Physica A 390 (2011) 614–618

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

DCCA cross-correlation coefficient: Quantifying level of cross-correlation

G.F. Zebende *

Computational Modeling Program - SENAI CIMATEC 41650-010 Salvador, Bahia, Brazil
Physics Department - UEFS 44036-900 Feira de Santana, Bahia, Brazil

ARTICLE INFO

Article history:
Received 4 August 2010
Received in revised form 3 October 2010
Available online 29 October 2010

Keywords:
Time series
Cross-correlation
DCCA method
DFA method

ABSTRACT

In this paper, a new coefficient is proposed with the objective of quantifying the level of cross-correlation between nonstationary time series. This cross-correlation coefficient is defined in terms of the DFA method and the DCCA method. The implementation of this cross-correlation coefficient will be illustrated with selected time series.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

There is a large number of situations, whether in the real world or not, where a given observable y_i is measured at successive time intervals, forming a time series $\{y_i\}$ [1]. Many of these time series are recorded simultaneously, and have the same length N. Some strategies for time series analysis have been developed. The most popular is the measurement of the Pearson correlation coefficient. However, this coefficient is not robust [2] and can be misleading if outliers are present, as in real-world data characterized by a high degree of nonstationarity [3]. Then, other statistical methods can be proposed for time series analysis [4,5]. However, if the time series exhibit complex behavior, such as self-affinity, we can characterize the auto-correlation (cross-correlation of the signal with itself) by power-laws [6,7]. In this way, we can identify universality for different kinds of problems [8,9]. Hurst was one of the first to identify a power-law in a real-world time series, specifically studying the Nile river and problems related to water storage, by R/S analysis [10]. He found that many records are very well described by the empirical relation $R/S \sim n^H$, where n is a time scale and H is the Hurst exponent. The values of the Hurst exponent range between 0.0 and 1.0. In terms of auto-correlations, $H = 0.5$ indicates that there is no auto-correlation in time series. If $0.5 < H < 1.0$, we have long-range auto-correlations with persistent behavior, and if $0.0 < H < 0.5$, we have long-range auto-correlations with anti-persistent behavior. Recently, many other papers have been investigated in order to investigate long-range auto-correlations in time series with new ideas, such as the DNA Walk method [11,12]. But an alternative method, Detrended Fluctuation Analysis (DFA), cited $\approx$ 1054 times, has been proposed to detect long-range auto-correlations embedded in a patch landscape and also avoids the spurious detection of apparent long-range auto-correlations [13]. The DFA method performs better than either the standard R/S analysis or the DNA walk in quantifying the scaling behavior of noisy signals for a wide range of correlations [14,15].

2. Discussion

The DFA method provides a relationship between $F_{\mathrm{DFA}}(n)$ (root mean square fluctuation) and the scale n, characterized for a power-law $F_{\mathrm{DFA}}(n) \propto n^{\alpha}$. In this way, α is the long-range auto-correlation scaling exponent, such that if $\alpha = 0.5$ the signal

* Corresponding author at: Computational Modeling Program - SENAI CIMATEC 41650-010 Salvador, Bahia, Brazil.
E-mail address: gfzebende@hotmail.com.

