



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Physica A 330 (2003) 1–7

PHYSICA A

[www.elsevier.com/locate/physa](http://www.elsevier.com/locate/physa)

# The effect of long-term correlations on the return periods of rare events

Armin Bunde<sup>a</sup>, Jan F. Eichner<sup>a</sup>, Shlomo Havlin<sup>b</sup>,  
Jan W. Kantelhardt<sup>a,\*</sup>

<sup>a</sup>*Institut für Theoretische Physik III, Justus-Liebig-Universität, Giessen, Heinrich-Buff-Ring 16, Giessen 35392, Germany*

<sup>b</sup>*Department of Physics, Minerva Center, Bar-Ilan University, Ramat-Gan, Israel*

---

## Abstract

The basic assumption of common extreme value statistics is that different events in a time record are uncorrelated. In this case, the return intervals  $r_q$  of events above a given threshold size  $q$  are uncorrelated and follow the Poisson distribution. In recent years there is growing evidence that several hydro-meteorological and physiological records of interest (e.g. river flows, temperatures, heartbeat intervals) exhibit long-term correlations where the autocorrelation function decays as  $C_x(s) \sim s^{-\gamma}$ , with  $\gamma$  between 0 and 1. Here we study how the presence of long-term correlations changes the statistics of the return intervals  $r_q$ . We find that (a) the mean return intervals  $R_q = \langle r_q \rangle$  are independent of  $\gamma$ , (b) the distribution of the  $r_q$  follows a stretched exponential,  $\ln P_q(r) \sim -(r/R_q)^\gamma$ , and (c) the return intervals are long-term correlated with an exponent  $\gamma'$  close to  $\gamma$ .

© 2003 Elsevier B.V. All rights reserved.

PACS: 05.40.-a; 05.45.Tp; 92.60.-e

Keywords: Long-term correlations; Time series; Rare events; Return periods; Fluctuation analysis

---

## 1. Introduction

By extreme value statistics (for reviews, see e.g. Refs. [1–3]) one wants to learn about the occurrence of rare events, for example floods, that usually are very uncomfortable and make a lot of damage. One usually considers a record of  $N$  elements  $x_i$ ,  $i = 1, 2, \dots, N$ , that can be, e.g., annual river flows or annual temperatures at a given

---

\* Corresponding author.

E-mail address: [jan.w.kantelhardt@theo.physik.uni-giessen.de](mailto:jan.w.kantelhardt@theo.physik.uni-giessen.de) (J.W. Kantelhardt).

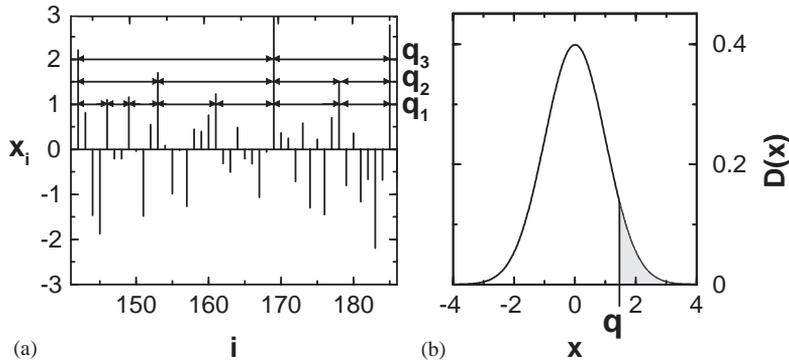


Fig. 1. (a) Illustration of the definition of the return intervals  $r_q(l)$ ,  $l=1, \dots, N_q$ , from a record  $x_i$ ,  $i=1, \dots, N$ . The return intervals for three threshold values  $q_1$ ,  $q_2$  and  $q_3$  are indicated by arrows. (b) Illustration of the distribution function  $D(x)$  of the values  $x_i$  in a record. Here, a Gaussian with zero mean and variance one is shown. The ratio between the grey area in the figure and the total area under the distribution function is the probability that an event of size greater than or equal to  $q$  occurs.

hydrological or meteorological station. The main quantity here is the time interval between events  $x_i$  that exceed a certain threshold  $q$ . When the threshold is small, the return intervals are short, when the threshold is large, the return intervals are long (see Fig. 1(a)).

