Available online at www.sciencedirect.com
ScienceDirect
Physica A 382 (2007) 9–15
PHYSICA A
www.elsevier.com/locate/physa

De trending moving average algorithm: A closed-form approximation of the scaling law

Sergio Arianos*[,1], Anna Carbone

Dipartimento di Fisica, Politecnico di Torino, Corso Duca degli Abruzzi 24, I-10129 Torino, Italy

Available online 1 March 2007

Abstract

The Hurst exponent H of long range correlated series can be estimated by means of the detrending moving average (DMA) method. The computational tool, on which the algorithm is based, is the generalized variance $\sigma^2_{DMA} = 1/(N-n)\sum_{i=n}^{N}[y(i) - \widetilde{y}_n(i)]^2$, with $\widetilde{y}_n(i) = 1/n\sum_{k=0}^{n} y(i-k)$ being the average over the moving window n and N the dimension of the stochastic series $y(i)$. The ability to yield H relies on the property of σ^2_{DMA} to vary as n^{2H} over a wide range of scales [E. Alessio, A. Carbone, G. Castelli, V. Frappietro, Eur. J. Phys. B 27 (2002) 197]. Here, we give a closed form proof that σ^2_{DMA} is equivalent to $C_H n^{2H}$ and provide an explicit expression for C_H. We furthermore compare the values of C_H with those obtained by applying the DMA algorithm to artificial self-similar signals.

Keywords: Hurst exponent; Moving average; DMA algorithm

1. Introduction

Long-memory stochastic processes are ubiquitous in fields as different as condensed matter, biophysics, social sciences, climate changes, finance [1–4]. The development of methods able to quantify the statistical properties and, in particular, to extract the Hurst exponent of long-range correlated signals continue therefore to draw the attention not only of the physicist community [5–15]. For long-memory correlated processes, the value of the Hurst exponent H ranges from $0<H<0.5$ and from $0.5<H<1$, respectively, for negative and positive persistence; $H = 0.5$ is found in fully uncorrelated signals. Several techniques have been proposed in the literature to study the scaling properties of time series. We limit ourselves to mention here only a few of them such as the seminal work by Hurst on rescaled range statistical (R/S) analysis, the modified R/S analysis, the multi-affine analysis, the detrended fluctuation analysis (DFA), the periodogram regression (GPH) method, the (m,k)-Zipf method, the detrended moving average (DMA) analysis. The challenge is to get the Hurst exponent H, that is related to the fractal dimension $D = 2 - H$, by means of more and more accurate

*Corresponding author.
E-mail addresses: arianos@to.infn.it (S. Arianos), anna.carbone@polito.it (A. Carbone).
URL: http://www.polito.it/noiselab (A. Carbone).
[1]Permanent address: Theoretical Physics Department, Universitá di Torino, Italy.

0378-4371/$ - see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.physa.2007.02.074

and fast algorithms. The methods of extraction of the scaling exponents from a random signal exploit suitable statistical functions of the series itself.

Recently, a method called DMA technique for the analysis of the persistence has been proposed. The striking difference between the DMA and other (R/S, DFA) variance methods is that the DMA algorithm does not need a division of the series in boxes. The scaling property is obtained by using a simple continuous function: the moving average. This fact makes the DMA algorithm highly efficient from the computational point of view. The scaling properties of the DMA variance have been studied and applications have been demonstrated in previous work [13,14].

The purpose of this work is to derive a closed form approximation of the scaling behavior of the DMA variance at large n, i.e., $\sigma^2_{DMA} \sim C_H n^{2H}$. Furthermore, the expression $C_H n^{2H}$ is compared with the data obtained by applying the DMA algorithm to surrogate series with assigned H. For such comparison, we use 30 samples of surrogate series with $N = 2^{23}$. Very long signals assure the consistency of the simulation results with the expression $C_H n^{2H}$ holding at large n. The surrogate series are generated by the random midpoint displacement (RMD) algorithm [16], that is preferable over other signal generators for its high execution speed though it suffers of poorer accuracy [17].