0378-4371/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2010.10.022

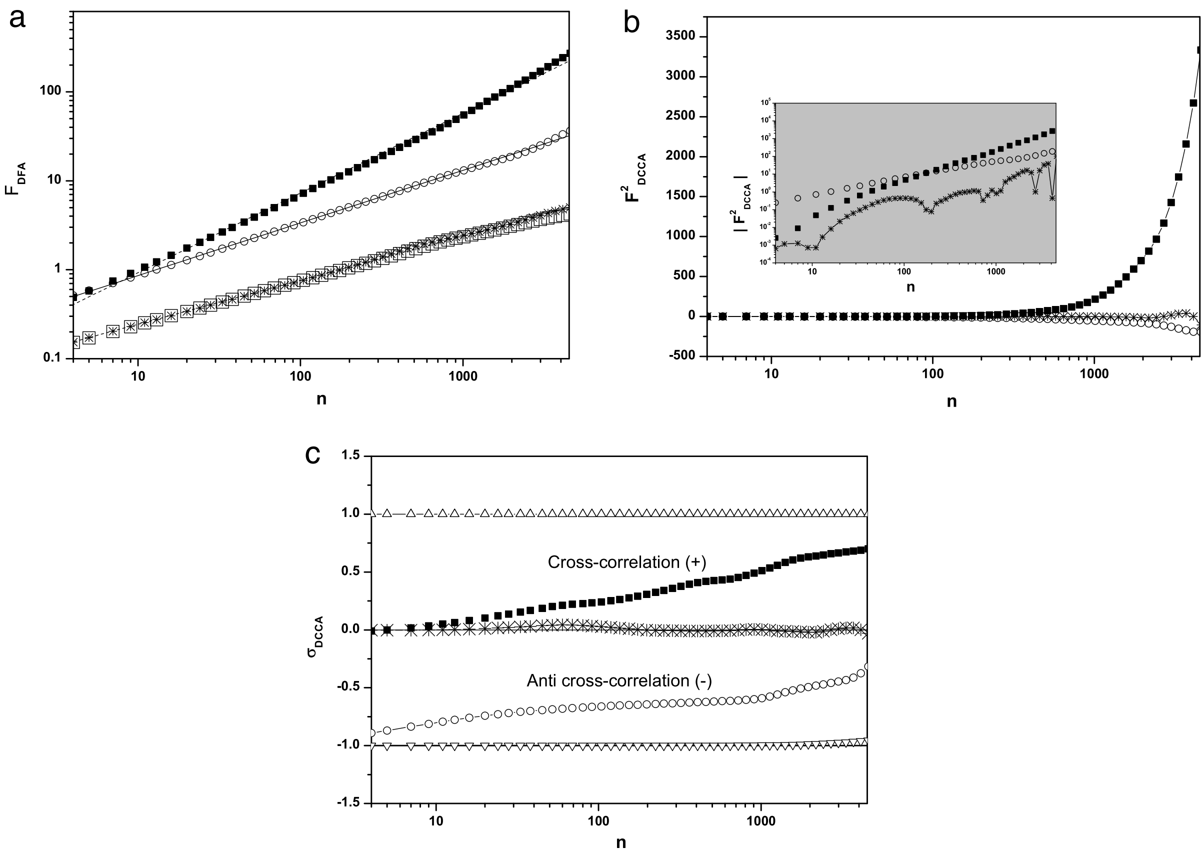

Fig. 1. Time series analysis in the simulated time series by the ARFIMA process. (a) Exhibit the DFA auto-correlation $F_{DFA} \times n$ in log–log scale for the original time series $\{y_{Rand1}\}$ ($*$), $\{y_{Rand2}\}$ ($\square$), and respectively their transformed, i.e., time series 1 with $\rho = 0.1$ ($\bigcirc$), and time series 2 with $\rho = 0.4$ ($\blacksquare$). These time series are very well fitted by power-laws $F_{DFA} \sim n^{\alpha}$, with $\alpha = 0.5$ (for the original random time series), $\alpha = 0.6$ (time series 1), and $\alpha = 0.9$ (time series 2). (b) Show the detrended cross-correlation covariance $F^2_{DCCA} \times n$ for: (i) $\rho = 0.1$ (time series 1), $\rho = 0.4$ (time series 2), with $W = 1.0$ and $\varepsilon_i \neq \varepsilon'_i$ ($*$), (ii) $\rho = 0.1$ (time series 1), $\rho = 0.4$ (time series 2), with $W = 0.85$, and $\varepsilon_i \neq \varepsilon'_i$ ($\blacksquare$), and (iii) for $\rho = -0.1$ (time series 1), $\rho = 0.4$ (time series 2), with $W = 0.80$ and $\varepsilon_i = -\varepsilon'_i$ ($\bigcirc$). Finally, (c) gives the DCCA cross-correlation coefficient σ_{DCCA} as a function of n for the time series analyzed in (b) case. The value $\sigma_{DCCA} = 1$ implies a perfect cross-correlation (upper solid line in the figure), $\sigma_{DCCA} = -1$ implies a perfect anti cross-correlation (lower solid line), and $\sigma_{DCCA} = 0$ implies that there is no cross-correlation (middle solid line).

