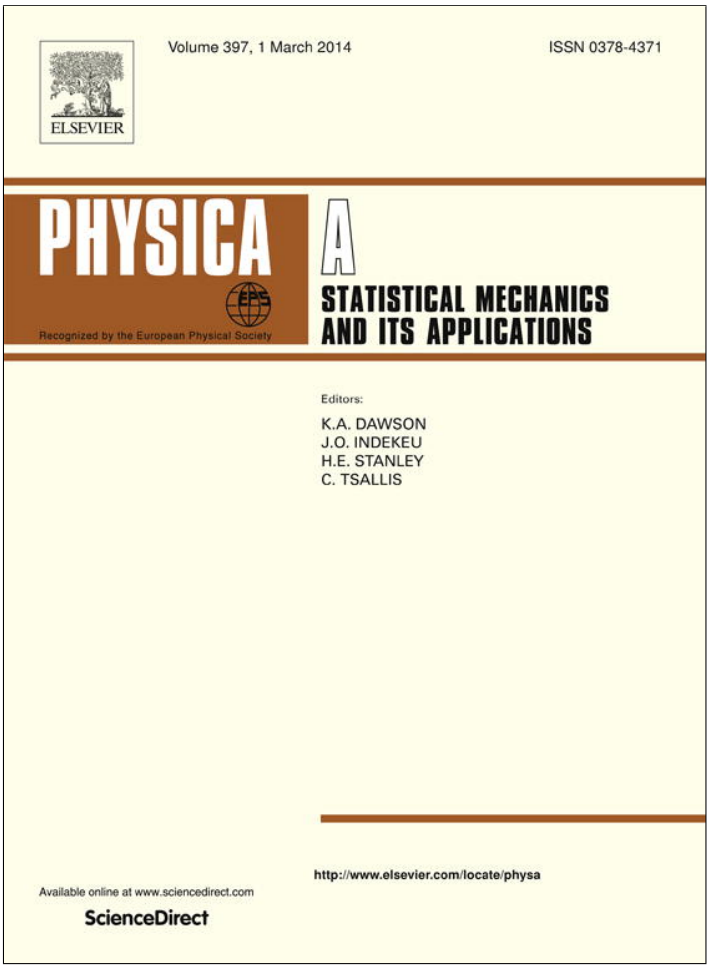


Physica A 397 (2014) 17–30

Contents lists available at ScienceDirect

Physica A

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

A criterion for the determination of optimal scaling ranges in DFA and MF-DFA

Damián Gulich [a,b,c,*], Luciano Zunino [a,b]

[a] Centro de Investigaciones Ópticas (CONICET La Plata - CIC), C.C. 3, 1897 Gonnet, Argentina
[b] Departamento de Ciencias Básicas, Facultad de Ingeniería, Universidad Nacional de La Plata (UNLP), 1900 La Plata, Argentina
[c] Departamento de Física, Facultad de Ciencias Exactas, Universidad Nacional de La Plata (UNLP), 1900 La Plata, Argentina

HIGHLIGHTS

- We develop a criterion to determine optimal fitting regions in DFA and MF-DFA.
- We use it on several artificial series and find good agreement with known parameters.
- We show that crossover scales can be objectively estimated with this method.
- We also examine monthly sunspot data and confirm previous scaling results.

ARTICLE INFO

Article history:
Received 27 May 2013
Received in revised form 22 October 2013
Available online 6 December 2013

Keywords:
Scaling range
Scaling exponent
Detrended Fluctuation Analysis
Multifractal Detrended Fluctuation Analysis
Crossover time scales
Time series analysis

ABSTRACT

We develop a criterion based on a brute-force algorithm to systematically determine optimal fitting regions for fluctuation functions in Detrended Fluctuation Analysis (DFA) and Multifractal Detrended Fluctuation Analysis (MF-DFA). We analyze and compare results with several artificially generated time series with known parameters as illustrative examples of this technique. We show that crossover time scales can also be objectively and efficiently estimated using the introduced algorithm. Finally, we employ the proposed methodology to study the scaling behavior of a natural time series: monthly sunspot data previously filtered with empirical mode decomposition (EMD); in this case we find results comparable with those found in the literature.

1. Introduction

Since the seminal work on Detrended Fluctuation Analysis (DFA) by Peng et al. [1] and its multifractal generalization (MF-DFA) by Kantelhardt et al. [2] these techniques have been widely used in time series analysis. Both of them have been implemented in such diverse fields as econophysics [3,4], seismology [5], biology [1], medicine [6,7], cosmology [8], condensed matter [9], and music [10–14] among others. There are more than 1400 papers related to DFA and more than 600 on MF-DFA,[1] and these numbers are indeed growing.

Several papers have tried to shed some light on the performance of these techniques on short time series (finite-size effect) [15,16]. Actually, W.-X. Zhou [17] has introduced the concept of *effective multifractality*, defined as the apparent multifractality after removing the finite-size effect. Also, the spurious effects due to the presence of noise, short-term

* Corresponding author at: Centro de Investigaciones Ópticas (CONICET La Plata - CIC), C.C. 3, 1897 Gonnet, Argentina. Tel.: +54 221 471 4341; fax: +54 221 471 2771.
E-mail addresses: dgulich@ciop.unlp.edu.ar, damian.gulich@gmail.com (D. Gulich), lucianoz@ciop.unlp.edu.ar (L. Zunino).

[1] Information extracted from the Scopus bibliographic database (www.scopus.com/scopus/home.url) accessed in September, 2013.

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.11.029

memory and periodicities in the data have been carefully analyzed [18,19]. Both methodologies, i.e. DFA and MF-DFA, require the identification of power-law behaviors in a range of scales in order to estimate the associated scaling exponents. However, fitting regions are sometimes very hard to determine without a theoretical model for scaling behaviors; even with a model the determination of such regions is subjective and therefore results derived from different analysis of the same series are difficult to compare. As stated by Shao et al. [20]: "there is no consensus on an objective determination approach of the scaling range, which plays a crucial role in the estimation of the scaling exponents". This is also a subtle issue when dealing with series which have more than one scaling behavior. Multi-scaling behaviors are often found in natural and man-made systems [21].

It is therefore necessary to develop a user-independent criterion to estimate the optimal fitting regions of fluctuation functions. One possibility is to study the local exponents by means of the log derivative plot $(d(\log(F_q(s)))/d(\log(s))$ versus $\log(s))$, as suggested by Govindan et al. [22], Bashan et al. [23] and Lopez et al. [16], looking for constant value regions. However, experimental series very usually have log-derivative discontinuities for small q when dealing with MF-DFA, making this method very hard to apply in such cases. Michalski [24] has identified the optimal minimal and maximal scale sizes for persistent processes (fractional Brownian motion and fractional Gaussian noise) through a number of extensive Monte Carlo simulations. The scaling region for DFA of artificial correlated and uncorrelated series as a function of series length, Hurst exponent and goodness of fluctuation function linear fit (R^2) at given confidence levels has been studied by Grech et al. [25]. The same study has been very recently extended with success to other techniques of fluctuation analysis [26]. Following a related approach, we propose to use R^2 as a way to locate optimal scaling regions. This criterion based on goodness of linear fit may be naturally extended to other techniques requiring linear fits. In fact, modified detrending methods, such as the Detrended Moving Average (DMA) [27], Centering Moving Average (CMA) [28], Modified Detrended Fluctuation Analysis (MDFA) [29], continuous DFA (CDFA) [30], Detrended Cross-Correlation Analysis (DCCA) [31], Multifractal Detrended Cross-Correlation Analysis (MF-DCCA) [32], Multifractal Detrending Moving Average (MF-DMA) [33], Multifractal Detrending Moving-Average Cross-Correlation Analysis (MFXDMA) [34], DFA method based on varying polynomial order [35], and EMD-based MF-DFA [36], might be actually benefited from its implementation. As will be discussed in detail below, crossover phenomena, i.e. the presence of crossover scales separating regimes with different scaling exponents, can also be efficiently unveiled by applying this methodology. The identification of these characteristic scales is relevant for a complete understanding of the underlying multiscale dynamics [21].

The remainder of this work is organized as follows: in Section 2 we briefly describe the fractal and multifractal detrending techniques used. In Section 3 we introduce the proposed criterion for DFA and then we generalize it for MF-DFA also discussing the identification of crossover scales. Some illustrative examples are included in Section 4. In Section 5 we employ the proposed algorithm for the analysis of a natural time series: the monthly sunspot time series. Finally, the conclusions of this work are detailed in Section 6.

