on log–log plot in the τ domain to $(\Delta\nu/\nu_o)_{\text{rms}}$, the square root of noise spectrum, $\sqrt{S_y(f)}$, in the Fourier-frequency domain. QPM is quantized phase modulation. QSFM(f_0) means QSFM at f_0.

Noise type	$\widehat{\text{Theo1}}$-dev level a_μ	$\widehat{\text{Theo1}}$-dev slope $\frac{\mu}{2}$	$\sqrt{S_y(f)}$ or $\left(\frac{\Delta\nu}{\nu_o}\right)_{\text{rms}}$ versus f
QPM[a]	a_{-2}	-1	$f\frac{1}{2\sqrt{3f_h}}\tau a_{-2}$
WH/FL PM[a]	a_{-2}	-1	$f^{\frac{1}{2}}\frac{\pi}{\sqrt{3f_h}}\tau a_{-2}$
WH FM	a_{-1}	$-\frac{1}{2}$	$\sqrt{2\tau}a_{-1}$
FL FM	a_0	0	$f^{-\frac{1}{2}}\frac{1}{\sqrt{ln2}}a_0$
RW FM	a_{+1}	$+\frac{1}{2}$	$f^{-1}\frac{1}{\pi}\sqrt{\frac{6}{\tau}}a_{+1}$
Drift	a_{+2}	$+1$	$f^{-\frac{3}{2}}\frac{2}{\tau}a_{+2}$
QSFM(f_0)	$\frac{1}{2}\frac{\Delta\nu}{\nu_o}\left(\frac{\sin^2\pi f_o\tau}{\pi f_o\tau}\right)$	-1 (avg)	$\frac{\Delta\nu}{\nu_o}(f_o)$

[a] Requires f_h (high-frequency cutoff), $2\pi f_h\tau \gg 1$.

Correlation of octave τ-values using $\widehat{\text{Theo1}}$ variance is equivalent to that of Avar and is qualitatively tested and compared by measuring each statistic's expected $1/\tau^2$ response to a phase spike [25].

3.2. Smooth frequency response of $\widehat{Theo1}$

The $\widehat{\text{Theo1}}(\tau)$ sampling function shown in figure 1 is nearly impossible to interpret in the time domain but easier to understand in the frequency domain [7]. Frequency response of a statistic is the square root of the sum of the squares of the real and imaginary Fourier transforms (FTs) of the statistic's sampling sequence. For the multiple sequences of $\widehat{\text{Theo1}}(\tau)$, responses are averaged. Frequency-response functions associated with $\widehat{\text{Theo1}}$, Total Avar and Avar are shown in figure 2 to compare the effect of their sampling functions. These plots were obtained by sweeping the input of each statistic with a sine wave and observing the output response. Results agreed with the FT method. The thin line in figure 2 is the response of a constant-Q, one-octave pass-band filter considered to be ideal for extracting typical power-law noise levels [26–30]. Total Avar implements a circular convolution of Avar's frequency response, resulting in reducing the depth of Avar's periodic nulls. But $\widehat{\text{Theo1}}$'s frequency response is by far the most impressive approximation to the response of an ideal pass-band filter.

3.3. Improved confidence of $\widehat{Theo1}$

Table 2 compares edf between Avar and $\widehat{\text{Theo1}}$ variance from 100 simulation trials in which $N_x = 1025$ using the three FM noises in the mid- to long-term range $16 < m < 1024$. In all cases, the increased edf using $\widehat{\text{Theo1}}$ is significant. Table 2 is interpreted to mean that at larger τ-values, unlike $\hat{\sigma}_y^2(\tau)$'s uncertainty, which grows quickly as $\tau \to \frac{1}{2}T$, evaluations

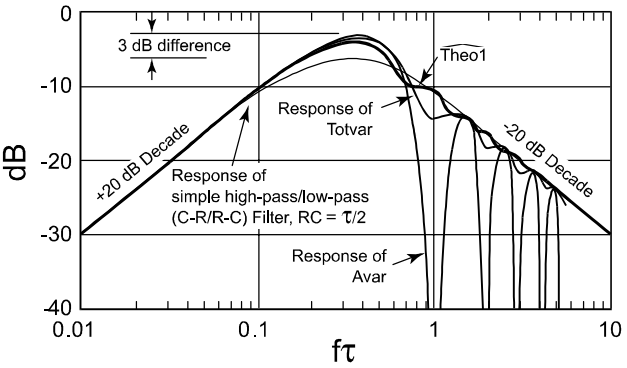

Figure 2. A comparison of frequency responses of $\widehat{\text{Theo1}}$ (——), Total Avar (shown as Totvar) and Avar and a passband variance consisting of a simple cascade of a single-pole high-pass followed by a low-pass filter with identical break points at $RC = \tau/2$ (- - - - [27]). Frequency responses are in [7] and were provided courtesy of Chuck Greenhall using his generalized autocovariance theory.

using $\widehat{\text{Theo1}}(\tau)$ have significantly lower uncertainty as long as $\tau_0 \ll \tau$ and whose uncertainty increases smoothly and nearly monotonically as $\tau \to \frac{3}{4}T$ [31]. An empirically-determined moderate starting m-sample size is $m \geqslant 10$ since $\widehat{\text{Theo1}}$'s advantage over Avar begins with $\widehat{\text{Theo1}}(7.5\tau_0)$.

