Preprint, Proceedings of the Precise Time and Time Interval Systems and Applications Meeting, Albuquerque, New Mexico, November 17-19, 2009.

NEGATIVE POWER LAW NOISE, REALITY VS MYTH

VICTOR S. REINHARDT, RAYTHEON SPACE AND AIRBORNE SYSTEMS, EL SEGUNDO, CA, USA

Abstract

In this paper, it is shown that two popular conceptions about the behavior of negative power law (neg-p) noise—that is, noise with a PSD $L_p(f) \propto |f|^p$ for p<0—are based on myth and that the reality is quite different. The first myth is that one can "fix" a neg-p divergence problem in a variance like a standard or N-sample variance simply by replacing it with an Allan or Hadamard variance without further action. The paper will show that each type of variance has a different interpretation as an error measure and that such arbitrary swapping merely masks the true problem. In the process, we will show that such variance divergences are true indicators of severe system or modeling problems that must be physically addressed not ignored. The second myth is that one can use ensemble based statistical estimation techniques like least squares and Kalman filters to properly estimate polynomial deterministic behavior in data containing non-highpass filtered neg-p noise. It is demonstrated that such noise can generate highly anomalous fitting results because non-highpass filtered neg-p noise is both infinitely correlated and non-ergodic. Thus, non-p noise is shown to act more like systematic error than conventional noise in such cases.

I. INTRODUCTION

This paper will show that two popular conceptions in dealing with negative power law noise (neg-p) noise are based on myth and that the reality is quite different. By neg-p noise, we mean noise with a single sideband (SSB) power spectral density (PSD) $L_p(f) \propto |f|^p$ for p<0 [1,2]. This paper is not questioning the reality that higher order Δ-variances [3], like Allan [1] and Hadamard variances [4] are convergent measures of neg-p noise [1,4]. What the paper will show is that it is myth that one can "fix" neg-p divergence problems in common variances like standard and sample variances [5] simply by replacing them with Δ-variances without further action. We will show that each type of variance is a statistical answer to a different type of error question and that arbitrarily changing variances is misleading in that it doesn't fix the divergent answer to the original question. Furthermore, we will show that such variance divergences have physical significance—that they are valid indicators of real problems that must be fixed by changing the system or the question being asked, not mathematical artifacts to be ignored.

A second myth we will address is that one can use ensemble based statistical estimation theory, such as least squares [5] and Kalman [7] filters, on data containing neg-p noise to properly estimate true or deterministic polynomial behavior also contained in the data, unless the neg-p noise is sufficiently highpass filtered [8-11]. We will demonstrate that fitting results in such cases cannot separate the true behavior from much of the noise, because non-highpass filtered neg-p noise is both infinitely correlated and non-ergodic (ensemble averages are not equal to time averages).

II. MYTH 1: ONE CAN ARBITRARILY SWAP VARIANCES TO "FIX" NEG-P DIVERGENCE PROBLEMS

In this section, we will show that each type of variance is a statistical answer to a different type of error question. Thus arbitrarily swapping variances misleadingly changes the question and does not eliminate a divergent answer to the original question.

A. STATISTICAL ESTIMATION

Statistical error measures like variances are generally defined in the context of statistical estimation. Fig. 1 and Table I describe the truth model and variables we will use in discussing statistical estimation. This model applies to least squares fitting (LSQF) [5] and Kalman filters [6] in *a posteriori* form [12], as well as other similar statistical estimation techniques. We will briefly summarize this model here, and the reader is referred to [7-11] for more detail. In this model, $x(t)$ is general data variable (not necessarily the time error) whose samples $x(t_n)$ are collected over an interval T. t and t_n here are ideal continuous and discrete observation times and are considered error-free. $x(t)$ in our model is the sum of $x_c(t)$ the true or deterministic behavior and $x_r(t)$ the contaminating error or measurement noise. In the model, an unspecified estimation technique generates a "best" estimate of $x_c(t)$ by adjusting M-parameters a_m in a model function $x_{a,M}(t)$ based on some fit over $x(t_n)$. Note that we will use $x_{a,M}(t)$ both to describe the model function with adjustable parameters and the final fit depending on the context. We also note that $x_c(t)$, $x(t)$, and $x_{a,M}(t)$ can be functions of other time dependent variables, such as temperature and pressure [13]. For simplicity, these other variables are not shown.

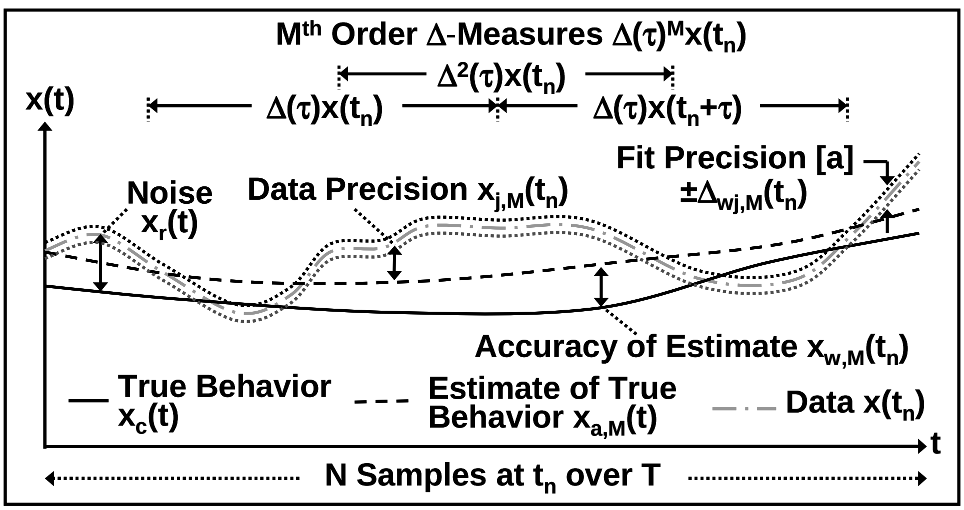


Fig. 1. Truth model and variables for statistical estimation.

