ELSEVIER

Physica A 271 (1999) 448–457

PHYSICA A

www.elsevier.com/locate/physa

Growth of companies and water-level fluctuations of the river Danube

Imre M. Jánosi[a,d,*], Jason A.C. Gallas[b,c,d]

[a]*Department of Physics of Complex Systems, Eötvös University, Puskin u. 5-7, H-1088 Budapest, Hungary*

[b]*Grupo de Meteorologia e Climatologia, Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal*

[c]*Instituto de Física, Universidade Federal do Rio Grande do Sul, 91501-970 Porto Alegre, Brazil*

[d]*John von Neumann Institut für Computing (NIC), Forschungszentrum Jülich, D-52425 Jülich, Germany*

Received 22 May 1999

Abstract

Recent studies on growth rate of publicly traded companies revealed interesting scaling properties and universality. Based on the statistical analysis, different models have been proposed to cast light into the inner workings of companies responsible for the phenomena observed. The purpose of this paper is to point out that the properties of the analysed economic data might be present in a wider class of complex systems producing strongly correlated noisy time series. As an illustration, we report an investigation of the daily water-level fluctuations of the river Danube over a period of 87 years, as measured in Nagymaros, Hungary. The Danube data shows similar characteristics to those seen in the company growth. This suggests that a universal description of the statistics should exist for both systems.

PACS: 02.50.Ey; 05.40.+j; 89.90.+n

Keywords: Universal fluctuations; Growth-rate statistics; Scaling

1. Introduction

In a series of recent papers [1–5], the scaling behaviour of growth of profit oriented economic units was studied by means of modern tools of statistical physics.

* Corresponding author.
E-mail address: janosi@lecso.elte.hu (I.M. Jánosi)

0378-4371/99/$ - see front matter
PII: S0378-4371(99)00246-0

The empirical results for company growth were extracted from the Compustat database, comprising all publicly traded United States manufacturing firms within the years 1974–1993 [1,2]. An extension of this study evaluated the fluctuations in gross domestic product (GDP) of 152 countries for the period 1950–1992 [5]. (For the sake of simplicity we use the term "firm" for both of company and domestic economy.)

It was found that the distribution of firm sizes remained stable during the years considered, i.e. the mean value and the standard deviation are approximately constant. The distribution of sizes of new companies in each year is well approximated by a log-normal form. Further, Stanley et al. [1] and Amaral et al. [2] found that the distribution of annual (logarithmic) growth rates has an exponential shape, and the spread in the distribution of rates decreases with increasing sales as a power law over seven orders of magnitude.

In a subsequent paper, Buldyrev et al. [3] (see also Ref. [1]) proposed models intended to shed light into the behaviour observed with the scaling approach. They first studied a model in which the growth rate of a company is affected by a tendency to retain an "optimal size". Such modelling leads to an exponential distribution of the logarithm of the growth rate, in agreement with the empirical results. Then, a hierarchical tree-like model of a company was studied [3]. Two parameters of the model were related to an exponent β, which describes the dependence of the standard deviation of the distribution of growth rates on size. One of the main results is that $\beta = -\ln \Pi / \ln z$, where z defines the mean branching ratio of the hierarchical tree, and Π is the probability that the lower levels follow the policy of higher levels in the hierarchy. Note that the exponent value obeys $\beta \in [0, 0.5]$ for all probabilities and branching ratios, whenever scaling is present, and the empirical value for firms is $\beta_{emp} \approx \frac{1}{6}$. The distribution of growth rates of this hierarchical model reproduced the exponential form too.

Amaral et al. introduced a third model to explain the same scaling behaviour of company growth [4]. Here a system of interacting subunits (firms) are studied. These firms have a complex internal structure, with each firm composed of subunits (divisions). The size of the subunits evolves according to a random multiplicative process, and the growth rates of different divisions are independent of one another. Interaction among the units is treated in a "mean field" approximation through the imposition of a minimum size for the subunits. The size statistics is in agreement with the empirical data, again.