2. Method

First we describe the main steps of the DMA algorithm. The technique is based on the function:

$$\sigma^2_{DMA} = \frac{1}{N-n} \sum_{i=n}^{N} [y(i) - \tilde{y}_n(i)]^2, \tag{1a}$$

$$\tilde{y}_n(i) = \frac{1}{n} \sum_{k=0}^{n} y(i-k). \tag{1b}$$

Eq. (1a) defines a generalized variance of the random path $y(i)$ with respect to the moving average $\tilde{y}_n(i)$ (Eq. (1b)). The function $\tilde{y}_n(i)$ is calculated by averaging the n-past value in each sliding window of length n. In so doing, the reference point of the averaging process is the last point of the window. The dynamic averaging process and the DMA algorithm can be, however, referred to any point within the window, by generalizing Eqs. (1a), (1b) as follows:

$$\sigma^2_{DMA} = \frac{1}{N-n} \sum_{i=n(1-\theta)}^{N-n\theta} [y(i) - \tilde{y}_n(i)]^2, \tag{1c}$$

$$\tilde{y}_n(i) = \frac{1}{n} \sum_{k=-n\theta}^{n(1-\theta)} y(i-k). \tag{1d}$$

Upon variation of the parameter θ in the range $[0, 1]$, the reference point of $\tilde{y}_n(i)$ is accordingly set within the moving window n. In particular, we will consider the following three relevant cases: (i) $\theta = 0$ corresponding to calculate $\tilde{y}_n(i)$ over all the past points within the window n; (ii) $\theta = \frac{1}{2}$ corresponding to calculate $\tilde{y}_n(i)$ over $n/2$ past and $n/2$ future points within the window n and (iii) $\theta = 1$ corresponding to calculate $\tilde{y}_n(i)$ over all the future points within the window n.

In order to calculate the Hurst exponent of the series, the DMA algorithm is implemented as follows. The moving average $\tilde{y}_n(i)$ is calculated for different values of the window n, with n ranging from 2 to a maximum value n_{max} depending upon the size of the series. The σ_{DMA}, defined by Eqs. (1), is then calculated for all the windows n over the interval $[n, N]$. For each $\tilde{y}_n(i)$, the value of σ_{DMA} corresponding to each $\tilde{y}_n(i)$ is plotted as a function of n on log–log axes. The most remarkable property of the log–log plot is to exhibit a power-law dependence on n, i.e., $\sigma^2_{DMA} \sim n^{2H}$, allowing thus to calculate the scaling exponent H of the signal $y(i)$.

3. Derivation of the scaling relationship at large *n*

A closed form approximation of Eqs. (1) will be deduced in the limit of large n using the properties of the fractional Brownian path. We will obtain the following expression:

$$\sigma^2_{DMA} \sim C_H n^{2H}, \quad n \gg 1, \tag{2}$$

with

$$C_H = \frac{(1-\theta)^{2H+1} + \theta^{2H+1}}{2H+1} - \frac{1}{2(H+1)(2H+1)}. \tag{3}$$

By simple transformations, Eq. (1c) can be written as

$$(N-n)\sigma^2_{DMA} = \sum_{i=n-\theta n}^{N-\theta n} y^2(i) - \frac{2}{n}\sum_{i=n-\theta n}^{N-\theta n} y(i) \sum_{k=-\theta n}^{n-\theta n} y(i-k) + \frac{1}{n^2}\sum_{i=n-\theta n}^{N-\theta n}\left(\sum_{k=-\theta n}^{n-\theta n} y(i-k)\right)^2. \tag{4}$$

Let us consider each term on the right-hand side of Eq. (4) separately. The first term writes

$$\sum_{i=n-\theta n}^{N-\theta n} y^2(i) = \sum_{i=n-\theta n}^{N-\theta n} i^{2H} \simeq \frac{1}{2H+1}[(N-\theta n)^{2H+1} - (n-\theta n)^{2H+1}]. \tag{5}$$

