Statistics of return intervals in long-term correlated records

Jan F. Eichner,[1] Jan W. Kantelhardt,[1,2] Armin Bunde,[1] and Shlomo Havlin[3]
[1]*Institut für Theoretische Physik III, Justus-Liebig-Universität Giessen, 35392 Giessen, Germany*
[2]*Fachbereich Physik und Zentrum für Computational Nanoscience, Martin-Luther-Universität Halle-Wittenberg, 06099 Halle (Saale), Germany*
[3]*Minerva Center and Department of Physics, Bar-Ilan University, Ramat-Gan 52900, Israel*
(Received 27 June 2006; published 26 January 2007)

We consider long-term correlated data with several distribution densities (Gaussian, exponential, power law, and log normal) and various correlation exponents γ ($0<\gamma<1$), and study the statistics of the return intervals r_j between events above some threshold q. We show that irrespective of the distribution, the return intervals are long-term correlated in the same way as the original record, but with additional uncorrelated noise. Due to this noise, the correlations are difficult to observe by the detrended fluctuation analysis (which exhibits a crossover behavior) but show up very clearly in the autocorrelation function. The distribution $P_q(r)$ of the return intervals is characterized at large scales by a stretched exponential with exponent γ, and at short scales by a power law with exponent $\gamma-1$. We discuss in detail the occurrence of finite-size effects for large threshold values for all considered distributions. We show that finite-size effects are most pronounced in exponentially distributed data sets where they can even mask the stretched exponential behavior in records of up to 10^6 data points. Finally, in order to quantify the clustering of extreme events due to the long-term correlations in the return intervals, we study the conditional distribution function and the related moments. We find that they show pronounced memory effects, irrespective of the distribution of the original data.

DOI: 10.1103/PhysRevE.75.011128 PACS number(s): 02.50.−r, 89.75.Da, 05.40.−a

I. INTRODUCTION

The statistics of return intervals between well defined extremal events is a powerful tool to characterize the temporal scaling properties of observed time series and to derive quantities for the estimation of the risk for hazardous events like floods, very high temperatures, or earthquakes. For Gaussian distributed data it has been shown recently that the long-term correlations [1,2], inherent, e.g., in river flows and temperature records, represent a natural mechanism for the clustering of the hazardous events [3,4]. The distribution density of the return intervals strongly depends on the history, and can be well approximated by a stretched exponential [3,5,6]. In addition, the mean residual time to the next extreme event increases with the elapsed time and depends strongly on the previous return intervals [3].

In this paper we extend these results in several directions. We focus on long-term correlated signals with different distributions of the values (Gaussian, exponential, power-law, and log-normal distributions) and show numerically that (i) the stretched exponential decay of the return interval distribution density occurring for large return intervals is complemented by a power-law decay regime for small return intervals, (ii) both regimes are hardly affected by the distribution density of the original data, but discreteness corrections as well as finite-size corrections have to be taken into account for very short and long return intervals, respectively, (iii) the conditional return interval distribution densities, i.e., the distribution densities of those return intervals that follow immediately a preceding return interval of a given size in the series of the return intervals, are hardly affected by the distribution densities of the original data, (iv) the long-term correlation properties of the return interval series can be estimated more reliably by a classical autocorrelation analysis rather than by the more sophisticated approaches based on random walk theory (e.g., detrended fluctuation analysis), and (v) in practical applications the calculation of double-conditional return periods, i.e., the average return interval following a preceding and prepreceding return interval of a given size, can yield an improved risk estimation for the occurrence of hazardous events.

There is growing evidence that many natural records exhibit long-term persistence [1,2]. Prominent examples include hydrological data [7–10], meteorological and climatological records [11–15], turbulence data [16,17], as well as physiological records [18–20], and DNA sequences [21,22]. Long-term correlations have also been found in the volatility of economic records [23]. In long-term persistent records (x_i), $i=1,\ldots,N$ with mean $\bar{x}$ and standard deviation σ_x the autocorrelation function decays by a power law

$$C_x(s)=\frac{1}{\sigma_x^2}\langle(x_i-\bar{x})(x_{i+s}-\bar{x})\rangle\sim s^{-\gamma}, \tag{1}$$

where $\langle\cdots\rangle$ denotes the average over $i=1,\ldots,N-s$, and γ denotes the correlation exponent, $0<\gamma<1$. Such correlations are called "long term" since the mean correlation time $T=\int_0^\infty C_x(s)ds$ diverges for infinitely long series, i.e., in the limit $N\to\infty$. Power-law long-term correlations according to Eq. (1) correspond to a power spectrum $S(\omega)\sim\omega^{-\beta}$ with $\beta=1-\gamma$ according to the Wiener-Khintchin theorem.

In addition to long-term correlations, natural records are characterized by the distribution density $D(x)$ of their values x_i. In previous work we focussed on Gaussian distributed data [3,5,24] although broader distributions are also common in natural records. For example, if the underlying process is multiplicative rather than additive, one should rather expect a log-normal distribution density. Here we study in detail how

 ©2007 The American Physical Society

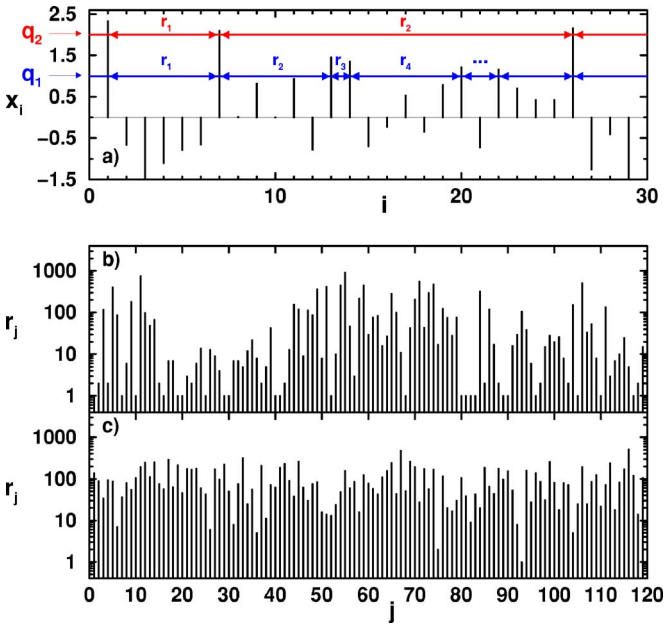

FIG. 1. (Color online) Return intervals: (a) Illustration of the definition of return intervals r_j between events x_i above given thresholds $q_1=1$ and $q_2=2$. The smallest possible r value is $r=1$, which occurs if $x_i>q$ and $x_{i+1}>q$. (b) Sequence of return intervals r_j for long-term correlated data x_i with $\gamma=0.4$ for a fixed threshold $q=2.327$ so that the average r value is $R_q=100$. Note the logarithmic scale of the r-axis. (c) Sequence of return intervals for uncorrelated data [shuffled x_i of (b)]. In (b) more epochs with r values of small and large size appear compared to (c).

different long-term correlation behavior and different distribution densities $D(x)$ affect the statistics of return intervals between extreme events.

