Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Physica A 342 (2004) 308–314
PHYSICA A
www.elsevier.com/locate/physa

Return intervals of rare events in records with long-term persistence

Armin Bunde[a,*], Jan F. Eichner[a], Shlomo Havlin[b], Jan W. Kantelhardt[a,1]

[a] *Institut für Theoretische Physik III, Justus-Liebig-Universität, Heinrich-Buff-Ring 16, 35392 Giessen, Germany*

[b] *Minerva Center and Dept. of Physics, Bar-Ilan University, Ramat-Gan, Israel*

Received 1 December 2003

Available online 17 May 2004

Abstract

Many natural records exhibit long-term correlations characterized by a power-law decay of the auto-correlation function, $C(s) \sim s^{-\gamma}$, with time lag s and correlation exponent $0 < \gamma < 1$. We study, how the presence of such correlations affects the statistics of the return intervals r_q for events above a certain threshold value q. We find that (a) the mean return interval R_q does not depend on γ, (b) the distribution of r_q follows a stretched exponential, $\ln P_q(r) \sim -(r/R_q)^{\gamma}$, and (c) the return intervals are also long-term correlated with the exponent γ, yielding clustering of both small and large return intervals. We provide indications that both the stretched exponential behaviour and the clustering of rare events can be seen in long temperature records.

PACS: 05.40.−a; 05.45.Tp; 92.60.Wc

Keywords: Long-term correlations; Time series; Rare events; Return periods; Clustering; Temperature records

1. Introduction

In recent years there is growing evidence that many natural records exhibit long-term persistence [1]. Prominent examples include heartbeat records [2–4], DNA sequences

* Corresponding author. Tel.: +49-641-99-33360; fax: +49-641-99-33369.
E-mail address: armin.bunde@theo.physik.uni-giessen.de (A. Bunde).

[1] Present address: Computational Nanoscience, Fachgruppe Theorie, Fachbereich Physik, Martin-Luther-University Halle-Wittenberg, 06099 Halle (Saale), Germany.

0378-4371/$ - see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.physa.2004.01.069

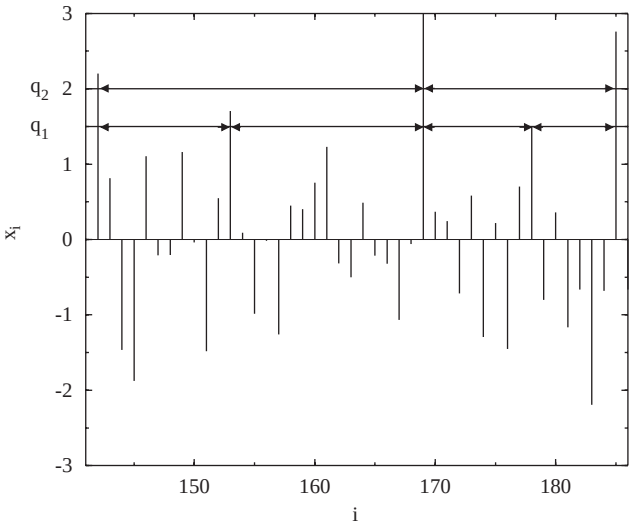

Fig. 1. Illustration of the return intervals $r_q(l)$, $l=1,\ldots,N_q$, of a record x_i, $i=1,\ldots,N$. The return intervals for two threshold values q_1 and q_2 are indicated by arrows.

[5,6], hydrological data [7,8], meteorological and climatological [9–12] records, as well as turbulence records [13,14]. Long-term correlations have also been found in the volatility of economic records [15].

An important problem in extreme value statistics (for reviews, see, e.g., Refs. [16–19]), is the re-occurrence of rare events exceeding a threshold q (see Fig. 1). The aim is to predict catastrophic events such as floods, droughts or stock crashes. The basic assumption is that the events are uncorrelated. In this case, the statistics of the return intervals r_q is solely determined by the tail of the distribution $D(x)$ of the elements x_i in the considered record. When the tail of the distribution $D(x)$ is determined, the mean return interval R_q is given by $R_q=\langle r_q\rangle=\{\int_q^{\infty} D(x)\,\mathrm{d}x/\int_{-\infty}^{+\infty} D(x)\,\mathrm{d}x\}^{-1}$. For uncorrelated records, the return intervals are also uncorrelated and obey the Poisson distribution $P_q(r)=(1/R_q)\exp(-r/R_q)$. Accordingly, one of the main challenges of the traditional extreme value statistics has been to develop appropriate methods to evaluate the tail of $D(x)$ accurately.