The robustness of the above results is remarkable: Three models of rather different spirit reproduced the same statistics. This suggests that other complex systems may show the same behaviour, and no special properties of interacting economic units are necessary for a proper description. Motivated by this observation, we repeated the same statistical analysis on a completely different time series that was available to us: Daily water-level fluctuations of the river Danube recorded between January 1, 1901 and December 31, 1987. For this data set the "growth rate" can be defined as the ratio h_{t+1}/h_t, where h_t is the water level on day t.

Surprisingly, our time series produces curves (shown below) displaying the same three basic characteristics which were extracted in Refs. [1–5]:

(i) The data displays similar exponential probability density distribution and generates the same typical tent-shape curves in double logarithmic plots.

(ii) The higher the initial value of a given bin, the narrower the distribution.

(iii) Tent widths also scale with the initial value producing a characteristic exponent, in our case with $\beta = 0.58 \pm 0.03$.

Our data set has a very different nature than the economic data in Refs. [1–5], and, in our opinion, none of the aforementioned models can be installed easily to support a similar interpretation.

It is the purpose of the present paper, first, to characterize shortly our data, to report the results found by repeating the analysis discussed in Refs. [1,2,4,5] and, finally, to present our conclusions offering an interpretation of the underlying characteristics that we believe to be common characteristics of both sets of data as well as probably many others. We emphasize here that the focus is not on the analysis of water level fluctuations, therefore we do not study the "classical" scaling properties of river data observed first by Hurst [6].

2. The Danube data: standard analysis

The basic data available to us is a set of 31 863 measurements providing daily records of the water level of the river Danube, as observed at Nagymaros, north of Budapest, between 1 January 1901 and 31 December of 1987 [7]. The water level is measured in cm with respect to a fixed reference level, as usual everywhere.

The data is stored as a two-dimensional array to which we associate the *height* variable $h_k(t)$, where the subscript k, $1901 \leqslant k \leqslant 1987$, refers to the year of the measurement, and t, $1 \leqslant t \leqslant 365$, refers to the day of the measurement within a given year k. A subset of this data, displaying the first 20 000 measurements, is shown in Fig. 1. As seen from the figure, the time series of Fig. 1a fluctuates bounded around a stable average (of 236 cm), as it must be for obvious reasons: The absolute minimum is limited by the dry river bed, and the absolute maximum by the height of the dams along the river.

From the individual water levels $h_k(t)$ we compute the 87-year average height $H(t)$ for every day $t = 1, 2, \ldots, 365$:

$$H(t) = \frac{1}{87} \sum_{k=1901}^{1987} h_k(t) . \qquad (1)$$

The 366th day of the leap years (29 February) is discarded. The yearly average is shown in Fig. 1b. As one recognizes easily from this figure, there is a strong seasonality in the data, which is a common characteristic of the rivers in the region [8]. The steep rise of the average water level in Spring is a simultaneous result of the snow-melt in the Alps and in the Carpatian mountains, and the intense rainfalls in the season. At the