is uncorrelated, if $\alpha < 0.5$ the signal is anti-persistent, and $\alpha > 0.5$ represents a persistent signal. Now if we have two time series, $\{y_i\}$ and $\{y'_i\}$, respectively, the study of cross-correlation between these time series can be applied. For example, in the case of cross-correlations in finance many papers have been analyzed [16–23]. However, we can apply a generalization of the DFA method, called Detrended Cross-Correlation Analysis method (DCCA) [24] to study cross-correlations between time series. In the DCCA paper, the authors illustrate the method by selected examples from physics, physiology, and finance, where they reported power-law cross-correlations in absolute values of logarithmic changes in price between Dow Jones and NASDAQ. In Ref. [25] the authors found power-law magnitude cross-correlations between absolute values of price changes and trading-volume changes. The DCCA method is designed to investigate power-law cross-correlations between different simultaneously recorded time series of equal length N, in the presence of nonstationarity, by means of the detrended covariance function $F^2_{DCCA}(n)$. If long-range cross-correlation appears between these two time series then $F_{DCCA} \sim n^{\lambda}$, with $\lambda \approx (\alpha_{DFA} + \alpha'_{DFA})/2$. In the case of cross-correlation between $\{y_i\}$ with itself, the detrended covariance function $F^2_{DCCA}(n)$ reduces to the detrended variance function $F^2_{DFA}(n)$ [24]. The λ exponent quantifies long-range power-law correlations and also identifies seasonality [26], but λ does not quantify the level of cross-correlations. Moreover, according Zho [27], there is no clear relationship between λ_{DCCA} and α_{DFA}. An immediate question arises, is it possible to quantify the level of cross-correlation with the DFA and the DCCA method? The answer is yes, and to discuss this question, in this paper a new coefficient is proposed. The DCCA cross-correlation coefficient is defined as the ratio between the detrended covariance function F^2_{DCCA} and the detrended variance function F_{DFA}, i.e.,

$$\sigma_{DCCA} \equiv \frac{F^2_{DCCA}}{F_{DFA}\{y_i\}} F_{DFA}\{y'_i\}. \tag{1}$$

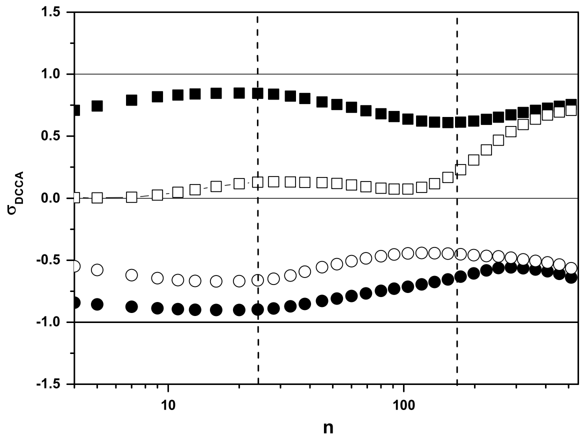

Fig. 2. DCCA cross-correlation coefficient for a simultaneous climate surface data record, collected every hour by the INMET meteorological station in the city of Salvador, Bahia (Brazil). These data were recorded from 12 March 2010 to 10 June 2010. In this figure we have $\sigma_{\text{DCCA}} = 1$ for: air temperature × solar radiation (■), air temperature × relative humidity (●), relative humidity × solar radiation (○), and wind speed × wind direction (□). Vertical dashed lines correspond to one day and one week respectively in time scale.

Table 1
σ_{DCCA} in terms of level of cross-correlation.

