FRACTIONAL DIFFERENCE PREWHITENING IN ATOMIC CLOCK MODELING

L. Schmidt, L. Breakiron, and C. Ekstrom
U.S. Naval Observatory
E-mail: *schmidt.lara@usno.navy.mil*

Abstract

A new approach to atomic clock classification has been enabled by the application of long-memory, or fractionally integrated, noise constructs. While the spectral properties of the long-memory noises are consistent with the historical $1/f^\alpha$ approach, they also allow a range of estimation strategies, in both spectral and time domains, for the classification of atomic clock behavior. These fractionally integrated noises are analyzed and applied to atomic timescales in this research, with particular emphasis on a prewhitening technique using fractional differencing which allows the separation of clock noise autocorrelation from clock rate and drift. Results from simulation studies show the utility of the fractional differencing approach both for simple fractionally integrated processes, and more complex processes which are more characteristic of atomic clock noise.

FRACTIONALLY INTEGRATED NOISE

The class of so-called long-memory or fractionally integrated processes can be used to describe the correlations seen in data from fields such as physics, chemistry, astronomy, and other sciences. Long-memory processes exhibit autocorrelation functions which decay to zero at a much slower rate than the typical autoregressive moving average (ARMA) process. Early work with a family of fractional Gaussian processes developed expressly to describe properties witnessed in physical systems (including the $1/f^\alpha$ noises) exhibits such long memory **[1]**. The well-known ARMA processes can be generalized **[2]** to accommodate long-term persistence, while allowing the short-term correlations to be described by an ARMA process. These fractionally integrated ARMA processes are often abbreviated ARFIMA(p, d, q), where d is the fractional-integration or long-memory parameter, and p and q describe the orders of the AR and MA components respectively. When $p = q = 0$, the process is termed a fractionally integrated process, or $I(d)$ process, of order d. A description of the fractionally integrated processes can be found in **[3]** and applications to atomic timekeeping using $I(d)$ noises are found in **[4]** and **[5]**. Although brief definitions are given here, the reader is referred to the above for a complete treatment.

A discrete time fractionally integrated process, $Y_t \sim I(d)$, of order d, is defined by

$$Y_t = \nabla^{-d} Z_t = (1-B)^{-d} Z_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \tag{1}$$

where Z_t is a normally distributed sequence of independent, identically distributed random variables with zero mean and variance σ^2, B is the backshift operator, and $\psi_j = \Gamma(j+d)/(\Gamma(d)\Gamma(j+1))$, with $d \in (-\infty, 1/2), d \neq 0, -1, -2, ...$ As usual, the Gamma function is defined by $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ if $x > 0$ and is analytically continued to the negative real axis by $\Gamma(x+1) = x\Gamma(x)$ with poles at $x = 0, -1, -2, ...$ These $I(d)$ processes are called fractionally integrated due to their construction by summing an uncorrelated noise process, Z_t. Consistent with the integration idea, $I(d)$ processes can also be viewed as a series that, when differenced d times, results in a white noise process. This relationship is given by:

$$Z_t = \nabla^d Y_t = (1-B)^d Y_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j} \tag{2}$$

where $\pi_j = \Gamma(j-d)/(\Gamma(j+1)\Gamma(-d))$. Here, π_j can be derived by viewing $(1-B)^d$ by means of the binomial expansion. A similar derivation yields the expression for ψ_j.

The above forms the basis for the statistical treatment of long-memory processes, and can be found, for example, in **[2]**,**[6]** and **[7]**. This construct is the discrete time analogue to the fractional Gaussian noise in **[1]**.

The power spectral density (PSD) for a fractionally integrated process, Y, is given by

$$S_Y(\omega) = \frac{\sigma^2}{2\pi|4\sin^2(\omega/2)|^d}, \quad |\omega| \leq \pi, \tag{3}$$

$$\approx \frac{\sigma^2}{2\pi|\omega|^{2d}}, \quad |\omega| \leq \delta < \pi, \tag{4}$$

where δ defines an interval of low frequencies such that the small angle approximation $\sin(x) \approx x$ holds, and σ^2 is the variance of the white noise process in (1). The PSD can be derived (see, for example, **[7]**) by treating Y_t as a linear filter on white noise.

