Assessment of long-range correlation in time series: How to avoid pitfalls

Jianbo Gao,[1,*] Jing Hu,[1] Wen-Wen Tung,[2] Yinhe Cao,[3] N. Sarshar,[4] and Vwani P. Roychowdhury[4]
[1]*Department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida 32611, USA*
[2]*Department of Earth & Atmospheric Sciences, Purdue University, West Lafayette, Indiana 47907, USA*
[3]*BioSieve, 1026 Springfield Drive, Campbell, California 95008, USA*
[4]*Electrical Engineering Department, UCLA, Los Angeles, California 90095, USA*
(Received 24 August 2005; published 13 January 2006)

Due to the ubiquity of time series with long-range correlation in many areas of science and engineering, analysis and modeling of such data is an important problem. While the field seems to be mature, three major issues have not been satisfactorily resolved. (i) Many methods have been proposed to assess long-range correlation in time series. Under what circumstances do they yield consistent results? (ii) The mathematical theory of long-range correlation concerns the behavior of the correlation of the time series for very large times. A measured time series is finite, however. How can we relate the fractal scaling break at a specific time scale to important parameters of the data? (iii) An important technique in assessing long-range correlation in a time series is to construct a random walk process from the data, under the assumption that the data are like a stationary noise process. Due to the difficulty in determining whether a time series is stationary or not, however, one cannot be 100% sure whether the data should be treated as a noise or a random walk process. Is there any penalty if the data are interpreted as a noise process while in fact they are a random walk process, and vice versa? In this paper, we seek to gain important insights into these issues by examining three model systems, the autoregressive process of order 1, on-off intermittency, and Lévy motions, and considering an important engineering problem, target detection within sea-clutter radar returns. We also provide a few rules of thumb to safeguard against misinterpretations of long-range correlation in a time series, and discuss relevance of this study to pattern recognition.

DOI: 10.1103/PhysRevE.73.016117 PACS number(s): 02.50.−r, 05.45.Tp, 05.40.Ca

I. INTRODUCTION

Of the types of activity that characterize complex systems, the most ubiquitous and puzzling is perhaps the appearance of $1/f^{\alpha}$ noise, a form of temporal or spatial fluctuation characterized by a power-law decaying power spectral density. Some of the classical literature on this subject can be found, for example, in Press [1], Bak [2], and Wornell [3]. Some of the more recently discovered $1/f^{\alpha}$ processes are in traffic engineering [4–6], DNA sequences [7–9], human cognition [10], coordination [11], posture [12], dynamic images [13,14], and the distribution of prime numbers [15]. The dimension of such processes usually cannot be reduced by principal component analysis, since the rank-ordered eigenvalue spectrum decays as a power law [16]. An important subclass of the $1/f^{\alpha}$ noise is those with long-range temporal correlation (or long memory).

Since data with long memory appear quite frequently in many different areas of science and engineering, many methods have been proposed to estimate the key scaling parameter, the Hurst parameter H. When $1/2<H<1$, the process is said to have persistent correlation; when $H=1/2$, the process is memoryless or only has short-range correlation; when $0<H<1/2$, the process is said to have antipersistent correlation [17]. While this field seems to be mature, there still exist many important issues unresolved. We consider three here. (i) Often, researchers assume that the methods for estimating H should yield consistent results when applied to model systems. In practice, however, they resort to the detrended fluctuation analysis (DFA) [18] more frequently than other methods. Under what circumstances can the above assumption and practice be justified? Note that this issue was recently partially examined by Rangarajan and Ding [19], by studying two methods, the spectral method and the rescaled range analysis. (ii) The mathematical theory of long-range correlation concerns the behavior of the correlation of the time series for very large time. A measured time series is finite, however. How can we relate the fractal scaling break at a specific time scale to important parameters of the data? (iii) An important technique in assessing long-range correlation in a time series is to construct a random walk process from the data, under the assumption that the data are like a stationary noise process. Due to the difficulty in determining whether a time series is stationary or not (see [20] and many references therein), however, one cannot be 100% sure whether the data should be treated as a noise or a random walk process. Is there any penalty if the data are interpreted as a noise process while in fact they form a random walk process, and vice versa?

In this paper, we seek to gain important insights into these issues by examining three model systems, the autoregressive (AR) process of order 1 [21], on-off intermittency, and Lévy motions. The first is usually denoted as AR(1). It is the simplest and most commonly used model for colored noise. We shall also consider an important engineering problem, target detection within sea-clutter radar returns.

*Electronic address: gao@ece.ufl.edu

 ©2006 The American Physical Society

II. THEORETICAL BACKGROUND

Let $X=\{X_t: t=0,1,2,\ldots\}$ be a covariance stationary stochastic process with mean μ, variance σ^2, and autocorrelation function $r(k)$, $k \geqslant 0$. The process is said to have long-range correlation [22] if $r(k)$ is of the form

$$r(k) \sim k^{2H-2} \quad \text{as } k \to \infty, \tag{1}$$

where $0<H<1$ is the Hurst parameter. Note that when $1/2<H<1$, $\Sigma_k r(k)=\infty$. This justifies the term "long-range correlation."

Next we construct a new covariance stationary time series $X^{(m)}=\{X_t^{(m)}: t=1,2,3,\ldots\}$, $m=1,2,3,\ldots$, obtained by averaging the original series X over nonoverlapping blocks of size m,

$$X_t^{(m)} = (X_{tm-m+1} + \cdots + X_{tm})/m, \quad t \geqslant 1. \tag{2}$$

Note that the length of $\{X_t^{(m)}\}$ is $[N/m]$, where N is the length of $\{X_t\}$, and [ ] denotes the greatest integer function.

There are several useful relationships between the autocorrelation functions of the original process and its averaged version [23]. Using the stationarity properties of the process, a general formula for the autocorrelation function $r^{(m)}(k)$ of $X^{(m)}$ can be stated as

$$r^{(m)}(k) = \frac{(k+1)^2 V_{(k+1)m} - 2k^2 V_{km} + (k-1)^2 V_{(k-1)m}}{2V_m}, \tag{3}$$

where $V_m=\mathrm{var}(X^{(m)})$. Using this relationship, it is straightforward to verify that the variance of $X^{(m)}$ satisfies

$$\mathrm{var}(X^{(m)}) = \sigma^2 m^{2H-2} \tag{4}$$

if and only if the autocorrelation function of the long-range-dependent (LRD) process satisfies

$$r(k) = \frac{1}{2}[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}]. \tag{5}$$

Moreover, one can verify that if X satisfies Eq. (4), then the autocorrelation function $r^{(m)}(k)$ of the process $X^{(m)}$ satisfies

$$r^{(m)}(k) = r(k), \quad k \geqslant 0. \tag{6}$$

A process X defined by Eq. (4) [or equivalently, Eqs. (1), (4), and (6)] is often referred to as an exactly second-order self-similar process. On the other hand, if one relaxes Eq. (4) as

$$\lim_{k\to\infty} \frac{r(k)}{k^{2H-2}} = c_1,$$

where $0<c_1$ is an arbitrary constant, then one can show that

$$\lim_{k\to\infty} \frac{\mathrm{var}(X^{(m)})}{m^{2H-2}} = c_2 \tag{7}$$

for some constant $c_2>0$. Such a process is often referred to as an asymptotically second-order self-similar process.

