Analysis of clusters formed by the moving average of a long-range correlated time series

A. Carbone,[1] G. Castelli,[1] and H. E. Stanley[2]
[1]*Dipartimento di Fisica and INFM, Politecnico di Torino, Corso Duca degli Abruzzi 24, I-10129 Torino, Italy*
[2]*Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA*
(Received 9 October 2003; published 19 February 2004)

We analyze the stochastic function $C_n(i) \equiv y(i) - \tilde{y}_n(i)$, where $y(i)$ is a long-range correlated time series of length $N_{\rm max}$ and $\tilde{y}_n(i) \equiv (1/n)\Sigma_{k=0}^{n-1} y(i-k)$ is the moving average with window n. We argue that $C_n(i)$ generates a stationary sequence of self-affine clusters $\mathcal{C}$ with length ℓ, lifetime τ, and area s. The *length* and the *area* are related to the *lifetime* by the relationships $\ell \sim \tau^{\psi_\ell}$ and $s \sim \tau^{\psi_s}$, where $\psi_\ell = 1$ and $\psi_s = 1 + H$. We also find that ℓ, τ, and s are power law distributed with exponents depending on H: $P(\ell) \sim \ell^{-\alpha}$, $P(\tau) \sim \tau^{-\beta}$, and $P(s) \sim s^{-\gamma}$, with $\alpha = \beta = 2 - H$ and $\gamma = 2/(1+H)$. These predictions are tested by extensive simulations on series generated by the midpoint displacement algorithm of assigned Hurst exponent H (ranging from 0.05 to 0.95) of length up to $N_{\rm max} = 2^{21}$ and n up to 2^{13}.

DOI: 10.1103/PhysRevE.69.026105 PACS number(s): 89.75.Da, 89.75.Fb, 05.10.−a, 05.40.−a

Long-range correlated time series, such as fractional Brownian motion, have been widely used for the theoretical description of diverse phenomena. The variance at large t scales as a power law:

$$\sigma^2 \sim t^{2H}. \tag{1}$$

Here H, the Hurst exponent, ranges from 0 to 1, with $H = 0.5$ corresponding to ordinary uncorrelated Brownian motion. H is related to the fractal dimension D by $D = 2 - H$. The Hurst exponent has been successfully exploited for practical purposes in fields as different as biophysics, econophysics, and climate physics [1–9]. For example, heartbeat intervals of healthy and sick hearts can be distinguished on the basis of the value of H [3–5,9]. The stock price volatility shows a degree of persistence ($0.7 < H < 0.8$) larger than that of the price series ($H \sim 0.5$) [6]. The validation of climate models is based on the analysis of a long-term correlation of atmospheric series [7].

A number of approaches are currently used to obtain accurate estimates of H. Such procedures generally consist of calculating appropriate statistical functions from the entire signal. Each procedure produces a slightly different estimate, so in order to obtain the most reliable estimates of H it is useful to apply as many approaches as possible, preferably combining techniques working in the spectral and time domains [10]. Here we propose an approach motivated by detrended moving average analysis, which was recently developed [11,12] as an alternative to the detrended fluctuation analysis technique [14]. One begins by defining the function

$$\sigma_{\rm MA} \equiv \sqrt{\frac{1}{N_{\rm max} - n} \sum_{i=n}^{N_{\rm max}} C_n(i)^2}, \tag{2}$$

where $N_{\rm max}$ is the length of the series,

$$C_n(i) \equiv y(i) - \tilde{y}_n(i), \tag{3}$$

and

$$\tilde{y}_n(i) \equiv \frac{1}{n} \sum_{k=0}^{n-1} y(i-k) \tag{4}$$

is the moving average of window size n, i.e., the average of the signal over n points. It is a linear operator, whose output are the low-frequency components of the signal, which are selected on the basis of the window amplitude n [13]. The function $\sigma_{\rm MA}$ shows a power-law dependence on n, i.e., $\sigma_{\rm MA} \sim n^H$ [11,12].