4. Formulation of $\widehat{\text{Theo1}}$ that is unbiased relative to long-term Avar

In statistics, for an estimator of a parameter, its bias is defined as the difference between the expectation of the estimator and the true value of the parameter. Bias in the context of this writing means $\text{Theo1}(m, \tau_0)$ relative to the Allan variance as a ratio of $\sigma_y^2(\tau)$ to $\text{Theo1}(m, \tau_0)$ given by a normalized bias as

$$\text{nbias}(m, \tau_0) = \frac{\sigma_y^2(\tau)}{\text{Theo1}(m, \tau_0)} - 1. \tag{4}$$

It is the unique data sampling of Theo1 in figure 1 that causes bias with respect to zero-dead-time sampling of $\sigma_y^2(\tau)$. $\widehat{\text{Theo1}}$ is defined to be an unbiased estimator of $\sigma_y^2(\tau)$ in the case of the white FM (WHFM) noise type. A method for determining other noise types is to use Avar's bias, or B_1, function [16,32]. This function is undefined beyond $T/2$ since Avar is undefined, which leads us to a difficulty in using this method for $\tau > T/2$. Since B_1 does not exist for the longest τ values of $\widehat{\text{Theo1}}$, we cannot determine the noise type and hence correct for $\widehat{\text{Theo1}}$ bias. Other estimates of noise type are obtainable [33, 34], although use of these methods is limited at this time. An *ad hoc* method of determining noise type and removing bias previously reported in [35] has since been found to have its bias slightly dependent on m [36]. The most accurate dependence is given by the function

$$\text{nbias}(m) = a + \frac{b}{m^c} - 1, \tag{5}$$

where a, b and c are empirically derived constants. These constants are summarized for each of the five integer-slope noise types in table 3. Taken together, the above-mentioned methods assume (1) negligible uncertainty of the noise type at a particular value of m, (2) that only one noise type occurs and (3) that the noise type fits neatly into one of the five integer-slope noise models.

For these reasons, an estimate of (4) was investigated in regions where essentially all τ-values of $\widehat{\text{Theo1}}$ and Avar overlap. For a given τ_0, uncertainty reduces as the data run increases and the mean-ratio is proportionately weighted toward the longest overlapping τ-values because the number of point estimates in (4) increases with τ. Both these effects are generally desirable.

Starting at the longest Avar τ-values, an unbiased version of $\widehat{\text{Theo1}}$ variance, called $\widehat{\text{TheoBR}}$ variance (for 'Theo bias-removed'), is estimated by

$$\widehat{\text{TheoBR}}(m, \tau_0, N_x) = \left[\frac{1}{n+1}\sum_{i=0}^{n}\frac{\text{Avar}(m = 9 + 3i, \tau_0, N_x)}{\widehat{\text{Theo1}}(m = 12 + 4i, \tau_0, N_x)}\right] \times \widehat{\text{Theo1}}(m, \tau_0, N_x), \tag{6}$$

where $n = \lfloor (0.1N_x/3) - 3 \rfloor$ (where $\lfloor\cdot\rfloor$ denotes the floor function). $\widehat{\text{Theo1}}$ is defined as (3) and Avar has its usual definition [2, 3]. The range in the formula above is based on extensive empirical tests that looked at minimizing bias and uncertainty under the widest range of conditions. Avar is limited to τ no longer than 20% of the data-run length as recommended by [3] due to nulls in Avar's frequency response shown in figure 2 that potentially cause excessive variability and downshoots in the long term [9, 11]. For $n \geqslant 0$, $N_x \geqslant 90$ ensures at least one term in the summand for computing a bias of sufficient confidence. The method of determining $\widehat{\text{TheoBR}}$-variance in (6) does not require *a priori* assumption of long-term noise type and works with non-integer slope types.

Figure 3 plots Adev, $\widehat{\text{Theo1}}$-dev and $\widehat{\text{TheoBR}}$-dev for a data run of a mix of WHFM and RWFM noise types whose length $N_x = 4097$, thus, a data-run time of $T = 4096\tau_0$. Note the effect of bias removal from $\widehat{\text{Theo1}}$-variance using (6) results in $\widehat{\text{TheoBR}}$-dev properly resuming where Adev leaves off. $\widehat{\text{TheoBR}}$-dev, like $\widehat{\text{Theo1}}$-dev, can extend to 3/4 data-run length and is not subject to downshoots [31].

5. Definition of ThêoH

Define a hybrid frequency stability estimator called ThêoH as a composite of $\text{Avar}(m, \tau_0, N_x)$ and $\widehat{\text{TheoBR}}(m, \tau_0, N_x)$, namely,

$$\text{ThêoH}(m, \tau_0, N_x) = \begin{cases} \text{Avar}(m, \tau_0, N_x), & \text{for } 1 \leqslant m < \frac{k}{\tau_0}, \\ \widehat{\text{TheoBR}}(m, \tau_0, N_x), & \text{for } \frac{k}{0.75\tau_0} \leqslant m \leqslant N_x - 1, \\ m \text{ even}, & \end{cases} \tag{7}$$

where k is the largest $\tau \leqslant 10\%T$ where $\text{Avar}(m, \tau_0, N_x)$ has sufficient confidence. ThêoH is $\text{Avar}(m, k, N_x)$ up to 10% of the length of the data run and $\widehat{\text{TheoBR}}(m, k, N_x)$ thereafter. As Avar's statistical properties become poorer in the long term, we switch to $\widehat{\text{TheoBR}}$ variance, which has twice the range of m-values as that of Avar and different dependence on m of

Table 2. Comparison of edfs of Avar and $\widehat{\text{Theo1}}$ with 100 simulation trials of $\{x_n\}$ series consisting of $N_x = 1025$ points each. Note that Avar extends to a limit of 512 points, while $\widehat{\text{Theo1}}$ extends to 1024 points given by (3).

	WHFM		FLFM		RWFM	
m	edf Avar(m)	edf $\widehat{\text{Theo1}}(m)$	edf Avar(m)	edf $\widehat{\text{Theo1}}(m)$	edf Avar(m)	edf $\widehat{\text{Theo1}}(m)$
16	81.00	205.2	70.94	149.1	48.57	94.91
32	45.50	101.5	32.45	72.23	27.44	46.74
64	26.13	58.72	14.67	32.51	12.70	26.28
128	10.54	30.87	6.77	14.37	6.34	11.86
256	4.54	14.28	3.63	7.34	2.15	5.90
512	0.94	6.02	1.11	4.33	1.02	2.08
1024		1.57		1.34		1.11

Table 3. Constant values for the bias functions of $\widehat{\text{Theo1}}$ defined in equation (5).

