THE INFLUENCE OF A SINGLE-TONE SINUSIOD OVER HURST ESTIMATORS

Süleyman Baykut [1], Tolga Esat Özkurt [2], Melike Erol [3] and Tayfun Akgül [1]

[1]Departments of Electronics and Communications Engineering, [2]Computer Science, and [3]Computer Engineering
Istanbul Technical University, 34469, Maslak, Istanbul, Turkey
phone: + (90) 212 285 3568, fax: + (90) 212 285 3565
e-mail: baykut@ehb.itu.edu.tr, web: www.ehb.itu.edu.tr/~baykut

ABSTRACT

Hurst parameter is a common measure of the degree of self-similarity. Its estimation is an important issue where numerous methods have been proposed in the past and is still being proposed. In particular, estimation becomes intricate when the data contains periodicity, trend, noise, etc. These factors considerably affect the accuracy of Hurst parameter estimators. In this paper, we focus on periodicity and explore the behavior of three estimation methods under additive periodic component. For comparison, we choose a time domain estimator, Higuchi's Method, a frequency domain estimator, Wavelet Based Method and finally, a recently proposed, eigen domain estimator, Principal Component Analysis Based Method. We derive the analytical expressions for each of these estimators considering a *1/f* signal with a single tone sinusoid buried into it. We also verify our results with simulations using a single tone sinusoid added to fractional Brownian motion (fBm) trace. We show that the magnitude and the frequency of the periodic component clearly affect the Hurst estimation.

1. INTRODUCTION

Hurst parameter, H, is used to characterize the correlation structure of Statistically Self-Similar (SSS) processes. Its importance arises from self-similar processes' being popular models for many natural and artificial man-made data. SSS processes are distributionally invariant along scales i.e., their statistical behaviour along scales does not change. An SSS random process $x(t)$ has the scaling form given below:

$$x(t) \overset{p}{=} a^{-H} x(at) \tag{1}$$

where a is a positive real constant, ^{p} denotes statistical equivalence and H is the so-called Hurst parameter in the interval $(0,1)$ [1].

Self-similar processes have an alternative interpretation on the frequency domain regarding their spectral behavior. They obey the so-called power-law relationship and as a consequence, they are also called *1/f* processes,

$$S_x(\omega) \approx \frac{\sigma_x^2}{|\omega|^{\gamma}} \tag{2}$$

where $S_x(\omega)$ is the empirical power spectrum of $x(t)$, σ_x^2 is the variance, ω is the angular frequency and γ is the spectral exponent [1].

1/f processes are frequently modeled both by fractional Brownian motion (fBm) and its incremental version, namely, fractional Gaussian noise (fGn). It is relevant to mention that they are the most widely preferred models of *1/f* processes. Both processes are normally distributed, zero-mean, while fBm is nonstationary and fGn is stationary [2].

Considering Eq. (1) and Eq. (2), the connection between γ and H is shown to be $\gamma=2H+1$ for fBm and $\gamma=2H-1$ for fGn [2].

The significance of Hurst parameter has made its estimation an attractive topic [3]. Especially, estimation of H, for real data –which may include inferences like periodicity, noise, trend etc., requires a careful study. Periodic components within a self-similar process are observed in some signal processing applications such as computer networks [4, 5]. For instance, in a recent study [4], it is mentioned that a periodic component may mislead the estimators.

In this paper, we study the behavior of estimators for *1/f* data mixed with a periodic component, in a systematic manner. We use three estimators namely, Higuchi's Method (HM), Wavelet Based Method (WBM) and Principal Component Analysis Method (PCAM). First, we derive explicit expressions for these estimators in Section 2 and next we present our simulation results using data sets with a single tone sinusoid buried into fBm in Section 3. We summarize our conclusions in Section 4.

2. INFLUENCE OF PERIODICITY ON THE ESTIMATION METHODS

In this section, we give a brief review on the H estimators and we derive explicit expressions to show how they are affected when a periodic component exists within a *1/f* process.

2.1 Higuchi's Method

HM, a time domain estimation method, uses the length of the fractal path to estimate H parameter [6]. To show this, let us consider a *1/f* process, $x(t)$, which satisfies the equation below [7]:

$$E\{\Delta x(t)^2\} = c\Delta^{2H} \tag{3}$$

where c is a constant, Δ is the lag, $E\{.\}$ denotes the expectation operator and $\Delta x(t) = x(t+\Delta) - x(t)$. The logarithm of both sides of Eq. (3) yields:

$$2\log(E\{|\Delta x(t)|\}) = \log c + 2H\log(\Delta) \tag{4}$$

Clearly, when $E\{|\Delta x(t)|\}$ versus Δ is plotted logarithmically, the slope of the straight line yields the H parameter.

