Using detrended fluctuation analysis for lagged correlation analysis of nonstationary signals

Jose Alvarez-Ramirez, Eduardo Rodriguez,* and Juan Carlos Echeverria
División de Ciencias Basicas e Ingenieria, Universidad Autonoma Metropolitana-Iatapalapa, Apartado Postal 55-534, Mexico D.F. 09340, Mexico
(Received 14 October 2008; published 20 May 2009)

Here we propose a method, based on detrended fluctuation analysis (DFA), to investigate lagged correlations for nonstationary time series. The aim is to show that the largest correlation can be found at positive lags, reflecting the existence of underlying delays in the evolution of real time sequences. The performance of the lagged DFA method is illustrated by selected real examples.

DOI: 10.1103/PhysRevE.79.057202 PACS number(s): 05.45.Tp

Stochastic time series from real systems may contain hidden correlations due the complex interaction of diverse mechanisms. The characterization of these correlations is of prime importance for gaining insights into the nature of such mechanisms and for developing suitable models for simulation and forecasting purposes. A distinctive feature of many time series is their nonstationarity, which hampers the usage of classical statistical methods mainly based on a stationarity assumption. To confront this situation, different methods have been proposed in recent decades to be applied for nonstationary processes. The traditional method for assessing scaling correlations is the R/S analysis [1]. However, the results from R/S analysis can be strongly affected by the presence of nonstationarities and long-run trends. The detrended fluctuation analysis (DFA) [2] has emerged as an important alternative for nonstationary time series, finding acceptation and application in diverse fields, including heart rate dynamics [3], neuron spiking [4], long-time weather records [5], financial time-series [6], expressway traffic flow [7], earthquake data [8], heart rate variability in fetuses [9], and many more cases. Examples of interesting DFA applications for time series that are not obviously stationary are seismic dynamics and spatial distributions [10], fluctuations of the North-Atlantic hurricane frequencies [11], behavioral sequences of wild chimpanzees [12], and characterization of electroencephalography as a measure of anesthesia depth [13]. Several recently introduced methods for the detection of long-range correlations in data series were compared based on similar ideas as the well-established DFA, indicating the ability of such methods for revealing different time and frequency features of complex time series [14]. Extensions of the DFA for high-dimensional analysis have also been proposed and applied for image scaling studies [15]. Summing up, the DFA is a widely use method for characterizing different aspects (e.g., scaling, correlations, etc.) of complex nonstationary time series.

An assumption inherited from classical statistical methods is that scaling correlations of time series are always presented without delay. That is, it is assumed that the correlations should be found for zero lags because the correlation function is a monotonic nonincreasing and positive-definite function of the lag. But this is not the case for time series generated autorecursively in the form $x_i=\phi(x_{i-\theta},x_{i-\theta-1},\ldots)$, $\theta\geq 1$. In this case, the time series $\{x_i\}$ should present maximum correlation for the particular lag θ. In this work, we modify the traditional DFA method to explore the existence of maximum correlations for nontrivial lags. The lagged DFA follows the same steps as the traditional method but the detrended fluctuations are computed by a lagged convolution. Four examples will be used to illustrate the functioning of this method and to show that the location of a maximum correlation at nontrivial lags can be related to the underlying physics of the system.

The DFA is a method designed to estimate long-range correlations for nonstationary sequences [2]. The reader is referred to the Physionet internet site [16] for a detailed description of the method and a list of publications showing the use of DFA to characterize correlations. In fact, the proposed DFA modifications presented here to account for lagged causality effects can be described as follows. Let us suppose that x_k is a series of length N and that this series is of compact support. The support is defined as the set of the indices k with nonzero values x_k. The value of $x_k=0$ is interpreted as having no value at this k. The lagged DFA is a straightforward modification of the standard DFA [2] and is implemented following the next steps:

(i) Step 1. Determine the profile,

$$Y_i=\sum_{k=1}^{i}[x_k-\langle x\rangle],\quad i=1,\ldots,N.$$

For convenience, define the profile $Z_i=Y_{i+\theta}$, $i=1,\ldots,N-\theta$. Note that Y_i is lagged θ times from Z_i.

