A Physical Interpretation of Difference Variances

Victor S. Reinhardt

Raytheon Space and Airborne Systems
El Segundo, CA 90245-902, USA

Abstract— **Most variances of the time error x(t) used in industry can be written as $\sigma_{x,M}^2(\tau)$ M^{th} order difference variances of x(t) over the interval τ or their finite sample statistics. $\sigma_{x,M}^2$ is normalized so all M-orders are equivalent when x(t) consists solely of white-x noise. However, all M-orders are not equivalent when negative power law (neg-β) noise is present, and the question arises as to which order of $\sigma_{x,M}^2$ is most appropriate to use in a given problem. The paper shows that $\sigma_{x,M}^2$ can be interpreted as an approximate measure of $\sigma_{r,M}^2$ the variance of the residual error after an $(M-1)^{th}$ order polynomial $x_{a,M}(t, \mathbf{A})$ with adjustable coefficients A is removed from x(t) by least squares fitting $x_{a,M}(t, \mathbf{A})$ to samples of x(t) over the interval $T = M\tau$. This interpretation of $\sigma_{x,M}^2$ then provides an objective rationale for choosing the appropriate M-order of $\sigma_{x,M}^2$ based on the order of the $x_{a,M}(t, \mathbf{A})$ removed from x(t) in the problem addressed. It is further noted that $x_{a,M}(t, \mathbf{A})$ can represent either a causal aging function or more generally a polynomial representation of any causal information extracted from x(t). The aging interpretation explains the sensitivity of Allan based variances ($\sigma_{x,2}^2$) to frequency drift and the insensitivity of Hadamard based variances ($\sigma_{x,3}^2$) to such drift. The information extraction interpretation leads to an even more important conclusion, that the process of extracting information from data highpass filters the residual noise with increasing efficiency as the complexity of the extracted data (as measured by M) increases. Another consequence of interpreting $\sigma_{x,M}^2$ as a measure of $\sigma_{r,M}^2$ is that the M-order of $\sigma_{x,M}^2$ is not a free parameter that one can arbitrarily change to avoid a divergence problem caused by the presence of neg-β noise. The paper shows that such divergences are indications of real problems in the design, specification, or analysis of a system, and examples are given of how to properly mitigate such divergences without arbitrarily changing M. It is finally noted that the results of this paper can be applied to variances of other variables such as $\phi(t)$ and y(t).**

Post-publication corrections: Equation (3)

I. INTRODUCTION

Most variances of the time error x(t) [2] used in industry can be written as "theoretical" M^{th} order difference variances given by

$$\sigma_{x,M}^2(\tau) = \lambda_M^{-1} E\{[\Delta(\tau)^M x(t)]^2\} \qquad (1)$$

or their finite sample statistics, where $\Delta(\tau)$ the forward difference operator [3] over the interval τ is given by

$$\Delta(\tau)x(t) = \Theta(\tau)x(t) - \Theta(0)x(t) = x(t+\tau) - x(t) \qquad (2)$$

$\Theta(\tau)$ is the forward shift operator [3] defined implicitly in (2), and λ_m is a normalization that will be defined later in this section. In this paper, we define a theoretical variance as one written in terms of a statistical expectation value or ensemble average E{..}. We use E{..} in this paper to define the theoretical variance rather the more often used infinite average time average [2], because E{..} is more appropriate in dealing with the mix of random and causal time-dependent behavior discussed in this paper (One can link the two types of averages for random noise by assuming ergodicity.). We also define a finite sample statistic of $\sigma_{x,M}^2$ in the usual manner as an approximate measure of $\sigma_{x,M}^2$, where E{..} is replaced by an arithmetic average over a finite set of data samples (such as TVAR [1][2] for $\sigma_{x,2}^2$).

TABLE I. TYPICAL VARIANCES OF THE TIME ERROR

M	$\Delta(\tau)^M x(t)$	λ_M	Associated Names	$\sigma_{x,M}^2$
0	x(t)-μ (μ=E{x(t)}) [*]	1	Standard, Sample	$E\{[x(t)-\mu]^2\}$
1	x(t+τ)-x(t)	2	TIErms [1]	$TIErms^2/2$
2	x(t+2τ)-2x(t+τ)+x(t)	6	$\sigma_x^2(\tau)$, $\sigma_y^2(\tau)$, TVAR [2]	$\tau^2\sigma_y^2(\tau)/3$
3	x(t+3τ)- 3x(t+2τ)+3x(t+τ)-x(t)	20	$_H\sigma_y^2(\tau)$ = Hadamard [4]	$3\tau^2{}_H\sigma_y^2(\tau)/10$
[*] M=0 is degenerate case. $\Delta(\tau)^0 x(t) \equiv x(t) - \mu$. $\Delta(\tau)\mu = 0$ for M > 0.				

