Eliminating finite-size effects and detecting the amount of white noise in short records with long-term memory

Sabine Lennartz* and Armin Bunde
Institut für Theoretische Physik III, Justus-Liebig-Universität Giessen, 35392 Giessen, Germany
(Received 24 October 2008; revised manuscript received 23 January 2009; published 5 June 2009)

Long-term memory is ubiquitous in nature and has important consequences for the occurrence of natural hazards, but its detection often is complicated by the short length of the considered records and additive white noise in the data. Here we study synthetic Gaussian distributed records x_i of length N that consist of a long-term correlated component $(1-a)y_i$ characterized by a correlation exponent γ, $0<\gamma<1$, and a white-noise component $a\eta_i$, $0\le a\le 1$. We show that the autocorrelation function $C_N(s)$ has the general form $C_N(s)=[C_\infty(s)-E_a]/(1-E_a)$, where $C_\infty(0)=1$, $C_\infty(s>0)=B_a s^{-\gamma}$, and $E_a=\{2B_a/[(2-\gamma)(1-\gamma)]\}N^{-\gamma}+O(N^{-1})$. The finite-size parameter E_a also occurs in related quantities, for example, in the variance $\Delta_N^2(s)$ of the local mean in time windows of length s: $\Delta_N^2(s)=[\Delta_\infty^2(s)-E_a]/(1-E_a)$. For purely long-term correlated data $B_0\cong(2-\gamma)(1-\gamma)/2$ yielding $E_0\cong N^{-\gamma}$, and thus $C_N(s)=[\frac{(2-\gamma)(1-\gamma)}{2}s^{-\gamma}-N^{-\gamma}]/[1-N^{-\gamma}]$ and $\Delta_N^2(s)=[s^{-\gamma}-N^{-\gamma}]/[1-N^{-\gamma}]$. We show how to estimate E_a and $C_\infty(s)$ from a given data set and thus how to obtain accurately the exponent γ and the amount of white noise a.

DOI: 10.1103/PhysRevE.79.066101 PACS number(s): 89.75.Da, 64.60.an, 95.75.Wx

I. INTRODUCTION

Long-term correlations are ubiquitous in nature. They occur, for example, in climate [1–9], physiology [10–12] and financial markets [13,14], and possibly also in earthquakes [15,16]. Long-term memory has important consequences for the occurrence of extreme events. Since the return intervals between consecutive events above some threshold q are also long-term correlated, long-term memory leads to a clustering of extreme events and thus is important for risk estimation [17].

The most direct quantities that characterize the long-term memory in records of a given length N are the autocorrelation function (ACF) $C_N(s)$ and the variance of the local mean of s successive data points $\Delta_N^2(s)$. For $N\to\infty$ both quantities scale as $s^{-\gamma}$ with $0<\gamma<1$. Usually the observational records of interest have a length between $N=10^3$ and $N=10^5$ data points. It is known that in these comparatively short records an accurate determination of γ, via $C_N(s)$ or $\Delta_N^2(s)$, is difficult to achieve, in particular when γ is small.

To test for long-term memory, one thus is led to use less direct methods as, e.g., the various detrending fluctuation [12,18–20] and wavelet techniques [21], which also have the advantage of eliminating polynomial trends in the data. In the presence of considerable additive white noise [see Eq. (2) with a close to 1], these methods usually cannot be used and thus one has to rely solely on the ACF [19,20].

In this paper we focus on $C_N(s)$ and $\Delta_N^2(s)$. We derive an analytical expression for the a, γ, N, and s dependences of $C_N(s)$, which shows how $C_N(s)$ is related to the ACF in the thermodynamic limit. We describe how to extract $C_\infty(s)$ and thus the proper γ value as well as the amount of white noise a from the data. We also derive a simple scaling relation between $C_N(s)$ and $\Delta_N^2(s)$ and show this way that both quantities are fully equivalent, with the same information content.

*sabine.lennartz@uni-giessen.de

II. FINITE-SIZE EFFECTS IN THE ACF

We consider a long-term correlated, equidistant record of N data x_i, $i=1,\ldots,N$ with the mean $\langle x\rangle_N=\frac{1}{N}\sum_{i=1}^N x_i$ and the variance $\sigma_N^2=\frac{1}{N}\sum_{i=1}^N(x_i-\langle x\rangle_N)^2$. The ACF is defined as

$$C_N(s)=\frac{1}{\sigma_N^2(N-s)}\sum_{i=1}^{N-s}(x_i-\langle x\rangle_N)(x_{i+s}-\langle x\rangle_N), \qquad (1)$$

with $C_N(0)=1$. In the limit of $N\to\infty$, $C_N(s)$ scales as $C_\infty(s)\sim s^{-\gamma}$, where $0<\gamma<1$ is the correlation exponent.