Here we study, how the statistics of the return intervals is modified in the presence of long-term persistence. We consider records $\{x_i\}$, $i=1,\ldots,N$, standardized to zero mean and unit variance, that are long-term correlated with an auto-correlation function $C_x(s)=\langle x_i x_{i+s}\rangle \equiv (1/(N-s))\sum_{i=1}^{N-s} x_i x_{i+s}$ that decays by a power law,

$$C(s)\sim s^{-\gamma}, \quad 0<\gamma<1\,. \tag{1}$$

We are interested in the return intervals and their statistics. In the following we show, how the mean return interval R_q, the distribution $P_q(r)$ of the return intervals, and the correlations between subsequent return intervals are affected by the presence of long-term correlations.

2. Mean return interval and distribution of the return intervals

For simplicity, we assume that the x_i-values ($i=1,\ldots,N$) are chosen from a Gaussian distribution. First we consider the mean return interval R_q of a given record. For a certain threshold q, there exist N_q return intervals $r_q(l)$, $l=1,\ldots,N_q$. For $1 \ll N_q$, we have $\sum_{l=1}^{N_q} r_q(l) \cong N$ (for N approaching infinity, the equal sign holds). When the data are shuffled, the long-term correlations are destroyed, but the sum rule still applies (with the same N_q value). Accordingly, for both cases, $R_q \equiv (1/N_q)\sum_{l=1}^{N_q} r_q(l) \cong N/N_q$, i.e., the mean return interval is not affected by the long-term correlations and can be obtained directly from the tail of the distribution function $D(x)$. Accordingly, there is a one-by-one correspondence between q and R_q, which we also confirmed numerically.

Next, we consider the distribution $P_q(r)$ of the return intervals as a function of the correlation exponent γ. In the numerical study, we have generated long records up to length $N=2^{21}$, for various values of γ, by the Fourier-transform technique (see, e.g., Ref. [20] and references therein). For each γ we calculated $P_q(r)$ for several threshold values q. A representative result for $\gamma=0.4$ and $q=2.0$ ($R_q \simeq 44$) is shown in Fig. 2(a) (in grey). In the semi-logarithmic plot, $P_q(r)$ for $\gamma=0.4$ differs considerably from the Poisson distribution of the shuffled data that we show for comparison. The probabilities of having return intervals well below R_q and well above R_q are strongly enhanced in the correlated record. To determine the functional form of $P_q(r)$, we have plotted in a double-logarithmic fashion $-\ln(P_q(r)/P_q(1))$ as a function of r/R_q. The results for $\gamma=0.1$, 0.4, 0.7 and $q=1.5$, 2.5 (corresponding to $R_q \simeq 15$ and 161), as well as for the shuffled data are shown in Fig. 2(b). For each value of γ, the curves with different q values collapse to a single line. The slopes of all lines coincide, within the error bars, with the values of the correlation exponent γ. Accordingly, we conclude that the distribution function of the return intervals has the form of a *stretched* exponential,

$$\ln P_q(r) \sim -(r/R_q)^{\gamma}, \quad 0 < \gamma \leqslant 1 \,. \tag{2}$$

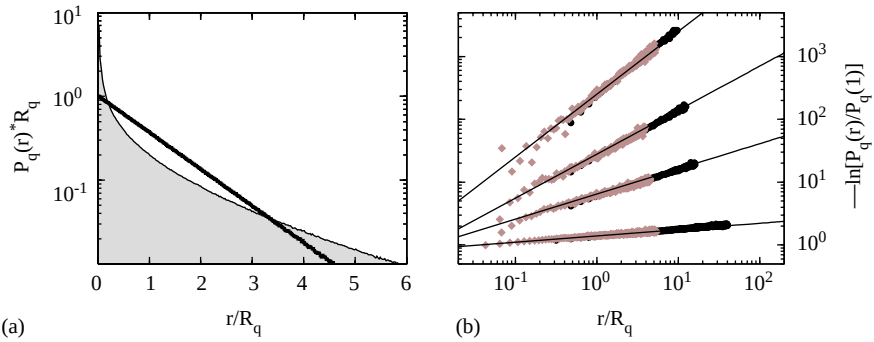

Fig. 2. (a) Distribution $P_q(r)$ of the return intervals r for the threshold $q=2.0$ ($R_q \simeq 44$) for long-term correlated data ($\gamma=0.4$, grey histogram) and for shuffled data (filled squares). (b) Double-logarithmic plot of $-\ln(P_q(r)/P_q(1))$ as a function of r/R_q for $\gamma=0.1$, 0.4, 0.7 as well as for the shuffled data (from bottom to top) and $q=1.5$ (black), 2.5 (grey). The straight lines are shown for comparison and have the slope γ for the long-term correlated data and one for the uncorrelated shuffled data.

