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 $\psi_2^2(\tau_s, \tau)$, a variance that separates the Allan variance’s averaging time ($\tau$) from its sampling interval ($\tau_s$), also called ‘stride’, as applied to normalized fractional-frequency $y(t)$. The $\tau$-length average of $y(t)$ is denoted $\tau y(t)$, and every permissible squared difference of $\tau y(t)$-values in $\psi_2^2(\tau_s, \tau)$ is averaged in $m \tau_0$ segments of $y(t)$ to compute Theo1($m, \tau_0$). Although biased relative to the Allan variance, Theo1($m, \tau_0$) mimics all statistical properties of the Allan variance and works to very long stride. Its estimator $\hat{\text{Theo}}_1(\tau_s)$ 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 $\hat{\text{Theo}}_1$ variance, called ‘$\hat{\text{TheoBR}}$’ variance, yields ThêoH variance (‘H’ to indicate high confidence and/or hybrid Allan and TheoBR functions). ThêoH deviation combines the Allan deviation in short term and 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 $\tau$, 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 $\tau$-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 $\tau$-values, where the extension is greatest, while essentially computing plain Avar at mid- and short-term $\tau$-values, where the extension becomes progressively smaller and smaller [7–10]. Total Avar($\tau, T$) has dependence on $\tau$ and data-run length $T$. The spectral filter response associated with Total Avar’s sampling for large $\tau$-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), FLMF 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 FLMF 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 \([11]\).

Théo\(\text{H}\) 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 WFM, Théo\(\text{H}\) variance has a six-fold increase in edfs, hence, substantially improved confidence, on the reported level of frequency stability at the maximum-\(\tau\) value of Avar \([13]\). Théo\(\text{H}\) variance also gives levels of frequency stability with good confidence beyond the maximum-\(\tau\) of Avar \([14, 15]\).

Section 2 defines a generic first-difference variance called ‘psi-variance’ \(\psi^2(\tau_s, \tau)\) in which measurement dead time is a free parameter given by sampling interval \(\tau_s\) 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 ‘Théo\(\text{I}\) variance’ with recognized benefits \([12–15, 20]\). Throughout this paper, the maximum-overlap estimator of Theo\(\text{I}\) is designated ‘Théo\(\text{I}\).’ The properties of Théo\(\text{I}\) 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 Théo\(\text{H}\) variance defined in section 5. Section 6 gives Theo\(\text{I}\) edf formulae. Section 7 shows that by using Théo\(\text{H}\) 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: \(\psi^2(\tau_s, \tau)\)

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

\[
\psi^2(\tau_s, \tau) \equiv \frac{\tau}{\tau_s} \left( \langle [\tau(t) - \tau(t - \tau_s)]^2 \rangle \right)
\]

where \(\langle \cdot \rangle\) is an average over duration \(\tau = m\tau_0\) and \(\langle \cdot \rangle\) denotes an ensemble average of realizations of sequential \(\tau(t)\) \([21]\). In terms of the time-error function \(x(t)\) between two oscillators, \(\tau(t) = (x(t) - x(t - \tau))\). Definition (1) is based on taking sequential mean frequency measurements spaced \(\tau_s\) apart, differencing them, and computing the mean square \([22]\). The \(\tau/\tau_s\) makes the expected value the same for WFM irrespective of \(\tau\) and \(\tau_s\) for a constant interval. Figure 1 shows the sampling-sequence function associated with \(\psi^2(\tau_s, \tau)\) acting

\[\text{(a)}\]

\[\text{(b)}\]

\[\text{(c)}\]

\[\text{(d)}\]

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

2.2. Definition of Theo\(\text{I}\)

A new class of long-stride statistics is introduced in this writing by ‘Theo\(\text{I}\)’ that averages \(\psi^2(\tau_s, \tau)\) over combinations of \(\tau_s\) and \(\tau\). \(\tau\), \(\tau_s\), \(x(t)\) 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 \(\{x_n : n = 1, \ldots, N_n\}\) with a sampling period between adjacent observations given by \(\tau_0\), we define the \(\alpha\tau_0\)-average fractional-frequency deviate as

\[
\bar{\tau}_a(\alpha) \equiv \frac{1}{\alpha} \sum_{j=0}^{\alpha-1} y_{n-j},
\]

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 \(N_n\) and sample-rate period \(\tau_0\) are used in discrete-continuous versions of actual measurements \([23]\).