(ii) Step 2. Divide the profiles Y_i and Z_i into $N_s=\text{int}[(N-\theta)/s]$ nonoverlapping segments of equal length s.

(iii) Step 3. Calculate the local trends $\tilde{Y}_{(v-1)s+i}$ and $\tilde{Z}_{(v-1)s+i}$ for each of the N_s segments by a least-squares fit of the series. Then determine the lagged variance

$$V_v(s;\theta,q)^q=\frac{1}{s}\sum_{i=1}^{s}|Y_{(v-1)s+i}-\tilde{Y}_{(v-1)s+i}|^{q/2}|Z_{(v-1)s+i}-\tilde{Z}_{(v-1)s+i}|^{q/2}. \quad (1)$$

The trends $\tilde{Y}_{(v-1)s+i}$ and $\tilde{Z}_{(v-1)s+i}$ can be computed from a linear, quadratic, or higher order polynomial fit of the profile for each segment. Note that by taking $N_s=\text{int}[(N-\theta)/s]$ nonoverlapping segments in step 2, the values of $Z_{(v-1)s+i}$ are well defined for all $i\in[1,s]$ and all $\theta\geq 0$.

*Corresponding author; rofe@xanum.uam.mx

(iv) Step 4. Average over all segments to obtain the qth order fluctuation function,

$$F(s;\theta,q) = \left\{ \frac{1}{N_s} \sum_{v=1}^{N_s} V_v(s;\theta,q)^q \right\}^{1/q}. \qquad (2)$$

(v) Step 5. Determine the scaling behavior of the fluctuation functions by analyzing the log-log plots $F(s;\theta,q)$ versus s for each value of q. If the series x_i is long-range power-law correlated, $F(s;\theta,q)$ increases, for large values of s, as a power law

$$F(s;\theta,q) \sim s^{\alpha(\theta,q)}. \qquad (3)$$

The scaling exponent $\alpha(\theta,q)$ depends on the parameter q to account for multifractal effects and on the lag θ to account for any delayed causality. If $\alpha(\theta,q)$ is not dependent on q, the signal $\{x_i\}$ is associated with a monofractal process. If for a given q the scaling exponent $\alpha(\theta,q)$ is maximum at $\theta_{\max} \geq 1$, the time series involves a type of delayed feedback effect. If $\theta=0$ and $q \in R$, the above procedure is reduced to the standard DFA method [17]. As it is well known, the time series $\{x_i\}$ is noncorrelated if $\alpha(\theta,q)=0.5$, persistent if $\alpha(\theta,q)>0.5$ and antipersistent if $\alpha(\theta,q)<0.5$.

If $\theta=0$, one has that $Z_i \equiv Y_i$, and the procedure described above reduces to the standard DFA. In this form, the lagged DFA can be seen as a generalization of existing detrended fluctuation analysis schemes to account for the existence of maximum correlations for a positive lag θ. In order to clarify how the method operates, several selected real signals will be considered. In this initial contribution, four single time series are considered to illustrate the fact that maximal autocorrelations can be found at nonzero lags. For simplicity in the presentation, linear fits were used to compute the linear trend for each box.

Logistic map. The logistic map $x_i = \rho x_{i-1}(1-x_{i-1})$, $x_i \in [0,1]$, displays chaotic behavior for values of ρ close to 4. It is clear that the signal x_i has a first-order recursiveness, so one can expect a maximum correlation degree for $\theta=1$. For $q=2$, $\rho=3.9$, and 10^5 samples, Fig. 1(a) displays the fluctuation function $F(s;\theta,2)$ for $\theta=0$ and $\theta=1$, showing that the scaling exponent is larger for $\theta=1$. For $\rho=3.9$ and $\rho=3.95$, Fig. 1(b) depicts the scaling exponent for different values of the lag θ, showing a maximum scaling exponent at $\theta_{\max}=1$. This result confirms that with the lagged DFA is possible to recover the delayed recursiveness of single signals.

