Proceedings of the 2009 Joint Meeting of the EFTF and the IEEE IFCS, Besançon, France, April 21-24, 2009.

The Profound Impact of Negative Power Law Noise on the Estimation of Causal Behavior

Victor S. Reinhardt, Raytheon Space and Airborne Systems, El Segundo, California, USA

Abstract—This paper demonstrates that highly correlated negative power law (neg-p) noise—noise with a PSD ∝ $|f|^p$ when p<0—and causal behavior contained in data cannot be properly separated from each other by *any* causal fitting or estimation technique. It then shows: (a) that this leads such techniques to generate anomalous estimates of the true causal behavior and to underreport the true fit error, and (b) that these anomalies cannot be corrected by adding modeling or increasing the data collection interval T. Further consequences of the above are also discussed. These include anomalous behavior in noise whitening, M-corner hat, and cross-correlation techniques and the non-observability of unbiased measures of pure random error when neg-p noise is present. The paper also investigates the relationship between ergodic-like behavior—the approximate equality of time averages over a finite T and ensemble averages—and such anomalous behavior in general statistical processing techniques. It is shown that such ergodic-like behavior is a necessary condition for practical implementations of these techniques to behave like their theoretical counterparts and, furthermore, a necessary condition for ergodic-like behavior is $T>>\tau_c$, where τ_c is the correlation time of the noise process involved. Finally, it is shown that this is unachievable for neg-p noise, because $\tau_c = \infty$ for such noise (unless system highpass filtering makes τ_c finite for the system filtered noise variable).

INTRODUCTION

It is well-known that systematic or correlated error cannot be properly separated from causal behavior in data using fitting or estimation techniques (i.e., a least squares fit (LSQF) [1] or a Kalman filter [2]). But the profound implications of this in dealing with highly correlated negative power law (neg-p) noise (noise with a $\text{PSD} \propto |f|^p$ for p<0 [3,4]) have not been fully appreciated. This paper explores such implications.

A. The Estimation of Causal Behavior Imbedded in Data

This paper will primarily deal with the problem of generating $x_{a,M}(t)$ an M-parameter estimate of true causal behavior $x_c(t)$ imbedded in data samples $x(t_n)$ over the data collection interval T when noise $x_r(t)$ is also present (See Fig. 1 and Table I). For more detail, also see [5-8], but be warned that there are notational differences. For the techniques that generate $x_{a,M}(t)$ from the data, we will only assume: (i) that the superposition principle holds [9], and (ii) that $x_{a,M}(t) \rightarrow x_c(t)$ when $x_r(t) = 0$ and there is no model error ($x_{a,M}(t)$ can track $x_c(t)$ over T with the appropriate parameter adjustment [1]). Note that x(t) is *any* observable continuous data variable and the $x(t_n)$ are its data samples. As a conceptual aid, one can view x(t) as a time error-like variable [3,10], though this is not a condition for the paper's results. By *observable* we mean that x(t) is the measured output of an instrument, such as $x(t) = x_{Tx}(t-\tau) - x_{Tx}(t)$ the range output of a 2-way ranging receiver, where $x_{Tx}(t)$ is the *unobservable* true time error of its timing source and t is the ideal observation time [5-8]. Note also from the table that x(t) can be a function of other t-dependent independent variables $\underline{\upsilon}(t)$, but for most of the paper, we will subsume these in x(t).

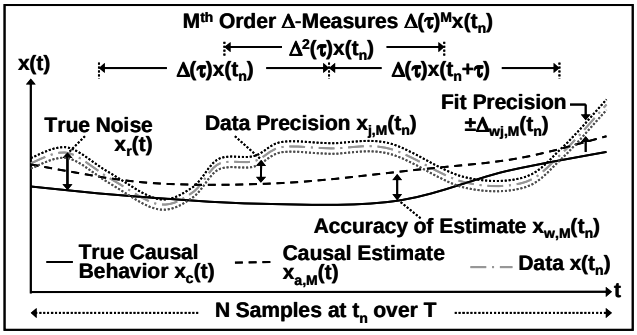


Fig. 1. Model for the estimation causal behavior from noisy data.

TABLE I. MODEL DEFINITIONS [5-8].

Data variable: $x(t) = x_c(t) + x_r(t)$ Noise variable: $x_r(t)$ Causal behavior: $x_c(t)$ ($= x_c(t,\underline{\upsilon}(t))$, $\underline{\upsilon}(t)$ = other variables included in $x_c(t)$
$x_{a,M}(t)$ = M-parameter estimate of $x_c(t)$ $x_{poly,M}(t)$ = (M-1)th order polynomial
True accuracy of fit: $x_{w,M}(t_n) = x_{a,M}(t_n) - x_c(t_n)$ $\mathcal{E}$ = ensemble average Point variance (Kalman): $\Delta^2_{w,M}(t_n) = \mathcal{E}[x_{w,M}(t_n) - \mathcal{E}\{x_{w,M}(t_n)\}]^2$ Average variance over T (LSQF): $\sigma^2_{w,M} = \sum_n \xi_n \Delta^2_w(t_n)$ [n = 1 : N]
Data precision: $x_{j,M}(t_n) = x(t_n) - x_{a,M}(t_n)$ $\Delta^2_{w,M}(t_n) = \mathcal{E}[x_{w,M}(t_n) - \mathcal{E}\{x_{w,M}(t_n)\}]^2$ $\sigma^2_{j,M} = \sum_n \xi_n \Delta^2_{j,M}(t_n)$
Fit precision (point variance): $\Delta^2_{wj,M}(t_n) = \rho_d(t_n)\Delta^2_{j,M}(t_n)$ Fit Precision (average variance): $\sigma^2_{wj,M} = \rho_d \sigma^2_{j,M}$
M^{th} - order Δ - measures: $\Delta(\tau)^M x(t_n)$ $\Delta(\tau)x(t_n) \equiv x(t_n+\tau) - x(t_n)$ $\Delta^2_{\tau,M}(t_n) = \lambda_M^{-1}\mathcal{E}[\Delta(\tau)^M x(t_n)]^2$ $\sigma^2_{\tau,M} = \xi_0 \sum_n \Delta^2_{\tau,M}(t_n)$ $\lambda_M = \sum_{m=0}^{M}\left(\frac{M!}{m!(M-m)!}\right)^2$