2. Theoretical backgrounds

MF-DFA [2] is based on the traditional DFA [1]. It may be briefly summarized as follows[2]:

1. Let $x_1, x_2, \ldots, x_N$ be a series of N equidistant measurements. Given its mean value $\langle x\rangle$ we determine a new series of $Y(1), \ldots, Y(N)$ values given by $Y(i) = \sum_{k=1}^{i} (x_k - \langle x\rangle)$.
2. We split the series of $Y(1), \ldots, Y(N)$ into N_s non-overlapping windows of s data points where $N_s = \lfloor N/s \rfloor$.[3] If N is not divisible by s, there will be some remaining r values ($r = N - N_s s$) at the end of the series. To solve this, we take other N_s segments but starting from $Y(r+1)$. In this way we get $2N_s$ windows of s values.
3. Let ν be the index of the $2N_s$ windows ($\nu = 1, 2, \ldots, 2N_s$). For each one of the windows we shall take the polynomial $y_\nu(i)$ of degree m that best fits the data in the window where i is the data index. Then we find the N_s local variances for $\nu = 1, \ldots, N_s$:

$$F^2(\nu, s) = \frac{1}{s} \sum_{i=1}^{s} \{Y[(\nu - 1)s + i] - y_\nu(i)\}^2,$$

m is the same for every step of this technique and determines the detrending polynomial order of the analysis, hence MF-DFA1 means $m = 1$, etc.[4] If N is not divisible by s, then we also have to find the variances of the other N_s windows ($\nu = (N_s + 1), \ldots, 2N_s$) according to

$$F^2(\nu, s) = \frac{1}{s} \sum_{i=1}^{s} \{Y[N - (\nu - N_s)s + i] - y_\nu(i)\}^2.$$

Since N not divisible by s is the most usual situation, if N is divisible by s we simply repeat the N_s values of $F^2(\nu, s)$ so that we shall always work with $2N_s$ values of $F^2(\nu, s)$.

[2] We refer the interested readers to Ref. [37] for further details.

[3] Where $\lfloor w \rfloor$ = floor (w) is the largest integer not greater than w.

[4] It has been recently shown that the choice of a suitable detrending polynomial order is crucial to avoid spurious findings [38].

4. We calculate the fluctuation functions $F_q(s)$ where

$$F_q(s) = \begin{cases} \left\{ \frac{1}{2N_s} \sum_{\nu=1}^{2N_s} \left[F^2(\nu, s)\right]^{q/2} \right\}^{1/q}, & q \neq 0 \\ \exp\left\{ \frac{1}{4N_s} \sum_{\nu=1}^{2N_s} \ln\left[F^2(\nu, s)\right] \right\}, & q = 0. \end{cases}$$

   These functions are only defined for $s \geq m + 2$. As a note, $q = 2$ yields the traditional DFA.
5. We repeat steps 2–4 for different values of s, usually in the $[10, \lfloor N/4 \rfloor]$ range and chosen to be equally spaced in a log scale (this will be necessary for step 6).
6. We finally determine the scaling behavior of the fluctuation functions analyzing the log–log plot of all the $F_q(s)$ versus s. If the original series of x_i values has long range correlations, there exists a range of scales, $s_{\min} < s < s_{\max}$, in which $F_q(s) \sim s^{h(q)}$ where $h(q)$ is the generalized Hurst exponent; it can be estimated as the slope of the log–log plot of $F_q(s)$. If the series is monofractal and stationary, then $h(q)$ is constantly equal to the Hurst exponent H, i.e. independent of q. Otherwise, for a multifractal time series, the generalized Hurst exponent is a decreasing function of q and the Hurst exponent can be estimated from the second moment ($q = 2$).

3. Criterion based on goodness of linear fit

3.1. DFA case

Suppose x_t is a time series and we have M given scales[5] $s : s \equiv \{s_1, s_2, \ldots, s_M\}$ for which we have already calculated the fluctuation function $F(s) \equiv \{F(s_k)\}$ $(1 \leq k \leq M)$. We have chosen $M = 100$ for all examples in this work. The following algorithm determines the optimum scale fitting range for the fluctuation function associated with the series:

1. Define a minimum number of data points δ. In practice, this parameter should be no less than 10 so that a linear fit makes statistical sense; we have found that values of δ between $\lfloor M/4 \rfloor$ and $\lfloor M/3 \rfloor$ will discriminate regions of at least that number of data points, which is a reasonable partition of the data points domain.
2. Calculate the logarithmic series[6] $L_s = \{\log(s_k)\}$ and $L_F = \{\log(F(s_k))\}$ $(1 \leq k \leq M)$.
3. Define a matrix $r_{M \times M}$ of default values equal to 0.
4. Compute all non-zero elements of r according to

$$r_{ij} = R^2\left(s_i, s_j\right) \tag{1}$$

   where $R^2\left(s_i, s_j\right)$ is the coefficient of determination (R^2) of the linear fit of L_F versus L_s between (and including) $\log(s_i)$ and $\log\left(s_j\right)$. Also $1 \leq i \leq (M - \delta + 1)$ and $(i + \delta - 1) \leq j \leq M$. By definition, every $r_{ij} \leq 1$ and non-zero values are in the diagonal superior region of the matrix.
5. Sort all non-zero values r_{ij} in decreasing order while keeping a record of their original subindices.
6. If there are repeated values in r, sorting requires an extra criterion; based on length of interval choose first the longest one (in amount of data points).

The first element of this list will then provide the best linear fit interval ('*Dominant*') and the rest will follow it in decreasing quality. Since the slope of linear fit in the range estimates the scaling exponent h ($F \sim s^h$), we have found the best estimation range for this parameter (this will be further discussed in Section 3.3).

This algorithm is brute force in nature since it actually performs all possible fits of at least δ data points between scales s_1 and s_M; this is not computationally prohibitive since we are dealing with relatively few points (please see the Appendix for a detailed analysis). MATLAB implementation for this algorithm is available upon request. If interested, please contact Damián Gulich (e-mail: dgulich@ciop.unlp.edu.ar).

3.2. Generalization to MF-DFA

In MF-DFA [2] fluctuation functions depend on exponent $q = \left\{q_1, \ldots, q_Q\right\} : F_q(s)$ (thus having Q fluctuation functions). The previous algorithm may be generalized as follows:

[5] Equally log spaced.

[6] $\log \equiv \log_{10}$ in this work.

1. Define a minimum number of points δ as in step 1 of the DFA case. This δ will remain constant in the following steps.
2. Repeat steps 2–5 of the previous subsection for each $F_q(s)$ obtaining the generalized version of Eq. (1):
$$r_{ij}^{(n)} \equiv R_{q_n}^2\left(s_i, s_j\right) \tag{2}$$
with $1 \leq n \leq Q$.
3. With Eq. (2) compute the mean matrix $\langle r \rangle$:
$$\langle r \rangle \equiv \langle r \rangle_{ij} = \frac{1}{Q}\sum_{n=1}^{Q} r_{ij}^{(n)}. \tag{3}$$
4. Apply the sorting described in step 5 of the previous subsection to $\langle r \rangle$.
5. If there are repeated values in $\langle r \rangle$, sorting requires an extra criterion; based on length of interval choose first the longest interval (in number of data points).

Note that this generalization will produce the same output for DFA. Another possibility is to calculate fluctuation functions for many values $\{q_1, \ldots, q_Q\}$ and apply this technique only to a subset $\{q'_1, \ldots, q'_Z\} \subseteq \{q_1, \ldots, q_Q\}$ $(Z \leq Q)$.

3.3. Different scaling behaviors

In several situations the correlations of data do not follow the same scaling law for all considered scales s, and different scaling regimes with one or more crossover scales are observed [39]. For example, the presence of long-range correlations on small scales ($s < s_x$, s_x being the crossover scale) and a practically uncorrelated behavior on larger scales ($s > s_x$) has been shown (please see Fig. 11 in Ref. [40]). Moreover, a multifractal to monofractal transition at a well-defined crossover scale in traffic speed fluctuations has been verified [41]. Undoubtedly, the identification of these crossover scales can provide very useful information about the underlying dynamics. Indeed, differences in scaling for short and long time scales can be employed as potential indicators for distinguishing normal from pathological time series [6]. It should be remarked here that Ge and Leung [21] have very recently introduced a rigorous statistical approach, called the *scaling-identification regression model*, that is able to identify crossover time scales with confidence intervals.

It is interesting to note that the determination of good quality non-overlapping intervals implies different scaling behaviors: $F \sim s^{h_1}, s^{h_2}, \ldots$ for the DFA case and $F_q(s) \sim s^{h_1(q)}, s^{h_2(q)}, \ldots$ for the MF-DFA case. The methodology proposed in this paper will give first the best fitting range for both DFA and MF-DFA; nearby different regimes may be found searching for the closest intervals on the list that (including border elements)

- end before the optimal interval's beginning
- begin after the optimal interval's end.

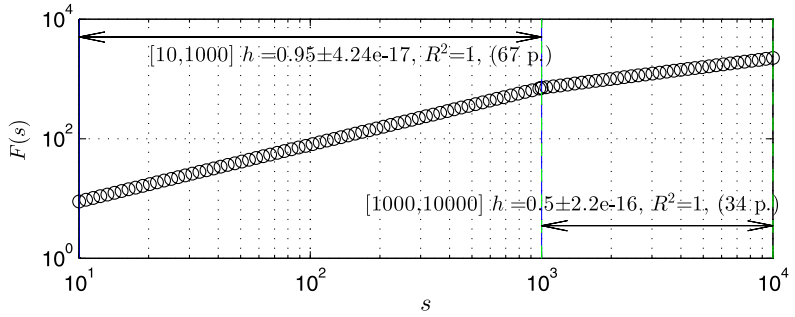
This process may be iterated to cover all scales involved, thus discriminating regions of well-defined power-law behavior. The first interval to the right (left) will be called '*Next*' ('*Previous*'). Other regions may exist to the right (left) of 'Next' ('Previous') regions as long as they satisfy the δ condition ($\lfloor M/4 \rfloor \leq \delta \leq \lfloor M/3 \rfloor$).