Applications of these results to boundary layer wind fields, market volatilities, and climatological and hydrological data can be found in Refs. [3,24–26]. Since the analysis of long daily hydroclimatological data is complicated by the presence of seasonal trends that cannot be eliminated completely, the studies have focused on annual data.

The paper is organized as follows. In Sec. II we discuss the different scaling regimes and corrections that govern the (unconditional) means and distribution densities of return intervals. Section III is devoted to the long-term correlation properties of the return intervals and to the critical discussion of the results based on different methods. Section IV deals with conditional return interval distribution densities and conditional mean return intervals, and in Sec. V we briefly conclude.

II. MEANS AND DISTRIBUTIONS OF THE RETURN INTERVALS

For describing the recurrence of rare events exceeding a certain threshold q, defined in units of the standard deviations of the original distribution $D(x)$, we investigate the statistics of the return intervals r between these events as illustrated in Fig. 1(a). Specifically, a return interval r occurs if $x_i>q$ and $x_{i+r}>q$ while $x_j\leq q$ for $i<j<i+r$. Figure 1(b) shows a section of the sequence of the return intervals for Gaussian long-term correlated data for a threshold q (quantile) chosen such that the mean return interval R_q ("return period") is approximately 100. Figure 1(c) shows the same section, but the original data was shuffled before, destroying the correlations. One can see that there are more large r values and many more short r values in Fig. 1(b) compared to the uncorrelated case in Fig. 1(c), although the mean return interval R_q is the same. The long and short return intervals in (b) appear in clusters [3], creating epochs of cumulated extreme events caused by the short r values, and also long epochs of few extreme events caused by the long r values. In the following we show, how the return period R_q, the standard deviation σ_r, and the distribution density $P_q(r)$ of the return intervals are affected by the presence of long-term correlations as well as by different distribution densities $D(x)$ of the data.

In our numerical procedure, we generated sequences of random numbers (x_i) of length $N=2^{21}$ with either Gaussian, exponential, power-law, or log-normal distribution with unit variance. The corresponding distribution densities $D(x)$ are given by

$$D_{\text{Gauss}}(x)=\frac{1}{\sqrt{2\pi\sigma}}\exp(-x^2/2\sigma), \tag{2}$$

$$D_{\text{exp}}(x)=\frac{1}{x_0}\exp(-x/x_0), \tag{3}$$

$$D_{\text{power}}(x)=(\delta-1)x^{-\delta}, \tag{4}$$

$$D_{\text{log-norm}}(x)=\frac{1}{\sqrt{2\pi x}}\exp[-(\ln x+\mu)^2/2]. \tag{5}$$

Here, we choose, without loss of generality, $\sigma=1$, $x_0=1$, and $\mu=0.763$ in Eqs. (2), (3), and (5) and select $\delta=5.5$ in Eq. (4).

For each distribution density $D(x)$ we generated 150 data sets using 1000 iterations, restoring the desired power spectrum by Fourier filtering and restoring the desired distribution by rank-ordered replacement of the values in each iteration until convergence is achieved [27]. A full explanation of this iterative procedure is given in the appendix. For the Gaussian data one iteration is sufficient since Fourier filtering preserves the Gaussian distribution. We tested the quality of the long-term correlations of the data with detrended fluctuation analysis (DFA) [19,28] and autocorrelation function analysis.

A. Mean return interval and standard deviation

First we consider the mean return interval R_q. For a given threshold q, there exist N_q return intervals r_j, $j=1,2,\ldots,N_q$, which satisfy the sum rule $\Sigma_{j=1}^{N_q}r_j=N$ for periodic boundary conditions. When the data are shuffled, the long-term correlations are destroyed, but the sum rule still applies with the same value of N_q. Accordingly, for both long-term correlated and uncorrelated records, $R_q\equiv(1/N_q)\Sigma_{j=1}^{N_q}r_j$ is simply $R_q=N/N_q$, i.e., the mean return interval is not affected by the

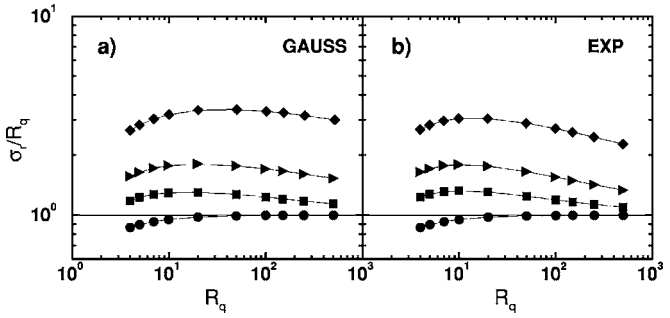

FIG. 2. Standard deviation of r values divided by R_q as a function of R_q for (a) Gaussian and (b) exponential uncorrelated data (circles) and long-term correlated data (squares $\gamma=0.7$, triangles $\gamma=0.4$, and diamonds $\gamma=0.2$). The $\sigma_r(R_q)$ dependences are representative also for data with other distribution densities $D(x)$.

long-term correlations. This statement can also be considered as the time series analogous of Kac's Lemma [6,29]. Hence, R_q can be obtained directly from the tail of the (normalized) distribution density $D(x)$ of the values x_i via

$$\frac{1}{R_q} = \frac{N_q}{N} \cong \int_q^\infty D(x)dx. \tag{6}$$

The larger q is, the larger the average return interval R_q will be. Accordingly, there is a one-by-one correspondence between q and R_q, which depends only on the distribution density $D(x)$ but not on the correlations.

However, there is no one-by-one correspondence between q and the variance $\sigma_r^2 \equiv (1/N_q)\Sigma_{j=1}^{N_q}[r_j - R_q]^2$ of the return intervals. Figure 2 shows (in units of R_q) the standard deviation σ_r for uncorrelated data and long-term correlated data with three correlation exponents. Due to the appearance of larger return intervals with increasing correlations (i.e., decreasing correlation exponent γ), the standard deviation also increases. The decrease of σ_r for small and large R_q is due to discretization effects and finite-size effects, respectively, which will be discussed in the following sections.