We explore the properties of the function $C_n(i)$ which generates, for each $\tilde{y}_n(i)$, a sequence of *clusters* $\mathcal{C}$, each corresponding to the region delimited by two consecutive intersections between $y(i)$ and $\tilde{y}_n(i)$ (see Fig. 1). Three quantities can be defined:

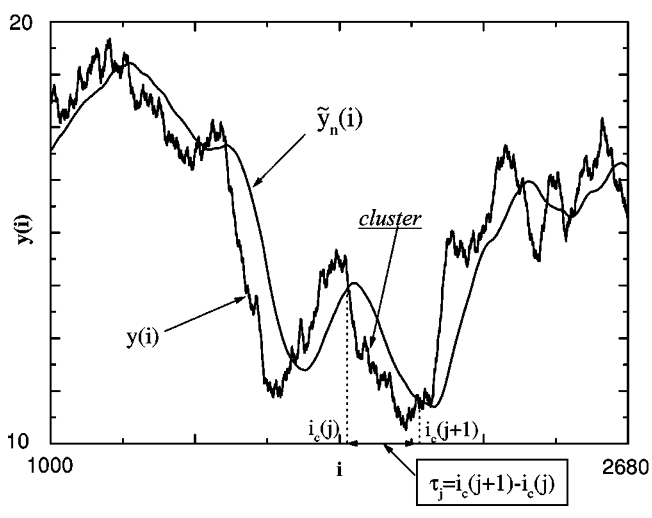

FIG. 1. Stochastic series $y(i)$ of length $N_{\rm max} = 2^{19}$ obtained by the random midpoint displacement algorithm with $H = 0.8$. Also shown is the moving average $\tilde{y}_n(i)$, with box dimension $n = 30$. The time interval between two subsequent crossing points $y(i)$ and $\tilde{y}_n(i)$ define the length ℓ_j, the duration τ_j, and the area s_j of the cluster according to Eqs. (5), (6), and (7).

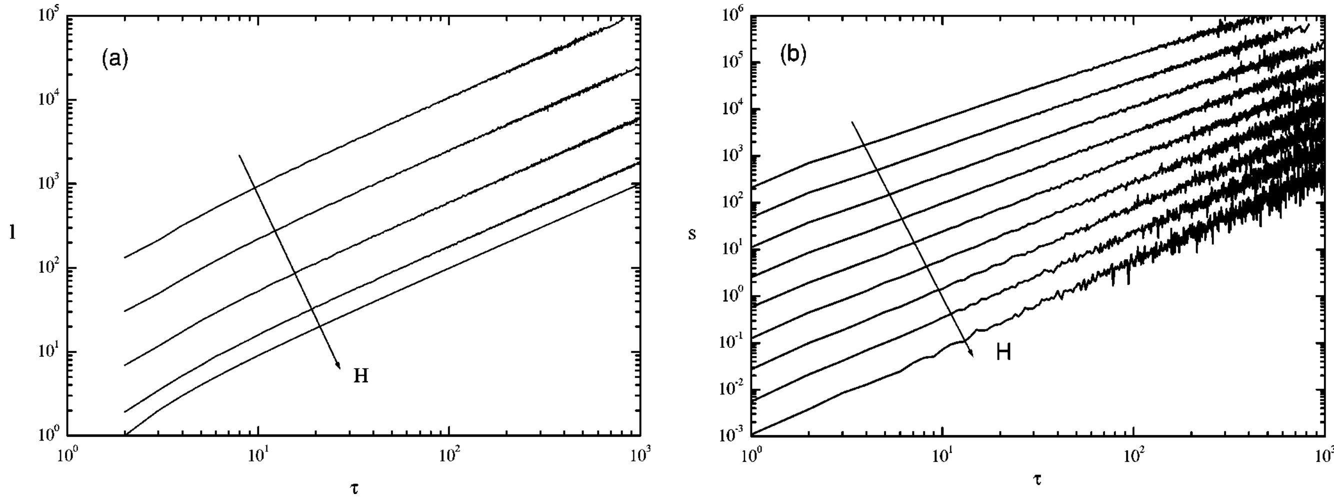

FIG. 2. (a) Log-log plot of the cluster length ℓ vs the cluster lifetime τ for series having different H, $H=0.2$, 0.3, 0.4, 0.5, and 0.8. (b) Log-log plot of the cluster area s vs the cluster lifetime τ for H varying between $H=0.1$ and 0.9 in steps of size 0.1.