Typical $\sigma_{x,M}^2(\tau)$ used in industry are listed in Table I [1][2][4]. As indicated in Table I, the M=0 variance $\sigma_{x,0}^2$ is the standard variance. This order is degenerate, because the

ensemble mean $\mu = E\{x(t)\}$ cannot be ignored for $M=0$ (For $M>0$, μ can be ignored because the finite difference of a constant is zero.). We thus use the special definition of $\Delta(\tau)^0 x(t)$ as $x(t) - \mu$ to include this case in our general equations. As is also indicated in the table, $\sigma_{x,1}^2$ is related to the square of the RMS time interval error (TIErms) [1], $\sigma_{x,2}^2$ is the theoretical form of the Allan variance of the time error [2], and $\sigma_{x,3}^2$ is related to the theoretical Hadamard variance as defined in [4] rather than in [5].

λ_M is defined as

$$\lambda_M = \sum_{m=0}^{M} [\frac{M!}{m!(M-m)!}]^2 \qquad (3)$$

in order to make all $\sigma_{x,M}^2$ equal for white-x or f^0 noise when the bandwidth is sufficiently wide to make the filtered noise uncorrelated over the interval τ. Thus, for white-x noise, all orders of $\sigma_{x,M}^2$ are equivalent. However, all $\sigma_{x,M}^2$ are not equivalent when negative power law (neg-β) noise is present. We note that neg-β noise is defined in the usual manner [2] as random noise with a single sideband (or double-sided) power spectral density (PSD) of x(t) given by

$$L_x(f) = 0.5 g_\beta f^\beta \qquad (4)$$

for $\beta < 1$.

Because all $\sigma_{x,M}^2$ are not equivalent when neg-β noise is present, the question arises as to which order of $\sigma_{x,M}^2$ is most appropriate as a measure of residual random error for a given problem. There has been much discussion about this question, and there is no consensus across the electronics community. Many in the IEEE UFFC society favor the sole use of $\sigma_{x,2}^2$ and related Allan based variances [2]. However, while other IEEE societies and the ITU accept the use of Allan based variances in some applications, they recommend the use of other variances also [1][6][7][8].

This paper addresses this question by investigating the properties of $\sigma_{r,M}^2$ the variance of the residual error after $x_{a,M}(t,\mathbf{A})$ an $(M-1)^{th}$ order polynomial with adjustable coefficients $\mathbf{A}$ is removed from x(t) by least squares fitting $x_{a,M}(t,\mathbf{A})$ to samples of x(t) over the interval $T = M\tau$. In the process, it is shown that the $\sigma_{x,M}^2$ and their sample statistics can be physically interpreted as approximate measures of $\sigma_{r,M}^2$. This provides a rationale for the choice of the order of $\sigma_{x,M}^2$ (or the order of its sample statistic) for a particular problem based on the order of $x_{a,M}(t,\mathbf{A})$ used in the problem.

$x_{a,M}(t)$ can represent either a causal aging function removed from data or more generally a polynomial representation of any causal information extracted from data over T. The aging interpretation explains the sensitivity of Allan based variances ($\sigma_{x,2}^2$) to frequency drift and the insensitivity of Hadamard based variances ($\sigma_{x,3}^2$) to such drift. Thus, one must match the order of $\sigma_{x,M}^2$ to that of the aging polynomial used, where its order can be either specified by the problem statement or is the one most appropriate to fit the actual aging behavior over T.

An information extraction interpretation leads to an even more important conclusion, that the process of extracting information from data highpass filters the residual noise $\sigma_{r,M}^2$ with increasing efficiency as the complexity of the extracted data (as measured by the equivalent polynomial order) increases. This conclusion is based the on the well-known spectral properties of the $\sigma_{x,M}^2$, which have spectral integral kernels proportional to f^{2M} for $f<<1$, and the approximation of $\sigma_{r,M}^2$ by $\sigma_{x,M}^2$.

Another consequence of interpreting $\sigma_{x,M}^2$ as a measure of $\sigma_{r,M}^2$ is that one is not free to arbitrarily change to a different variance order simply to avoid a divergence problem caused by the presence of neg-β noise. The paper shows that such divergences are indications of real problems in the design, specification, or analysis of the system under consideration when neg-β noise is present. Thus, the proper mitigation of such divergences involves correcting the real source of the problem and not arbitrarily changing the order of the variance. The paper further shows how to properly mitigate such divergences by using examples.

Before proceeding to the next section, we note that the linear operators $\Delta(\tau)$ and $\Theta(\tau)$ defined in (2) have been introduced because they are useful in simplifying the derivation of formulas. An example is the expansion of $\Delta(\tau)^M x(t)$ in terms of M+1 data points $x(t+n\tau)$ (n = 0 to M). Using the properties of $\Delta(\tau)$ and $\Theta(\tau)$ and the binomial theorem, it is straightforward to show for M>0 that

$$\Delta(\tau)^M x(t) = [\Theta(\tau) - \Theta(0)]^M x(t) = \sum_{n=0}^{M} C(M,n)x(t+n\tau) \quad (5)$$

where

$$C(M,n) = \frac{M!}{n!(M-n)!}(-1)^{M-n} \qquad (6)$$

Similarly, it is straightforward to show that (3) is the normalization which makes all $\sigma_{x,M}^2$ equivalent for uncorrelated white-x noise. This is accomplished by using (5) and the fact that $E\{x(t+n\tau)x(t+n'\tau)\} = \sigma_o^2 \delta_{n,n'}$.
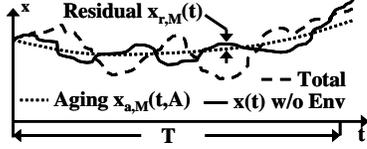
II. MODELING TIME ERROR OVER A LONG INTERVAL T

Figure 1.   Separation of time error into causal and random contributions.

