TOTAL VARIANCE AS AN EXACT ANALYSIS OF THE SAMPLE VARIANCE*

Donald B. Percival
Applied Physics Laboratory
University of Washington
Box 355640
Seattle, WA 98195–5640 USA
E-mail: dbp@apl.washington.edu

David A. Howe
Time and Frequency Division
National Institute of
Standards and Technology
325 Broadway
Boulder, CO 80303 USA
E-mail: dhowe@nist.gov

Abstract

Given a sequence of fractional frequency deviates, we investigate the relationship between the sample variance of these deviates and the total variance (Totvar) estimator of the Allan variance. We demonstrate that we can recover exactly twice the sample variance by renormalizing the Totvar estimator and then summing it over dyadic averaging times 1, 2, 4, ..., 2^J along with one additional term that represents variations at all dyadic averaging times greater than 2^J. This decomposition of the sample variance mimics a similar theoretical decomposition in which summing the true Allan variance over all possible dyadic averaging times yields twice the process variance. We also establish a relationship between the Totvar estimator of the Allan variance and a biased maximal overlap estimator that uses a circularized version of the original fractional frequency deviates.

1 INTRODUCTION

The goal of this paper is to explore the relationship between the sample variance of a sequence of fractional frequency deviates $\{y_n : n = 1, \ldots, N_y\}$, namely,

$$\hat{\sigma}_y^2 \equiv \frac{1}{N_y} \sum_{n=1}^{N_y} (y_n - \overline{y})^2, \text{ where } \overline{y} \equiv \frac{1}{N_y} \sum_{n=1}^{N_y} y_n,$$

and a new estimator of the Allan variance called "Totvar" ("total variance" – see the companion article by Howe and Greenhall [1] in these Proceedings for additional details). The Totvar estimator is based upon the hypothesis that reasonable surrogates for unobserved deviates y_n, $n < 1$ or $n > N_y$, can be formed by tacking on reversed versions of $\{y_n\}$ at the beginning and end of the original series. The Totvar estimator makes use of certain of these surrogate values in order to come up with a new estimator of the Allan variance that has better mean-squared error properties than the usual Allan variance estimator at the very largest sampling times (Howe and Greenhall [1]). Here we show that a renormalized version of the Totvar estimator can be used to exactly decompose twice the sample variance. Except for the factor of two (an historical artifact due to the original definition of the Allan variance), this decomposition of the sample variance is very much similar to the one afforded by traditional spectral analysis estimators, which exactly decompose the sample variance across different Fourier frequencies. By comparison our results show that the (renormalized) Totvar estimator decomposes the sample variance across dyadic averaging times (i.e., averaging times of the form $2^j\tau_0$, where τ_0 is the sample period for $\{y_n\}$). Our result thus says that the Allan variance can be regarded as an example of an analysis of variance technique, which is one of the most widely used data analysis methods in modern statistics.

The remainder of this paper is organized as follows. In Section 2 we recall that in fact a very early estimator of the Allan variance (the nonoverlapped estimator) exactly decomposes twice the sample variance for the special cases when N_y is a power of two. Because of its poor variance properties, the nonoverlapped estimator is very seldom used, so we discuss in Section 3 what is generally considered to be the preferred estimator, namely, the maximal overlap estimator. The usual formulation of this estimator does *not* yield a decomposition of twice the sample variance; however, if we view this estimator as the mean-squared output of a circular filtering operation, we can augment the estimator with additional terms (namely, ones that make explicit use of the circularity assumption) and come up with an biased version of the maximal overlap estimator that does yield a decomposition of twice the sample variance for any sample size N_y. Because of the potential mismatch between y_1 and y_{N_y}, this circularity assumption can lead to serious biases. Thus, in Section 4 we consider using the biased maximal overlap estimator with the series of length $2N_y$ formed by tacking on a reversed version of $\{y_n\}$ at the end of the original series. This new estimator can be written as a renormalized version of the Totvar estimator. In Section 5 we summarize our results and conclude with a few comments.