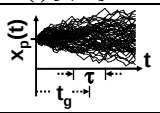
$x_{w,M}(t_n)$ as defined in Table I is the *only* true detailed measure of the accuracy of $x_{a,M}(t)$. From $x_{w,M}(t_n)$, we form two statistical measures: (i) $\Delta^2_{w,M}(t_n)$ the point variance used in a Kalman filter, and (ii) $\sigma^2_{w,M}$ the average variance used in a LSQF [7,8]. These accuracy measures, however, are

unobservable, since a priori knowledge of $x_c(t)$ is required to generate them. The basic *observable* error we will use is the data precision $x_{j,M}(t_n)$ and its variances $\Delta^2_{j,M}(t_n)$ and $\sigma^2_{j,M}$. From these one can generate the fit precision measures $\Delta^2_{wj,M}(t_n)$ and $\sigma^2_{wj,M}$, which estimate $\Delta^2_{w,M}(t_n)$ and $\sigma^2_{w,M}$ based on theoretical knowledge of the conversion factors $\rho_d(t_n)$ and ρ_d. We will later show how to properly calculate ρ_d and $\rho_d(t_n)$ when the neg-p order is known. We note, for uncorrelated $x_r(t_n)$ and a uniform LSQF, that $\rho_d \rightarrow \rho_{do}=M/(N-M)$ [1] and that this ρ_{do} and its equivalent $\rho_{do}(t_n)$ are often used when correlated noise is present. This, we will later show, can lead to very misleading accuracy estimates when neg-p noise is present.

The third set of error measures we will discuss are M^{th} order Δ-measures $\Delta(\tau)^M x(t_n)$ over the interval τ and their variances $\Delta^2_{\tau,M}(t_n)$ and $\sigma^2_{\tau,M}$ [3,5-7]. These variances are proportional to the Allan variance of the time error [3] when M=2 and to the Hadamard variance of the time error [11] when M=3. We will later show that such Δ-measures can be viewed as either stability or precision measures under the right fitting conditions [5-8].

B. Representations of Neg-p Noise

TABLE II. REPRESENTATIONS OF NEG-P NOISE OF x(t) [4,12]

Non-Stationary (NS) Picture ($x_p(t) = 0$ for $t<0$) t_g = Global time from start of noise process τ = local or difference time	
Covariance or Correlation Fn ($x_p(t)$ real, $\mathcal{E}x_p(t) = 0$)	$R_r(t_g,\tau) = \mathcal{E}\, x_p(t_g+\tau/2)x_p(t_g-\tau/2)$ $R_p(t_g,\tau) = 0$ for $t_g < 0$ or $\lvert\tau\rvert > 2\, t_g$
Wigner-Ville Function	$W_p(t_g,f) = \mathcal{F}_{f,\tau} R_p(t_g,\tau)$
Loève Spectrum	$L_p(f_g,f) = \mathcal{F}_{f_g,t_g} W_p(t_g,f)$
(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 ($x_p(t) \neq 0$ for all t)	
Covariance or Correlation Function	$R_p(\tau) = \lim_{t_g\rightarrow\infty} R_p(t_g,\tau) \rightarrow \infty$ for all τ
SSB PSD	$L_p(f) = \lim_{t_g\rightarrow\infty} W_p(t_g,f) \propto \lvert f\rvert^p \quad [p<0]$

Neg-p noise $x_p(t)$ gets its name from $L_p(f)$ the single sideband wide-sense stationary (WSS) PSD given in Table II [3,4]. Note that the SSB PSD is used here for mathematical convenience [5-9,12], not the DSB $S_p(f)$ used in much time and frequency (T&F) literature [3]. Neg-p noise, however, is in reality non-stationary (NS), as is summarized in Table II [4]. In this NS picture, $x_p(t) = 0$ for $t < 0$, and its covariance or correlation function $R_p(t_g,\tau)$ (the same here because $\mathcal{E}x_p(t)=0$) is a function of both τ the difference time between the covariant arguments and t_g the global time from the start of $x_p(t)$ [4,12]. A neg-p $R_p(t_g,\tau)$ has two fundamental properties: (i) $R_p(t_g,\tau) < \infty$ for $t_g < \infty$, and (ii) $R_p(t_g,\tau)\rightarrow\infty$ as $t_g\rightarrow\infty$. In the NS picture, there are two spectral forms, the Wigner-Ville Function $W_p(t_g,f)$ and the Loève Spectrum $L_p(f_g,f)$. Note from the table that the neg-p WSS covariance $R_p(\tau)\rightarrow\infty$ for all τ, because it is the $t_g\rightarrow\infty$ limit of $R_p(t_g,\tau)$ and property (ii). Finally note that $L_p(f)$ can be defined without the use of $R_p(\tau)$ as the $t_g\rightarrow\infty$ limit of $W_p(t_g,f)$.