Table I. Statistical Estimation Model & Error Measures

Basic variables
Measured data : $x(t_n) = x_c(t_n) + x_r(t_n)$ Sample times : t_n (over data period T)
True or deterministic behavior: $x_c(t)$ True noise: $x_r(t)$
M parameter model function and final estimate of $x_c(t)$: $x_{a,M}(t)$
(M-1)th order polynomial model function: $x_{poly,M}(t) = \Sigma_m a_m(t - t_0)^m$ [m = 0:M-1]
Basic Error Measures
True accuracy of fit: $x_{w,M}(t_n) = x_{a,M}(t_n) - x_c(t_n)$
Data precision: $x_{j,M}(t_n) = x(t_n) - x_{a,M}(t_n)$
M^{th} order Δ-measures (stability & precision under certain conditions): $\Delta(\tau)^M x(t_n)$
1^{st} forward difference: $\Delta(\tau) = x(t_n+\tau) - x(t_n)$
Variances
Point variance (Kalman) : $\Delta^2_\varsigma(t_n) = \mathcal{E}[x_\varsigma(t_n) - \mathcal{E}\{x_\varsigma(t_n)\}]^2$ $[\varsigma = w, M \text{ or } j, M]$
Average variance over T (LSQF) : $\sigma^2_\varsigma = \sum_n \xi_n \Delta^2_\varsigma(t_n)$ $[n = 1 : N]$
$\Delta^2_{\tau,M}(t_n) = \lambda_M^{-1}\mathcal{E}[\Delta(\tau)^M x(t_n)]^2$ $\sigma^2_{\tau,M} = \sum_n \xi_n \Delta^2_{\tau,M}(t_n)$ $\lambda_M = \sum_{m=0}^{M}\left(\frac{M!}{m!(M-m)!}\right)^2$
$\mathcal{E}$ = Ensemble average ξ_n = Data weighting over T
Derived Error Measures
Fit precision (point variance): $\Delta^2_{wj,M}(t_n) = \rho_d(t_n)\Delta^2_{j,M}(t_n)$
Fit Precision (average variance): $\sigma^2_{wj,M} = \rho_d \sigma^2_{j,M}$
ρ_d and $\rho_d(t_n)$ are theoretically calculated from $\rho_d(t_n) = \Delta^2_{w,M}(t_n)/\Delta^2_{j,M}(t_n)$ and
$\rho_d = \sigma^2_{w,M}/\sigma^2_{j,M}$ based on a specific error model and assuming no model error.

B. BASIC ERROR MEASURES

In Fig. 1 and Table I, we define the true accuracy of the fit at t_n as $x_{w,M}(t_n)$; that is, the accuracy is the difference between $x_{a,M}(t_n)$ and $x_c(t_n)$. $x_{w,M}(t_n)$ is of course *unobservable* from the data alone, since *a priori* knowledge of $x_c(t)$ is required to generate it. The basic *observable* error measure at t_n is the data precision $x_{j,M}(t_n)$ defined as the difference between $x(t_n)$ and $x_{a,M}(t_n)$, also given in Table I. From $x_{w,M}(t_n)$ and $x_{j,M}(t_n)$, one can form two types of theoretical variances (see Table I):

(a) $\Delta^2_{w,M}(t_n)$ and $\Delta^2_{j,M}(t_n)$ we will call point variances. These are generally used in a Kalman filter [6].

(b) $\sigma^2_{w,M}$ and $\sigma^2_{j,M}$ we will call average variances over T weighted by ξ_n. These are generally used in a LSQF [5].

Note that $\sigma^2_{w,M}$ is also called the standard variance and $\sigma^2_{j,1}$ is called the sample variance when ξ_n=1/(N-1) and the fit solution is the sample mean $x_{a,1}(t) = N^{-1}\Sigma_n x(t_n)$ [5]. Note also that the above are "theoretical" or ensemble variances formed by averaging over an ensemble of data sets [14], that is, by using an ensemble averaging operator $\mathcal{E}$ as opposed to an infinite time average operator <…> over a single ensemble member [14]. Finally note that $\sigma^2_{w,M}$ and $\sigma^2_{j,M}$ may have implicit t-dependence, because both $x_{a,M}(t_n)$ and $x_c(t)$ are not generally time invariant [9,10].

Another set of measures used to describe random error we will call Mth order Δ-measures $\Delta(\tau)^M x(t_n)$ [3]. These and their theoretical variances $\Delta^2_{\tau,M}(t_n)$ and $\sigma^2_{\tau,M}$ are defined in Table I [10,11]. These are measures of x(t) variations over the interval τ. We note that $\Delta^2_{\tau,M}(t_n)$ and $\sigma^2_{\tau,2}$ is related to the Allan variance of the time error [1], and $\sigma^2_{\tau,3}$ is related to the Hadamard variance of the time error [4,15].