Equation (4) [or more generally, Eq. (7)] is often called the variance-time relation. It provides a simple and precise way of quantifying the "little smoothing" behavior. For example, when $H=0.5$, $\mathrm{var}(X^{(m)})$ drops to $10^{-2}\sigma_0^2$ when $m=100$, where σ_0^2 is the variance of the original process; when $H=0.75$, in order for $\mathrm{var}(X^{(m)})$ to drop as much, m has to be 10 000.

Note that the power spectral density (PSD) for X is

$$S_X(f) \sim f^{-(2H-1)}. \tag{8}$$

Therefore, X is called a $1/f^\alpha$ process. Its integration, called the random walk process (see below), has PSD $f^{-(2H+1)}$.

The prototypical model for the $1/f^\alpha$ process is the fractional Brownian motion (FBM) process $B_H(t)$, where H is the Hurst parameter [17]. It is a Gaussian process with mean 0, stationary increments, variance

$$E[(B_H(t))^2] = t^{2H}, \tag{9}$$

and covariance

$$E[B_H(s)B_H(t)] = \frac{1}{2}(s^{2H} + t^{2H} - |s-t|^{2H}). \tag{10}$$

The increment process of the FBM, $X_i=B_H[(i+1)\Delta t] - B_H(i\Delta t)$, $i \geqslant 1$, where Δt can be considered a sampling time, is called fractional Gaussian noise. It is a zero-mean stationary Gaussian time series, with autocorrelation function

$$\gamma(k) = E(X_i X_{i+k})/E(X_i^2) = \frac{1}{2}[(k+1)^{2H} - 2k^{2H} + |k-1|^{2H}], \quad k \geqslant 0. \tag{11}$$

Since $\gamma(k)$ is independent of Δt, without loss of generality, we can take $\Delta t=1$. In particular, we have $\gamma(1) = 1/2(2^{2H}-2)$. The notions of persistent and antipersistent correlations come from the fact that $\gamma(1)$ is positive when $1/2<H<1$, but negative when $0<H<1/2$.

We now consider estimation of H. A convenient framework is based on the random walk process y, defined as

$$y_k = \sum_{i=1}^{k} (X_i - \bar{X}), \tag{12}$$

where $\bar{X}$ is the mean of X. We then examine whether the following scaling law holds or not:

$$F(m) = \langle |y(i+m) - y(i)|^2 \rangle^{1/2} \sim m^H, \tag{13}$$

where the average is taken over all possible pairs of $(y(i+m), y(i))$. This method is often called fluctuation analysis (FA). As will be explained in the Appendix, FA is a special case of $q=2$ in the structure-function-based multifractal formalism. For this reason, H may also be denoted as $H(2)$. In the Appendix, we shall also explain how one can readily prove that in the light of multifractal formalism, many methods for estimating H are equivalent to FA.

Next we explain DFA [18]. It works as follows. First divide a given random walk of length N into $\lfloor N/m \rfloor$ nonoverlapping segments (where the notation $\lfloor x \rfloor$ denotes the largest integer that is not greater than x); then define the local trend in each segment to be the ordinate of a linear least-squares fit for the random walk in that segment; finally compute the "detrended walk," denoted by $y_m(n)$, as the difference between the original walk $y(n)$ and the local trend. Then one examines

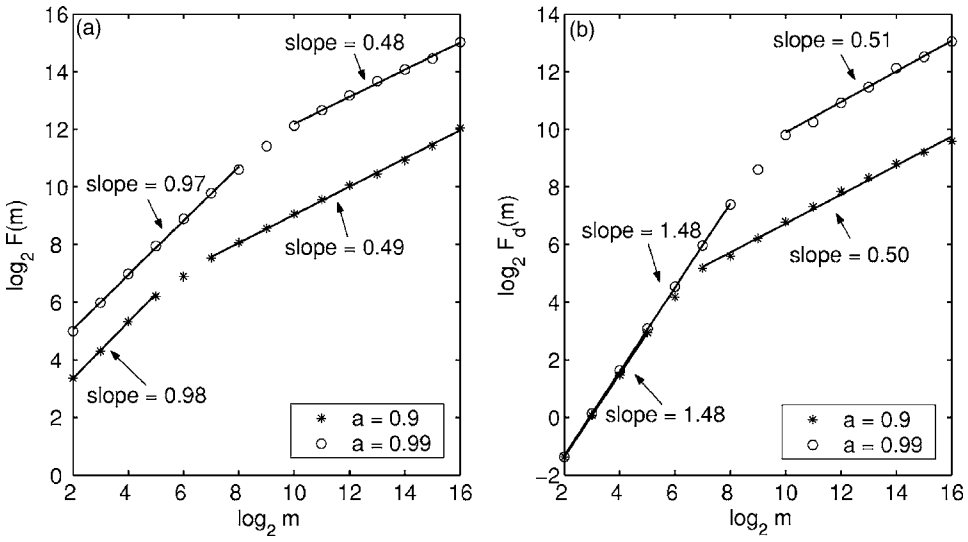

FIG. 1. H parameter for AR(1) model. (a) is for FA and (b) for DFA. The Hurst parameters are obtained as the slopes of the lines.

$$F_d(m) = \left\langle \sum_{i=1}^{m} y_m(i)^2 \right\rangle^{1/2} \sim m^H \tag{14}$$

where the angular brackets denote the ensemble average of all the segments and $F_d(m)$ is the average variance over all segments.

Finally, we discuss the wavelet-based H estimator [24]. The method directly works on the original data instead of the random walk process. Denote a scaling function by ϕ_0 and a mother wavelet by Ψ_0. Let the maximum scale level be J. A discrete wavelet transform is a mapping from $x(t)$ to the wavelet coefficients (a_x, d_x):

$$x(t) \to \{a_x(J,k),\ d_x(j,k),\ j \in [1,J],\ k \in Z\},$$

where the $a_x(J,k)$'s are called approximation coefficients and the $d_x(j,k)$'s detailed coefficients, defined by

$$a_x(J,k) = \langle x, \phi_{J,k} \rangle, \quad d_x(j,k) = \langle x, \psi_{j,k} \rangle,$$

$$\phi_{j,k} = 2^{-j/2}\phi_0(2^{-j/2} - k), \quad \Psi_{j,k} = 2^{-j/2}\Psi_0(2^{-j/2} - k), \quad k \in Z,$$

$$x(t) = \sum_k a_x(J,k)\phi_{J,k}(t) + \sum_{j=1}^{J} \sum_k d_x(j,k)\Psi_{j,k}(t).$$

Let

$$\Gamma_j = \frac{1}{n_j} \sum_{k=1}^{n_j} |d_x(j,k)|^2, \tag{15}$$

where n_j is the number of coefficients at level j; then the Hurst parameter is given by

$$\log_2 \Gamma_j = (2H - 1)j + c_0, \tag{16}$$

where c_0 is some constant.