It is important to note that $x_p(t)$ is not the same as $x_r(t)$. This is because variables such $x(t)$, $x_r(t)$, and $x_c(t)$ are generally filtered by a system. We represent this using the system response function $H_s(f)$ described in Table III [5-8]. Further note that such $H_s(f)$ have highpass (HP) behavior as well as lowpass (LP) behavior for typical systems [5-8]. Finally, note one can define τ_c the correlation time of $x_r(t)$ as given in Table III [13]. One can show $\tau_c\rightarrow\infty$ for neg-p noise, unless $|H_s(0)|^2$ has a zero of sufficient order to suppress the pole in $L_p(0)$ [4]. This will be important in our later discussions.

TABLE III. THE SYSTEM FILTERING OF DATA VARIABLES [8-11]

$t-$domain: $x_r(t) = h_s(t-t') \otimes_{t'} x_p(t')$ $R_r(t_g,\tau) = \mathcal{E}\, x_r(t_g+\tau/2)x_r(t_g-\tau/2)$	$x(t) \otimes_t y(t) = \int_{-\infty}^{+\infty} dt\, x(t)y(t)$
$f-$domain: $X_r(f) = \mathcal{F}_{f,t} x_p(t) = H_s(f)X_p(f)$ $X_p(f) = \mathcal{F}_{f,t} x_p(t)$ $L_r(f_g,f) = H_s(f+0.5f_g)H_s^*(f-0.5f_g)L_p(f_g,f)$	$H_s(f) = \mathcal{F}_{f,t} h_s(t)$ $L_r(f_g,f) = \lvert H_s(f)\rvert^2 L_p(f)$ Similar equations for x(t) & $x_c(t)$
Filtered correlation time τ_c	$\tau_c = \lim_{t_g\rightarrow\infty} 0.5R_r(t_g,0)^{-1} \otimes_\tau R_r(t_g,\tau)$

ANOMALOUS FITTING BEHAVIOR DUE TO NEG-P NOISE

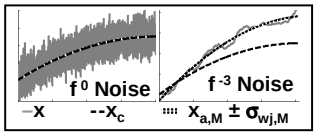


Fig. 2. Uniform LSQF with p = 0 and p = -3 (N=2048).

The major thesis of this paper is that neg-p noise correlations prevent *any* causal fitting or estimation technique from properly separating $x_r(t)$ from $x_c(t)$ in $x_{a,M}(t)$ and that this causes the fitted $x_{a,M}(t)$ to significantly deviate from $x_c(t)$ for any N. Fig. 2 demonstrates such behavior for f^{-3} noise when: (i) a uniform LSQF is used, (ii) both $x_c(t)$ and $x_{a,M}(t)$ are 2^{nd} order polynomials (M=3), (iii) $H_s(f)$ is an ideal Nyquist LP filter, and (iv) ρ_{do} is used to predict the error bars $x_{a,M}(t)\pm\sigma_{wj,M}$ (which are so small in the figure that they appear coalesced with $x_{a,M}(t)$). The f^0 case is also shown for reference, where we note the fit behaves as is classically expected. For the f^{-3} case, one can see that $x_{a,M}(t)$ significantly deviates from $x_c(t)$ and $\sigma_{wj,M}$ does not properly predict $\sigma_{w,M}$. What is happening here is that the $x_r(t_n)$ ensemble member here has behavior that is substantially quadratic, which the fit includes in $x_{a,M}(t_n)$. This causes $x_{a,M}(t_n)$ to veer off from $x_c(t)$ and the true error to be underreported, a well-known effect of correlated systematic error in LSQF theory [3]. This 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 [14].

A further consequence of this is that noise whitening, a technique used to determine the structure of $x_c(t)$ by increasing M until $x_{j,M}(t_n)$ is white (uncorrelated) [1], is a faulty procedure to use when neg-p noise is present. This is because part of correlated portion of the neg-p noise will be included in $x_{a,M}(t_n)$, making the conclusion about the nature of

$x_c(t)$ faulty. In other words, noise whitening is based on a white noise truth model and the noise here is not white.

This anomalous behavior is further demonstrated in a Kalman filter, as is shown in Fig. 3. Here, simulation results are shown with both f^0 and f^{-2} measurement noise, when $x_c(t)$ and the causal model are both quadratic and (a) uncorrelated and (b) correlated measurement noise models [2] are used for f^{-2} noise [15]. For both (a) and (b), one again sees that $x_{a,M}(t)$ veers off from $x_c(t)$ and $\Delta_{wj,M}(t_n)$ underestimates $\Delta_{w,M}(t_n)$, but (b) yields better results. This better result is due to the fact that the f^{-2} ensemble member shown substantially deviates from quadratic behavior, and thus the correlated Kalman noise model is partially effective. One would expect poorer correlated noise model results for f^{-3} noise, because such noise tends to be much more quadratic-like, as in Fig. 2.

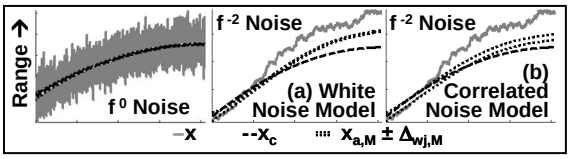


Fig. 3. Anomalous neg-p behavior in a Kalman filter.