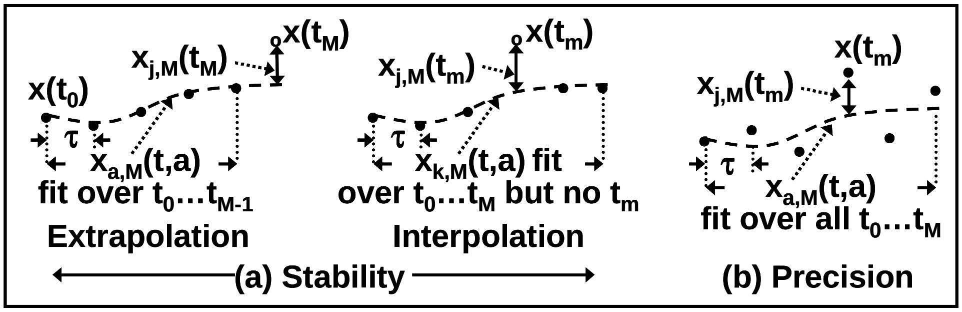


Fig. 2. Δ-measures as stability and precision measures

Δ-measures are generally interpreted as measures of Mth order stability [1,4,15]. To understand this interpretation, let us precisely define what we mean by Mth order stability. Consider Fig. 2(a). Here, we show M+1 data points $x(t_{m'})$ where we have passed a model function $x_{a,M}(t)$ exactly through M of the M+1 points, excluding the point at t_m. This is possible because there are M points and M adjustable parameters in $x_{a,M}(t)$, so there are zero degrees of freedom [5]. We then define the Mth order stability as the data precision $x_{j,M}(t_m)$ at the excluded point. Note from the figure that $x_{j,M}(t_m)$ can be either an extrapolation or interpolation error depending on t_m. What is important about this is one can show that

$$x_{j,M}(t_m) \propto \Delta(\tau)^M x(t_0) \tag{1}$$

when: (a) $x_{a,M}(t)$ is $x_{poly,M}(t)$ an (M-1)th order polynomial, and (b) the $t_{m'}$ are separated by the time interval τ [3]. Thus, Δ-variances are measures of such Mth order stability.

Fig. 2(b) shows the data precision when all M+1 $x(t_{m'})$ are used to determine $x_{poly,M}(t)$ with an unweighted LSQF. In this case, one can also show that (1) is true [7,8,16]. Thus, Mth order Δ-variances can also be considered data precision measures under these conditions. The proportionality constants relating $x_{j,M}(t_m)$ to $\Delta(\tau)^M x(t_0)$ for both the stability and precision have been published [7,8]. In the precision case, the published constant is derived semi-empirically [7], but we note that Charles A. Greenhall has provided the author with a totally analytical derivation of this constant [16]. Thus, the Allan variance can also be interpreted as a measure of data precision for M+1 $x(t_m)$ spaced by τ when a time and frequency offset are removed from the data by an unweighted LSQF [6,7]. Similarly, the Hadamard variance can also be interpreted as a measure of such data precision when a time and frequency offset <u>and</u> the frequency drift are removed from the data by an unweighted LSQF [6,7]. This explains the sensitivity of the Allan variance and insensitivity of the Hadamard variance to deterministic frequency drift, since such drift is an unmodeled error term for M = 2 but not for M = 3 [7,8].

C. THE FIT PRECISION (ERROR BARS)-A DERIVED ERROR MEASURE

From the data precision variances, one can generate what we will call the fit precision (deviate) or error bars $\Delta_{wj,M}(t_n)$ and $\sigma_{wj,M}$, as given in Table I [9-11]. These are statistical estimates of the accuracy based both on the observable data precision and the ratios $\rho_d(t_n)$ and ρ_d, which are theoretically calculated using a a *specific* noise model (and assuming that $x_{a,M}(t)$ would precisely reproduce $x_c(t)$ over T if no noise were present). For example, ρ_{do}=M/(N-M) is the ρ_d for uncorrelated or white $x_r(t_n)$ and an unweighted LSQF [5]. We note that this ρ_{do} does not apply when the $x_r(t_n)$ are correlated [7-11]. We will later show that this misuse of the white ρ_{do} is one of the sources of unexpected fitting results when neg-p noise is present.

D. THE NEG-P CONVERGENCE PROPERTIES OF VARIANCES

It is well-known that an average variance σ_ς^2 can be represented using the spectral integral [14,17]

$$\sigma_\varsigma^2 = \int_{-\infty}^{+\infty} df\, K_\varsigma(f)\,|H_s(f)|^2\, L_p(f) \quad [\zeta = w,M;\ j,M;\ \tau,M;\ \ldots] \tag{2}$$

Here, $H_s(f)$ is a response function that describes the noise filtering properties of the system, and $K_\zeta(f)$ is a spectral kernel that describes the $L_p(f)$ filtering properties of the variance in question independent of $H_s(f)$ [3,17]. It is well known that the Δ-variance kernel $K_{\tau,M}(f)$ has f^{2M} highpass (HP) filtering properties for |f| << 1 [1,3,4]. Less well known is the fact that the M^{th} order data precision kernel $K_{j,M}(f)$ has the same f^{2M} highpass (HP) filtering properties for |f| << 1 when $x_{a,M}(t)$ is an $(M\text{-}1)^{th}$ order polynomial $x_{poly,M}(t)$ [7,8]. This result is true for general fitting techniques given only minimal restrictions [7,8]. Fig. 3 shows this $K_{j,M}(f)$ HP behavior when both a weighted and unweighted LSQF are used as the fitting technique [7,8]. Thus both $\sigma_{j,M}^2$ and $\sigma_{\tau,M}^2$ are guaranteed to converge for neg-p noise when 2M ≥ |p| [3,5,6-8].