To study $C_N(s)$ we have generated, for several values of γ, ten data sets with length $N=2^{21}$ by the classical Fourier transform technique (see [22]) and divided these sequences into (i) 10×2^3 subsequences of length $N=2^{18}$, (ii) 10×2^5 subsequences of length $N=2^{16}$, and (iii) 10×2^7 subsequences of length $N=2^{14}$. When calculating the ACF we averaged over all subsets of fixed length, such that the statistics do not depend on the length N of the data sets.

Figure 1(a) shows the resulting ACF for $\gamma=0.4$ and $N=2^{21}$, 2^{18}, 2^{16}, and 2^{14} (upper set of curves from top to bottom). For comparison, the dashed straight lines show the expected power-law behavior of $C_\infty(s)$. While for the longest data set the ACF shows approximately the expected slope; the ACF bends down for the shorter data sets. This effect becomes stronger with decreasing system size N.

Figures 1(b)–1(d) show the same results for $\gamma=0.3$, 0.2, and 0.1. For decreasing γ the finite-size effect increases. Especially for $\gamma=0.1$ the finite-size effect is so strong that even for $N=2^{21}$ the curve does not exhibit the correct slope. For all four curves the effective exponent, measured between $s=10^1$ and 10^2, is between 0.2 and 0.3 instead of 0.1.

Next, we consider noisy records x_i, $i=1,\ldots,N$, which consist of a mixture of long-term correlated data y_i and uncorrelated data (white noise) η_i, i.e.,

$$x_i=(1-a)y_i+a\eta_i, \qquad (2)$$

where $\langle\eta\rangle_N=\langle y\rangle_N=\langle\eta y\rangle_N=0$ and $\langle\eta^2\rangle_N=\langle y^2\rangle_N=1$. The results for $a=0.8$ are shown in Figs. 1(a)–1(d), lower sets of curves.

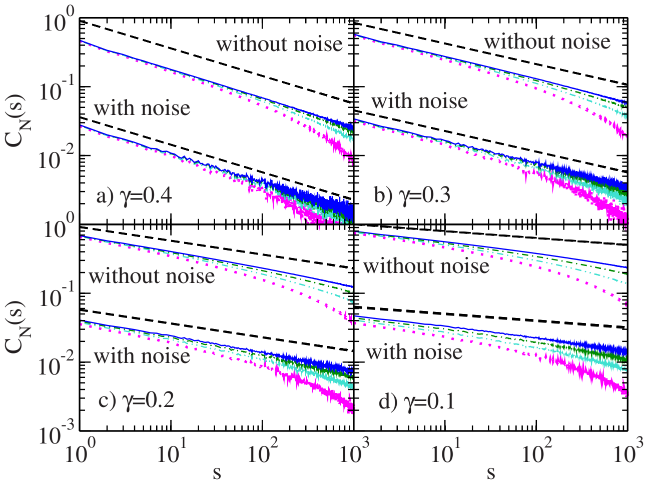

FIG. 1. (Color online) Autocorrelation function $C_N(s)$ of long-term correlated records x_i with correlation exponent (a) $\gamma=0.4$, (b) $\gamma=0.3$, (c) $\gamma=0.2$, and (d) $\gamma=0.1$ and lengths $N=2^{21}$, 2^{18}, 2^{16}, and 2^{14} (from top to bottom) without white noise (upper sets of curves) and with 80% white noise (lower sets of curves). The dashed lines are power laws with exponent $-\gamma$ for comparison.

For the ACF, the additive noise results in a vertical shift in the double logarithmic plot, and the finite-size effects are nearly the same as for the data without white noise.

We like to note that the correlation exponent γ can also be determined from fluctuation functions $F^2(s)$, which are either related to the variance $\Delta_N^2(s)$ (see Sec. IV) or have been obtained from detrending techniques such as the DFA [12,18–20]. For purely long-term correlated data $F^2(s)$ scales as $F^2(s)\sim s^{2\alpha}$, with $\alpha=1-\gamma/2$ and an accurate determination of γ is possible. This changes in the presence of additive white noise where an extra term $\sim s$ occurs in $F^2(s)$. When the white noise is dominant then this term contributes considerably to $F^2(s)$ also for large s values, making a determination of γ impossible [19,20].

A further method to detect γ is to calculate the power spectrum $S(k)$ which scales as $S(k)\sim k^{-(1-\gamma)}$ for purely long-term correlated data. For additive white noise a constant value is added. If the white noise is sufficiently large, γ cannot be obtained from the decay of $S(k)$ at small k values.