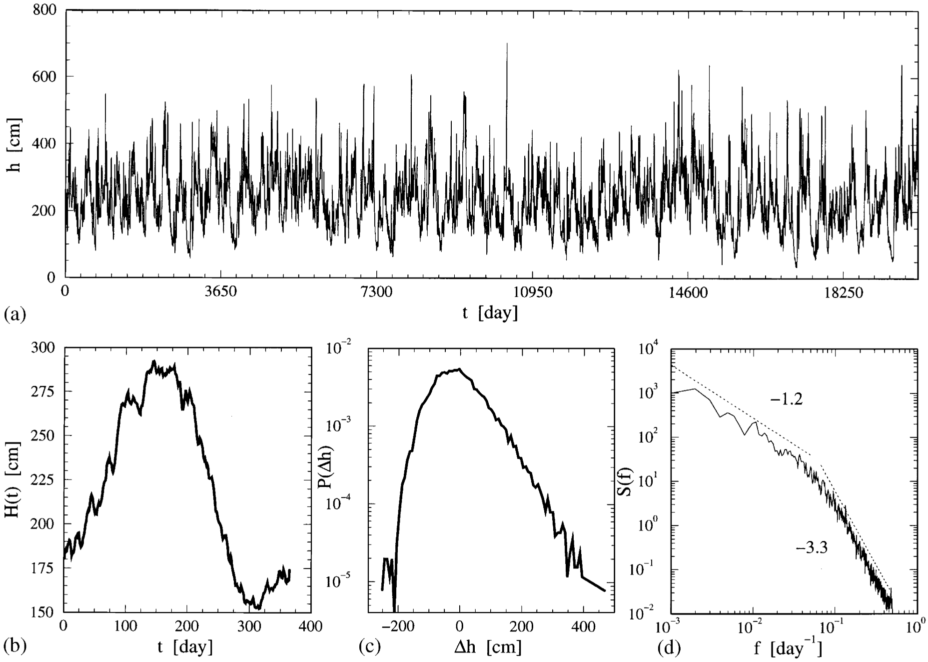

Fig. 1. (a) 20 000 data of daily water level of the river Danube, measured at Nagymaros from 1st of January 1901 [7]. (b) Seasonality of the average daily water for one year [see Eq. (1)]. (c) Probability density distribution $P(\Delta h)$ of the water level fluctuations Δh [see Eq. (2)]. Note that the vertical scale is logarithmic. (d) Power spectrum of the detrended time series obtained by the standard FFT method. Dotted lines show two scaling regimes, at low frequencies ($f < 0.05$ day^{-1}) the characteristic exponent is 1.2 ± 0.1, at large frequencies ($f > 0.1$ day^{-1}) the exponent value is 3.3 ± 0.1.

measuring station Nagymaros, the average water level variation is about 1.5 m while peak-to-peak variations may be as high as 7 m, depending on the year.

To overcome the natural nonstationarity of the data series due to seasonal trends, we have determined the fluctuations of the daily water level Δh with respect to the daily average over 87 years $H(t)$ as

$$\Delta h(t) = h_k(t) - H(t) \,. \tag{2}$$

Similar to Eq. (1), the leap days are omitted. The histogram (or empirical probability density distribution) $P(\Delta h)$ for the fluctuations Δh is shown in Fig. 1c. The asymmetric peak is clearly far from being a Gaussian, and a log-normal fit works only for the tails, too. We have not found a simple function which would give a satisfactory fit, the best result was achieved by the convolution of nine partial distributions having exponential shape. Fig. 1d shows the power spectrum $S(f)$ of the fluctuations $\Delta h(t)$ obtained by a standard FFT algorithm. The shape of the curve suggest scaling for two different frequency ranges. At low frequencies ($f < 0.05$ day^{-1}) the characteristic exponent is 1.2 ± 0.1, at large frequencies ($f > 0.1$ day^{-1}) the exponent value is 3.3 ± 0.1. The breakpoint around $f \approx 0.05$–0.06 day^{-1} (i.e. $t \approx 17$–20 days) is probably associated with the typical length of the time intervals in the dry summer and winter seasons, when the water level decreases monotonously as a consequence of low water supply.

3. Statistics of the rate of changes

The key quantity in the statistical analysis of the growth of companies [2,3] is the one year growth rate, which is defined in economics as the size of a given company with respect to some measure (sales, assets, number of employees, etc.) divided by the size in the preceding year. We can define an analogous quantity "growth rate" or rather "rate of change" for our time series as the ratio of the water level between consecutive days, h_{t+1}/h_t and, associated with it, the logarithmic rate of change $r = \ln(h_{t+1}/h_t)$. In the following we repeat the analysis step by step following Stanley et al. [1,2] to show that the rate of change in the case of water levels behaves very similarly to the growth rate that was observed for companies.