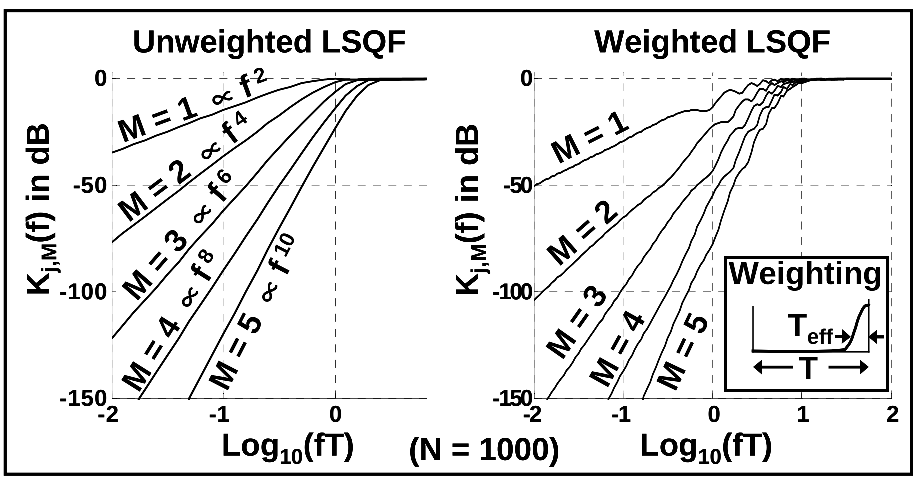


Fig. 3. $K_{j,M}(f)$ HP filtering behavior for $x_{poly,M}(t)$.

On the other hand, the accuracy kernel $K_{j,M}(f)$ has no HP filtering properties. (In fact, one can show that $K_{j,M}(f) + K_{w,M}(f) = 1$ for a LSQF [11].) Thus, $\sigma_{w,M}^2$ relies totally on the HP filtering properties of $|H_s(f)|^2$ for its convergence in the presence of neg-p noise [7-11]. Because of this, there is the obvious temptation to "fix" a neg-p accuracy variance divergence simply by replacing it with one of the convergent variances without further action. In the next section, we will show that this is improper.

E. EACH VARIANCE ADRESSES A DIFFERENT ERROR QUESTION

We have just shown that each type of variance addresses a different statistical error question:

(a) Accuracy: What is the error of the fit from the true behavior without noise or other error present?
(b) Data Precision: What are the data fluctuations from the fitted behavior?
(c) Fit Precision: What is the estimated fit accuracy based on the measured data and a noise model?
(d) Stability: What is the extrapolation or interpolation error to an additional point from a perfect M point polynomial fit?

From this, it is obvious that one cannot eliminate a divergence problem in one type of variance simply by arbitrarily replacing it with another type, since all this does is misleadingly change the question and leaves the divergence intact for the original question.

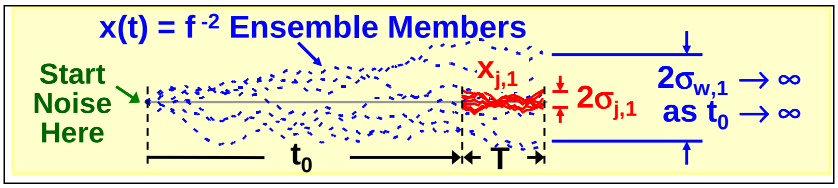


Fig. 4. A physical interpretation of a neg-p accuracy divergence.

This leaves the problem of how to interpret the physical meaning of such a neg-p accuracy divergence. Fig. 4 shows one such interpretation. Here, we show several data ensemble members where the data consists entirely of f^{-2} or random walk noise [2]. Note that the f^{-2} noise process is started at finite time (t = 0), which is called the non-stationary (NS) picture (to be discussed later) [2,11,14]. At a time t_0 after the noise process has started, we then perform a one-parameter LSQF on the data over T to generate our fit $x_{poly,1}(t) = a_0$. One can immediately see that something is amiss. We note that $\sigma_{j,1}$ is significantly less than $\sigma_{w,1}$. Thus a white noise based fit precision $\sigma_{wj,1}$ will severely underestimate the true accuracy $\sigma_{w,1}$ when N >> 1. Furthermore, one can easily show that $\sigma^2_{w,1} \to \infty$ as $t_0 \to \infty$, while $\sigma^2_{j,1}$ will remain finite [7-11]. This is a physical meaning of a neg-p accuracy variance infinity: *that the true accuracy of a fit will become severely inaccurate when t_0 is large* (which is the typical physical case). One can obviously see from this example that using the wrong $\sigma_{wj,1}$ and $\sigma_{\tau,M}$ to represent $\sigma_{w,1}$ in this case will just mask the problem, not fix it.

The proper response here would be: (a) to theoretically analyze ρ_d using the correct noise model [7-11], (b) to identify that the $\sigma_{w,1}$ infinity will occur before performing the experiment, and (c) to redesign the system ($H_s(f)$) and/or reformulate the question so that the accuracy infinity will not occur [7-11]. This last step often involves the introduction of periodic calibration [11].

For $|f|^{-3}$ noise, note that $\sigma^2_{j,1} \to \infty$ as $t_g \to \infty$ but that $\sigma^2_{j,2}$ will remain finite [7,8]. Thus, if one is interested in obtaining a finite data precision, the proper response to a $\sigma^2_{j,M}$ infinity is to change the estimation model M-order, not to arbitrarily switch to an Allan or Hadamard variance and leave the model function $x_{a,M}(t)$ untouched.