4. Illustrative examples

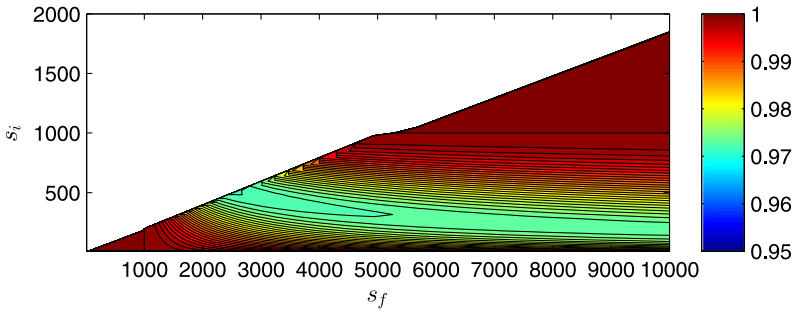
4.1. DFA of simple artificial series

We generated 99 equally spaced points u_k in the [1, 4] interval and added point 3 to have 100 points in total. Then we calculated
$$v_k = \begin{cases} 0.95 \cdot u_k & u_k \leq 3 \\ 1.35 + 0.5 \cdot u_k & u_k > 3. \end{cases}$$
We also generated a set of 100 uniform random numbers e_k in the interval $[-0.035, 0.035]$ to be used as additive noise, generating a series $\widetilde{v}_k = v_k + e_k$ $(1 \leq k \leq 100)$. Then we generated the artificial fluctuation functions $F(s)$ with series y_k and $\widetilde{y}_k$ and scales s according to:
$$s_k = 10^{u_k}$$
$$y_k = 10^{v_k}$$
$$\widetilde{y}_k = 10^{\widetilde{v}_k}$$
$(1 \leq k \leq 100)$. A log–log plot of (s_k, y_k) shows straight lines of slope 0.95 for scales $\leq 10^3$ and slope 0.5 for scales greater than 10^3 respectively (Fig. 1(a)), while both regimes are not so clear for $(s_k, \widetilde{y}_k)$ (Fig. 2(a)). Figs. 1 and 2 show results of the algorithm applied to both series ($\delta = 25$). Obtained results are compared in Table 1.

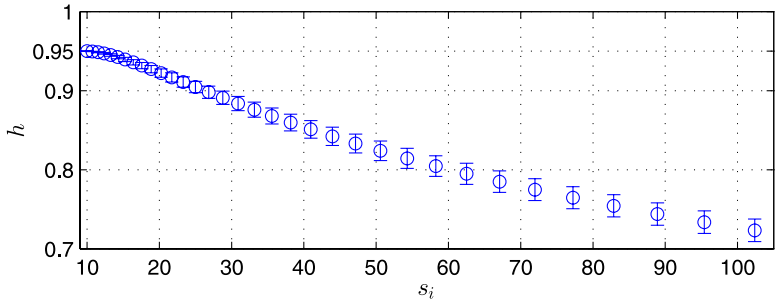
It is interesting to note that in both cases detected regions will remain the same for δ values greater than 25 and smaller than the width of the 'Next' region in data points. When the width of the 'Next' region (34 data points without noise and 31 with noise) becomes smaller than δ the criterion reestimates that region. In these cases, 'Dominant' regions will be unaffected up to $\delta = 67$ (without noise) and $\delta = 69$ (with noise). Both 'Dominant' and 'Next' regions remain approximately unchanged in the range of δ between $\lfloor N/4 \rfloor$ and $\lfloor N/3 \rfloor$.



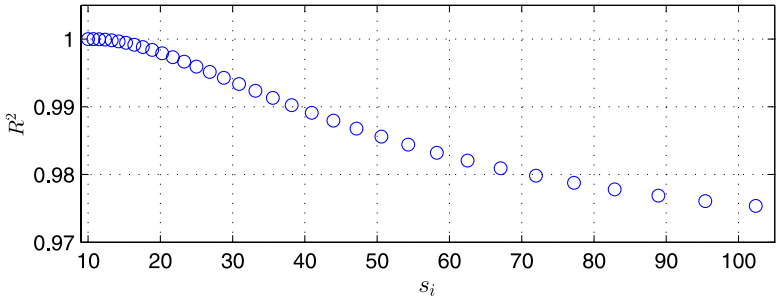
(a) Series without noise.



(b) R^2 array—non-zero values.



(c) h fit from s_i, 67 data points wide.

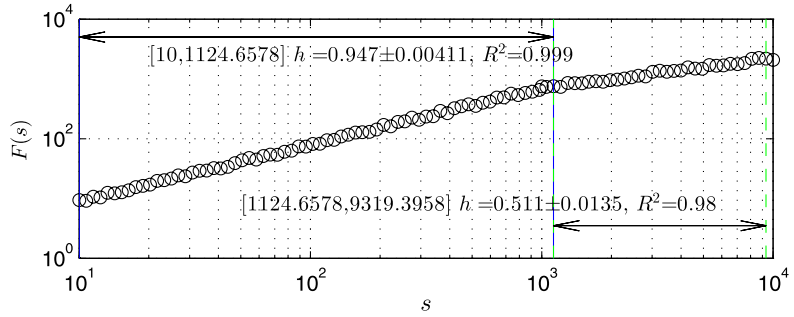


(d) R^2 fit from s_i, 67 data points wide.

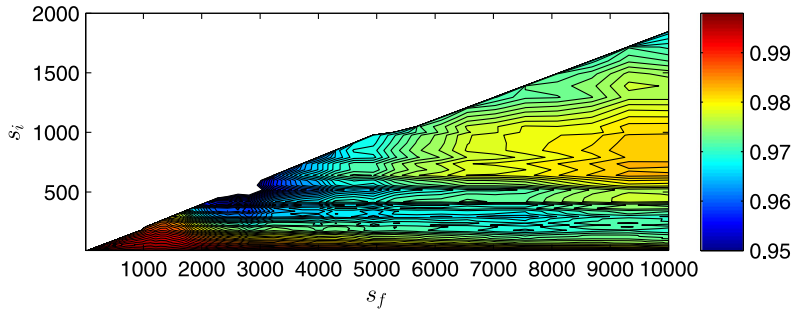
Fig. 1. Simulated fluctuation functions studied with the algorithm (DFA approach), $\delta = \lfloor 100/4 \rfloor = 25$. (a) Simulated fluctuation function without noise. (b) Contour plot of the $\langle r \rangle$ matrix. The 'Dominant' region was found to be $\left[10,\ 10^3\right]$, $R^2 = 1$ (67 data points wide); the 'Next' region was $\left[10^3,\ 10^4\right]$, $R^2 = 1$ (34 data points wide). (c) Estimated h fitting 67 consecutive data points and starting at s_i; error bars increase as the fitting interval gets more data from the 'Next' region. (d) R^2 of fits calculated as (c); quality of fit decreases. Best fit is in the 'Dominant' region and estimates $h = 0.950\ldots \pm 4.238 \cdot 10^{-17}$ (expected value is $h = 0.95$); and for the 'Next' region $h = 0.50\ldots \pm 2.20 \cdot 10^{-16}$ (expected value is $h = 0.5$).

4.2. Systematic analysis of fractional Gaussian noises

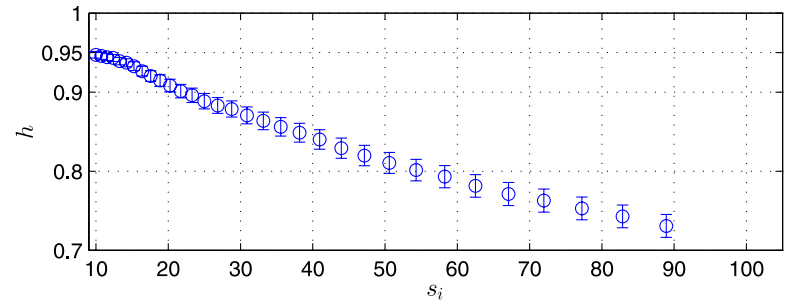
We have performed a systematic analysis of fractional Gaussian noises (fGns) with the criterion introduced in this work. It is well known that the long-range correlations associated with these monofractal Gaussian stationary stochastic processes are fully characterized by the Hurst exponent H. For this purpose we have generated 100 independent numerical realizations



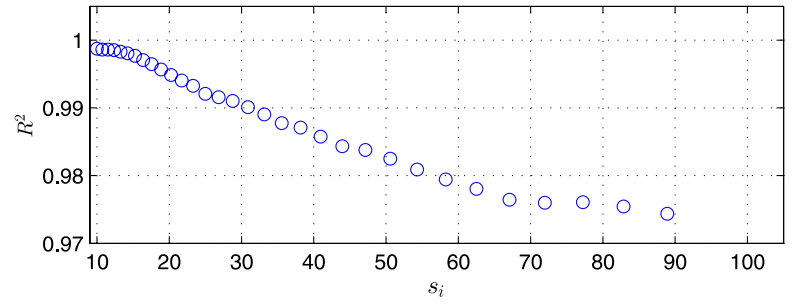
(a) Series with noise.



(b) R^2 array—non-zero values.



(c) h fit from s_i, 69 data points wide.



(d) R^2 fit from s_i, 69 data points wide.