The first quantity one is interested in here, is the mean return time of an event of size  $q$  or greater, which we denote by  $R_q$ . The second quantity is the distribution of the return intervals, which we denote by  $P_q(r)$ . Note that the mean return interval is just the first moment of  $P_q(r)$ . And finally, we are interested in the correlation behaviour of the return intervals. If they are correlated, the corresponding autocorrelation function  $C_r(s)$  between return intervals separated by  $s - 1$  return intervals, is non-zero in a certain range of  $s$ . If they are uncorrelated, the correlation function is zero except for  $s = 0$ .

The basic assumption in the classical extreme value statistics is that the events are uncorrelated, at least when they are far away from each other. In this case, one can obtain the mean return time of large events of size greater than  $q$  just by considering the distribution function  $D(x)$  of the  $x_i$  record. In the example in Fig. 1(b),  $D(x)$  is a Gaussian with zero mean and variance one. The ratio between the grey area in the figure and the total area under the distribution function is the probability that an event of size greater or equal  $q$  occurs, and the mean return time is just the inverse of this probability. When the events are uncorrelated, also the return intervals are uncorrelated, and the distribution  $P_q(r)$  is the well-known Poisson distribution,

$$P_q(r) = (1/R_q) \exp(-r/R_q). \quad (1)$$

In this case, the only quantity that remains to be calculated is the mean return time, which directly follows from the tail of the distribution  $D$ . However, when  $q$  is large, an event of size  $\geq q$  is a rare event. In this case, the tail of the distribution function

is not well known, and a great amount of work of the conventional extreme value statistics has been invested in finding reasonable extrapolations of  $D(x)$  for very large values of  $x$  (see e.g. Refs. [1–5]).

Nothing is known, however, to the best of our knowledge, of the effect of long-term correlations on the statistics of the return intervals  $r$ , and this is, what we consider next.

## 2. Long-term correlated signals

Long-term correlated sequences  $x_i$ , with zero mean, are characterized by an auto-correlation function

$$C_x(s) = \langle x_i x_{i+s} \rangle = \frac{1}{N-s} \sum_{i=1}^{N-s} x_i x_{i+s} \quad (2)$$

that decays very slowly, by a power law, with the time span  $s$  separating the pairs of elements,

$$C_x(s) \sim s^{-\gamma}, \quad 0 < \gamma < 1. \quad (3)$$

The mean correlation time is the integral over  $C_x(s)$ . It is easy to see that this time diverges when  $\gamma$  is between 0 and 1, and that is why we call these records *long-term correlated*. In the last years it has become clear that several physiological and hydro-meteorological records are long-term correlated. For example, the sequence of heartbeat intervals of a healthy person, during wake [6,7] and during REM sleep [8] is long-range correlated with an exponent  $\gamma$  close to 0.3 for REM sleep. We also know, actually already since the pioneering work of Hurst [9], that daily or annual river flows measured at a hydrological station are usually long-term correlated with  $\gamma$  between 0.4 and 0.5 (see also Refs. [10,11]). Similar values have also been obtained for many other geophysical records [12,13] (see also Ref. [14]). Other examples include the volatility of stock prices with  $\gamma \approx 0.15$  [15], traffic in the internet [16,17], and atmospheric as well as sea-surface temperatures [18–21]. For temperature records at continental stations,  $\gamma$  is always very close to 0.7 [18,19]. For island stations as well as for sea-surface temperatures,  $\gamma$  is around 0.4 [20,21]. The fact that for all continental stations, irrespective their climate zone and their distance from the oceans, the range of  $\gamma$  values is very narrow [22], has made this power law an ideal (and uncomfortable) test bed for global climate models [23].

Next we consider, how the mean return interval  $R_q$ , the distribution  $P_q(r)$  of the return intervals and their correlation behaviour is changed in the presence of long-term correlations. We will always assume that the  $x_i$  values are chosen from a Gaussian distribution with zero mean and unit variance. We have generated the long-term correlated series using the Fourier transform technique (see, e.g. Ref. [24] and references therein). First, we consider the mean return interval  $R_q$ .