Finally, we note that $\sigma^2_{w,M} \propto \ln(f_h t_0)$ for $|f|^{-1}$ noise and $H_s(f)$ given by a low pass cut-off f_h [2]. Thus, even though the $|f|^{-1}$ contribution to $\sigma^2_{w,M}$ is strictly infinite as $t_0 \to \infty$, practically, this $|f|^{-1}$ contribution is often smaller than the white noise contribution, even when t_0 is the age of the universe [17].

III. MYTH 2: ONE CAN OBTAIN PROPER ESTIMATION RESULTS FROM (NON-HP FILTERED) NEG-P NOISE CONTAINING DATA

In this section, we will show that non-HP filtered neg-p behaves more like systematic error than conventional noise. Thus, estimation techniques like least squares and Kalman filters can generate severely anomalous results when such noise is present, since estimation techniques have difficulty separating systematic error from true behavior [5]. We will show that this systematic-like behavior is due to the non-ergodic and infinitely correlated behavior of non-HP filtered neg-p noise. In the next sections, will show that neg-p noise has these properties using the non-stationary (NS) and wide-sense stationary (WSS) pictures of a random process [2,10,11,14,18].

A. THE NS AND WSS PICTURES OF A RANDOM PROCESSES

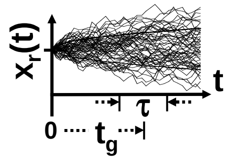
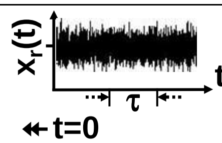
Two covariant representations or pictures are generally used when discussing a random process $x_r(t)$. The more inclusive but less familiar one is the non-stationary (NS) picture summarized in Table II [2,18]. In the NS picture, $x_r(t)$ is zero for $t<0$ or some other finite value. Because of this finite start time, the noise covariance or autocorrelation function of the process $R_r(t_g,\tau)$ [14,18] (the same here because we are assuming $\mathcal{E}x_r(t)=0$) has two time arguments: t_g the global or average time from the start of the noise process ($t=0$), and τ the difference or local time between the covariant x_r arguments [18]. The other less inclusive picture is the more familiar wide-sense stationary (WSS) one [2,12,14,19]. This picture is also summarized in Table II. Here, $x_r(t)$ is non-zero for all, t and $x_r(t)$ is assumed to be statistically time invariant so that the autocorrelation function is now given by $R_r(\tau)$. We note that $x_r(t)$ must also be statistically bounded for the process to be WSS, because many WSS theorems require such bounded behavior for their proof [19]. This bounded behavior is often ignored in descriptions of neg-p noise, which can lead to erroneous conclusions. Even when the underlying noise process is inherently WSS, note that one must use the NS $R_r(t_g,\tau)$ for small t_g, because of initial start-up transients [2].

As shown in Table II, an NS process has three different covariant spectral functions that are formed by the Fourier Transforms (FTs) of $R_r(t_g,\tau)$ with respect to various combinations of t_g and τ [18]. The Wigner-Ville function $W_r(t_g,f)$ in the table can be interpreted as a t_g dependent PSD and is the most physically intuitive for neg-p noise analysis. The Loève Spectrum $L_r(f_g,f)$ in the table, on the other hand, is useful for simplifying analytical expressions [2,9]. The Ambiguity Function $A_r(f_g,\tau)$ in the table is included here for completeness and is used in signal processing [18]. In the WSS picture, the SSB PSD $L_r(f)$ defined in Table II is the well-known spectral function formed by taking the complex Fourier transform of $R_r(\tau)$ with respect to τ [12,14,19]. Note that one can also use the double sideband PSD $S_r(f)$, as is common practice in time and frequency papers [1].

An important measure of the behavior of a random process is its correlation time τ_c, which is defined in Table II. Note that this definition is an extension of a WSS one [20] to include the NS picture in the limit of $t_g\to\infty$. τ_c is an important parameter in statistical estimation, because $N_i=T/\tau_c$ represents the number of statistically independent samples over T [20]. Thus, averaging over N samples reduces errors by some power of N_i (not N) when the relevant noise process is correlated, and only when $T >> \tau_c$ [10,11]. This will be very important in later discussions.

Finally, note from Table II that one can relate the NS picture to the WSS picture by letting $t_g\to\infty$ ($t_g\to\infty$ is equivalent to letting the $x_r(t)$ start time go to $-\infty$) [2,21]. Two important NS to WSS theorems based on this are also given in Table II [2,11]. These theorems will play a prominent role in understanding the true statistical properties of neg-p noise, as we will discuss in the next section.