Theo\(\text{I}\) variance averages \(\psi^2(\tau_s, \tau)\) in interval \(m\tau_0\). Its expectation value, in terms of symmetric differences of
time-error values \(\{x_\tau\}\), is

\[
\overline{\text{Theo}}(m, \tau_0) = \frac{1}{2} \left( \frac{2}{0.75m} \sum_{\delta=0}^{m-1} \frac{1}{m(\tau - \delta)} \times \frac{1}{\tau_0} \left[ (x_{\tau+\delta} - x_{\tau+\delta+\tau}) - (x_{\tau-\delta} - x_{\tau-\delta-\tau}) \right]^2 \right),
\]

(2)

where \(\tau_0\) is the minimum sampling interval. \(\overline{\text{Theo}}(m, \tau_0)\) 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 \(\overline{\text{Theo}}\)

The hat ‘\(\hat{a}\)’ over Theo1 variance denotes an estimator for length \(N_x\), consistent with nomenclature ‘\(\hat{\sigma}_2^2(\tau)\)’, 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

\[
\overline{\text{Theo}}(m, \tau_0, N_x) = \frac{1}{0.75(N_x - m)(m\tau_0)} \sum_{i=1}^{N_x-m} \frac{1}{m(\tau - \delta)} \times [(x_{i+m} - x_{i-\delta+\xi}) - (x_{i-\delta} - x_{i-\delta-\xi})]^2,
\]

(3)

for \(m\) even, \(10 \leq m \leq N_x - 1\), where frequency stability is evaluated, in essence, at an average sampling interval \(\tau = 0.75m\tau_0\). Theo1 variance estimates mean \(\hat{\psi}_2^2(\tau_0, \tau)\) by summing over all \(\tau_0\) and \(\tau_0\) in span \(m\tau_0\). This means that the longest \(\tau = 0.75(N_x - 1)\tau_0\) or an averaging time corresponding to 3/4 of the duration of a data-run length. This is 50% longer than the longest \(\tau_0\)-value of maximum-overlap estimator \(\hat{\sigma}_2^2(\tau)\) in which m’s range is only \(1 \leq m \leq (N_x - 1)/2 [2, 3]\).

In (3), the symmetry of Theo1’s sampling (see figure 1) means that \(m\) can take only even values, that is, \(m = 2, 4, 6, \ldots\). Thus, \(\overline{\text{Theo}}(\tau)\)’s evaluations actually occur at \(\tau = 1\tau_0, 3\tau_0, 5\tau_0, \ldots\). An evaluation of \(\overline{\text{Theo}}(m - 1, \tau_0, N_x)\) in a time-domain setting simply means \(\overline{\text{Theo}}(\tau - \tau_0)\). As a practical matter, since \(\tau_0\) and its effect can be made arbitrarily small, \(\overline{\text{Theo}}(\tau_0) = \overline{\text{Theo}}(\tau)\) just as \(\hat{\sigma}_2^2(\tau_0) = \hat{\sigma}_2^2(\tau)\) with Avar.

3. Properties of \(\overline{\text{Theo}}(\tau)\)

3.1. Compatibility with Allan deviation

The square-root of Theo1-variance of (3), or \(\overline{\text{Theo}}\)-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 \(\tau_0\) in seconds that defines time interval \(\tau\) such that \(\tau = 0.75m\tau_0\). Thus, notation \(\overline{\text{Theo}}(\tau)\) means \(\overline{\text{Theo}}(\tau = \tau_0) = 0.75m\tau_0\).

Table 1 gives coefficients for transforms of \(\overline{\text{Theo}}\)-deviation level and slope on log-log plot to the square root of noise spectrum, \(\sqrt{S_\tau(f)}\), that is, carrier-signal rms fractional frequency fluctuations \((\Delta f/\nu_0)_{\text{rms}}\) in a 1 Hz bandwidth evaluated at Fourier frequency \(f\). ‘QPM’ is the quantized phase modulation, and ‘QSF\(\text{M}(f_0)\)’ means quasi-sinusoidal frequency modulation (QSF\(\text{M}\)) at \(f_0\). Coefficients are consistent with straight-line (on log–log plots) Allan-deviation transforms to the Fourier-frequency domain [24].