Pseudorandom number generators. Portable pseudorandom number generators are based on high-order recursive procedures. In principle, an exact random number sequence should be free of correlations, so that $\alpha(\theta,q)=0.5$ for all lag $\theta \geq 0$. We have tested three pseudorandom number generators that are commonly used for, e.g., Monte Carlo simulation purposes. These generators are the RAN1 function taken from the Numerical Recipes© book [18], the RAND function of the MATLAB™ 7.0 package, and the RAN function of the Compaq Visual Fortran™ package. The results are displayed in Fig. 2 for 10^5 samples, showing that, despite having the zero-lag scaling exponent $\alpha(0,2) \approx 0.5$ as expected, significative deviations from the noncorrelated behavior is observed for nonzero lags. For instance, a maximum autocorrelation degree is found at about $\theta_{\max}=12$ for the Visual Fortran Package™ generator, yet it becomes apparent that this generator outperforms the other two because the scaling exponent shows the smallest deviations from 0.5 for all lag values. Error bars in Fig. 2 are only shown for one case, indicating that the error bars for the other cases have similar behavior.

Heart rate variability. As a third example for lagged autocorrelations, heart rate fluctuations from presumably healthy subjects are considered. The data, consisting of the

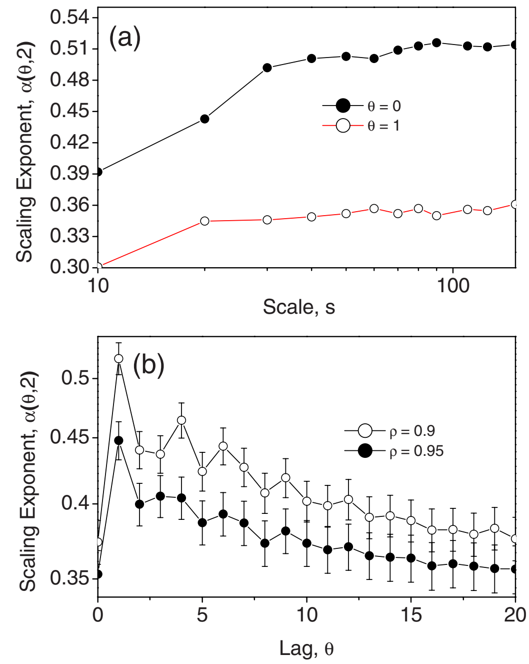

FIG. 1. (Color online) (a) Local scaling exponent for lags $\theta=0$ and $\theta=1$, showing that a scaling region for scales larger than about $s=20$. (b) Scaling exponent as a function of the lag θ for two different values of the logistic map's parameter ρ. The error bars were computed from the standard deviation. Note that the maximum correlation is preserved at $\theta=1$ for the two parameter values ρ.

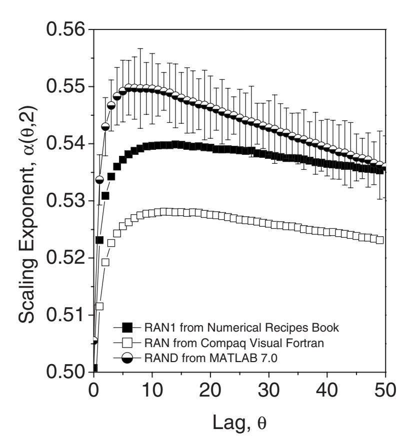

FIG. 2. Lagged scaling exponent for three different pseudorandom number generators. In principle, an accurate pseudorandom number generator should provide $\alpha(q,\theta) \approx 0.5$ for all $q \in R$ and $\theta \in R$. The results in this figure show that a certain correlation degree can be found for nonzero lags.

