Stochastic modeling of daily temperature fluctuations

Andrea Király and Imre M. Jánosi
Department of Physics of Complex Systems, Eötvös University, P.O.Box 32, H-1518 Budapest, Hungary
(Received 26 November 2001; revised manuscript received 15 February 2002; published 26 April 2002)

Classical spectral, Hurst, and detrended fluctuation analysis have been revealed asymptotic power-law correlations for daily average temperature data. For short-time intervals, however, strong correlations characterize the dynamics that permits a satisfactory description of temperature changes as a low order linear autoregressive process (dominating the texts on climate research). Here we propose a unifying stochastic model reproducing correlations for all time scales. The concept is an extension of a first-order autoregressive model with power-law correlated noise. The inclusion of a nonlinear ‘‘atmospheric response function’’ conveys the observed skew for the amplitude distribution of temperature fluctuations. While stochastic models cannot help to understand the physics behind atmospheric processes, they are capable to extract useful features promoting to benchmark physical models, an example is shown. Possible applications for other systems of strong short-range and asymptotic power-law correlations are discussed.

DOI: 10.1103/PhysRevE.65.051102 PACS number(s): 05.40.−a, 05.45.Tp, 89.75.Da, 95.75.Wx

I. INTRODUCTION

Climate research aims to identify and utilize persistent features of the atmosphere for predicting climatic anomalies. Any substantial anomaly in temperature or precipitation may result from a relatively small number of significant weather events localized in space and time. That is why numerical models for the coupled atmosphere-ocean system retain fine temporal and spatial scales even if the goal is long-term global prediction [1]. A better physical understanding gradually achieved by modeling requires a proper description for correlation properties of local variables, such as daily average temperature, on a scale spanning from days to decades.

Various methods are used to characterize quantitatively the fluctuations of high frequency meteorological data. Power density spectra [2] have been routinely computed for decades. Pelletier [3] determined the temperature spectra for hundreds of stations and ice core records, and identified different power-law behavior for continental and maritime locations. Koscielny-Bunde *et al.* [4] observed a near universal exponent value in the fluctuations of daily temperatures, however, Talkner and Weber [5,6] found differences depending on the altitude of the meteorological station. In general, scaling can be identified asymptotically from the longest available records only, shorter-time correlations are almost fully explained by using first-or second-order linear autoregressive models [7]. Here we propose a unified description reproducing the observed correlation properties both for short and long times.

The data we use for illustration are historical records of daily mean temperatures measured at 16 weather stations in Hungary in the period from 1 January 1951 until 31 December 1989. Amplitude distributions, power spectra, and autocorrelation functions are analyzed earlier essentially for the same data set by Jánosi and Vattay [8]. Note that clear scaling was observed only for shorter times with an exponent characteristic for Markovian random walk processes.

II. DFA ANALYSIS AND MODELING

Here we exploit the method of detrended fluctuation analysis (DFA) [9] that has proven useful in revealing the extent of long-range correlations in time series. Power-law correlations are obtained for diverse systems such as cardiac dynamics [10], DNA sequences [9,11], economic time series [12], and meteorology [4,5,13,14]. Recently, Talkner and Weber [5], and Heneghan and McDarby [15] pointed out that DFA and traditional power spectra (see, e.g., [2]) provide equivalent characterizations of correlated stochastic signals, apart from that DFA can effectively filter out slow trends. The situation for stationary signals is similar to the connection between power spectra and autocorrelation functions manifested by the Wiener-Khintchin theorem: the information mathematically is the same, but in many cases spectral density functions are more sensitive and better exploratory tools for real data (see, e.g., [5,7]). From this point of view, recent empirical analyses on synthetic time series with different correlation properties and background trends by Talkner and Weber [5], Kantelhardt *et al.* [16], and Hu *et al.* [17] are very illuminating.