For stationary $I(d)$ processes, the correlation between elements Y_t and Y_{t+h} for lag h is defined as follows **[7]**. The autocovariance function of an $I(d)$ process as in (1) with $-1/2 < d < 1/2$ is given by

$$\gamma(h) = \frac{(-1)^h\Gamma(1-2d)\sigma^2}{\Gamma(h-d+1)\Gamma(1-h-d)}, \quad h = 0, 1, 2, 3, ... \tag{5}$$

Thus, the memory of a stationary $I(d)$ process can be described simply as a function of the memory parameter, d, and lag, h. It is known **[2]** that a long-memory process, $Y \sim I(d)$, is stationary for $-1/2 < d < 1/2$, and that **[5]** the k^{th} difference (for integer k), denoted $\nabla^k Y$, is a fractionally integrated process of order $d-k$, namely $\nabla^k Y \sim I(d-k)$.

Power-law noises, i.e. $1/f^\alpha$ noises, are so-named due to the shape of their power spectral density. Similarly for long-memory processes, an indicator of long-term persistence is the shape of the spectral density of these processes as $f \to 0$. Like the $1/f^\alpha$ processes, $I(d)$ processes exhibit a power-law PSD shape near zero frequency, namely $S_Y(f) \sim 1/f^\alpha$. The $I(1/2)$ process is identified **[2]** with the flicker process having spectral density $1/f$. Similarly, the $I(d)$ models for other $1/f^\alpha$ noise processes can be identified by choosing $d = \alpha/2$, as seen in Table 1, where Y is in the fractional frequency domain. A proof of this relationship between α and d is available in **[8]**. Therefore, by employing the fractionally integrated noise construct, the traditional clock model can now be restated.

THE $I(d)$ ATOMIC CLOCK MODEL

Consider a sequence of observations of the phase differences between an atomic clock and a reference clock. Let Y be the corresponding process in the fractional frequency domain. A new representation of the clock model, written in matrix notation, is

$$Y = X\beta + \epsilon, \quad \epsilon = \sum_{i=1}^{k} \epsilon_i, \quad \epsilon_i \sim I(d_i), \quad \epsilon_i \perp \epsilon_j, \tag{6}$$

where $Y = (y(t_1), y(t_2), ..., y(t_n))'$ is the vector of observations, X is the design matrix, β is the vector of parameters, and ϵ is the vector of additive noise with independent components, ϵ_i. Here, the $n \times 2$ matrix X with rows $(1, t_i)$ and the 2×1 parameter vector, $\beta' = (b_0, b_1)$, corresponds to a linear model. The additive noise process, $\epsilon(t)$, is modeled as a sum of k independent fractionally integrated noises with, potentially, k different distributions. For tractability, it is assumed that each ϵ_i has a Normal distribution, namely N(0,$\Sigma_i = {\sigma_i}^2 \Psi_{d_i}$).

Table 1: Identification of α and d

Noise Type	α in $S_Y(f) \sim 1/f^{\alpha}$	d in $Y \sim I(d)$
White Phase	-2	-1
Flicker Phase	-1	$-1/2$
White Frequency	0	0
Flicker Frequency	1	1/2
Random Walk Frequency	2	1

ESTIMATION

Estimation of a linear trend in the presence of additive noise, a classic problem in statistics, is most frequently approached by a least-squares technique. The most prevalent such technique, ordinary least squares (OLS), is only optimal when the additive noise is serially uncorrelated. When the additive noise is serially correlated, the estimated OLS regression coefficients, $\hat{\beta}$, are still unbiased, but no longer have the minimum variance property, and may be quite inefficient. Additionally, the estimates of both σ^2 and the variance of $\hat{\beta}$ may seriously underestimate the true variances; therefore, confidence intervals and results of hypothesis tests for β are no longer reliable. Given that the clock model is not limited to uncorrelated additive noise, application of OLS must be made with great care.

The remainder of this work describes a prewhitening process which allows the application of the OLS technique even in the presence of the long-memory correlations typical of atomic clock measurements. By modeling and removing these strong correlations, a residual process which is uncorrelated (white) can be obtained. It is shown below that the prewhitening process removes only the noise autocorrelation, and does not perturb the underlying deterministic structure (clock rate and drift) that needs to be estimated. The subsequent application of OLS is not only appropriate (since the input noise process is uncorrelated) but also yields the best estimators of clock rate and drift in both a maximum-likelihood and mean-squared-error sense. The prewhitening approach described below is applicable to both stationary and non-stationary noise processes.