B. Stretched exponential and finite-size effects for large return intervals

For uncorrelated data ("white noise"), the return intervals are also uncorrelated and (according to the Poisson statistics) exponentially distributed with $P_q(r) = \frac{1}{R_q}\exp(-r/R_q)$, where R_q is the mean return interval for the given threshold q (see, e.g., Ref. [30]). When introducing long-term correlations in Gaussian data the shape of $P_q(r)$ for large values of r, $r > R_q$, is changed to a stretched exponential [3,5,6,24] (see also Ref. [31])

$$P_q(r) \cong \frac{a_\gamma}{R_q}\exp[-b_\gamma(r/R_q)^\gamma], \tag{7}$$

where the exponent γ is the correlation exponent, and the parameters a_γ and b_γ are independent of q. Their dependence upon γ can be determined from the two normalization conditions that must hold for the (discrete) distribution density $P_q(r)$:

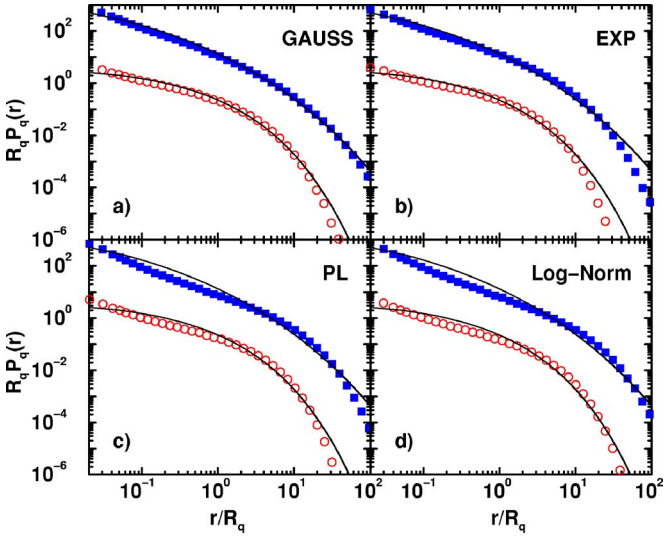

FIG. 3. (Color online) Normalized distribution density function $R_qP_q(r)$ of r values with $R_q=100$ as a function of r/R_q for long-term correlated data with $\gamma=0.4$ (open symbols) and $\gamma=0.2$ (filled symbols; we multiplied this data by a factor 100 to avoid overlapping curves). In (a) the original data was Gaussian distributed, in (b) exponentially distributed, in (c) power-law distributed with $\delta=5.5$, and in (d) log-normally distributed. All four figures follow quite well stretched exponential curves [solid lines, Eq. (7)] over several decades. For small r/R_q values a power-law regime seems to dominate, while on large scales deviations from the stretched exponential behavior are due to finite-size effects.

$$\sum_{r=1}^{\infty} P_q(r) = 1, \quad \sum_{r=1}^{\infty} rP_q(r) = R_q. \tag{8}$$

If we replace the sums for the discrete values of the return interval r by integrals and disregard the deviations from Eq. (7) for small values of r (see Sec. II C), we can solve these two equations for a_γ and b_γ, obtaining

$$a_\gamma = \frac{\gamma\Gamma(2/\gamma)}{\Gamma^2(1/\gamma)} \quad \text{and} \quad b_\gamma = \frac{\Gamma(2/\gamma)}{\Gamma(1/\gamma)}, \tag{9}$$

where $\Gamma(x)$ is the Gamma function. Hence, a_γ and b_γ are not free parameters but rather determined only by the correlation exponent γ (see also Ref. [6]). Hence, if γ is determined independently by correlation analysis, we can obtain a data collapse of all curves for different values of q by plotting $R_qP_q(r)$ versus r/R_q according to Eq. (7) [3], since the dependence of R_q on q is given by Eq. (6). However, we have to note that Eq. (7) does not hold for small values of r (see Sec. II C), causing some deviations in the values of parameters a_γ and b_γ and the data collapse.

Figure 3 shows the rescaled distribution density function of the return intervals for the four different types of distributions of the original data [Gaussian, exponential, power-law, and log-normal according to Eqs. (2)–(5)] with correlation exponents $\gamma=0.2$ and 0.4 in a double-logarithmic scale. The shape of the corresponding stretched exponentials (7) is also

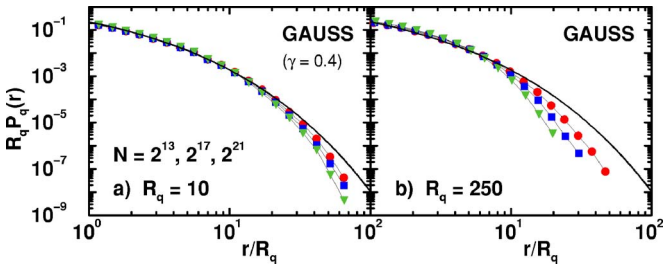

FIG. 4. (Color online) Finite-size effects on $P_q(r)$. Deviations from the stretched exponential curve for large r/R_q and (a) small $R_q=10$ as well as (b) large $R_q=250$ in data with $\gamma=0.4$ for three lengths $N=2^{13}$ (triangles), 2^{17} (squares), and 2^{21} (circles) indicate finite-size effects. For large values of q and small lengths N the occurrence of very large r is actually impossible ($r \leq N-N/R_q$). The finite-size effects for Gaussian data are representative for all four distribution densities $D(x)$ we considered.

plotted. The agreement is best for the Gaussian data, but also for the other distributions the stretched exponential fit is a good approximation over several decades.

The deviations from the stretched exponential at the right tail of the curves in Fig. 3 (i.e., for $r \gg R_q$) are due to finite-size effects. Figures 4(a) and 4(b) show that finite-size effects seem to cause a decrease of $P_q(r)$ for $r \gg R_q$ compared with the stretched exponential Eq. (7), unless we consider very long series. The series analyzed here have the lengths $N=2^{13}$ (triangles), 2^{17} (squares), and 2^{21} (circles). The plots suggest a convergence in the asymptotic regime (large r) towards the stretched exponential curve (solid line). For small values of R_q the deviations are weaker since the number of return intervals, $N_q=N/R_q$, is much larger and the statistics is better.

C. Power-law regime and discretization effects for small return intervals

Next we consider the deviations from the stretched exponential form at small values of r, $r<R_q$ (see Fig. 3). Figure 5 shows the left parts of $P_q(r)$ for three return periods R_q and again the four distribution densities $D(x)$. The figure reveals the occurrence of an additional scaling regime for $0.01<r/R_q<1$ for Gaussian and exponentially distributed data, see (a) and (b), and for $0.1<r/R_q<1$ for power-law and log-normally distributed data, see (c) and (d). The scaling behavior with R_q still holds in this new regime, such that data for different values of q collapse onto each other as long as $r \gg 1$. The scaling behavior in this regime might be characterized by a power law, giving rise to the two-branched ansatz

$$R_q P_q(r) \sim \begin{cases} \left(\frac{r}{R_q}\right)^{\gamma'-1} & \text{for } 1 \ll r \leq R_q, \\ \exp\left[-\left(b\frac{r}{R_q}\right)^{\gamma}\right] & \text{for } R_q < r \ll N \end{cases} \tag{10}$$

replacing Eq. (7). For Gaussian and exponential distribution densities $D(x)$, $\gamma' \approx \gamma$ seems to be consistent with the data.