III. ANALYSIS OF MODEL SYSTEMS

To better appreciate the notion of finite fractal scaling and consistency of H estimators, we consider three model systems, the AR(1) model, on-off intermittency, and Levy motions.

A. AR(1) model

We first consider the AR(1) model $z_{n+1} - \bar{z} = a(z_n - \bar{z}) + \eta_n$, where $\bar{z}$ is the mean of z_n, η_n is a white Gaussian noise of mean 0 and variance σ_η^2, and a is a coefficient satisfying the condition $|a| < 1$. We may rewrite $x_n = z_n - \bar{z}$ and have

$$x_{n+1} = ax_n + \eta_n. \tag{17}$$

It is well known that the autocorrelation for $\{x_n\}$ decays exponentially, $C(m) = [\sigma_\eta^2/(1-a^2)]a^{|m|}$. When the time lag m is large, the correlation is essentially zero; we can expect H to be 1/2. However, when the coefficient a is only slightly smaller than 1, $C(m)$ will be close to $\sigma_\eta^2/(1-a^2)$ for a considerable range of m. In this case, we have almost perfect correlation. One thus might expect $H \approx 1$ for a not too large time scale. This seems to be verified when one applies the variance-time relation to analyze the generated time series, or equivalently applies FA to the random walk process constructed from the data. The latter is shown in Fig. 1(a), where we observe that H (as the slopes of the lines) is close to 1 when m is not too large, and close to 0.5 for large m. However, there is a problem here—if we employ DFA, then we obtain $H \approx 1.5$ for a not too large time scale, as shown in Fig. 1(b). How shall we understand this difference?

To find an answer, let us examine which one is consistent with the PSD of the AR(1) process. Using the Wiener-Khintchine theorem, we can readily find the PSD $S_x(\omega)$ of the AR(1) process by taking the Fourier transform of the autocorrelation function,

$$S_x(\omega) = \sum_{m=-\infty}^{\infty} \frac{\sigma_\eta^2}{1-a^2} a^{|m|} e^{-j\omega m} = \frac{\sigma_\eta^2}{1 + a^2 - 2a\cos\omega},$$

$$0 \leqslant \omega \leqslant \pi, \tag{18}$$

where $j^2 = -1$. Alternatively, we may take the Fourier transform of both sides of Eq. (17) to obtain

$$X(\omega) = ae^{-j\omega}X(\omega) + \sigma_\eta,$$

where $X(\omega)$ denotes the Fourier transform of the left side of Eq. (17), and the coefficient $ae^{-j\omega}$ is due to delay by one unit of time. Then,

$$S_x(\omega) = |X(\omega)|^2.$$

Let us now simplify Eq. (18). Expanding $\cos\omega = 1-\omega^2/2+\cdots$ and noticing $\omega=2\pi f$, we have

$$S_x(f) \approx \frac{\sigma_\eta^2}{(1-a)^2 + a(2\pi)^2 f^2}, \quad 0 \leqslant f \leqslant 1/2.$$

At the low-frequency end, the term $a(2\pi)^2 f^2$ can be dropped, and we have a flat spectrum, consistent with $H=1/2$. At the high-frequency ($f \to 1/2$) end, since a is close to 1, the term $(1-a)^2$ can be dropped, and we have

$$S_x(f) \propto f^{-2}.$$

The transition frequency f_* is found by equating the two terms,

$$(1-a)^2 = a(2\pi)^2 f_*^2,$$

from which, we get

$$T^* = 1/f_* = \frac{2\pi\sqrt{a}}{1-a}.$$

To more precisely find the frequency ranges where the PSD is flat or decays as f^{-2}, we may require $(1-a)^2 \geqslant a(2\pi)^2 f^2$ when $f \leqslant f_1$, and $(1-a)^2 \leqslant a(2\pi)^2 f^2$ when $f_2 \leqslant f \leqslant 1/2$. Quantitatively, $f_{1,2}$ may be defined by the following conditions:

$$(1-a)^2 = \theta a(2\pi)^2 f_1^2$$

and

$$\theta(1-a)^2 = a(2\pi)^2 f_2^2,$$

where the parameter θ is on the order of 10. The two time scales defined by $f_{1,2}$ are $T_{1,2}=1/f_{1,2}$, with

$$T_1 = T^*/\sqrt{\theta}, \quad T_2 = T^*\sqrt{\theta}.$$

Let us now examine Fig. 1 again. From either FA or DFA plots, we indeed observe that around T^*, the scaling changes from a large H (1 for FA and 1.5 for DFA) to $H=1/2$. More precisely, when $m < T_1$, H is close to 1 for FA and 1.5 for DFA; when $m > T_2$, H is close to 0.5 for both FA and DFA. For a $1/f^\alpha$ noise, $\alpha=2H-1$. Now that $\alpha=2$, we have to conclude $H=1.5$ for $m < T_1$. Therefore, DFA is consistent with the spectrum, but FA is not.

At this point, it is important to note that the AR(1) model with coefficient a very close to 1 has been proposed as a (pseudo)model for LRD traffic with $H=1$ [25], and a convenient model for exact $1/f$ noise [26]. The former misinterpretation is indeed due to misuse of FA (or the variance-time relation) with the data. One cause of the latter misinterpretation may be the following: the magnitude response of the Fourier transform of the process $|X(\omega)|$ scales with f as f^{-1} when $f \to 1/2$. When $|X(\omega)|$ is mistaken for the PSD, then one would claim that the model generates an exact $1/f$ spectrum. While this error might not be relevant in [26],[1] it is a major cause of controversy around the development of self-organized criticality [27–32].

The discrepancy between FA and DFA is due to the fact that the Hurst parameter estimated by FA and related methods saturates at 1. See the Appendix for a proof.