In order to obtain our rationale for choosing the order of $\sigma_{x,M}^2$, let us examine how the time error over a long time interval T is modeled. As shown in Figure 1, the time error is modeled over T by separating it into causal environmental sensitivity and aging factors [9] and a residual error $x_{r,M}(t)$ after causal factors are removed. In this paper, we assume that $x_{r,M}(t)$ would consist of random neg-β noise components if the causal factors were perfectly known. Thus, we ignore causal contributions to $x_{r,M}(t)$, such as spurs, but the presence of such residual causal factors will not generally affect the validity of the paper's results.

Environmental sensitivity to temperature, power supply voltage, etc., is modeled by causal functions of these parameters [9]. Environmental coefficients for these sensitivity functions can be determined with medium term tests ($\ll$T), where the errors from the other factors can be kept small. Thus, the removal of such environmental factors from x(t) will have minimal impact on the long term behavior of x(t). From this point on, we will therefore assume that x(t) has such environmental factors removed and that this removal has little effect on our discussion.

Aging is modeled by a causal function of time. For this paper, we will model the aging over the interval $[t_o, t_o+T]$ as an $(M-1)^{th}$ order polynomial given by

$$x_{a,M}(t,\mathbf{A}) = \sum_{m=0}^{M-1} a_m(t-t_o)^m \qquad (7)$$

(for M>0) where $\mathbf{A}$ is the coefficient vector $(a_0, a_1, \ldots, a_{M-1})$. The determination of the aging coefficient vector $\mathbf{A}$ is inherently different from that of the environmental coefficients. This is because aging coefficient determination must be accomplished over an interval comparable to T for the determination to be accurate. Otherwise, terms that affect the time error over T cannot be properly estimated. This will be crucial to the discussions that follow, because long term correlations from neg-β noise will impact the determination of $\mathbf{A}$ from data over T. It is noted that we will assume $\mathbf{A}$ is independent of time over the interval T (but not necessarily over longer periods).

We noted in the introduction that $x_{a,M}(t,\mathbf{A})$ can also be more generally interpreted as a function representing the extraction of information from data over T. For brevity in most of the discussion that follows, however, we will refer to $x_{a,M}(t,\mathbf{A})$ solely as an aging function with the understanding that the equations also apply to the information extraction interpretation.

The residual time error after removing aging (and environmental factors) is

$$x_{r,M}(t) = x(t) - x_{a,M}(t,\mathbf{A}) \ (x(t)-\mu \text{ for } M=0) \qquad (8)$$

We will assume this residual can be represented by the sum of neg-β noise components when the true or actual aging $x_{a,M}(t,\mathbf{A}^{(o)})$ is perfectly known. Thus, we can characterize $x_{r,M}(t)$ using a variance given by

$$\sigma_{r,M}^2 = E\{x_{r,M}(t)^2\} \qquad (9)$$

where the coefficients $\mathbf{A}$ used to generate the residual $x_{r,M}(t)$ are determined by a fit of $x_{a,M}(t,\mathbf{A})$ to discrete x(t) samples over T. The behavior of this aging fit will lead us to our rationale for choosing the order of $\sigma_{x,M}^2$.

III. BEHAVIOR OF AGING FIT WHEN NEG-β NOISE IS PRESENT

Consider x(t) given by a true or actual polynomial $x_{a,M}(\mathbf{A}^{(o)},t)$ plus neg-β noise. One can use a least squares fit (LSQF) to simultaneously estimate $\mathbf{A}$ and $\sigma_{r,M}^2$ from N samples of x(t) over T. In a numerical simulation, one can also calculate the difference between the estimated aging and the true aging given by

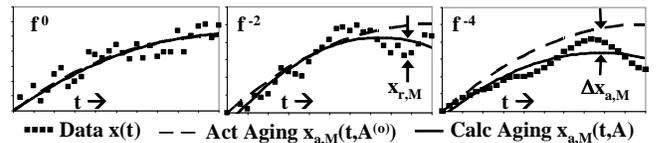$$\Delta x_{a,M}(t) = x_{a,M}(t,\mathbf{A}) - x_{a,M}(t,\mathbf{A}^{(o)}) \qquad (10)$$

Figure 2.   Behavior of LSQF for various types of f^β noise.