(1) *cluster length* ℓ_j,

$$\ell_j \equiv \sum_{i=i_c(j)}^{i_c(j+1)} y(i), \tag{5}$$

(2) *cluster lifetime* τ_j,

$$\tau_j \equiv i_c(j+1) - i_c(j), \tag{6}$$

and (3) *cluster area* s_j,

$$s_j \equiv \sum_{i=i_c(j)}^{i_c(j+1)} |y(i) - \tilde{y}_n(i)| \Delta i, \tag{7}$$

where the index j refers to each cluster, $i_c(j)$ and $i_c(j+1)$ are the values of the index i corresponding to two subsequent intersections between $\tilde{y}_n(i)$ and $y(i)$ and Δi is the time interval corresponding to an elementary fluctuation in the time series. Finally, let ℓ and s indicate the value of the length and of the area obtained by averaging ℓ_j and s_j over the subset of clusters $\mathcal{C}$ having the same value of lifetime τ. Figures 2(a) and 2(b) show log-log plots of the cluster length ℓ and the cluster area s plotted against the cluster lifetime τ for long-range correlated time series constructed with the random midpoint displacement technique and with different values of H. The log-log plots are consistent with linearity over more than two decades, i.e., with the power law relationships

$$\ell \sim \tau^{\psi_\ell} \quad (\psi_\ell = 1), \tag{8}$$

and [15]

$$s \sim \tau^{\psi_s} \quad (\psi_s = 1 + H). \tag{9}$$

The values of ψ_ℓ and ψ_s are plotted as functions of H respectively in Figs. 3(a) and 3(b), and compared with the theoretical predictions [15].

Next we calculate the probability density function (PDF) of the cluster lifetime τ [see Fig. 4(a)] of the cluster length ℓ

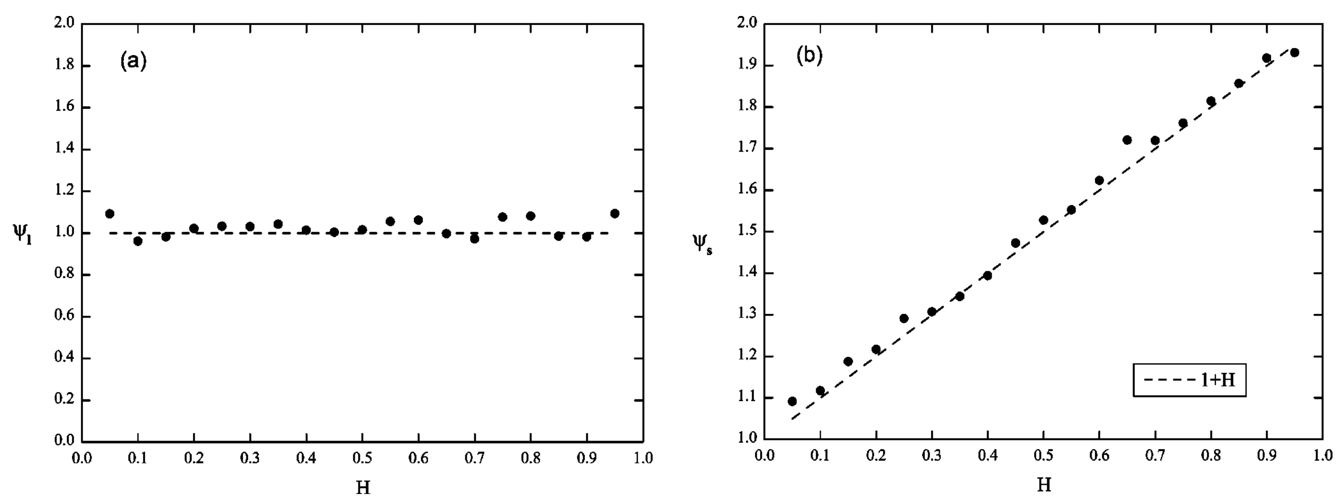

FIG. 3. Plot of the exponents (a) ψ_ℓ vs H [Eq. (8)] and (b) ψ_s vs H [Eq. (9)] for series having different H [from $H=0.05$ to 0.95 in steps of size 0.05 (circles)]. The relationships $\psi_l = 1$ and $\psi_s = 1+H$ are shown (dashed lines).