Fig. 2 and Fig. 3 also demonstrate another very important principle: ***(P1)*** *A noise process must be ergodic-like over T for the practical realization of a fitting technique to behave like its theoretical counterpart* [15]. By ergodic-like, we mean that $\langle..\rangle_T$ a finite time average over T for a single ensemble member data set approximates the equivalent $\mathcal{E}$ average over all ensemble members. We borrow here from (strict) ergodicity, where $\langle..\rangle_T \to \mathcal{E}$ as $T\to\infty$ [16]. One should note that this distinction between $\langle..\rangle_T$ and $\mathcal{E}$ is often overlooked in discussing fitting processes [1,2,9]. As a corollary to this, we state a second principle: ***(P2)*** *One must have $T/\tau_c >> 1$, for a noise process to be ergodic-like* [15]. Both these principles are stated as Ansätze that we demonstrate by examples, such as those in Fig. 2, Fig. 3, and Fig. 4, rather than as formal theorems. However, these principles are intuitively obvious, because the decorrelation of samples as $T\to\infty$ is a necessary condition for strict ergodicity [16].

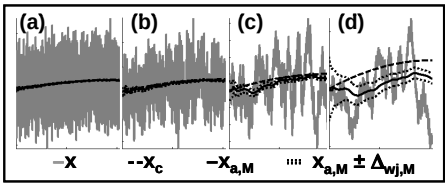


Fig. 4. Kalman filter simulations for a correlated Gauss-Markov process: (a) T/τ_c=2000, (b) T/τ_c=200, (c) T/τ_c=20, (d) T/τ_c=2.

To further illustrate these principles, let us examine Fig. 4. Here we show Kalman filter simulations of correlated but stationary Gauss-Markov noise [9,15] with various ratios of T/τ_c. It is obvious from the figure that one must have $T/\tau_c >> 1$ for a practical Kalman solution to act like its ensemble averaged counterpart. Thus such a correlated process has intermediate ergodicity, rather than local ergodicity (ergodicity as $T\to 0$) [17]. Principal (P2) is well-known in another form as the fact that the number of statistically independent samples over a data set is given by $N_{eff}=T/\tau_s$, not $N=T/t_s$, when correlated noise is present [9,13]. Thus, a practical fit becomes anomalous when T approaches τ_c, because there is a deficiency of independent samples. The problem of course in dealing with neg-p noise is that $\tau_c=\infty$ for such noise; that is, it is non-ergodic. Thus: ***(P3)*** *Neg-p noise generates anomalous fitting behavior for all T, unless $H_s(f)$ is sufficient to make τ_c finite for $x_r(t)$* [15].

The above is also true for the practical realization of any statistical processing technique. Thus M-corner hat [18] and cross-correlation techniques [19,20] should exhibit similar anomalous behavior when neg-p noise is present. For example, both these techniques assume that ensemble independence of sources implies zero $\langle..\rangle_T$ cross-correlations for large N. For independent sources that are white, non-zero $\langle..\rangle_T$ cross-correlation fluctuations decrease as $N^{-1/2}$, but for such sources that are internally correlated, it is obvious from the above that these fluctuations decrease as $(\tau_c/T)^{1/2}$ and only when $T >> \tau_c$. Thus, neg-p noise is again problematic unless there is sufficient $H_s(f)$ HP filtering to make τ_c finite. This inability to handle accidentally correlated behavior in independent sources is a recognized limitation in M-corner hat techniques, but its scope has not been fully appreciated in dealing with neg-p noise. For cross-correlation techniques, τ_c is finite for $p \geq -2$. ($|H_s(f)|^2 \propto f^2$ for $|f| << B_L$ the loop bandwidth of the 1st order PLL used in this technique [19].) However, $\tau_c\to\infty$ for $p \geq -2$, which means that anomalous behavior should appear for f^{-3} noise, but with f^{-1} anomalous properties, which only grow as ln(T) [4]. This needs further investigation.

SPECTRAL REPRESENTATION OF ACCURACY AND PRECISION

Table IV summarizes a previously presented Green's function methodology for calculating $\Delta^2_{\zeta,M}(t_n)$ and $\sigma^2_{\zeta,M}$ (ζ = w or j) and ρ_d and $\rho_d(t_n)$ in terms of $L_p(f)$ or $L_p(f_g,f)$ [6-8]. The fit here is only assumed to have the general properties given in Section A of the Introduction [6-8]. (Note that our notation here is somewhat different from that in [6-8] and that Table IV corrects typos in (A.23) of [6] and (A.11) of [7].)

Table V augments Table IV with new material specific to various types of LSQF.

In these tables, the $x_{\zeta,M}(t)$ are related to x(t) using continuous Green's functions $g_{\zeta,M}(t,t')$ and their Fourier transforms $G_{\zeta,M}(t,f)$. Note that the t'-integration of $g_{w,M}(t,t')x(t')$ here is in actuality a discrete t_n-summation of $g^{(N)}_{w,M}(t,t_n)x(t_n)$, but the continuous form is used so that $\Delta^2_{\zeta,M}(t_n)$ and $\sigma^2_{\zeta,M}$ can be formally written as continuous spectral integrations. Also note that no model error is assumed here, so all errors are purely a function of $x_r(t)$. It is a simple matter to add such model error to this formulation if so desired. Finally, the spectral estimation kernels $|G_{\zeta,M}(t,f)|^2$ and $K_{\zeta,M}(f)$ together with $H_s(f)$ completely specify $\Delta^2_{\zeta,M}(t_n)$ and

$\sigma^2_{\varsigma,M}$ once $L_p(f)$ or $L_p(f_g,f)$) are known, and thus ρ_d and $\rho_d(t_n)$ are determined given only the p-order of the noise.