Correlation of octave \(\tau\)-values using 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 \(\overline{\text{Theo}}\)

The \(\overline{\text{Theo}}(\tau)\) 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 \(\overline{\text{Theo}}(\tau)\), responses are averaged. Frequency-response functions associated with 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 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 \(\overline{\text{Theo}}\)

Table 2 compares edf between Avar and 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 Theo1 is significant. Table 2 is interpreted to mean that at larger \(\tau\)-values, unlike \(\hat{\sigma}_2^2(\tau)\)’s uncertainty, which grows quickly as \(\tau \rightarrow 1/2 T\), evaluations
which leads us to a difficulty in using this method for
Since $B$ monotonically as $m$ moderate starting
advantage over Avar begins with $7$
long-term Avar
4. Formulation of Théo1 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 Théo1($m$, $\tau_0$) relative to the Allan variance as a ratio of $\sigma^2_\tau^\tau(\tau)$ to Théo1($m$, $\tau_0$) given by a normalized bias as

$$nbias(m, \tau_0) = \frac{\sigma^2_\tau^\tau(\tau)}{Théo1(m, \tau_0)} - 1. \tag{4}$$

It is the unique data sampling of Théo1 in figure 1 that causes bias with respect to zero-dead-time sampling of $\sigma^2_\tau^\tau(\tau)$. Théo1 is defined to be an unbiased estimator of $\sigma^2_\tau^\tau(\tau)$ 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 us to a difficulty in using this method for $\tau \geq T/2$. Since $B_1$ does not exist for the longest $\tau$ values of Théo1, we cannot determine the noise type and hence correct for Théo1 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 on $m$ is given by the function

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

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 $\tau$-values of Théo1 and Avar overlap. For a given $\tau_0$, uncertainty reduces as the data run increases and the mean-ratio is proportionately weighted toward the longest overlapping $\tau$-values because the number of point estimates in (4) increases with $\tau$. Both these effects are generally desirable.

Starting at the longest Avar $\tau$-values, an unbiased version of Théo1 variance, called TheoBR variance (for ‘Theo-bias-removed’), is estimated by

$$TheoBR(m, \tau_0, N_k) = \left[ \frac{1}{n+1} \sum_{i=0}^{n} Avar(m = 9 + 3i, \tau_0, N_k) \right] \times Théo1(m, \tau_0, N_k), \tag{6}$$

where $n = \lceil(0.1N_k/3) - 3 \rceil$ (where $\lceil \cdot \rceil$ denotes the floor function). Théo1 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 $\tau$ 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 $n > 0$, $N_k > 90$ ensures at least one term in the summand for computing a bias of sufficient confidence. The method of determining 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, Théo1-dev and TheoBR-dev for a data run of a mix of WHFM and RWFM noise types whose length $N_k = 4097$, thus, a data-run time of $T = 4096\tau_0$. Note the effect of bias removal from Théo1-variance using (6) results in TheoBR-dev properly resuming where Adev leaves off. TheoBR-dev, like Théo1-dev, can extend to 3/4 data-run length and is not subject to downshoots [31].

5. Definition of ThéoH
Define a hybrid frequency stability estimator called ThéoH as a composite of Avar ($m$, $\tau_0$, $N_k$) and TheoBR($m$, $\tau_0$, $N_k$), namely,

$$ThéoH(m, \tau_0, N_k) = \begin{cases} Avar(m, \tau_0, N_k), & \text{for } 1 \leq m < \frac{k}{\tau_0}, \\ TheoBR(m, \tau_0, N_k), & \text{for } \frac{k}{\tau_0} \leq m \leq N_k - 1, \\ \end{cases} \tag{7}$$

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

Figure 2. A comparison of frequency responses of Théo1 (——), 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 $RC = T/2$ (- - - [27]). Frequency responses are in [7] and were provided courtesy of Chuck Greenhall using his generalized autocovariance theory.

Metrologia, 43 (2006) S322–S331 S325
Avar extends to a limit of 512 points, while TheoBR in figure 3. recipe in (7) to obtain TheoH deviation for the same data used as an example of usage of the hybrid statistic by applying the edf. Several potential definitions for TheoH 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 TheoBR variance have negligible effect since overlapping τ-values are closely matched anyway.