Figure 2 shows the results of such a numerical simulation for various β values. For white-x or $\beta=0$ noise, one can see that both $\sigma_{r,M}^2$ and $\Delta x_{a,M}$ behave in the classically expected manner for an uncorrelated noise input. That is, they both go to zero as N$\rightarrow\infty$ (assuming the bandwidth of the system also goes to infinity, so the points remain uncorrelated). On the other hand, for neg-β noise with $\beta\leq$-2, one can see from the figure that $\sigma_{r,M}^2$ and $\Delta x_{a,M}$ don't go to zero as N$\rightarrow\infty$. This is because long term noise correlations from the neg-β noise strongly impact the aging solution. It is noted that f^{-1} noise is a borderline case and behaves more like white-x noise. Thus,

in future references to neg-β noise, we will implicitly assume that $\beta \le -2$ unless otherwise stated.

A consequence of this long term correlation error when neg-β noise is present is that one cannot recover the true aging over T below the level of the long term random error in a single measurement over T. One must ensemble average over many data sets to determine the true aging below the level of the long term noise error. For real devices, such as frequency standards or oscillators, an ensemble average over many devices is not generally physically realizable. This is because aging coefficients generally vary significantly from device to device, and one cannot substitute a time average over a single device for the ensemble average unless **A** is time invariant over times much greater than T.
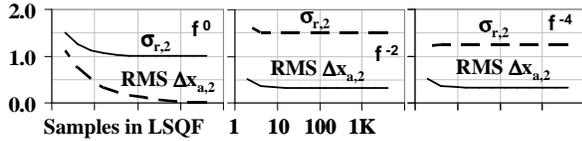
Figure 3.   Errors in LSQF for $x_{a,2}(\mathbf{A},t)=a_o+a_1 t$.

Figure 3 shows some further properties of LSQF solutions in the presence of neg-β noise for a 1st order aging polynomial (M=2) from a numerical simulation over a total of 4K points. Here, we've plotted $\sigma_{r,M}$ and RMS $\Delta x_{a,M}$ verses the number of samples in an LSQF used to determine **A**. For this simulation, $\sigma_{r,M}$ and RMS $\Delta x_{a,M}$ were calculated   by averaging over: (a) LSQF solutions over 2 to 2K samples spaced over the 4K points (sample separations from 2K to 2 points) and all possible initial point offsets, and (b) all 4K points regardless of the number of samples in the LSQF. There are two important conclusions that one can draw from the figure. First, $\sigma_{r,M}$ is relatively insensitive to the number of samples in the LSQF (especially for neg-β noise). Second, for neg-β noise, the RMS $\Delta x_{a,M}$ is greater than $\sigma_{r,M}$. Thus, when neg-β noise is present, error estimates of the aging coefficients based on the LQSF residual do not truly reflect the actual error in the determination of **A**. This is a well-known property of LSQF solutions when correlated errors are present.

One can also theoretically investigate the properties of LQSF solutions in the presence of neg-β noise in terms of the random noise autocorrelation function $R_x(\tau)=E\{x(t+\tau)x(t)\}$ using Green's functions. However, this does not lead us to our goal of obtaining a physical interpretation of $\sigma_{x,M}^2$ and closed form solutions for $\sigma_{x,M}^2$ in terms or $R_x(\tau)$ become increasingly complex as M increases. A simpler approach for obtaining a general M solution that will lead us to our goal is the topic of our next section.
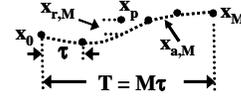
IV. THE ZERO DEGREE OF FREEDOM APPROXIMATION

Figure 4.   Zero Degree of Freedom Problem (M>0).

One can use the fact that $\sigma_{r,M}^2$ is relatively insensitive to the number of samples in the LSQF to simplify the calculation of $\sigma_{r,M}^2$ by using a zero-degree-of-freedom (N=M) approximation.   The zero-degree-of-freedom problem is shown in Figure 4. Here, we have M+1 points over T = Mτ given by

$$x_n = x(t_o + n\tau) \quad (n = 0 \text{ to } M) \tag{11}$$

in which we determine $x_{a,M}(t,\mathbf{A})$ from M points by excluding the pth point. **A** can then be determined simply by inverting the matrix equation

$$x_n = \sum_{\substack{m=0 \\ m \ne p}}^{M} a_m (n\tau)^m \tag{12}$$

However, because the zero degree of freedom solution exactly passes through the points x_n $(n \ne p)$, one can determine $x_{r,M}(t+p\tau)$ directly from the x_n without actually solving for **A** or removing $x_{a,M}(t,\mathbf{A})$. The result, which is a generalization of Newton's Forward Difference Formula [10], is given by

$$x_{r,M}(t_o + p\tau) = \Delta(\tau)^M x(t_o) / C(M,p) \tag{13}$$

and is derived in Appendix A. (It is noted that Newton's original Forward Difference Formula is (13) with p=M.) Using (13) one can write

$$\sigma_{r,M}^2(t_o + p\tau) = \lambda_M \, \sigma_{x,M}^2(\tau) / C(M,p)^2 \tag{14}$$

for this restricted problem.

