Multifractal detrended cross correlation analysis between Darwin and Tahiti stations

Sebastián Jaroszewicz,[1] Cristina Mariani,[1] and Cristina Mariani[2,3]

[1]*Comisión Nacional de Energía Atómica, Bs. As., Argentina*
[2]*UTEP, El Paso, United States*
[3]*Faculty of Physics, Mathematics and Computer Science, Cracow University of Technology, Kraków, Poland*

(Dated: October 13, 2022)

Escribir abstract

I. INTRODUCTION

The El Niño Southern Oscillation (ENSO) is a disrruption of the ocean atmospheric system in the tropical Pacificic. This complex atmosferic and oceanographic phenomenon is produced from the climate variability that is generated by an interaction between the atmosphere and ocean circulations [**? ? ?**]. El Niño is a warming event of the sea surface temperature (SST) in the central and eastern equatorial Pacific Ocean that is repeated in periods ranging from three to seven years. During these periods, the surface waters of a large strip of the tropical Pacific Ocean, warm or cool between 1 ° C and 3 ° C, compared to normal. This oscillating warming and cooling pattern constitutes the ENSO cycle. El Ni no and La Ni na are the extreme phases of the ENSO cycle. The name El Ni no (referring to the baby Jesus) was given by Peruvian fishermen to a warm current that appears periodically around Christmas. It was not until the 1960s that it was notic ed that this was not a local Peruvian phenomenon, and it was associated with changes throughout the tropical Pacific and beyond. The warm phase of El Niño usually lasts between 8-10 months. The entire ENSO cycle generally lasts between 3 and 7 years, and often includes a cold phase (La Ni na) that can be equally strong, as well as some years that are not abnormally cold or warm. However, the cycle is not a regular oscillation, and can be highly variable in both intensity and duration.

****

El Niño, and its counterpart La Niña, are two phases of one phenomenon (recurring at irregular intervals of between two and seven years, separated by years of neutral conditions. Under normal conditions, the warm waters of the western Pacific heat the air, which rises, causes precipitation, and circulates eastward as it reaches the stratosphere. It then descends and moves west, where it accumulates warm water. During El Niño episodes, warm waters heat the air above them, in this way the temperature changes produced in the water are coupled with the atmosphere. In this way, the central-eastern region of the Pacific becomes the main engine of an atmospheric circulation pattern called the Walker circulation. In the presence of an intense source of rising humid air, the surface winds along the equator weaken and sometimes change direction, blowing from west to east, causing warm water to enter the eastern Pacific Ocean, which brings with it increased precipitation in that region. In contrast, during La Niña, those winds strengthen and warm water moves westward, making the eastern Pacific cooler and drier. Therefore, El Niño represents the warm phase of a cycle of warming and cooling of the surface waters of the Tropical Pacific while La Niña represents the cold phase of the cycle. This atmospheric response forms the El Niño-Southern Oscillation (ENSO) and represents a phenomenon of great importance that contributes to El Niño maintaining and reinforcing itself. The phenomenon has great repercussions since, due to the great extension of the tropical Pacific, any variation that occurs triggers a domino effect in the rest of the world.

****

El Niño has important consequences for weather along the globe, non only in the tropical Pacific but in many regions of the world. It can cause droughts in places as diverse as Australia, Indonesia, India, Kenya, Morocco, Canada or Mexico, etc., and floods in the central and eastern Pacific regions, parts of South America close to Argentina, Chile, Peru, Ecuador , etc. As a consequence, local economies related to ecosystems, fisheries and agriculture are greatly affected.

By the early 1980s it was evident that El Niño was intimately linked to an atmospheric phenomenon called the Southern Oscillation, discover by Sir Gilbert Walker in 1923 [**?**]. The Southern Oscillation is a seesaw of atmospheric pressures between the Pacific and Indian Oceans. The strenght of this oscillating atmospheric phenomenon is measured by the Southern Oscillation Index (SOI) which is defined as the difference in the monthly Surface air pressure between two meteorological stations one at Darwin, Australia and other at the island of Tahiti. It is verified that El Ni no episodes are associated with negative values of the SOI, while La Ni na episodes are associated with positive values of the index.