2 THE NONOVERLAPPED ESTIMATOR OF THE ALLAN VARIANCE

For this section only we assume that the sample size is a power of two ; i.e., we can write $N_y = 2^J$ for a positive integer J. Given a sequence of τ_0-average fractional frequency deviates $\{y_n : n = 1, \ldots, N_y\}$ with a sampling period between adjacent observations given by τ_0 also, let us define the $m\tau_0$-average fractional frequency deviate as

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

If we regard $\{\overline{y}_n(m) : n = m, \ldots, N_y\}$ as a realization of one portion of the stochastic process $\{\overline{Y}_n(m) : n = 0, \pm1, \pm2, \ldots\}$, the Allan variance for averaging time $m\tau_0$ is defined as

$$\sigma_y^2(m) \equiv \frac{1}{2}E\left\{\left[\overline{Y}_n(m) - \overline{Y}_{n-m}(m)\right]^2\right\},$$

where we assume that the stochastic process is such that the expectation above in fact depends on the averaging time index m, but not on the time index n (this will be true if the first difference process $\{\overline{Y}_n(1) - \overline{Y}_{n-1}(1)\}$ is a stationary process).

For $m = 2^j$ for $j = 0, 1, \ldots, J-1$, let us form the so-called nonoverlapped estimator of the Allan variance:

$$\hat{\sigma}^2_{y,\text{nooo}}(2^j) \equiv \frac{2^j}{N_y}\sum_{k=1}^{\frac{N_y}{2^{j+1}}}\left[\overline{y}_{2k2^j}(2^j) - \overline{y}_{(2k-1)2^j}(2^j)\right]^2.$$

For example, if $j = 0$ so that $m = 1$, the above reduces to

$$\hat{\sigma}^2_{y,\text{nono}}(1) = \frac{1}{N_y} \sum_{k=1}^{\frac{N_y}{2}} [y_{2k} - y_{2k-1}]^2, \tag{1}$$

so that each y_n contributes to exactly one term in the sum of squares above (hence the origin of the name "nonoverlapped estimator"). At the other extreme when $j = J - 1$ so that $m = \frac{N_y}{2}$, we have

$$\hat{\sigma}^2_{y,\text{nono}}(\tfrac{N_y}{2}) = \frac{1}{2}\left[\overline{y}_{N_y}(\tfrac{N_y}{2}) - \overline{y}_{\frac{N_y}{2}}(\tfrac{N_y}{2})\right]^2.$$

The nonoverlapped estimator can be interpreted in terms of an orthonormal transform of the column vector $\mathbf{y}$ whose elements are given by $\{y_n\}$. For $N_y = 8$, this transform is given by the following 8×8 matrix:

$$\mathcal{W} \equiv \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \end{bmatrix}$$

(for other N_y, the $N_y \times N_y$ matrix $\mathcal{W}$ is formulated in an analogous manner and and is one version of the discrete Haar wavelet transform – for details, see, e.g., [4]). Letting $\mathbf{w} = \mathcal{W}\mathbf{y}$ and letting $\{w_n\}$ denote the elements of $\mathbf{w}$, it follows that

$$\begin{aligned} w_1^2 + w_2^2 + w_3^2 + w_4^2 &= 4\hat{\sigma}^2_{y,\text{nono}}(1) \\ w_5^2 + w_6^2 &= 4\hat{\sigma}^2_{y,\text{nono}}(2) \\ w_7^2 &= 4\hat{\sigma}^2_{y,\text{nono}}(4) \\ w_8^2 &= 8\overline{y}^2 \end{aligned}$$

Because $\mathcal{W}$ is an orthonormal transform, we must have $\|\mathbf{w}\|^2 = \|\mathbf{y}\|^2$, where $\|\mathbf{x}\|$ is the usual Euclidean norm of the vector $\mathbf{x}$. It follows that

$$\hat{\sigma}^2_y \equiv \frac{1}{8}\sum_{n=1}^{8}(y_n - \overline{y})^2 = \frac{1}{8}\sum_{n=1}^{7} w_n^2 = \frac{1}{2}\sum_{j=0}^{2}\hat{\sigma}^2_{y,\text{nono}}(2^j).$$

For general $N_y = 2^J$, the corresponding result is

$$\hat{\sigma}^2_y = \frac{1}{2}\sum_{j=0}^{J-1}\hat{\sigma}^2_{y,\text{nono}}(2^j) \text{ or, equivalently, } \sum_{j=0}^{J-1}\hat{\sigma}^2_{y,\text{nono}}(2^j) = 2\hat{\sigma}^2_y;$$

i.e., summing the nonoverlapped Allan variance estimator over all dyadic averaging times less than or equal to $\frac{N_y}{2}$ yields exactly twice the sample variance (for additional details and some historical background, see Section III of [3]).