PREWHITENING

Recall the filter given in the definition of fractional integration (1) and consider forming an inverse filter. The operation opposite fractional integration is termed fractional differencing. Define the fractional difference of a discrete data set as follows. For a finite data set $Y_1, Y_2, ..., Y_n$, its d^{th} fractional difference is the set $\nabla^d Y_1, \nabla^d Y_2, ..., \nabla^d Y_n$ where each $\nabla^d Y_t$ is given by $\nabla^d Y_t = \sum_{j=1}^{k} \pi_j Y_{t-j} = \sum_{j=1}^{k} \Gamma(j-d)/(\Gamma(j+1)\Gamma(-d))Y_{t-j}$, where k is a suitably chosen constant. It can be determined via simulation studies which value of k is adequate for the removal of long-memory structure in synthetic data. It is easily seen that the fractional difference operator applied to model (1) prewhitens the noise process when d is correctly chosen. In general, however, d is unknown. Numerical procedures, such as the Hildreth-Lu technique **[9]**, are available for estimating d via a search over several possible prewhitening filters. An often-used criterion for selecting the best prewhitening filter is that of minimizing the sum of the squared errors. In the present application, however, the goal is to identify a data transformation that results in white additive noise; thus, the criterion for selecting the best prewhitening filter focuses upon minimizing the intradependence of the additive noise process. The best prewhitening filter, P, as a function of d, is given by:

$$P_d = \operatorname{argmin}(\sum_{h>0} \hat{\gamma}_e^2(h)) \tag{7}$$

where e are the residuals obtained from OLS applied to the fractionally differenced data (with respect to d), and $\hat{\gamma}_e^2(h)$ is the sample autocovariance function of the residuals. Here, P_d is chosen such that the difference between the autocovariance of the OLS residuals and the autocovariance of white noise (i.e., 0) is minimized across all non-zero lags. Simulation studies **[5]** reveal that d can be reliably estimated by this minimization technique. It is held that the estimated value of d shall not fluctuate with time.

Fractional differencing, as in (2), is an exact inverse of fractional integration only in the case of a single $I(d)$ noise. The fractional difference prewhitening approach proposed above is an approximate technique useful for removing long-memory structure in the case of the composite $I(d_i)$ clock model. It is also the case that the fractional difference transformation can be decoupled from the underlying linear function. That is, when $Y = X\beta + \epsilon$ is transformed via $\nabla^d Y = \nabla^d(X\beta + \epsilon)$, ∇^d acts upon $X\beta + \epsilon$ in a predictable way which can be undone to reveal β in the original units **[5]**. Consider the data set $Y = (y_1, y_2, ..., y_n)'$, which is linearly related to the time vector $(t_1, t_2, ..., t_n)'$ by

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = m \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{pmatrix} + \begin{pmatrix} b \\ b \\ \vdots \\ b \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}. \tag{8}$$

Fractionally differencing both sides yields

$$\nabla^d \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = m\nabla^d \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{pmatrix} + \nabla^d \begin{pmatrix} b \\ b \\ \vdots \\ b \end{pmatrix} + \nabla^d \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \tag{9}$$

which effectively prewhitens the ϵ process while preserving the slope, m. The intercept, b, is, however, perturbed since $\nabla^d b = b\sum_{j=0}^{k} \pi_j$, which requires a simple division to recover the original intercept, b. Thus, the fractional differencing transformation does not interfere with the estimation of clock rate and drift, as both can be recovered after employing the transformation.

Simulation studies show that the value of the fractional differencing parameter, d, can be reliably estimated via the fractional difference transformation described above, with minimal bias and scatter. Table 2 shows the results of 500 simulations for each of five values of d. The average value of $\hat{d}$, maximum deviation from the true value, and standard deviation of $\hat{d}$ are shown. In all cases, the minimization procedure searched for $\hat{d}$ in the range (-1,1). It can be seen that a slight tendency to underestimate d exists. However, the bias is not large and is likely preferable to a tendency to overestimate d. Nonetheless, bias reduction techniques could be investigated in future research.

Table 2: Estimation of the Memory Parameter, d

True d	Average $\hat{d}$	Max $\lvert\hat{d} - d\rvert$	Standard Deviation of $\hat{d}$
.01	.01	.12	.02
.11	.08	.14	.05
.21	.19	.13	.04
.31	.29	.15	.04
.41	.39	.19	.04

WHITENESS OF RESIDUALS

The true test of the performance of the prewhitening transformation is the extent to which the residual process is uncorrelated. Simulation studies were conducted in two cases to test the whiteness of the residuals. First, simple $I(d)$ processes were prewhitened and, second, a sum of two $I(d_i)$ processes was prewhitened using a single $\hat{d}$ in the fractional difference transformation. The results are discussed below.