σ_{DCCA}	Condition
1	Perfect cross-correlation
0	No cross-correlation
−1	Perfect anti cross-correlation

Eq. (1) leads to a new scale of cross-correlation in nonstationary time series. Here, σ_{DCCA} is a dimensionless coefficient that ranges between $-1 \leq \sigma_{\text{DCCA}} \leq 1$. A value of $\sigma_{\text{DCCA}} = 0$ means there is no cross-correlation, and it splits the level of cross-correlation between the positive and the negative case (see Table 1).

3. Results

With the objective of testing the utility of the DCCA cross-correlation coefficient, selected time series, σ_{DCCA} as a function of n is presented below. Firstly, σ_{DCCA} was tested for simulated time series generated by using a two-component fractionally autoregressive integrated moving average (ARFIMA) stochastic process, in order to investigate power-law auto-correlations and power-law cross-correlations, in this case each variable depends not only on its own past, but also on the past values of the other variable [28,29],

$$y_i = W \sum_{n=1}^{\infty} a_n(\rho_1) y_{i-n} + (1-W) \sum_{n=1}^{\infty} a_n(\rho_2) y'_{i-n} + \varepsilon_i,$$
$$y'_i = (1-W) \sum_{n=1}^{\infty} a_n(\rho_1) y_{i-n} + W \sum_{n=1}^{\infty} a_n(\rho_2) y'_{i-n} + \varepsilon'_i. \tag{2}$$

Here, ε_i and ε'_i denote two independent and identically distributed (i.i.d.) Gaussian variables with zero mean and unit variance, $a_n(\rho)$ are statistical weights defined by $a_n(\rho) = \Gamma(n-\rho)/(\Gamma(-\rho)\Gamma(1+n))$, where Γ denotes the Gamma function, ρ are parameters ranging from −0.5 to 0.5 (related to the DFA exponent, $\alpha = 0.5 + \rho$ [28,29]), and W is a free parameter ranging from 0.5 to 1.0 and controlling the strength of power-law cross-correlations between y_i and y'_i. By using the two-component ARFIMA process of Eq. (2), we generate a new time series y_i and y'_i [30] characterized by different values of $a_n(\rho_{1,2})$ and W. In this sense, taking into account the ARFIMA process, was tested the value of σ_{DCCA} for extreme cases where there are (or not) cross-correlations between the time series, specifically where we have perfect cross-correlation, perfect anti cross-correlation, and no cross-correlation. Therefore, for $\rho = 0.1$ (time series 1), $\rho = 0.4$ (time series 2), $W = 0.5$ (maximum strength of power-law cross-correlations), and $\varepsilon_i = \varepsilon'_i$, we have a perfect cross-correlation $\sigma_{\text{DCCA}} = 1$ (△ in Fig. 1(c)). In the case of $\rho = 0.1$ (time series 1), $\rho = 0.1$ (time series 2), $W = 1.0$ (minimum strength of power-law cross-correlations), and $\varepsilon_i = -\varepsilon'_i$ there is a perfect anti cross-correlation $\sigma_{\text{DCCA}} = -1$ (∇ in Fig. 1(c)). Now, in the case of $\rho = 0.1$ (time series 1), $\rho = 0.4$ (time series 2), $W = 1.0$ and $\varepsilon_i \neq \varepsilon'_i$ there is no cross-correlation $\sigma_{\text{DCCA}} = 0$ (∗ in Fig. 1(c)). These results are entirely in agreement with Refs. [24,28,29]. The next step was to test the loss (expansion) of cross-correlations, and for this purpose a two-component ARFIMA stochastic process was generated. Initially, the expansion of cross-correlations between the time series, i.e., for $\rho = 0.1$ (time series 1), $\rho = 0.4$ (time series 2), $W = 0.85$, and $\varepsilon_i \neq \varepsilon'_i$