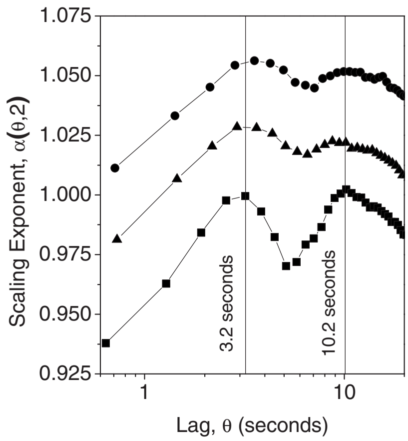

FIG. 3. Lagged scaling exponent of the heart rate variability for three different healthy subjects. Observe the maximum correlations at lags of about 3 and 10 s. Presumably, these lags can be related to the delays in the action of the sympathetic branch of the autonomic nervous system.

beat-to-beat dynamics during 6 h, were extracted from the Physionet database [16]. It is well known that the heart rate variability for healthy subjects exhibits scaling behavior with scaling exponent about $\alpha(0,2)=0.9-1.0$ [2]. Figure 3 shows the lagged DFA results for three different cases [23], where for convenience the horizontal scale was normalized by the individual mean heart rate, such that scales are given in seconds. It can be observed that the larger autocorrelations are not found at zero lag but for two positive lags. The cause of these maximum lagged autocorrelations may be explained by the underlying delays associated with the activity of the autonomic nervous system. Typically there is a 1–3 s delay before sympathetic activity begins to increase the heart rate, and there is a further 10–20 s before that effect becomes complete [19,20]. Since the maximum autocorrelation lags are located at about 3 and 10 s, it is probable that the lagged autocorrelation analysis recovers the delays associated with the sympathetic control action that is established, in principle, by the time constants of the norephinephrine release, vascular response, and dissipation of vascular effects [21]. Such consideration seems also supported by findings showing that a decreased vagal outflow with increased sympathetic activity strengthens the fractal correlation properties of heart rate variability data [22].

Dow Jones index daily changes. We analyze the daily closing values of the Dow Jones financial indices. We analyze the time series of absolute values of the differences of logarithms for successive days. Figure 4 shows the scaling exponent as a function of the lag θ for $q=2$ and for scales smaller than a year. One can see that the absolute index changes are correlated with scaling exponent of about 0.68 for $\theta=0$ in agreement with previous studies [6]. However, the largest correlation degree is found for lags about 6 business days with a scaling exponent up to 0.78. This suggests that the construction of forecasting models should account for delays in order to improve the prediction capability. That is, a better prediction performance can be obtained if the predicted absolute return, say r_t, is computed as a function of past delayed absolute returns $\{r_{t-\theta_{\max}}, r_{t-2\theta_{\max}}, \ldots\}$. The maximum lag $\theta_{\max}$ of about 5–7 days seems to correspond to the feedback effect of weekly cycles in the operation of the stock markets.

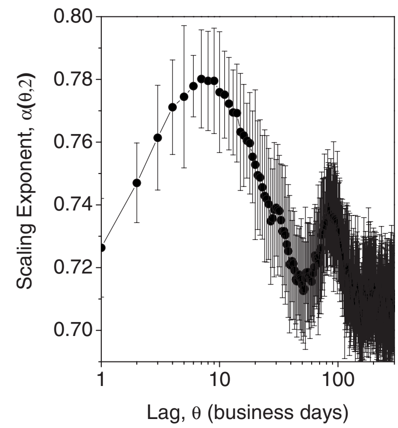

FIG. 4. Lagged scaling exponent for the daily Dow Jones absolute changes. The DFAs were carried out for scales smaller than a year and $q=2$. Note the maximum correlations at about 5–7 days.

North-Atlantic oscillations. The North-Atlantic oscillation (NAO) is a climatic phenomenon in the North-Atlantic Ocean reflecting fluctuations of the difference of sea-level pressure between the Icelandic low and the Azores high. Through east-west rocking motions of the Icelandic low and the Azores high, NAO affects the strength and direction of westerly winds as well as storm tracks across the North-Atlantic. Similar to the El Niño phenomenon in the Pacific Ocean, the NAO is one of the most important drivers of climate fluctuations in the North-Atlantic and surrounding humid climates [11]. The average difference in the pressure at the Iceland and Azores stations is known as NAO index, which is reported on a daily basis. The daily NAO index, from January 1, 1950 to June 30, 2008 (a time series of length 21 366), was extracted from the U.S. National Oceanic and Atmospheric Administration site [24]. Figure 5 presents the scaling exponent, showing that the larger autocorrelations are found for $\theta_{\max}=8$ days. This suggests that a kind of delayed feedback effect in the NAO evolution. The purpose of increasing the DFA parameter q is to magnify the