The second term writes

$$\begin{aligned} -\frac{2}{n}\sum_{i=n-\theta n}^{N-\theta n} y(i) \sum_{k=-\theta n}^{n-\theta n} y(i-k) &= -\frac{2}{n}\sum_{i=n-\theta n}^{N-\theta n} y(i) \sum_{j=i-n+\theta n}^{i+\theta n} y(j) \\ &\simeq -\frac{1}{2H+1}[(N-\theta n)^{2H+1} - (n-\theta n)^{2H+1}] \\ &\quad - \frac{1}{2(H+1)(2H+1)}\frac{1}{n}[N^{2H+2} - n^{2H+2} - (N-n)^{2H+2}] \\ &\quad + \frac{n^{2H}}{2H+1}\left[(1-\theta)^{2H+1} + \theta^{2H+1}\right](N-n). \end{aligned} \tag{6}$$

The third term writes

$$\begin{aligned} \frac{1}{n^2}\sum_{i=n-\theta n}^{N-\theta n}\left(\sum_{k=-\theta n}^{n-\theta n} y(i-k)\right)^2 &= \frac{1}{n^2}\sum_{i=n-\theta n}^{N-\theta n}\left[\sum_{j=i+\theta n-n}^{i+\theta n} y(j)\right]^2 \\ &\simeq \frac{1}{2(H+1)(2H+1)}\frac{1}{n}[N^{2H+2} - n^{2H+2} - (N-n)^{2H+2}] \\ &\quad - \frac{n^{2H}}{2(H+1)(2H+1)}(N-n). \end{aligned} \tag{7}$$

Summing the contributions from each term, one obtains

$$\sigma^2_{DMA} \simeq \left[\frac{(1-\theta)^{2H+1} + \theta^{2H+1}}{2H+1} - \frac{1}{2(H+1)(2H+1)}\right] n^{2H}. \tag{8}$$

One can easily check that the term in square brackets in Eq. (8) takes, respectively, the following expressions:

- for $\theta = 0$ or $\theta = 1$, i.e., when the moving average is referred to the last or to the first point of the window, it is:

$$C_H = \frac{1}{2(H+1)}, \tag{9}$$

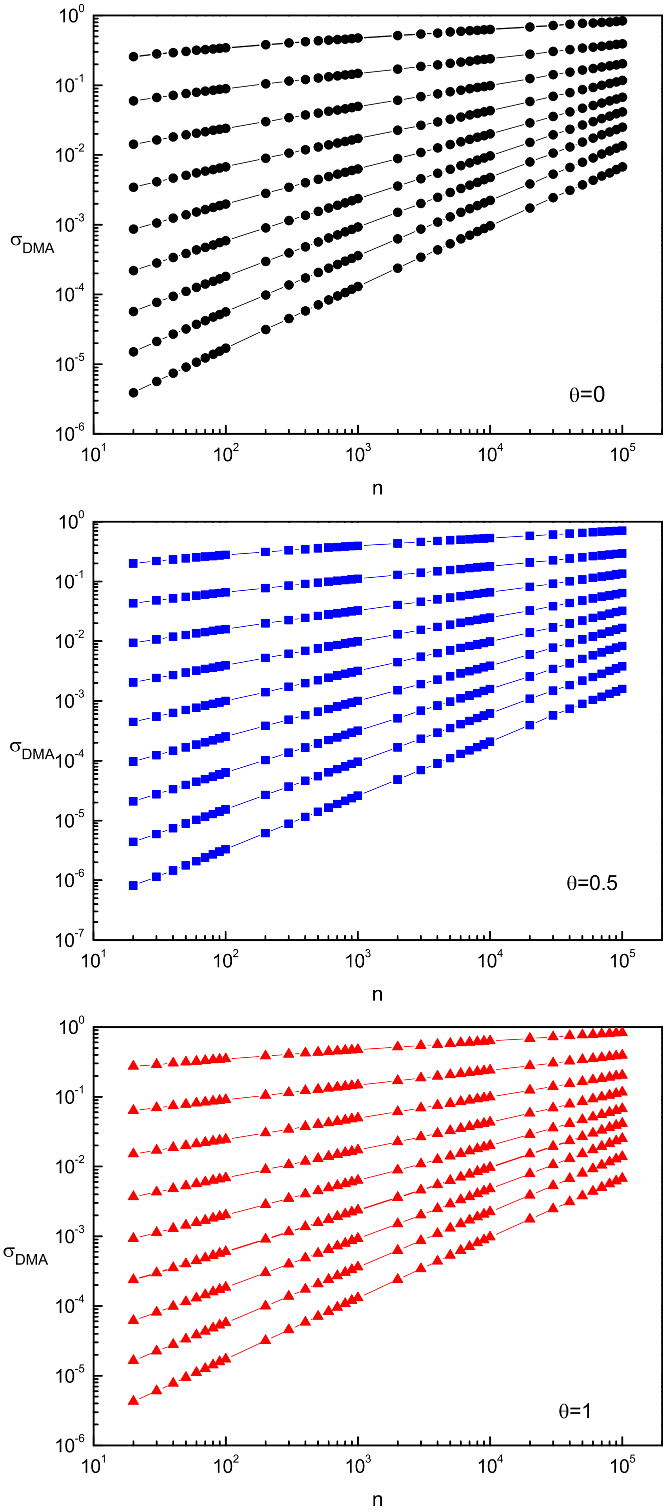

Fig. 1. (Color online). Log–log plot of the function σ_{DMA} defined by the Eq. (1) for artificial series generated by the random midpoint displacement (RMD) algorithm. The series have length $N = 2^{23}$ and Hurst exponent varying from 0.1 to 0.9 with step 0.1. The parameter θ is taken equal to 0, 0.5 and 1, respectively.

- for $\theta = \frac{1}{2}$ (i.e., when the moving average is referred to the center of the window) it is:

$$C_H = \frac{1}{2(H+1)} - \frac{1-2^{-2H}}{2H+1} \tag{10}$$

The above calculations have been performed for a fractional Brownian motion with variance $\sigma^2 = t^{2H}$. For the general case of a fractional Brownian motion with variance $\sigma^2 = D_H t^{2H}$, Eq. (8) asymptotically behaves as $\sigma^2_{DMA} = D_H C_H n^{2H}$.

4. Results and discussion

In this section, the values of C_H obtained by calculating the DMA variance of artificial fractional Brownian motions are compared with those calculated using Eq. (8).

In Fig. 1, the results of the DMA algorithm implemented over artificial fractional random walks generated by the RMD algorithm are shown. We apply the DMA algorithm to 30 samples of random walks with $N = 2^{23}$ and the Hurst exponent ranging from 0.1 to 0.9 with step 0.1. The curves in the three figures refer, respectively, to three values of the parameter θ, namely $\theta = 0$, $\theta = 0.5$ and $\theta = 1$. The slopes of the logarithms of the data plotted in Fig. 1 are shown in Fig. 2. The slopes and the intercepts of the logarithms of the data plotted in Fig. 1 are reported in Table 1. From the data shown in Table 1, it is possible to deduce that the DMA with $\theta = 0.5$ performs better with positively correlated signals with $0.5 < H < 1$, while the DMA with $\theta = 0$ and $\theta = 1$ performs better with negatively correlated signals with $0 < H < 0.5$. In Fig. 3, the theoretical values of C_H, calculated by using Eq. (8), are compared with those obtained by the intercepts of the curves plotted in Fig. 1 for $\theta = 0.5$ (data of the 2nd column of Table 1). Since the fractional Brownian motions, used for the simulations plotted in Fig. 1, have been generated by the RMD algorithm, in the calculation of C_H it must be kept in mind that $\sigma^2_{RMD} = (1 - 2^{2H-2})/2^{2Hv}\sigma^2_{Gauss}$, v being the number of steps of the RMD algorithm.

It is interesting to compare Eqs. (9), (10) with the corresponding ones obtained for the DFA algorithm. According to the DFA method, the integrated profile $y(i)$ is divided into boxes of equal length n. In each box, the signal $y(i)$ is best-fitted by an ℓ-order polynomial $y_{n,\ell}(i)$, which represents the local trend in that box. The different order of the DFA-ℓ (e.g., DFA-0 if $\ell = 0$, DFA-1 if $\ell = 1$, DFA-2 if $\ell = 2$, etc.) is obtained according to the order of the polynomial fit. Finally, the variance:

$$\sigma^2_{DFA-\ell} \equiv \frac{1}{N}\sum_{i=1}^{N}[y(i) - y_{n,\ell}(i)]^2 \tag{11}$$