At first, we determine the conditional probability density distributions $P(r|h_0)$ of the one day logarithmic rate of change r with a given initial height h_0. Fig. 2a shows the result for three different initial values. The characteristic tent-shape seen in the figures suggests that the distribution has an exponential form

$$P(r|h_0) = \frac{1}{\sqrt{2}\sigma(h_0)} \exp\left(-\frac{\sqrt{2}|r - \bar{r}(h_0)|}{\sigma(h_0)} \right) , \tag{3}$$

where $\sigma(h_0)$ and $\bar{r}(h_0)$ are, respectively, the width and the average value, both quantities depending on the given bin centred at h_0. The smaller h_0, the larger the width $\sigma(h_0)$, similar to the company data [2,1]. Note that the centre of the partial distributions $\bar{r}(h_0)$ has typically a small negative value for the river, while slight positive values characterise the company data [3]. Fig. 2b shows how the width $\sigma(h_0)$ depends on the initial value h_0. For large h_0 values the scaling breaks down, but for $h_0 \leqslant 300$ cm a power-law assumption gives a satisfactory fit (also shown in the inset of Fig. 2b):

$$\sigma(h_0) \sim h_0^{-\beta}, \quad \beta = 0.58 \pm 0.03 . \tag{4}$$

Fig. 3 shows the rescaled probability density distribution $P' = \sqrt{2}\sigma(h_0)P(r|h_0)$ as a function of the rescaled logarithmic rate of change $r' = \sqrt{2}[r - \bar{r}(h_0)]/\sigma(h_0)$ for the three sets of data shown in Fig. 2a. All three data sets collapse into a single curve

$$P' = \exp(-|r'|) , \tag{5}$$

as predicted by Eqs. (3) and (4).

The characteristic differences between the statistics of company growth and water level changes can be summarized as follows. Firstly, the centre of the partial distributions $\bar{r}(h_0)$ is a small negative value for the Danube, while it is positive for the company growth. Secondly, the scaling exponents are different; β=0.58 for the Danube, and $\beta \approx \frac{1}{6}$ for the firms. Thirdly, the systematic deviation from the ideal exponential shape defined by Eq. (5) has an opposite sign: For $r' > 0$ the slopes of the rescaled