Several studies reported that the SOI index has multifractal characteristics. Various methods were applied to extract the empirical multifractal properties in metereologicla data sets, for instance, Wavelet Transform, Modulus Maxima (WTMM) [55–57], and the Multifractal Detrended Fluctuation Analysis (MFDFA) [58], Multifractal Detrended Cross Correlation analysis [59], Multifractal Detrending Moving-Average Cross Correlation Analysis [60], Multifractal Cross-Correlation Analysis [61]. Although there are many works that study both the statistical and multifractal properties of the SOI, there is no precedent where the same is done with the values of the time series generated by the Darwin and Tahiti stations. We believe that the study of these series will allow a better understanding of the dynamics of the El Ni no phenomenon.

This is way the main purpose of this article is to study the cross-correlations between the time series of the surface air pressure measured at the Tahiti and Darwin stations under the fractal framework theory. As far as we know, there are no published works that analyze the cross-correlations between both

time series. In this article we analyze the cross-correlations between the Darwin and Tahiti monthly surface air pressure by the means of the MultiFractal Detrended Fluctuation cross-correlation Analysis (MFDFA) method.

The paper is organized as follows: in Section 2 the data used are presented. In Section 3 we describe the methods we use in this article. In section 4 we show and discuss the obtained results. Finally we present conclusions in the last section.

********** Agregar un parrafo donde se comente que en trabajos previos se encontró una ley de potencias con exponentes *** pero que no se realizaron estudios estadísticos adecuados o a los que les falta fundamento estadísitico. **********

II. DATA

The Southern Oscillation Index (SOI) is one measure of the fluctuations in air pressure occurring between the western and eastern tropical Pacific. Traditionally, this index has been calculated based on the differences in air pressure at sea level between the island of Tahiti and the city of Darwin, Australia. The existence of sustained negative values of the SOI frequently indicate episodes of the El Niño phenomenon. Those negative values are usually accompanied by a sustained warming in the central and eastern tropical Pacific Ocean, a decrease in the strength of the Pacific winds and a reduction in rainfall in eastern and northern Australia. On the other hand, positive SOI values are associated with strong winds from the Pacific and a warming of the sea north of Australia, popularly called La Niña. These positive values are generally accompanied by cooling of the waters of the central and eastern tropical Pacific Ocean and an increased probability that eastern and northern Australia will be wetter than normal.

Time series in present study were obtained from the Climatic Research Unit, University of East Anglia. The time period spanning from January 1866 to December 2021. The Data contains the surface air pressure measured at the Tahiti (17° 40 S, 149° 25 W) and Darwin stations (12° 27 S, 130° 50 E) as well as the SOI index values. The SOI, Thaity and Darwin time series are plotted as a function of time in Figs. 1 and 2.