3 MAXIMAL OVERLAP ESTIMATORS OF THE ALLAN VARIANCE

The nonoverlapped estimator of the Allan variance is rarely used in practice because it does not take advantage of certain "information" regarding $\sigma_y^2(m)$. To see in what sense this is true, let us consider the form of the maximal overlap estimator for $m = 1$:

$$\hat{\sigma}^2_{y,\text{mo}}(1) = \frac{1}{2(N_y - 1)} \sum_{k=2}^{N_y} [y_k - y_{k-1}]^2 .$$

Note that, whereas each y_n appears exactly once in the nonoverlapped estimator of Equation (1), the variables $y_2, \ldots, y_{N_y-1}$ appear twice because now, in addition to terms like $[y_2 - y_1]^2$, $[y_4 - y_3]^2$ and $[y_6 - y_5]^2$ that appear in $\hat{\sigma}^2_{y,\text{no}}(1)$, the maximal overlap estimator also includes terms like $[y_3 - y_2]^2$ and $[y_5 - y_4]^2$. For general m the maximal overlap estimator takes the form

$$\hat{\sigma}^2_{y,\text{mo}}(m) = \frac{1}{2(N_y - 2m + 1)} \sum_{k=2m}^{N_y} [\overline{y}_k(m) - \overline{y}_{k-m}(m)]^2 .$$

Even if we were to restrict the sample size N_y to be a power of two 2^J, it can be argued that in general

$$\sum_{j=0}^{J-1} \hat{\sigma}^2_{y,\text{mo}}(2^j) \neq 2\hat{\sigma}^2_y,$$

so the usual maximal overlap estimator does *not* constitute an analysis of twice the sample variance. There is, however, an interesting way to define a variation on the maximal overlap estimator that in fact does yield an exact analysis of variance, as the following argument shows.

We start with two filters $\{\tilde{h}_{0,l}\}$ and $\{\tilde{g}_{0,l}\}$ defined as follows:

$$\tilde{h}_{0,l} \equiv \begin{cases} \frac{1}{2}, & l = 0; \\ -\frac{1}{2}, & l = 1; \text{ and} \\ 0, & l < 0 \text{ or } l \geq 2; \end{cases} \quad \text{and } \tilde{g}_{0,l} \equiv \begin{cases} \frac{1}{2}, & l = 0; \\ \frac{1}{2}, & l = 1; \text{ and} \\ 0, & l < 0 \text{ or } l \geq 2 \end{cases}$$

(in the wavelet literature, $\{\tilde{h}_{0,l}\}$ and $\{\tilde{g}_{0,l}\}$ are two versions of what are called the Haar wavelet and scaling filters – for details, see, e.g., [4]). Let $\tilde{H}_0(\cdot)$ and $\tilde{G}_0(\cdot)$ be the transfer functions for $\{\tilde{h}_{0,l}\}$ and $\{\tilde{g}_{0,l}\}$:

$$\tilde{H}_0(f) \equiv \sum_{l=-\infty}^{\infty} h_l e^{-i2\pi fl} = ie^{-i\pi f} \sin(\pi f) \quad \text{and} \quad \tilde{G}_0(f) \equiv \sum_{l=-\infty}^{\infty} g_l e^{-i2\pi fl} = e^{-i\pi f} \cos(\pi f);$$

i.e., $\tilde{H}_0(\cdot)$ is the discrete Fourier transform (DFT) of $\{\tilde{h}_{0,l}\}$. Note that we have

$$|\tilde{H}_0(f)|^2 + |\tilde{G}_0(f)|^2 = 1 \text{ for all } f. \tag{2}$$

We want to circularly filter $\{y_n\}$ separately using the filters $\{\tilde{h}_{0,l}\}$ and $\{\tilde{g}_{0,l}\}$. Formally, we do so by defining $\{\tilde{h}^\circ_{0,l} : l = 0, \ldots, N_y - 1\}$ and $\{\tilde{g}^\circ_{0,l} : l = 0, \ldots, N_y - 1\}$, which are said to be $\{\tilde{h}_{0,l}\}$ and $\{\tilde{g}_{0,l}\}$ periodized to length N_y. By definition,

$$\tilde{h}^\circ_{0,l} = \sum_{k=-\infty}^{\infty} \tilde{h}_{0,l+kN_y}, \quad l = 0, \ldots, N_y - 1,$$

with a similar definition for $\{\tilde{g}^\circ_{0,l}\}$. If $N_y \geq 2$, we have