Noise	a	b	c
WHPM	0.09	0.74	0.40
FLPM	0.14	0.82	0.30
WHFM	1	0	0
FLFM	1.87	−1.05	0.79
RWFM	2.70	−1.53	0.85

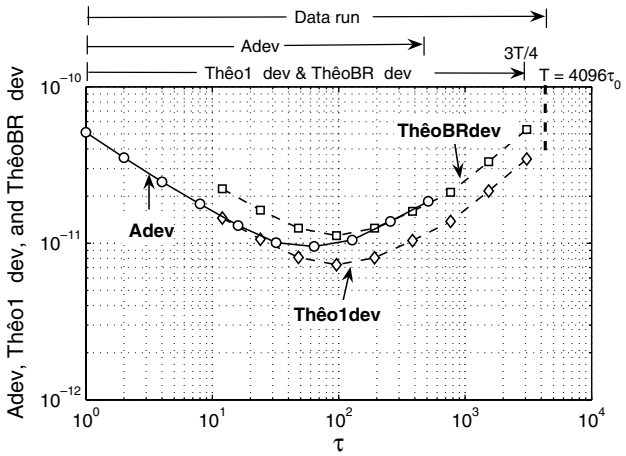

Figure 3. Adev, $\widehat{\text{Theo1}}$-dev and $\widehat{\text{TheoBR}}$-dev for a data-run time of $T = 4096\tau_0$. Data are a mix of WHFM and RWFM noise types. Long-term bias removed (BR) from $\widehat{\text{Theo1}}$-variance is $\widehat{\text{TheoBR}}$-variance using (6). $\widehat{\text{TheoBR}}$-dev, like $\widehat{\text{Theo1}}$-dev, can extend to 3/4 data-run time.

$\tau = 0.75m\tau_0$ and $\tau = m\tau_0$, respectively. Figure 4 shows an example of usage of the hybrid statistic by applying the recipe in (7) to obtain ThêoH deviation for the same data used in figure 3.

It is important to note that the points computed using $\widehat{\text{TheoBR}}(m, k, N_x)$ are based on a new class of long-stride statistics and are not a mere extrapolation of Adev. Several potential definitions for ThêoH were investigated before deciding on (7). Tests used a thorough range of simulated and actual clock data [35], including steep, mixed integer and non-integer noise types that change beyond the longest τ of Avar, into its unobservable long-τ region [37]. Slight variations in where the switch occurs from Avar to $\widehat{\text{TheoBR}}$ variance have negligible effect since overlapping τ-values are closely matched anyway.

Online MatLab code is available [38] in addition to the

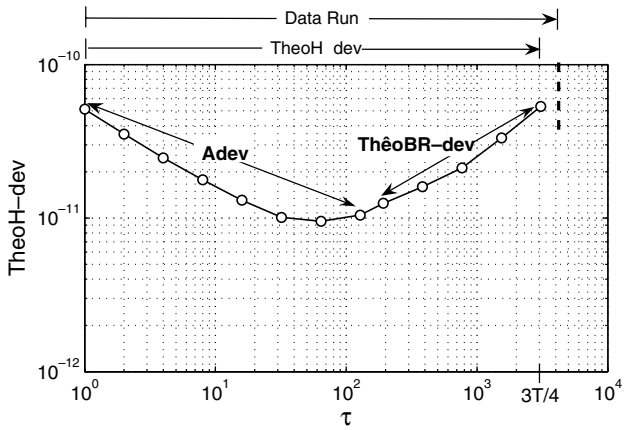

Figure 4. ThêoH-dev combines Adev and $\widehat{\text{TheoBR}}$-dev to provide one 'Allan-compatible' plot from shortest to longest τ-values.

aforementioned commercially-available, tested software (see footnote 1) that executes ThêoH and other data-analysis tools[2]. Figure 5 shows recent frequency stability results of a free-running 563 nm laser that is ultimately locked to a mechanically stable optical cavity at approximately 2.80×10^{14} Hz with the methodology and measurement described in [39]. This is an interesting illustration of exceptional short-term frequency stability in the presence of steep τ^{+1} slope that might be interpreted as frequency drift. However, systematic drift is not indicated in the raw frequency data shown at the top of figure 5. Indeed, ThêoH deviation indicates a long-term slope change towards apparent RWFM whose level can be estimated, even for this limited data run.

Characterization of models of the five noise types in the frequency domain, $S_y(f)$, at extremely low f, is possible using ThêoH with traditional mapping coefficients [2, 3, 24] rather than with table 1.

6. Confidence intervals for ThêoH

An uncertainty (estimation error) can be assigned to ThêoH by use of edf. ThêoH values have edfs that are substantially greater than those using the max-overlap estimates of the Allan variance (Avar), particularly at long-term τ-values. For computing confidence intervals, empirical formulae of edfs

[2] MatLab is a registered trademark of MathWorks, Inc. No endorsement is implied. Shared code is subject to change and is a cooperative service for users. Any such sharing is expressly exempt from any liability.

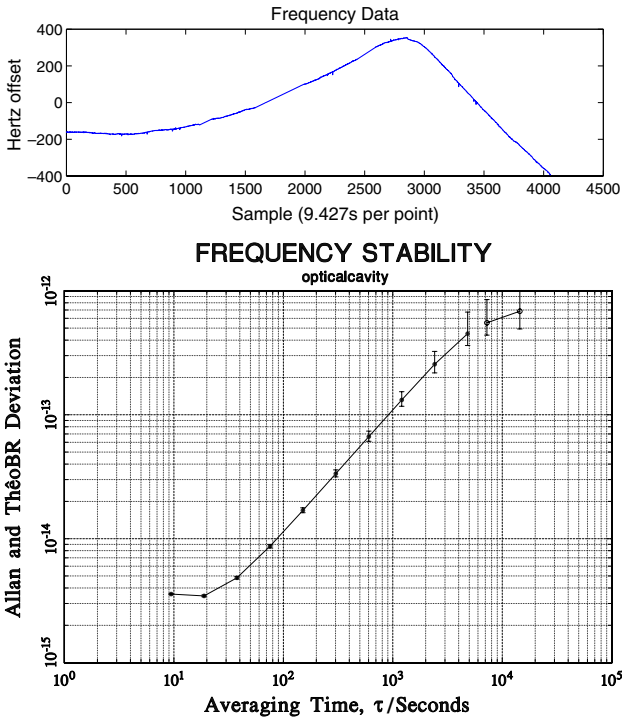

Figure 5. Allan-compatible plot from shortest to longest-possible τ-values of a laser frequency that is locked to a stable optical cavity. Data are courtesy of J Bergquist and S Diddams, June 2006.