It is easy to summarize the methodology of DFA based on the theoretical papers [4,5,9,15–17]. We consider a fluctuating time series $x_i, (i=1,\ldots,N)$ sampled at equidistant times $i\Delta t$. We assume that x_i are increments of a random walk process around the average $\langle x\rangle = N^{-1}\Sigma_{i=1}^{N} x_i = 0$, thus the ‘‘trajectory’’ or ‘‘profile’’ of the signal is given by

$$y_j = \sum_{i=1}^{j} x_i . \tag{1}$$

We divide the profile into nonoverlapping segments of equal length n indexed by $k=1,\ldots,[N/n]$. In each segment, the local trend is fitted by a polynomial of order p $f_k^{(p)}(y)$, and the profile is detrended by subtracting this local fit: $Y_i^{(p)} = y_i - f_k^{(p)}(y_i)$ A possible measure of fluctuations can be given by the root mean square

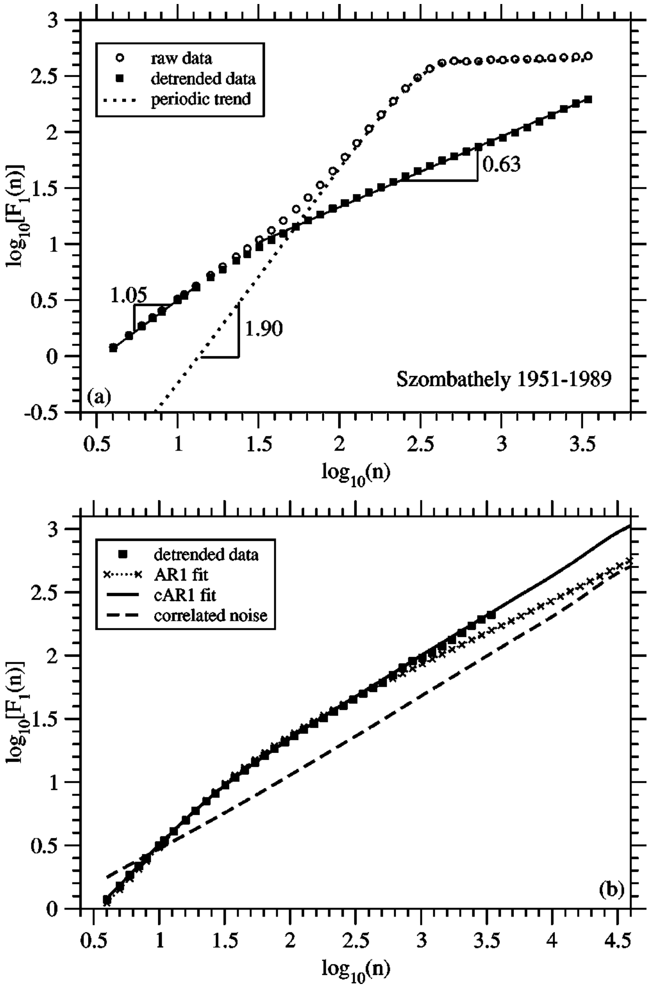

FIG. 1. (a) DFA1 results for the raw mean-temperature data (empty circles), temperature anomaly series (squares), and the harmonic component Eq. (4) (dotted line). Characteristic slopes are indicated. (b) DFA1 results for temperature anomalies (squares), fitted AR1 series Eq. (6) (crosses), CAR1 process driven by power-law correlated noise Eq. (13) (heavy line), and pure power-law correlated noise with fitted amplitude distribution (dashed line), see text.

$$F_p(n)=\sqrt{\frac{1}{n[N/n]}\sum_{i=1}^{n[N/n]}(Y_i^{(p)})^2} \tag{2}$$

for a given segment length n. A power-law relationship between $F_p(n)$ and n indicates scaling with an exponent δ (DFAp exponent),

$$F_p(n)\sim n^{\delta}. \tag{3}$$

Figure 1(a) shows the results of DFA1 analysis for a given station. All the other series give very similar curves: Hungary (in the middle of the Carpathian Basin) can be considered as climatically uniform. We emphasize that higher order DFA tests did not reveal any further structure in the time series. The curve for raw data [empty circles in Fig. 1(a)] indicates a strong periodic background in complete agreement with the analysis of Hu *et al.* [17]. This yearly trend can be removed with the seasonal variation function

$$\overline{T}(d)=T_{av}+A\cos\left(\frac{2\pi}{365}d+\varphi\right), \tag{4}$$

where $\overline{T}(d)$ is the average mean temperature of a given calendar day $d=1,\ldots,365$ (leap days are omitted), A and φ are amplitude and phase parameters, T_{av} is the long-time average temperature. The temperature anomaly series ΔT_i is given by the difference between the actual temperature and the average temperature of that particular day: $\Delta T_i=T_i-\overline{T}(d)$. DFA1 analysis for the anomaly data [heavy squares in Fig. 1(a)] reveals gradually decreasing level of correlations, the asymptotic exponent $\delta=0.63\pm0.02$ is consistent with other reported values for continental climate [4–6,14]. Note that the overall mean-temperature variance is dominated by seasonal changes, as clearly indicated by the horizontal plateau in Fig. 1(a). The same level of variance for temperature anomalies could be extracted from a ~ 200-year-long time series, if scaling with the same exponent holds.

