IS PCA RELIABLE FOR THE ANALYSIS OF FRACTIONAL BROWNIAN MOTION?

Tolga Esat Özkurt [1], Tayfun Akgül [2]

Department of Computer Science[1], Department of Electronics and Communications Engineering[2]
Istanbul Technical University, Maslak, Istanbul, Turkey
[1]phone: + (90) 212 285 6056-114, fax: + (90) 212 285 7073,
email: tolgaesat@be.itu.edu.tr, web: www.be.itu.edu.tr/~tolgaesat

ABSTRACT

Estimation of the self-similarity parameter, also known as Hurst (H) parameter, is an important issue. In this paper, we study one of the H parameter estimation methods, namely the Principal Component Analysis (PCA) and show that this method may not give reliable results for the persistent part ($H>0.5$) of the fractional Brownian motion. Moreover, when the results are unreliable, the eigenvalue progression seriously deviates from linearity. Thus, with a linear-fit error threshold, one can comment on the reliability for the results of the PCA method.

1. INTRODUCTION

In order to model fractal time series and hence the *1/f* processes, the fractional Brownian motion (fBm) is suggested as one of the most popular, simple, nonstationary and normally distributed processes. Such signals are observed in diverse fields i.e., meteorology, biomedicine, finance and computer networks.

In literature, there are several methods suggested to estimate Hurst parameter (H) which is a significant parameter that characterizes the fBm sequences. Among them, the Principal Component Analysis (PCA) has recently been proposed as an efficient H estimator for *1/f* processes [1]. PCA is known to be a useful and popular tool in signal and image processing applications such as dimension reduction, signal enhancement and face recognition. In our study, we show that PCA has different behaviour for antipersistent ($H < 0.5$) and persistent ($H > 0.5$) parts of *1/f* processes. It is observed that this estimation method may give unreliable results for the persistent part of the process. Moreover, since the time series we deal with are not ideal fBm's (they have lower and upper scale limits); the choice of the lag may have an effect on the accuracy of the estimation.

In this work, we propose an approach to explore whether the PCA based estimator gives a valid result or not; and comment on the lag parameter for eigenanalysis of fBm processes.

This paper is organized as follows. In Section 2, we briefly introduce the PCA of *1/f* processes and describe our main contribution. In Section 3, we give simulation examples to show the performance of the estimator and the linear-fit error is proposed as a criterion of validation. Finally, we include our conclusion in Section 4.

2. PRINCIPAL COMPONENT ANALYSIS OF THE FRACTIONAL BROWNIAN MOTION DATA

The PCA method relies on the eigenanalysis of the following relation:

$$\mathbf{R}\boldsymbol{\varphi}_i = \lambda_i \boldsymbol{\varphi}_i, \qquad i=1,2,\ldots,N \tag{1}$$

where λ_i's are eigenvalues and $\boldsymbol{\varphi}_i$'s are the corresponding eigenvectors of the autocorrelation matrix $\mathbf{R}$ of the discretized process of $x(t)$ with length N.

In general the fBm processes are considered in three different regions as $0 < H < 0.5$, $H = 0.5$ and $0.5 < H < 1$ [2]. Note that, as H increases, the process becomes more regular. In [1], only for $H = 0.5$, the analytical result is provided while some experimental study is supplied for the other regions.

The Karhunen-Loéve expansion is the analog version of PCA and hence uses autocorrelation function instead of the autocorrelation matrix. Its fundamental equation is given as [3] :

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

where $\varphi(t)$'s are orthonormal eigenfunctions corresponding to the eigenvectors. Autocorrelation function $R(t_1,t_2)$ of an fBm process $x(t)$ is given as [2]:

$$R(t_1,t_2) = \frac{1}{2}\sigma_H^2\left\{|t_1|^{2H} - |t_1-t_2|^{2H} + |t_2|^{2H}\right\} \tag{3}$$

where σ_H^2 is defined as

$$\sigma_H^2 = E\{x(1)^2\} = \Gamma(1-2H)\frac{\cos(\pi H)}{\pi H} \tag{4}$$

Here Γ (.) is the well-known gamma function.