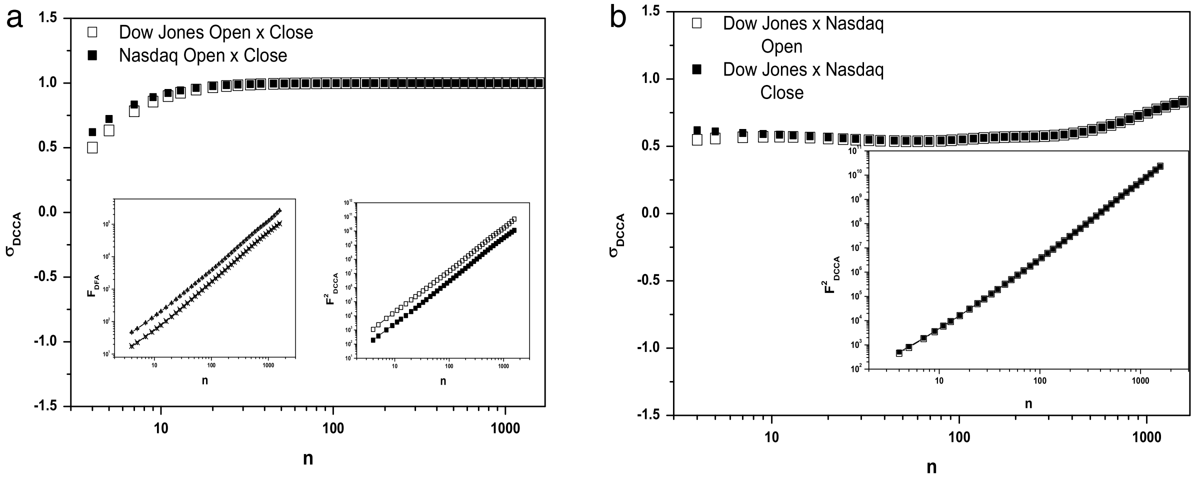

Fig. 3. DCCA cross-correlation coefficient σ_{DCCA} for Dow Jones and NASDAQ Open–Close indices. (a) Gives the individual index between opening and closing of trading, more specifically for Dow Jones (□) and NASDAQ (■). The inset figure to the left gives a log–log plot of $F_{\mathrm{DFA}} \times n$, with long-range power-law DFA exponents $\alpha_{DJopen} = \alpha_{DJclose} = 1.47$ and $\alpha_{NDopen} = \alpha_{NDclose} = 1.50$. In the right position in this figure we have a log–log plot of $F^2_{\mathrm{DCCA}} \times n$ with $\lambda_{\square} = 1.50$ and $\lambda_{\blacksquare} = 1.50$. (b) The cross-correlations between Dow Jones and NASDAQ Opening (□) and also Dow Jones and NASDAQ Closing (■). The inset figure gives a log–log plot of $F^2_{\mathrm{DCCA}} \times n$ with long-range cross-correlation power-law DCCA exponents $\lambda_{\square} = 1.49$ and $\lambda_{\blacksquare} = 1.51$.

Table 2
Mean values of σ_{DCCA}, with seasonal components.

Cross-correlation	$n \leq 24$	$n \geq 168$
Air temperature × solar radiation	0.81	0.68
Air temperature × relative humidity	−0.88	−0.59
Relative humidity × solar radiation	−0.64	−0.49
Wind speed × wind direction	0.05	0.52