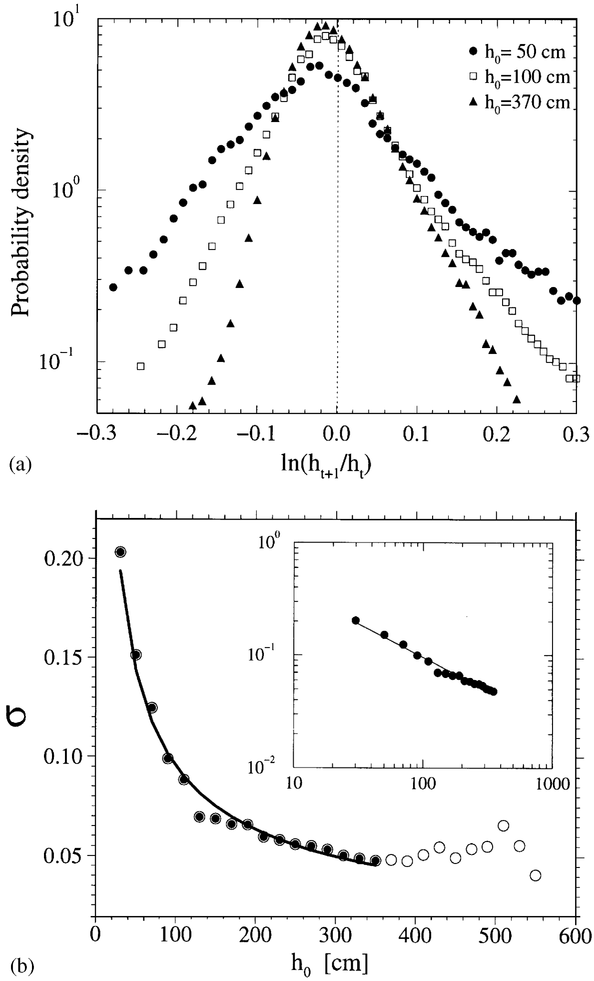

Fig. 2. (a) Probability density distribution of the one day logarithmic rate of change $\ln(h_{t+1}/h_t)$ for three different bins centred around $h_0 = 50, 100$, and 370 cm. Each bins have the same spread $h - 20 < h_0 < h + 20$ cm. The distributions have an exponential shape [Eq. (3)] centred around small negative values. (b) Width of the exponential distributions $\sigma(h_0)$ [see Eq. (3)] as a function of different initial values h_0. The solid line is a power-law fit for the data points denoted by heavy dots, also shown in the inset. For large initial values ($h_0 > 300$ cm) the scaling breaks down (empty circles). The exponent value is $\beta = 0.58 \pm 0.03$ [see Eq. (4)].

data for the companies seem to be larger than predicted by Eq. (5) (cf. Fig. 3 of [2]), whereas for the Danube data the same is visible but for $r' < 0$.

4. Discussion

It is obvious that the two systems compared above are of fundamentally different nature. Nevertheless, the similarities in the statistics of rate of changes suggest the

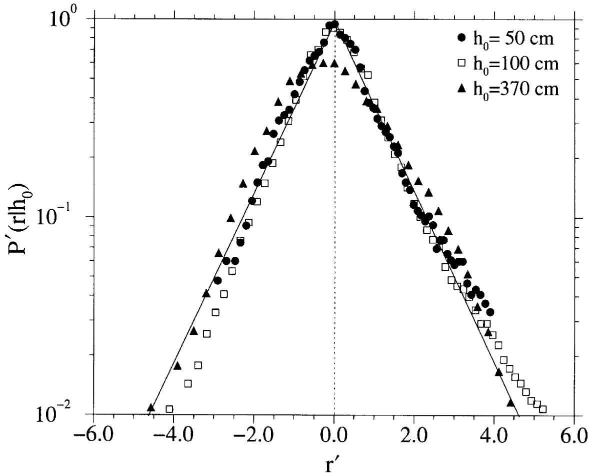

Fig. 3. Rescaled probability density distribution $P' = \sqrt{2}\sigma(h_0)P(r|h_0)$ as a function of the rescaled logarithmic rate of change $r' = \sqrt{2}[r - \bar{r}(h_0)]/\sigma(h_0)$ for the data shown in Fig. 2a. The data approximately collapse upon the universal curve Eq. (5) (thin solid line).

existence of some common features underlying both sets of data. One of the common characteristics is that both discrete time series have the form of

$$x_{t+1} = x_t \pm \varepsilon_t \,, \tag{6}$$

where ε_t is the change of the quantity x_t at time t. The logarithmic rate of change obeys the series expansion

$$r_t = \ln\left(\frac{x_{t+1}}{x_t}\right) = \ln\left(1 \pm \frac{\varepsilon_t}{x_t}\right) \approx \pm\frac{\varepsilon_t}{x_t} - \frac{1}{2}\left(\frac{\varepsilon_t}{x_t}\right)^2 \pm \frac{1}{3}\left(\frac{\varepsilon_t}{x_t}\right)^3 \cdots \,, \tag{7}$$

if $\varepsilon_t/x_t \ll 1$ holds. This is not a strict restriction, because relative changes can be small in many random-walk-like processes, such as the growth of companies or water level fluctuations. Eq. (7) means that the statistics of logarithmic rate of changes is dominated by the statistics of small relative changes. Let us discuss a few trivial examples.