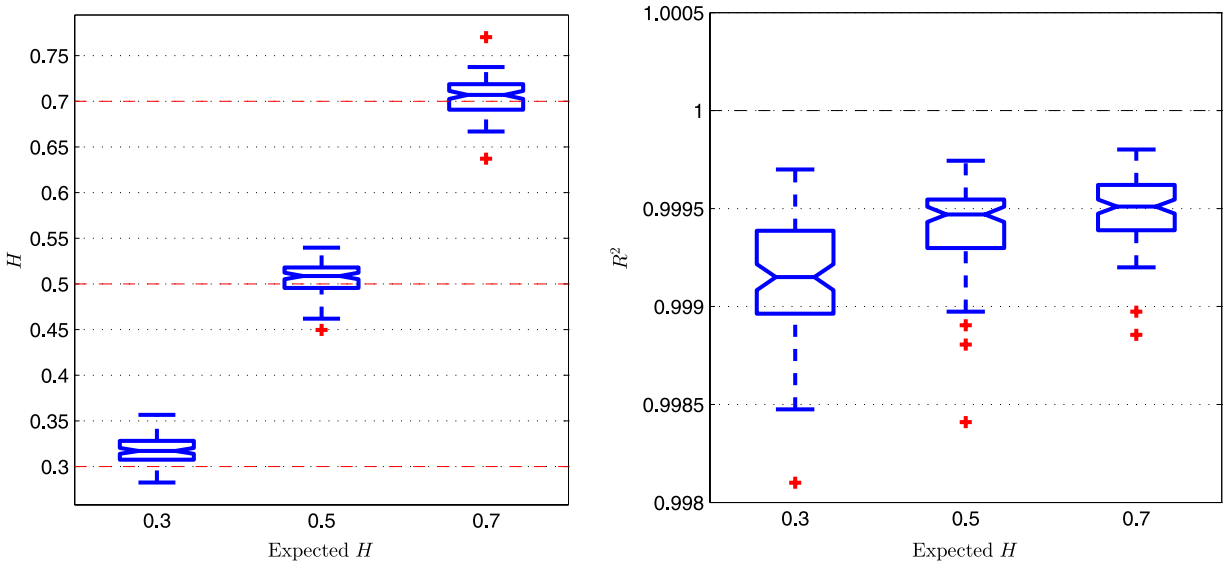
Fig. 2. Simulated fluctuation functions studied with the algorithm (DFA approach), $\delta = \lfloor 100/4 \rfloor = 25$. (a) Simulated fluctuation function with additive noise. (b) Contour plot of the $\langle r \rangle$ matrix. The 'Dominant' region was found to be $[10, 1124.658]$, $R^2 = 0.999$ (69 data points wide); the 'Next' region was $[1124.658, 9319.396]$, $R^2 = 0.980$ (31 data points wide). (c) Estimated h fitting 69 consecutive data points and starting at s_i; error bars increase as the fitting interval gets more data from the 'Next' region. (d) R^2 of fits calculated as (c); quality of fit decreases. The best fit is in the 'Dominant' region and estimates $h = 0.947 \pm 0.004$ (expected value is $h = 0.95$); and for the 'Next' region $h = 0.511 \pm 0.014$ (expected value is $h = 0.5$).

of fGns with $H \in \{0.3, 0.5, 0.7\}$ of length $N = 10^4$ datapoints by using the MATLAB *wfbm* function.[7] The results obtained of this automated study by implementing DFA with a detrending polynomial order $m = 1$, $M = 100$ and $\delta = 25$ are shown in Fig. 3. The quality of fit (R^2) increases with H while it decreases in dispersion (Fig. 3(b)).

[7] This function simulates fractional Brownian motion (fBms) following the algorithm proposed by Abry and Sellan [42]. FGns were obtained through successive differences of the generated fBms.

Table 1
Summary of results derived from Fig. 1.

Parameter	Theoretical (no noise)	Without noise (algorithm)	With noise (algorithm)
'Dominant' region	$[10, 10^3]$	$[10, 10^3]$	[10, 1124.658]
R^2 ('Dominant')	–	1	0.999
h ('Dominant')	0.95	$0.950 \ldots \pm 4.24 \cdot 10^{-17}$	0.947 ± 0.004
Data points ('Dominant')	67	67	69
'Next' region	$[10^3, 10^4]$	$[10^3, 10^4]$	[1124.658, 9319.396]
R^2 ('Next')	–	1	0.980
h ('Next')	0.5	$0.50 \ldots \pm 2.20 \cdot 10^{-16}$	0.511 ± 0.014
Data points ('Next')	34	34	31



(a) H DFA1 boxplot – fGn – ($M = 100 - \delta = 25 -$ Dom.). (b) R^2 boxplot–fGn.

Fig. 3. Automated study of fGns. (a) Boxplot of estimated H values calculated in the 'Dominant' regions versus expected H. (b) R^2 boxplot for expected H values: quality of fit increases with H.

4.3. Identification of crossover time scales

An analysis of the performance of the proposed criterion regarding the estimation of crossovers follows. With this aim in mind artificial time series with two well-localized crossovers at scales s_1 and s_2 are generated. These testbed records can be easily simulated following the recipe described by Schumann and Kantelhardt in Ref. [43]. Segmentation of the original time series into non-overlapping boxes of length s_u and the subsequent random shuffling of these boxes allows us to destroy long-range correlations on scales above the scale s_u [23]. In such a way, generated surrogate data is monofractal and uncorrelated ($h(q) = 1/2$) for $s > s_u$. Analogously, correlations on short scales $s < s_v$ can be eliminated after random shuffling of the data within blocks of size s_v while the order of the blocks are kept unchanged. Consequently, $h(q) = 1/2$ for $s < s_v$.

We have generated 20 numerical independent realizations of fGns with $H = 0.6$ and $N = 10^5$ by using the same algorithm as described in Section 4.2. On the one hand, the underlying correlation is eliminated for shorter scales ($s < s_1 = 10^2$) by shuffling the original data within blocks of this size. On the other hand, an uncorrelated behavior is simulated for larger scales ($s > s_2 = 10^3$) shuffling whole blocks of this length. Thus, original correlation is only preserved for an intermediate scale range, i.e. $s_1 < s < s_2$.

Crossover scales are estimated from the intersections of the linear fits in the different regions. Please see Fig. 4 where results obtained for the estimated crossovers by implementing DFA with a detrending polynomial order $m = 1$ and $\delta = 25$ are depicted. The three different scaling regimes are identified. The first crossover ('Dominant–Next1') has a lower dispersion than the second one ('Next1–Next2'); this may be explained by considering that there is a lower number of windows for larger scales (this has a noise-like effect in fluctuation functions for large s).

4.4. Binomial multifractal series

Following Ref. [2] we generated a series of 2^{16} values calculated with

$$x_k = a^{n(k-1)} (1 - a)^{n_{\max} - n(k-1)}$$

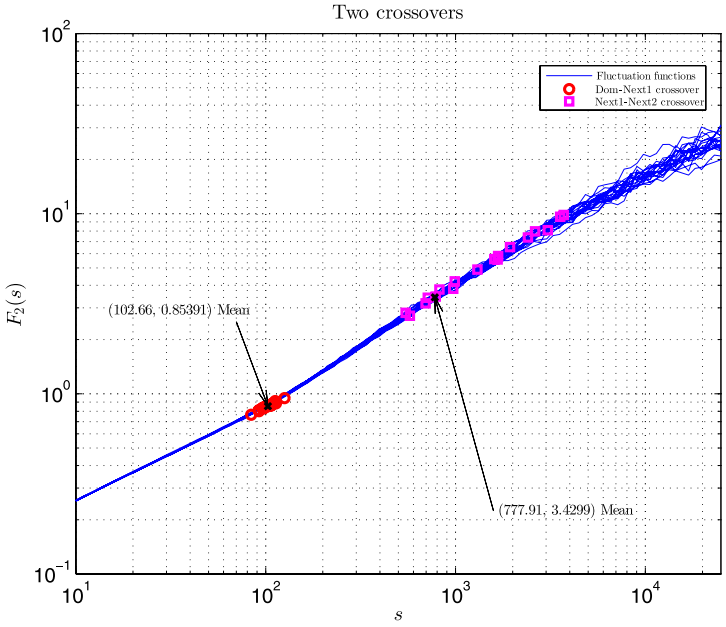


Fig. 4. Results of the criterion applied to 20 series with theoretical crossover scales at $s_1 = 10^2$ and $s_2 = 10^3$. All observed crossover points are of the type 'Dominant–Next1' and 'Next1–Next2'.

where $1 \leq k \leq 2^{16}$ and $n(k)$ is the number of digits equal to 1 in the binary representation of k. This series has a theoretical multifractal spectrum given by

$$h(q) = \frac{1}{q} - \frac{\ln\left[a^q + (1-a)^q\right]}{q \ln(2)}.$$

Fig. 5 shows results of the algorithm applied to the series with $a = 0.75$ (we used MF-DFA3). The 'Dominant' region $\left[10, 2^{14}\right]$ has a better agreement with theoretical values than other regions (Fig. 5(d1)). We have also confirmed a good agreement with theoretical curves for $a = 0.6$ and $a = 0.9$.