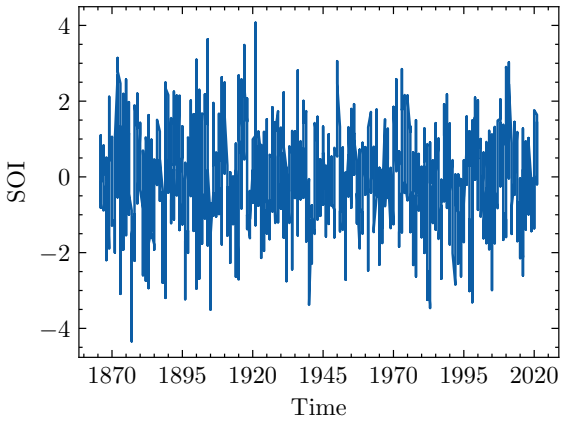

FIG. 1. Monthly values of the SOI from January 1866 to December 2021

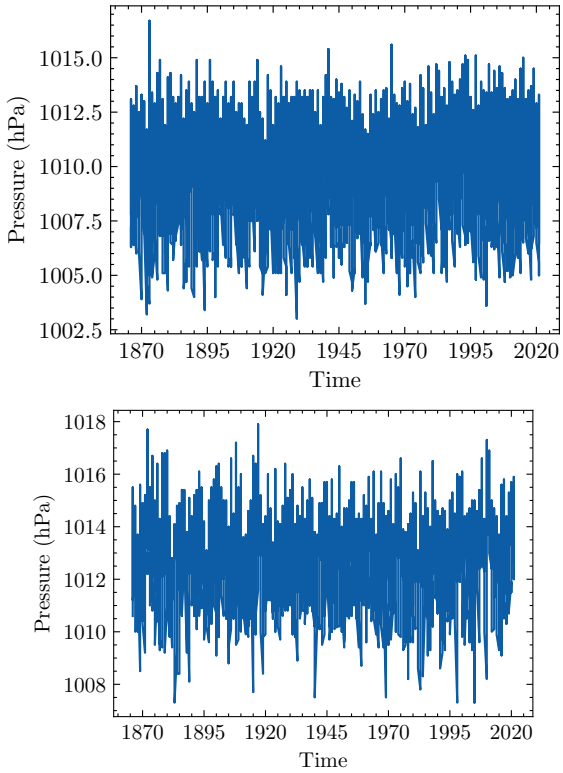

FIG. 2. Manthly values of the sea level preassure in hPa at (a) Darwin station and (b) Tahity station.

III. METHODOLOGY

A. Multifractal Detrended Fluctuation Analysis

As mentioned in the introduction, in this study MFDFA method was applied to study the multifractal characteristics of the analized temporal series. The MFDFA is a well-known technique, and is commonly used to detect multifractality in a time series [1] which can be summarized as follows.

Let $x(i)$ for $i = 1, \ldots, N$ be a possibly non-stationary time series, where N indicates its length. We, first, construct the "trajectory" or the "profile" by integration after subtracting from the time series its average x_{ave}.

$$X(i) = \sum_{k=1}^{i} [x(k) - x_{ave}]. \tag{1}$$

The profile is sub-divided into $N_s = int(N/s)$ non-overlapping windows of equal length s. Since the length N of the series may not be an integer multiple of the window size s, and a short part of the profile $Y(i)$ at the end may be

disregarded by the procedure, the sub-division is performed also starting from the opposite end, obtaining a total of $2N_S$ segments. A polynomial of degree m fits the profile in each of the $2N_S$ windows and the variance is calculated by using the following formula:

$$F^2(s,v) = \frac{1}{s}\sum_{i=1}^{s}\{X[(v-1)s+i] - y_v(i)\}^2. \quad (2)$$

For each segment v, $v = 1, \ldots, N_s$ and

$$F^2(s,v) = \frac{1}{s}\sum_{i=1}^{s}\{X[N-(v-N_s)s+i] - y_v(i)\}^2 \quad (3)$$

where $v = N_s+1, \ldots, 2N_s$. Here $y_v(i)$ is the fitting polynomial in segment v.

Then, averaging over all segments the q_{th} order fluctuation function is computed

$$F_q(s) = \left\{\frac{1}{2N_s}\sum_{v=1}^{2N_s}[F^2(s,v)]^{\frac{q}{2}}\right\}^{1/q} \quad (4)$$

where, in general, the index variable q can take any real value except zero. $F_q(s)$ will increase with increasing s and if $F_q(s)$ behave as a power-law of s the series is scaling for that specific q. In this case

$$F_q(s) \propto s^{h_q}. \quad (5)$$

The exponent h_q is called generalized Hurst exponent due to the equivalence between h_2 and the Hurst exponent (H) for stationary series, leading to consider the well know DFA [2] a particular case of the MFDFA for $q = 2$ [15, 16]. For $q = 0$ the value h_0 corresponds to the limit h_q for $q \to 0$, and is obtained through a logarithmic averaging procedure:

$$F_0(s) \equiv exp\left\{\frac{1}{4N_s}\sum_{v=1}^{2N_s} ln[F^2(s,v)]\right\} \propto s^{h_0}. \quad (6)$$