B. On-off intermittency

On/off intermittency is a ubiquitous and important phenomenon. For example, a queueing system or a network can alternate between idle and busy periods; a fluid flow can switch from a turbulent motion to a regular (called laminar) one. Let us denote an on period by 1 and an off period by 0. We study three types of on-off trains where on and off periods are independent and both have the same (i) exponential distribution, (ii) Pareto distribution, and (iii) truncated Pareto distribution. The Pareto distribution is defined as

$$P[X \geqslant x] = \left(\frac{b}{x}\right)^\alpha, \quad x \geqslant b > 0, \quad \alpha > 0 \tag{19}$$

where α and b are called the shape and the location parameters, respectively. For Pareto distributions, we choose two α: 1.6 and 0.6. Truncation is achieved by simply requiring $x \leqslant L$, where L is a parameter. When $1 \leqslant \alpha \leqslant 2$, it can be proven [33] that

$$H = (3-\alpha)/2. \tag{20}$$

One of our purposes here is to check whether Eq. (20) can be numerically verified. For this purpose, we apply FA and DFA to the integrated data of an on-off train. The on-off train is sampled in such a way that in a total of about 1000 on-off periods, on average a few tens of points of an on or off period are sampled. The results for FA and DFA are shown in Figs. 2(a)–2(d), respectively. We observe that for all these three cases, for small time scale (determined by the average length of an on or off period), H (as the slopes of the lines) are close to 1 by FA and 1.5 by DFA. By simple analytical analysis or numerical simulation, one can readily find that for high frequency, the PSD for an on-off train scales with the frequency as f^{-2}, just as at the high-frequency end of an AR(1) model. Therefore, for time scales not longer than the average on or off period, DFA is consistent with the PSD, but FA is not. For larger scales, for case (i), we observe H from both FA and DFA is 0.5 [with regard to Eq. (20), this amounts to taking $\alpha=2$]; while for cases (ii) and (iii), we observe that Eq. (20) is correct with FA, when $1 \leqslant \alpha \leqslant 2$, and correct with DFA for the entire range of admissible α: $0 \leqslant \alpha \leqslant 2$. When $0 \leqslant \alpha < 1$, due to saturation, FA always gives $H=1$. When the power-law distribution is truncated, H eventually becomes 1/2, by both FA and DFA.

[1] The transition frequencies reported in [26] depend on the variance of the process. This is incorrect, since in a log-log plot of the PSD, the variance can only cause the PSD curve to shift upward or downward.

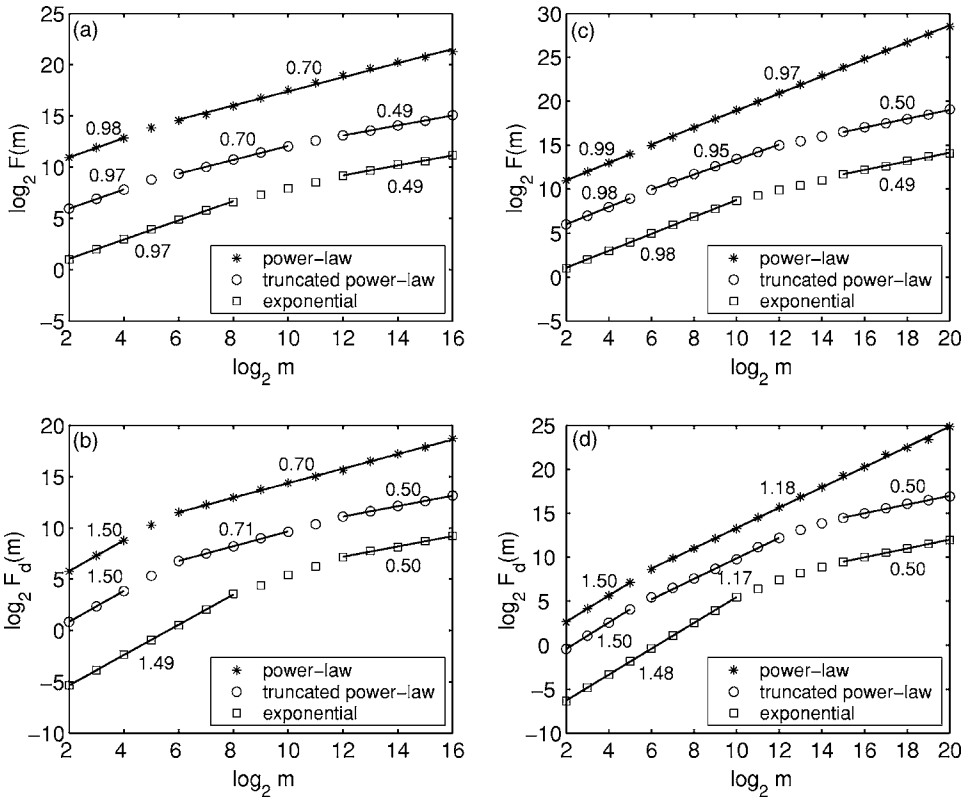

FIG. 2. H parameter for on-off model. The α parameter is 1.6 for (a),(b) and 0.6 for (c),(d). (a),(c) are for FA, and (b),(d) for DFA. The Hurst parameters are obtained as the slopes of the lines.

This case study again shows that FA and DFA are consistent when $H<1$. Otherwise, the results by FA cannot be trusted.

C. Lévy motions

A Lévy motion is a stochastic process defined through stable laws [34]. A stable law can be conveniently defined through a characteristic function. For our purpose, it suffices to note that the tail of a stable law is a power law $P[X \geqslant x] \sim x^{-\alpha}$ when $x \to \infty$. There are two types of Lévy motions. One is Lévy flights, which are random processes consisting of many independent steps, each step being characterized by a stable law, and consuming a unit time regardless of its length. The other is Lévy walkers, where each step takes a time proportional to its length. A Lévy walker can be viewed as sampled from a Lévy flight with a uniform speed.

Intuitively, we expect Lévy flights to be memoryless, simply characterized by $H=1/2$, irrespective of the value of the exponent α characterizing the stable laws. This is indeed the case, as is shown in Fig. 3. The correlation structure of a Lévy walker, however, is more complicated. We observe from Fig. 4 that when the scale is small, corresponding to "walking" along a single step of a Lévy flight, H is close to 1 by FA, and close to 1.5 by DFA. Analysis by Fourier transform shows that the PSD at the high-frequency end again decays as f^{-2}. Therefore, for time scales not longer than the average on or off period, DFA is consistent with the PSD, but FA is not. On larger scales, corresponding to constantly "switching" from one step of a Lévy flight to another, H is given by Eq. (20) for FA when $1 \leqslant \alpha \leqslant 2$, and for DFA when $0 \leqslant \alpha \leqslant 2$. Again due to saturation, FA always yields $H=1$ when $0 \leqslant \alpha < 1$. While these observations are similar to those found for the on-off trains discussed above, we note a difference between Figs. 2 and 4. That is, for a Lévy walker, the transition from a larger H at small scale to a smaller H at large scale is more gradual. This difference is due to the difference between a stable law and a Pareto distribution.