$$\tilde{h}^\circ_{0,l} \equiv \begin{cases} \frac{1}{2}, & l=0; \\ -\frac{1}{2}, & l=1; \text{ and} \\ 0, & 2 \leq l \leq N_y - 1; \end{cases} \quad \text{and} \quad \tilde{g}^\circ_{0,l} \equiv \begin{cases} \frac{1}{2}, & l=0; \\ \frac{1}{2}, & l=1; \text{ and} \\ 0, & 2 \leq l \leq N_y - 1; \end{cases}$$

if, however, $N_y = 1$ so that $\{\tilde{h}^\circ_{0,l}\}$ and $\{\tilde{g}^\circ_{0,l}\}$ each have but a single term, then $\tilde{h}^\circ_{0,0} = 0$ and $\tilde{g}^\circ_{0,0} = 1$. It is an easy exercise to show that the DFT of $\{\tilde{h}^\circ_{0,l}\}$ can be obtained by subsampling the DFT for $\{\tilde{h}_{0,l}\}$; i.e.,

$$\sum_{l=0}^{N_y-1} \tilde{h}^\circ_{0,l} e^{-i2\pi kl/N_y} = H_0(\tfrac{k}{N_y}), \quad k = 0, \ldots, N_y - 1.$$

The finite sequences $\{\tilde{h}^\circ_{0,l}\}$ and $\{H_0(\frac{k}{N_y})\}$ thus constitute a Fourier transform pair, a relationship we will express as

$$\{\tilde{h}^\circ_{0,l} : l = 0, \ldots, N_y - 1\} \longleftrightarrow \{H_0(\tfrac{k}{N_y}) : k = 0, \ldots, N_y - 1\}.$$

Similarly we have

$$\{\tilde{g}^\circ_{0,l} : l = 0, \ldots, N_y - 1\} \longleftrightarrow \{G_0(\tfrac{k}{N_y}) : k = 0, \ldots, N_y - 1\}.$$

Let us now define

$$\widetilde{w}_{0,n} \equiv \sum_{l=0}^{N_y-1} \tilde{h}^\circ_{0,l} y_{(n-l) \bmod N_y} \text{ and } \widetilde{v}_{0,n} \equiv \sum_{l=0}^{N_y-1} \tilde{g}^\circ_{0,l} y_{(n-l) \bmod N_y}, \quad n = 1, \ldots, N_y,$$

where we define $n \bmod N_y$ to be n if $1 \leq n \leq N_y$ and to be $n + kN_y$ otherwise, where k is the unique nonzero integer such that $1 \leq n + kN_y \leq N_y$ (thus $-1 \bmod N_y = N_y - 1$; $0 \bmod N_y = N_y$; $1 \bmod N_y = 1$; $\ldots$; $N_y \bmod N_y = N_y$; $N_y + 1 \bmod N_y = 1$; etc.). By construction we have

$$\frac{2}{N_y - 1} \sum_{n=2}^{N_y} \widetilde{w}^2_{0,n} = \hat{\sigma}^2_{y,\text{mo}}(1); \tag{3}$$

i.e., we have expressed the maximal overlap estimator of the Allan variance for $m = 1$ in terms of a sum of squares of the output from circular filtering $\{y_n\}$ with $\{\tilde{h}^\circ_{0,l}\}$.

An important point to note is that $\hat{\sigma}^2_{y,\text{mo}}(1)$ does not involve the entire output from the filter: it is missing $\widetilde{w}_{0,1} \propto y_{N_y} - y_1$, which is the only term that explicitly makes use of the circularity assumption. Inclusion of this term is one of the two keys to defining a version of the maximal overlap estimator that constitutes an analysis of variance. The other key is to recognize that $\hat{\sigma}^2_{y,\text{mo}}(2^j)$ for $j = 1, 2, \ldots$ can be obtained by further filtering of $\{\widetilde{v}_{0,n}\}$ so that, whereas $\{\widetilde{w}_{0,n}\}$ contains information about the variations of $\{y_n\}$ at τ_0 averaging times, the series $\{\widetilde{v}_{0,n}\}$ contains information about variations of $\{y_n\}$ at all dyadic averaging times higher than τ_0 (i.e., $2\tau_0$, $4\tau_0$, etc.). Accordingly, let $\widetilde{\mathbf{w}}_0$ be an N_y dimensional vector whose elements are $\{\widetilde{w}_{0,n}\}$, and define $\widetilde{\mathbf{v}}_0$ to contain $\{\widetilde{v}_{0,n}\}$. Letting $\{\mathcal{Y}_k\}$ be the DFT of $\{y_n\}$, we have (from a standard theorem in filtering theory)

$$\{\widetilde{w}_{0,n}\} \longleftrightarrow \{\widetilde{H}_0(\tfrac{k}{N_y})\mathcal{Y}_k\} \text{ and } \{\widetilde{v}_{0,n}\} \longleftrightarrow \{\widetilde{G}_0(\tfrac{k}{N_y})\mathcal{Y}_k\}.$$