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

<table>
<thead>
<tr>
<th>(m)</th>
<th>edf Avar(m)</th>
<th>edf Theo1(m)</th>
<th>edf Avar(m)</th>
<th>edf Theo1(m)</th>
<th>edf Avar(m)</th>
<th>edf Theo1(m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>81.00</td>
<td>205.2</td>
<td>70.94</td>
<td>149.1</td>
<td>48.57</td>
<td>94.91</td>
</tr>
<tr>
<td>32</td>
<td>45.50</td>
<td>101.5</td>
<td>32.45</td>
<td>72.23</td>
<td>27.44</td>
<td>46.74</td>
</tr>
<tr>
<td>64</td>
<td>26.13</td>
<td>58.72</td>
<td>14.67</td>
<td>32.51</td>
<td>12.70</td>
<td>26.28</td>
</tr>
<tr>
<td>128</td>
<td>10.54</td>
<td>30.87</td>
<td>6.77</td>
<td>14.37</td>
<td>6.34</td>
<td>11.86</td>
</tr>
<tr>
<td>256</td>
<td>4.54</td>
<td>14.28</td>
<td>3.63</td>
<td>7.34</td>
<td>2.15</td>
<td>5.90</td>
</tr>
<tr>
<td>512</td>
<td>0.94</td>
<td>6.02</td>
<td>1.11</td>
<td>4.33</td>
<td>1.02</td>
<td>2.08</td>
</tr>
<tr>
<td>1024</td>
<td>1.57</td>
<td></td>
<td>1.34</td>
<td></td>
<td>1.11</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Constant values for the bias functions of Theo1 defined in equation (5).

<table>
<thead>
<tr>
<th>Noise</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHFM</td>
<td>0.09</td>
<td>0.74</td>
<td>0.40</td>
</tr>
<tr>
<td>FLFM</td>
<td>0.14</td>
<td>0.82</td>
<td>0.30</td>
</tr>
<tr>
<td>WHFM</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FLFM</td>
<td>1.87</td>
<td>−1.05</td>
<td>0.79</td>
</tr>
<tr>
<td>RWFM</td>
<td>2.70</td>
<td>−1.53</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Figure 3. Adev, Theo1-dev and 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 Theo1-variance is TheoBR-variance using (6). TheoBR-dev, like Theo1-dev, can extend to \(3/4\) data-run time.

\[ \tau = 0.75m\tau_0 \text{ and } \tau = m\tau_0, \text{ respectively.} \]

Figure 4. TheoH-dev combines Adev and TheoBR-dev to provide one ‘Allan-compatible’ plot from shortest to longest τ-values.

6. Confidence intervals for TheoH

An uncertainty (estimation error) can be assigned to TheoH by use of edf. TheoH 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

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.
distribution functions are narrower using ThéoH Allan variance for calculating confidence intervals, but the accuracy of the fit to simulation results is of Adev and ThéoH values for a short RWFM noise simulation, values of 90% confidence intervals versus edf as a percentage a little practical use. An empirical formula to obtain percentage formulas corresponding to section 3.3 and stipulated in (3):

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

\[
edf_{\text{FLFM}} = \frac{2.7N_x^2 - 1.3N_x - 3.5m}{N_x m} \left( \frac{m}{m^{3/2} + 5.45} \right)
\]

\[
edf_{\text{RWFM}} = \frac{4.4N_x - 2}{2.175m}
\]

\[
\times \left( \frac{(4.4N_x - 1)^2 - 6.45m(4.4N_x - 1) + 6.413m^2}{(4.4N_x - 3)^2} \right)
\]

\[
edf_{\text{WHFM}} = \frac{0.86(N_x + 1)(N_x - m)}{N_x - 0.75m} \left( \frac{m}{m + 1.52} \right)
\]

\[
edf_{\text{FLFM}} = \frac{5.54N_x^2 - 5.52N_x - 10.727m}{(m + 48.8)^{1/2}(N_x - 0.75m)} \left( \frac{m}{m + 0.4} \right)
\]

Accuracy of the fit to simulation results is ±10%.

Chi-square distribution functions are used for the Allan variance for calculating confidence intervals, but the distribution functions are narrower using ThéoH(0.75mτ0) (which is another of its benefits) at long averaging periods in which τ0 ≪ T [31]. Figure 6 shows exact upper-bound values of 90% confidence intervals versus edf as a percentage of Adev and ThéoH 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

\[
\text{Percentage error using ThéoH}(\tau) = \frac{100}{\sqrt{2(\text{edf} + 6.65)}}
\]