Short-time correlations in meteorological records are usually explained by low order autoregressive processes [7]. One can assume, e.g., that the dynamics obeys a first-order ordinary linear differential equation

$$a_1\frac{dx(t)}{dt}+a_0x(t)=\xi(t), \tag{5}$$

where $\xi(t)$ represents uncorrelated Gaussian noise of unit variance, and a_0 and a_1 are constants. Standard time discretization yields a first-order autoregressive (AR1) process

$$x_i=\alpha_1 x_{i-1}+\epsilon\xi_i \tag{6}$$

with $\alpha_1=a_1/(a_0+a_1)$ and $\epsilon=1/(a_0+a_1)$. An important property of such a process is that the autocorrelation function $C(\tau)=\langle x_{i+\tau}x_i\rangle$ decays as

$$C_{AR1}(\tau)=\alpha_1^{\tau}, \tag{7}$$

where stationarity condition warrants $\alpha_1<1$. Fitting the parameters α_1 and ϵ for an empirical data set is a trivial task, all statistical software packages include this feature [18].

DFA1 test for a simulated time series with fitted AR1 parameters is shown in Fig. 1(b) (crosses). Numerical values are $\alpha_1=0.805\pm0.010, \epsilon=2.1\pm0.19$, the errors reflect weather-station dependence. It is remarkable that the changing DFA slope is explained by such a simple process, deviation is visible only for the largest window sizes n. The asymptotic DFA slope for an AR1 process is 1/2 as the consequence of uncorrelated noise ξ. For the sake of comparison, we show the DFA1 result for a pure power-law correlated (colored) noise [dashed line in Fig. 1(b)] with an amplitude distribution fitted to the empirical data. Its autocorrelation function reads

$$C_{cn}(\tau)=\tau^{2\rho-1}, \tag{8}$$

where $\rho \in (0,0.5)$. Power-law correlated data sets were generated with the algorithm developed by Pang, Yu, and Halpin-Healy [19]. We verified that the DFA1 exponent δ depends on the ρ parameter of the autocorrelation function Eq. (8) as expected [4,5]

$$\delta_{cn} = \rho + \frac{1}{2}. \tag{9}$$

It is clear from Fig. 1(b) that a pure colored noise strongly underestimates the variance (note again that the amplitudes are fitted), but the asymptotic slope is well reproduced.

One of the main findings of Hu *et al.* [17] is that DFA results on signals with different correlation properties and background trends can be fully explained by the assumption of variance superposition. Based on this observation, we propose a simple extension of the AR1 process Eq. (6) in order to reproduce the observed correlation properties for temperature time series. We assume that

$$x_i = \alpha_1' x_{i-1} + \epsilon \eta_i, \tag{10}$$

where the noise term η is power-law correlated according to Eq. (8) with a Gaussian amplitude distribution $P(\eta) = (1/\sqrt{2\pi})\exp(-\eta^2/2)$. We expect for such a colored noise driven autoregressive (CAR1) process that correlations for short times are determined by the autoregressive part according to Eq. (7), but for long times the correlation properties of the noise η will dominate. Indeed, the result shown in Fig. 1(b) (heavy line) is rather convincing, DFA slopes for the observed data are reproduced for all window sizes.

We emphasize here that the noise term η in Eq. (10) has a "true" power-law autocorrelation upto the sample size. This approach is different from the traditional modeling of colored noise by means of an autoregressive moving average process given by

$$x_n = \sum_{i=1}^{M} \alpha_i x_{n-i} + \sum_{i=0}^{N} \epsilon_i \xi_{n-i}, \tag{11}$$

where long-memory effects can be modeled, but correlations have always finite range for any finite N and M [2].