### 3. Mean return interval $R_q$

For simplicity, we assume periodic boundary conditions. In this case, for a given threshold  $q$ , there exist  $N_q$  return intervals  $r_q(l)$ ,  $l = 1, 2, \dots, N_q$ , which then satisfy the sum rule

$$\sum_{l=1}^{N_q} r_q(l) = N. \tag{4}$$

When we shuffle the data randomly, thereby destroying the long-term correlations, the sum rule also applies (with the same  $N_q$  value). Accordingly, for both long-term correlated and uncorrelated records,  $R_q$  is given by

$$R_q = \frac{1}{N_q} \sum_{l=1}^{N_q} r_q(l) = \frac{N}{N_q}, \tag{5}$$

i.e., the mean return interval is not changed by the presence of long-term correlations, and the techniques developed for uncorrelated data to calculate  $R_q$  can be used also for records with long-term correlations. Next, we consider the distribution  $P_q(r)$  of the return intervals.

### 4. Distribution function $P_q(r)$ of the return intervals

To investigate  $P_q(r)$  as a function of  $\gamma$ , we have generated long records ( $N = 10^6$ ) for  $\gamma = 0.4, 0.7$ , and without correlations ( $\gamma = \infty$ ). For each  $\gamma$  value we calculated  $P_q(r)$  for several threshold values  $q$ . The result for  $\gamma = 0.4$  and  $q = 1.5$  is shown in Fig. 2(a) (in grey), where also  $P_q(r)$  for uncorrelated data is shown (in black) for comparison.

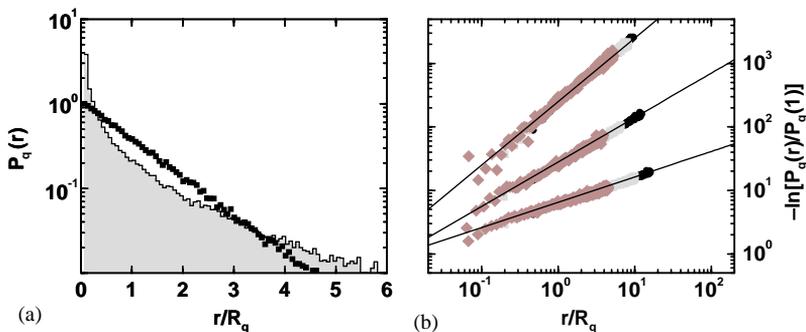


Fig. 2. (a) Distribution  $P_q(r)$  of the return intervals  $r$  for the quantile  $q = 1.5$  for uncorrelated data (filled squares) and for long-term correlated data ( $\gamma = 0.4$ , grey histogram). (b) Double-logarithmic plot of  $-\ln(P_q(r)/P_q(1))$  as a function of  $r/R_q$  for  $\gamma = 0.4, 0.7$  and  $\infty$  (from bottom to top) and  $q = 1.5$  (black),  $2.0$  (light grey),  $2.5$  (dark grey). In this presentation, the slope of the curves corresponds to the exponent  $\gamma$  in the stretched exponential (6). The straight lines are shown for comparison and they have the slope  $\gamma$  for the long-term correlated data and slope one for the uncorrelated data ( $\gamma = \infty$ ).

This semi-logarithmic plot  $P_q(r)$  for  $\gamma = 0.4$  differs considerably from the Poisson distribution for uncorrelated data. The probability of having return intervals well below  $R_q$  and well above  $R_q$  is strongly enhanced compared with the uncorrelated case. 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  $\gamma = 0.4, 0.7$  and  $\infty$  and  $q = 1.5, 2.0$  and  $2.5$  are shown in Fig. 2(b). For each value of  $\gamma$ , the curves with different  $q$  values collapse to a single line. Each line, in the double-logarithmic plot, has a slope very close to  $\gamma$ . Therefore, for long-term correlated records with exponent  $\gamma$  between 0 and 1, the distribution function of the return intervals becomes simply a stretched exponential,

$$\ln P_q(r) \sim -(r/R_q)^\gamma, \quad (6)$$

where the exponent is identical with the correlation exponent.