For the investigation of the influence of periodicity, we consider a signal,

$$y(t) = x(t) + s(t) \tag{5}$$

where $x(t)$ is a zero-mean *1/f* process and $s(t)$ is a single tone sinusoid with amplitude A_c and a particular angular frequency ω_c:

$$s(t) = A_c \sin(\omega_c t) \tag{6}$$

By substituting $E\{\Delta y(t)^2\}$ in Eq. (3) we obtain:

$$\begin{aligned} E\{\Delta y(t)^2\} &= E\{(y(t+\Delta) - y(t))^2\} \\ &= E\{\Delta x(t)^2\} + E\{\Delta s(t)^2\} + 2E\{x(t)\}E\{s(t)\} \end{aligned} \tag{7}$$

Since $s(t)$ and $x(t)$ are assumed to be zero mean signals, the last term of Eq. (7) is omitted. The impact of periodicity is observed only by the term $E\{\Delta s(t)^2\}$:

$$E\{\Delta s(t)^2\} = A_c^2 E\{(\sin(\omega_c t + \omega_c\Delta) - \sin(\omega_c t))^2\} \tag{8}$$

Using simple trigonometric properties Eq. (8) reduces to:

$$E\left\{\Delta s(t)^2\right\} = A_c^2 - A_c^2 \cos(\omega_c \Delta) \tag{9}$$

and hence Eq. (7) can be rewritten as:

$$E\left\{\Delta y(t)^2\right\} = c\Delta^{2H} + A_c^2 - A_c^2 \cos(\omega_c \Delta) \tag{10}$$

which shows that the effect of periodicity is dependent on the lag (Δ) value, the amplitude (A_c) and the frequency (ω_c) of the sinusoid. When ω_c is small enough, $cos(\omega_c\Delta)$ approaches to “1” and $E\{\Delta s(t)^2\}$ reduces to zero, i.e., the periodic component does not have any influence on the estimated H value. However, for higher ω_c values, $cos(\omega_c\Delta)$ oscillates between -1 and +1, i.e., $E\{\Delta s(t)^2\}$ changes periodically between $2A_c^2$ and zero depending on the lag. This oscillatory behavior can be observed in the regression plots of Section 3. (See: *Figures 3.a and 4.a.*)

2.2 Wavelet Based Method

WBM is a popular frequency domain method which can be used to estimate H from the progression of the variances of the wavelet coefficients since they should follow the power law relationship given in Eq. (2) for any $1/f$ process. In other words, applying the wavelet transform to a $1/f$ process, $x(t)$, whose spectrum has a spectral exponent γ, then the wavelet coefficients x_n^m will be mutually uncorrelated, zero-mean, Gaussian random variables with variances:

$$Var x_n^m = \sigma^2 2^{-\gamma m} \tag{11}$$

By taking the logarithm and fitting a straight line to the variance regression, the slope provides an estimate of γ.

In order to explore the influence of a single tone sinusoid on this method, we consider Eq. (5) once again where the wavelet coefficients of $y(t)$, y_n^m, can be calculated by projecting the orthogonal wavelet basis onto $y(t)$:

$$y_n^m = \int_{-\infty}^{\infty} y(t)\psi_n^m(t)dt \tag{12}$$

where $\psi_n^m(t)$ is the normalized dilations (m) and translations (n) of the wavelet basis function $\psi(t)$ (mother wavelet):

$$\psi_n^m(t) = 2^{m/2}\psi(2^m t - n) \tag{13}$$

Eq. (12) can be rewritten as:

$$y_n^m = \int_{-\infty}^{\infty} x(t)\psi_n^m(t)dt + \int_{-\infty}^{\infty} s(t)\psi_n^m(t)dt \tag{14}$$

$$y_n^m = x_n^m + s_n^m$$

where x_n^m and s_n^m are the wavelet coefficients of $x(t)$ and $s(t)$, respectively; and in practice y_n^m are obtained by the filter-and-sample operation using the following filter-bank [1]:

$$y_n^m = \left\{y(t) * \psi_0^m(-t)\right\}\Big|_{t=2^{-m}n} \tag{15}$$

Here, $\psi_0^m(t)$ is the impulse response of a bandpass filter at the m^{th} scale, and “ * ” denotes the convolution operation. The wavelet coefficients are then the estimates of the frequency components of the signal in a particular frequency region. In theory, there are infinite number of scales, but in practice, since the data is truncated at length N, there remains only $M = log_2(N)$ available scales. The magnitude frequency responses of the bandpass filters $|\Psi_0^m(\omega)|$, $m=1, 2, \ldots, M$, are given in *Fig. 1* for Daubechies’ wavelets of order 5. The bandwidths of these filters depend on the scale m which corresponds to the frequency band $2^{m-M-1}\pi < \omega < 2^{m-M}\pi$.

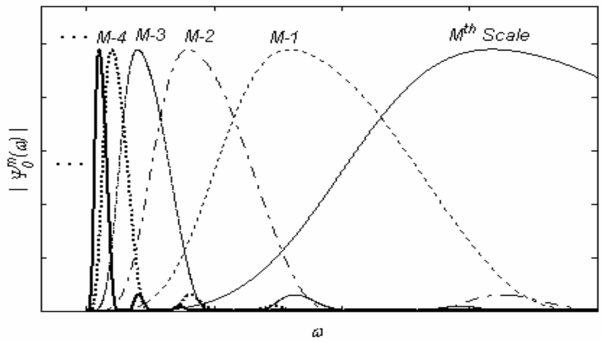

Figure 1: The magnitude frequency responses of the bandpass filters for Daubechies’ wavelets of order 5.

It is apparent that if ω_c is within the frequency band corresponding to m_c^{th} scale, the sinusoidal component will cause a jump in the variance of wavelet coefficients:

$$var(y_n^m) = \begin{cases} var(x_n^{m_c}) + var(s_n^{m_c}) \ , & m = m_c \\ var(x_n^m) & , m \neq m_c \end{cases} \tag{16}$$

However, in practice, since the filters in the filter bank are not ideal, there will be significant spectral overlaps which can be reflected as the spectral leakage (See: *Fig. 1*) affecting the variances of the wavelet coefficients at scales nearby m_c.

2.3 Principal Component Analysis Method

The use of PCA method for the estimation of the H parameter has been recently proposed [8]. This method relies on the eigen-analysis of a discrete-time sequence through its autocorrelation matrix $\boldsymbol{R}$:

$$\boldsymbol{R}\boldsymbol{\varphi_i} = \lambda_i \boldsymbol{\varphi_i} \tag{17}$$

where λ_i’s are eigenvalues and $\boldsymbol{\varphi_i}$’s are eigenvectors of $\boldsymbol{R}$. It can be shown that the progression of the eigenvalues (sorted in decreasing order) versus indices yields,

$$\lambda_i \approx i^{-\gamma} \tag{18}$$

This expression is analytically proven for $H=0.5$ in [8], and later on for $H < 0.5$ in [9]. Then the estimation method simply relies on the extraction of the spectral exponent γ using a log-log plot.

In order to investigate the effect of periodicity on the PCA method, we consider the autocorrelation function of $y(t)$ of Eq. (5):

$$R_Y(t_1,t_2) = R_X(t_1,t_2) + R_s(t_1,t_2) \tag{19}$$

where the autocorrelation function of $s(t)$ corresponds to:

$$R_S(t_1,t_2) = \frac{A_c^2}{2}\left\{cos(\omega_c(t_1 - t_2)) - cos(\omega_c(t_1 + t_2))\right\} \tag{20}$$

The continuous time version of the PCA method is known as the Karhunen-Loéve expansion, i.e.,

$$\int_{-T/2}^{T/2} R_y(t_1,t_2)\varphi(t_2)dt_2 = \lambda\varphi(t_1) \tag{21}$$

for $-T/2 < t_1$, $t_2 < T/2$. Here, $\varphi(t)$’s are orthonormal eigenfunctions corresponding to the eigenvectors and T denotes the length of the process. Assuming that the eigenfunctions of Eq. (21) are complex exponentials [9]:

$$\varphi_i(t) = exp(j\phi_i t), \ \phi_i = \frac{2\pi}{T} i \tag{22}$$

we obtain the expression for the Karhunen-Loéve expansion of the sinusoidal component as:

$$\int_{-T/2}^{T/2} R_s(t_1,t_2)e^{j\phi_i t_2}dt_2 = \frac{A_c^2}{2}\left\{2\pi e^{-j\phi_i t_1}\left[\delta(-\omega_c-\phi_i)+\delta(-\omega_c+\phi_i)\right]+ 2\pi e^{j\phi_i t_1}\left[\delta(-\omega_c+\phi_i)+\delta(\omega_c+\phi_i)\right]\right\} \quad (23)$$

where its influence is observed only when $\varphi_i = \omega_c$. Then Eq. (21) becomes:

$$\int_{-T/2}^{T/2} R_x(t_1,t_2)e^{j\phi_i t_2}dt_2 + 2\pi A_c^2 \cos(\omega_c t_1) = \lambda_i e^{j\phi_i t_1} \quad (24)$$

which shows that only the i^{th} eigenvalue is affected by the periodic component when i=c. Here, the index c is the integer part of $\omega_c L_{max}/2\pi$, where L_{max} denotes the maximum lag value. Since the eigenvectors are not ideally complex exponentials, the influence of the periodic component can be witnessed along the neighboring indices. This may be seen in the regression plots presented in Section 3.

3. SIMULATION RESULTS

In this section, we present the simulation results where we use fBm to model the *1/f* processes. We use a wavelet-based algorithm to generate the fBm traces of a fixed length $N = 4096$ for various H. A trace of fBm with $H = 0.6$ is shown as an example in *Fig. 2.a*.

We add single tone sinusoids (with relatively low $\omega_c = \pi/128$ and high $\omega_c = \pi/16$ frequency components) to the fBm traces. Note that, we consider two separate mixtures where Signal-to-fBm ratio (SNR) is -1 dB and -2 dB. A trace of fBm+Sinusoid with $H = 0.6$, $\omega_c = \pi/128$ and SNR = -1dB is shown in *Fig. 2.b*.

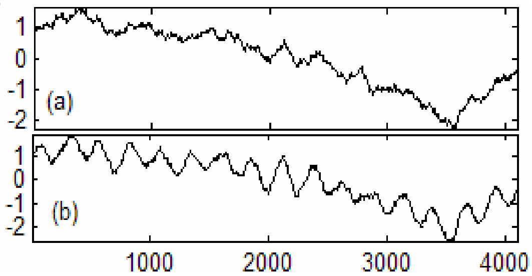

Figure 2: (a) Artificial fBm data ($H = 0.6$), *(b)* fBm + Sinusoid (SNR = -*1*dB, $\omega_c = \pi/128$)

The regression plots of HM, WBM, and PCAM for the fBm (H=0.6) and fBm+Sinusioid signals (SNR = -1dB, $\omega_c = \pi/16$) are given in *Fig.3.a, b, c*, respectively. Clearly, they are consistent with the theoretical results derived in Section 2. For example, the result of HM, where we observe the effect of additive sinusoidal component scattered through all lags periodically, is given in *Fig.3.a*.

The regression plot in *Fig.3.b* corresponds to the result of WBM where we observe a jump at scale 9. This scale corresponds to the frequency band *(π/16-π/8)* which includes ω_c (frequency of the added sinusoid). We also notice a slight jump in the neighboring 8th and 10th scales.

In *Fig.3.c*, the PCAM result is plotted where the eigenvalues with indices between 2 and 30 are observed to be affected. The actual location for the sinusoidal component is at index $i = 9$.

In *Fig.4,* we show the regression plots when the frequency of the sinusoid is changed to $\omega_c = \pi/128$.

In *Fig.4.a* the same periodic scattering is observed as in *Fig.3.a* for HM. However, the oscillation in the regression plot has a lower frequency compared to the previous one.

In *Fig.4.b*, for WBM, a similar jump is observed but this time at scales 5 and 6. Although the SNR value is the same for both *Fig.3.b* and *Fig.4.b,* the jump is smaller as expected.

In *Fig.4.c,* for PCAM, the indices of the affected eigenvalues are shifted towards 2 and 8.