that fit simulation can be used [24, 40]. Here are the edf formulae corresponding to $\widehat{\text{Theo1}}$, hence, ThêoH, for the five noises with the condition that $\tau_0 \leqslant T/10$, as discussed in section 3.3 and stipulated in (3):

$$\underbrace{edf}_{\text{WHFM}} = \left[\frac{5.5N_x + 1.07}{m} - \frac{3.1N_x + 6.5}{N_x}\right]\left(\frac{m^{3/2}}{m^{3/2} + 8}\right)$$

$$\underbrace{edf}_{\text{FLFM}} = \left(\frac{2.7N_x^2 - 1.3N_x m - 3.5m}{N_x m}\right)\left(\frac{m^3}{m^3 + 5.45}\right)$$

$$\underbrace{edf}_{\text{RWFM}} = \left(\frac{4.4N_x - 2}{2.175m}\right) \times \left(\frac{(4.4N_x - 1)^2 - 6.45m(4.4N_x - 1) + 6.413m^2}{(4.4N_x - 3)^2}\right)$$

$$\underbrace{edf}_{\text{WHPM}} = \left(\frac{0.86(N_x + 1)(N_x - m)}{N_x - 0.75m}\right)\left(\frac{m}{m + 1.52}\right)$$

$$\underbrace{edf}_{\text{FLPM}} = \left(\frac{5.54N_x^2 - 5.52N_x m + 10.727m}{(m + 48.8)^{1/2}(N_x - 0.75m)}\right)\left(\frac{m}{m + 0.4}\right)$$

Accuracy of the fit to simulation results is $\pm 10\%$.

Chi-square distribution functions are used for the Allan variance for calculating confidence intervals, but the distribution functions are narrower using ThêoH($0.75m\tau_0$) (which is another of its benefits) at long averaging periods in which $\tau_0 \ll T$ [31]. Figure 6 shows exact upper-bound values of 90% confidence intervals versus edf as a percentage of Adev and ThêoH values for a short RWFM noise simulation, and lower-bound values skew below computed values and have little practical use. An empirical formula to obtain percentage

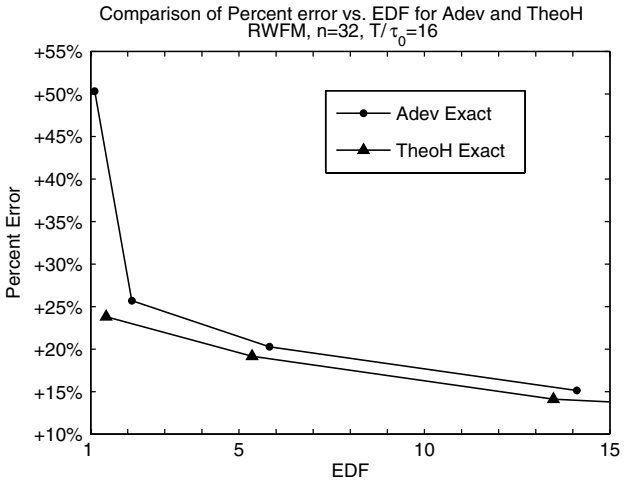

Figure 6. Upper bound of 90% confidence interval for Adev uncertainty (exact values from inverse chi-square) and ThêoH uncertainty (exact values from [31]) in terms of a percentage error of the computed value. Note the substantially lower error in ThêoH at critical low edf, corresponding to long-term τ.

upper-bound ThêoH($\tau = 0.75m\tau_0$) uncertainty is

$$\text{percentage error using ThêoH}(\tau) = \frac{100}{\sqrt{2(edf + 6.6)}}, \quad (8)$$

where $edf \geqslant 1$ is calculated from the formulae above. This uncertainty is meant to provide some information on the spread of possible values of ThêoH($0.75m\tau_0$) that are consistent with 90% confidence. It does not address values obtained for a specific case.

7. Improvements to estimating slopes

One of the main goals of τ-domain analysis is determining noise type, level and other instabilities and efficiently finding departures from stationary noise processes. The first occurence of flicker FM behaviour and estimating its lowest level, or 'floor,' is significant since the sampling interval, or stride, of 'best stability' is often regarded as the threshold of 'best reproducibility' which sets limits on accuracy [13–15]. Statistically, the FM noise is classed as stationary typified by WHFM, or not stationary signalling the onset of FLFM or RWFM processes. In the nonstationary case, a theoretical (or infinite) *mean frequency* cannot be estimated reliably [41, 42].

If an oscillator exhibits flicker FM, accuracy or reproducibility claims are no longer valid since the mean frequency changes proportionately as the sampling interval increases. To see whether the flicker-like noise exists or not, we must wait long enough for the confidence intervals to narrow the range of possible slopes so that a determination of noise type can be made. Accordingly, measurements of frequency stability of 10^{-16}, given such frequency standards, can require extremely long averaging times, perhaps years, to verify that a 'mean frequency' and accuracy apply.