The fitting procedure of Eq. (10) for empirical data needs some remarks. First of all, it is obvious that a power-law correlated noise $x_i = \eta_i$ cannot be represented as an autoregressive process of any order. Nevertheless, standard ARp fitting algorithms detect linear correlations for such data, too. Figure 2 shows the results for $p=1$, i.e., the relationship

$$\eta_i = c\,\eta_{i-1} + \xi_i \tag{12}$$

is tested. For small and large ρ values [see Eq. (8)] the variance is very large (different realizations can be produced from different random seeds), however, the empirical trend allows an estimate for the apparent AR1 coefficient of colored noise. Variance superposition assumption yields

$$x_i = (\alpha_1 - c) x_{i-1} + \epsilon \eta_i, \tag{13}$$

where α_1 and c are the fitted AR1 coefficients for the measured time series and for the computer generated noise signal, respectively. Summarizing, the steps of the fitting procedure are the following: (i) extract AR1 parameters α_1 and ϵ from the temperature record; (ii) measure asymptotic DFA slope; (iii) construct power-law correlated noise series with the desired DFA slope; (iv) measure apparent AR1 coefficient c for the noise; (v) produce a data set by Eq. (13). The result is shown in Fig. 1(b) with heavy line.

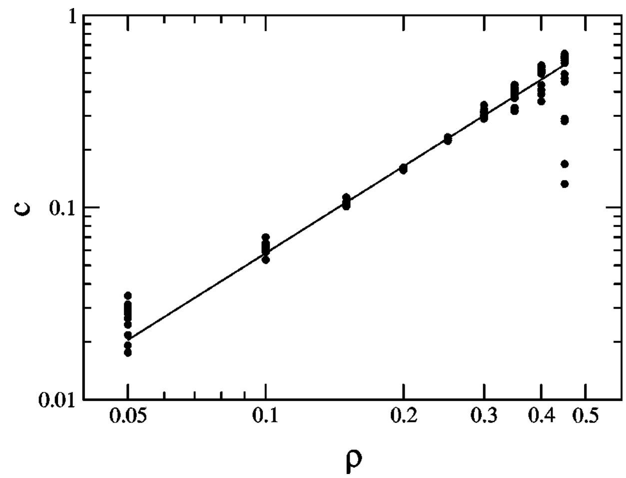

FIG. 2. Fitted AR1 coefficient c [see Eq. (12)] for power-law correlated series with different autocorrelation parameter ρ [see Eq. (8)]. Several realizations with different random seeds are evaluated. The straight line (not fitted) obeys $c = 2\rho^{3/2}$.

III. ATMOSPHERIC RESPONSE FUNCTION

Further details can be revealed by direct statistical tests of the basic assumptions. In Fig. 3 we plotted the average temperature step $\langle T_{i+1} - T_i \rangle$ and its standard deviation σ as a function of temperature anomaly ΔT_i (recall that this is the measure of deviation from the long-term average temperature for the particular calendar day). Apparently this function represents a negative feedback: the larger the anomaly, the more probable is a step backward on the next day. It is remarkable that this "atmospheric response function" is strongly asymmetric (positive excursions are hindered stronger) and strictly nonlinear. The best empirical representations we found are a fifth-order polynomial for $\langle T_{i+1} - T_i \rangle$ with coefficients $b_1 = -0.1898$, $b_2 = -0.002\,141\,3$, $b_3 = 0.000\,314\,8$, $b_4 = -3.2005e-05$, and $b_5 = -4.3807e-06$, and a quadratic function $\sigma(\Delta T) = 2.049 - 0.0058\Delta T + 0.0094(\Delta T)^2$. By means of this response function the process Eq. (13) can be refined as

$$\Delta T_{i+1} = [f(\Delta T_i) - c]\Delta T_i + \sigma(\Delta T_i)\,\eta_i, \tag{14}$$

where the functional forms $f(\Delta T_i)$ and $\sigma(\Delta T_i)$ are obtained from the fits shown in Fig. 3. Note that a linear approximation around the origin $f(\Delta T_i) = -(1-\alpha_1)$ and a constant $\sigma(\Delta T_i) = \epsilon$ give back the linear process Eq. (13).