IV. APPLICATION: TARGET DETECTION WITHIN SEA-CLUTTER RADAR RETURNS

Sea clutter is the backscattered returns from a patch of the sea surface illuminated by a radar pulse. Robust detection of

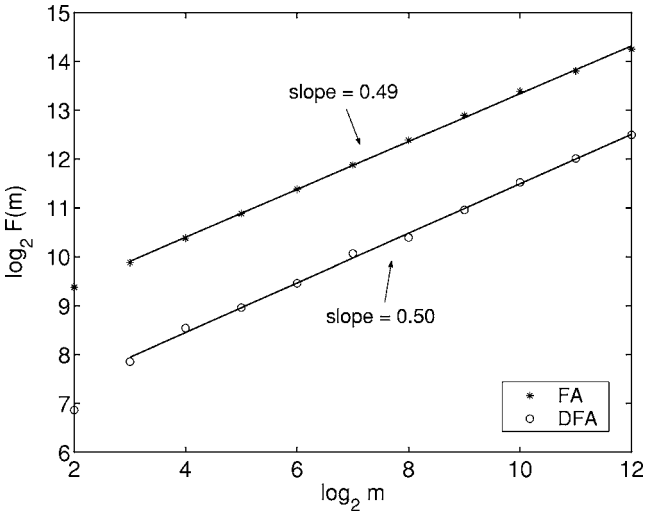

FIG. 3. H parameter for Lévy flights. H, as the slopes of the lines, are independent of the parameter α.

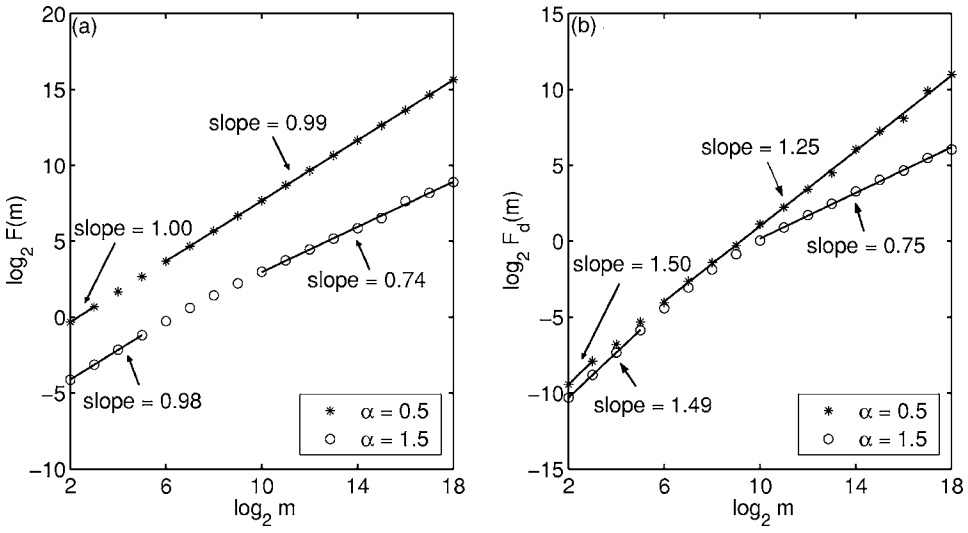

FIG. 4. H parameter for Lévy walks. (a) is for FA and (b) for DFA. The Hurst parameters are obtained as the slopes of the lines.

targets from sea-clutter radar returns is an important problem in remote sensing and radar signal processing applications. This is a difficult problem, because of the turbulent wave motions on the sea surface as well as multipath propagation of radar pulses massively reflected from the sea surface. In the past several decades, tremendous effort has been made to understand the nature of sea clutter as well as to detect targets within sea clutter [35–43]. However, novel, simple, and reliable methods for target detection are yet to be developed. In this section, we show that the H parameter together with the notion of the finite fractal scaling range offer a very simple and effective method to detect low-observable targets within sea clutter. This case study also vividly illustrates what kind of penalty may result if one misinterprets a noise process as a random walk process and vice versa.

First we briefly describe the data. Fourteen sea-clutter data sets were obtained from a website maintained by Haykin [44,45]. The measurement was made using the McMaster IPIX radar at the east coast of Canada, from a clifftop near Dartmouth, Nova Scotia. The operating (or carrier) frequency of the radar is 9.39 GHz (and hence a wavelength of about 3 cm). Data of two polarizations, HH (horizontal transmission, horizontal reception) and VV (vertical transmission, vertical reception), were analyzed here. The grazing angle varied from less than 1° to a few degrees. The wave height in the ocean varied from 0.8 to 3.8 m (with peak height up to 5.5 m). The wind conditions varied from still to 60 km/h (with gusts up to 90 km/h). Each data set contains 14 spatial range bins of HH as well as 14 range bins of VV data sets. Therefore, there are a total of 392 sea-clutter time series. A few of the range bins hit a target, which was made of a spherical block of styrofoam of diameter 1 m, wrapped with wire mesh. This is a very small target, more difficult to detect than, say, a ship. Each range bin of data contains 2^{17} complex numbers, with a sampling frequency of 1000 Hz. We analyze the amplitude data. Figure 5 shows two examples of the sea-clutter amplitude data without and with the target. Note that similar signals have been observed in many different fields. Therefore, the analysis below may also be applicable to those fields.

Let us denote the sea clutter amplitude data by $u_1, u_2, \ldots$, the integrated data by $v_1, v_2, \ldots$, and the differenced data by $w_1, w_2, \ldots$. First we apply DFA to $v_1, v_2, \ldots$. A typical result for a measurement (which contains 14 range bins) is shown in Fig. 6(a). From it, one would conclude that the data have excellent fractal scaling behavior. However, this is an illusion due to the large y-axis range in the figure. If one reduces the y-axis range by plotting $\log_2[F_d(m)/m]$ vs $\log_2 m$ (which can be viewed as *detrended Fano factor analysis*; see the Appendix), then one finds that the curves for sea-clutter data without target change abruptly around $m_1=2^4$ and $m_2=2^{12}$. Since the sampling frequency is 1000 Hz, they correspond to time scales of about 0.01 and 4 s. It turns out that if one fits a straight line to the $\log_2[F_d(m)/m]$ vs $\log_2 m$ curves in this

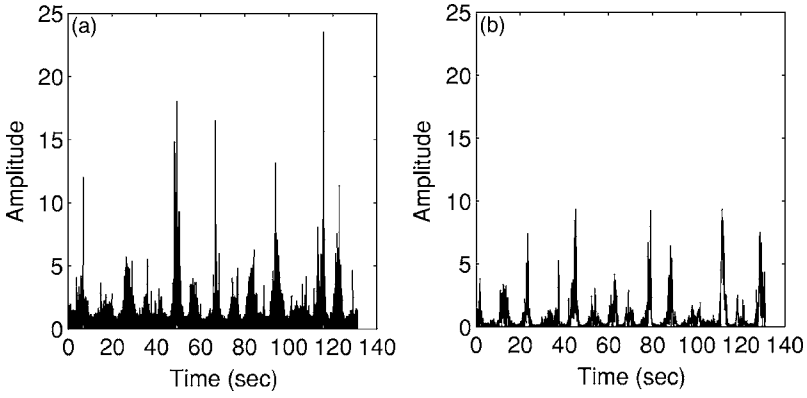

FIG. 5. Examples of the sea-clutter amplitude data (a) without and (b) with target.