Stability computations that occur in power-of-2 increments (also called 'octave' increments) of m are used for determining power-law noise types since these points are sufficiently independent, and an 'eyeball' estimate of slope-per-octave

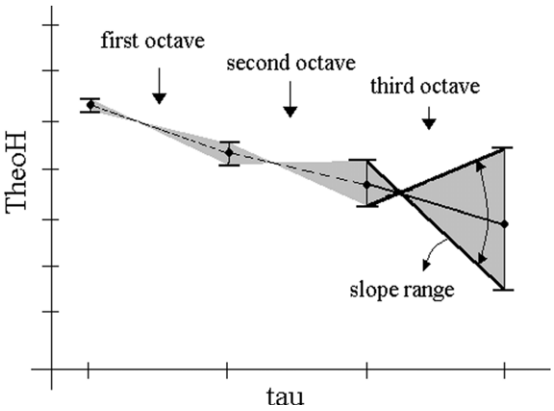

Figure 7. For the confidence interval associated with the value of ThêoH at each octave of τ, that is, $0.75m\tau_0 = \tau$, $m = 1, 2, 4, 8...$, a range of slopes corresponds to the upper and lower confidence factors. This 'slope range' is illustrated as the darkened region between τ values.

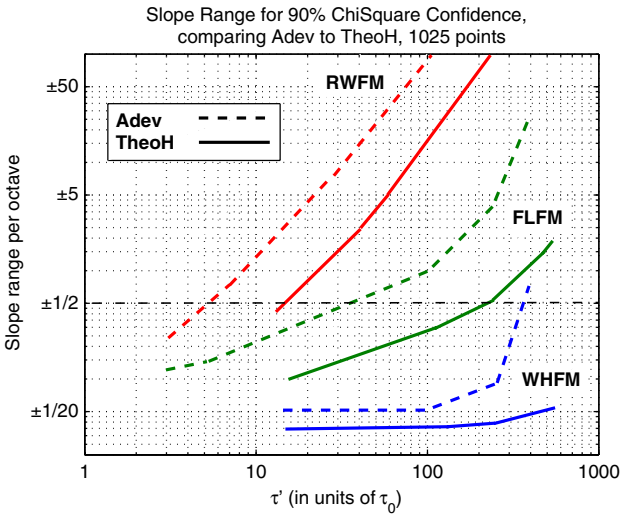

Figure 8. Spread of possible slopes-per-octave for simulated noise consisting of 1025 data points of frequency noises (WHFM, FLFM and RWFM) respectively proportional to $\tau^{-1/2}$, τ^{0}, $\tau^{+1/2}$. τ' is the midpoint between two computations of Adev and ThêoH at τ_{upper} and τ_{lower} separated by an octave. In the plots above, a spread greater than $\pm\frac{1}{2}$ means that prevailing power-law noise type is indistinguishable from its neighbouring type.

from a log–log representation is an adequate starting point. A weighted, linear least-squares regression applied to Adev and ThêoH versus tau provides a probabilistic interpretation in the 'spread' or range of slopes on a per-octave basis [43]. The regression is 'weighted' in the sense that the confidence intervals are different for different τ-values. Figure 7 depicts what is meant by the slope range that corresponds to the expected group or band of slopes.

At long term, ThêoH deviation has enough narrowness in its range of slopes as compared with Adev so that it enables faster detection of the onset of nonstationary noise types of FLFM or RWFM. To generalize this assertion, a range or band of slopes can be established based on the 90% confidence intervals of an Adev versus tau plot compared with ThêoH deviation versus tau for each octave of sampling interval τ as just previously discussed and as illustrated in figure 7. Figure 8 shows results of slope ranges for WHFM, FLFM and RWFM of 1025-point simulated data runs and using Adev and ThêoH. The plots show that for increasing averaging time, ThêoH deviation has significantly narrower slope spread than Adev. As an example, for FLFM, if the slope range is above $\pm\frac{1}{2}$, then FLFM is not readily distinguishable from RWFM or WHFM. This occurs for Adev at $\tau' \cong 40\tau_0$, or about 4% of the data run, whereas for ThêoH deviation, prevailing FLFM continues to be detected as such to $\tau' \cong 200\tau_0$ or about 1/5 of the data run. For WHFM, the range is below $\pm\frac{1}{2}$, except at the longest averaging times, noting only that white noise is detected better using ThêoH deviation. All bands use 90% confidence intervals based on chi-square distribution properties (see section 6).

The results above show promise when any number of random noise components may be present in frequency-stability data, depending on the type of test and reference oscillators being compared and the environment in which the data are obtained. If a ThêoH deviation plot has nulls, 'structure' or oscillations over a particular range of τ, then it is probably not a power-law process and one must try to estimate the spectral density $L(f)$ or $S_y(f)$ using a phase- or frequency-noise measurement method and a DFT or FFT analysis in addition to an analysis using ThêoH [44, 45].

8. Conclusion

$\widehat{\text{Theo1}}$ variance, the genesis of ThêoH variance, is the estimator of a new class of theoretical variance which averages every permissible squared second-difference of time errors in a given period $m\tau_0$. ThêoH deviation consists of Adev for short and mid τ-values in combination with points at very long τ-values of $\widehat{\text{Theo1}}$ with BR in such a way that mixed noise types can be accurately characterized. 'H' of ThêoH is used to indicate a high confidence and/or hybrid statistic (Adev in short term and TheoBR (bias removed) deviation in long term) with evaluations that extend to 50% longer than possible using Adev alone. ThêoH deviation provides the following:

- unprecedented confidence in estimating frequency stability,
- automatic correction for bias relative to Adev, even for non-integer power-law noises,
- improved power-law noise determination at very long τ-values using ThêoH-versus-τ slopes,
- characterization to longer τ-values than with Adev, and
- characterization of $S_y(f)$ to extremely low f using existing Allan mapping coefficients.

Long-term frequency stability can be obtained in essentially one-third less time. For example, a two-month stability can be obtained with three months of data, rather than four months of data that is usually required for such a point.

Acknowledgments

The author thanks Jennifer McGee, Tiffany Tasset, Trudi Peppler, Tom Parker and Marc Weiss of NIST for assistance and suggestions in completing this work. He especially thanks Chuck Greenhall of JPL, Don Percival of University of Washington APL, Francois Vernotte of Besançon Observatory and

Bill Riley of Hamilton Technical Services for discussions, interpretations and tests of Total Avar that motivated development of $\widehat{\text{Theo1}}$, $\widehat{\text{TheoBR}}$ and ThêoH. Lastly, he thanks David Allan of Allan's Time for many years of engaging discussions.