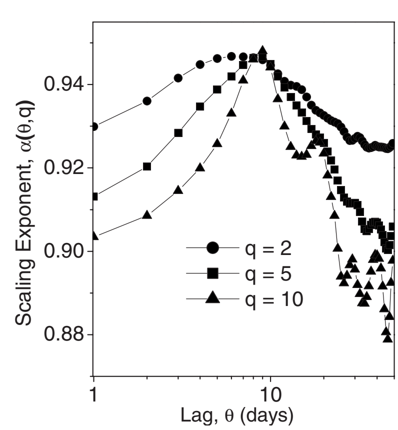

FIG. 5. Lagged scaling exponent for the NAO, showing that the largest autocorrelation is found for $\theta_{\max}=8$ days.

oscillation magnitude, such that larger fluctuations play a more important effect in the DFA computations. In this form, the results in Fig. 5 also display the scaling exponent for three different values of q, showing that $\theta_{\max}$ is unchanged for large NAO variations. Since the maximum correlations at $\theta_{\max}=8$ days are more pronounced for large values of q, the lagged correlations should be in particular linked to larger oscillations.

Summing up, the widely used DFA was adapted to explore the existence of lagged maximal correlation in nonstationary time series. The lagged DFA follows the same steps as the traditional method but the detrended fluctuations are computed as a lagged convolution. The application of the method for different examples, from physics to physiology, showed that the highest scaling exponents or maximum correlation can be found at nonzero lags. This suggests that the stochastic processes underlying the generation of such diverse time series may involve a lagged recursive procedure. That is, a signal $\{x_t\}$ must be generated from a recursive procedure of the form $x_t=\phi(x_{t-\theta_{\max}},x_{t-2\theta_{\max}},\ldots)$.

[1] R. Mandelbrot and M. Taqqu, Bull. Internat. Statist. Inst. **48**, 59 (1979).
[2] C. K. Peng *et al.*, Chaos **5**, 82 (1995); J. W. Kantelhardt *et al.*, Physica A **316**, 87 (2002).
[3] H. E. Stanley *et al.*, Physica A **270**, 309 (1999).
[4] S. Blesic *et al.*, Physica A **268**, 275 (1999).
[5] E. Koscielny-Bunde *et al.*, Phys. Rev. Lett. **81**, 729 (1998).
[6] Y. H. Liu *et al.*, Physica A **245**, 437 (1997).
[7] S. Tadaki *et al.*, J. Phys. Soc. Jpn. **75**, 034002 (2006).
[8] G. Afshar *et al.*, Geophysical Research Abstracts **9**, 04835 (2005).
[9] A. Kikuchi *et al.*, Gynecol. Obstet. Invest. **65**, 116 (2008).
[10] L. Telesca *et al.*, Chaos, Solitons Fractals **21**, 335 (2004).
[11] J. B. Elsner *et al.*, J Clim. **12**, 427 (1999).
[12] C. L. Alados and M. A. Huffman, Ethology **106**, 105 (2008).
[13] M. Jospin *et al.*, IEEE Trans. Biomed. Eng. **54**, 840 (2007).
[14] A. Bashan *et al.*, Physica A **387**, 5080 (2008).
[15] G. F. Gu and W. X. Zhou, Phys. Rev. E **74**, 061104 (2006).
[16] www.physionet.org/physiotools/dfa/citations.shtml
[17] J. W. Kantelhardt *et al.*, Physica A **316**, 87 (2002).
[18] W. H. Press *et al.*, *Numerical Recipes in C: The Art of Science Computing* (Cambridge University Press, New York, New York, 1990).
[19] E. Magosso *et al.*, Cardiovasc. Eng. **1**, 101 (2001).
[20] S. C. Malpas, Am. J. Physiol. Heart Circ. Physiol. **282**, H6 (2002).
[21] D. L. Eckberg, Ann. Med. **32**, 341 (2000).
[22] M. P. Tulppo *et al.*, Circulation **112**, 314 (2005).
[23] Similar behavior was obtained for the cases contained in the Physionet database.
[24] www.noaa.gov