When $x_{a,M}(t)\rightarrow x_{poly,M}(t)$, it has been proven that the estimation kernels have the $2M^{th}$ order LP and HP filtering properties also listed in Table IV [6,7]. Fig. 5 demonstrates this for $K_{j,M}(f)$ with a LSQF. The fact that a fit can be viewed as a filter is well-known [9], but this specific $2M^{th}$ order form for $x_{poly,M}(t)$ estimation is what is new here. A consequence of this is that $\Delta^2_{j,M}$ and $\sigma^2_{j,M}$ are always guaranteed to converge for any neg-p order given an appropriate choice of $x_{poly,M}(t)$, independent of the HP properties of $H_s(f)$ [6,7].

TABLE IV. GREEN'S FUNCTION REPRESENTATION OF A FITTING PROCEDURE

$x_{a,M}(t)=g_{w,M}(t,t')\otimes_{t'} x(t')$ $\quad g_{w,M}(t,t')=\sum_n g^{(N)}_{w,M}(t,t_n)\delta(t'-t_n)$

$g_{j,M}(t,t')=\delta(t-t')-g_{w,M}(t,t')$ $\quad G_{\varsigma,M}(t,f)=\mathcal{F}_{f,t'}\, g_{\varsigma,M}(t,t')$ $\quad (\varsigma = w \text{ or } j)$

Point Variances $\rho_d(t_n)=\Delta^2_{w,M}(t_n)/\Delta^2_{j,M}(t_n)$ (zero model error)

WSS: $\Delta^2_{\zeta,M}(t_n)=|G_{\zeta,M}(t_n,f)|^2|H_s(f)|^2\otimes_f L_p(f)$

NS: $\Delta^2_{\zeta,M}(t_n)=G_{\zeta,M}(t_n,f+0.5f_g)G^*_{\zeta,M}(t_n,f-0.5f_g)$

$\cdot\otimes_{f_g}\otimes_f H_s(f+0.5f_g)H^*_s(f-0.5f_g)L_p(f_g,f)$

Average Variances $\rho_d=\sigma^2_{w,M}/\sigma^2_{j,M}$ (zero model error)

WSS: $\sigma^2_{\zeta,M}=K_{\zeta,M}(f)|H_s(f)|^2\otimes_f L_p(f)$ $\quad K_{\zeta,M}(f)=\sum_n \xi_n|G_{\zeta,M}(t_n,f)|^2$

NS: $\sigma^2_{\zeta,M}=K_{\zeta,M}(f_g,f)H_s(f+0.5f_g)H^*_s(f-0.5f_g)\otimes_{f_g}\otimes_f L_p(f_g,f)$

$K_{\zeta,M}(f_g,f)=\sum_n \xi_n G_{\zeta,M}(t_n,f+0.5f_g)G^*_{\zeta,M}(t_n,f-0.5f_g)$

Spectral Behavior: For $x_{a,M}(t) = x_{poly,M}(t)$ an $(M-1)^{th}$ polynomial, $G_{w,M}(t,f)$ and $G_{j,M}(f)$ are M^{th} order LP & HP filters and $K_{w,M}(f)$ and $K_{j,M}(f)$ are $2M^{th}$ order LP & HP filters, all with knee frequencies on the order of f_T [6,7]

Noise bandwidth of fit: $f_T=0.5\otimes_f K_{w,M}(f)/K_{w,M}(0)$

TABLE V. LSQF FORMULAS

Weighted LSQF $x_{a,M}(t)=\underline{u}(t)^\dagger\underline{a}$ $\quad \underline{u}(t)^\dagger\otimes_t\rho_\xi(t)\underline{u}(t)=\underline{1}$ **(orthonormal)**

$\chi^2=[x(t)-\underline{a}^\dagger\underline{u}(t)]\otimes_t\rho_\xi(t)[x(t)-\underline{u}(t)^\dagger\underline{a}]$ $\quad \underline{u}(t)=(u_1(t)..u_M(t))^\dagger$

$\underline{a}=\underline{u}(t)\otimes_t\rho_\xi(t)x(t)$ $\quad x_{a,M}(t,\underline{a})=\underline{u}(t)^\dagger\underline{u}(t')\otimes_{t'}\rho_\xi(t')x(t')$ $\quad \underline{a}=(a_1..a_M)^\dagger$

$g_{w,M}(t,t')=\underline{u}(t)^\dagger\underline{u}(t')\rho_\xi(t')$ $\quad \rho_\xi(t)=\sum_n\xi_n\delta(t-t_n)$

$G_{w,M}(t,f)=\underline{u}(t)^\dagger\underline{H}^{(u)}(f)$ $\quad \underline{H}^{(u)}(f)=\mathcal{F}_{f,t}\rho_\xi(t)\underline{u}(t)$ $\quad {}^\dagger$= Transpose

$K_{w,M}(f)=\underline{H}^{(u)}(f)^\dagger\underline{H}^{(u)}(f)$ $\quad K_{j,M}(f)=1-K_{w,M}(f)$

$\mathrm{Cov}(\underline{a})=\mathcal{E}\underline{a}\underline{a}^\dagger=\underline{H}^{(u)}(f)\underline{H}^{(u)}(f)^\dagger\otimes_f|H_s(f)|^2\,L_p(f)$ (no model error)

For all $\xi_n=\xi_0$ & $N\rightarrow\infty$: $\underline{u}(t)$ = normalized Legendre Polynomials.