Using this expansion and the autocorrelation function of the standard Brownian motion* (Bm) (in this case $H=0.5$), the relationship between the eigenvalues λ_i and their indices i (sorted in decreasing order) can be analytically shown as [3,4]

$$\lambda_i \approx i^{-(2H+1)} \tag{5}$$

If we consider the eigenfunctions as complex exponentials

*Brownian motion is the version of fBm where $H=0.5$. It is also the cumulated sum of white Gaussian noise.

$$\varphi_i(t) = exp(j\omega_i t)\ ,\ \ \omega_i = \frac{2\pi}{T} i \tag{6}$$

and substitute Eq. (3) into Eq. (2) and then apply the Karhunen-Loéve expansion, after some algebra, we obtain

$$\lambda_i \omega_i = \sigma_H^2 H \left\{ \int_0^\infty 2\sin(\omega_i t) t^{2H-1} dt \right\} \tag{7}$$

This equation is convergent only for $0< H <0.5$ and the solution of the integration yields :

$$\lambda_i \omega_i = \frac{\omega_i^{-2H} H\pi}{\Gamma(1-2H)\cos(\pi H)} \sigma_H^2 \tag{8}$$

This can be easily transformed into a simple expression between the sorted eigenvalues and indices of the eigenvalues :

$$\lambda_i = \frac{1}{\omega_i^{2H+1}} \approx i^{-(2H+1)} \tag{9}$$

Note that the integration diverges for $0.5< H < 1$.

Finally through a logarithmic linear-fit of Eq. (9), the H parameter can be easily estimated.

3. EXPERIMENTS

In this section, we present the performance of the PCA based method using synthetic and real data and give a criterion to determine whether the estimator provides reliable results or not.

3.1 Spectral Synthesis

In the synthesis method [5], first, in order to obtain the magnitude spectrum $|X(\omega)|$, we simply construct the power-law relationship for *1/f* processes, i.e.,

$$|X(\omega)| = \frac{\sigma_x^2}{|\omega|^{(\frac{2H+1}{2})}} \tag{10}$$

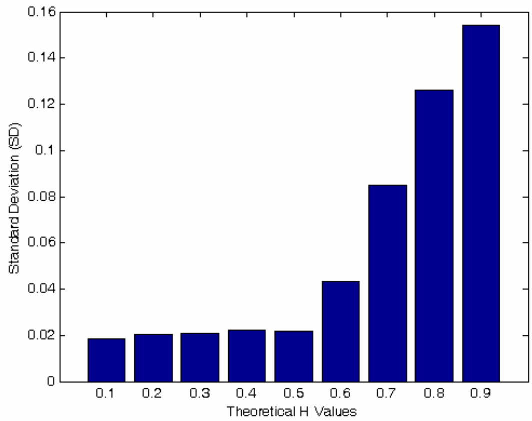
where $x(t)$ is the corresponding process, σ_x^2 is the variance and ω is the angular frequency. Then the phase ξ is generated by uniformly distributed random numbers between [-π, π]. Finally, by taking the inverse Fourier transform of $|X(\omega)| exp(j\xi)$, the corresponding fBm sequence is synthesized.

3.2 Monte Carlo Simulations for $0 < H <1$

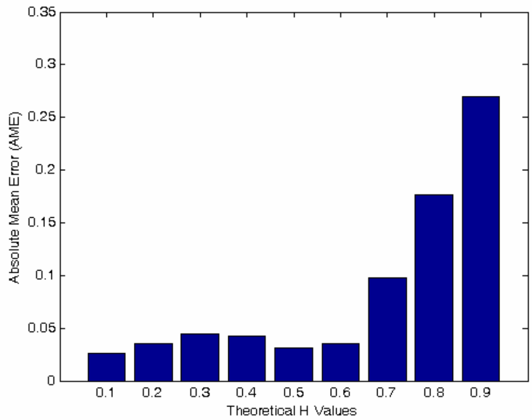
A set of fBm data ($K=100$) having lengths N_1=1024 and N_2=8192, meanwhile H is between 0.1 and 0.9 with increments of 0.1, has been synthesized by the spectral synthesis method. After estimating the H values by the PCA method, the absolute mean error (AME) and the standard deviation (SD)

$$AME = \frac{1}{K}\sum_{k=1}^{K} |H - \hat{H}_k|$$
$$SD = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(\hat{H}_k - \bar{H}_k)^2} \tag{11}$$

are computed for each data set where $\hat{H}_k$ is the estimated value and $\bar{H}_k$ is the mean value of the estimates. Performance of the estimator can be observed by the behaviour of the AME and SD plots given in Fig. 1 and Fig. 2. For example, in Fig 1 (a) and (b), the SD and AME plots indicate that the method yields inaccurate estimates when $H > 0.5$ for N_1=1024.