Parseval's theorem tells us that

$$\|\widetilde{\mathbf{w}}_0\|^2 = \frac{1}{N}\sum_{k=0}^{N_y-1} |\widetilde{H}_0(\tfrac{k}{N_y})|^2|\mathcal{Y}_k|^2 \text{ and } \|\widetilde{\mathbf{v}}_0\|^2 = \frac{1}{N}\sum_{k=0}^{N_y-1} |\widetilde{G}_0(\tfrac{k}{N_y})|^2|\mathcal{Y}_k|^2,$$

which in turn yields

$$\|\widetilde{\mathbf{w}}_0\|^2 + \|\widetilde{\mathbf{v}}_0\|^2 = \frac{1}{N}\sum_{k=0}^{N_y-1} |\mathcal{Y}_k|^2 \left(|\widetilde{H}_0(\tfrac{k}{N_y})|^2 + |\widetilde{G}_0(\tfrac{k}{N_y})|^2\right) = \frac{1}{N}\sum_{k=0}^{N_y-1} |\mathcal{Y}_k|^2 = \|\widetilde{\mathbf{y}}\|^2,$$

where we have made use of Equation (2) and a second application of Parseval's theorem.

Let us now define the following estimator of the Allan variance for $m = 1$:

$$\tilde{\sigma}^2_{y,\text{maxo}}(1) \equiv \frac{2}{N_y}\sum_{n=1}^{N_y} \widetilde{w}^2_{0,n} = \frac{2}{N_y}\|\widetilde{\mathbf{w}}_0\|^2.$$

We refer to this estimator as the biased maximal overlap estimator of $\sigma^2_y(1)$ based on $\{y_n\}$. It differs from the standard maximal overlap estimator (Equation (3)) because of an additional term proportional to $(y_{N_y} - y_1)^2$. Although this estimator is in general a biased estimator of the true Allan variance, it is in fact unbiased when $\{y_n\}$ is a white noise process. It satisfies the analysis of variance condition

$$\tilde{\sigma}^2_{y,\text{maxo}}(1) + \tilde{\eta}^2_{y,\text{maxo}}(2) = 2\hat{\sigma}^2_y, \text{ where } \tilde{\eta}^2_{y,\text{maxo}}(2) \equiv \frac{2}{N_y}\|\widetilde{\mathbf{v}}_0\|^2 - 2\overline{y}^2.$$

We can regard the second piece of the decomposition $\tilde{\eta}^2_{y,\text{maxo}}(2)$ as being related to variations in $\{y_n\}$ at dyadic averaging times of 2 and greater.

Just as $\{y_n\}$ was split into the components $\{\widetilde{w}_{0,n}\}$ and $\{\widetilde{v}_{0,n}\}$, we now split $\{\widetilde{v}_{0,n}\}$ into two components, namely, $\{\widetilde{w}_{1,n}\}$ and $\{\widetilde{v}_{1,n}\}$. The first component $\{\widetilde{w}_{1,n}\}$ will be used to construct an estimator of $\sigma^2_y(2)$, while the second component is related to variations in $\{y_n\}$ at dyadic averaging times of 4 and greater. The filters that accomplish the desired split are N_y-periodized versions of ones whose transfer functions are defined by $\widetilde{H}_0(2f)$ and $\widetilde{G}_0(2f)$ – the impulse response sequence for these filters can be formed by taking the original filters $\{\tilde{h}_{0,l}\}$ and $\{\tilde{g}_{0,l}\}$ and inserting a single zero after each element, a procedure that is known as upsampling in the engineering literature [4]. For example, since the $l = 0, 1, 2, 3$ and 4 values of the impulse response sequence for $\widetilde{H}_0(f)$ are given by $\frac{1}{2}$, $-\frac{1}{2}$, 0, 0 and 0, the corresponding values for $\widetilde{H}_0(2f)$ are given by $\frac{1}{2}$, 0, $-\frac{1}{2}$, 0 and 0. We can also obtain $\{\widetilde{w}_{1,n}\}$ and $\{\widetilde{v}_{1,n}\}$ by directly filtering $\{y_n\}$:

$$\widetilde{w}_{1,n} = \sum_{l=0}^{N_y-1} \tilde{h}^\circ_{1,l} y_{(n-l) \bmod N_y} \text{ and } \widetilde{v}_{1,n} = \sum_{l=0}^{N_y-1} \tilde{g}^\circ_{1,l} y_{(n-l) \bmod N_y}, \quad n = 1, \ldots, N_y,$$

where $\{\tilde{h}^\circ_{1,l}\}$ is the circular filter such that

$$\{\tilde{h}^\circ_{1,l}\} \longleftrightarrow \{\widetilde{H}_0(\tfrac{2k}{N_y})\widetilde{G}_0(\tfrac{k}{N_y})\}; \text{ likewise, } \{\tilde{g}^\circ_{1,l}\} \longleftrightarrow \{\widetilde{G}_0(\tfrac{2k}{N_y})\widetilde{G}_0(\tfrac{k}{N_y})\}.$$

Note that the impulse response sequence for $\{\tilde{h}^\circ_{1,l}\}$ is the circular convolution of the impulse response sequence for $\widetilde{G}_0(\frac{k}{N_y})$ and $\widetilde{H}_0(\frac{2k}{N_y})$, i.e., $\frac{1}{2}$, $\frac{1}{2}$, 0, …convolved with $\frac{1}{2}$, 0, $-\frac{1}{2}$, 0, which yields $\frac{1}{4}$, $\frac{1}{4}$, $-\frac{1}{4}$, $-\frac{1}{4}$, 0, …(as long as $N_y > 4$). This latter filter is seen to be proportional to the filtering operation commonly used in estimates of $\sigma_y^2(2)$.

Define $\tilde{\mathbf{w}}_1$ as the N_y dimensional vector containing $\{\tilde{w}_{1,n}\}$, and let $\tilde{\mathbf{v}}_1$ contain $\{\tilde{v}_{1,n}\}$. By a simple variation on the argument used to establish $\|\tilde{\mathbf{w}}_0\|^2 + \|\tilde{\mathbf{v}}_0\|^2 = \|\tilde{\mathbf{y}}\|^2$, we have

$$\|\tilde{\mathbf{w}}_1\|^2 + \|\tilde{\mathbf{v}}_1\|^2 = \|\tilde{\mathbf{v}}_0\|^2.$$

Now define the following estimator of the Allan variance for $m = 2$:

$$\tilde{\sigma}^2_{y,\text{mmo}}(2) \equiv \frac{2}{N_y}\sum_{n=1}^{N_y} \tilde{w}^2_{1,n} = \frac{2}{N_y}\|\tilde{\mathbf{w}}_1\|^2.$$

This is the standard maximal overlap estimator with three additional terms – these are proportional to $(y_1 + y_{N_y} - y_{N_y-2} - y_{N_y-1})^2$, $(y_1 + y_2 - y_{N_y-1} - y_{N_y})^2$ and $(y_2 + y_3 - y_1 - y_{N_y})^2$. In general this estimator is a biased estimator of $\sigma_y^2(2)$. We have the analysis of variance condition

$$\tilde{\sigma}^2_{y,\text{mmo}}(1) + \tilde{\sigma}^2_{y,\text{mmo}}(2) + \tilde{\eta}^2_{y,\text{mmo}}(4) = 2\hat{\sigma}^2_y, \text{ where } \tilde{\eta}^2_{y,\text{mmo}}(4) \equiv \frac{2}{N_y}\|\tilde{\mathbf{v}}_1\|^2 - 2\overline{y}^2.$$

We can now state the result for general J, a proof of which follows from an easy inductive argument. We define

$$\tilde{w}_{j,n} = \sum_{l=0}^{N_y-1} \tilde{h}^\circ_{j,l} y_{(n-l) \bmod N_y} \text{ and } \tilde{v}_{j,n} = \sum_{l=0}^{N_y-1} \tilde{g}^\circ_{j,l} y_{(n-l) \bmod N_y}, \quad n = 1, \ldots, N_y,$$

where

$$\{\tilde{h}^\circ_{j,l}\} \longleftrightarrow \{\widetilde{H}_0(\tfrac{2^j k}{N_y})\widetilde{G}_0(\tfrac{2^{j-1}k}{N_y})\widetilde{G}_0(\tfrac{2^{j-2}k}{N_y})\cdots\widetilde{G}_0(\tfrac{k}{N_y})\}$$

and

$$\{\tilde{g}^\circ_{j,l}\} \longleftrightarrow \{\widetilde{G}_0(\tfrac{2^j k}{N_y})\widetilde{G}_0(\tfrac{2^{j-1}k}{N_y})\widetilde{G}_0(\tfrac{2^{j-2}k}{N_y})\cdots\widetilde{G}_0(\tfrac{k}{N_y})\}.$$