Since we've shown that the value of $\sigma_{r,M}^2$ is relatively insensitive to the number of points (especially for neg-β noise), (14) can be used as an approximation for the general N-point LSFQ solution for $\sigma_{r,M}^2$ at $t_o + p\tau$. This is our physical interpretation of $\sigma_{x,M}^2$, that $\sigma_{x,M}^2$ is an approximate measure of $\sigma_{r,M}^2$ for any number of points in the LSQF solution. Given this interpretation, our rationale for choosing the order M of $\sigma_{x,M}^2$ is then that it should match the order of $x_{a,M}(t,\mathbf{A})$ which is specified by the problem, or the lowest

order of $x_{a,M}(t,\mathbf{A})$ which fits the actual causal behavior of $x(t)$ over the time T (by using Occam's razor) when $x_{a,M}(t,\mathbf{A})$ is not specified by the problem.

V. INTERPRETING $\sigma_{x,M}^2$ AS A MEASURE OF RESIDUAL ERROR AFTER AGING REMOVAL

TABLE II.   AGING REMOVAL ASSOCIATED WITH TYPICAL $\sigma_{x,M}^2$

M	Associated Names	$\sigma_{x,M}^2$	Aging Removed
0	Standard, Sample	$E\{[x(t)-\mu]^2\}$	None
1	TIErms (ITU-T G.810)	TIErms^2/2	a_0
2	$\sigma_x^2(\tau)$, $\sigma_y^2(\tau)$, TVAR	$\tau^2\sigma_y^2(\tau)/3$	$a_0 + a_1 t$
3	$_H\sigma_y^2(\tau)$ = Hadamard	$3\tau^2{}_H\sigma_y^2(\tau)/10$	$a_0 + a_1 t + a_2 t^2$

No removal (Degenerate Case)
$$x_{r,0} = \Delta(\tau)^0 x_0 \equiv x_0 - \mu$$
$$\cdots \quad \bullet \quad x_0$$
$$x_r$$
$$\cdots \quad —\quad \mu$$
$$\sigma_{r,0}^2 = \text{Standard}$$

0th order aging (time offset) removal
$$x_{r,1} = \Delta(\tau)^1 x_0$$
$$x_1 = \cdots \ x_r$$
$$x_0 \bullet \cdots\cdots \bullet$$
$$x_a$$
$$\sigma_{r,1}^2 = \text{TIErms}^2$$

1st order aging (time & freq offset) removal
$$x_{r,2} = 0.5\Delta(\tau)^2 x_0$$
$$x_1$$
$$x_0 \bullet \bullet$$
$$x_a \quad x_2$$
$$\sigma_{r,2}^2 = 0.5\tau^2\sigma_y^2(\tau)$$
$$p \cong 0.5M \text{ for min } \sigma_{r,M}^2$$

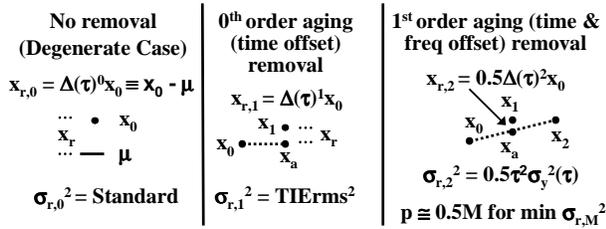Figure 5.   Physical interpretation of typical $\sigma_{x,M}^2$.

Table II lists the aging removal associated with typical $\sigma_{x,M}^2$, and Figure 5 shows the physical interpretation of some of these $\sigma_{x,M}^2$ as aging removed residuals. Thus: (a) the standard variance is associated with no aging removal, (b) TIErms is associated with 0th order aging or time offset removal, (c) the Allan variance is associated with 1st order aging or time and frequency offset removal, and (d) the Hadamard variance is associated with time and frequency offset and frequency drift removal. A corollary of our interpretation is that aging will contaminate $\sigma_{x,M}^2$ when the $(M-1)^{th}$ aging model is not sufficient to account for the actual aging over T. This explains the sensitivity of Allan based variances to frequency drift and the insensitivity of Hadamard based variances to such frequency drift.

One can extend the definition of $\sigma_{x,M}^2$ to include finite sample statistics, such as those using overlapping, modified, and total averaging techniques [2]. The following is the overlapping statistic defined for a sample data set over an interval $T_d > T$ in the limit of N→∞ but T=Ndt finite

$$\sigma_{x,M,ov}^2 = \lambda_M^{-1}(T_d - M\tau)^{-1}\int_0^{T_d-M\tau} dt\left[\Delta(\tau)^M x(t)\right]^2 \quad (15)$$

The above is just the finite time average of $\left[\Delta(\tau)^M x(t)\right]^2$. Under the same conditions, the modified statistic can be written as

$$\sigma_{x,M,mod}^2 = \lambda_M^{-1}(T_d - M\tau)^{-1}\int_0^{T_d-M\tau} dt[\Delta(\tau)^M\overline{x}_\tau(t)]^2 \quad (16)$$

where

$$\overline{x}_\tau(t) = \tau^{-1}\int_t^{t+\tau} dt\ x(t) \quad (17)$$

is the time average of $x(t)$ over τ. Similarly, one can generate a total average statistic for $\sigma_{x,M}^2$.