(a)



(b)

Fig. 1. (a) Absolute mean errors of the estimated *H* values by PCA of fBm (*N*=1024), **(b)** Standard deviation of the estimated *H* values

In our simulations, the lag value is chosen as 256 and hence the dimension of the autocorrelation matrix is 256 by 256. This is a reasonable choice for such experiments. From the AME results for $N_2=8192$, given in Fig 2, it is clear that the accuracy of the estimates does not change significantly with the increase of the length.

3.3 Criterion for Validation

Linear progression of logarithmic eigenvalues gives us a clue on the reliability of the estimator. That is why we measure the linear-fit error by summing the squares of differences of logarithmic eigenvalues and the fitted line. We then normalize it by the lag value. When we follow this procedure, especially for $H > 0.5$, if the fit error is greater than a threshold (according to our observations the threshold value is around 0.1), the estimator is considered to be unreliable. In Fig. 3 and Fig. 4, examples are given for synthetic data. It is observed that although the theoretical H is 0.8 for both Fig. 3 and Fig. 4, the linear-fit error increases for the invalid estimation given in Fig. 4.

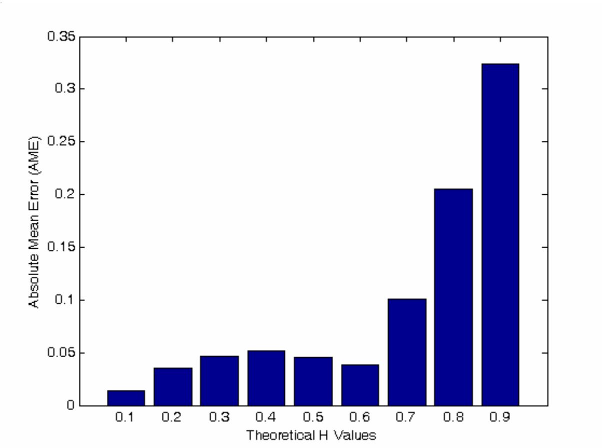


Fig. 2. Absolute mean errors of the estimated H values by PCA of fBm (N=8192)

We apply the same criterion for real meteorological data. The daily average wind speed data sets gathered by two separate stations in the Republic of Ireland are tested [6]. We construct the wind speed data by subtracting the mean value and then take the cumulative sum in order to analyze the data's fractal properties. We then repeat the identical procedure stated above for the synthetic data. Although the exact H parameter is unknown, we estimated it (by using various methods like Higuchi's or Wavelet-based) as approximately 0.8. As shown in Fig 5, the PCA estimation for this parameter is as 0.8392 while the fit error is under 0.1. In Fig 6, it is observed that the estimated H is 0.4893 which does not match with the other estimators' results of 0.8. Note that the fit error is greater than 0.1 for this invalid estimation.

4. CONCLUSION

For the cases $0< H <0.5$ and $H=0.5$, the PCA method can be used to estimate the Hurst parameter of fBm processes. This is also verified from the derivations using the Karhunen-Loéve expansion of fBm processes. However, the experiments show that the accuracy of the PCA method is questionable when $0.5< H <1$. We suggest to check whether the progression of eigenvalues follows the power-law relation or not, simply by calculating a linear fit-error. That is because we observe the following: When the fit-error is greater than a threshold value (experimentally 0.1), the estimation results may be unreliable for both real and synthetic data.

REFERENCES

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

[2] B. B. Mandelbrot, J. W. Van Ness, "Fractional Brownian Motions, Fractional Noise and Applications," *SIAM Review*, 10(4), pp. 422-437, 1968.

[3] H. L. L. Van Trees, *Detection, Estimation and Modulation Theory: Part I*, John Wiley & Sons, 1968.

[4] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw Hill, 1991.

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

[6] J. Haslett, A. E. Raftery, "Space-time Modelling with Long-memory Dependence: Assessing Ireland's Wind Power Resource", *Applied Statistics*, 38, pp. 1-50, 1989.

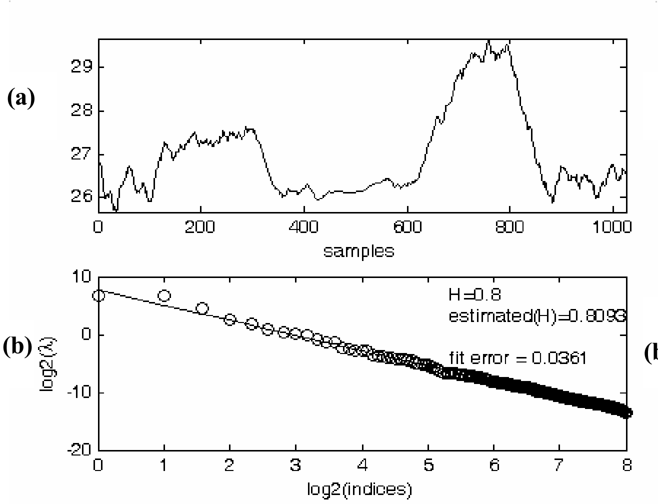


Fig. 3. (a) Synthetic fBm data for *H=0.8*, **(b)** Logarithmic eigenvalue progression, $\hat{H}$ *=0.8093, error=0.0361*