An inductive argument can be used to show that $\{\tilde{h}^\circ_{j,l}\}$ is the usual filter involved in estimating $\sigma_y^2(2^j)$. Letting $\tilde{\mathbf{w}}_j$ and $\tilde{\mathbf{v}}_j$ be N_y dimensional vectors containing $\{\tilde{w}_{j,n}\}$ and $\{\tilde{v}_{j,n}\}$, define

$$\tilde{\sigma}^2_{y,\text{mmo}}(2^j) \equiv \frac{2}{N_y}\sum_{n=1}^{N_y} \tilde{w}^2_{j,n} = \frac{2}{N_y}\|\tilde{\mathbf{w}}_j\|^2,$$

which is the biased maximal overlap estimator of $\sigma_y^2(2^j)$ based upon $\{y_n\}$ – it differs from the standard maximal overlap estimator due to $2^{j+1} - 1$ additional terms involving explicit circular use of $\{y_n\}$. For any J, the biased maximal overlap estimators satisfy the analysis of variance condition

$$\sum_{j=0}^{J} \tilde{\sigma}^2_{y,\text{mmo}}(2^j) + \tilde{\eta}^2_{y,\text{mmo}}(2^{J+1}) = 2\hat{\sigma}^2_y \text{ with } \tilde{\eta}^2_{y,\text{mmo}}(2^{J+1}) \equiv \frac{2}{N_y}\|\tilde{\mathbf{v}}_J\|^2 - 2\overline{y}^2,$$

where the term $\tilde{\eta}^2_{y,\text{mmo}}(2^{J+1})$ represents variations in $\{y_n\}$ at dyadic averaging times of 2^{J+1} and greater.

4 ESTIMATION OF THE ALLAN VARIANCE USING A CIRCULARIZED SERIES

For most models for $\{y_n\}$ of interest for actual frequency standards, a circularity assumption can yield an unacceptably large bias in the estimator $\tilde{\sigma}^2_{y,\text{mmo}}(2^j)$ due to the fact that y_1 and y_{N_y} can be quite different. To solve this problem, we construct a series $\{y^*_n\}$ of length $2N_y$

$$y^*_n \equiv \begin{cases} y_n, & 1 \le n \le N_y; \text{ and} \\ y_{2N_y+1-n}, & N_y + 1 \le n \le 2N_y. \end{cases}$$

For example, if $N_y = 3$ so that only y_1, y_2 and y_3 are observed, the values of $y^*_1, \ldots, y^*_6$ are given, respectively, by $y_1, y_2, y_3, y_3, y_2, y_1$. Note that, by construction, the sample mean and variance of $\{y_n\}$ and $\{y^*_n\}$ are identical. We now apply the estimation procedure of Section 3 to $\{y^*_n\}$ to obtain the following estimator of $\sigma^2_y(2^j)$:

$$\tilde{\sigma}^2_{y^*,\text{mmo}}(2^j) \equiv \frac{1}{N_y} \|\tilde{\mathbf{w}}^*_j\|^2,$$

where $\tilde{\mathbf{w}}^*_j$ is a vector of length $2N_y$ formed by circularly filtering $\{y^*_n\}$ with the circular filter of length $2N_y$ whose DFT is given by

$$\tilde{H}_0(\tfrac{2^j k}{2N_y})\tilde{G}_0(\tfrac{2^{j-1}k}{2N_y})\tilde{G}_0(\tfrac{2^{j-2}k}{2N_y})\cdots\tilde{G}_0(\tfrac{k}{2N_y}), \quad k = 0, \ldots, 2N_y - 1.$$

We refer to $\tilde{\sigma}^2_{y^*,\text{mmo}}(2^j)$ as the biased maximal overlap estimator of $\sigma^2_y(2^j)$ based on the circularized series $\{y^*_n\}$ (note, however, that this estimator is in fact unbiased for the special case where $\{y_n\}$ is a white noise process). This biased estimator satisfies the following analysis of variance condition for all J and all sample sizes N_y:

$$\sum_{j=0}^{J} \tilde{\sigma}^2_{y^*,\text{mmo}}(2^j) + \tilde{\eta}^2_{y^*,\text{mmo}}(2^{J+1}) = 2\hat{\sigma}^2_y, \text{ where } \tilde{\eta}^2_{y^*,\text{mmo}}(2^{J+1}) \equiv \frac{2}{N_y}\|\tilde{\mathbf{v}}^*_J\|^2 - 2\overline{y}^2.$$

Finally we note that there is a very simple relationship between $\tilde{\sigma}^2_{y^*,\text{mmo}}(2^j)$ and Totvar estimator (Greenhall, 1997, private communication):

$$\text{Totvar}(2^j, N_y, \tau_0) = \frac{N_y}{N_y - 1}\tilde{\sigma}^2_{y^*,\text{mmo}}(2^j),$$

where Totvar is defined as in Equation (4) of Howe and Greenhall [1].