The first-order nonlinear autoregressive process with correlated noise (NLCAR1) given by Eq. (14) results in a DFA1 curve indistinguishable from that of Eq. (13). However, it describes much better the empirical probability distribution for temperature fluctuations, as shown in Fig. 4.

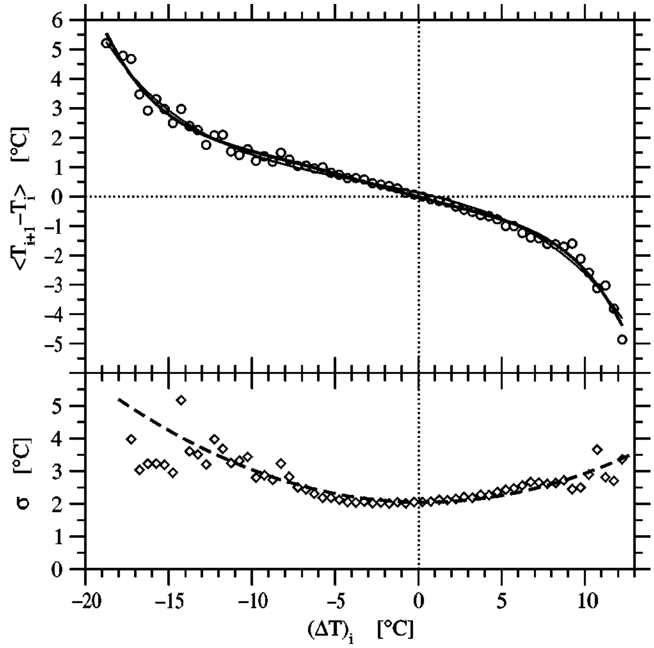

FIG. 3. The components of the empirical ‘‘atmospheric response function’’ $f(\Delta T)$ for the 16 weather stations in Hungary: the average temperature step $\langle T_{i+i}-T_i \rangle = f(\Delta T_i)\Delta T_i$ (top) and its standard deviation σ (bottom) as a function of temperature anomaly ΔT_i. Cubic (thin line) and fifth-order polynomial (thick line) fits for the temperature steps, as well as a quadratic fit for σ are indicated (see text).

We evaluated both basic representations (AR1 and NLCAR1) by means of the more traditional power spectral and autocorrelation methods. The difference between power spectra (not shown here) are not very large, they reproduce the gradually increasing (negative) slope of the empirical data. However, the spectrum of an AR1 process flattens out for small frequencies by rule, while the nonlinear process with power-law correlated noise Eq. (14) approximates this range much better. The difference between the autocorrelation functions is more pronounced (Fig. 5). The measured autocorrelation for the AR1 process can be perfectly fitted by Eq. (7), the NLCAR1 model reproduces the slow decay for the empirical data up to a time lag of five months (further comparison is not possible because of the noise level of real data).

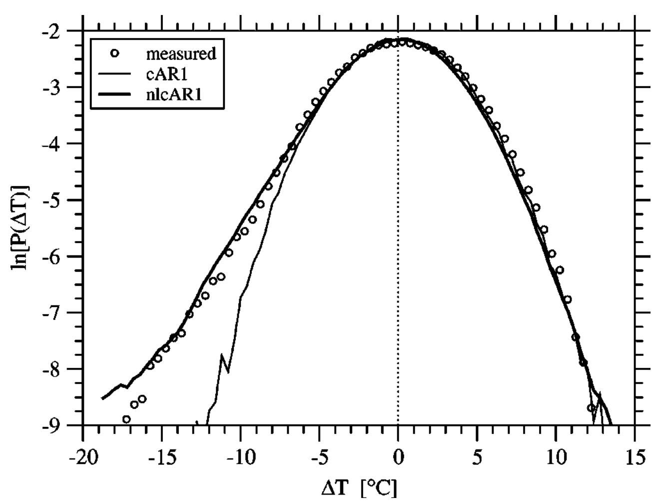

FIG. 4. Probability distribution of temperature fluctuations measured for the 16 stations (circles), fitted by a CAR1 process Eq. (13) (thin line), and by NLCAR1 Eq. (14) (thick line).