For monofractal series $h(q)$ is independent of q because the behavior of F^2 in 4 is independent of q. On the contrary, for multifractal series the exponent h_q will depend on q, and it monotonically decreases with the increase of q, the series is multifractal. For positive values of q the function $h(q)$ describes the scaling behavior of the segments with large fluctuations, whereas for negative values of q, the scaling behavior of the segments with small fluctuations is described. Based on $h(q)$ the mass exponent $\tau(q)$ (also called Rényi exponent) can be calculated as follows

$$\tau(q) = qh_q - 1 \quad (7)$$

The application of the Legendre transform to $\tau(q)$

$$\alpha = d\tau/dq \quad (8)$$

$$f(\alpha) = q\alpha - \tau(q) \quad (9)$$

allows us to characterize a multifractal time series through its multifractal spectrum. In the above equation α is the Hölder exponent and $f(\alpha)$ is the singularity spectrum that indicates the dimension of the subset of the series that is characterized by α. The multifractal spectrum indicates how much dominant are the various fractal exponents present in the series. The width of the spectrum, as well as the range of the generalized Hurst exponent $(max(h_q) - min(h_q))$, are often used to quantitatively measure the degree of multifractality of the series. Thus, the wider the spectrum the more multifractal the series is.

In order to characterize the complexity of the studied process, it is possible to extract a set of parameters from the spectrum. To do this, we adjust the singularity spectrum to a fourth degree polynomial:

$$f(\alpha) = A + B(\alpha-\alpha_0) + C(\alpha-\alpha_0)^2 + D(\alpha-\alpha_0)^3 + E(\alpha-\alpha_0)^4 \quad (10)$$

and we calculate the following parameters of the multifractal spectrum: the position of the maximum of the spectrum α_0, the width of the spectrum $\omega = \alpha_{max} - \alpha_{min}$, where α_{max} indicates the value of the most extreme events and α_{min}that of the softest ones, and, finally, the asymmetry parameter given by $a_s = (\alpha_0 - \alpha_{min})/(\alpha_{max} - \alpha_0)$. The parameter α_0 provides an estimate of the value of the Hurst exponent, in general a value of $\alpha_0 > 0.5$ indicates a correlated or persistent process, and $\alpha_0 < 0.5$ an anti-correlated or anti-persistent process, while $\alpha_0 = 0.5$ indicates a totally random process. The parameter ω, as indicated above, is the width of the spectrum and measures the amplitude of the fractal exponents necessary to describe the signal. The greater the multifractality of the process, the greater the width of the spectrum. The asymmetry parameter a_s allows us to measure the skewness of $f(\alpha)$, $a_s < 1$ indicates a rigth skewed spectrum while $a_s > 1$ a left skewed one.The importance of this parameter resides in that it provides us with information the prevalence of small and large fluctuations in the multifractal spectrum. If $a_s = 1$, the spectrum is symmetric and both large and small fluctuations contribute equally to multifractality. On the other hand, if the spectrum is asymmetric to the right, the greatest contribution to the multifractal spectrum is given by the subsets with small fluctuations and finally, an asymmetric spectrum to the left indicates that the subsets with large fluctuations are those that contribute the most to the multifractal spectrum.

In summary, these parameters allow us to evaluate the complexity of the process: a time series with a high value of α_0, a wide range of fractal exponents, and a right-skewed spectrum can be considered more complex than one with opposite characteristics.

B. DCCA and MF-DCCA

DCCA was developed by Podobnik et al. [10] from the well-known DFA method [2] with the objective of quantifying the long-term cross-correlation of two non-stationary time series.