The simplest case is when the variation of x does not depend either on the time t or the value of x. This is a one-dimensional discrete random walk of stepsize $\varepsilon = \Delta x$, i.e. $x_{t+1} = x_t \pm \Delta x$. For any particular value of x far enough from the origin, the width of the logarithmic rate of change "distribution" σ_r scales with x like $\sigma_r = \Delta x/x \sim x^{-1}$. The other limiting case is when the variation of x is strictly proportional to x as $x_{t+1} = x_t(1 \pm c)$, where c is a small positive constant. In this case the width of the logarithmic rate of change "distribution" is $\sigma_r = c \sim x^0$ for $c \ll 1$. It is easy to see that any scaling exponent can be realized between zero and one, if we assume that the variation obeys a power law

$$\varepsilon = cx^{\beta} \,, \tag{8}$$

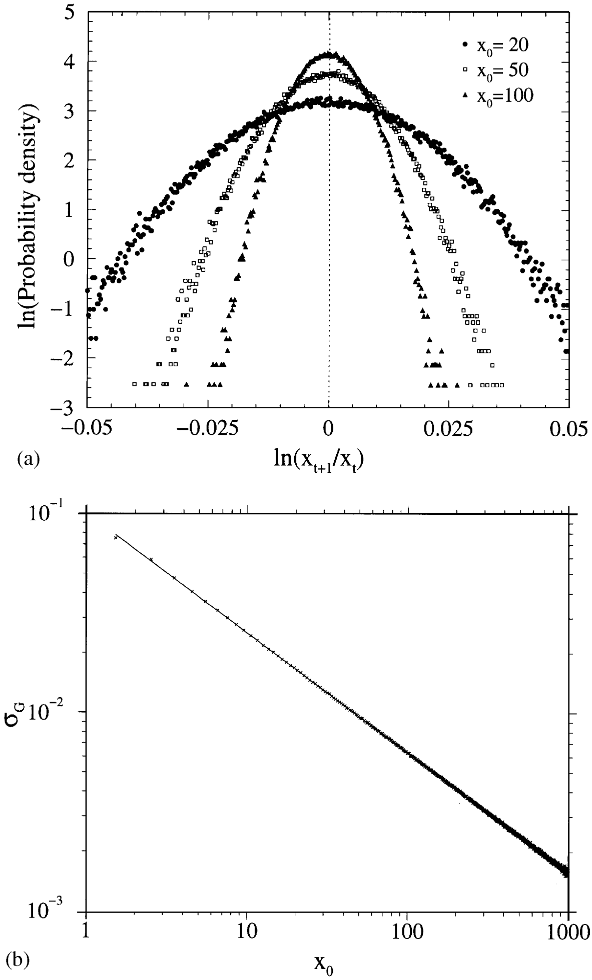

Fig. 4. (a) Partial probability density distribution of logarithmic growth rate $\ln(x_{t+1}/x_t)$ for a Gaussian random walk, where the standard deviation of the stepsize depends on the distance from the origin as $\sigma(x) \sim x^{0.4}$. Statistics for three different bins of unit width centred around $x_0 = 20, 50$, and 100 are shown. (Note the only values $x_t > 1$ are evaluated for obvious reasons.) (b) Standard deviation of the distribution $\sigma_G(x_0)$ as a function of bin centre x_0 for the same process. The solid line is an exact power law of exponent -0.6.

where $0 \leqslant \beta \leqslant 1$ and $c \ll 1$. For small relative variations we get for the width

$$\sigma_r = cx^{\beta - 1} \; . \tag{9}$$

This consideration holds also if the variation ε has no strict functional dependence on x, but it is a random variable of a given probability density. We should assume only that the average $\langle \varepsilon_x \rangle = 0$, and the standard deviation of it scales as $\sigma(x) \sim x^{\beta}$. It is also clear that the shape of the probability density function does not play any role in scaling. In order to illustrate this fact, we generated a random walk $x_{t+1} = x_t \pm \varepsilon(x_t)$ with Gaussian random increments $\varepsilon(x_t)$ of zero mean and standard deviation $\sigma(x_t) = 0.1 x_t^{0.4}$.