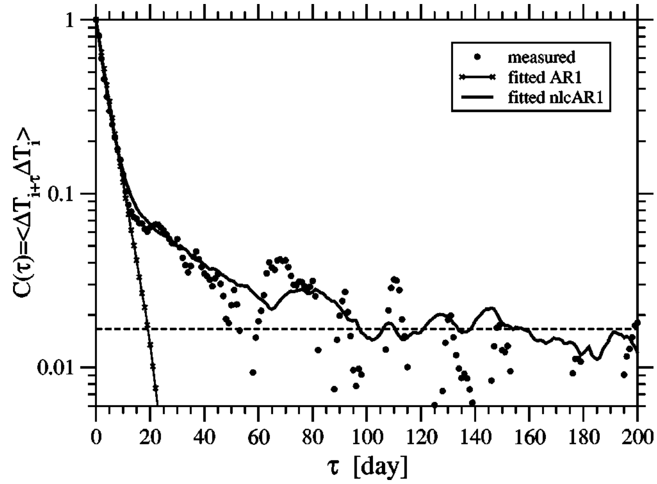

FIG. 5. Autocorrelation function for measured data at station Békéscsaba, for AR1 fitted series Eq. (6), and fitted by Eq. (14). Dashed line indicates the 95% confidence limit for the empirical data.

IV. DISCUSSION

At first we discuss shortly the relevance of a more accurate model representation for data of asymptotic power-law correlations. As for daily temperature fluctuations, the simplest AR1 process performs rather well by reproducing DFA1 curve [Fig. 1(b)], power spectrum, and even autocorrelation for a few days (Fig. 5). Nevertheless it is obvious that many important problems, related, e.g., to climatic changes, involve much longer time scales than a couple of days or months. In an otimal case, computed data can be compared directly with observations. Govindan *et al.* [14] illustrated that four state-of-the-art coupled atmosphere-ocean models used to estimate future warming failed to reproduce the correlation properties of two long time series of monthly mean temperatures, even for the time interval covered by observations.

An improved representation of fluctuating signals can help to extract features that are directly not available from records of limited lengths. As an example, we computed the so called run-length distribution [7] for the empirical and model data. The run-length distribution gives the probability $P(L)$ for a sequence of temperature anomalies of the same sign with length L, i.e., the statistics of zero crossing. (Between two zero-crossing events of L time units the temperature remains either above or below the long-time average value.) The result is plotted in Fig. 6. If we consider daily temperature anomalies, there is no difference between the AR1 and NLCAR1 models [Fig. 6(a)], since short-time correlations are dominated by the autoregressive part of the process. Clear difference appears on a longer time scale. Figure

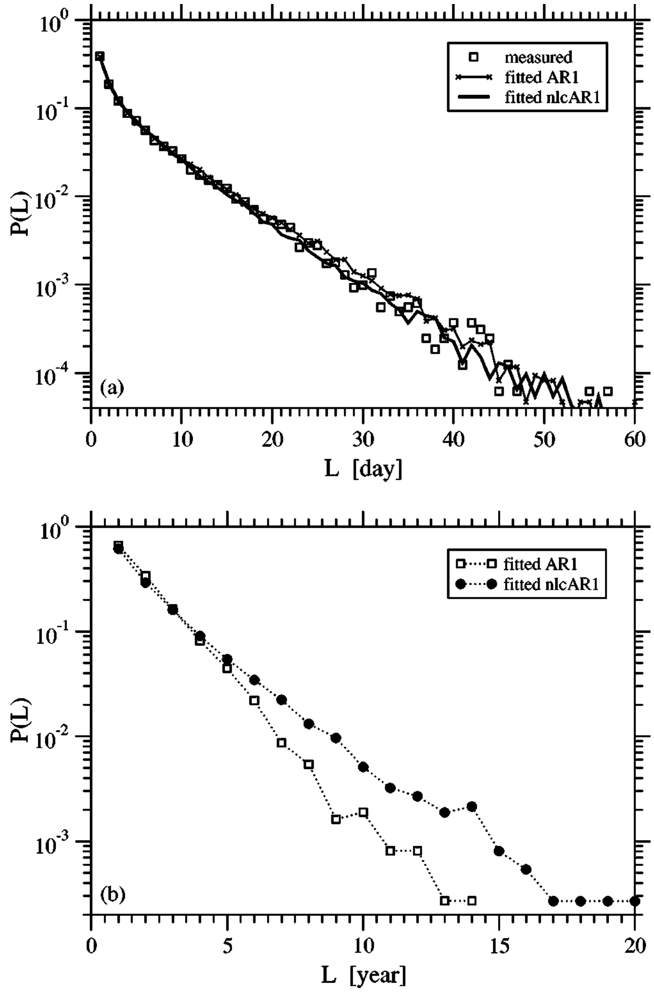

FIG. 6. (a) Normalized probability density $P(L)$ of run-length L for daily average temperature anomalies. (b) The same as above for annual average temperatures.