Polynomials $\underline{z}(t)=(1..t^{M-1})^\dagger$ $\quad x_{a,M}(t,\underline{a})=\underline{z}(t)^\dagger\underline{b}$ $\quad \underline{u}(t)=\underline{w}\underline{z}(t)$

$G_{w,M}(t,f)=\underline{z}(t)^\dagger\underline{w}^\dagger\underline{w}\underline{H}^{(z)}(f)$ $\quad K_{w,M}(f)=\underline{H}^{(z)}(f)^\dagger\underline{w}^\dagger\underline{w}\underline{H}^{(z)}(f)$ $\quad \underline{b}=\underline{w}^\dagger\underline{a}$

$\underline{H}^{(z)}(f)=\mathcal{F}_{f,t}\rho_\xi(t)\underline{z}(t)$ $\quad H^{(z)}_m(f)=(j\partial/\partial\omega)^{m-1}\sum_n\xi_n\exp(-j\omega t_n)$

Evaluate $\underline{H}^{(z)}(f)$ with high precision for $f<<1$ because of term cancellation.

Uniform weighting $\xi_n=\xi_0=1/N$ $\quad H^{(z)}_m(f)=(j\partial/\partial\omega)^{m-1}\dfrac{\exp(-j\omega t_0)}{N}\dfrac{1-\exp(-j\omega T)}{1-\exp(-j\omega t_s)}$

Uncorrelated $x_r(t_n)$ $\mathcal{E}x_r(t_n)x_{r'}(t_{n'})=\sigma^2\delta_{n,n'}$ & uniform LSQF ($\xi_n=\xi_0$)

$\sigma^2_{w,M}=M\xi_0\sigma^2$ $\quad \sigma^2_{j,M}=(N-M)\xi_0\sigma^2$ $\quad \rho_{do}=M/(N-M)$

C. *The Graphical Represention of ρ_d*

In order to gain an intuitive understanding of the behavior of $\sigma^2_{w,M}$, $\sigma^2_{j,M}$, and ρ_d for neg-p noise, it is useful to graph the spectral integrations in Table IV, as is shown in Fig. 6. Here, we have simplified $K_{w,M}(f)$ and $K_{j,M}(f)$, representing them as sharp cut-off LP and HP filters with knee frequencies f_T ($N>>M>>1$). We have also simplified $H_s(f)$ representing it as a simple sharp LP filter with a knee frequency at f_h. Fig. 6(a) shows the $L_0(f)$ white noise case, where one can easily see that $\rho_d=f_T/(f_h-f_T)$. Thus, even for white noise and a uniform LQSF ($\xi_n=\xi_0$), one does *not* obtain the uncorrelated $\rho_{do}=M/(N-M)$ unless $H_s(f)$ is an ideal Nyquist sampling filter with $f_h=(2t_s)^{-1}$ [9] ($N=T/t_s$ and one can show $f_T=M/(2T)$). This is a restatement of the well-known correlation effect that occurs in oversampling for the analog bandwidth [9].

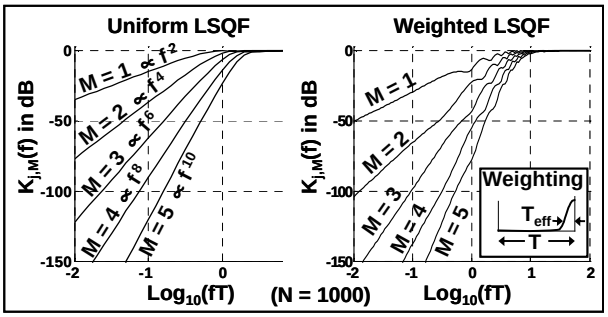


Fig. 5. $K_{j,M}(f)$ HP filtering behavior for $x_{poly,M}(t)$ [9-10].

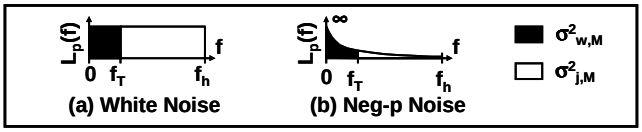


Fig. 6. WSS spectral representations of $\sigma^2_{w,M}$ and $\sigma^2_{j,M}$.

Now consider the Fig. 6(b) case, where neg-p noise is present. It is obvious from the graph that $\sigma^2_{w,M}\rightarrow\infty$ and $\sigma^2_{j,M}$ is finite ($\rho_d\rightarrow\infty$) for any T. In the past, this WSS $\sigma^2_{w,M}$ infinity has been considered a mathematical defect of $\sigma^2_{w,M}$ [3], but we will show later, using the NS picture, that this infinity indicates a real physical problem.

D. *The effect of an HP filtering System Response Function*

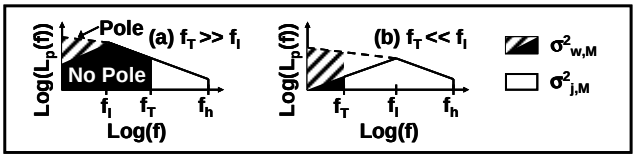


Fig. 7. The effect of HP $H_s(f)$ on $\sigma^2_{w,M}$ and $\sigma^2_{j,M}$ for neg-p noise.