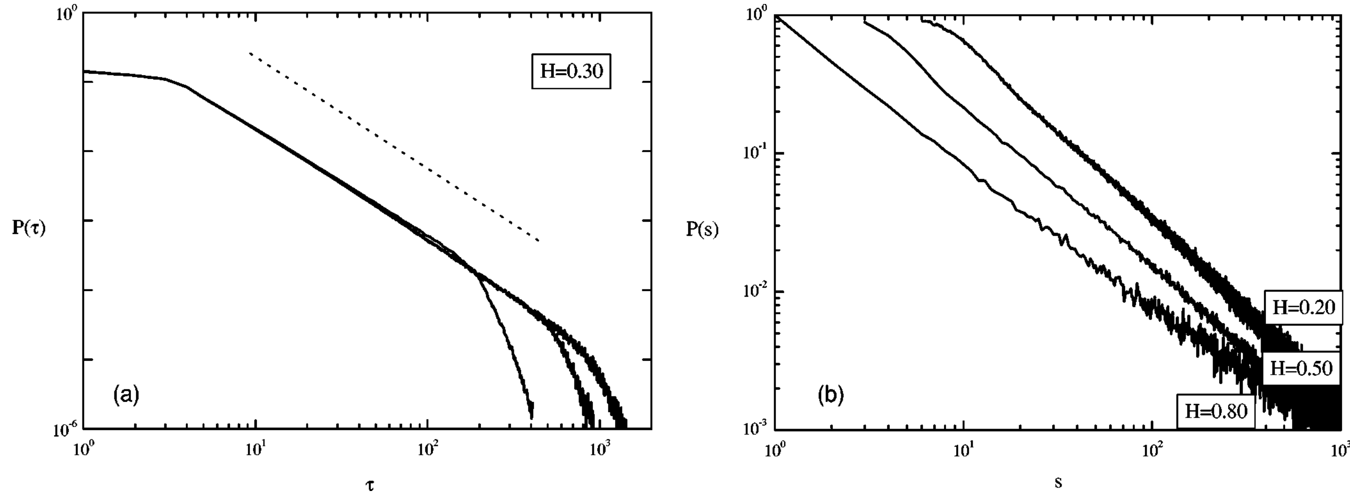

FIG. 4. (a) The PDF $P(\tau)$ of the cluster lifetime τ for a time series with $H=0.3$; the results are consistent with a power-law dependence $P(\tau)\sim\tau^{-\beta}$. The curves, from left to right, are obtained for window sizes $n=200$, 600, and 1000. The onset of the finite-size effect is visible when τ is approximately equal to the moving average window n. (b) The PDF $P(s)$ of cluster area s for $n=1000$, consistent with a power-law dependence $P(s)\sim s^{-\gamma}$, and three different values of H, $H=0.3$, 0.5, and 0.8.

and of the cluster areas [see Fig. 4(b)]. The results are consistent with a power-law behavior:

$$P(\tau)\sim\tau^{-\beta}. \tag{10}$$

$P(\tau)$ is the first return probability distribution [16–18] of the crossing points between $\tilde{y}_n(i)$ and $y(i)$, with exponent β:

$$\beta=2-H. \tag{11}$$

Equations (8) and (9) allow us to relate the probability density functions $P(\ell)$ and $P(s)$ to $P(\tau)$:

$$P(\ell)=P(\tau(\ell))\frac{d\tau}{d\ell}\sim\ell^{-\alpha}, \tag{12}$$

$$P(s)=P(\tau(s))\frac{d\tau}{ds}\sim s^{-\gamma}. \tag{13}$$

By using Eqs. (8) and (12), the exponent α can be written in terms of the exponent β as

$$\alpha=\beta=2-H. \tag{14}$$

Analogously, using Eqs. (9) and (13), γ can be expressed in terms of β and ψ_s as

$$\gamma=\frac{\beta+1-\psi_s}{\psi_s}=\frac{2}{1+H}. \tag{15}$$

To test the predictions of Eqs. (14) and (15), we have calculated the exponents α, β, and γ for a wide range of parameters: $N_{\max}$ ranges from 2^{14} to 2^{21} while n ranges from 2^3 up to 2^{13}. The exponents β and γ are plotted against H in Figs. 5(a) and 5(b), and compared with thepredictions of Eqs. (14) and (15).