where edf ≥ 1 is calculated from the formulae above. This uncertainty is meant to provide some information on the spread of possible values of ThéoH(0.75mτ0) 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 occurrence 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 ‘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 classified 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 ThéoH at each octave of \( \tau \), that is, 0.75m\( \tau_0 \), is the mid point between two computations of Adev and ThéoH at \( \tau \)-values of an Adev versus \( \tau \) plot compared with ThéoH of slopes can be established based on the 90% confidence FLFM or RWFM. To generalize this assertion, a range or band of slopes as compared with Adev so that it enables faster detection of the onset of nonstationary noise types of its range of slopes as compared with Adev. A weighted, linear least-squares regression applied to Adev and ThéoH versus \( \tau \) provides a probabilistic interpretation in terms of 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 \( \tau \)-values. Figure 7 depicts what is meant by the slope range that corresponds to the expected range of slopes per octave for simulated noise consisting of 1025-point simulated data runs and using Adev and ThéoH. The plots show that for increasing averaging time, ThéoH 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 \( \tau \) ≥ 400\( \tau_0 \), or about 4% of the data run, whereas for ThéoH deviation, prevailing FLFM continues to be detected as such to \( \tau \) ≥ 200\( \tau_0 \), or about 1/5 of the data run. For WHFM, the range is below \( \pm \frac{1}{4} \), except at the longest averaging times, noting only that white noise is detected better using ThéoH deviation. All bands use 90% confidence intervals based on chi-square distribution properties (see section 6).

Figure 8. Spread of possible slopes-per-octave for simulated noise of 1025 data points of frequency noises (WHFM, FLFM and RWFM) respectively proportional to \( \tau^{-1/2} \), \( \tau^0 \), \( \tau^{+1/2} \). \( \tau \) is the midpoint between two computations of Adev and ThéoH at \( \tau_{upper} \) and \( \tau_{lower} \) 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 ThéoH versus \( \tau \)-values 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 \( \tau \)-values. Figure 7 depicts what is meant by the slope range that corresponds to the expected group or band of slopes.

At long term, ThéoH 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 ThéoH deviation versus \( \tau \) for each octave of sampling interval \( \tau \) 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 ThéoH. The plots show that for increasing averaging time, ThéoH 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 \( \tau \) ≥ 400\( \tau_0 \), or about 4% of the data run, whereas for ThéoH deviation, prevailing FLFM continues to be detected as such to \( \tau \) ≥ 200\( \tau_0 \), or about 1/5 of the data run. For WHFM, the range is below \( \pm \frac{1}{4} \), except at the longest averaging times, noting only that white noise is detected better using ThéoH deviation. All bands use 90% confidence intervals based on chi-square distribution properties (see section 6).

8. Conclusion

Théo1 variance, the genesis of ThéoH 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 \( m\tau_0 \). ThéoH deviation consists of Adev for short and mid \( \tau \)-values in combination with points at very long \( \tau \)-values of Théo1 with BR in such a way that mixed noise types can be accurately characterized. ‘H’ of ThéoH 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. ThéoH 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 \( \tau \)-values using ThéoH-versus-\( \tau \) slopes,
- characterization to longer \( \tau \)-values than with Adev, and
- characterization of \( S_1(f) \) 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 Theo1, TheoBR and TheoH. 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 Theo1(τ)

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 \( N_x = 4 \), so we have only the inner or second summation to compute, and it has two terms corresponding to indices \( \delta = 0, 1 \). They are

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

We obtain

\[
\text{Theo1}(3, 86,400 \text{ s}, 5) = \left( \frac{1}{0.75} \right) \frac{1}{16(86,400)^2} \left\{ \frac{1}{2} \times [(x_1 - x_3) + (x_5 - x_3)]^2 \right. \\
+ \left. 1 \times [(x_1 - x_2) + (x_5 - x_4)]^2 \right\}.
\]

(9)

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 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

\[
x_1 = 1.08, \quad x_2 = 0.5, \quad x_3 = 2.2, \quad x_4 = 4.68, \quad x_5 = 3.29.
\]

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

\[
x_1 - x_3 = -1.12, \quad x_5 - x_3 = 1.09
\]

and

\[
x_1 - x_2 = 0.58, \quad x_5 - x_4 = -1.39
\]

Using these values and taking the square root of Theo1, we get

\[
\text{Theo1} = \sqrt{\left( \frac{0.75}{259,200} \right) \left( 0.00045 + 0.656 \right)^{\frac{1}{2}}} = 1.18 \text{ ns}
\]

\[
\text{TheoBR} = \sqrt{\left( \frac{0.75}{259,200} \right) \left( 0.00045 + 0.656 \right)^{\frac{1}{2}}} = 1.18 \text{ ns}
\]

\[
\text{TheoH} = \sqrt{\left( \frac{0.75}{259,200} \right) \left( 0.00045 + 0.656 \right)^{\frac{1}{2}}} = 1.18 \text{ ns}
\]