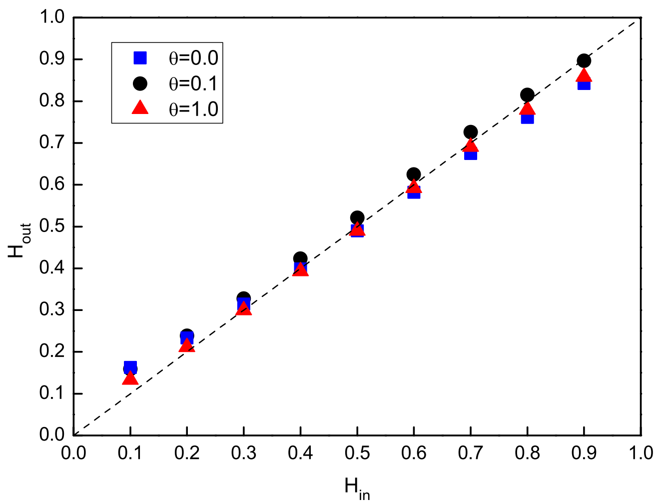

Fig. 2. (Color online). Values of the slopes of the function σ_{DMA} plotted in Fig. 1 for three different values of the parameter θ, respectively, equal to 0, 0.5 and 1.

Table 1
Intercept (A) and slope (B) of the logarithms of the data plotted in Fig. 1

H	$\theta = 0$		$\theta = 0.5$		$\theta = 1$	
	A	B	A	B	A	B
0.1	−0.73211	0.13220	−0.84280	0.14151	−0.71759	0.12861
0.2	−1.48755	0.21675	−1.62576	0.22023	−1.47214	0.21294
0.3	−2.23912	0.30962	−2.41792	0.30952	−2.22270	0.30555
0.4	−2.99192	0.40900	−3.20887	0.40212	−2.97326	0.40421
0.5	−3.72180	0.50752	−4.00245	0.50058	−3.70336	0.50301
0.6	−4.45987	0.61307	−4.80327	0.60187	−4.44063	0.60843
0.7	−5.17084	0.71296	−5.60555	0.70333	−5.14974	0.70776
0.8	−5.84043	0.79625	−6.40763	0.79829	−5.82297	0.79277
0.9	−6.51500	0.87257	−7.27298	0.89698	−6.49215	0.86705

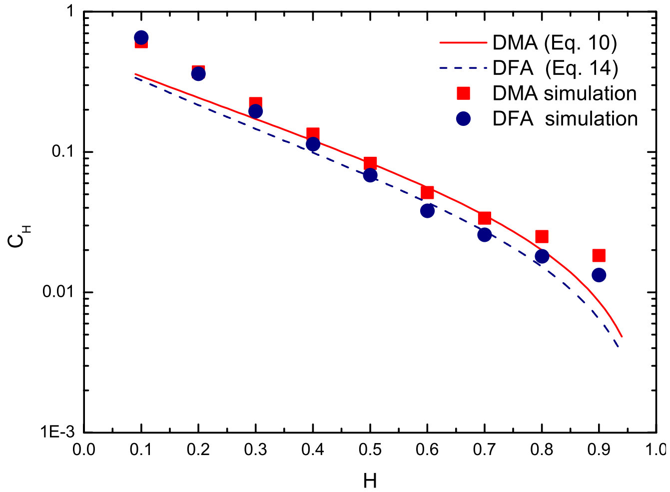

Fig. 3. (Color online). Values of C_H for the DMA with $\theta = 0.5$ (red squares) and for the DFA-1 (blue circles) algorithms applied to the same series. The solid lines represent the values of C_H calculated by using expressions (8) and (14), respectively.

is calculated for each box n. The calculation is then repeated for different box lengths n, yielding the behavior of $\sigma_{DFA-\ell}$ over a broad range of scales. For scale-invariant signals with power-law correlations, the following relationship between the function $\sigma_{DFA-\ell}$ and the scale n holds:

$$\sigma^2_{DFA-\ell} \sim n^{2H}. \tag{12}$$

The asymptotic behavior of the $DFA-0$ and $DFA-1$ functions has been derived in [17,18]. The following relation (Eq. (21) of Ref. [18]) has been worked out for the σ_{DFA-0} function:

$$\sigma^2_{DFA-0} \simeq \left[\frac{1}{2H+1} - \frac{1}{2(H+1)}\right] n^{2H}. \tag{13}$$

It is easy to check that the "scaled windowed variance without any trend correction" in Ref. [18] is indeed equivalent to the $DFA-0$ variance. The function obtained by fitting the random walks $y(i)$ by constant segment in each box corresponds indeed to a *zero-order approximation* of the trend of $y(i)$.