5. Application to sunspot time series analysis

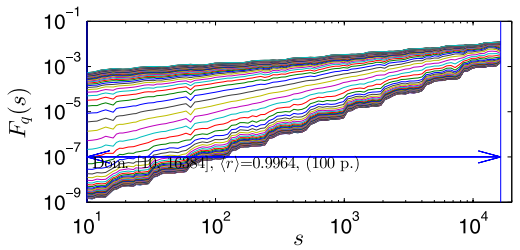
The analysis of natural time series represents a harder challenge than previous examples in this work because they are inherently much more complex. In this section we apply the proposed criterion for characterizing the fractal and multifractal structure associated with the sunspot number fluctuations.

Although sunspots have been observed for millennia [44,45], only telescopic observation since 1610 allowed the systematic study of this phenomenon [44]. The number of these spots of relatively lower temperature on the surface of the sun during a given period of time is a good indicator of its activity [46]. The link between solar dynamical activity and Earth climate has been extensively investigated [46–50], therefore making this area of research very active. General cycles have been identified [46,51], while more recent efforts have been focused on the characterization of the 'noise' in the time series mounted on these general cycles [52,53].

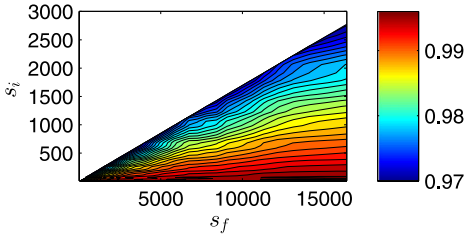
We have analyzed the monthly International Sunspot Number series publicly available at the SIDC's website (http://sidc.oma.be/sunspot-data/). This dataset begins in January of 1749 and is regularly updated. It has been studied with DFA (and MF-DFA) in the past: up to the year 2006 by Movahed et al. [54], by Hu et al. [52] (until 2008), and more recently by Zhou et al. [53] (until 2009). The monthly dataset used in this article spans the January 1749–November 2012 period (3167 months).

It is well known that trends notably affect the estimation of the real correlation nature related to the intrinsic fluctuations [55]. Consequently, for a reliable detection of the scaling behavior in data series, it is essential to separate trends, usually due to external effects and not *a priori* known, from intrinsic fluctuations. DFA and MF-DFA allow a systematic elimination of polynomial trends of different order [23]. On the other hand, oscillatory trends, e.g. seasonal cycles in weather and climate series [56], may strongly disturb the correlation analysis and deserve a special treatment [57,18]. More precisely, extrinsic periodicities induce the presence of spurious crossovers in the associated fluctuation functions; this prevents a reliable estimation of the intrinsic scaling behavior.

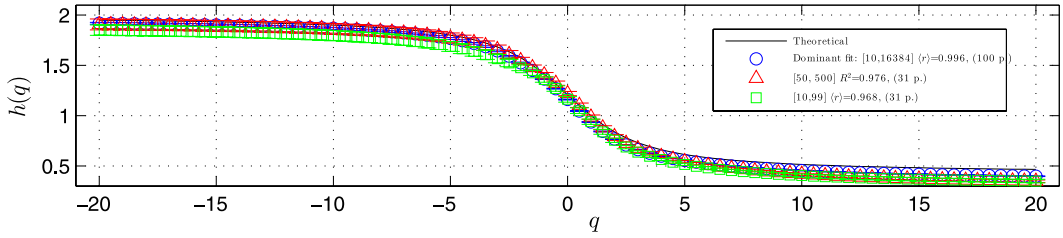
Taking into consideration the presence of the well-known 11 year period in solar dynamics, we have filtered the monthly sunspot record by implementing empirical mode decomposition (EMD) [58]. This technique has proven to be effective to deal with periodic trends [53]. Given a signal $x(t)$, the EMD algorithm decomposes the original series into a set of basic components called *intrinsic mode functions* (IMF_k) (a detailed description can be found in Refs. [58,59]). The series was



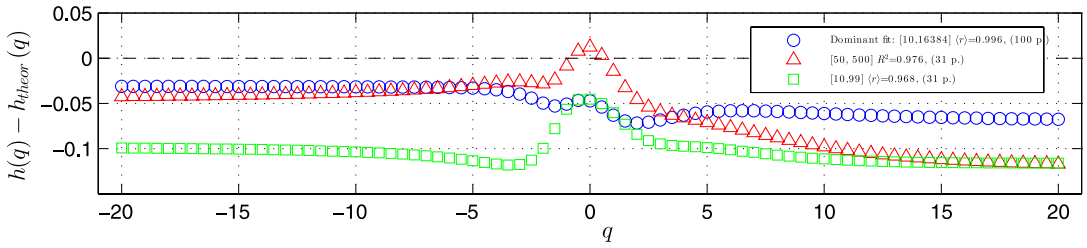
(a) MF-DFA3 fluctuation function binomial cascade ($a = 0.75$).



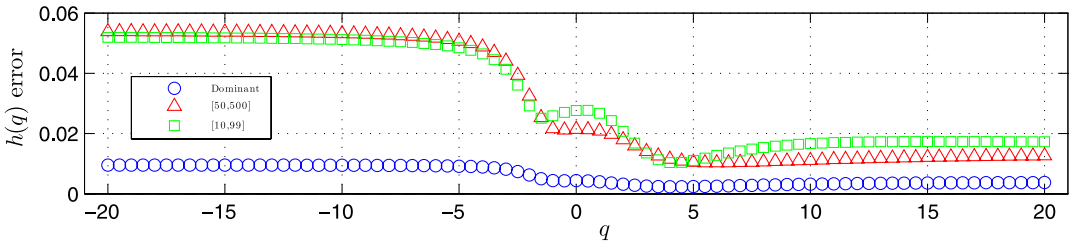
(b) $\langle r \rangle$ array, non-zero values.



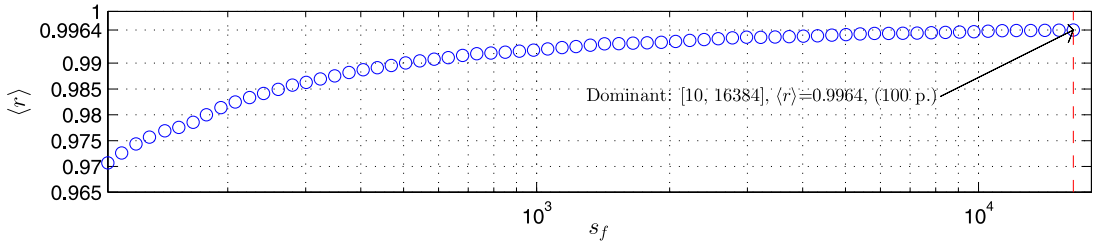
(c) MF-DFA3 multifractal spectrum.



(d1) MF-DFA3 (estimated spectrum—theoretical).



(d2) MF-DFA3—error magnitude.



(e) MF-DFA3—$\langle r \rangle$, $s_i = 10$.

Fig. 5. MF-DFA3 for the binomial multifractal ($a = 0.75$). (a) Fluctuation functions, $q = -20, -19.5, \ldots, 20$. (b) Contour plot of the $\langle r \rangle$ matrix of the series (non-zero values, $\delta = 25$). (c) Theoretical multifractal spectrum and spectra estimated: 'Dominant' region ($[10, 2^{14}]$)—100 data points fit ($\langle r \rangle \simeq 0.996$); a mid-range example [50, 500]—31 data points fit ($\langle r \rangle \simeq 0.976$), and [10, 99]—31 data points fit ($\langle r \rangle \simeq 0.968$). (d1) Quality of the three spectra in (c) shown as the difference between estimated and theoretical $h(q)$. (d2) Estimated error magnitude of the three spectra. (e) A section of contour plot of the $\langle r \rangle$ matrix (subfigure (b)): values of multifractal spectrum estimation fit between $s_i = 10$ and ending at s_f (variable); the final value is the absolute maximum of $\langle r \rangle \simeq 0.996$ and indicates the 'Dominant' region (red dashed vertical line) also depicted in (a): $[10, 2^{14}]$.

filtered using the MATLAB empirical mode decomposition toolbox by Rilling and Flandrin from Laboratoire de Physique CNRS and ENS Lyon (France) [60], downloaded from http://perso.ens-lyon.fr/patrick.flandrin/emd.html, and following the analysis performed by Zhou et al. [53] (including the iteration stop criterion). The EMD decomposition of the original monthly