was simulated. We can see that there is long-range cross-correlation between these time series (Fig. 1(b) (■)), but because $W = 0.85$ this cross-correlation is not perfect (Fig. 1(c) (■)). Afterwards, the negative case (anti cross-correlation), with cross-correlations loss was tested too, and for this case $\rho = -0.1$ (time series 1), $\rho = 0.4$ (time series 2), $W = 0.80$ and $\varepsilon_i = -\varepsilon'_i$. Again, there is long-range cross-correlation between these time series (Fig. 1(b) (◯)). But, even starting from $\sigma_{\mathrm{DCCA}} = -1$, the final value of σ_{DCCA} is below 0, because $W = 0.80$ (Fig. 1(c) (◯)). Thus, if we intend to quantify the level of cross-correlation, it will be impossible to use only the DCCA method, while σ_{DCCA} gives us this information directly (see Fig. 1(c)). We can see the potential utility of σ_{DCCA} in the next two real examples, the first produced by climatology and the second produced by the stock market. The first example, for climatological time series analysis, comes from the Brazilian National Institute of Meteorology, INMET [31], which is an important institute that has data in accordance with the World Climate Program [32]. INMET has 447 climatological stations working in all Brazilian territories. A climatological station consists of a central memory unit ("data logger") connected to various sensors of the meteorological parameters, integrating the observed values automatically every hour. In this paper we chose the climatological station in the city of Salvador (Bahia, Brazil), located at S 13.0053°, W 58.5058°, 51.41 m above sea level. The data are from 13 March 2010 to 10 June 2010, with $N = 2173$ points. In this station five variables are measured simultaneously every hour: air temperature (°C), relative humidity (%), solar radiation (kJ/m^2), wind speed (m/s), and wind direction (°). Then, the DCCA cross-correlation coefficient σ_{DCCA} was calculated between: air temperature × solar radiation (■), air temperature × relative humidity (•), relative humidity × solar radiation (∘), and wind speed × wind direction (□); see Fig. 2. In this figure we can identify positive, negative, and also zero cross-correlations. Now, if we check the cross-correlation between air temperature and humidity, we can identify $\sigma_{\mathrm{DCCA}} < 0$ (anti cross-correlated) and it is not perfect (mainly for large n); see Fig. 2 (•). Similar results are found if we analyze the cross-correlation between relative humidity and solar radiation (Fig. 2 ◯), but with a lower level of cross-correlation. The cross-correlation between air temperature and solar radiation is always positive and not perfect (Fig. 2 ■), while there is no cross-correlation between wind speed and wind direction until $n \simeq 100$ (Fig. 2 □). Finally, if we look at σ_{DCCA} in terms of n, we can identify seasonal components. Seasonality is most evident for $n = 24$ (day scale) and $n = 168$ (week scale) (see Fig. 2 vertical lines). In this way we can compose Table 2.

The last example is associated with the daily opening and closing values of the Dow Jones and the NASDAQ stock market indices, collected from 11 October 1984 to 5 May 2010. Fig. 3 depicts a complete overview for DFA, DCCA, and σ_{DCCA}. Specifically, σ_{DCCA} for the Dow Jones (Open × Close) (□) and NASDAQ (Open × Close) (■) have the same behavior; see Fig. 3(a). Again, in this figure we can identify seasonal components, i.e., $n = 30$ (one month) divides σ_{DCCA} into perfect cross-correlation ($n > 30$) or not ($n < 30$). Fig. 3(b) describes the cross-correlation between (Dow Jones × NASDAQ) Open–Open

(□) as well as (Dow Jones × NASDAQ) Close–Close (■). This figure informs us that if we analyze the cross-correlation between the Dow Jones and NASDAQ for opening or the Dow Jones and NASDAQ for closing, we have the same behavior, but with seasonal components at $n = 365$ (one year). Thus, $\sigma_{\rm DCCA} \simeq 0.5$ for $n < 365$ and $\sigma_{\rm DCCA} \to 1.0$ (perfect cross-correlation) for $n > 365$.

In conclusion, this paper proposes a new cross-correlation coefficient, defined by $\sigma_{\rm DCCA}$ in order to quantify the level of cross-correlation between nonstationary time series. The coefficient $\sigma_{\rm DCCA}$ is based on the DFA and the DCCA method. It proved to be robust, and also succeeded in identifying seasonal components, in both types of cross-correlations, positive and negative. For this purpose $\sigma_{\rm DCCA}$ was tested in selected nonstationary time series (simulated and real). Finally, is possible to generalize $\sigma_{\rm DCCA}$ to multi-time series analysis, in a possible DCCA cross-correlation matrix.

Acknowledgements

I wish to thank A. Machado Filho, and P.A. da Silva for discussions. This work received financial support from FAPESB (Bahia Research Agency).