The asymptotic behavior of the DFA-1 function, as reported in the Appendix of Ref. [17], is

$$\sigma^2_{DFA-1} \simeq \left[\frac{2}{2H+1} + \frac{1}{H+2} - \frac{2}{H+1}\right] n^{2H}. \tag{14}$$

In Fig. 3, the values of C_H for the DMA (with $\theta = 0.5$) and the $DFA - 1$ are shown. It can be observed that the behavior of C_H obtained from the simulations (square and circles) follows quite well the analytical curves (solid lines) around $H \simeq 0.5$. Deviations are observed at the extrema of the H range. Such deviations might be related to the accuracy either of the DFA and DMA techniques or of the RMD signal generator.

5. Conclusions

We have derived the asymptotic scaling behavior of the DMA algorithm (Eq. (8)) for an arbitrary value of the reference point of the function $\tilde{y}_n(i)$. The values of C_H are compared with those yielded by the simulations of fractional Brownian paths with assigned values of H generated by the RMD algorithm. A comparison between the behavior of C_H for the DMA (with $\theta = 0.5$) and the DFA (with $\ell = 1$) functions is also provided.

References

[1] R.N. Mantegna, H.E. Stanley, An Introduction to Econophysics: Correlations and Complexity in Finance, Cambridge University Press, New York, 2000.
[2] W. Willinger, R. Govindan, S. Jamin, V. Paxson, S. Shenker, PNAS 99 (2002) 2573.
[3] W.S. Lam, W. Ray, P.N. Guzdar, R. Roy, Phys. Rev. Lett. 94 (2005) 010602.
[4] S.O. Ferreira, et al., Appl. Phys. Lett. 88 (2006) 244102.
[5] J. Feder, Fractals, Plenum, New York, 1988.
[6] H.E. Hurst, Trans. Am. Soc. Civil Eng. 116 (1951) 770.
[7] B.B. Mandelbrot, J.R. Wallis, Water Resources Res. 5 (2) (1969) 321.
[8] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Phys. Rev. E 49 (1994) 1685.
[9] N. Vandewalle, M. Ausloos, Phys. Rev. E 58 (1998) 6832.
[10] K. Ivanova, M. Ausloos, Physica A 265 (1999) 279;
K. Ivanova, M. Ausloos, Eur. Phys. J. B 8 (1999) 665.
[11] G. Rangarajan, M. Ding, Phys. Rev. E 61 (2000) 004991.
[12] C. Heneghan, G. McDarby, Phys. Rev. E 62 (2000) 6103.
[13] E. Alessio, A. Carbone, G. Castelli, V. Frappietro, Eur. J. Phys. B 27 (2002) 197.
[14] A. Carbone, G. Castelli, H.E. Stanley, Phys. Rev. E 69 (2004) 0161;
A. Carbone, G. Castelli, H.E. Stanley, Physica A 344 (2004) 267.
[15] T. Di Matteo, T. Aste, M.M. Dacorogna, J. Banking Finance 29 (2005) 827.
[16] R. Voss, Random fractal forgeries, in: R.A. Earnshaw (Ed.), Fundamental Algorithms for Computer Graphics, NATO ASI Series F, vol. 17, Springer, New York, 1985, pp. 805–835.
[17] P. Abry, P. Flandrin, M. Taqqu, D. Veitch, Long Range Dependence: Theory and Applications, Birkhäuser, Boston, 2003;
M.S. Taqqu, V. Teverovsky, W. Willinger, Fractals 3 (1995) 785.
[18] G.M. Raymond, J.B. Bassingthwaighte, Physica A 265 (1999) 85.