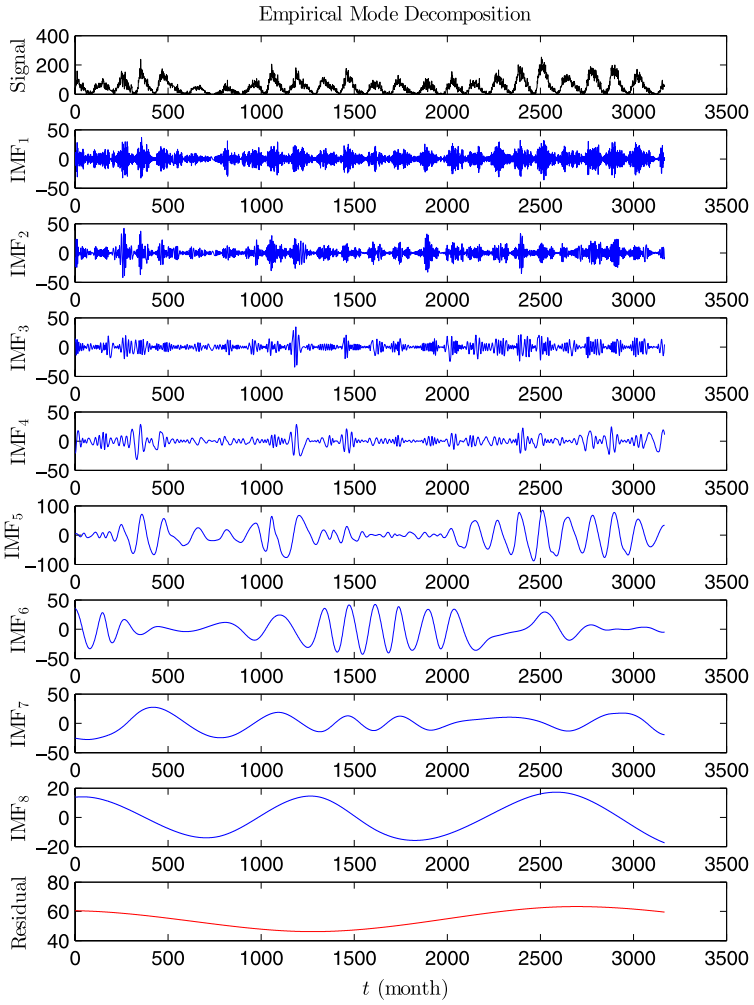


Fig. 6. EMD decomposition of monthly sunspot time series. IMF_1 to IMF_8 are depicted from top to bottom; the residual is considered as IMF_9.

sunspot time series is depicted in Fig. 6. Partial sums of reconstruction are $S\,(i, j)$ according to

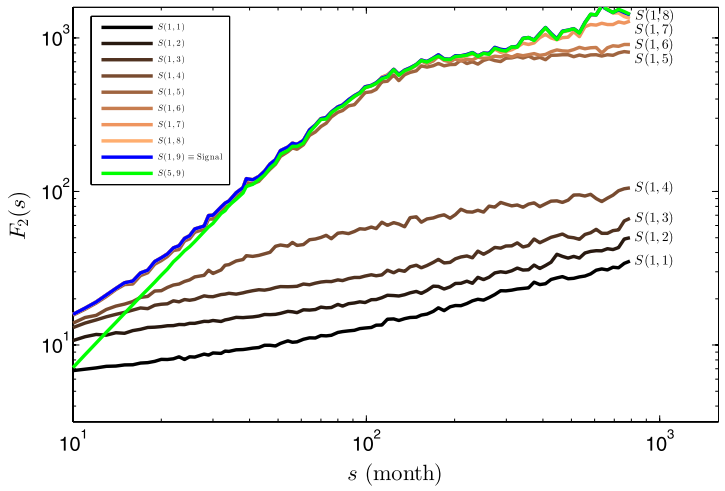
$$S\,(i, j) = \sum_{k=i}^{j} \mathrm{IMF}_k.$$

In Fig. 7(a) we plot DFA fluctuation functions for $S\,(1, j)$ $(1 \leq j \leq 9)$. An abrupt transition can be easily observed between the scaling obtained for $S\,(1, 4)$ and $S\,(1, 5)$. These results are comparable with those reported by Zhou et al. [53]. Scaling exponents and crossover scales are estimated from $S\,(1, 4)$ (Fig. 7(b)) since all cycles longer than 11 years are filtered in this reconstruction. The 'Dominant' region was found to be [10, 56] with $h = 0.702 \pm 0.007$ ($R^2 = 0.996$, 34 data points); the 'Next' region was found to be [56, 581] with $h = 0.306 \pm 0.008$ ($R^2 = 0.969$, 54 data points). The crossover scale is therefore estimated to be $s_c = 57.441$ months in complete agreement with the position of the crossover point ($s_c = 57.544$ months) given in Ref. [53].

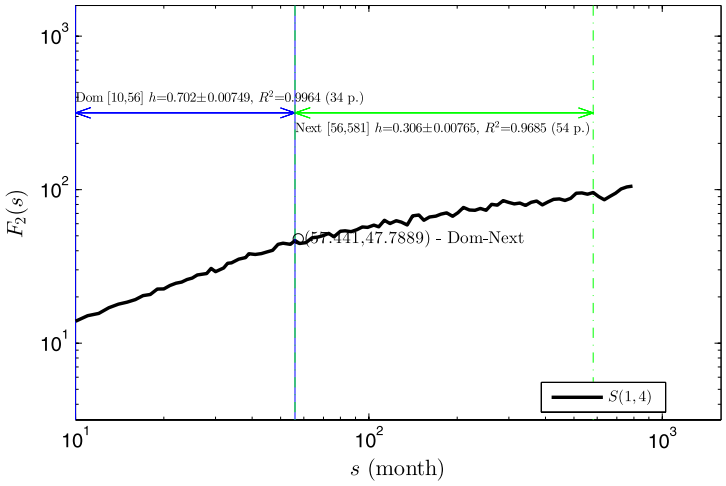
MF-DFA results for $S\,(1, 4)$ (MF-DFA1 to MF-DFA4) are shown in Fig. 8. At the top we plot the fluctuation functions and show the 'Dominant' interval determined by the MF-DFA generalization of the criterion; below we plot the multifractal spectrum of each 'Dominant' region. Results confirm a multifractal behavior of the series for negative q and a monofractal behavior for positive values of q; this general behavior does not change much from MF-DFA2 on. It is worth remarking here that this finding is in line with the multifractal spectrum obtained by Hu et al. (please compare with Fig. 12 of Ref. [52]).

6. Conclusions

We have proposed and employed a conceptually simple criterion to systematically determine optimal fitting regions for fluctuation functions in DFA; and generalized it for MF-DFA. This method provides a classification for such optimal scaling regions ('Previous', 'Dominant', 'Next') based on the quality of a linear fit of the fluctuation function in a log–log plot. When used in DFA, this technique helps to identify crossover scales between different power law regimes. When used in MF-DFA, it estimates the fitting range of the multifractal spectrum with the smallest error bars.



(a) DFA1 versus level of reconstruction.



(b) $S(1, 4)$-DFA1.

Fig. 7. (a) DFA fluctuation functions for $S(1, j)$ $(1 \leq j \leq 9)$. Note the transition between the scaling obtained for $S(1, 4)$ and $S(1, 5)$. (b) DFA fluctuation function for $S(1, 4)$. The 'Dominant' region is $[10, 56]$ with $h = 0.702 \pm 0.007$ ($R^2 = 0.996$, 34 data points); the 'Next' region is $[56, 581]$ with $h = 0.306 \pm 0.008$ ($R^2 = 0.969$, 54 data points). Crossover scale is $s_c = 57.441$ months.

Fit regions are a critical point in DFA and MF-DFA and their optimal identification is a crucial task for a genuine determination of the underlying scaling phenomenon. Both mono- and multifractal cases are often scale-size dependent; the criterion proposed in this work helps to identify scaling and multi-scaling behaviors in a more precise, reliable and objective way. Moreover, crossover scales that separate two or more regimes can be easily estimated.

This criterion can be automated for the analysis of large sets of time series since it does not require any previous knowledge about the raw data. Furthermore, real-time variants of the DFA introduced for monitoring and forecasting dynamically changing signals (medical, meteorological, stock market, etc.) [61] and local DFA and MF-DFA analysis devised to follow the fractal and multifractal time evolution properties in data [62] can be particularly benefited from its implementation. Even though it is brute force in nature, the number of calculations involved renders it computationally practical.

In a future work, we plan to analyze in depth a potential link between the optimal fit region derived from the proposed algorithm and the range of temporal correlation present in the system under study. As Drożdż et al. [63] and W.-X. Zhou [17] have shown, genuine multifractality only originates from nonlinear temporal correlations. Consequently, a significant correlation between these two ranges is expected. The confirmation of this hypothesis could help to shed some light on the reasons behind scaling range and the presence of cutoffs in empirical fractals [64–66].

Acknowledgments

Damián Gulich and Luciano Zunino were supported by Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina. The authors would like to thank the reviewers for their helpful comments, remarks and suggestions that led to a significant improvement of the manuscript.