We note that (15) and (16) are statistical measures of the residual error after an Mth order aging polynomial is removed over the interval T (not T_d) and that the implicit fit value of $\mathbf{A}$ changes as t is slid over the integrals. It is also noted that that the dummy variable t in the integrals (15) and (16) replaces t_o (7). Thus, these forms enable $\sigma_{r,M}^2$ to be determined for an aging model that is valid over the interval T but not T_d. Finally, it is again noted that $\sigma_{x,M}^2$ and its statistics allow the calculation of $\sigma_{r,M}^2$ without actually removing the aging from the x(t) data.

One can define a spectral integral for $\sigma_{x,M}^2$ as [11]

$$\sigma_{x,M}^2 = 2\int_0^\infty L_x(f)|H_s(f)|^2 K_{x,M}(f)df \quad (18)$$

where $H_s(f)$ in (18) is a system response function that replaces the simple high frequency cut-off f_h used in previous spectral formulations [2]. $H_s(f)$ is determined by examining the specific spectral properties of the system under consideration, and is a more exact description of the spectral behavior of a system than just f_h. The advantage of using $H_s(f)$ in (18) is one can show that $H_s(f)$ has a kth order zero at f=0 for many systems [11]. This zero improves the convergence properties of low order $\sigma_{x,M}^2$ for many systems, and often allows low-order $\sigma_{x,M}^2$ to converge in the presence of neg-β noise where the f_h formulation does not.

In (18), $K_{x,M}(f)$ is an x-kernel that is the spectral equivalent of the time-domain operations which generate $\sigma_{x,M}^2$ from x(t) [11]. Using the Fourier transformed properties of $\Delta(\tau)$ and $\Theta(\tau)$, one can show that $K_{x,M}(f)$ is given by

$$K_{x,M}(f) = \lambda_M^{-1}2^{2M}\sin^{2M}(\pi f\tau) \quad (19)$$

which is well-known for lower order $\sigma_{x,M}^2$ [1][2].

From (19), one can see that $K_{x,M}(f)$ is proportional to f^{2M} for f<<1, leading to the well-known highpass (HP) filtering of $L_x(f)$ that increases in order as M increases. This HP filtering gives the higher order $\sigma_{x,M}^2$ their well-known improved

convergence properties in the presence of neg-β noise. An important consequence of this is that the extraction of aging or information from data as represented by the removal of $x_{a,M}(t,\mathbf{A})$ from x(t) by fitting to data over T will HP filter the residual noise. This is based on the presumption that the spectral properties of $\sigma_{r,M}^2$ for any N can be approximated by those of $\sigma_{x,M}^2$. However, an outstanding question that remains is whether this f^{2M} behavior for f<<1 is exactly replicated by x-kernels of general N-point $\sigma_{r,M}^2$. It is also well-known that the statistics of $\sigma_{x,M}^2$ (for M>0) have f<<1 properties that are similar to $K_{x,M}(f)$ [2][11][12]. The statistic of $\sigma_{x,0}^2$, the sample variance, is a special case because of the presence of μ [11].

A final consequence of our interpretation is that one cannot not <u>arbitrarily</u> change the order of $\sigma_{x,M}^2$ to avoid a divergence problem due to the presence of neg-β noise, if one wants $\sigma_{x,M}^2$ to represent a measure of the appropriate $\sigma_{r,M}^2$. This means that one must find other ways of dealing with such divergences. This is the subject of our next section.

VI. DEALING WITH DIVERGENCES WHEN NEG-β NOISE IS PRESENT

It is well-known that $\sigma_{x,0}^2$ and $\sigma_{x,1}^2$ are susceptible to divergences in the presence of neg-β noise. It has therefore been argued that these variances should not be used. However, from our prior discussion, $\sigma_{x,0}^2$ is the correct measure of $\sigma_{r,0}^2$ (no aging removal), and $\sigma_{x,1}^2$ is the correct measure of $\sigma_{r,1}^2$ (0th order aging or time-offset removal). Thus, if one wants these difference variances to represent the appropriate measures of $\sigma_{r,M}^2$, one cannot change them without changing the statement of the problem (or system specification) under consideration.

The following example illustrates this point. Consider the specification of the output time error for a coherent frequency synthesizer (evaluated without its frequency source). Should this synthesizer produce a frequency offset error (or its time drift error equivalent), it is considered a performance degradation problem. Thus $\sigma_{x,2}^2$ is not an appropriate error measure for specifying the residual time error in such a synthesizer. This is because $\sigma_{x,2}^2$ removes the average frequency offset over T in calculating the residual time error. Thus, depending on whether a time offset error is also of concern or not, ether $\sigma_{x,0}^2$ or $\sigma_{x,1}^2$ are the appropriate residual time error measures for such a synthesizer. $\sigma_{x,2}^2$ is more appropriate for specifying the residual error of a free running frequency source when the actual frequency is not important, but only the source's frequency stability properties.