5 SUMMARY AND COMMENTS

We have developed a relationship between the sample variance $\hat{\sigma}^2_y$ and the Totvar estimator Totvar(m, N_y, τ_0) of the Allan variance $\sigma^2_y(m)$, where m sets the averaging time $m\tau_0$, N_y is the number of τ_0-average fractional frequency deviates $\{y_n\}$, and τ_0 is the basic sampling and averaging time of the observed deviates. For any sample size N_y and any positive integer J, we have demonstrated that

$$\frac{N_y - 1}{N_y}\sum_{j=0}^{J} \text{Totvar}(2^j, N_y, \tau_0) + \tilde{\eta}^2_{y^*,\text{mmo}}(2^{J+1}) = 2\hat{\sigma}^2_y,$$

where $\tilde{\eta}^2_{y^*,\text{maxo}}(2^{J+1})$ can be interpreted in terms of variations in $\{y_n\}$ at all dyadic averaging times greater than 2^J. We have also shown that the Totvar estimator is related to a biased maximal overlap estimator of the Allan variance that is based upon $\{y_n^*\}$, which is a sequence of length $2N_y$ formed by tacking onto $\{y_n\}$ a reversed version of itself.

In closing, we make the following comments about our results, some of which will be expanded upon in future research.

- It can be shown that, if $\{y_n\}$ is a portion of a realization of a stationary or nonstationary process $\{Y_n\}$ for which the Allan variance is well-defined, then we have

$$\sum_{j=0}^{\infty} \sigma_y^2(2^j) = 2\sigma_y^2,$$

  where σ_y^2 is the process variance of $\{Y_n\}$ (this is taken to be infinite if $\{Y_n\}$ is nonstationary). The $\tilde{\sigma}^2_{y^*,\text{maxo}}(2^j)$ estimator is the first "modern" estimator of the Allan variance to mimic this important property.

- Because higher order Daubechies wavelet filters also satisfy Equation (2), the above development extends trivially to higher order wavelet variances (the Allan variance is essentially twice the Haar wavelet variance). These higher order wavelet variances are suitable substitutes for some of the variations on the Allan variance that have been proposed and studied in the literature (an example is the modified Allan variance). For details, see [2].

- In addition to plotting $\tilde{\sigma}^2_{y^*,\text{maxo}}(2^j)$ versus $2^j\tau_0$ on a log-log scale, we suggest that $\tilde{\eta}^2_{y^*,\text{maxo}}(2^{J+1})$ be plotted (with a separate symbol) versus $2^{J+1}\tau_0$ – this will indicate how much of $2\hat{\sigma}_y^2$ has not been accounted for by estimates of the Allan variance.

- In theory J can be made as large as desired, but there will be serious biases in $\tilde{\sigma}^2_{y^*,\text{maxo}}(2^J)$ for any J such that $2^J > N_y$. Because of its close relationship to Totvar, the results of Howe and Greenhall [1] indicate that $\tilde{\sigma}^2_{y^*,\text{maxo}}(2^j)$ outperforms traditional estimators of the Allan variance for averaging times close to $\frac{N_y}{2}$.

References

[1] D. A. Howe and C. A. Greenhall 1998, *"Total Variance: a Progress Report on a New Frequency Stability Characterization,"* **29th PTTI Meeting**, these Proceedings.

[2] D. A. Howe and D. B. Percival 1995, *"Wavelet Variance, Allan Variance, and Leakage,"* **IEEE Trans. Instrum. Meas. IM-44**, 94–97.

[3] D. B. Percival 1991, *"Characterization of Frequency Stability: Frequency-Domain Estimation of Stability Measures,"* **Proc. IEEE 79**, 961–972.

[4] M. Vetterli and J. Kovačević 1995, **Wavelets and Subband Coding** (Englewood Cliffs, New Jersey: Prentice Hall).