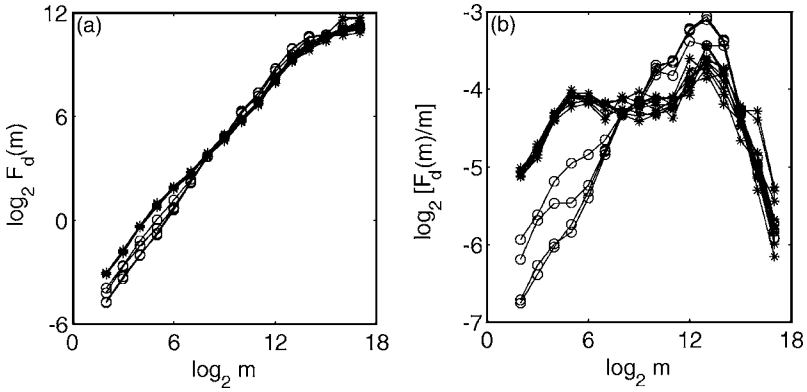

FIG. 6. Target detection within sea clutter using DFA. Open circles designate data with target, while crosses are for data without target.

m range, then the H parameter can completely separate data with and without the target, as shown in Fig. 7. The last statement simply says that the H-based method achieves very high accuracy in detecting targets within sea clutter.

Let us now make a few comments. (i) The time scales of 0.01 and a few seconds have specific physical meanings: below 0.01 s, the data are fairly smooth and hence cannot be fractal; above a few seconds, the wave pattern on the sea surface may change, and hence the data may change to a different behavior (possibly another type of fractal). With the available length of the data (which is about 2 min), the latter cannot be resolved, however. (ii) If one tries to estimate H from other intervals of time (which would be the case when one tries to apply, say, maximum likelihood estimation), then H fails to detect targets within sea clutter. (iii) The fractal scaling in the identified time scale range is actually not excellently defined, especially for data without a target. This implies that sea-clutter data are more complicated than what fractal scaling can characterize. (iv) If one applies DFA to the u_i process, the original sea-clutter amplitude data, then the estimated H_u is about H_v-1, and the H-based method for target detection still works [the result is not shown here, since it is similar to that obtained by FA, which is shown in Fig. 8(a)]. (v) When FA is applied to the u_i process, the obtained H are similar to those by DFA. See Fig. 8(a). Hence, FA is consistent with DFA. However, FA fails to work when it is applied to the integrated data, the v_i process, since all the estimated H_v cannot be larger than 1 [Fig. 8(b)]. (vi) The wavelet H estimator is the most versatile. The H values obtained by applying the method to the u_i and v_i process as well as the w_i process can all be used to detect the target, as shown in Fig. 9. From the figure, in fact, H is increased by 1, progressing from w_i to u_i, and from u_i to v_i. Neither FA nor DFA gives useful result when applied to the w_i process, because of saturation of H at 0 (see the Appendix). (vii) H for some data sets with targets is close to 1/3, the very H corresponding to the famous Kolmogorov energy spectrum of turbulence. This may be due to the development of wave-turbulence interactions around the target, under favorable weather and sea conditions.

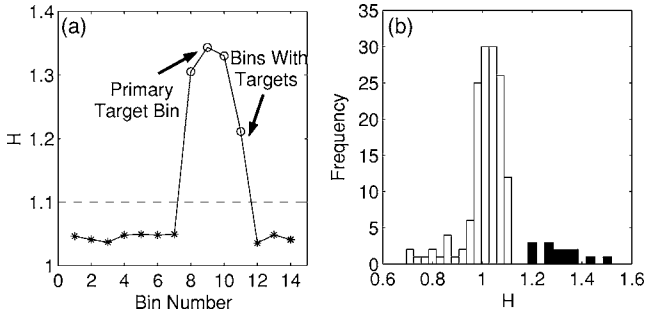

FIG. 7. (a) H parameter estimated by DFA of the v_i process of the 14 range bins of a measurement; (b) histogram (equivalent to probability density function) for the H parameter for all the measurements. Black boxes are for data with target, while open boxes are for data without target.

V. CONCLUDING REMARKS

In this paper, we have shown that methods for assessing long-range correlation in a time series can be grouped into four classes: the spectral method, FA and related methods, DFA, and the wavelet method. Furthermore, we have shown that H estimated by FA and related methods has to lie in the unit interval, H estimated by DFA lies in between 0 and 2 (see the Appendix), while H estimated by the spectral method and the wavelet method can assume any value. For model systems, so long as the estimated H lies in between 0 and 1, all the methods are consistent. However, when H is close to 1 by FA and related methods, or close to 0 by FA and related methods, as well as by DFA, then the result might be incorrect. When inconsistency arises, it is desirable to adopt the following rules of thumb.

Rule of thumb 1. When a time series is treated as a noise

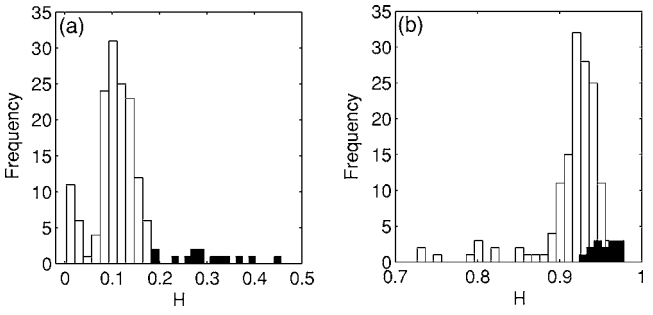

FIG. 8. Histogram for the H parameter estimated by FA from the original sea-clutter data (a) and their integration (b). Black boxes are for data with target, while open boxes are for data without target.