Accordingly, in the presence of white noise, the ACF method is the only method to detect long-term correlations, irrespective of its pronounced finite-size effects. In the following, we derive an analytical expression for $C_N(s)$, which allows to overcome these finite-size effects and thus to identify the long-term memory as well as the amount of the additive white noise.

III. ELIMINATING FINITE-SIZE EFFECTS IN THE ACF AND DETERMINATION OF THE AMOUNT OF ADDITIVE WHITE NOISE

To identify the origin of the finite-size effects in the ACF, we consider the identity

$$
\begin{aligned}
0 &= \frac{1}{N^2\sigma_N^2}\sum_{l,m=1}^{N} x_l x_m - \frac{\langle x\rangle_N^2}{\sigma_N^2} \\
&= \frac{1}{N^2\sigma_N^2}\sum_{l,m=1}^{N} (x_l-\langle x\rangle_N)(x_m-\langle x\rangle_N) \\
&= \frac{1}{N} + \sum_{s=1}^{N-1}\frac{2(N-s)C_N(s)}{N^2},
\end{aligned}
\tag{3}
$$

which results in the *sum rule*

$$\sum_{s=1}^{N-1}\frac{2(N-s)}{N}C_N(s) = -1. \tag{4}$$

We like to note that this relation holds in general, irrespective of the kind of correlations and distributions considered. It is clear that this sum rule cannot be satisfied for any ACF with $C_N(s)$ positive. In particular, it is inconsistent with the assumption $C_N(s)=\delta_{s,0}$ for uncorrelated records ($a=1$) and with the assumption that the ACF of a long-term correlated record of length N has the form $C_N(s)=C_\infty(s)$, with

$$C_\infty(s) = \begin{cases} 1 & ,s=0 \\ B_a s^{-\gamma} & ,s>0. \end{cases} \tag{5}$$

For noisy long-term correlated records [Eq. (2)],

$$B_a = B_0(1-a)^2/[a^2+(1-a)^2] \equiv B_0 p, \tag{6}$$

where B_0 refers to the pure long-term correlated record ($a=0$). It is easy to verify that the ansatz

$$C_N(s) = \frac{1}{1-E}[C_\infty(s)-E], \tag{7}$$

with

$$E = \frac{1}{N} + \sum_{l=1}^{N-1}\frac{2(N-l)}{N^2}C_\infty(l) \tag{8}$$

satisfying the sum rule [Eq. (4)]. For uncorrelated records, Eq. (7) reduces to $C_N(s)=\delta_{s,0}-\frac{1}{(N-1)}(1-\delta_{s,0})$.

For finite records with long-term memory, E may become of the order of unity and cannot be neglected. Substitution of Eq. (8) into Eq. (7) yields the general relation

$$C_N(s) = \frac{C_\infty(s) - \frac{1}{N} - \sum_{l=1}^{N-1}\frac{2(N-l)}{N^2}C_\infty(l)}{1-\frac{1}{N}-\sum_{l=1}^{N-1}\frac{2(N-l)}{N^2}C_\infty(l)}, \tag{9}$$

which replaces $C_\infty(s)$ for data sets of finite length N.

We like to note that Eqs. (7)–(9) are general and hold for all functions $C_\infty(s)$, for which the sum in Eq. (8) tends to zero in the thermodynamic limit $N\to\infty$.

In the following, we assume that the data are described by Eq. (2), and thus $C_\infty(s)$ is described by Eq. (5). In this case, $E\equiv E_a$ becomes

$$E_a = \frac{1}{N} + B_a\sum_{s=1}^{N-1}\frac{2(N-s)}{N^2}s^{-\gamma}. \tag{10}$$

For $N\gg 1$, one can easily verify that

$$E_a = \frac{2B_a}{(2-\gamma)(1-\gamma)} N^{-\gamma} + O(N^{-1}) \tag{11}$$

$$\approx B_a \left(\frac{s_x}{N}\right)^{\gamma}, \tag{12}$$

with

$$s_x = \left(\frac{2}{(2-\gamma)(1-\gamma)}\right)^{1/\gamma}. \tag{13}$$

Accordingly, $C_N(s)$ becomes

$$C_N(s) = \frac{B_a}{1 - B_a(s_x/N)^{\gamma}} s^{-\gamma}\left[1 - \left(\frac{s s_x}{N}\right)^{\gamma}\right], \tag{14}$$

with B_a from Eq. (6). Equation (14) shows, how the autocorrelation function $C_N(s)$, up to $O(N^{-1})$, depends on N, γ, s, and the amount of white noise a in the data. In the following, we suggest a method how the exponent γ and the prefactor B_a can be determined efficiently. Then we determine B_0 as a function of γ and show how to determine a from the data set.