Divergences due to neg-β noise are in fact indicators of real design, specification, or analysis problems that occur when neg-β noise is present and should be addressed by fixing

these problems, not by <u>arbitrarily</u> change the variance without further consideration. Changing the variance may be appropriate, but there should be a logical rationale supporting the change. The following discussion will illustrate this point.

Such divergences can be categorized as non-essential and essential ones. We define a non-essential divergence as one which only requires an analysis change. That is, the divergence is due to an analysis error and is not inherent in the design or specification of the system in question. One example of such an analysis error is the use of the improper system response function $H_s(f)$ for the system under consideration (or f_h instead $H_s(f)$). Thus the identification of the appropriate $H_s(f)$ will add highpass filtering that enables variance used to converge if the divergence is non-essential. If the divergence remains (and the correct variance for the problem is being used), then the divergence is an essential one which requires a change in the system design or spec.
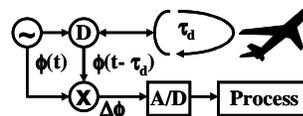

Figure 6. Coherent radar phase response.

Another example of a non-essential divergence is the inadvertent choice of the wrong variance for the system. This is demonstrated by the coherent radar example shown in Figure 6. A coherent radar, has a "delay" $|H_s(f)|^2$ given by [11]

$$|H_s(f)|^2 = 4\sin^2(\pi f\tau_d) \qquad (20)$$

which is proportional to f^2 for f<<1. In radar analysis, the standard variance is often identified as the theoretical phase error ($\phi(t) = \omega_0 x(t)$ [2]) measure to be used in such a system. This variance diverges when f^{-3} noise is present, because $K_{x,0}(f) = 1$ and $|H_s(f)|^2 \propto f^2$ are not sufficient to overcome the divergent behavior of $L_x(f) \propto f^{-3}$ in (18). To fix this problem, radar analysts use a heuristic spectral approximation of the standard variance (which is really a bandpass variance [11]) given by

$$\sigma_{x,bp}^2 = 2\int_{1/T_c}^{f_h} L_x(f)df \qquad (21)$$

where T_c is the interval over which data blocks are coherently processed. (21) solves the convergence problem, but there is no rigorous justification for replacing the standard variance with a bandpass variance.

The rigorous justification is provided by noting that the standard variance is not the correct residual error measure for this problem. The proper residual error measure is the sample variance over T_c, when a single average value is extracted

from the data, or a higher order statistic when more complex information is extracted. Thus, the x-kernel of the variance must at least behave as f^2 for f<<1, and the correct variance converges in the presence of f^{-3} noise.
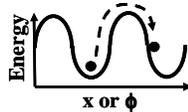

Figure 7.   Cycle Slip in PLL.

An example of an essential physical divergence is a 1st order phase locked loop (PLL) in the presence of f^{-3} noise. For a 1st order PLL, $|H_s(f)|^2 \propto f^2$ when f<<1 [11]. Thus, the standard variance again diverges when f^{-3} is present for the linear (small x(t)) PLL model. This divergence is essential and is an indicator of cycle slips which occur in the actual (non-linear) PLL as shown in Figure 7.

This physical problem can be properly mitigated using two approaches. The first approach is to change the hardware design so there is a 2nd order PLL in the system. For a 2nd order PLL, $|H_s(f)|^2 \propto f^4$ [11], and this fourth order zero is sufficient to allow the standard variance to converge in the linear model and to remove cycle slips in the actual system. Note that no change in variance is necessary with this approach.

The second approach is to allow cycles slips because the cumulative time or phase error is not of concern (hardware design is not changed). However, one should add a mean time to cycle slip to the specification to ensure that such cycle slips are rare enough to avoid becoming a nuisance, and one must specify that the residual time or phase error variance is to be evaluated using a sample variance measured over a specified interval with any data containing cycle slips excluded. We finally note that this second approach changes the variance but with a rationale to support the change.

SUMMARY AND CONCLUSIONS

We summarize the conclusions of this paper as follows.

1) $\sigma_{x,M}^2(\tau)$ can be interpreted as an approximate measure of the residual error $\sigma_{r,M}^2$ when an $(M-1)^{th}$ order polynomial $x_{a,M}(t,\mathbf{A})$ is fit to data over $T = M\tau$. This explains the sensitivity of Allan based variances to frequency drift and the insensitivity of Hadamard based variances to such frequency drift. When using this interpretation, $\sigma_{x,M}^2$ should be matched to the equivalent $(M-1)^{th}$ order polynomial removed from the data.

2) Removing a polynomial $x_{a,M}(\mathbf{A},t)$ from data by least squares fitting spectrally highpass (HP) filters the residual noise. Furthermore, the order of the HP filter is determined by the order of the removed polynomial. This conclusion is based on the use of $\sigma_{x,M}^2$ as an approximate measure of $\sigma_{r,M}^2$.