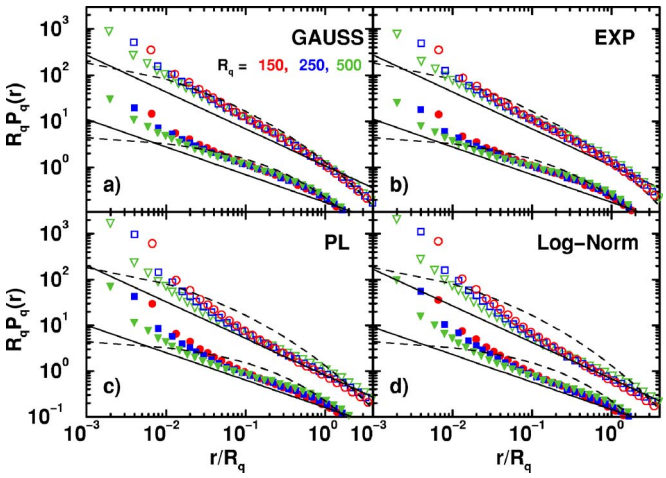

FIG. 5. (Color online) Power-law decay of $P_q(r)$ for $r<R_q$ for $R_q=150$ (circles), 250 (squares), and 500 (triangles) for (a) Gaussian, (b) exponentially, (c) power-law, and (d) log normally distributed data with $\gamma=0.4$ (solid symbols) and $\gamma=0.2$ (open symbols, shifted upwards by a factor ten to avoid overlapping curves). The dashed line is the stretched exponential Eq. (7), the solid line is a power law with slope $\gamma-1$ as a guide to the eye. The power-law effect in (a) and (b) appears in the scale range $0.01<r/R_q<1$ and in (c) and (d) in $0.1<r/R_q<1$. For even smaller r/R_q values ($r=1,2,3,\ldots$) all four figures show an upward trend which is due to the discreteness of the r values.

However, we cannot fully exclude the possibility that γ' might depend slightly on the quantile q, in particular for power-law and log-normal distribution densities $D(x)$. The normalization factors and the parameter b cannot be calculated exactly for Eq. (10), which is a drawback of the two-branched distribution density.

Figures 3–5 show that the behavior of $P_q(r)$ as described by Eq. (10) becomes visible only for very long data sets. If short and, e.g., exponentially distributed data are studied, the stretched exponential regime shrinks and an asymptotic pure exponential decay of $P_q(r)$ can be observed. In this case, $P_q(r)$ displays a combination of (i) discretization effects for very small r values, (ii) power-law behavior for intermediate r values, and (iii) an exponential decay for large r values, resembling the return interval distribution obtained for seismic activity [32–34].

For very small return intervals (r close to 1) the continuous distribution density Eq. (10) has to be replaced by a discrete distribution. Figure 5 clearly shows the deviations from the power-law regime because of the discreteness of the r values. The first few points move upwards ($r=1$ and 2, maybe up to $r=5$), because they accumulate probability from (impossible) smaller noninteger intervals. Hence, the data points for r close to 1 do not obey the scaling of the distribution density with R_q, and no data collapse can be achieved for them. For the power-law distributed data and the log-normal distributed data, Figs. 5(c) and 5(d), the discretization effects even seem to suppress the full development of the power-law scaling regime.

Return intervals $r=1$ are of particular relevance when answering the question if two consecutive events x_i and x_{i+1} will surpass the threshold q. Figures 6(a) and 6(b) show the

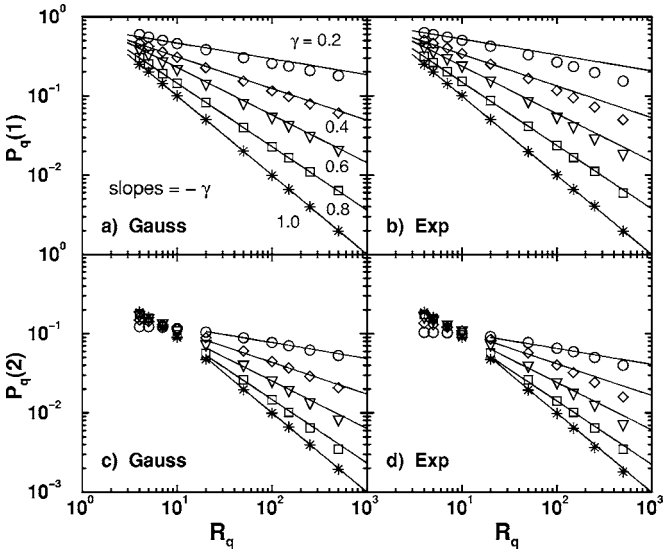

FIG. 6. Dependence of $P_q(1)$ and $P_q(2)$ on γ for various quantiles q. The distribution of the return intervals of size $r=1$ is plotted versus R_q for (a) Gaussian and (b) exponentially distributed data with correlation exponents $\gamma=0.2$ (circles), 0.4 (diamonds), 0.6 (triangles), 0.8 (squares) as well as uncorrelated data (stars). The stronger the correlations, i.e., the smaller γ, the higher the probability that two consecutive events x_i and x_{i+1} exceed the given threshold q. The straight lines indicate power laws with exponents $-\gamma$, suggesting the discretization effects in Fig. 5 to be independent of q or R_q. Note that the larger R_q, the stronger the correlation effect. Figures (c) and (d) show the same for return intervals of size $r=2$.

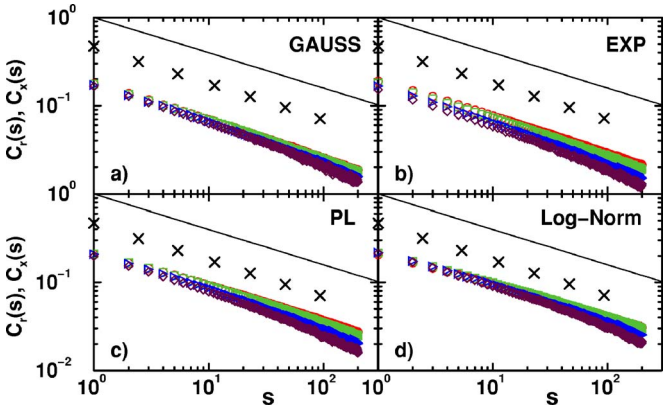

FIG. 7. (Color online) Autocorrelation functions $C_x(s)$ of the original data ($\times$) and $C_r(s)$ of the return interval series for (a) Gaussian, (b) exponentially, (c) power-law, and (d) log-normally distributed data with $\gamma=0.4$ for $R_q=$ 10 (circles), 20 (squares), 50 (triangles), and 100 (diamonds). All curves show a power-law behavior indicating long-term correlations in the sequence of return intervals with the same correlation exponent γ (straight lines). Curves for large R_q show weaker correlations at large scales s due to finite-size effects. All results were averaged over 150 configurations of original data with length $N=2^{21}$.