## 5. Correlations among the return intervals

Next we studied the type of correlations between the return intervals when the original record exhibits long-term correlations. To this end, we did not evaluate the auto-correlation function  $C_r(s)$  of the return intervals  $r_q(l)$  directly, but instead evaluated the fluctuation function  $F_r(s)$  to determine—for long-term correlated data—the correlation exponent  $\gamma'$  in  $C_r(s) \sim s^{-\gamma'}$  (compare with Eq. (3)). To obtain  $F_r(s)$ , one considers the record  $r_q(1) - R_q, r_q(2) - R_q, \dots, r_q(N_q) - R_q$ , and regards the  $l$ th element as  $l$ th increment of a random walker on a linear chain. The walker steps to the left (right) if  $r_q(l) - R_q$  is negative (positive), and the step length is  $|r_q(l) - R_q|$ . The fluctuation function  $F_r(s)$  is just the root-mean-square displacement of the random walker, after  $s$  steps. It is known that (see e.g. Refs. [8,18])

$$F_r(s) \sim s^\alpha, \quad \alpha = 1 - \gamma'/2 \quad (7)$$

if the data are long-term correlated with an exponent  $\gamma'$  between 0 and 1. Otherwise,  $F_r(s) \sim s^{1/2}$ . Accordingly, from the asymptotic behaviour of  $F_r(s)$  we can learn, if the considered record of return intervals is long-term correlated. The result is shown in Fig. 3 for  $\gamma = 0.4$  and  $0.7$ . In the double-logarithmic presentation, both curves are asymptotic straight lines with exponents  $\alpha = 0.8$  for  $\gamma = 0.4$  and  $\alpha = 0.65$  for  $\gamma = 0.7$ . This suggests that also the return intervals are long-term correlated, with an exponent  $\gamma'$  that is close to the exponent  $\gamma$  of the original records. In long-term correlated records, long sequences can occur where the data are well above or below their average value, and it is generally a difficult task to distinguish between long-term correlations and trends. When the return intervals are long-term correlated, there is a much higher probability than for uncorrelated intervals to obtain sequences where the  $r_q$  are well above  $R_q$  followed by sequences where the  $r_q$  are well below  $R_q$ . Our results suggest that such a behaviour may not necessarily be the consequence of a trend (global warming etc.), but may arise in a natural way when the data are long-term correlated, which is the case for river flows and temperature records.

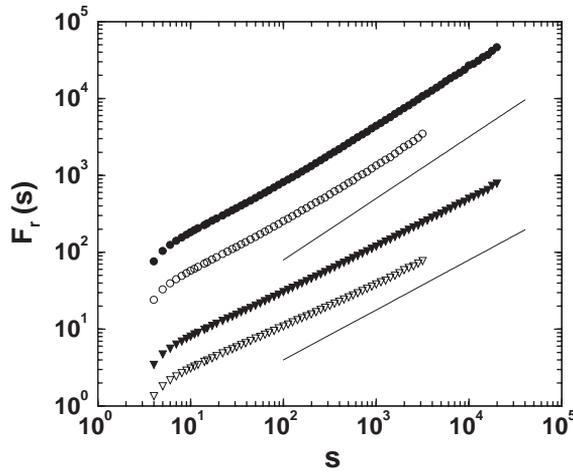


Fig. 3. Fluctuation analysis for the records of detrended return intervals,  $r_q(l) - R_q$ ,  $l = 1, \dots, N_q$  for  $q = 1.5$  (filled symbols),  $q = 2.5$  (open symbols) for long-term correlated data with  $\gamma = 0.4$  (circles) and  $0.7$  (triangles). The fluctuation function  $F_q(s)$  is shown versus time scale  $s$  in a double logarithmic plot, where the slope  $\alpha$  of the data is related to the correlation exponent  $\gamma' = 2 - 2\alpha$  of the return intervals. Both curves asymptotically approach the straight lines with slopes  $\alpha = 0.8$  for  $\gamma = 0.4$  and  $\alpha = 0.65$  for  $\gamma = 0.7$ . This suggests that also the return intervals are long-term correlated, with an exponent  $\gamma'$  that is close to the exponent  $\gamma$  of the original records. The length of the original records was  $N = 2 \times 10^6$ .