We noted before that $H_s(f)$ can have HP filtering properties. Here we will discuss the impact of these HP properties on $\sigma^2_{w,M}$, $\sigma^2_{j,M}$, and ρ_d when neg-p noise is present [6-8]. Fig. 7 again graphs $\sigma^2_{w,M}$ and $\sigma^2_{j,M}$ when neg-p noise is present but with $|H_s(f)|^2$ approximated by a $2M^{th}$ order HP filter with a knee frequency f_l plus a sharp LP cutoff at f_h. Combinations of two sets of conditions are shown in the

figure: (i) the $|H_s(f)|^2$ HP roll-off suppresses the $L_p(0)$ pole, or (ii) it does not, and (a) $f_T >> f_l$, or (b) $f_T << f_l$. For (a-i) and (b-i), one can see that ρ_d is finite, but will deviate from the uncorrelated ρ_{do}. For (a), $\rho_d > \rho_{do}$, and interestingly for (b), $\rho_d < \rho_{do}$. We also note for these cases, that τ_c is finite ($\tau_c \cong 1/(2f_l)$) for the filtered neg-p noise, and this noise is both WSS and intermediate ergodic. Thus, a fitting technique will be well-behaved for such filtered neg-p noise. For combinations (a-ii) and (b-ii), both $\sigma^2_{w,M}$ and ρ_d are infinite. In these cases, τ_c is infinite, but the effective neg-p order is reduced for $f < f_l$.

NEG-P ACCURACY AND PRECISION IN THE NS PICTURE

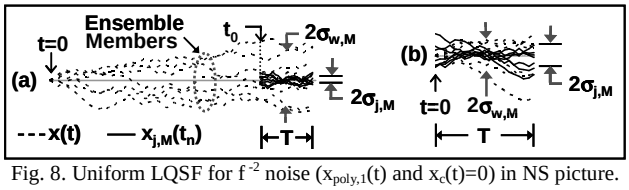


Fig. 8. Uniform LQSF for f^{-2} noise ($x_{poly,1}(t)$ and $x_c(t)=0$) in NS picture.

To gain a better understanding of what is physically happening when $\sigma^2_{w,M} \to \infty$ in the WSS picture, consider the NS picture of the similar situation in Fig. 8(a). Here: (i) multiple ensemble members of $x(t) = x_r(t)$ (so $x_c(t) = 0$) start at the finite time $t = 0$, (ii) the data collection starts at $t_0 > 0$, and (iii) $H_s(f)$ is a simple Nyquist LP filter. From the figure, one can see that $\sigma^2_{w,M} > \sigma^2_{j,M} >> \rho_{do}\sigma^2_{j,M}$ and $\sigma^2_{w,M} \to \infty$ as $t_0 \to \infty$, because $x_r(t_0)$ becomes unbounded as $t_0 \to \infty$, but $\sigma^2_{j,M}$ remains well behaved as $t_0 \to \infty$, because $x_r(t_0)$ is treated as part of $x_{a,M}(t)$. Thus, the infinity in the WSS $\sigma^2_{w,M}$ is the indication of a true problem with the fit accuracy in a real situation; that is, the inaccuracy of $x_{a,M}(t)$ truly becomes unbounded when t_0 is very large in a physical situation. The proper response to this infinity is therefore to investigate and try to correct the real problem, not to ignore it by using an inappropriate error measure that happens to converge. Finally, note that f^{-1} noise is a borderline case, because $\sigma^2_{w,M}$ only grows as $\ln(t_0 f_h)$ [21]. Thus, even when t_0 is as large as the age of the universe, the f^{-1} contribution to $\sigma^2_{w,M}$ can be smaller than that due to the white noise [22].

Fig. 8(b) demonstrates a pitfall that occurs when using simulations to predict neg-p noise errors. Here, t_0 is small compared with τ_c ($t_0 = 0$ in this case). Note that $\sigma^2_{w,M}$ and $\sigma^2_{j,M}$ are roughly comparable here, which misrepresents the situation when t_0 is large. Thus, unless $t_0 >> \tau_c$, such simulations will not properly represent the physical situation being simulated [15]. Since this is problematic for neg-p noise when $H_s(f)$ cannot make τ_c finite, such neg-p noise simulations will always misrepresent the true physical situation.

Δ-VARIANCES AS STABILITY AND PRECISION MEASURES

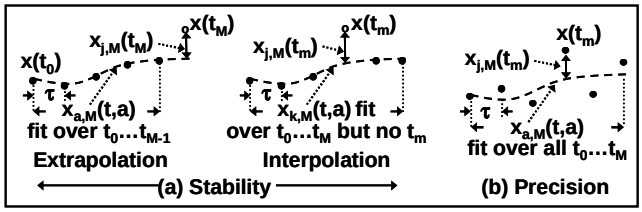


Fig. 9. Δ-measures as stability and precision measures

In Fig. 9, we summarize how M^{th} order Δ-measures can be interpreted as either stability or precision measures [5-8]. In Fig. 9(a), we *define* M^{th} order stability as the extrapolation or interpolation error $x_{j,M}(t_m)$ under the following conditions: (i) There are M+1 data points $x(t_m' = t_0 + m'\tau)$ for m' = 0 to M, (ii) $x_{a,M}(t_m) \to x_{poly,M}(t_m)$, and (iii) $x_{poly,M}(t_m)$ is fit to only M of these points $x(t_{m'})$ excluding m'=m so that $x_{poly,M}(t_m)$ exactly passes though these $x(t_{m'})$. (There are zero degrees of freedom here [5].) Under these conditions, it has been shown that $x_{j,M}(t_M) \propto \Delta(\tau)^M x(t_0)$ [5]. Thus, M^{th} order Δ-variances, such as Allan and Hadamard variances, can be rigorously interpreted as statistical measures of such stability. Fig. 9(b) shows how one can also interpret M^{th} order Δ-variances as data precision measures when all M+1 $x(t_{m'})$ are used to determine $x_{poly,M}(t)$ with a uniform LSQF. Here, one can also show that $x_{j,M}(t_M) \propto \Delta(\tau)^M x(t_0)$ [6,7].