Table II. Representations of a Random Process

Non-Stationary (NS) Picture **$x_r(t) = 0$ for $t<0$** t_g = Global or average time from start of noise process τ = Local or difference time $\mathcal{E}$ = Ensemble average	
Covariance or Correlation Fn ($x_r(t)$ real, $\mathcal{E}x_r(t) = 0$)	$R_r(t_g,\tau) = \mathcal{E}\, x_r(t_g+\tau/2)x_r(t_g-\tau/2)$ $R_r(t_g,\tau) = 0$ for $t_g < 0$ or $\lvert\tau\rvert > 2\, t_g$
Wigner-Ville Function	$W_r(t_g,f) = \mathcal{F}_{f,\tau} R_r(t_g,\tau)$
Loève Spectrum	$L_r(f_g,f) = \mathcal{F}_{f_g,t_g} W_r(t_g,f)$
Ambiguity Function	$A_p(f_g,\tau) = \mathcal{F}_{f_g,t_g} R_r(t_g,\tau)$
Correlation time	$\tau_c = \lim\limits_{t_g\to\infty} 0.5 R_r(t_g,0)^{-1} \int_{-\infty}^{+\infty} d\tau\, R_r(t_g,\tau)$
Number of statistically independent samples over T	$N_i = T/\tau_c$
(Complex) Fourier Transform	$V(f) = \mathcal{F}_{f,t} v(t) \equiv \int_{-\infty}^{+\infty} dt \exp(-j\omega t) v(t) \quad [\omega = 2\pi f]$
Wide-Sense Stationary (WSS) Picture **$x_r(t) \neq 0$ for all t**	
Covariance or Correlation Fn ($x_r(t)$ real, $\mathcal{E}x_r(t) = 0$)	$R_r(\tau) = \mathcal{E}\, x_r(t_g+\tau/2)x_r(t_g-\tau/2)$
(SSB) Power Spectral Density	$L_r(f) = \mathcal{F}_{f,\tau} R_r(\tau)$
NS → WSS Theorems [4,21]	
$R_r(\tau) = \lim\limits_{t_g\to\infty} R_r(t_g,\tau) \quad L_r(f) = \lim\limits_{t_g\to\infty} W_r(t_g,f) = \lim\limits_{f_g\to 0} j\omega_g L_r(f_g,f)$ $L_p(f_g,f)$ form of $L_r(f)$ derived from the Laplace Final Value Theorem.	

B. THE STATISTICAL PROPERTIES OF NEG-P NOISE

Table III. Properties of Non-highpass Filtered Neg-p Noise

NS Picture: $R_p(t_g,\tau) < \infty$ for $t_g < \infty$	$R_p(t_g,\tau) = \infty$ for $t_g \to \infty$
$W_p(t_g,f) < \infty$ for all $f, t_g < \infty$	Note bandlimiting needed to make $W_p(t_g,\tau)$ finite for $p\geq-1$
WSS Picture: $R_p(\tau) = \lim\limits_{t_g\to\infty} R_p(t_g,\tau) = \infty$	WSS $R_p(\tau)$ is undefinable for all τ.
$L_p(f) = \lim\limits_{t_g\to\infty} W_p(t_g,f) \propto \lvert f\rvert^p$	$R_p(\tau)$ not needed to define $L_p(f)$
Neg-p noise has an **infinite correlation time:** $\tau_c = \infty$	
Neg-p noise is **inherently non-ergodic:** $\mathcal{E}x_p(t) \neq \langle x_p(t)\rangle_{T\to\infty}$.	

One can show that non-HP filtered neg-p noise has the basic properties listed in Table III [2]. Note that the

WSS $R_p(\tau)$ is infinite for all τ, because $R_p(\tau) = \lim_{t_g \to \infty} R_p(t_g, \tau)$ as given in Table II. Thus the WSS $R_p(\tau)$ is strictly indefinable. However, because $L_p(f) = \lim_{t_g \to \infty} W_p(t_g, f)$ and this limit is well-behaved for $f \neq 0$, one can properly define the WSS $L_p(f)$ for $f \neq 0$ without the use of the WSS $R_p(\tau)$. Thus, one can interpret equations such as (2) as the $t_g \to \infty$ limit of the NS picture and properly apply them to neg-p problems.

Another important property of non-HP filtered neg-p noise listed in

Table III is that its τ_c is infinite. This means that N_i is effectively zero for all T, which is one of the factors that leads to the anomalous neg-p fitting behavior that we will discuss later.

A very important (and not well-known property) of non-HP filtered neg-p noise is that it is intrinsically *non-ergodic*, that is the infinite time average $< x_p(t)>_{T\to\infty}$ is not equal to the ensemble average $\mathcal{E}x_p(t)$ for such noise [10,11,14,22]. This is a consequence of a theorem stating that an NS random process $x_p(t)$ is ergodic if and only if $R_p(t_g,\tau)$ is bounded for all t_g (including ∞) and the last $x_p(t)$ point in $<x_p(t)>_T$ becomes decorrelated with $<x_p(t)>_T$ as $T\to\infty$ [22]. This decorrelation property can be shown to imply that τ_c must be finite. This is another factor that leads to the anomalous fitting behavior of neg-p noise, which we will now discuss.

C. ANOMALOUS FITTING BEHAVIOR IN FINITE DATA SETS

Now let us investigate how the non-ergodicity and infinite τ_c of neg-p noise effects practical implementations of fitting techniques, such as least squares and Kalman filters. Fig. 5 shows a simulated unweighted LSQF for both p = 0 and p = -3 noise when: (a) both the true behavior $x_c(t)$ and the model function $x_{a,M}(t)$ are 2^{nd} order polynomials (M=3), (b) $H_s(f)$ is an ideal Nyquist LP filter, and (c) the uncorrelated ρ_{do} is used to predict the error bars $x_{a,M}(t)\pm\sigma_{wj,M}$ (which are so small in the figures that they appear coalesced with $x_{a,M}(t)$). In the p = 0 case shown on the left of Fig. 5, note that the fit behaves as ensemble-based white-noise fitting theory predicts; that is, $x_c(t)$ and $x_{a,M}(t)$ fall on top of each other for the large N used, and $\sigma_{wj,M}$ properly predicts $\sigma_{w,M}$.

For the p = -3 case on the right of the figure, however, note that $x_{a,M}(t)$ significantly deviates from $x_c(t)$ while the white-noise based error bars do not properly predict this deviation.

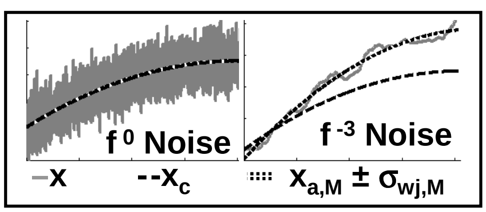


Fig. 5. Unweighted LSQF with p = 0 and p = -3 (N=2048).