The article is a contribution of the National Institute of Standards and Technology and is not subject to US copyright.

Appendix A

A.1. Example calculation of $\widehat{Theo1}(\tau)$

To illustrate (3), consider a short segment of data consisting of a series of five time-error measurements $\{x_1, x_2, x_3, x_4, x_5\}$ taken once a day; thus $\tau_0 = 1\,\text{d} = 86\,400\,\text{s}$, $3\tau_0 = 3\,\text{d} = 259\,200\,\text{s}$ in this 5 d worth of data, thus our evaluation is at $m = 4$ with $N_x = 5$. The outer (leading or first) summation in (3) consists of only one term, noting that i goes from 1 to $N_x - 4 = 1$, so we have only the inner or second summation to compute, and it has two terms corresponding to indices $\delta = 0, 1$. They are

$$[(x_1 - x_3) + (x_5 - x_3)]^2 \text{ and } [(x_1 - x_2) + (x_5 - x_4)]^2.$$

We obtain

$$\begin{aligned}\widehat{\text{Theo1}}&(3, 86\,400\,\text{s}, 5)\\ &= \left(\frac{1}{0.75}\right) \underbrace{\frac{1}{16\tau_0^2}}_{16(86\,400\,\text{s})^2} \left\{\frac{1}{2} \times [(x_1 - x_3) + (x_5 - x_3)]^2\right.\\ &\quad \left. + 1 \times [(x_1 - x_2) + (x_5 - x_4)]^2\right\}. \qquad (9)\end{aligned}$$

Figure 9 shows a representative set of time-error values sampled from function $x(t)$. These are example values used in Annex C3 of [3], but we have taken the liberty of changing the sampling to 1 sample/day rather than 1 sample/s in order to discuss how $\widehat{\text{Theo1}}$-variance operates on long-term data in a way that is different from how the Allan variance does.

In units of nanoseconds, we have

$$\begin{aligned} x_1 &= 1.08, \qquad x_2 = 0.5, \qquad x_3 = 2.2,\\ x_4 &= 4.68, \qquad x_5 = 3.29.\end{aligned}$$

The required first differences and operations on series $\{x_n\}$ in (9) are

$$\left.\begin{aligned} x_1 - x_3 &= -1.12\\ &+\\ x_5 - x_3 &= 1.09\end{aligned}\right\} = -0.03;\ \frac{1}{2}\times(-0.03)^2 = 0.000\,45,$$

and

$$\left.\begin{aligned} x_1 - x_2 &= 0.58\\ &+\\ x_5 - x_4 &= -1.39\end{aligned}\right\} = -0.81;\ 1 \times (-0.81)^2 = 0.656.$$

Using these values and taking the square root of $\widehat{\text{Theo1}}$, we get

$$\begin{aligned}\sqrt{\widehat{\text{Theo1}}}&(\tau = (0.75)259\,200\,\text{s})\\ &= \frac{1\,\text{ns}}{4\sqrt{0.75}(86\,400\,\text{s})}\{0.000\,45 + 0.656\}^{\frac{1}{2}}\\ &= \frac{0.81\,\text{ns}}{4\sqrt{0.75}(86\,400\,\text{s})} = 3.15 \times 10^{-15},\end{aligned}$$

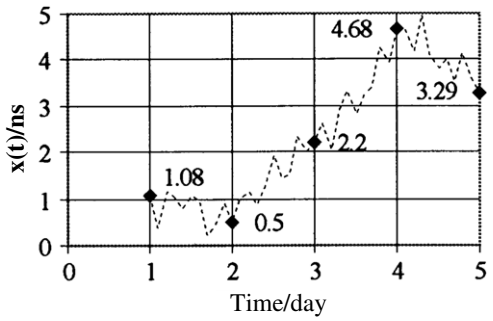

Figure 9. Time-error measurements taken once each day for a five-day data run for four fractional-frequency error-values. Three-day frequency stability, while impossible to compute with the Allan deviation, is 3.12×10^{-15} using $\sqrt{\widehat{\text{Theo1}}}$ ($\tau = 2.25$ d). Values above are test data from Annex C3 of [3] with data spacing τ_0 changed to 1 d.

which is our estimated frequency stability for a 2.25-day or 194 400 s interval based on this 5-day data set. The Allan deviation is undefined for this long τ-value, but for comparison, Adev at $\tau = 2\,\text{d}$ has a value of 1.23×10^{-16}, which is almost 20 times lower and unrealistically low based on the data set.

This small number of sampled measurements makes a few issues apparent. First, the total number of terms in the inner summation in (3) for obtaining $\widehat{\text{Theo1}}$-variance depends on the range or extent of index δ, which is inversely proportional to the ratio of data spacing to stride, that is, τ_0-to-τ. Shorter τ_0 means larger ratio m and more terms to average in the inner summation for an interval τ. Additionally, for a fixed τ, the number of overlapping terms for the outer summation to average in (3) is inversely proportional to τ_0-to-T. Shorter τ_0 provides a greater ratio T/τ_0. Second, the longest computable interval is $\tau = 0.75(T - \tau_0)$. If we want to obtain an evaluation at the longest possible τ-value, or $\widehat{\text{Theo1}}(\tau)$ just shy of T itself, data spacing τ_0 should be as short as possible. Third, recall that for an evaluation at $\widehat{\text{Theo1}}(\tau)$, we must actually calculate (3) at one-higher increment of data-spacing $m + 1$ corresponding to $\tau + \tau_0$. This preciseness is needed when m is small, but in the case of long-term frequency-stability measurements, the dimensionless quantity $m = \tau/0.75\tau_0$ can often be made arbitrarily large by simply sampling the data at a faster rate. Based on measurements of edf and exact computations of confidence intervals [31], $m \geqslant 10$ is recommended. In conclusion, for a given τ-value, making data spacing τ_0 short allows $\widehat{\text{Theo1}}$ to extract or 'squeeze out' more information about noise type and level. Our example data set in figure 9 is therefore not ideal, but illustrates how $\widehat{\text{Theo1}}$-variance operates on $\{x_n\}$ and is computed. In the example data, it would be better to sample time-error function $x(t)$ at 10 samples/day or faster rather than 1 sample/day over a 5-day run for all of the reasons given if we wish to estimate frequency stability in the interval beyond 2 days. In contrast, Avar shows no substantial improvement when measuring the time-error function between two oscillators at a sample period (data spacing) of τ_0 much shorter than its longest limit of $\tau = T/2$, unless of course you specifically want evaluations of shorter-term stability.