Consider now the problem of defining *unbiased* observable statistics of purely random neg-p error. Theoretically, σ^2_r an unbiased variance of purely random neg-p error is easily formed by replacing $x(t)$ with $x_r(t)$. However, we have just shown that $x(t)$ with $x_{a,M}(t)$ removed over T is not ensemble equivalent to $x_r(t)$ when neg-p noise is present, even when $N \to \infty$. Thus, observable statistics of σ^2_r formed by removing $x_{a,M}(t)$ from $x(t)$ are not even asymptotically unbiased statistics, let alone unbiased ones for finite N. But this is precisely the implication in many standards documents, which state, without further comment, that statistics of random error are to be formed by removing frequency offsets and drifts from the data first [3,10]. Specifically, we have just shown that such observable statistics will underestimate σ^2_r by removing portions of $L_p(f)$ below f_T for $x_{poly,M}(t)$ fitting. For Δ-variance statistics of random error, one can show that such underestimation will peak as τ approaches T. This can explain the well-known tendency of non-total [23,3] statistics of Δ-variances with $x_{poly,M}(t)$ removed to generate downwardly biased results for neg-p noise when τ approaches T. For total statistics [23,3], which double the data to remove such downward biases, it is conjectured that the interaction between the doubling and fitting process causes the appearance bias removal without truly increasing the statistical confidence of the estimate. This, however, needs further investigation.

SECONDARY ACCURACY RELATIVE TO CALIBRATION

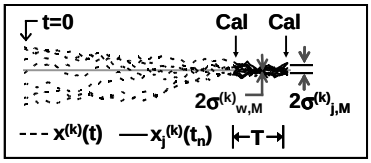


Fig. 10. Secondary accuracy relative to calibration.

Δ and precision measures are well-behaved in the presence of neg-p noise precisely because their spectral kernels HP filter $L_p(f)$ sufficiently to suppress its pole at f=0 [3,5,11,23]. One can extend this to accuracy by using the concept of secondary accuracy relative to a primary calibration standard, as is illustrated in Fig. 10. Here, we define calibrated secondary data as

$$x^{(k)}(t_n, \underline{\upsilon}(t_n)) = x(t_n \underline{\upsilon}(t_n)) - x^{(k)}_{a,M'}(t_n) \quad (1)$$

where $x^{(k)}_{a,M'}(t)$ is some function generated by a calibration process. (Note the dependence of the data on $\underline{\upsilon}(t_{n'})$ here is made explicit.) We define this calibration process as: (i) taking N' calibration measurements of the data $x(t_{n'},\underline{\upsilon}(t_{n'}))$ against a primary standard, which we will assume is perfect, when $\underline{\upsilon}_n(t_{n'}) = \underline{\upsilon}_o$, and (ii) generating $x^{(k)}_{a,M'}(t)$ from the $x(t_{n'},\underline{\upsilon}(t_{n'}))$ using some suitable fitting technique, which we will represent using a calibration Green's function $g^{(k)}_{w,M'}(t,t')$. (1) thus describes a classic experiment where the measurement accuracy of the causal dependencies of $x(t_n,\underline{\upsilon}(t_n)$ on $\underline{\upsilon}(t_{n'})$ is improved by utilizing periodic calibration when $\underline{\upsilon}_n(t_{n'}) = \underline{\upsilon}_o$. Here, the issue of the divergence of $x_r(t_0)$ as $t_0 \to \infty$ is sidestepped because such divergences are removed over T (regardless of t_0) by a suitably defined $x^{(k)}_{a,M'}(t)$. Using the previous theory, one can show that the spectral properties of $\sigma^{(k)}_{w,M}$ the secondary accuracy of this difference (defined as before but with $x(t) \to x^{(k)}(t)$) has an additional HP filtering kernel $K^{(k)}_{j,M'}$ similar in properties to $K_{j,M}$. This $K^{(k)}_{j,M'}$ ensures the convergence of $\sigma^{(k)}_{w,M}$ for any neg-p order simply by choosing the appropriate calibration process. In fact, with the appropriate system and measurement configuration, one can show that only the stability of secondary reference over T (and the primary standard, if it is imperfect), not its accuracy, is relevant to the determination of the $\underline{\upsilon}(t_{n'})$ dependencies.

CONCLUSIONS

In this paper, we have demonstrated that the correlated nature neg-p noise prevents *any* fitting procedure from properly separating neg-p noise and true causal behavior in data and have investigated the profound consequences this in many areas. These include anomalous behavior in fitting and other statistical data processing techniques, as well as the unobservability of unbiased neg-p noise measures when causal behavior is also present.

REFERENCES

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

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

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

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

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

[6] 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.

[7] 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.

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

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

[10] *ITU Definitions and Terminology for Synchronization Networks*, ITU-T G.810, International Telecommunications Union, 1996.

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

[12] 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.

[13] 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.

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

[15] 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.

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

[17] 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.

[18] J. E. Gray and D. W. Allan, "A method for estimating the frequency stability of an individual oscillator," *in Proc. Frequency Control Symp., 1974*, pp. 243-246.

[19] W, F. Walls, "Cross-Correlation Phase Noise Measurements," *in Proc. IEEE FCS, 1992*, pp 257-261.

[20] E. Rubiola and V. Giordano, "Correlation-based phase noise measurements," *Rev. Sci. Instr., v. 71, # 8, Aug, 2000*, pp. 3085-3091.

[21] J. A. Barnes, "Atomic Timekeeping and the Statistics of Precision Signal Generators," Proc. IEEE v. 54, no. 2, 1966, pp 207-220.

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

[23] 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.