What is happening here is that the particular $x_r(t_n)$ ensemble member in this example has behavior that is 2^{nd} order polynomial-like and thus substantially correlated with $x_c(t)$ so that LSQF cannot separate this correlated noise component from the true behavior of $x_c(t)$ [5]. This is what generates the anomalous fitting results; the fit interprets the correlated noise as being part of $x_c(t)$. Such anomalous behavior is well-known as result of correlated systematic error [5] and is a specific example of a more general principal—that linearly dependent variables cannot be separated by *any* solution technique, because the determinant of the solution matrix goes to zero [24]. This systematic-like behavior in the neg-p noise case is a direct result of its infinitely correlated and non-ergodic nature. Thus, $\mathcal{E}$-averaged theory predictions do not represent the behavior of individual ensemble members over the data collection interval.

An important consequence of the above is that noise whitening, a procedure meant to determine the true structure of $x_c(t)$ by increasing M in the model function $x_{a,M}(t)$ until the residuals $x_{j,M}(t_n)$ are uncorrelated [5], will <u>not</u> properly identify $x_c(t)$ when non-HP filtered neg-p noise is present. That is, the truth model for such noise whitening is *uncorrelated* noise plus true behavior, and neg-p noise is highly uncorrelated. Note that p = -1 noise can be a marginal case when white noise is also present, because of the slow growth of this noise as t_g becomes large [17].

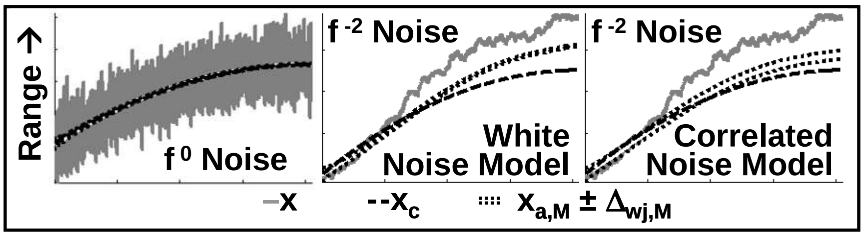


Fig. 6. Anomalous neg-p behavior in a Kalman filter.

Fig. 6 shows that Kalman filters also exhibit such anomalous neg-p behavior when neg-p noise is present. Here, simulation results are shown for both p = 0 and p = -2 noise (measurement noise not process noise [6,12]) when $x_c(t)$ and $x_{a,M}(t)$ are again both quadratic polynomials. In the middle graph, where p = -2 and an uncorrelated noise model is used [4], one again sees the characteristic veering off of $x_{a,M}(t)$ from $x_c(t)$ and the gross underestimation of the true errors by the error bars. What is even more interesting is the right graph. Here again p = -2, but the Kalman filter is augmented with a random walk measurement noise model, which is supposed to correct for such anomalous behavior [6]. One can see that $x_{a,M}(t)$ is closer to $x_c(t)$ and the error bars do a better job of estimating the true error. However, there still are substantial deviations of $x_{a,M}(t)$ from $x_c(t)$ and the error bars still underestimate the true error. The culprit here is the non-ergodic-like behavior of the neg-p noise and the mimicking of the true behavior by the neg-p noise. It is expected that p = -3 noise would exhibit even more significant anomalous behavior with an augmented Kalman filter, because p = -3 noise looks like a slowly changing random drift [2]. However, the author has not demonstrated this yet.

D. ERGODICITY, τ_C, AND PROPER FITTING BEHAVIOR

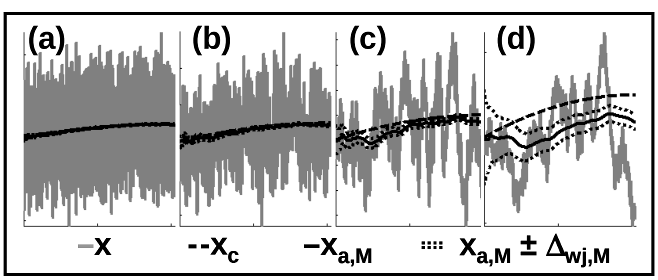


Fig. 7. Kalman filter simulations for a correlated Gauss-Markov process:
(a) T/τ_c=2000, (b) T/τ_c=200, (c) T/τ_c=20, (d) T/τ_c=2.

Fig. 7 illustrates that neg-p-like anomalous fitting behavior also occurs in ergodic WSS but correlated processes when one does not have $T/\tau_c >> 1$. Shown here is a (non-augmented) Kalman simulation using stationary Gauss-Markov noise (single pole lowpass filtered white noise [12,14]). This noise is both ergodic and WSS but has a τ_c related to the reciprocal of the lowpass knee frequency of the noise filter.

One can observe from the figure that the Kalman results behave as theoretically expected for $T/\tau_c >> 1$ but become more and more anomalous as T/τ_c approaches 1. This occurs because the single noise ensemble members averaged over T here do not behave like their ensemble averaged counterparts when T approaches τ_c. That is, the noise is not *ergodic-like* over T ($\mathcal{E}\ldots \neq <\ldots>_T$) when we don't have $T/\tau_c >>1$ [11]. One can see here that *strict ergodicity*, $\mathcal{E}\ldots = <\ldots>_{T\to\infty}$ [14,22] does not guarantee such ergodic-like behavior over any T. Another way to view this, is that there are not enough statistically independent samples $N_i = T/\tau_s$ when we don't have $T/\tau_c >> 1$ for the fit to be statistically meaningful. We note that works on ensemble based fitting theory [5,6,12] often implicitly assume $\mathcal{E}\ldots \cong <\ldots>_T$ as $N\to\infty$ for any T, but we have just shown this is not the case for correlated noise processes. This assumption for $T\to 0$ is called *local ergodicity* [25]. To coin a phrase, a noise process with a substantial τ_c is *intermediate ergodic*; that is, one must have $T/\tau_c >> 1$ for the process to be ergodic-like.