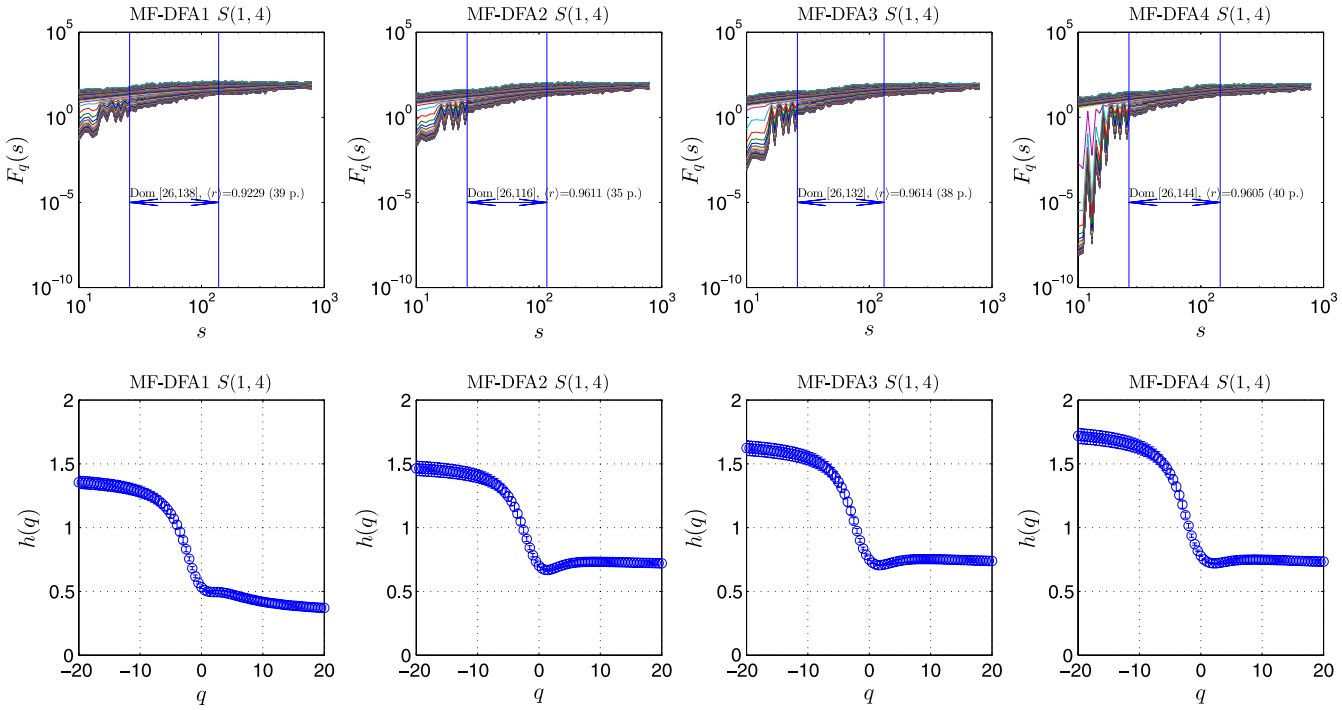


Fig. 8. Above: fluctuation functions of monthly sunspot data based on MF-DFA with growing detrending polynomial order; 'Dominant' ranges are indicated. Below: multifractal spectra calculated in those ranges.

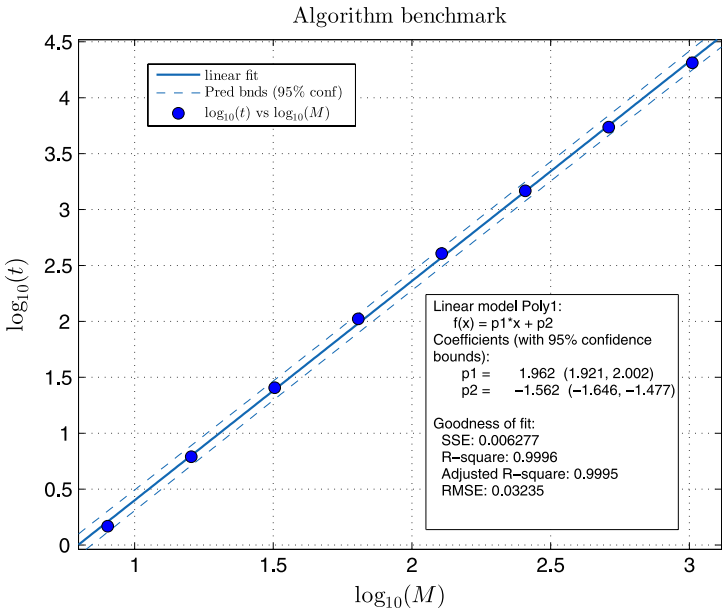


Fig. 9. Time performance of algorithm implementation in MATLAB.

Appendix. Benchmark

We have tested the performance of the implementation of the algorithm used in this paper by means of the MF-DFA3 analysis of a binomial multifractal series as described in Section 4.4 (we chose MF-DFA to study the most computationally intensive version of the analysis). We generated a series of 2^{16} values with $a = 0.75$ in a desktop PC.[8] We then calculated fluctuation functions ($q = -20, -19.5, \ldots, 20$) with $M = 8, 16, 32, \ldots, 1024$ scales and ran the MATLAB implementation using $\delta = M/4$ in every case.

[8] Processor AMD FX(tm)-8350 eight-core processor, 4013 MHz, four cores, eight logical processors; total physical memory 15.9 GB; available physical memory 11.7 GB; total virtual memory 31.8 GB; available virtual memory 27.0 GB. Only one core (out of eight) was charged during the test.

We have found that (for this implementation) the computing time t (measured in seconds) in the tested system (please see Fig. 9) may be approximated by

$$t\,(M, \#C) \simeq 0.0316\ \mathrm{s}\frac{M^2}{\#C},$$

where $\#C$ is the number of active cores. This inverse proportionality is due to the fact that elements in the $\langle r \rangle$ matrix are independent of each other, therefore enabling the possibility to split the processing task.

References

[1] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Mosaic organization of DNA nucleotides, Phys. Rev. E 49 (2) (1994) 1685–1689.
[2] Jan W. Kantelhardt, Stephan A. Zschiegner, Eva Koscielny-Bunde, Shlomo Havlin, Armin Bunde, H.Eugene Stanley, Multifractal detrended fluctuation analysis of nonstationary time series, Physica A 316 (1–4) (2002) 87–114.
[3] L. Zunino, B.M. Tabak, A. Figliola, D.G. Pérez, M. Garavaglia, O.A. Rosso, A multifractal approach for stock market inefficiency, Physica A 387 (26) (2008) 6558–6566.
[4] Petre Caraiani, Evidence of multifractality from emerging European stock markets, PLoS One 7 (7) (2012) e40693.
[5] Luciano Telesca, Vincenzo Lapenna, Maria Macchiato, Mono- and multi-fractal investigation of scaling properties in temporal patterns of seismic sequences, Chaos Solitons Fractals 19 (1) (2004) 1–15.
[6] C.-K. Peng, Shlomo Havlin, H.Eugene Stanley, Ary L. Goldberger, Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series, Chaos 5 (1) (1995) 82–87.
[7] Nandan Das, Subhasri Chatterjee, Jalpa Soni, Jaidip Jagtap, Asima Pradhan, Tapas K. Sengupta, Prasanta K. Panigrahi, I. Alex Vitkin, Nirmalya Ghosh, Probing multifractality in tissue refractive index: prospects for precancer detection, Opt. Lett. 38 (2) (2013) 211–213.
[8] M.Sadegh Movahed, F. Ghasemi, Sohrab Rahvar, M. Reza Rahimi Tabar, Long-range correlation in cosmic microwave background radiation, Phys. Rev. E 84 (2) (2011) 021103.
[9] G.Z. dos Santos Lima, M.A. Corrêa, R.L. Sommer, F. Bohn, Multifractality in domain wall dynamics of a ferromagnetic film, Phys. Rev. E 86 (6) (2012) 066117.
[10] Leonardo Dagdug, Jose Alvarez-Ramirez, Carlos Lopez, Rodolfo Moreno, Enrique Hernandez-Lemus, Correlations in a Mozart's music score (k-73x) with palindromic and upside-down structure, Physica A 383 (2) (2007) 570–584.
[11] G.R. Jafari, P. Pedram, L. Hedayatifar, Long-range correlation and multifractality in Bach's inventions pitches, J. Stat. Mech. 2007 (4) (2007) P04012.
[12] Luciano Telesca, Michele Lovallo, Revealing competitive behaviours in music by means of the multifractal detrended fluctuation analysis: application to Bach's Sinfonias, Proc. R. Soc. A 467 (2134) (2011) 3022–3032.
[13] Holger Hennig, Ragnar Fleischmann, Anneke Fredebohm, York Hagmayer, Jan Nagler, Annette Witt, Fabian J. Theis, Theo Geisel, The nature and perception of fluctuations in human musical rhythms, PLoS One 6 (10) (2011) e26457.
[14] Luciano Telesca, Michele Lovallo, Analysis of temporal fluctuations in Bach's sinfonias, Physica A 391 (11) (2012) 3247–3256.
[15] Dariusz Grech, Grzegorz Pamuła, Multifractal background noise of monofractal signals, Acta Phys. Polon. A 121 (2B) (2012) B34–B39.
[16] Juan Luis López, Jesús Guillermo Contreras, Performance of multifractal detrended fluctuation analysis on short time series, Phys. Rev. E 87 (2) (2013) 022918.
[17] Wei-Xing Zhou, Finite-size effect and the components of multifractality in financial volatility, Chaos Solitons Fractals 45 (2) (2012) 147–155.
[18] Josef Ludescher, Mikhail I. Bogachev, Jan W. Kantelhardt, Aicko Y. Schumann, Armin Bunde, On spurious and corrupted multifractality: the effects of additive noise, short-term memory and periodic trends, Physica A 390 (13) (2011) 2480–2490.
[19] Damián Gulich, Luciano Zunino, The effects of observational correlated noises on multifractal detrended fluctuation analysis, Physica A 391 (16) (2012) 4100–4110.
[20] Ying-Hui Shao, Gao-Feng Gu, Zhi-Qiang Jiang, Wei-Xing Zhou, Didier Sornette, Comparing the performance of FA, DFA and DMA using different synthetic long-range correlated time series, Sci. Rep. 2 (2012) 835.
[21] Erjia Ge, Yee Leung, Detection of crossover time scales in multifractal detrended fluctuation analysis, J. Geogr. Syst. 15 (2) (2013) 115–147.
[22] R.B. Govindan, J.D. Wilson, H. Preißl, H. Eswaran, J.Q. Campbell, C.L. Lowery, Detrended fluctuation analysis of short datasets: an application to fetal cardiac data, Physica D 226 (1) (2007) 23–31.
[23] Amir Bashan, Ronny Bartsch, Jan W. Kantelhardt, Shlomo Havlin, Comparison of detrending methods for fluctuation analysis, Physica A 387 (21) (2008) 5080–5090.
[24] Sebastian Michalski, Blocks adjustment—reduction of bias and variance of detrended fluctuation analysis using Monte Carlo simulation, Physica A 387 (1) (2008) 217–242.
[25] Dariusz Grech, Zygmunt Mazur, On the scaling ranges of detrended fluctuation analysis for long-term memory correlated short series of data, Physica A 392 (10) (2013) 2384–2397.
[26] Dariusz Grech, Zygmunt Mazur, Scaling range of power laws that originate from fluctuation analysis, Phys. Rev. E 87 (5) (2013) 052809.
[27] E. Alessio, A. Carbone, G. Castelli, V. Frappietro, Second-order moving average and scaling of stochastic time series, Eur. Phys. J. B 27 (2) (2002) 197–200.
[28] Jose Alvarez-Ramirez, Eduardo Rodriguez, Juan Carlos Echeverría, Detrending fluctuation analysis based on moving average filtering, Physica A 354 (2005) 199–219.
[29] Ken Kiyono, Zbigniew R. Struzik, Naoko Aoyagi, Fumiharu Togo, Yoshiharu Yamamoto, Phase transition in a healthy human heart rate, Phys. Rev. Lett. 95 (5) (2005) 058101.
[30] M. Staudacher, S. Telser, A. Amann, H. Hinterhuber, M. Ritsch-Marte, A new method for change-point detection developed for on-line analysis of the heart beat variability during sleep, Physica A 349 (3–4) (2005) 582–596.
[31] Boris Podobnik, H.Eugene Stanley, Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series, Phys. Rev. Lett. 100 (8) (2008) 084102.
[32] Wei-Xing Zhou, Multifractal detrended cross-correlation analysis for two nonstationary signals, Phys. Rev. E 77 (6) (2008) 066211.
[33] Gao-Feng Gu, Wei-Xing Zhou, Detrending moving average algorithm for multifractals, Phys. Rev. E 82 (1) (2010) 011136.
[34] Zhi-Qiang Jiang, Wei-Xing Zhou, Multifractal detrending moving-average cross-correlation analysis, Phys. Rev. E 84 (1) (2011) 016106.
[35] D. Horvatic, H.E. Stanley, B. Podobnik, Detrended cross-correlation analysis for non-stationary time series with periodic trends, Europhys. Lett. 94 (1) (2011) 18007.
[36] Xi-Yuan Qian, Gao-Feng Gu, Wei-Xing Zhou, Modified detrended fluctuation analysis based on empirical mode decomposition for the characterization of anti-persistent processes, Physica A 390 (23–24) (2011) 4388–4395.
[37] J.W. Kantelhardt, Fractal and multifractal time series, arXiv:0804.0747 [physics.data-an].
[38] P. Oświęcimka, S. Drożdż, J. Kwapień, A.Z. Górski, Effect of detrending on multifractal characteristics, Acta Phys. Polon. A 123 (3) (2013) 597–603.
[39] Jan W. Kantelhardt, Eva Koscielny-Bunde, Henio H.A. Rego, Shlomo Havlin, Armin Bunde, Detecting long-range correlations with detrended fluctuation analysis, Physica A 295 (3–4) (2001) 441–454.
[40] V. Livina, Z. Kizner, P. Braun, T. Molnar, A. Bunde, S. Havlin, Temporal scaling comparison of real hydrological data and model runoff records, J. Hydrol. 336 (1–2) (2007) 186–198.