In the first set of simulations for simple $I(d)$ noise, 500 simulations were run to estimate d and prewhiten the data. In order to determine if the prewhitened data are indeed white, one may use the property that the asymptotic distribution of the sample autocorrelation function, $\hat{\rho}(h)$, tends to the Normal distribution with mean 0 and variance $1/n$ as the number of points in the data set, n, tends to infinity. That is,

$$\hat{\rho}(h) \;\rightarrow\; N(0, 1/n). \tag{10}$$

Therefore, by comparing the sample autocorrelation function to the 95% confidence limits from the Normal distribution, one can test for whiteness of the prewhitened data set. In simulation studies, application of this technique resulted in the conclusion that the data were successfully prewhitened approximately 85% of the time. This is, unfortunately, less than the 95% expected by chance, indicating that improvements can still be made to the prewhitening algorithm. Nonetheless, the procedure yields white noise in the majority of cases and is certainly superior to *no* prewhitening.

In the simulations when the input process is a sum of two $I(d_i)$ processes, the results are again encouraging. White noise is obtained in approximately 81% of the cases; thus, the fractional differencing transformation has successfully prewhitened the data a majority of the time. It is worth noting that in the case of a sum of fractionally integrated processes (as in clock noise), the fractional difference transformation is *not* analytically an exact inverse filter. It is by virtue of the minimization of the autocorrelation structure that the transformation yields white noise. And although the use of fractional difference prewhitening is simply a convenient approximate technique in this case, the results are quite good and provide a means for OLS estimation of clock rate and drift when no such estimation is otherwise available.

CONCLUSIONS

The fractional difference prewhitening transformation has been proven through simulation studies to be a viable technique when the process noise is long-memory in nature. It is also useful as an approximate technique when the additive noise is comprised of a sum of long-memory processes. Estimates of the memory parameter, d, are found to be reliable and the prewhitened process is verified to be "white" in a large majority of the cases simulated. Therefore, the application of ordinary least squares is justified and the resulting confidence intervals for the regression coefficients are reliable, allowing the application of standard hypothesis testing and/or statistical process monitoring techniques.

REFERENCES

[1] B. Mandlebrot and J. W. Van Ness, 1968, **SIAM Review, 10**, 422-437.

[2] J. R. M. Hosking, 1981, **Biometrika, 68**, 165-176.

[3] D. Percival, 2003, **Metrologia**, to appear.

[4] L. Schmidt, 2003, **Metrologia**, to appear.

[5] L. S. Schmidt, 2001, *Estimation in the presence of fractionally integrated noise*, Ph.D. dissertation, American University, Washington, D.C.

[6] C. W. J. Granger and R. Joyeux, 1980, **Time Series Analysis, 1**, 15-29.

[7] P. J. Brockwell and R. A. Davis, 1991, **Time Series: Theory and Methods**, 2nd ed. (Springer, Heidelberg).

[8] J. Geweke and S. Porter-Hudak, 1983, **Journal of Time Series Analysis, 4**, 221-238.

[9] J. Neter, M. H. Kutner, J. Nachtsheim, and W. Wasserman, 1990, **Applied Linear Regression Models**, 3rd ed. (Irwin, Chicago).

QUESTIONS AND ANSWERS

STEVEN HUTSELL (Second Space Operations Squadron): Excellent work, Lara, as always. I had a question on what your thoughts were in terms of applications. Do you see that this is more applicable towards batch processing, or recursive real-time processing systems, or both?

LARA SCHMIDT: I have it experimentally set up where it runs automatically, and right now I have it set up to help me identify breakpoints in where you might suspect a model change over this region to another region for a single clock. I am still evaluating that, and we are looking into it. But eventually I hope this can run automatically. It is just a matter of finding the memory parameter that is appropriate, and you can run that iteratively or you can fix it and have it always use the same grammar.

JUDAH LEVINE (National Institute of Standards and Technology): How do you tell the difference between the fractionally pre-whitened data and a time series that's really just not stationary?

SCHMIDT: Yes, that is a hard call. When you look at the autocorrelation function itself, as I put up early in the talk, if you are using a short-memory process, and that is what you mindset is for modeling, it looks like a nonstationary short-memory process. So you have to know something about the kind of time series you are dealing with and, since we are assuming it is the power law noise spectrum, that we could have nonstationarity or stationarity; you just have to go with the shape of the autocorrelation function.

DAVE HOWE (NIST): Lara, have you looked at the effects of periodicities? Because I think that goes to the question that Judah was asking.

SCHMIDT: Right. No, I have not simulated any periodic data. All of my data have just been long-memory, short-memory, or white.