Finally, what is obvious from this above discussion is that non-highpass filtered neg-p noise will generate such anomalous polynomial fitting behavior for any T, because $\tau_c = \infty$. Again, $|f|^{-1}$ noise can often be a marginal exception, because t_0 is large but not infinite and white noise effects can dominate.

IV. CONCLUSIONS

In this paper, we have shown that one cannot simply swap variances to "fix" a neg-p divergence problem without further action. We have shown that each type of variance is a statistical answer to a different error question and such arbitrary swapping merely masks the true problem that caused the divergence. We have also shown that such neg-p variance divergences are true indicators of estimation problems that must be physically addressed, not ignored. Furthermore, we have shown that non-highpass filtered neg-p noise acts like systematic error and generates anomalous behavior in statistical estimation techniques like least squares and Kalman filters when the estimation functions consist of polynomials. It has also been shown that this systematic behavior is due to the non-ergodic and infinitely correlated nature of such neg-p noise. As a final note, the surest way to reduce the anomalous effects of neg-p noise is develop frequency standards with lower neg-p noise. This is good news for frequency standards developers.

REFERENCES

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

[2] V. S. Reinhardt, "Modeling Negative Power Law Noise," *in Proc. 2008 IEEE IFCS*, pp. 685-592.

[3] V. S. Reinhardt, "A Physical Interpretation of Difference Variances," *in the Proc.. 2007 IFCS* , pp. 961-968.

[4] R.A., Baugh, "Frequency Modulation Analysis with the Hadamard Variance," *in Proc. 25th Annual Frequency Control Symposium*, *1971*, pp. 222–225.

[5] J. R. Wolberg, *Prediction Analysis*, Princeton, NJ: D. Van Nostrand and Co, 1967.

[6] H. W. Sorenson, "Kalman Filtering Techniques," in *Advances in Control Systems* (Vol. 3), C. T. Leondes (Ed.), New York: Academic Press, 1966.

[7] V, S. Reinhardt, “How Extracting Information from Data Highpass Filters its Additive Noise,” *in Proc. 39th PTTI Systems and Applications Meeting, Dec., 2007*, pp. 559-580.

[8] V. S. Reinhardt, “On Difference Variances as Residual Error Measures in Geolocation,” *in Proc. Institute of Navigation 2008 National Technical Meeting, Jan., 2008,* pp. 763-772.

[9] V. S. Reinhardt, “Characterizing the Impact of Time Error on General Systems,” *in Proc. 2008 IEEE IFC*, pp. 677-684.

[10] V. S. Reinhardt, “Zero Mean Noise Processes that Do Not Appear to be Zero Mean,” *in Proc. Institute of Navigation International Technical Meeting 2009, January 26-28, 2009*¸ pp 384-390.

[11] V. S. Reinhardt, “The Profound Impact of Negative Power Law Noise on the Estimation of Causal Behavior,” *in Proc. 2009 IEEE IFC,* pp. 322-327.

[12] R. G. Brown, Introduction to Random Signal Analysis and Kalman Filtering, New York: John Wiley & Sons, 1983.

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

[14] A. Papoulis and S. U. Pillai, Probability, Random Variables, and Stochastic Processes, 4^{th} ed., Boston MA: McGraw-Hill, 2002.

[15] D. A. Howe, R. L. Beard, C. A. Greenhall, F. Vernotte, W. J. Riley, and T. K. Peppler, “Enhancements to GPS Operations and Clock Evaluations Using a ‘Total’ Hadamard Deviation,” *IEEE trans. UFFC*, v. 52, #8, Aug. 2005, pp. 1253-1261.

[16] C. A. Greenhall, Jet Propulsion Laboratory, *private communication, June, 2009*.

[17] Victor S. Reinhardt, “A Review of Time Jitter and Digital Systems” *in Proc. 2005 IEEE IFC*, pp. 38-45.

[18] L. L. Scharf, B. Firedlander, and D. J. Thomson, 1998, “Covavariant Estimators of Time-Frequency Descriptors for Nonstationary Random Processes,” *in Proc. 32nd Asilomar Conference on Signals, Systems, and Computers, v1*, Pacific Grove, CA, 1998, pp 808-811

[19] W. B. Davenport, Jr. and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*, New York, NY: McGraw-Hill, 1958.

[20] C. A. Greenhall, “Recipes for Degrees of Freedom of Frequency Stability Estimators,” *IEEE Trans. I&M, vol. 40, no. 6, Dec., 1991*, pp. 994-999.

[21] M. S. Keshner, “1/f Noise,” Proc. IEEE, v. 70, no, 3, March, 1982, pp. 212-218.

[22] E. Parzan, *Stochastic Processes*, San Francisco, CA: Holden-Day, 1966.

[23] D. M. Young*, Iterative Solution of Large Linear Systems,* New York, NY: Academic Press, 1971.

[24] V F Gapoškin, “The Local Ergodic Theorem for Groups of Unitary Operators and Second Order Stationary Processes,” Math. USSR Sb. 39, 1981, pp,227-242.

[25] M. Abramowitz and I, A. Stegun, Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables, 1st ed., Mineola, New York: Dover, 1965.