According to Eq. (14), deviations from the power-law behavior become large when the expression in the brackets significantly deviates from 1, for example, when it is equal to $(1-\delta)$. This happens for $s > k(\gamma,\delta)N$ with $k(\gamma,\delta) = \delta^{1/\gamma}/s_x$. For a simple estimate, we choose $\delta = 1/5$. Then for $\gamma = 0.1$, 0.2, 0.3, and 0.4 we have $k(\gamma, 1/5) \simeq 2\times 10^{-8}$, 6×10^{-5}, 8×10^{-4},and 3×10^{-3}, respectively. Accordingly, for N of the order of 10^5, $k(\gamma,1/5)N$ is well above unity only for $\gamma \geq 0.4$, and only in this case one can obtain a good guess of γ from the initial slope of $C_N(s)$ obtained in the first decade [as can be seen in Fig. 1(a)]. For smaller γ-values, however, the finite-size effects are so strong that even in very large records of length 10^6 the proper exponent γ cannot be observed. In this case, to extract γ (and thus s_x) from Eq. (14), we have to take into account the functional form of $C_N(s)$ in the whole s regime. To do this in the most efficient way, we consider $C_N(s)$ for three s values, s_1, s_2, and $s_3 = s_2^2/s_1$, and combine them. This yields

$$\frac{C_N(s1) - C_N(s_2)}{C_N(s_2) - C_N(s_2^2/s_1)} = \frac{s_1^{-\gamma} - s_2^{-\gamma}}{s_2^{-\gamma} - (s_2^2/s_1)^{-\gamma}} = (s_2/s_1)^{\gamma}, \tag{15}$$

and thus

$$\gamma = \frac{\ln\left(\frac{C_N(s1) - C_N(s_2)}{C_N(s_2) - C_N(s_2^2/s_1)}\right)}{\ln(s_2/s_1)}. \tag{16}$$

For obtaining a reliable numerical estimation for γ, it is necessary to consider pairs (s_1 and s_2) such that the differences in Eq. (16) are not too small. This can be achieved, for example, by $s_1 = 1$ (or 2) and s_2 well above 10, but such that $s_2^2/s_1 \ll N$. When γ is known, the parameter B_a can also be extracted from Eq. (14),

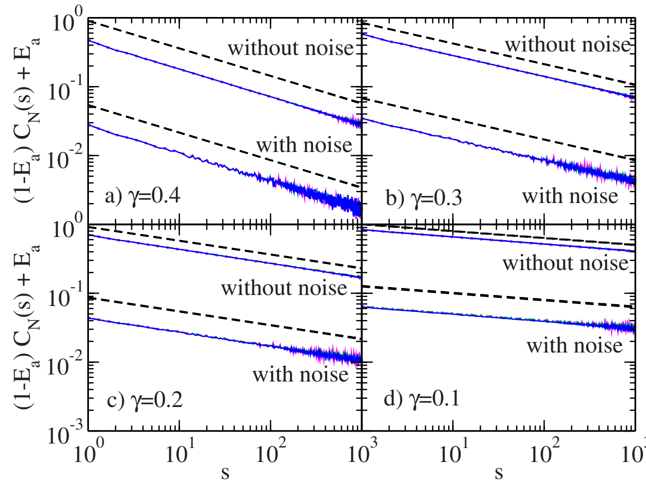

FIG. 2. (Color online) $(1-E_a)C_N(s)+E_a$ for the same long-term correlated records x_i as in Fig. 1. The curves for different datalengths collapse. The dashed lines are power laws with exponent $-\gamma$ for comparison.

$$B_a = \frac{C_N(s)}{s^{-\gamma} - (s_x/N)^{\gamma}[1 - C_N(s)]}, \tag{17}$$

$$\approx C_N(s) s^{\gamma}\{1 + (s_x/N)^{\gamma}[1 - C_N(s)]\}, \tag{18}$$

which then yields E_a with Eq. (10).

We have applied this approach to our long-term correlated data with $\gamma = 0.4$, 0.3, 0.2, and 0.1, each with length $N = 2^{21}$, 2^{18}, 2^{16}, and 2^{14}, and with and without additive white noise ($a = 0.8$ and 0). As before, in order to keep the statistics the same, we averaged $C_N(s)$ over 10, 10×2^3, 10×2^5, and 10×2^7 records of length $N = 2^{21}$, $N = 2^{18}$, $N = 2^{16}$, and $N = 2^{14}$, respectively. To obtain γ we used Eq. (16), where we averaged over the pairs (s_1 and s_2) with $s_1 = 1$, $s_2 = 15, 16, 17, \ldots, 25$ and $s_1 = 2$, $s_2 = 16, 18, 20, \ldots, 34$. To obtain B_a, we then used Eq. (17), where we averaged over the first 50 s values, respectively.