In summary, the method consists of dividing the previously integrated time series $X(k), Y(k)$ into N_n segments of equal length n, and in each of them applying an ordinary linear regression to capture the local trend. The integrate series $y_{n,s}(k)$ is then detrended by substracting the local trend from the data in each box and the detrended covariance is calculated as

$$F^2_{DCCA}(n) = \frac{1}{nN_n} \sum_{s=0}^{N_n-1} \sum_{k=ns+1}^{n(s+1)} [X(k) - X_{n,s+1}(k)\,[Y(k) - Y_{n,s+1}(k)] \tag{11}$$

The relationship between the length n of the segments and F_{DCCA} is obtained by repeating the calculation for all segments sizes. In case that cross-correlation between series decay as a power law, then the detrending covarianze grows with the time scale as

$$F_{DCCA} \sim n^{\lambda} \tag{12}$$

where is the DCCA cross-correlation exponent. The value of λ the scaling exponent can be calculated from a linear regression on a plot of $\log F$ vs $\log \lambda$. If,however, the detrended covariance oscillates around zero as a function of the time scale, there are no long-range cross-correlations.

MF-DCCA is a generalization of the DCCA introduced by Zhou [11] and differs from MFDFA in that there are now two detrended signals $\tilde{X}_v(s,i)$ and $\tilde{Y}v(s,i)$ in each window. Therefore, the variance defined in equations 2 and 3 becomes covariance

$$F^2_{XY}(s,v) = \frac{1}{s} \sum_{i=1}^{s} \tilde{X}_v(s,i)\tilde{Y}_v(s,i) \tag{13}$$

Now the q_{th} order fluctuation function is computed by

$$F_q(s) = \left\{ \frac{1}{N_s} \sum_{v=1}^{N_s} [F_{XY}(s,v)]^{q/2} \right\}^{1/q} \tag{14}$$

Generally, q can take any real value, except zero. For $q = 0$, equation (14) becomes:

$$F_0(s) = \exp\left(\frac{1}{2N_s} \sum_{v=1}^{N_s} \ln F_{XY}(s,v) \right) \tag{15}$$

For $q = 2$, the standard DCCA is retrieved.

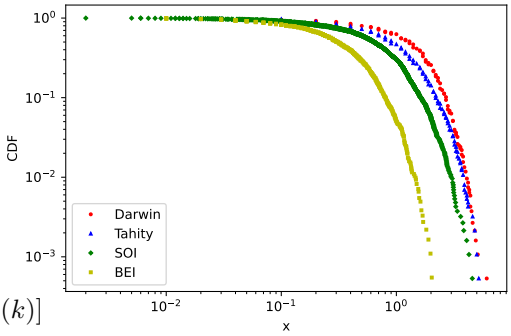

FIG. 3. Complemenatary cumulative distribution functions

Index	$< x >$	σ	x_{min}	α	σ_α
Darwin	-0.00037	1.652	4.400	11.549	3.73
Tahity	0.00112	1.360	1.799	4.512	0.19
SOI	0.00109	1.018	1.060	3.679	0.12
BEI	-0.00036	0.498	1.099	6.997	0.69

TABLE I.

In the case that there is a correlation between the two series, we can express the relationship between F and s as

$$F_q(s) \sim n^{\lambda}(q) \tag{16}$$

If $\lambda(q) = 0.5$ the two series are not correlated, which means that changes in one do not cause changes in the other. The case $\lambda(q) > 0.5$ indicates the presence of a persistent cross-correlation, which means it is likely that the increase in one series is followed by an increase in the other. Finally if the cross correlation is anti-persistent $\lambda(q) < 0.5$.

Like for the MFDFA if $\lambda(q)$ does not change with q, the cross-correlations between the series are monofractal. If $\lambda(q)$ is a decreasing function of q , the cross-correlation between these two time series is multifractal. Like for the MFDFA if the multifractal spectrum ends up being a point, there is no presence of multifractality. A wider multifractal spectrum indicates a stronger degree of multifractality.