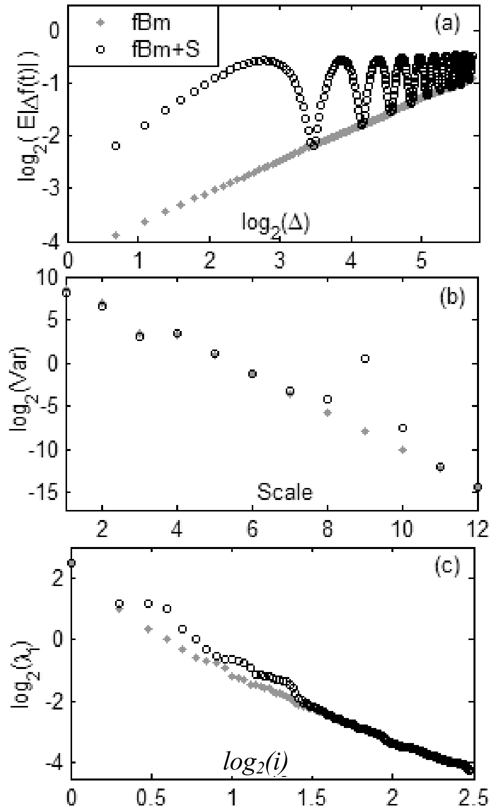

Figure 3: The regression plots of fBm (H = 0.6) and fBm+S signals (SNR = -1dB, $\omega_c = \pi/16$) *(a)* HM ($\hat{H}_{fbm+S} = 0.1408$), *(b)* WBM ($\hat{H}_{fbm+S} = 0.5150$), *(c)* PCAM ($\hat{H}_{fbm+S} = 0.6969$).

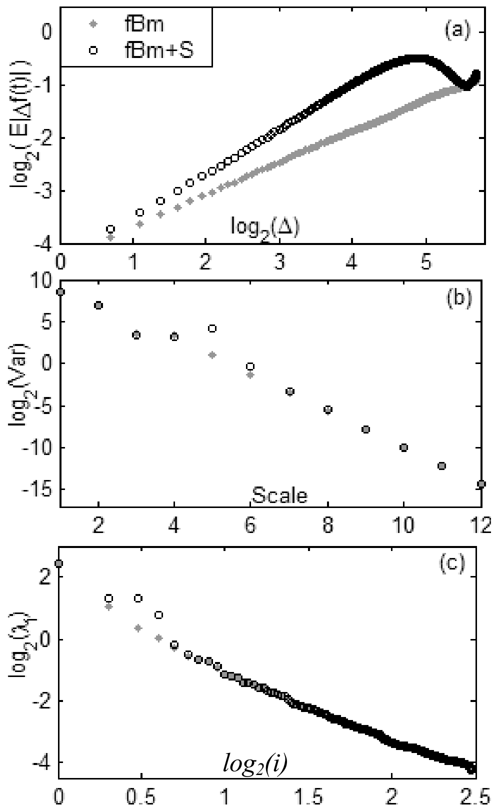

Figure 4: The regression plots of fBm (H = 0.6) and fBm+S signals (SNR = -1dB, $\omega_c = \pi/128$) *(a)* HM ($\hat{H}_{fbm+S} = 0.4081$), *(b)* WBM ($\hat{H}_{fbm+S} = 0.7534$), *(c)* PCAM ($\hat{H}_{fbm+S} = 0.6096$).

Both theoretical and empirical results indicate that the effect of periodicity is apparent over all lags for HM, whereas it is bursty and localized for WBM and PCAM.

HM is more sensitive to the periodic mixture than the other two methods. Note that, for HM, when the maximum lag is chosen relatively small, the oscillatory behaviour may be seen at the far end of the regression plot or it may not be observed at all. However, this does not mean that H estimation is not influenced.

Note that, in practice, all of the methods use a statistically appropriate region of regression plots in line-fit. In our simulations, we choose the maximum lag = 300 for HM and PCAM to be nearly *8%* of the data length N. In WBM, the lower scales (1-4) are omitted.

The regression plots are similar for other H values but they are not included here due to space limitations.

In *Table 1*, for the three methods: HM, WBM, and PCAM, the mean of the estimated H values ($\hat{H}$) for *100* fBm traces are given for the cases: SNR = -1dB, -2dB and $\omega_c = \pi/128, \pi/16$.

The estimated H values of HM are highly affected by the sinusoidal component. The higher the SNR or the higher the frequency may cause misleading HM results.

WBM is also influenced to some degree, however PCAM seems to be the least affected method.