Figure 2 shows the result for $(1-E_a)C_N(s)+E_a$, which should be identical to $C_\infty(s)$. In the figure the curves for different data lengths cannot be distinguished since they all collapse, for each γ value, to a single line. The dashed lines are power laws with exponent $-\gamma$ for comparison. In all these figures $(1-E_a)C_N(s)+E_a$ shows the expected power-law behavior, confirming Eqs. (7) and (14).

Next, we show how the amount of white noise a can be determined. Using Eq. (6) we obtain

$$a = [1 + \sqrt{B_a/(B_0 - B_a)}]^{-1} \tag{19}$$

when B_a and B_0 are known. To determine B_0, we utilize Eq. (12) and plot (in Fig. 3) E_0 as a function of $N^{-\gamma}$ for $\gamma = 0.1$, 02, 03, 04, 0.6, and 0.8, and $N = 2^{21}$, 2^{18}, 2^{16}, and 2^{14}. The figure shows that all data collapse to a single line with slope 1, which then yields

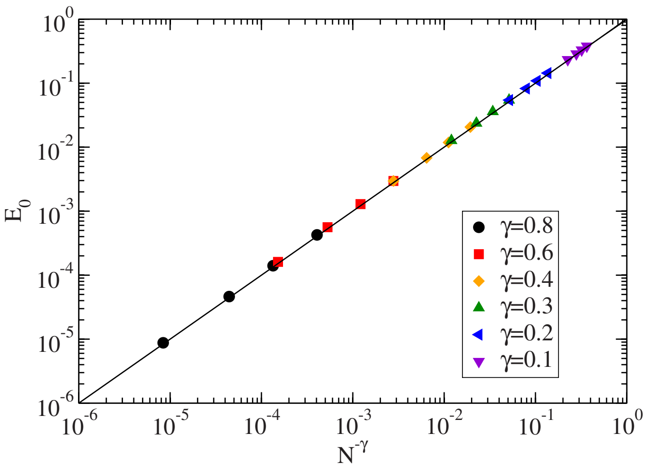

FIG. 3. (Color online) Finite-size correction E_0 (symbols) as a function of $N^{-\gamma}$ for γ=0.8, 0.6, 0.4, 0.3, 0.2, and 0.1, and $N=2^{21}$, 2^{18}, 2^{16}, and 2^{14}. All values of E_0 collapse to a single line with slope 1.

$$B_0 \cong s_x^{-\gamma} = \frac{(2-\gamma)(1-\gamma)}{2}. \tag{20}$$

Thus we obtain for $C_N(s)$, in the absence of additive white noise, the remarkable simple result

$$C_N(s) = \frac{1}{1-N^{-\gamma}}\left(\frac{(2-\gamma)(1-\gamma)}{2}s^{-\gamma} - N^{-\gamma}\right). \tag{21}$$

By measuring B_a via Eq. (17) and using Eqs. (19) and (20), we can directly obtain the amount of white noise in the data.

We like to note, that this procedure also can be done in the presence of weak trends, which by definition affect $C_N(s)$ only for very large s values. This is obvious for $\gamma \gtrsim 0.4$, where for obtaining γ and B_a only the initial slope of $C_N(s)$ in the first decade is needed. For smaller γ values, the detection of γ is more difficult within the proposed method and can only be done for very weak trends where $C_N(s_2^2/s_1)$ is not affected by the trend.

IV. FINITE-SIZE EFFECT IN RELATED QUANTITIES

Next we consider the variance of the local mean of s successive data points,

$$\Delta_N^2(s) = \left\langle \left(\frac{1}{s}\sum_{i=1}^{s} \frac{(x_i - \langle x\rangle_N)}{\sigma_N}\right)^2 \right\rangle, \tag{22}$$

where the average has been taken over all windows of size s in the record. By definition, $R_N^2(s) = s^2\Delta_N^2(s)$ can also be interpreted as the mean-square displacement of a random walker along a linear chain where $|(x_i - \langle x\rangle_N)/\sigma_N|$ is the length of the ith step and its sign determines the direction. In this interpretation, we are interested to see how the mean-square displacement $R_N^2(s)$ of the random walker depends on the time lag s and on the total number of time steps N he performed.

We like to note, that $R_N(s)$ is identical to the FA fluctuation function, which has been used before to detect γ, and in conjunction with the DFA, possible trends in the data [5,9,18].