We like to note that a similar stretched exponential behaviour has been obtained analytically [21], when considering the problem of zero-level crossing in the presence of long-term correlations. For $\gamma \geqslant 1$, the correlations are short-ranged, and the distribution of large return intervals is described by the Poisson distribution.

3. Correlation of the return intervals

Eq. (2) indicates that return intervals both well below and well above R_q are considerably more frequent for long-term correlated data than for uncorrelated data. It does not quantify, however, if the return intervals themselves are arranged in a correlated or in an uncorrelated fashion. To study the correlations, we evaluated the auto-correlation function $C_r(s) = \langle r_q(l) r_q(l+s) \rangle - R_q^2$. The results for $\gamma = 0.4$ and 0.7 and three q-values each ($q = 1$ and 2) are shown in Fig. 3(a). In the double-logarithmic presentation, for each γ-value, the three curves are parallel straight lines. This suggests that also the return intervals are long-term correlated, with the same exponent γ as in the original record. Accordingly, large and small return intervals are not arranged in a random fashion. Instead, we expect them to form clusters.

The calculation of the auto-correlation functions requires record lengths of more than 10^6 data points. Real records consist of considerably less data points. Thus, for quantifying the clustering of extreme events in real data, we need to consider quantities that require considerably less statistics. Such a quantity is the conditional mean return interval R_{q,r_0}, which is defined as the mean value of those intervals in the record that immediately follow an interval of length r_0. For uncorrelated systems, subsequent return intervals are independent of each other and $R_{q,r_0} = R_q$. For long-term correlated records, we expect $R_{q,r_0}/R_q < 1$ for r_0 well below R_q and $R_{q,r_0}/R_q > 1$ for r_0 well above R_q.

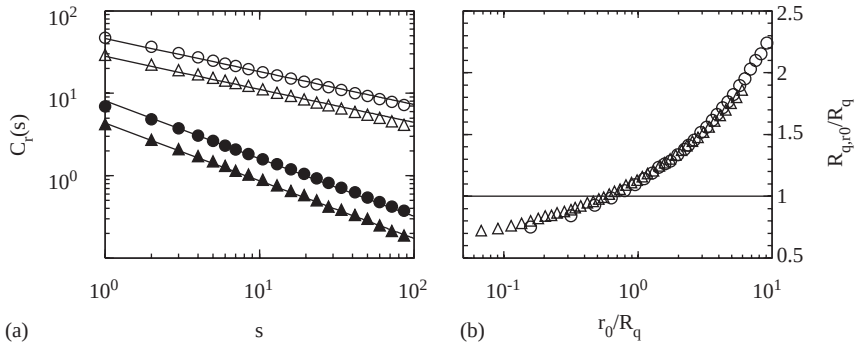

Fig. 3. (a) Auto-correlation function $C_r(s)$ of the return intervals for $q=1.0$ (circles) and $q=2.0$ (triangles) for long-term correlated data with $\gamma = 0.4$ (open symbols) and $\gamma = 0.7$ (filled symbols). In the double logarithmic plot, the slopes of the straight lines are equal to γ. The plot suggests that also the return intervals are long-term correlated, with the exponent γ of the original records. (b) Mean conditional return interval R_{q,r_0} versus preceding return interval r_0 (normalized by R_q) in a double-logarithmic plot for $\gamma = 0.4$, and for two values of q [as in (a)]. The straight line is the result for the shuffled uncorrelated data. The length of the series was $N = 2^{21}$, and averages over 1000 configurations have been performed.

Fig. 3(b) shows the numerical results for $R_{q,r_0}/R_q$ as a function of r_0/R_q, for $\gamma = 0.4$ and $q = 1$ and 2. Within the numerical accuracy, the data for different q-values scale. The figure quantifies the way, clumping and stretching of large events occurs when the data are long-term correlated. For $\gamma = 0.4$, for example, R_{q,r_0} can be as low as $0.7R_q$ for $r_0 \simeq R_q/5$ and as large as $1.8R_q$ for r_0 close to $5R_q$.