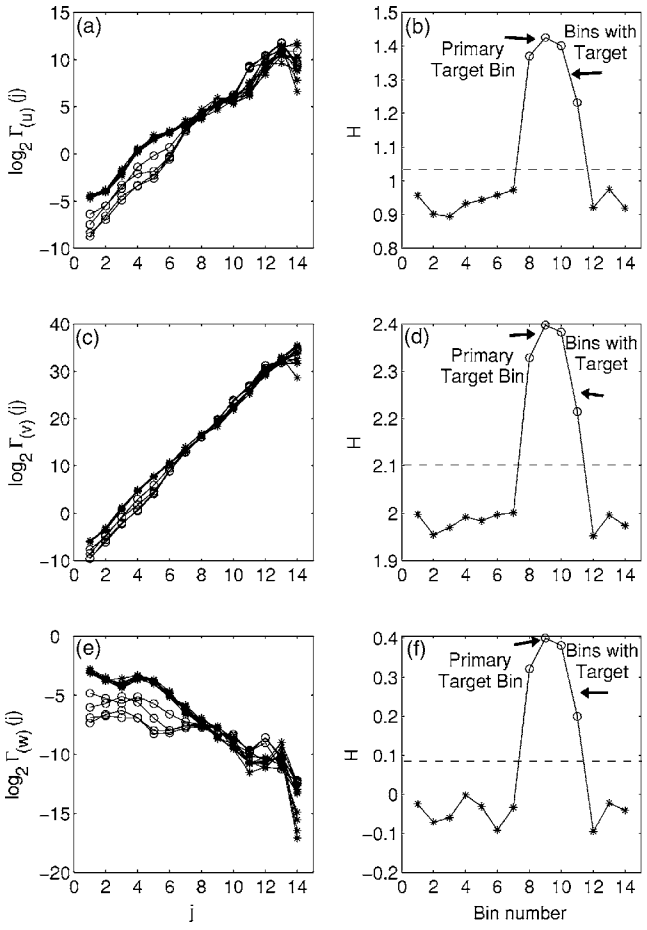

FIG. 9. Wavelet H estimator on (a),(b) the u_i process, (c),(d) the v_i process, and (e),(f) the w_i process.

process and the estimated H parameter is close to 1, question your result; redo the analysis by treating the data as a random walk process.

Rule of thumb 2. When a time series is treated as a random walk process and the estimated H parameter is close to 0, do not trust your result; redo the analysis by integrating the data.

Rule of thumb 3. To be safe, perform DFA or the wavelet-based analysis on your data, along with FA (or other equivalent methods), and check the consistency of the results based on different methods.

We have paid particular attention to the relation between the parameter of a fractal process and the time scale where the fractal scaling breaks. By considering an important engineering problem, target detection within sea-clutter radar returns, we have shown that for pattern recognition purposes, sometimes the fractal scaling break is at least as important as the behavior of fractal scaling. This feature is particularly important in practice, since experimental data are always finite, and therefore may not conform to the ideal mathematical definition of fractal processes with long-range correlations.

We should emphasize that while theoretically the spectral method is a reliable method, there are pitfalls associated with it. Besides the error that the magnitude response is mistaken as the PSD, when the data are short and noisy, it may be difficult to identify a suitable scaling region to estimate H. Therefore, the spectral method is most useful when used for cross checking, but may not be as useful as one might hope when it is used alone.

APPENDIX: STRUCTURE-FUNCTION-BASED MULTIFRACTAL ANALYSIS AND ESTIMATION OF THE HURST PARAMETER

After the random walk process is constructed, it is straightforward to extend FA to a multifractal formalism [46]. That is, for each real q, we examine whether the following scaling relation holds or not:

$$F^{(q)}(m) = \langle |y(i+m) - y(i)|^q \rangle^{1/q} \sim m^{H(q)}, \tag{A1}$$

where the average is taken over all possible pairs of $(y(i+m), y(i))$. Negative and positive q values emphasize small and large absolute increments of $y(n)$, respectively. When the power-law scaling for some q exists, we say the process under study is a fractal process. Furthermore, if $H(q)$ is not a constant function of q, we say the process is a multifractal.

To understand various methods of estimating H, we first note that FA is given by $q=2$. Also note that FA is equivalent to the *variance-time relation* described by Eq. (4), noticing that $\langle |y(i+m)-y(i)|^2 \rangle = m^2 \, \mathrm{Var}(X^{(m)})$.

Closely related to the variance-time relation is the Fano factor analysis, which is quite popular in neuroscience [47–49]. In the context of analysis of the interspike interval of neuronal firings, the Fano factor is defined as

$$F(T) = \frac{\mathrm{Var}[N_i(T)]}{\mathrm{Mean}[N_i(T)]} \tag{A2}$$

where $N_i(T)$ is the number of spikes in the ith window of duration T. For a Poisson process, $F(T)$ is 1, independent of T. For a fractal process, one expects $\mathrm{Var}[N_i(T)] \propto T^{2H}$, while $\mathrm{Mean}[N_i(T)] \propto T$. Therefore, $F(T) \sim T^{2H-1}$. In other words, Fano factor can be viewed as examining the relation between $[\langle |y(i+m)-y(i)|^2 \rangle / m]$ and m instead of the relation between $[\langle |y(i+m)-y(i)|^2 \rangle]$ and m. This is why the relation between $\log_2[F_d(m)/m]$ and $\log_2 m$ can be viewed as a detrended Fano factor analysis.

We now discuss methods that employ $H(1)$ to estimate H. Two such approaches are reviewed by Taqqu *et al.* [50], namely, the *absolute values of the aggregated series approach* and *Higuchi's method*. In the former, one examines if the following scaling law holds:

$$\frac{1}{[N/m]} \sum_{k=1}^{[N/m]} |X^{(m)}(k)| \sim m^{H-1},$$

where N is the length of the time series, $X^{(m)}$ is the nonoverlapping running mean of X of block size m, as defined by Eq. (2), and $[\,]$ denotes the greatest integer function. Higuchi's method, on the other hand, examines if the following scaling law is true:

$$L(m) = \frac{N-1}{m^3} \sum_{i=1}^{m} [(N-i)/m]^{-1} \times \sum_{k=1}^{[(N-i)/m]} |y(i+km) - y(i+(k-1)m)| \sim m^{H-2},$$

where N again is the length of the time series, m is essentially a block size, and $y(i) = \Sigma_{j=1}^{i} X_j$. Note that the two methods are quite similar. In fact, the first summation of Higuchi's method (divided by m) is equivalent to the absolute values of the aggregated series approach. The second summation $\Sigma_{i=1}^{m}$ is another moving average, equivalent to taking overlapping running means of the original time series X. By now, it should be clear that both methods estimate the $H(1)$ parameter instead of the $H(2)$ parameter, when the time series X has mean zero. The reason that $H(1)$ can be used to estimate $H(2)$ is that typically they are quite close, even if the time series X is a multifractal. When the time series X is very much like a monofractal, or only weakly multifractal, then we see that any $H(q)$, $q \neq 2$, can be used to estimate $H(2)$. In this case, the structure-function-based technique provides infinitely many ways of estimating the Hurst parameter.

We note that if the mean of the time series X is not zero, then neither the absolute values of the aggregated series approach nor Higuchi's method estimates $H(1)$. When this is the case, one should remove the mean from the X time series first.

We have pointed out that Higuchi's method is equivalent to take overlapping running means when constructing $X^{(m)}$. We thus see that the condition of "nonoverlapping" for constructing $X^{(m)}$ when defining long memory is not essential.