distribution of the $r=1$ values for Gaussian and exponentially distributed data as a function of the return period R_q (i.e., of the threshold q). In uncorrelated data the probability to have another extreme event directly after any extreme event must be $1/R_q$ (stars in Fig. 6), because each data point x_i is an extreme event with probability $1/R_q$ and there are no correlations. In long-term correlated data the probability to find another extreme event directly after an extreme event will be higher due to the persistence of the data. The figure shows, how this probability changes with increasing correlations (i.e., decreasing γ). Except for small values of γ the shape of the curves seems to be a power law of the form $P_q(1)\approx R_q^{-\gamma}$, in agreement with Eq. (10) and the assumption $\gamma'=\gamma$ [35]. The larger R_q is chosen, the larger the difference between $P_q(1)$ in correlated and uncorrelated data. The distribution $P_q(2)$ of the $r=2$ values [Figs. 6(c) and 6(d)] shows a similar power-law for large values of R_q, but deviates for small R_q.

III. LONG-TERM CORRELATIONS OF THE RETURN INTERVALS

The form of the distribution density $P_q(r)$ of return intervals r between extreme events in long-term correlated data indicates that very short and very long return intervals are more frequent than for uncorrelated data. However, $P_q(r)$ does not quantify, if the return intervals themselves are arranged in a correlated fashion, and if clustering of rare events may be induced by long-term correlations. In our previous work [3,24] we reported that (i) long-term correlations in a Gaussian time series induce long-term correlations in the sequence of return intervals and (ii) that both correlation functions are characterized by the same correlation exponent γ. We showed that this leads to a clustering of extreme events, an effect that also can be seen in long climate records.

Here we present further support for our result (ii) showing that it is independent of the distribution density $D(x)$ of the data. We also compare the results of classical autocorrelation analysis and detrended fluctuation analysis (DFA) [19,28] and explain why in this case the classical autocorrelation analysis is more suitable than the more sophisticated DFA (see also Refs. [5,6]).

A. Autocorrelation analysis

In order to determine the autocorrelation behavior of the return interval series (r_j), $j=1,2,\ldots,N_q$, for a given quantile q we calculate the autocorrelation function of the return intervals [see Eq. (1) where x_i is replaced by r_i]. Figure 7 shows $C_r(s)$ and $C_x(s)$ for data characterized by the four distribution densities $D(x)$ and four mean return periods R_q. One can see that for each distribution, the data approximately collapse to a single line, exhibiting the same slope as the original data. This shows that the return intervals are also long-term correlated, with the same value of γ as the original data. There is, however, one important difference in the correlation behavior: For the return intervals, the autocorrelation function $C_r(s)$ is significantly below the autocorrelation function $C_x(s)$ of the original data (also shown in Fig. 7) by a factor between 2 and 3, depending on the distribution. Accordingly, there is additional white noise in the return interval sequences that only weakly depends on the return period R_q.

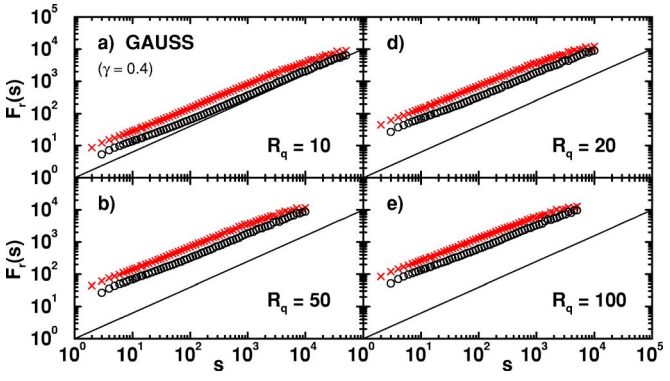

FIG. 8. (Color online) Fluctuation analysis (FA) and detrended fluctuation analysis (DFA) of the series of return intervals for (a) $R_q=10$, (b) 20, (c) 50, and (d) 100. The FA (×) and the DFA1 (○) fluctuation functions $F_r(s)$ of the sequence of return intervals are shown for Gaussian long-term correlated data with $\gamma=0.4$. The asymptotic slope α is related to γ via $\alpha=1-\gamma/2$. The solid lines have the slope $\alpha=0.8$ ($\gamma=0.4$). On small scales a slope close to $\alpha=0.5$ is observed due to the large uncorrelated component in the return interval series, while the relative prefactor of the long-term correlated component is decreasing with increasing R_q.

We believe that this uncorrelated component is a consequence of the way the return intervals are constructed. Tiny changes in the threshold q can lead to large changes of several return intervals in the records. Thus, small white noise in the original data leads to larger noise in the return intervals, and this causes the significant additional random component. There is no crossover in the scaling behavior of the autocorrelation function except for finite-size effects.

B. Detrended fluctuation analysis

Alternatively to the autocorrelation function or autocovariance function one can employ fluctuation analysis (FA) [13] or detrended fluctuation analysis (DFA) [19,28,36,37]. In DFA, one considers the cumulated sum $Y_i=\Sigma_{j=1}^{i}y_j$ of standardized records (y_j) (i.e., mean 0 and variance 1) and studies, in time windows of length s, the mean fluctuation $F(s)$ of Y_i around a best fit (e.g., quadratic fit for DFA2). For long-term correlated data, $F(s)$ scales as $F(s)\sim s^{\alpha}$, with the relation $\alpha=1-\gamma/2$. Figure 8 shows the fluctuation functions $F_r(s)$, determined with FA and DFA, for the return interval series (r_j) in Gaussian long-term correlated data with $\gamma=0.4$ and four values of R_q ($R_q=10$, 20, 50, and 100). On large scales s both curves approach the slope $\alpha=0.8$ corresponding to the correlations of the original data (solid lines).

Obviously, on smaller scales the observed slopes differ significantly from $\alpha=0.8$ and seem to be closer to $\alpha=0.5$ particularly for large R_q, suggesting uncorrelated behavior on short time scales s. Linear fitting of the fluctuation functions in log-log plots can thus lead to reduced values of α that might suggest an increasing loss of correlations for increasing thresholds q. However, it has been shown that a process consisting of a long-term correlated component and an uncorrelated noise component yields a fluctuation function of the form [39]

$$F^2(s)\sim As+Bs^{2\alpha}, \tag{11}$$

where A and B are the weighting prefactors of the uncorrelated and correlated components. The scale $s_{\times}\sim(A/B)^{1/(2\alpha-1)}$, where the contribution of both terms is equal, defines a crossover length between the two scaling regimes. For $s\ll s_{\times}$ the first term dominates, while for $s\gg s_{\times}$ the second one dominates. As a consequence, the asymptotic scaling behavior $F(s)\sim s^{\alpha}$ can be observed only at large values of s (as seen in Fig. 8), in particular when $B\gg A$.