Table 1: Average-Estimated H values ($\mu_{\hat{H}}$) of fBm and fBm + Sinusoid for 100 realizations ($N = 4096$).

			fBm+Sinusoid			
			$\omega_c = \pi/128$		$\omega_c = \pi/16$	
			SNR -2dB	SNR -1dB	SNR -2dB	SNR -1dB
	Theoretical H	(fBm) $\mu_{\hat{H}}$	$\mu_{\hat{H}}$	$\mu_{\hat{H}}$	$\mu_{\hat{H}}$	$\mu_{\hat{H}}$
HM	0.1	0.1421	0.1423	0.1430	0.1395	0.1197
	0.3	0.3033	0.3020	0.2922	0.2871	0.2051
	0.5	0.4777	0.4713	0.4299	0.4034	0.2325
	0.6	0.5640	0.5457	0.4797	0.4494	0.2308
	0.7	0.6505	0.6163	0.5056	0.4947	0.2491
	0.8	0.7310	0.6885	0.5333	0.4781	0.2142
	0.9	0.7902	0.7280	0.5267	0.4959	0.2233
WBM	0.1	0.0733	0.0734	0.1014	0.0720	0.0652
	0.3	0.2727	0.2778	0.3157	0.2680	0.2547
	0.5	0.4735	0.4837	0.5439	0.4583	0.4333
	0.6	0.5737	0.5888	0.6615	0.5480	0.5126
	0.7	0.6741	0.6975	0.7919	0.6358	0.5893
	0.8	0.7755	0.8072	0.9240	0.7127	0.6509
	0.9	0.8807	0.9390	1.0946	0.7909	0.7104
PCAM*	0.1	0.1160	0.1165	0.1238	0.1212	0.1396
	0.3	0.3215	0.3228	0.3324	0.3342	0.3607
	0.5	0.5095	0.5135	0.5271	0.5340	0.5722
	0.6	0.5691	0.5668	0.5825	0.6008	0.6469

*Only the results for $H \leq 0.6$ are included in the table since PCAM is considered as an unreliable estimator outside this interval [9].

4. CONCLUSION

In this paper, we derive the analytical expressions for three Hurst estimation methods and present the simulation results for fBm data sets mixed with a single tone sinusoid. We show that the SNR and the frequency of the periodic component are the key factors affecting the Hurst estimation for all three methods. Analytical and empirical results dictate that for HM, periodicity has an overall oscillatory effect over all lags, meanwhile for WBM and PCAM, this influence is local and bursty (appears to be a jump in the corresponding scales or indices.) We observe that the H estimates are more influenced by periodicity for HM whereas for WBM and PCAM they are less affected.

REFERENCES

[1] G. W. Wornell, "Wavelet-Based Representations for the *1/f* Family of Fractal Processes," *Proc. of IEEE*, vol. 81, no. 10, pp. 1428-1450, 1993.

[2] B. B. Mandelbrot, J. W. Van Ness, "Fractional Brownian Motions, Fractional Noise and Applications," *SIAM Review*, vol. 10, no. 4, pp. 422-437, 1968.

[3] M. S. Taqqu, V. Teverovsky, W. Willinger, "Estimators for Long-range Dependence: An Empirical Study," *Fractals*, vol. 3, no. 4, pp 785-798, 1995.

[4] T. Karagiannis, M. Faloutsos, and R.H. Riedi, "Long-Range Dependence: Now You See It, Now You Don't!," *Proc. IEEE Global Telecommunications Conf. Global Internet Symp*., 2002.

[5] A. Montanari, V. Teverovsky, M. S. Taqqu, "Estimating Long-range Dependence in the Presence of Periodicity: an Empirical Study", *J. of Mathematical and Computer Modelling*, vol. 29, no.10/12, pp. 217-228, 1999.

[6] T. Higuchi, "Approach to an Irregular Time Series on the Basis of the Fractal Theory," *Physica D*, vol. 31, no. 2, pp. 277-283, 1988.

[7] D. G. Manolakis, V. K. Ingle, S. M. Kogon, *Statistical and Adaptive Signal Processing*, McGraw Hill, 2000.

[8] J. B. Gao, Y. H. Cao, J. M. Lee, "Principal Component Analysis of *1/f* Noise," *Phys. Lett. A,* 314, pp. 392-400, 2003.

[9] T. E. Özkurt, T. Akgül, "Is PCA Reliable for the Analysis of Fractional Brownian Motion?", *Proc. of European Signal Processing Conference (EUSIPCO),* 2005.