6(b) shows the probability distribution for excursions in annual average temperatures (empirical data are too short for such a statistics). The NLCAR1 model process with power-law correlated noise indicates definitely higher probabilities for long excursions of annual average temperature anomalies.

This modeling procedure can be applied for any process obeying strong correlations for short times and asymptotic scaling. A particularly well discussed example in the literature is related to hydrological records [20,21]. Actually, the first systematic study of asymptotic power-law correlations is given in the classical work by Hurst *et al.* [22]. Modeling and prediction have very sophisticated tools in hydrology, nevertheless low order autoregressive representations are common, at least as a starting point [20].

We repeated our analysis for daily water levels (31 776 data points) of the river Danube recorded at Nagymaros (Hungary), for the general description of the data see Ref. [23]. The quantity of interest is the daily water level anomaly Δh_i [cm], that is the difference between the actual level and the long-time average value for the particular day. The simplest AR1 model obeyed weak convergence, however, the

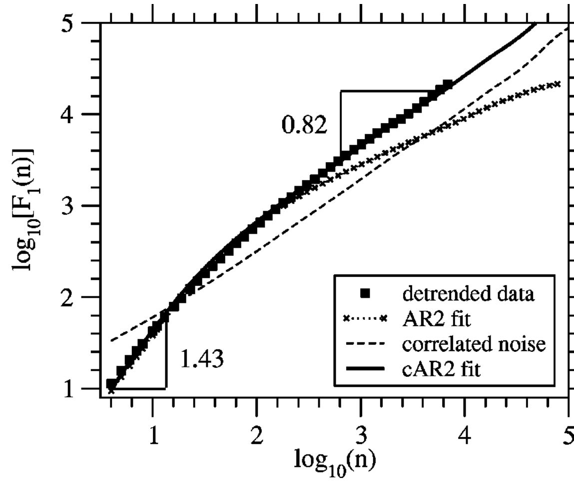

FIG. 7. Analysis of daily water level anomalies Δh for the Danube: DFA1 curves for the empirical data, for the best AR2 fit, for a pure power-law correlated noise and for the fitted CAR2 process Eq. (15). Characteristic slopes are indicated.

next order approximation (AR2) gave a satisfactory correlation behavior up to an interval of 5–6 months (see the crosses in Fig. 7). Here the difference between the asymptotic slopes of the DFA1 curves for the empirical data (squares in Fig. 7) and for the AR2 model is more pronounced. A pure power-law process η_i with $\rho=0.32$ [see Eq. (8)] reproduces the required asymtotic slope but fails for short times again (dashed line in Fig. 7). An extension of the AR2 Gaussian model

$$\Delta h_i=\beta_1\Delta h_{i-1}+\beta_2\Delta h_{i-2}+\gamma\eta_i \tag{15}$$

works quite well (solid line in Fig. 7), fitted numerical values are $\beta_1=1.547,\beta_2=-0.606$, and $\gamma=15.4$. Since we use this example merely to illustrate the applicability of the concept, more details are not presented here.

Finally we note that the linear approximations Eq. (10) and Eq. (15) are the first steps in the modeling only. They reproduce linear correlations detected by DFA, Hurst, spectral or autocorrelation methods. The response function Fig. 3 clearly indicates that the system is strictly nonlinear, thus the reproduction of finer details, such as the strongly skewed amplitude distribution, requires the incorporation of nonlinearities shown by Eq. (14).

In conclusion, our main result is a unified model explaining the correlation properties from days to decades for daily mean temperatures. Autoregressive processes have a huge number of variants [7], however, this is the first extension involving power-law correlated noise, as far as we know. We think that the analysis of asymptotic correlations and atmospheric response functions for different climates can contribute to the debate on possible universality in temperature fluctuations.