To find the way $\Delta_N^2(s)$ depends on N we start from the identity

$$\Delta_N^2(s) = \left\langle \left(\frac{1}{s}\sum_{i=1}^{s} \frac{(x_i - \langle x\rangle_N)}{\sigma_N}\right)^2 \right\rangle = \frac{2}{\sigma_N^2 s^2}\sum_{m=1}^{s-1}\sum_{i=1}^{s-m}\langle (x_i - \langle x\rangle_N)$$
$$\times (x_{i+m} - \langle x\rangle_N)\rangle + \frac{1}{s} = \frac{2}{s^2}\sum_{m=1}^{s-1}(s-m)C_N(m) + \frac{1}{s}. \tag{23}$$

Substituting Eq. (7) in Eq. (23) we obtain the general relationship

$$\Delta_N^2(s) = \frac{1}{1-E}[\Delta_\infty^2(s) - E], \tag{24}$$

where

$$\Delta_\infty^2(s) = \begin{cases} 1 & ,s=1 \\ \dfrac{2}{s^2}\displaystyle\sum_{m=1}^{s-1}(s-m)C_\infty(m) + \dfrac{1}{s} & ,s>1 \end{cases} \tag{25}$$

is the variance of the local mean in the limit of $N \to \infty$.

For $C_\infty(m)$ from Eq. (5) it can be shown easily that up to an error of $O(B_a\gamma s^{-1})$, $\Delta_\infty^2(s)$ is given by

$$\Delta_\infty^2(s) \approx B_a\left(\frac{s}{s_x}\right)^{-\gamma} + s^{-1}(1 - B_a s_x^\gamma) \tag{26}$$

with s_x from Eq. (13).

For purely long-term correlated data, $B_0 s_x^\gamma \approx 1$ [see Eq. (20)], which yields $\Delta_\infty^2(s) \approx s^{-\gamma}$ and

$$\Delta_N^2(s) = \frac{s^{-\gamma}}{1-N^{-\gamma}}[1-(s/N)^\gamma], \tag{27}$$

as well as

$$R_N^2(s) = \frac{s^{2-\gamma}}{1-N^{-\gamma}}[1-(s/N)^\gamma]. \tag{28}$$

Equations (27) and (28) show that the finite-size effects in $\Delta_N^2(s)$ and $R_N^2(s)$ are smaller than the finite-size effects in $C_N(s)$ [since $s_x > 1$, compare with Eq. (14)] but still large. This fact has already been observed earlier [without knowing the exact form of $\Delta_N^2(s)$ or $R_N^2(s)$] where it was suggested that $R_N^2(s)$ can only give reliable results for s below $N/10$ [5]. Equation (28) shows that for small γ values the reliability regime for s is even smaller.

Figure 4(a) shows $\Delta_N^2(s)$ for the same sets of data as in Figs. 1 and 2 without additive noise (a=0). Figure 4(b) shows $(1-E_0)\Delta_N^2(s) + E_0$, which should be identical to $\Delta_\infty^2(s)$. The fact that now, for each γ value, all curves collapse to the same curve $\Delta_\infty^2(s)$ confirms Eq. (24). The fact that $\Delta_\infty^2(s) \approx s^{-\gamma}$ confirms Eq. (26) and is also a further numerical proof for Eq. (20).

From Eqs. (5), (7), (24), and (26), it is easy to see that $\Delta_N^2(s)$ and $C_N(s)$ are related by the scaling relation

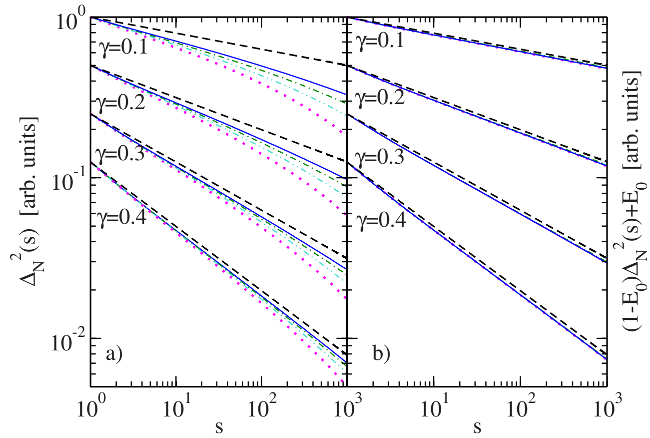

FIG. 4. (Color online) (a) $\Delta_N^2(s)$ and (b) $[(1-E_0)\Delta_N^2(s)+E_0]$ for the same long-term correlated records x_i as in the previous figures but only without additive white noise. The curves $(1-E_0)\Delta_N^2(s)+E_0$ collapse for different data lengths. The dashed lines are power laws with exponent $-\gamma$ for comparison. The curves for γ=0.2, 0.3, and 0.4 have been shifted downwards by factors 2, 4, and 8, respectively, for clarification.

$$\Delta_N^2(s) = C_N(s/s_x) + O(s^{-1}) \tag{29}$$

in the absence of additive white noise (a=0). Figure 5(a) shows that this relation holds quite general not only for large s values.