We suggest that this crossover in the scaling behavior of the FA and DFA fluctuation functions is the reason for the apparent loss of two-point correlations with increasing threshold value q observed in Refs. [5,6]. This problem is even more severe, if multifractal features of a record are being investigated by multifractal detrended fluctuation analysis (MF-DFA) [6,38], since the superposition rule Eq. (11) no longer applies for MF-DFA, and multifractal properties of two superimposed processes can become very complicated. In our opinion, previous DFA results on market-volatility records [26] also suffer from the same drawback, which yielded lower α-values for the return intervals than for the original records.

IV. CONDITIONAL RETURN INTERVALS

The long-term correlations in the sequence of the return interval series studied in the previous section lead (i) to clustering of rare events and (ii) to long epochs of very few extreme events—both because of clumping of both short and long return intervals. Hence, the probability of finding a certain return interval r depends on the history and in particular on the value of the preceding interval r_0. In the following we show that this effect is nearly independent of the distribution density $D(x)$ of the original data. We also study (double-) conditional return periods, which are the mean return intervals of those return intervals that immediately follow after preceding and preprecedding return intervals r_0 and r_{-1} of fixed values, and which might further improve predictions and risk estimations.

A. Conditional return interval distributions

In order to study the history effects on the distribution density $P_q(r)$ of the return intervals r we first consider the conditional distribution density $P_q(r|r_0)$, defined as the distribution density of all those return intervals that directly follow a given r_0 value in the sequence of return intervals. For uncorrelated data $P_q(r|r_0)=P_q(r)$. Figure 9 displays the ratio between $P_q(r|r_0)$ and $P_q(r)$ for Gaussian and exponentially distributed long-term correlated data with six R_q values (20, 50, 100, 150, 250, and 500), for $r_0/R_q\simeq 1/4$, 1, and 4, as a function of r/R_q. For fixed r_0/R_q all data points collapse onto single curves, independent of q and independent of the distribution of the original data. We thus find scaling behavior of the conditional distribution density function

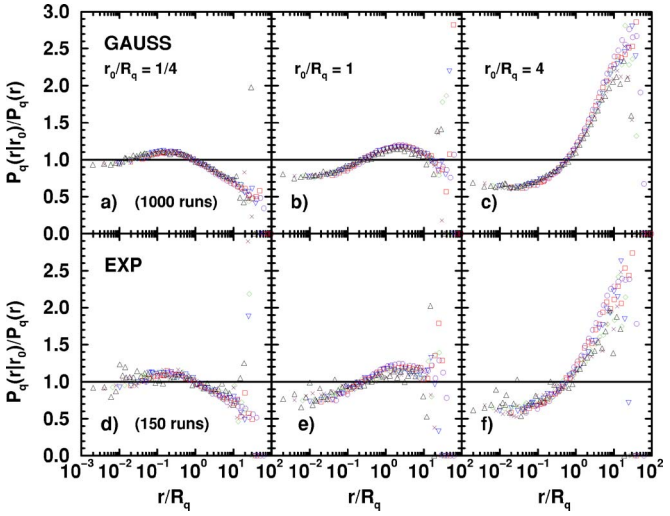

FIG. 9. (Color online) Conditional distribution density $P_q(r|r_0)$ divided by $P_q(r)$ for three different conditions $r_0/R_q=1/4$ (a), 1 (b), and 4 (c) for long-term correlated ($\gamma=0.4$) Gaussian distributed original data. The maximum of the ratio moves from small r/R_q to large r/R_q with increasing the condition r_0/R_q. The cutoff on the right-hand side is due to finite set length and therefore limited maximum r values. The data collapse consists of six curves for quantiles with $R_q=20$ (circles), 50 (squares), 100 (triangles down), 150 (diamonds), 250 (crosses), and 500 (triangles up). Parts (d) to (f) show the same for exponentially distributed data. The collapse seems noisier because the curves are averaged over only 150 configurations compared to 1000 configurations in (a) to (c). For uncorrelated data, $P_q(r|r_0)=P_q(r)$ exactly, as indicated by the straight line at the ratio one.

$$P_q(r|r_0) = \frac{1}{R_q} f_{r_0/R_q}(r/R_q) \tag{12}$$

as in Eq. (10). The long-term correlations in the sequence of r values cause a culmination of small r values for small r_0 and large r values for large r_0. The conditions $r_0=R_q/4$ and $r_0=R_q$ yield maxima at comparable r values, $r\approx 0.2R_q$ and $r\approx 2R_q$, respectively, see Figs. 9(a), 9(b), 9(d), and 9(e). For $r_0=4R_q$ a strong enhancement of large return intervals $r\approx 10R_q$ is observed, see Figs. 9(c) and 9(f). When the series of r values is shuffled, i.e., all correlations are destroyed, $P_q(r|r_0)$ and $P_q(r)$ are identical, but still show the stretched exponential shape. For uncorrelated original data $P_q(r)$ and $P_q(r|r_0)$ both follow a simple exponential decay.

B. Conditional mean return intervals

In contrast to the mean return interval R_q, the conditional mean return interval $R_q(r_0)=\Sigma_{r=1}^{\infty} rP_q(r|r_0)$, i.e., the average return interval of those r values that follow directly an interval of size r_0, clearly exhibits correlation effects. Figure 10 shows $R_q(r_0)$ in units of R_q as a function of r_0/R_q for four values of R_q (5, 10, 50, and 250), $\gamma=0.4$, and the four distribution densities $D(x)$ listed in Eqs. (2)–(5). The correlation effect becomes apparent: after small values of r_0/R_q the next expected return interval $R_q(r_0)$ is smaller than R_q, and after large r_0/R_q, $R_q(r_0)$ is much larger than R_q. Although the