ACKNOWLEDGMENTS

The authors thank Gábor Papp for computational assistance. This work was supported by the Hungarian Science Foundation (OTKA) under Grant No. T032437.

[1] *Ocean Circulation and Climate: Observing and Modeling the Global Ocean*, edited by G. Siedler, J. Gould, and J. Church (Academic Press, New York, 2001).
[2] S.L. Marple, *Digital Spectral Analysis with Applications* (Prentice-Hall, New Jersey, 1987); D.B. Percival and A.T. Walden, *Spectral Analysis for Physical Applications* (Cambridge University Press, Cambridge, 1993).
[3] J.D. Pelletier, J. Clim. **10**, 1331 (1997).
[4] E. Koscielny-Bunde *et al.*, Phys. Rev. Lett. **81**, 729 (1998); Philos. Mag. B **77**, 1331 (1998).
[5] P. Talkner and R.O. Weber, Phys. Rev. E **62**, 150 (2000).
[6] R.O. Weber and P. Talkner, J. Geophys. Res., [Atmos.] **17**, 20 131 (2001).
[7] H. von Storch and F.W. Zwiers, *Statistical Analysis in Climate Research* (Cambridge University Press, Cambridge, 1999).
[8] I.M. Jánosi and G. Vattay, Phys. Rev. A **46**, 6386 (1992); I.M. Jánosi, G. Vattay, and A. Harnos, J. Stat. Phys. **93**, 919 (1998).
[9] C.K. Peng *et al.*, Nature (London) **356**, 168 (1992); C.K. Peng *et al.*, Phys. Rev. E **49**, 1685 (1994); H.E. Stanley *et al.*, Physica A **200**, 4 (1996).
[10] P.Ch. Ivanov *et al.*, Nature (London) **383**, 323 (1996); P.Ch. Ivanov *et al.*, Physica A **249**, 587 (1998); H.E. Stanley *et al.*, *ibid.* **270**, 309 (1999); S. Havlin *et al.*, *ibid.* **273**, 46 (1999); A. Bunde *et al.*, Phys. Rev. Lett. **85**, 3736 (2000); Y. Ashkenazy *et al.*, *ibid.* **86**, 1900 (2001).
[11] R.N. Mantegna *et al.*, Phys. Rev. Lett. **73**, 3169 (1994); S.V. Buldyrev *et al.*, Phys. Rev. E **51**, 5084 (1995); X. Lu, Z. Sun, H. Chen, and Y. Li, *ibid.* **58**, 3578 (1998).
[12] Y. Liu *et al.*, Physica A **245**, 437 (1997); N. Vandewalle and M. Ausloos, *ibid.* **246**, 454 (1997); Phys. Rev. E **58**, 6832 (1998); Y. Liu *et al.*, *ibid.* **60**, 1390 (1999); I.M. Jánosi, B. Janecsko, and I. Kondor, Physica A **269**, 111 (1999).
[13] M. Ausloos and K. Ivanova, Phys. Rev. E **63**, 047201 (2001).
[14] R.B. Govindan *et al.*, Physica A **294**, 239 (2001).
[15] C. Heneghan and G. McDarby, Phys. Rev. E **62**, 6103 (2000).
[16] J.W. Kantelhardt *et al.*, Physica A **295**, 441 (2001).
[17] K. Hu *et al.*, Phys. Rev. E **64**, 011114 (2001).
[18] See, e.g., the Dataplot project by J.J. Filliben and A. Heckert (http://www.itl.nist.gov/div898/software/dataplot/) or the MacAnova project by G.W. Oehlert and Ch. Bingham (http://www.stat.umn.edu/macanova/).
[19] N.N. Pang, Y.K. Yu, and T. Halpin-Healy, Phys. Rev. E **52**, 3224 (1995).
[20] R.L. Bras, and I. Rodriguez-Itube, *Random Functions and Hydrology* (Addison-Wesley, Reading, Massachusetts, 1985).
[21] *Non-linear Variability in Geophysics, Scaling and Fractals*, edited by D. Schertzer and S. Lovejoy (Kluwer, Norwell, 1991).
[22] H.E. Hurst, R.P. Black, and Y.M. Simaika, *Long-Term Storage: An Experimental Study* (Constable, London, 1965).
[23] I.M. Jánosi and J.A.C. Gallas, Physica A **271**, 448 (1999).