Test suite

For purposes of illustration, we calculate one value of $\widehat{\text{Theo1}}$-dev from a test suite consisting of a short sequence of numbers. The sequence can be used as a basic test of computer programs. Consider the following sequence of twelve days of equispaced time error measurements $\{x_n : n = 1, \ldots, 12\}$ in units of nanoseconds:

$$\begin{aligned} x_1 &= -2.15 & x_7 &= -3.3 \\ x_2 &= -0.99 & x_8 &= 1.08 \\ x_3 &= 1 & x_9 &= 0.5 \\ x_4 &= 2.5 & x_{10} &= 2.2 \\ x_5 &= 0.65 & x_{11} &= 4.68 \\ x_6 &= -3.71 & x_{12} &= 3.29 \end{aligned}$$

Let us calculate $\widehat{\text{Theo1}}$-dev for $\tau = 8.5\,\text{d}$ $(7.344 \times 10^5\,\text{s})$. From (3), the inner summand terms for the sample variance $\widehat{\text{Theo1}}(10, 86\,400\,\text{s}, 12)$ for each index i are given by (for this value of τ, note that $m = 10$):

$$\begin{aligned} i = 1: \\ \delta = 0: &\quad \tfrac{1}{5}[(-2.15 - (-3.71)) + (4.68 - (-3.71))]^2 \\ \delta = 1: &\quad +\tfrac{1}{4}[(-2.15 - 0.65) + (4.68 - (-3.3))]^2 \\ \delta = 2: &\quad +\tfrac{1}{3}[(-2.15 - 2.5) + (4.68 - 1.08)]^2 \\ \delta = 3: &\quad +\tfrac{1}{2}[(-2.15 - 1) + (4.68 - 0.5)]^2 \\ \delta = 4: &\quad +1[(-2.15 - (-0.99)) + (4.68 - 2.2)]^2 \\ &\quad = 29.145, \\ i = 2: \\ \delta = 0: &\quad \tfrac{1}{5}[(-0.99 - (-3.3)) + (3.29 - (-3.3))]^2 \\ \delta = 1: &\quad +\tfrac{1}{4}[(-0.99 - (-3.71)) + (3.29 - 1.08)]^2 \\ \delta = 2: &\quad +\tfrac{1}{3}[(-0.99 - 0.65) + (3.29 - 0.5)]^2 \\ \delta = 3: &\quad +\tfrac{1}{2}[(-0.99 - 2.5) + (3.29 - 2.2)]^2 \\ \delta = 4: &\quad +1[(-0.99 - 1) + (3.29 - 4.68)]^2 \\ &\quad = 36.66, \end{aligned}$$

and, for the outer summation, we obtain $\sum_{i=1}^{2}(\cdot) = 65.81$.

$$\widehat{\text{Theo1}}(10, 86\,400\,\text{s}, 12) = \frac{1}{2(0.75)(10)^2}\sum_{i=1}^{2}(\cdot) = 0.4387,$$

and

$$\widehat{\text{Theo1}}\text{-dev}(10, 86\,400\,\text{s}, 12) = \sqrt{0.4387} = 0.6623.$$

Since the original measurements were in units of ns/day, we obtain $0.6623 \times 1\,\text{ns}/86\,400\,\text{s}$ for a sampling period of $\tau_0 = 1\,\text{d} = 86\,400\,\text{s}$. Therefore, for this set, $\widehat{\text{Theo1}}\text{-dev}(10, 86\,400\,\text{s}, 12) = 7.66 \times 10^{-15}$.

References

[1] Allan D W 1966 Statistics of atomic frequency standards *Proc. IEEE* **54** 221–30

[2] Sullivan D B, Allan D W, Howe D A and Walls F L (ed) 1990 Characterization of clocks and oscillators *Natl. Inst. Stand. Technol. Technical Note* 1337

[3] IEEE Std 1139-1999: Standard definitions of physical quantities for fundamental frequency and time metrology—random instabilities *IEEE-SA Standards Board* (J R Vig, chairperson) 1999

[4] Howe D A 1995 An extension of the Allan variance with increased confidence at long term *Proc. 1995 IEEE Int. Frequency Control Symp. (San Francisco, CA, 31 May–2 June)* pp 321–9

[5] Howe D A and Lainson K J 1996 Effect of drift on TOTALDEV *Proc. 1996 IEEE Int. Frequency Control Symp.* pp 883–9

[6] Howe D A 1997 Methods of improving the estimation of long-term frequency variance *Proc. European Frequency and Time Forum (Neuchâtel, Switzerland)* pp 91–9

[7] Howe D A and Greenhall C A 1997 Total variance: a progress report on a new frequency stability characterization *Proc. 29th Ann. Precise Time and Time Interval Meeting (Long Beach, CA)* pp 39–48

[8] Howe D A 1999 Total variance explained *Proc. 13th European Frequency and Time Forum* and *1999 IEEE Int. Frequency Control Symp. (Besançon, France)* pp 1093–9

[9] Howe D A 2000 Total deviation approach to long-term characterization of frequency stability *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **UFFC-47** 1102–10

[10] Howe D A and Peppler T 2001 Definitions of 'total' estimators of common time-domain variances *Proc. 2001 IEEE Int. Frequency Control Symp. (Seattle, WA)* pp 127–32

[11] Greenhall C A, Howe D A and Percival D B 1999 Total variance, an estimator of long-term frequency stability *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **UFFC-46** 1183–91