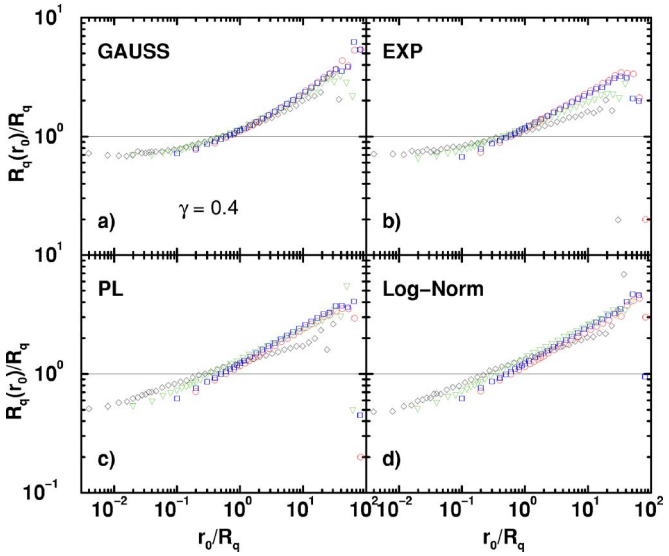

FIG. 10. (Color online) Conditional return periods $R_q(r_0)$ in units of R_q versus the condition r_0/R_q for R_q = 5 (circles), 10 (squares), 50 (triangles), and 250 (diamonds) for long-term correlated data with $\gamma=0.4$. All figures clearly display the memory effect in form of increasing $R_q(r_0)$ with increasing r_0, caused by the long-term correlations in the original data. In uncorrelated data, $R_q(r_0)=R_q$, as indicated by the horizontal lines at ratio one.

shapes of the (more or less) collapsing curves depend slightly on the original distribution, the tendency is the same for all four original distribution densities. We like to note again that finite-size effects, violating the scaling behavior, are most pronounced for the exponential distribution.

Due to the tight conditions regarding r_0, the conditional mean return interval $R_q(r_0)$ requires very large statistics and is thus not suitable for studies of real recordings. The quantity can be improved by integrating over two ranges of r_0 values, e.g., r_0 larger or smaller than the return period R_q, resulting in only two conditions. Therefore we define R_q^+ and R_q^- as the average return intervals that follow r_0 values either larger (+) or smaller (−) than R_q. Figure 11 shows R_q^+ and R_q^- (filled symbols) in units of R_q as a function of the correlation exponent γ of the data for all four distributions. For uncorrelated data (results shown at $\gamma=1$), R_q^+ and R_q^- coincide with R_q. The smaller γ, the stronger the correlations, and the more pronounced is the difference between R_q^+ and R_q^-.

Figure 11 also shows the average conditional return intervals R_q^{++} and R_q^{--} of those r-values that directly follow two return intervals, r_0 and r_{-1}, both larger (++) or smaller (−−) than R_q. As expected, the correlation effect is even stronger in R_q^{++} and R_q^{--} but the quantities require more statistics than R_q^+ and R_q^-. All curves in Fig. 11 look very similar and suggest that an effect of the shape of the distribution of the original data is minor, when the data is long-term correlated.

V. CONCLUSION

In this paper we have studied the statistics of return intervals between extreme events in long-term persistent data

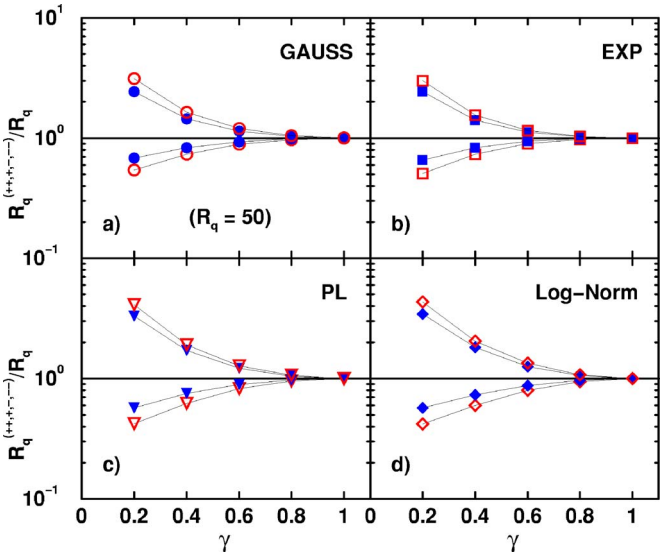

FIG. 11. (Color online) Conditional return periods for long-term correlated data with four values of γ and uncorrelated data (shown at $\gamma=1$); R_q is fixed at $R_q=50$. Figure (a) shows the single-conditional and the double-conditional return periods ($R_q^{+/-}$ and $R_q^{++/--}$) in units of R_q for Gaussian data. R_q^+ (filled symbols, upper curve) and R_q^- (filled symbols, lower curve) are the single-conditional return periods, where the predecessor r_0 is larger or smaller than R_q. R_q^{++} (open symbols, upper curve) and R_q^{--} (open symbols, lower curve) are the double-conditional return periods where the preceding and the prepreceding return intervals are both larger or both smaller than R_q [$R_q^{++} = R_q(r|r_0>R_q|r_{-1}>R_q)$, $R_q^{--} = R_q(r|r_0<R_q|r_{-1}<R_q)$]. The splitting of the curves is a consequence of the long-term correlations in the sequence of r values. In the uncorrelated case all four conditional return periods degenerate to R_q. Figure (b) shows the same for exponentially distributed data, (c) for power-law distributed data, and (d) for log-normally distributed data. All points were averaged over 20 configurations with roughly 40 000 r values.

with Gaussian, exponential, power-law, and log-normal distribution densities. We have shown that mainly the correlations rather than the distributions affect the return interval statistics, in particular the distribution density of return intervals, the conditional distribution density of return intervals, the conditional mean return intervals, and the correlation properties of the return interval series. The stretched exponential decay of the return interval distribution density for long return intervals is complemented by a power-law decay for small return intervals, which will dominate the behavior in short observational data. Still, the scaling behavior with the mean return interval holds in both regimes. In addition, discreteness corrections as well as finite-size corrections have to be taken into account for very short and long return intervals.

ACKNOWLEDGMENT

We would like to acknowledge financial support from the Bundesministerium für Bildung und Forschung and the Israel Science Foundation.

APPENDIX: DATA GENERATION

The long-term correlations in the random numbers were introduced by the Fourier-filtering technique, see, e.g., Ref. [40]. However, for the non-Gaussian distributions, (3)–(5), the shape of $D(x)$ is not preserved by Fourier filtering. In these cases we applied an iterative algorithm introduced by Schreiber and Schmitz [27]. The algorithm consists of the following steps: First one creates a Gaussian distributed long-term correlated data set with the desired correlation exponent γ by standard Fourier filtering [40]. The power spectrum $S_G(\omega)=F_G(\omega)F_G^*(\omega)$ of this Gaussian data set is considered as reference spectrum [where ω denotes the frequency in Fourier space and the $F_G(\omega)$ are the complex Fourier coefficients]. Next one creates an uncorrelated sequence of random numbers (x_i^{ref}), following a desired distribution, e.g., exponential distribution (3). The (complex) Fourier transform $F(\omega)$ of the (x_i^{ref}) is now divided by its absolute value and multiplied by the square root of the reference spectrum

$$F_{\mathrm{new}}(\omega) = \frac{F(\omega)\sqrt{S_G(\omega)}}{|F(\omega)|}. \tag{A1}$$

After the Fourier back-transformation of $F_{\mathrm{new}}(\omega)$, the new sequence (x_i^{new}) has the desired correlations (i.e., the desired γ), but the shape of the distribution has changed towards a (more or less) Gaussian distribution. In order to enforce the desired distribution, we exchange the (x_i^{new}) by the (x_i^{ref}), such that the largest value of the new set is replaced by the largest value of the reference set, the second largest of the new set by the second largest of the reference set and so on. After this the new sequence has the desired distribution and is clearly correlated. However, due to the exchange algorithm the perfect long-term correlations of the new data sequence were slightly altered again. So the procedure is repeated: the new sequence is Fourier transformed followed by spectrum adjustment, and the exchange algorithm is applied to the Fourier back-transformed data set. These steps are repeated several times, until the desired quality (or the best possible quality) of the spectrum of the new data series is achieved.