In the presence of additive white noise the term $\sim s^{-1}$ remains in Eq. (26) and dominates $\Delta_\infty^2(s)$ for $s < (s_x^{-\gamma}/B_a)^{1/(1-\gamma)}$. Since in this case $B_a \cong s_x^{-\gamma}p$, Eq. (26) becomes

$$\Delta_\infty^2(s) \approx ps^{-\gamma} + (1-p)s^{-1}, \tag{30}$$

and thus

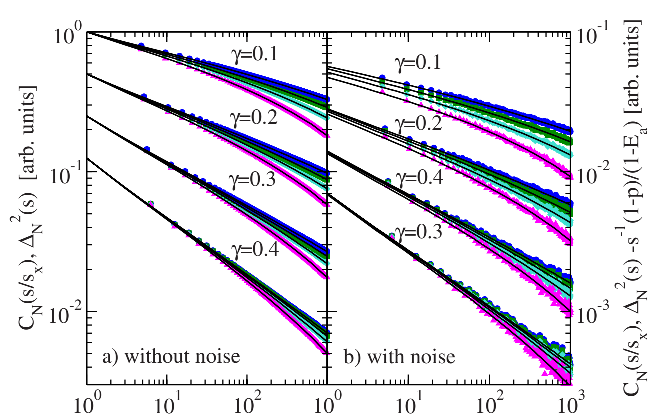

FIG. 5. (Color online) $C_N(s/s_x)$ (symbols) and $\Delta_N^2(s)-s^{-1}(1-p)/(1-E_a)$ (black full lines) for the same records as in the previous figures (a) without white noise (p=1) and (b) with white noise. Note that the black full lines and the symbols collapse. The curves for γ=0.2, 0.3, and 0.4 have been shifted downwards by factors 2, 4, and 8, respectively, for clarification.

$$\Delta_N^2(s) - s^{-1}(1-p)/(1-E) = C_N(s/s_x) + O(s^{-1}) \tag{31}$$

in the presence of additive white noise which extends Eq. (29). Figure 5(b) shows that this relation is also valid not only for large s values.

We like to note finally that in the absence of additive white noise the relevant parameters γ and B_0 can be obtained from both, $\Delta_N^2(s)$ and $C_N(s)$ since both quantities are described by simple power laws. In the presence of additive white noise, γ and B_a can only be obtained from $C_N(s)$ since the crossover in $\Delta_N^2(s)$ prohibits a straightforward calculation of these parameters. A similar crossover occurs also in the DFA fluctuation functions [19,20], and this is the reason why γ cannot be obtained accurately by DFA for very noisy records.

V. CONCLUSION

In summary, we have considered synthetic Gaussian distributed records of lengths N that are in the range of most observational records. The data consist of a long-term correlated component characterized by a correlation exponent γ and a white-noise component. We have derived analytical expressions for the ACF $C_N(s)$ and the related variance $\Delta_N^2(s)$ that specify the way these quantities depend on N, γ, s, and the amount of additive white noise a. For purely long-term correlated records the functional forms become remarkably simple, $C_N(s)=[\frac{(2-\gamma)(1-\gamma)}{2}s^{-\gamma}-N^{-\gamma}]/[1-N^{-\gamma}]$ and $\Delta_N^2(s)=[s^{-\gamma}-N^{-\gamma}]/[1-N^{-\gamma}]=C_N(s/s_x)$, with $s_x=[(2-\gamma)(1-\gamma)/2]^{-1/\gamma}$.

We have presented a procedure how to obtain $C_\infty(s)$ and $\Delta_\infty^2(s)$ in the thermodynamic limit for data with and without additive white noise and thus how to extract from $C_N(s)$ the correlation exponent γ and the amount of additive white noise a.

For records characterized by a large amount of white noise γ and a cannot be obtained from $\Delta_\infty^2(s)$ since $\Delta_\infty^2(s)$ exhibits a pronounced crossover behavior [see Eq. (30)]. This crossover behavior occurs also in the fluctuation functions obtained by the DFA (see [19,20]), which therefore cannot be used to determine these parameters in the case of very noisy long-term correlated data.