[12] Jefferts S R *et al* 2002 Accuracy evaluation of NIST-F1 *Metrologia* **39** 321–36

[13] Parker T E, Jefferts S R, Heavner T P and Donley E A 2005 Operation of the NIST-F1 caesium fountain primary frequency standard with a maser ensemble, including the impact of frequency transfer noise *Metrologia* **42** 423–30

[14] Heavner T P, Jefferts S R, Donley E A, Shirley J H and Parker T E 2005 NIST F1: recent improvements and accuracy evaluations *Metrologia* **42** 411–22

[15] Godone A, Micalizio S, Claudio C E and Levi F 2006 The pulsed rubidium clock *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **UFFC-53** 525–9

[16] Barnes J A 1969 Tables of bias functions, B_1 and B_2, for variances based on finite samples of processes with power law spectral densities *Nat. Bur. Stands. (US) Technical Note* TN-375

[17] Lesage P and Audoin C 1979 Effect of dead-time on the estimation of the two-sample variance *IEEE Trans. Instrum. Meas.* **IM-28** 6–10

[18] Parker T E 2001 Comparing and evaluating the performance of primary frequency standards: impact of dead time *Proc. 2001 IEEE Int. Frequency Control Symp. (Seattle, WA)* pp 57–62

[19] Howe D A and Hagn E E 1999 Limited live-time measurements of frequency stability *Proc. 13th European Frequency and Time Forum* and *1999 IEEE Int. Frequency Control Symp. (Besançon, France)* pp 1113–6

[20] Ekstrom C R and Koppang P A 2006 Error bars for three-cornered hats *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **UFFC-53** 876–9

[21] Birkhoff G D 1931 Proof of the ergodic theorem *Proc. Nat. Acad. Sci.* **17** 656–60

[22] Barnes J A *et al* 1971 Characterization of frequency stability *IEEE Trans. Instrum. Meas.* **IM-20** 105–20

[23] Howe D A 1983 Pitfalls in digitizing the data *Frequency Stability: Fundamentals and Measurements* ed V F Kroupa (New York: IEEE)

[24] Howe D A, Allan D W and Barnes J A 1981 Properties of signal sources and measurement methods *Proc. 35th Freq. Control Symp. (Philadelphia, PA)* pp 1–47 reprinted in [2]

[25] Riley W J and Greenhall C A 2004 Notes on calibration of Total Avar to Avar, private communications

[26] Rutman J 1974 Characterization of frequency stability: A transfer function approach and its application to

measurements via filtering of phase noise *IEEE Trans. Instrum. Meas.* **IM-23** 40–8
[27] Wiley R G 1977 A direct time-domain measure of frequency stability: the modified Allan variance *IEEE Trans. Instrum. Meas.* **IM-26** 38–41
[28] Allan D W, Weiss M A and Jespersen J L 1991 A frequency-domain view of time-domain characterizations of clocks and time and frequency distribution systems *Proc. 45th Freq. Control Symp. (Los Angeles, CA)* pp 667–78
[29] Percival D B 1991 Characterization of frequency stability: frequency-domain estimation of stability measures *Proc. IEEE* **79** 961–72
[30] Howe D A and Percival D B 1995 Wavelet variance, Allan variance, and leakage *IEEE Trans. Instrum. Meas.* **IM-44** 94–7
[31] Tasset T N, Howe D A and Percival D B 2004 $\widehat{\text{Theo1}}$ confidence intervals *Proc. 2004 IEEE Int. Freq. Control Symp. (Montréal, Canada)* pp 725–8
[32] Howe D A, Beard R L, Greenhall C A, Vernotte F, Riley W J and Peppler T K 2005 Enhancements to GPS operations and clock evaluations using a total hadamard deviation *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **UFFC-52** 1253–61
[33] Riley W J and Greenhall C A 2004 Power law noise identification using the lag 1 autocorrelation *Proc. 18th European Frequency and Time Forum (Guildford, UK)*
[34] Greenhall C A and Riley W J 2003 Uncertainty of stability variances based on finite differences *Proc. 35th Ann. Precise Time and Time Interval Meeting (San Diego, CA)* pp 267–79
[35] Howe D A and Peppler T K 2003 Very long-term frequency stability: estimation using a special-purpose statistic *Proc. 2003 Joint Mtg. IEEE Int. Freq. Control Symp. and EFTF Conf. (Tampa, FL)* pp 233–8
[36] Howe D A and Tasset T N 2004 $\widehat{\text{Theo1}}$: characterization of very long-term frequency stability *Proc. 18th European Frequency and Time Forum (Guildford, UK)*
[37] Greenhall C A 1997 A frequency-drift estimator and its removal from modified allan variance *Proc. 1997 IEEE Int. Frequency Control Symp. (Orlando, FL)* pp 428–32
[38] Senior K, Beard R and White J 2005 CANVAS: clock analysis, visualization, and archiving system *Proc. 37th Ann. Precise Time and Time Interval Meeting*
[39] Young B C, Cruz F C, Itano W M and Bergquist J C 1999 Visible lasers with subhertz linewidths *Phys. Rev. Lett.* **82** 3799–802
[40] Greenhall C A 1991 Recipes for degrees of freedom of frequency stability estimators *IEEE Trans. Instrum. Meas.* **IM-40** 994–9
[41] Vernotte F 1997 Estimation of the uncertainty of a mean frequency measurement *Proc. 11th European Frequency and Time Forum* pp 553–6
[42] Vernotte F, Delporte J, Brunet M and Tournier T 2001 Uncertainties of drift coefficients and extrapolation errors: Application to clock error prediction *Metrologia* **38** 325–42
[43] Percival D and Walden A 2000 *Wavelet Methods for Time Series Analysis* (Cambridge: Cambridge University Press) section 9.5–9.7
[44] Percival D B 2006 Spectral analysis of clock noise: a primer *Metrologia* **43** S299–310
[45] McGee J A and Howe D A 2006 TheoH and Allan deviation as power-law noise estimators *IEEE Trans. Ultrason., Ferroelectr. Freq. Control* at press