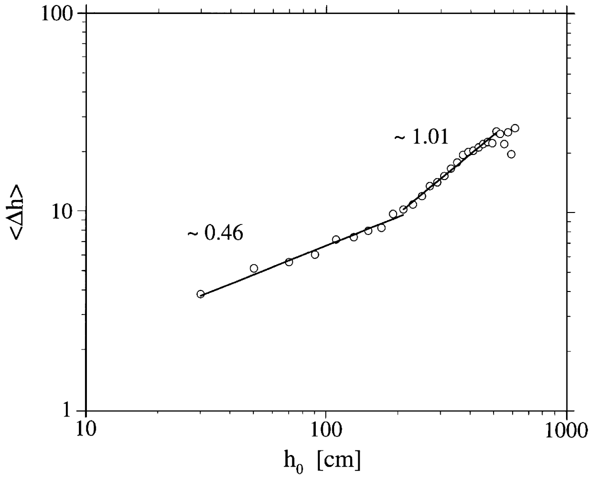

Fig. 5. Average daily absolute change $\langle \Delta h \rangle = \langle |h_{t+1} - h_t| \rangle$ as a function of actual water level h_0 for the Danube data on a double logarithmic scale.

The statistics for the logarithmic growth rate $r = \ln(x_{t+1}/x_t)$ is shown in Fig. 4. As expected, the shape is Gaussian for all bins (of unit width, particularly), and the scaling law Eq. (9) is restored.

Finally, we show in Fig. 5 the average daily changes as a function of water level for the Danube data. The statistics is not perfect, however two scaling regimes can be resolved with different exponents. The numerical values support our considerations: The evaluation of daily variations gives the same information as the statistics for rate of changes.

The above results illustrate that the characteristic patterns discovered by Stanley et al. [1] seem to be a general property of a wider class of phenomena than that of the economic time series. Strong time correlation is necessary to produce growth rate distributions centred around 1, however, this condition is automatically fulfilled for random walk-like processes of limited stepsize. Furthermore, if the amplitude of the fluctuations is limited, the partial distributions around different initial values must have different widths: The smaller the denominator, the larger the width. We have shown that scaling width for partial growth rate distributions can be a simple consequence of scaling step sizes. In our opinion, as is the case for water level variations and possibly of other sets of data also, see e.g. [9,10], the data presented by Stanley et al. manifests the characteristics of many strongly correlated stochastic processes.

Acknowledgements

I.M.J. thanks the Foundation for Research and Higher Education for a Zoltán Magyary scholarship. J.A.C.G. is partially supported by PRAXIS XXI (Portugal) and CAPES and CNPq (Brazil).

References

[1] M.H.R. Stanley, L.A.N. Amaral, S.V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M.A. Salinger, H.E. Stanley, Nature (London) 379 (1996) 804.

[2] L.A.N. Amaral, S.V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M.A. Salinger, H.E. Stanley, M.H.R. Stanley, J. Phys. I (France) 7 (1997) 621.

[3] S.V. Buldyrev, L.A.N. Amaral, S. Havlin, H. Leschhorn, P. Maass, M.A. Salinger, H.E. Stanley, M.H.R. Stanley, J. Phys. I (France) 7 (1997) 635.

[4] L.A.N. Amaral, S.V. Buldyrev, S. Havlin, M.A. Salinger, H.E. Stanley, Phys. Rev. Lett. 80 (1998) 1385.

[5] Y. Lee, L.A.N. Amaral, D. Canning, M. Meyer, H.E. Stanley, Phys. Rev. Lett. 81 (1998) 3275.

[6] H.E. Hurst, R.P. Black, Y.M. Simaika, Long-Term Storage: An Experimental Study, Constable, London, 1965.

[7] Hungarian National Bureau for Water Management, Budapest.

[8] Die Donau und ihr Einzugsgebiet, Regionale Zusammenarbeit der Donauländer, 1986, in German.

[9] I.M. Jánosi, G. Vattay, Phys. Rev. A 46 (1992) 6386.

[10] E. Koscielny-Bunde, A. Bunde, S. Havlin, Y. Goldreich, Physica (Amsterdam) A 231 (1996) 393.