Our analysis was restricted to data in the absence of trends and additional short-term correlations. But to a certain extend, the results presented here apply also to these cases. For example, weak trends only matter at very large scales (so that our results remain valid at smaller scales), and short-term correlations can be eliminated by averaging over time scales above the short-term correlation time. If this time is large the averaging procedure strongly reduces the number of data and thus statistical fluctuations will reduce the accuracy of the analysis. But this limitation also applies to the other methods, making it generally difficult to distinguish between pronounced short-term correlations and long-term memory.

ACKNOWLEDGMENT

We like to thank M. I. Bogachev for very valuable discussions.

[1] B. B. Mandelbrot and J. R. Wallis, Water Resour. Res. **5**, 321 (1969).

[2] Y. Ashkenazy, D. R. Baker, H. Gildor, and S. Havlin, Geophys. Res. Lett. **30**, 2146 (2003).

[3] E. Koscielny-Bunde, A. Bunde, S. Havlin, and Y. Goldreich, Physica A **231**, 393 (1996).

[4] J. D. Pelletier and D. L. Turcotte, J. Hydrol. **203**, 198 (1997); Adv. Geophys. **40**, 91 (1999).

[5] E. Koscielny-Bunde, A. Bunde, S. Havlin, H. E. Roman, Y. Goldreich, and H.-J. Schellnhuber, Phys. Rev. Lett. **81**, 729 (1998).

[6] R. O. Weber and P. Talkner, J. Geophys. Res., [Atmos.] **106**, 20131 (2001).

[7] J. W. Kantelhardt, E. Koscielny-Bunde, D. Rybski, P. Braun, A. Bunde, and S. Havlin, J. Geophys. Res., [Atmos.] **111**, D01106 (2006).

[8] J. F. Eichner, E. Koscielny-Bunde, A. Bunde, S. Havlin, and H.-J. Schellnhuber, Phys. Rev. E **68**, 046133 (2003).

[9] D. Rybski and A. Bunde, Physica A **388**, 1687 (2009).

[10] C.-K. Peng, J. Mietus, J. M. Hausdorff, S. Havlin, H. E. Stanley, and A. L. Goldberger, Phys. Rev. Lett. **70**, 1343 (1993).

[11] P. C. Ivanov, A. Bunde, L. A. N. Amaral, S. Havlin, J. Fritsch-Yelle, R. M. Baevsky, H. E. Stanley, and A. L. Goldberger, Europhys. Lett. **48**, 594 (1999).

[12] A. Bunde, S. Havlin, J. W. Kantelhardt, T. Penzel, J.-H. Peter, and K. Voigt, Phys. Rev. Lett. **85**, 3736 (2000).

[13] T. Lux, Appl. Econ. Lett. **3**, 701 (1996).

[14] Z. Ding, C. W. J. Granger, and R. F. Engle, J. Empirical Finance **1**, 83 (1993).

[15] S. Lennartz, V. N. Livina, A. Bunde, and S. Havlin, EPL **81**, 69001 (2008).

[16] V. N. Livina, S. Havlin, and A. Bunde, Phys. Rev. Lett. **95**, 208501 (2005).

[17] A. Bunde, J. F. Eichner, J. W. Kantelhardt, and S. Havlin, Phys. Rev. Lett. **94**, 048701 (2005).

[18] J. W. Kantelhardt, E. Koscielny-Bunde, H. H. A. Rego, S. Havlin, and A. Bunde, Physica A **295**, 441 (2001).

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

[20] J. F. Eichner, J. W. Kantelhardt, A. Bunde, and S. Havlin, Phys. Rev. E **75**, 011128 (2007).

[21] A. Arneodo, B. Audit, N. Decoster, J.-F. Muzy, and C. Vaillant, in *The Science of Disasters: Climate Disruptions, Heart Attacks, and Market Crashes*, edited by A. Bunde, J. Kropp, and H.-J. Schellnhuber (Springer, Berlin, 2002), p. 27.

[22] In this technique, one first generates uncorrelated Gaussian data g_i, $i=1,\ldots,N$ and determines its Fourier transform $\tilde{g}_k = (1/\sqrt{N})\Sigma_{l=1}^{N} g_l e^{-2\pi i k l/N}$ with $N/2 < k \leq N/2$ and $\tilde{g}_{-k} = \tilde{g}_k^*$. One defines $\tilde{x}_k = \tilde{g}_k \cdot (|k|/N)^{-(1-\gamma)/2}$ ($\tilde{x}_0$ is set to 0). The Fourier transform of $\tilde{x}_k$, $x_l = (1/\sqrt{N})\Sigma_{k=-(N/2-1)}^{N/2} \tilde{x}_k e^{2\pi i k l/N}$, is the desired long-term correlated record.