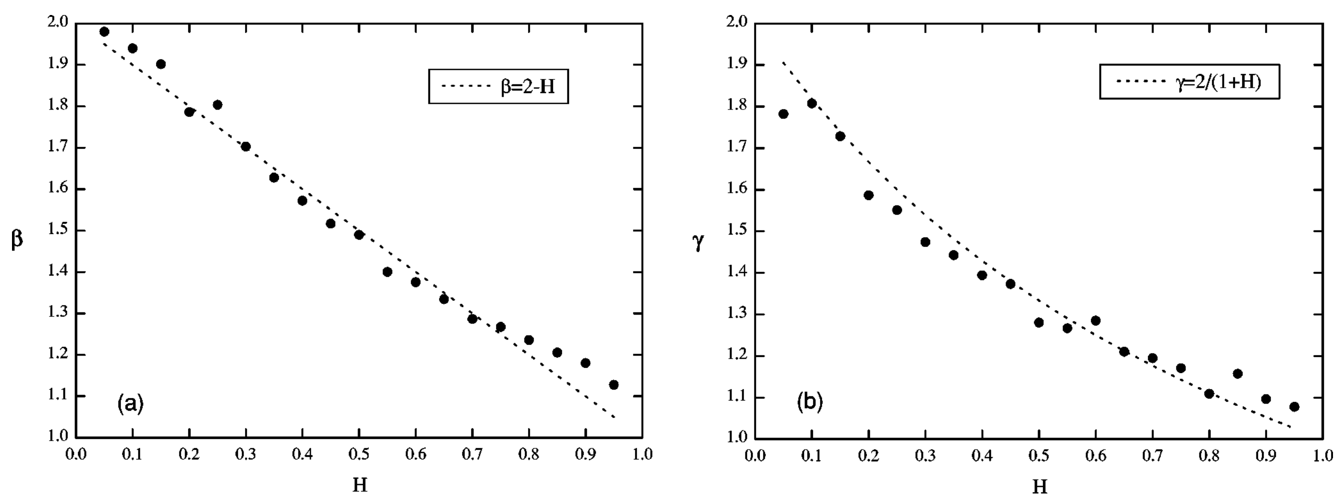

FIG. 5. (a) The exponent β vs H [Eq. (10)] for series having different H ($H=0.05$–0.95 with step size 0.05). The relationship $\beta=2-H$ is also shown (dashed line). (b) Plot of the exponent γ vs H [Eq. (15)] for series having different H ($H=0.05$–0.95 with step 0.05). The relationship $\gamma=2/(1+H)$ is also shown (dashed line).

In summary, the statistical properties of the sequence of stationary self-affine clusters $\mathcal{C}$ generated by the intersections of the time series $y(i)$ with the moving average $\tilde{y}_n(i)$ have been analyzed. For model series of length up to $N_{\max}=10^{21}$ we calculate the area $s\sim\tau^{\psi_s}$ and the PDFs $P(\ell)\sim\ell^{-\alpha}$, $P(\tau)\sim\ell^{-\beta}$, and $P(s)\sim s^{-\gamma}$. Our results are consistent with power laws whose exponents agree with the predictions $\psi_s=1+H$, $\alpha=\beta=2-H$, and $\gamma=2/(1+H)$ for a wide range of H $(0.05<H<0.95)$. It is noteworthy that the scaling properties of the $\mathcal{C}$ clusters are reminiscent of the self-organized criticality (SOC) model, proposed by Bak, Tang, and Wiesenfeld [19]. This similarity can be derived from the relation between the growth dynamics of the $\mathcal{C}$ clusters and of the steady-state SOC clusters. An in-depth discussion of such issue is however beyond the scope of the present work and will be developed elsewhere.

We thank L. A. N. Amaral, S. Havlin, and P. Ch. Ivanov for helpful discussions and NIH/National Center for Research Resources (P41RR13622) for support.

[1] D. Ben-Avraham and S. Havlin, *Diffusion and Reactions in Fractals and Disordered Systems* (Cambridge University Press, Cambridge, 2000).

[2] M.S. Taqqu, V. Teverosky, and W. Willinger, Fractals **3**, 785 (1995).