However, an outstanding question that remains is whether the exact x-kernels of general N-point $\sigma_{r,M}^2$ replicate the f^{2M} behavior for f<<1 that is the basis of this conclusion.

3) One type of variance is not appropriate for all problems.

4) One should not <u>arbitrarily</u> change to a different variance order simply to avoid a divergence problem due to negative power law noise. There must be a rationale behind any variance change.

5) Such a divergence is a signal to fix real problems in the system design, specification, or analysis. The divergence may be an essential one, requiring a design or spec change, or a non-essential one, requiring only an analysis change.

As a final note, the results of this paper can be applied to variances other variables such as $\phi(t)$ and y(t).

APPENDIX A. GENERALIZATION OF NEWTON'S FINITE DIFFERENCE FORMULA

An important property of the $\Delta(\tau)^M$ operator is

$$\Delta(\tau)^M t^{M'} = 0 \quad \text{for } M' < M \qquad (A.1)$$

Thus, $\Delta(\tau)^M x(t)$ filters out $(M-1)^{th}$ order polynomial behavior in x(t). (A.1) is easy to demonstrate by applying $\Delta(\tau)$ to $t^{M'}$ M times. To prove (13), we note that we can write (5) as

$$\Delta(\tau)^M x(t_o) = \sum_{m \neq p}^{M} \frac{(-1)^{M-m} M!}{m!(M-m)!} \Theta(\tau)^m x(t_o)$$
$$+ \frac{(-1)^{M-p} M!}{p!(M-p)!} \Theta(\tau)^p x(t_o) \qquad (A.2)$$

Since the zero degree of freedom solution for $x_{a,M}(t)$ passes through the data points $\Theta(\tau)^m x(t_o)$ for m≠p, we can substitute $\Theta(\tau)^m x_{a,M}(t_o)$ for $\Theta(\tau)^m x(t_o)$ in the first term of the right of (A.2) and obtain

$$\Delta(\tau)^M x(t_o) = \sum_{m \neq p}^{M} \frac{(-1)^{M-m} M!}{m!(M-m)!} \Theta(\tau)^m x_{a,M}(t_o)$$
$$+ \frac{(-1)^{M-p} M!}{p!(M-p)!} \Theta(\tau)^p x(t_o) \qquad (A.3)$$

Applying the Binomial Theorem, we can write (A.3) as

$$\Delta(\tau)^M x(t_o) = \Delta(\tau)^M x_{a,M}(t_o)$$

$$+ \frac{(-1)^{M-p} M!}{p!(M-p)!} \Theta(\tau)^p (x(t_o) - x_{a,M}(t_o)) \qquad \text{(A.4)}$$

Finally, we note from (A.1) that $\Delta(\tau)^M x_{a,M}(t_o) = 0$, so (A.4) can be written, using (8), as

$$\Delta(\tau)^M x(t_o) = \frac{(-1)^{M-p} M!}{p!(M-p)!} x_{r,M}(t_o + p\tau) \qquad \text{(A.7)}$$

This is our desired result (13).

REFERENCES

[1] ITU-T Recommendation G.810 (08/96), Definitions and Terminology for Synchronization Networks. ITU, 1996.

[2] IEEE 1139-1999: Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology—Random Instabilities, IEEE, 1999.

[3] G. E. Box, G. M. Jenkins, and G. C. Reinsel, Time Series Analysis, 3rd ed, Prentice-Hall, 1994.

[4] [Hutsell, 1995] S.T. Hutsell, "Relating the Hadamard Variance to MCS Kalman Filter Clock Estimation," Proc. 1995 PTTI Meeting, Dec. 1995.

[5] [Rutman, 1978] Rutman, J., "Characterization of Phase and Frequency Instabilities in Precision Frequency Sources: Fifteen Years of Progress," Proceedings of the IEEE, vol. 66, pp. 1048–1075, Sept. 1978.

[6] IEEE Std 181-2003, IEEE Standard on Transitions, Pulses, and Related Waveforms, IEEE 2003.

[7] IEEE Std 1057-1994 (R2001), IEEE Standard for Digitizing Waveform Recorders, IEEE 2001.

[8] IEEE Standard for Terminology and Test Methods for Analog-to-Digital Converters, IEEE Std 1241-2000, IEEE 2000.

[9] IEEE Std 1193™-2003, IEEE Guide for Measurement of Environmental Sensitivities of Standard Frequency Generators, IEEE 2003.

[10] Abramowitz, M. and Stegun, I. A. (Eds.). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing. New York: Dover, p. 880, 1972.

[11] Victor S. Reinhardt, The Properties of Time and Phase Variances in the Presence of Power Law Noise for Various Systems, Proceedings of the 2006 IEEE International Frequency Control Symposium, June 5-7, 2006, Miami, Florida.

[12] Vernotte, F., Lantz, E., Groslambert, J., and Gagnepain, J. J., "Oscillator Noise Analysis: Multi-Variance Measurement," IEEE Transactions on Instrumentation and Measurement, vol. IM-42, no. 2, pp. 342–350, Apr. 1993.