4. Application to temperature records

For testing the relevance of our results for real records, we have analyzed the 218-year maximum temperature record of Prague [11] and a 851-year long reconstructed record of annual northern hemisphere temperatures [22]. Both data sets are long-term correlated, with $\gamma \approx 0.7$ for the temperatures in Ref. [11] and $\gamma \approx 0.4$ for the temperature reconstruction.[1] For comparison with the theoretical results, we have standardized the records to zero mean and unit variance. The results are presented in Fig. 4. Figs. 4(a) and (b) show the distribution $P_q(r)$ of the return intervals r for the original data (filled circles) and for the shuffled data (open triangles). While the shuffled data clearly follow the Poisson distribution (dashed line) the original data are close to a stretched exponential decay (continuous line) with $\gamma = 0.7$ (in Fig. 4(a)) and $\gamma = 0.4$ (in Fig. 4(b)) in agreement with the prediction (2). The fluctuations of the results are due to the short length of the records (in particular for the annual data in Fig. 4(b)). In Fig. 4(a) strong short-range correlations (due to 'Grosswetterlagen') on time scales below 10 days cause an increased occurrence of very short return intervals. Hence, the normalization of the distribution $P_q(r)$ results in a shift of the distribution data for large r and we had to multiply the stretched exponential by the prefactor 0.2 to obtain agreement with the data.

Figs. 4(c) and (d) show the mean conditional return interval R_{q,r_0}, for both temperature records. The figures indicate clustering of small and large return intervals, in agreement with the predicted behaviour of Fig. 3(b). For the shuffled data, $R_{q,r_0}/R_q$ is close to one and the clustering disappears, as expected. The scattering on both the original and the shuffled data is due to the short length of the records.

In summary, we have shown that for long-term correlated records with a Gaussian distribution, the distribution of the return intervals follows a stretched exponential with an exponent γ identical to the correlation exponent. We also found that the return intervals are arranged in a long-term correlated fashion, again described by the exponent γ. It is important to emphasize, that both the distribution function and the long-term correlations between the return intervals scale the same for different q-values (Figs. 2(b) and 3(b)). Due to the scaling, we can evaluate the behaviour of return intervals of extremely rare events (very large q-values) from the behaviour of intermediate q-values that are accessible in the observational data. We also presented indications

[1] We studied the long-term correlations by a detrended fluctuation analysis (DFA), see e.g. Ref. [11], where the scaling of the fluctuation function $F(s) \sim s^{\alpha}$ is considered. The exponent α is related to the correlation exponent γ by $\alpha = 1 - \gamma/2$. We obtained $\alpha = 0.8$, from which $\gamma = 0.4$ follows.