[3] P.Ch. Ivanov, L.A.N. Amaral, A.L. Goldberger, S. Havlin, M.G. Rosenblum, R. Zbigniew, and H.E. Stanley, Nature (London) **410**, 242 (1999).

[4] Z. Chen, P.Ch. Ivanov, K. Hu, and H.E. Stanley, Phys. Rev. E **65**, 041107 (2002).

[5] K. Hu, P.Ch. Ivanov, Z. Chen, P. Carpena, and H.E. Stanley, Phys. Rev. E **64**, 011114 (2001).

[6] Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.-K. Peng, and H.E. Stanley, Phys. Rev. E **60**, 1390 (1999).

[7] *The Science of Disasters: Climate Disruptions, Heart Attacks and Market Crashes*, edited by A. Bunde, H. J. Schellnhuber, and J. Kropp (Springer, Berlin, 2002).

[8] N. Vandewalle and M. Ausloos, Phys. Rev. E **58**, 6832 (1998).

[9] Y. Ashkenazy, M. Lewkowicz, J. Levitan, S. Havlin, K. Saermark, H. Moelgaard, and P.E.B. Thomsen, Fractals **7**, 85 (1999).

[10] G. Rangarajan and M. Ding, Phys. Rev. E **61**, 04991 (2000).

[11] E. Alessio, A. Carbone, G. Castelli, and V. Frappietro, Eur. Phys. J. B **27**, 197 (2002).

[12] A. Carbone and G. Castelli, Proc. SPIE **5114**, 406 (2003).

[13] The moving average filter *dynamically* detrends the series: as the discrete index i increases by unity, the box window n switches its position by the same amount.

[14] C.K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, and A.L. Goldberger, Phys. Rev. E **49**, 1685 (1994); C.K. Peng, S. Ha, H.E. Stanley, and A.L. Goldberger, Chaos **5**, 82 (1995).

[15] Equation (8) follows if the constitutive relation of the fractional Brownian motion $\Delta y(i)\sim\Delta i^H$ is taken into account to calculate the length of the $y(i)$ segments limited by $i_c(j)$ and $i_c(j+1)$. Equation (9) follows if the relationship

$$\tilde{y}_n(i)-y(i)\approx D_H\frac{\Delta i}{\Delta n^2}\nabla^2 y(i) \qquad (16)$$

is taken into account, where D_H is the generalized diffusion coefficient for the fractional brownian motion and the other quantities have the usual meaning. The term on the right side of Eq. (16) is proportional to the average displacement of the random walker during a time interval Δi and thus varies as Δi^H. Using Eq. (16) to calculate the sum of terms $\tilde{y}(i)-y_n(i)$ over the time interval $\tau_j\equiv i_c(j+1)-i_c(j)$ [Eq. (7)], the behavior of s as τ^{H+1} is justified. It is worth mentioning that Eq. (16) has been worked out in J.R. Weimar and J.-P. Boon, Phys. Rev. E **49**, 1749 (1994).

[16] Equation (11) can be derived from the relation $\langle\tau\rangle\equiv N_{\max}/N_\times$ between the mean time interval $\langle\tau\rangle$ and the total number of crossing points $N_\times$. To express $\langle\tau\rangle$ in terms of $N_{\max}$, we write $\langle\tau\rangle\equiv\int_1^{N_{\max}}\tau P(\tau)d\tau\sim N_{\max}^{2-\beta}$. Similarly, we can express $N_\times$ in terms of $N_{\max}$ as $N_\times\sim N_{\max}^{1-H}$, which follows from the fact that the fractal dimension of the set of crossing points is $1-H$ (the codimension of the set of crossing points $d-d_\times$ is equal to the sum of the codimension $2-H$ of the signal and the codimension $2-1$ of the moving average). Thus $2-\beta=1-(1-H)$, from which Eq. (11) follows. A similar reasoning gives the value of the exponent $\alpha=2-H$ for the cluster length probability distribution function [Eqs. (12) and (14)].

[17] M. Ding and W. Yang, Phys. Rev. E **52**, 207 (1995).

[18] S. Maslov, M. Paczuski, and P. Bak, Phys. Rev. Lett. **73**, 2162 (1994).

[19] P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987).