[1] *Fractals in Science*, edited by A. Bunde and S. Havlin (Springer, Berlin, 1994).
[2] *The Science of Disasters—Climate Disruptions, Heart Attacks, and Market Crashes*, edited by A. Bunde, J. Kropp, and H.-J. Schellnhuber (Springer, Berlin, 2002).
[3] A. Bunde, J. F. Eichner, J. W. Kantelhardt, and S. Havlin, Phys. Rev. Lett. **94**, 048701 (2005).
[4] J. F. Eichner, J. W. Kantelhardt, A. Bunde, and S. Havlin, Phys. Rev. E **73**, 016130 (2006).
[5] A. Bunde, J. F. Eichner, J. W. Kantelhardt, and S. Havlin,

Physica A **330**, 1 (2003).
[6] E. G. Altmann and H. Kantz, Phys. Rev. E **71**, 056106 (2005).
[7] H. E. Hurst, R. P. Black, and Y. M. Simaika, *Long-term Storage: An Experimental Study* (Constable & Co. Ltd., London, 1965).
[8] B. B. Mandelbrot and J. R. Wallis, Water Resour. Res. **5**, 321 (1969).
[9] I. Rodriguez-Iturbe and A. Rinaldo, *Fractal River Basins—Change and Self-Organization* (Cambridge University Press, 1997).
[10] J. W. Kantelhardt, E. Koscielny-Bunde, D. Rybski, P. Braun, A. Bunde, and S. Havlin, J. Geophys. Res., [Atmos.] **111**, D01106 (2006).
[11] E. Koscielny-Bunde, A. Bunde, S. Havlin, and Y. Goldreich, Physica A **231**, 393 (1996).
[12] J. D. Pelletier and D. L. Turcotte, J. Hydrol. **203**, 198 (1997).
[13] E. Koscielny-Bunde, A. Bunde, S. Havlin, H. E. Roman, Y. Goldreich, and H.-J. Schellnhuber, Phys. Rev. Lett. **81**, 729 (1998).
[14] P. Talkner and R. O. Weber, Phys. Rev. E **62**, 150 (2000).
[15] J. F. Eichner, E. Koscielny-Bunde, A. Bunde, S. Havlin, and H.-J. Schellnhuber, Phys. Rev. E **68**, 046133 (2003).
[16] M. F. Shlesinger, B. J. West, and J. Klafter, Phys. Rev. Lett. **58**, 1100 (1987).
[17] R. R. Prasad, C. Meneveau, and K. R. Sreenivasan, Phys. Rev. Lett. **61**, 74 (1988).
[18] C.-K. Peng, J. Mietus, J. M. Hausdorff, S. Havlin, H. E. Stanley, and A. L. Goldberger, Phys. Rev. Lett. **70**, 1343 (1993).
[19] A. Bunde, S. Havlin, J. W. Kantelhardt, T. Penzel, J.-H. Peter, and K. Voigt, Phys. Rev. Lett. **85**, 3736 (2000).
[20] J. W. Kantelhardt, T. Penzel, S. Rostig, H. F. Becker, S. Havlin, and A. Bunde, Physica A **319**, 447 (2003).
[21] C.-K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons, and H. E. Stanley, Nature (London) **356**, 168 (1992).
[22] A. Arneodo, E. Bacry, P. V. Graves, and J. F. Muzy, Phys. Rev. Lett. **74**, 3293 (1995).
[23] Y. H. Liu, P. Cizeau, M. Meyer, C.-K. Peng, and H. E. Stanley, Physica A **245**, 437 (1997).
[24] A. Bunde, J. F. Eichner, J. W. Kantelhardt, and S. Havlin, Physica A **342**, 308 (2004).
[25] M. S. Santhanam and H. Kantz, Physica A **345**, 713 (2005).
[26] K. Yamasaki, L. Muchnik, S. Havlin, A. Bunde, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **102**, 9424 (2005).
[27] T. Schreiber and A. Schmitz, Phys. Rev. Lett. **77**, 635 (1996).
[28] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, Phys. Rev. E **49**, 1685 (1994).
[29] M. Kac, Bull. Am. Phys. Soc. **53**, 1002 (1947).
[30] H. v. Storch and F. W. Zwiers, *Statistical Analysis in Climate Research* (Cambridge University Press, Cambridge, 2001).
[31] When considering the (different) problem of zero level crossing in long term correlated Gaussian data it has been proven by G. F. Newell and M. Rosenblatt, Ann. Math. Stat. **33**, 1306 (1962), that the probability of having no zero level crossing after t time steps is bounded from above by a stretched exponential. See also I. E. Blake and W. C. Lindsay, IEEE Trans. Inf. Theory **19**, 295 (1973); S. N. Majumdar, Curr. Sci. **77**, 370 (1999).
[32] A. Corral, Phys. Rev. Lett. **92**, 108501 (2004).
[33] V. N. Livina, S. Havlin, and A. Bunde, Phys. Rev. Lett. **95**, 208501 (2005).
[34] R. Shcherbakov, G. Yakovlev, D. L. Turcotte, and J. B. Rundle, Phys. Rev. Lett. **95**, 218501 (2005).
[35] The methods applied to create long-term correlated data [27,40] become unprecise for very small values of γ and create data with an effective γ_{eff} slightly larger than γ.
[36] J. W. Kantelhardt, E. Koscielny-Bunde, H. H. A. Rego, S. Havlin, and A. Bunde, Physica A **295**, 441 (2001).
[37] K. Hu, P. Ch. Ivanov, Z. Chen, P. Carpena, and H. E. Stanley, Phys. Rev. E **64**, 011114 (2001).
[38] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley, Physica A **316**, 87 (2002).
[39] Z. Chen, P. Ch. Ivanov, K. Hu, and H. E. Stanley, Phys. Rev. E **65**, 041107 (2002).
[40] H. A. Makse, S. Havlin, M. Schwartz, and H. E. Stanley, Phys. Rev. E **53**, 5445 (1996).