[41] Pengjian Shang, Yongbo Lu, Santi Kamae, Detecting long-range correlations of traffic time series with multifractal detrended fluctuation analysis, Chaos Solitons Fractals 36 (1) (2008) 82–90.
[42] Patrice Abry, Fabrice Sellan, The wavelet-based synthesis for fractional Brownian motion proposed by F. Sellan and Y. Meyer: remarks and fast implementation, Appl. Comput. Harmon. Anal. 3 (4) (1996) 377–383.
[43] Aicko Y. Schumann, Jan W. Kantelhardt, Multifractal moving average analysis and test of multifractal model with tuned correlations, Physica A 390 (14) (2011) 2637–2654.
[44] Galileo Galilei, Christoph Scheiner, On Sunspots, University of Chicago Press, 2010.
[45] K.K.C. Yau, F.R. Stephenson, A revised catalogue of far Eastern observations of sunspots (165 BC to AD 1918), Q. J. R. Astron. Soc. 29 (1988) 175–197.
[46] X. Moussas, J.M. Polygiannakis, P. Preka-Papadema, G. Exarhos, Solar cycles: a tutorial, Adv. Space Res. 35 (5) (2005) 725–738.
[47] Samuli Helama, Jari Holopainen, Spring temperature variability relative to the North Atlantic oscillation and sunspots—a correlation analysis with a Monte Carlo implementation, Palaeogeogr. Palaeoclimatol. Palaeoecol. 326–328 (2012) 128–134.
[48] Jianjun Xu, Alfred M. Powell Jr., What happened to surface temperature with sunspot activity in the past 130 years? Theor. Appl. Climatol. 111 (3–4) (2013) 609–622.
[49] Pablo J.D. Mauas, Eduardo Flamenco, Andrea P. Buccino, Solar forcing of the stream flow of a continental scale South American river, Phys. Rev. Lett. 101 (16) (2008) 168501.
[50] Pablo J.D. Mauas, Andrea P. Buccino, Eduardo Flamenco, Long-term solar activity influences on South American rivers, J. Atmos. Sol.-Terr. Phys. 73 (2–3) (2011) 377–382.
[51] Stanisław Zieba, Zenon Nieckarz, Sunspot time series—relations inferred from the location of the longest spotless segments, Sol. Phys. 278 (2) (2012) 457–469.
[52] Jing Hu, Jianbo Gao, Xingsong Wang, Multifractal analysis of sunspot time series: the effects of the 11-year cycle and Fourier truncation, J. Stat. Mech. 2009 (2) (2009) P02066.
[53] Yu Zhou, Yee Leung, Empirical mode decomposition and long-range correlation analysis of sunspot time series, J. Stat. Mech. 2010 (12) (2010) P12006.
[54] M.Sadegh Movahed, G.R. Jafari, F. Ghasemi, Sohrab Rahvar, M. Reza Rahimi Tabar, Multifractal detrended fluctuation analysis of sunspot time series, J. Stat. Mech. 2006 (2) (2006) P02003.
[55] Kun Hu, Plamen Ch. Ivanov, Zhi Chen, Pedro Carpena, H.Eugene Stanley, Effect of trends on detrended fluctuation analysis, Phys. Rev. E 64 (1) (2001) 011114.
[56] Eva Koscielny-Bunde, Jan W. Kantelhardt, Peter Braun, Armin Bunde, Shlomo Havlin, Long-term persistence and multifractality of river runoff records: detrended fluctuation studies, J. Hydrol. 322 (1–4) (2006) 120–137.
[57] R. Nagarajan, R.G. Kavasseri, Minimizing the effect of sinusoidal trends in detrended fluctuation analysis, Int. J. Bifurcation Chaos 15 (5) (2005) 1767–1773.
[58] Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, Henry H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. A 454 (1971) (1998) 903–995.
[59] Zhaohua Wu, Norden E. Huang, Steven R. Long, Chung-Kang Peng, On the trend, detrending, and variability of nonlinear and nonstationary time series, Proc. Natl. Acad. Sci. USA 104 (38) (2007) 14889–14894.
[60] G. Rilling, P. Flandrin, P. Gonçalves, On empirical mode decomposition and its algorithms, in: IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03, 2003.
[61] András Hartmann, Péter Mukli, Zoltán Nagy, László Kocsis, Péter Hermán, András Eke, Real-time fractal signal processing in the time domain, Physica A 392 (1) (2013) 89–102.
[62] Luciano Telesca, Vincenzo Lapenna, Maria Macchiato, Multifractal fluctuations in earthquake-related geoelectrical signals, New J. Phys. 7 (2005) 214.
[63] S. Drożdż, J. Kwapień, P. Oświęcimka, R. Rak, Quantitative features of multifractal subtleties in time series, Europhys. Lett. 88 (6) (2009) 60003.
[64] Ofer Malcai, Daniel Lidar, Ofer Biham, David Avnir, Scaling range and cutoffs in empirical fractals, Phys. Rev. E 56 (3) (1997) 2817–2828.
[65] D. Avnir, Applied mathematics: is the geometry of nature fractal? Science 279 (5347) (1998) 39–40.
[66] B.B. Mandelbrot, Is nature fractal? Science 279 (5352) (1998) 783–784.