We now prove why the Hurst parameter estimated by FA saturates at 1. The idea lies in that if the process x has PSD $1/f^{\alpha_x}$, then its integration (i.e., the random walk process) y has PSD $1/f^{\alpha_y}$, with $\alpha_y = 2 + \alpha_x$. If we further integrate y to obtain z, then the PSD for z is $1/f^{\alpha_z}$, with $\alpha_z = 2 + \alpha_y$. The process of integration suggests that we may, without loss of generality, assume $y(n) \sim n^{\beta}$, $\beta > 1$. Then $\langle |y(n+m) - y(n)|^2 \rangle = \langle [(n+m)^{\beta} - n^{\beta}]^2 \rangle$ is dominated by terms with large n. When this is the case, $(n+m)^{\beta} = [n(1+m/n)]^{\beta} \approx n^{\beta}[1+\beta m/n]$. One then sees that $\langle |y(n+m) - y(n)|^2 \rangle \sim m^2$, i.e., $H(2) = 1$.

Similarly, one can prove that (i) H estimated by FA and related methods as well as by DFA has to be non-negative, and (ii) H estimated by DFA cannot be larger than 2 [51]. The wavelet H estimator is most flexible in the sense that there is no constraint on the value of H to be estimated. However, for practical purposes, DFA can be considered sufficient.

[1] W. H. Press, Comments. Astrophys. **7**, 103 (1978).
[2] P. Bak, *How Nature Works: The Science of Self-Organized Criticality* (Copernicus, New York, 1996).
[3] G. M. Wornell, *Signal Processing with Fractals: A Wavelet-Based Approach* (Prentice Hall, Englewood Cliffs, NJ, 1996).
[4] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, IEEE/ACM Trans. Netw. **2**, 1 (1994).
[5] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, IEEE Trans. Commun. **43**, 1566 (1995).
[6] V. Paxson and S. Floyd, IEEE/ACM Trans. Netw. **3**, 226 (1995).
[7] W. Li and K. Kaneko, Europhys. Lett. **17**, 655 (1992).
[8] R. F. Voss, Phys. Rev. Lett. **68**, 3805 (1992).
[9] C.-K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons, and H. E. Stanley, Nature (London) **356** 168 (1992).
[10] D. L. Gilden, T. Thornton, and M. W. Mallon, Science **267**, 1837 (1995).
[11] Y. Chen, M. Ding, and J. A. Scott Kelso, Phys. Rev. Lett. **79**, 4501 (1997).
[12] J. J. Collins and C. J. De Luca, Phys. Rev. Lett. **73**, 764 (1994).
[13] V. A. Billock, Physica D **137**, 379 (2000).
[14] V. A. Billock, G. C. de Guzman, and J. A. S. Kelso, Physica D **148**, 136 (2001).
[15] M. Wolf, Physica A **241**, 493 (1997).
[16] J. B. Gao, Y. H. Cao, and J. M. Lee, Physica A **314**, 392 (2003).
[17] B. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).
[18] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, Phys. Rev. E **49**, 1685 (1994).
[19] G. Rangarajan and M. Ding, Phys. Rev. E **61**, 4991 (2000).
[20] J. B. Gao, Phys. Rev. E **63**, 066202 (2001).
[21] G. Box, G. M. Jenkins, and G. Reinsel, *Time Series Analysis: Forecasting & Control*, 3rd ed. (Prentice Hall, Englewood Cliffs, NJ, 1994).
[22] D. R. Cox, in *Statistics: An Appraisal*, edited by H. A. David and H. T. Davis (Iowa State University Press, Ames, 1984), pp. 55–74.
[23] B. Tsybakov and N. D. Georganas, IEEE/ACM Trans. Netw. **5**, 397 (1997).
[24] P. Abry and D. Veitch, IEEE Trans. Inf. Theory **44**, 2 (1998).
[25] R. G. Addie, M. Zukerman, and T. Neame, in Proc. IEEE INFOCOM '95, Vol. 3, p. 977.
[26] B. Kaulakys, and T. Meskauskas, Phys. Rev. E **58**, 7013 (1998).
[27] P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987).
[28] H. M. Jaeger, C. H. Liu, and S. R. Nagel, Phys. Rev. Lett. **62**, 40 (1989).
[29] P. Helander *et al.*, Phys. Rev. E **59**, 6356 (1999).
[30] F. Dalton and D. Corcoran, Phys. Rev. E **63**, 061312 (2001).
[31] Hans-Jacob S. Feder and J. Feder, Phys. Rev. Lett. **66**, 2669 (1991); **67**, 283(E) (1991).
[32] H. J. Jensen, K. Christensen, and H. C. Fogedby, Phys. Rev. B **40**, 7425 (1989).
[33] M. S. Taqqu, W. Willinger, and R. Sherman, Comput. Commun. Rev. **27**, 5 (1997).
[34] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian*

Random Processes (Chapman and Hall, London, 1994).
[35] T. Thayaparan and S. Kennedy, IEE Proc., Radar Sonar Navig. **151**, 19 (2004).
[36] G. Davidson and H. D. Griffiths, Electron. Lett. **38**, 1128 (2002).
[37] T. Bhattacharya and S. Haykin, Electron. Lett. **28**, 1528 (1992).
[38] T. Bhattacharya and S. Haykin, IEEE Trans. Aerosp. Electron. Syst. **33**, 408 (1997).
[39] H. Leung, N. Dubash, and N. Xie, IEEE Trans. Aerosp. Electron. Syst. **38**, 98 (2002).
[40] N. Xie, H. Leung, and H. Chan, IEEE Trans. Geosci. Remote Sens. **GE-41**, 1491 (2003).
[41] C. P. Lin, M. Sano, S. Sayama, and M. Sekine, IEICE Trans. Commun. **E83B**, 1916 (2000).
[42] C. P. Lin, M. Sano, S. Obi, S. Sayama, and M. Sekine, IEICE Trans. Commun. **E83B**, 1955 (2000).
[43] M. Martorella, F. Berizzi, and E. D. Mese, IEEE Trans. Antennas Propag. **52**, 1193 (2004).
[44] S. Haykin, R. Bakker, and B. W. Currie, Proc. IEEE **90**, 860 (2002);S. Haykin, http://soma.ece.mcmaster.ca/ipix/dartmouth/datasets.html
[45] J. Hu, J. B. Gao, K. Yao, and U. S. Kim, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05), Philadelphia, March 2005, Vol. 5, p. 18–23; see also J. Hu, W. W. Tung, and J. B. Gao, in IEEE Trans. Antennas and Propagation (to be published).
[46] U. Frisch, *Turbulence—The Legacy of A. N. Kolmogorov* (Cambridge University Press, Cambridge, U.K., 1995).
[47] M. C. Teich, IEEE Trans. Biomed. Eng. **36**, 150 (1989).
[48] S. B. Lowen and M. C. Teich, J. Acoust. Soc. Am. **92**, 803 (1992).
[49] M. C. Teich and S. B. Lowen, IEEE Eng. Med. Biol. Mag. **13**, 197 (1994).
[50] M. S. Taqqu, V. Teverovsky, and W. Willinger, Fractals **3**, 785 (1995).
[51] P. Talkner and R. O. Weber, Phys. Rev. E **62**, 150 (2000).