## Acknowledgements

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

## References

- [1] E.J. Gumbel, *Statistics of Extremes*, Columbia University Press, New York, 1958.
- [2] J. Galambos, J. Lechner, E. Simin (Eds.), *Extreme Value Theory and Applications*, Kluwer, Dordrecht, 1994.
- [3] H.v. Storch, F.W. Zwiers, *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge, 2002.
- [4] R.A. Fisher, L.H.C. Tippett, *Proc. Camb. Philos. Soc.* 24 (1928) 180.
- [5] T. Antal, M. Droz, G. Györgyi, Z. Racz, *Phys. Rev. Lett.* 87 (2001) 240 601.
- [6] C.-K. Peng, J. Mietus, J.M. Hausdorff, S. Havlin, H.E. Stanley, A.L. Goldberger, *Phys. Rev. Lett.* 70 (1993) 1343.
- [7] P.Ch. Ivanov, A. Bunde, L.A.N. Amaral, S. Havlin, J. Fritsch-Yelle, R.M. Baevsky, H.E. Stanley, A.L. Goldberger, *Europhys. Lett.* 48 (1999) 594.
- [8] A. Bunde, S. Havlin, J.W. Kantelhardt, T. Penzel, J.-H. Peter, K. Voigt, *Phys. Rev. Lett.* 85 (2000) 3736.
- [9] H.E. Hurst, *Trans. Am. Soc. Civil Eng.* 116 (1951) 770.
- [10] Y. Tessier, S. Lovejoy, P. Hubert, D. Schertzer, S. Pecknold, *J. Geophys. Res. Atmos.* 101 (D21) (1996) 26427.

- [11] E. Koscielny-Bunde, J.W. Kantelhardt, P. Braun, A. Bunde, S. Havlin, *Water Resour. Res.*, submitted for publication, preprint physics/0305078 (2003).
- [12] H.E. Hurst, R.P. Black, Y.M. Simaika, *Long-term Storage: An Experimental Study*, Constable, London, 1965.
- [13] B.B. Mandelbrot, J.R. Wallis, *Water Resour. Res.* 5 (1969) 321.
- [14] J.W. Kantelhardt, D. Rybski, S.A. Zschiegner, P. Braun, E. Koscielny-Bunde, V. Livina, S. Havlin, A. Bunde, *Physica A* 330 (2003) doi:10.1016/j.physa.2003.08.019 [these proceedings].
- [15] Y. Liu, P. Cizeau, M. Meyer, C.-K. Peng, H.E. Stanley, *Physica A* 245 (1997) 437.
- [16] W.E. Leland, M.S. Taqqu, W. Willinger, D.V. Wilson, *IEEE Trans. Networking* 2 (1994) 1.
- [17] V. Paxson, S. Floyd, *IEEE Trans. Networking* 3 (1995) 226.
- [18] E. Koscielny-Bunde, A. Bunde, S. Havlin, Y. Goldreich, *Physica A* 231 (1996) 393.
- [19] E. Koscielny-Bunde, A. Bunde, S. Havlin, H.E. Roman, Y. Goldreich, H.-J. Schellnhuber, *Phys. Rev. Lett.* 81 (1998) 729.
- [20] J.D. Pelletier, D.L. Turcotte, *J. Hydrol.* 203 (1997) 198.
- [21] R.A. Monetti, S. Havlin, A. Bunde, *Physica A* 320 (2003) 581.
- [22] A. Bunde, J.F. Eichner, S. Havlin, E. Koscielny-Bunde, H.J. Schellnhuber, D. Vjushin, preprint physics/0305080 (2003).
- [23] R.B. Govindan, D. Vyushin, A. Bunde, S. Brenner, S. Havlin, H.-J. Schellnhuber, *Phys. Rev. Lett.* 89 (2002) 028 501.
- [24] H.A. Makse, S. Havlin, M. Schwartz, H.E. Stanley, *Phys. Rev. E* 53 (1996) 5445.