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 \( \tau \)-value, but for comparison, \( \text{Adev at } \tau = 2 \text{ d} \) has a value of \( 1.23 \times 10^{-16} \), 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 Theo1-variance depends on the range or extent of index \( \delta \), which is inversely proportional to the ratio of data spacing to stride, that is, \( \tau_0 \)-to-\( \tau \). Shorter \( \tau_0 \) means larger ratio \( m \) and more terms to average in the inner summation for an interval \( \tau \). Additionally, for a fixed \( \tau \), the number of overlapping terms for the outer summation to average in (3) is inversely proportional to \( \tau_0 \)-to-\( T \). Shorter \( \tau_0 \) provides a greater ratio \( T/\tau_0 \). Second, the longest computable interval is \( \tau = 0.75(T - \tau_0) \). If we want to obtain an evaluation at the longest possible \( \tau \)-value, or Theo1 (\( \tau \)) just shy of \( T \) itself, data spacing \( \tau_0 \) should be as short as possible. Third, recall that for an evaluation at 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 \( m = \tau_0/0.75\tau_0 \) 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], \( m \geq 10 \) is recommended. In conclusion, for a given \( \tau \)-value, making data spacing \( \tau_0 \) short allows 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 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 \( \tau_0 \) 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 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:

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

Let us calculate 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 Theo1(10, 86400 s,12) for each index \( i \) are given by (for this value of \( \tau \), note that \( m = 10 \)):

\[
i = 1 : \delta = 0:
\]

\[
\begin{align*}
&\delta = 0 : \frac{1}{2}((-2.15 - (-3.71)) + (4.68 - (-3.71)))^2 = 29.145, \\
&\delta = 1 : \frac{1}{2}((-2.15 - 0.65) + (4.68 - (-3.3)))^2 = 29.145, \\
&\delta = 2 : \frac{1}{2}((-2.15 - 2.5) + (4.68 - 1.08))^2 = 29.145, \\
&\delta = 3 : \frac{1}{2}((-2.15 - 1) + (4.68 - 0.5))^2 = 29.145, \\
&\delta = 4 : \frac{1}{2}((-2.15 - (-0.99)) + (4.68 - 2.2))^2 = 29.145,
\end{align*}
\]

\[
i = 2 : \delta = 0:
\]

\[
\begin{align*}
&\delta = 0 : \frac{1}{2}((-0.99 - (-3.3)) + (3.29 - (-3.3)))^2 = 36.66, \\
&\delta = 1 : \frac{1}{2}((-0.99 - (-3.71)) + (3.29 - 1.08))^2 = 36.66, \\
&\delta = 2 : \frac{1}{2}((-0.99 - 0.65) + (3.29 - 0.5))^2 = 36.66, \\
&\delta = 3 : \frac{1}{2}((-0.99 - 2.5) + (3.29 - 2.2))^2 = 36.66, \\
&\delta = 4 : \frac{1}{2}((-0.99 - 1) + (3.29 - 4.68))^2 = 36.66,
\end{align*}
\]

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

\[
\text{Theo1}(10, 86400 s, 12) = \frac{1}{2(0.75)(10)^2} \sum_{i=1}^{2} (\ldots) = 0.4387,
\]

and

\[
\text{Theo1-dev}(10, 86400 s, 12) = \sqrt{0.4387} = 0.6623.
\]

Since the original measurements were in units of ns/day, we obtain 0.6623 × 1 ns/86400 s for a sampling period of \( \tau_0 = 1 \text{ d} = 86400 \text{ s} \). Therefore, for this set, Theo1-dev(10, 86400 s, 12) = 7.66 × 10\(^{-15}\).

References

[13] Parker T E, Jefferts S R, Heavner T P and Donley E A 2005 Operation of the NIST-F1 cesium fountain primary frequency standard with a maser ensemble, including the impact of frequency transfer noise Metrologia 42 423–30
[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. Stand. (US) Technical Note TN-375
[25] Riley W J and Greenhall C A 2004 Notes on calibration of Total Avari to Avari, private communications


[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) section 9.5–9.7


[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