References

[1] http://www.physionet.org/physiobank/database/.
[2] R.R. Wilcox, Introduction to Robust Estimation and Hypothesis Testing, 2nd ed., Academic Press, San Diego, CA, 2005.
[3] S.J. Devlin, R. Gnanadesikan, J.R. Kettenring, Biometrika 62 (3) (1975) 531.
[4] P.J. Huber, Robust Statistics, Wiley, New York, 1981.
[5] M.H. Katz, Multivariable Analysis — A Practical Guide for Clinicians, 2nd ed., University of California, San Francisco, 2006.
[6] http://www.physionet.org/publications/#papers-with-data.
[7] M.A. Moret, G.F. Zebende, E. Nogueira Jr., M.G. Pereira, Phys. Rev. E 68 (2003) 041104.
[8] J. Feder, Fractals, Plenum Press, New York, 1988.
[9] P. Bak, How Nature Works: The Science of Self-Organized Criticality, Oxford University Press, Oxford, 1997.
[10] H.E. Hurst, Trans. Amer. Soc. Civil Engineers 116 (1951) 770.
[11] C.-K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H.E. Stanley, Nature (London) 356 (1992) 168.
[12] G.F. Zebende, P.M.C. de Oliveira, T.J.P. Penna, Phys. Rev. E 57 (1998) 3311.
[13] C.K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Phys. Rev. E 49 (1994) 1685.
[14] K. Hu, P.C. Ivanov, Z. Chen, P. Carpena, H.E. Stanley, Phys. Rev. E 64 (2001) 011114.
[15] Z. Chen, P.C. Ivanov, K. Hu, H.E. Stanley, Phys. Rev. E 65 (2002) 041107.
[16] R.N. Mantegna, Eur. Phys. J. B 11 (1999) 193.
[17] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, H.E. Stanley, Phys. Rev. Lett. 83 (1999) 1471.
[18] B. LeBaron, W.B. Arthur, R. Palmer, J. Econ. Dyn. Control 23 (1999) 1487.
[19] L. Kullmann, J. Kertesz, K. Kaski, Phys. Rev. E 66 (2002) 026125.
[20] T. Mizunoa, H. Takayasu, M. Takayasu, Physica A 364 (2006) 336.
[21] B. Podobnik, Dongfeng Fu, H.E. Stanley, P.Ch. Ivanov, Eur. Phys. J. B 56 (2007) 47.
[22] B. Podobnik, D. Horvatic, A. Lam, H.E. Stanley, P.Ch. Ivanov, Physica A 387 (2008) 3954.
[23] B. Podobnik, D. Wang, D. Horvatic, I. Grosse, H.E. Stanley, Europhys. Lett. (EPL) 90 (2010) 68001.
[24] B. Podobnik, H.E. Stanley, Phys. Rev. Lett. 100 (2008) 084102.
[25] B. Podobnik, D. Horvatic, A.M. Petersen, H.E. Stanley, Proc. Natl. Acad. Sci. USA 106 (2009) 22079.
[26] G.F. Zebende, A. Machado Filho, Physica A 388 (2009) 4863.
[27] W.-X. Zhou, Phys. Rev. E 77 (2008) 066211.
[28] J. Hosking, Biometrika 68 (1981) 165.
[29] B. Podobnik, I. Grosse, D. Horvatic, S. Ilic, P.Ch. Ivanov, H.E. Stanley, Eur. Phys. J. B 71 (2009) 243.
[30] For this purpose, firstly were built two simulated self-affine times series, with random numbers ($rand_i$) between 0 and 1, $\{y_{rand1}\}$ and $\{y_{rand2}\}$. These time series are independent and uncorrelated, with DFA exponents $\alpha_1 = \alpha_2 = 0.5$ (see Fig. 1(a) □ and ∗), after we apply the ARFIMA process. We introduce a cutoff lenght $l = 10^4$ in our simulations, and we let the sum from 1 to l.
[31] http://www.inmet.gov.br/.
[32] Calculation of monthly and annual 30-year standard normals, WMO Technical document, n. 341; WCDP, n. 10, Geneva, 1989.