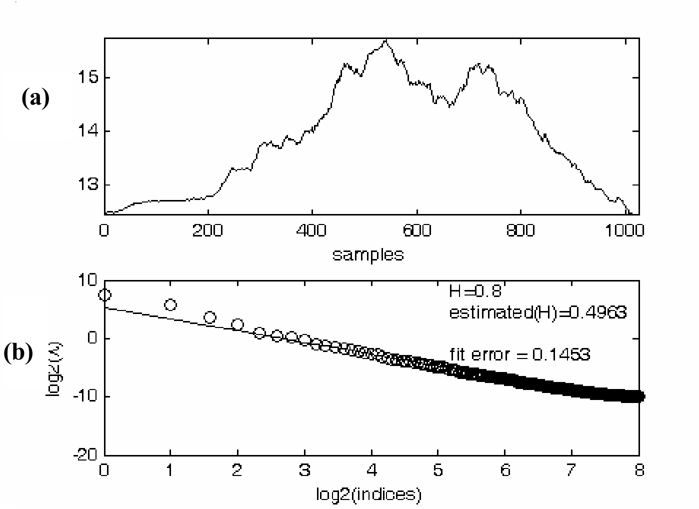


Fig. 4. (a) Synthetic fBm data for *H=0.8*, **(b)** Logarithmic eigenvalue progression, $\hat{H}$ *=0.4963, error=0.1453*

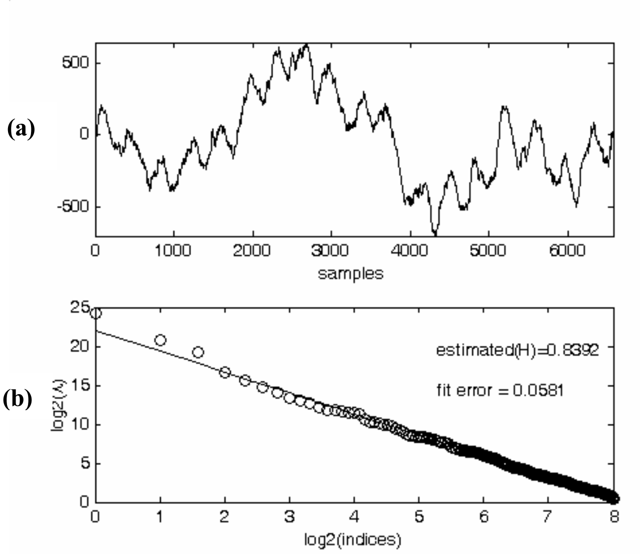


Fig. 5. (a) Daily wind speed data for the first station, **(b)** Logarithmic eigenvalue progression, $\hat{H}$ *=0.8392, error=0.0581*

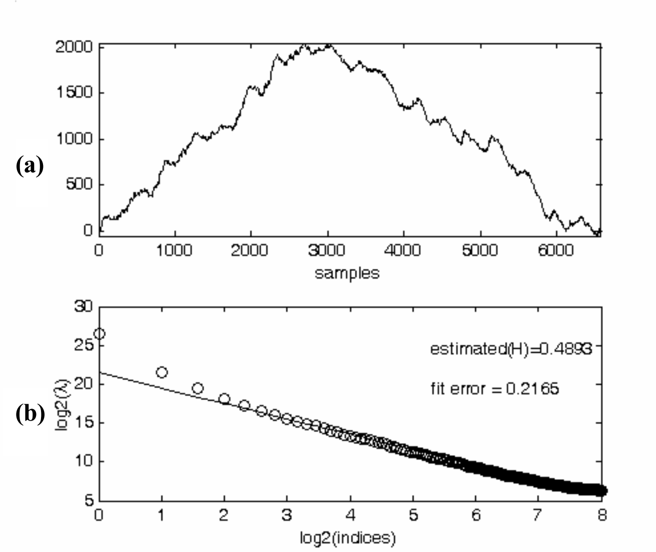


Fig. 6. (a) Daily wind speed data for the second station, **(b)** Logarithmic eigenvalue progression, $\hat{H}$ *=0.4893, error=0.2166*