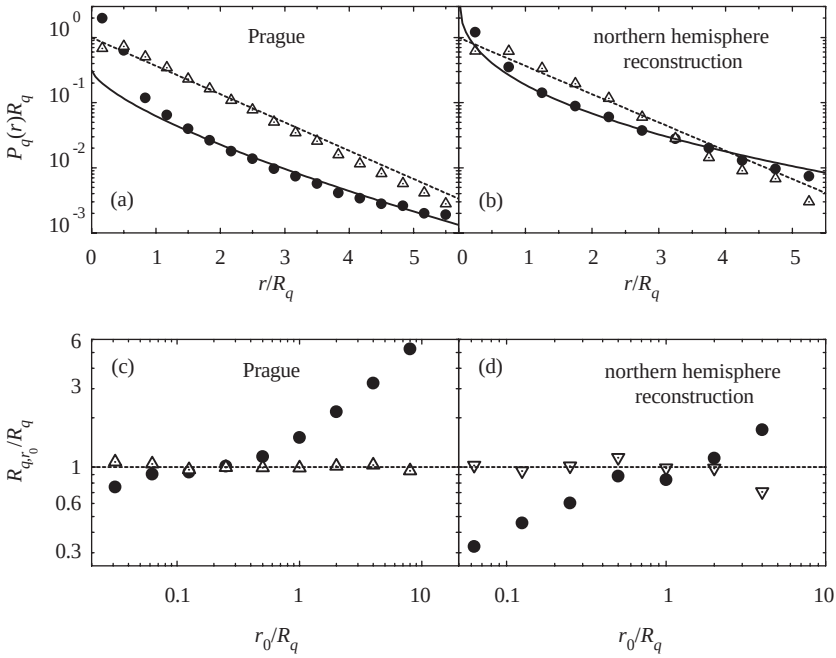

Fig. 4. Distribution function $P_q(r)$ and mean conditional return interval R_{q,r_0} for (a,c) daily temperature of Prague (CZ, 1775–1992) and (b,d) reconstructed annual temperature data of the northern hemisphere (1000–1850) [22]. The full circles represent the observed data sets, while the open triangles represent data sets that are obtained by randomly shuffling the observed original records. Both original records exhibit long-term correlations with $\gamma \approx 0.7$ (a) and 0.4 (b). Since $P_q(r)R_q$ as well as $R_{q,r_0}/R_q$ depend only on r/R_q [see Figs. 2(b) and 3(b)], we averaged $P_q(r)R_q$ and $R_{q,r_0}/R_q$ over several q-values in order to improve the statistics. For the reconstructed temperature record (b,d) we limited ourselves to data up to 1850 in order to exclude possible clustering of rare events due to global warming.

that both the stretched exponential behaviour and the clustering of rare events can be seen in long observational and reconstructed temperature records.

Acknowledgements

This work has been supported by the Bundesministerium für Bildung und Forschung and the Israel Science Foundation.

References

[1] A. Bunde, J. Kropp, H.-J. Schellnhuber (Eds.), The Science of Disasters—Climate Disruptions, Heart Attacks and Market Chrashes, Springer, Berlin, 2002.
[2] C.-K. Peng, J. Mietus, J.M. Hausdorff, S. Havlin, H.E. Stanley, A.L. Goldberger, Phys. Rev. Lett. 70 (1993) 1343.
[3] C. Schäfer, M.G. Rosenblum, J. Kurths, H.-H. Abel, Nature 392 (1998) 239.

[4] A. Bunde, S. Havlin, J.W. Kantelhardt, T. Penzel, J.-H. Peter, K. Voigt, Phys. Rev. Lett. 85 (2000) 3736.
[5] C.-K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H.E. Stanley, Nature 356 (1992) 168.
[6] A. Arneodo, E. Bacry, P.V. Graves, J.F. Muzy, Phys. Rev. Lett. 74 (1995) 3293.
[7] H.E. Hurst, R.P. Black, Y.M. Simaika, Long-Term Storage: An Experimental Study, Constable & Co. Ltd., London, 1965.
[8] B.B. Mandelbrot, J.R. Wallis, Water Resour. Res. 5 (1969) 321.
[9] E. Koscielny-Bunde, A. Bunde, S. Havlin, Y. Goldreich, Physica A 231 (1996) 393.
[10] J.D. Pelletier, D.L. Turcotte, J. Hydrol. 203 (1997) 198.
[11] E. Koscielny-Bunde, A. Bunde, S. Havlin, H.E. Roman, Y. Goldreich, H.-J. Schellnhuber, Phys. Rev. Lett. 81 (1998) 729.
[12] P. Talkner, R.O. Weber, Phys. Rev. E 62 (2000) 150.
[13] M.F. Shlesinger, B.J. West, J. Klafter, Phys. Rev. Lett. 58 (1987) 1100.
[14] R.R. Prasad, C. Meneveau, K.R. Sreenivasan, Phys. Rev. Lett. 61 (1988) 74.
[15] Y.H. Liu, P. Cizeau, M. Meyer, C.-K. Peng, H.E. Stanley, Physica A 245 (1997) 437.
[16] E.J. Gumbel, Statistics of Extremes, Columbia University Press, New York, 1958.
[17] M.R. Leadbetter, G. Lindgren, H. Rootzen, Extremes and Related Properties of Random Sequences and Processes, Springer, New York, 1983.
[18] J. Galambos, J. Lechner, E. Simin (Eds.), Extreme Value Theory and Applications, Kluwer Academic Publishers, Dordrecht, 1994.
[19] H.v. Storch, F.W. Zwiers, Statistical Analysis in Climate Research, Cambridge University Press, Cambridge, 2001.
[20] H.A. Makse, S. Havlin, M. Schwartz, H.E. Stanley, Phys. Rev. E 53 (1996) 5445.
[21] G.F. Newell, M. Rosenblatt, Ann. Math. Stat. 33 (1962) 1306.
[22] P.D. Jones, K.R. Briffa, T.P. Barnett, S.F.B. Tett, The Holocene 8 (1998) 455;
Data obtained from: IGBP PAGES/World Data Center-A for Palaeoclimatology Data Contribution Series #1998-039, NOAA/NGDC Palaeoclimatology Program, Boulder CO, USA.