IV. RESULTS

A. Power-Law Distribution

[1] Kantelhardt, J.W.; Zschiegner, S.A.; Koscielny-Bunde, E.; Havlin, S.; Bunde, A.; Stanley, H. Multifractal detrended fluctuation analysis of nonstationary time series. *Phys. A Stat. Mech. Appl.* **2002**, *316*, 87–114.

[2] Peng, C.-K.; Buldyrev, S.; Havlin, S.; Simons, M.; Stanley, H.E.; Goldberger, A.L. Mosaic organization of DNA nucleotides. *Phys. Rev. E* **1994**, *49*, 1685–1689

[3] Podobnik, B.; Stanley, H.E. Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series. *Phys. Rev. Lett.* **2008**, *100*, 084102.

[4] Peng, C.K.; Buldyrev, S.V.; Havlin, S.; Simons, M.; Stanley, H.E.; Goldberger, A.L. Mosaic organization of DNA nucleotides. *Phys. Rev. E* **1994**, *49*, 1685–1689.

[5] Wang, F.; Fan, Q.; Stanley, H. Multiscale multifractal detrended-fluctuation analysis of two-dimensional surfaces. *Phys. Rev. E* **2016**, *93*, 042213.

[6] Wang, F.; Liao, G.; Li, J.H.; Li, X.C.; Zhou, T.J. Multifractal detrended fluctuation analysis for clustering structures of electricity price periods. *Physical A Stat. Mech. Appl.* **2013**, *392*, 5723—5734.

[7] Igbawua, T.; Zhang, J.; Yao, F.; Ali, S. Long Range Correlation in Vegetation Over West Africa from 1982 to 2011. *IEEE Access* **2019**, *7*, 119151—119165

[8] Takaishi, T. Statistical properties and multifractality of Bitcoin. *Phys. A Stat. Mech. Its Appl.* **2018**, *506*, 507—519.

[9] Kalamaras, N.; Philippopoulos, K.; Deligiorgi, D.; Tzanis, C.G.; Karvounis, G. Multifractal scaling properties of daily air temperature time series. *Chaos Solitons Fractals* **2017**, *98*, 38—43.

[10] Podobnik, B.; Stanley, H. E. . Detrended Cross-Correlation Analysis: A New Method for Analyzing Two Nonstationary Time Series. *Phys. Rev. Lett.* **2008** *100* 1—11.

[11] Zhou,Wei-Xing. Multifractal detrended cross-correlation analysis for two nonstationary signals. *Phys. Rev. E* **2008***77*, 066211.

[12] Datos Abiertos del Ministerio de Transporte. Available online: https://archivos-datos.transporte.gob.ar/upload/Sube/total-usuarios-por-dia.csv

[13] JHU CSSE COVID-19 Data. Available online: https://github.com/CSSEGISandData/COVID-19.

[14] Dong E., Du H., Gardner L., An interactive web-based dashboard to track COVID-19 in real time. *Lancet Inf. Dis.* **2020**, *5*, 533–534.

[15] Kantelhardt, J.W.; Koscielny-Bunde, E.; Rybski, D.; Braun, P.; Bunde, A.; Havlin, S. Long-term persistence and multifractality of precipitation and river runoff records. *J. Geophys. Res. Atmos.* **2006**, *111*, D1.

[16] Zhang, Q.; Xu, C.-Y.; Yu, Z.; Liu, C.-L.; Chen, Y.-D. Multifractal analysis of streamflow records of the East river basin (Pearl river), China. *Phys. A Stat. Mech. Its Appl.* **2009**, *388*, 927—934.

[17] Dickey, D.A.; Fuller, W.A. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *J. Am. Stat. Assoc.* **1979**, *74*, 427—431.

[18] Dickey, D.A.; Fuller, W.A. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econom. J. Econom. Soc.* **1981**, *49*, 1057–1072.

[19] Theiler J, Galdrikian B, Longtin A, Eubank S, Farmer D. J. Using surrogate data to detect nonlinearity in time series. In: *Nonlinear modeling and forecasting*. Casdagli M., Eubank S.(eds) Addison-Wesley, Redwood City, CA, 1992, 163–188