ARE WEATHER FLUCTUATIONS LONG-RANGE CORRELATED?

Armin Bunde[1], Shlomo Havlin[1,2], Eva Koscielny-Bunde[1,2], H. Eduardo Roman[3], and Hans-Joachim Schellnhuber[4]

[1]Institut für Theoretische Physik III, Universität Giessen, D-35392 Giessen, Germany

[2] Minerva Center and Department of Physics, Bar-Ilan University, Ramat Gan, 52900 Israel

[3] Dipartimento di Fisica, Università di Milano, Via Celoria 16, 20133 Milano, Italy

[4] Potsdam Institute for Climate Impact Research, D-14412 Potsdam, Germany

Abstract

We have studied long time daily temperature records (typically 100 years) obtained from 14 meteorological stations in Europe and North-America, from various climatological zones. To search for correlations between the fluctuations of the daily temperatures around their average values, we applied several methods (wavelets methods and detrended fluctuation analysis) that can systematically overcome nonstationarities in the data. In addition, we have employed the Fourier-transform technique and compared our results with those obtained from a direct calculation of the relevant temperature autocorrelation function $C(\ell)$.We also analysed the distribution of the daily temperature fluctuations. Our analysis suggests that the persistence, characterized by the autocorrelation function $C(\ell)$, is long-ranged. For ℓ above one week (the time scale of a typical weather situation) $C(\ell)$ approximately decays as a power law, $C(\ell) \sim \ell^{-\gamma}$, with the same exponent $\gamma \cong 0.7$ for all stations considered. This universal persistence law seems to be valid at least for one decade of years, but we cannot exclude the possibility that the range of the power-law correlations even exceeds the range of the temperature series considered.

1. Introduction

It is well known in meteorology that the weather is persistent on short time scales. If one day is sunny and warm, there is a higher probability that the next day remains the same, and any "sophisticated" weather forecast must be better than the "trivial" one that predicts that the weather of tomorrow is the same as the weather of today [1].

To quantify the persistence, we have analysed the records of the maximum daily temperatures T_i of the following weather stations (the length of the records is written within the parentheses): Albany (90 y), Brookings (99 y), Huron (55 y), Luling (90 y), Melbourne (136 y), New York City (116 y), Pendleton (57 y), Prague (218 y), Sidney (117 y), Spokane (102 y), Tucson (97 y), Vancouver (93 y), Moscow (115 y), and St. Petersburg (111 y). The stations have been chosen randomly and represent the different climatological zones. We review the results from [2] and extend them using further complementary methods, such as Fast Fourier Transforms.

For each weather station, we consider the daily maximum temperature T_i. The total number N of days i available for a given weather station ranges typically from 20 000 days (Huron) to 80 000 days (Prague). For eliminating the periodic seasonal trends, we have considered the departures of the T_i, $\Delta T_i = T_i - \bar{T}_i$, from the mean maximum daily temperature $\bar{T}_i$ for each calendary date i, say 1st of April, which has been obtained by averaging over all years in the temperature series. We have used several mathematical techniques (random walk theory, wavelets, and Fourier-analysis) to analyse the ΔT_i time series. Our analysis suggests that the temperature fluctuations at days i and $i + \ell$ are long-range power-law correlated, i.e., the correlation function behaves like

$$C(\ell) \equiv \langle \Delta T_i \Delta T_{i+\ell} \rangle \sim \ell^{-\gamma} \tag{1}$$

with an apparently universal exponent $\gamma \cong 0.7$ for all weather stations considered. The brackets in (1) denote an average over all pairs of temperature data separated by ℓ days,

$$\langle \Delta T_i \Delta T_{i+\ell} \rangle = \frac{1}{N-\ell} \sum_{i=1}^{N-\ell} \Delta T_i \Delta T_{i+\ell}. \tag{2}$$

¿From our results we can conclude that, within the pertinent error bars, the power-law correlations set in after about one week (which is the typical time scale for a weather situation) and range at least over one decade of years. We did not find any evidence for a crossover to uncorrelated behavior at very large time scales, and cannot exclude the possibility that the range of the power-law correlations is larger than the range of the temperature series considered. In contrast to the universal behavior of the correlations, the distribution $\mathcal{H}(\Delta T)$ of the temperature variations does not exhibit a universal form.

The paper is organized as follows: In Section 2, we describe the methods for analysing the temperature series. In Section 3, we present representative results for the temperature fluctuations from various meteorological stations considered, and in Section 4, we end up with a discussion of the results.

2. Time series analysis techniques

For analysing the time series, we have applied, apart from the direct calculation of $C(\ell)$ (Eqs. (1,2)), three techniques: Detrended fluctuation analysis (DFA), wavelets methods (WL) and the fast Fourier transformation method (FFT). These methods have been found useful in the analysis of sequences of various types [3-7]. They have been instrumental, in particular, in detecting long-range correlations in a series, when (undesired) nonstationarities are present and it is hard to distinguish a priori between trends and fluctuations. The DFA has been developed originally by Peng et al. [3] to investigate long-range correlations in heart beat intervals and in DNA sequences, where nonstationarities similar to the nonstationarities in the temperature fluctuations [8] can occur. The wavelets methods in general are very convenient techniques to investigate fluctuating signals [4]. Wavelets techniques have been used by Arneodo et al. [5] to study the correlations in the noncoding part of DNA employing Gaussian wavelets and 1st-, 2nd-, and 3rd-order derivatives, by

Ivanov et al. [6] to study heart beat variability, and by Kantelhardt et al. [8] to analyse the fluctuations of random distribution functions within the context of multifractality employing discrete wavelets.

The DFA and the wavelets techniques will not be used here to analyse directly the ΔT_i or the correlation function $C(\ell)$. Instead, for reducing the level of noise present in the finite temperature series, we consider the "temperature profile"

$$Y_n \equiv \sum_{i=1}^{n} \Delta T_i. \tag{3}$$

The fluctuations of the profile, on a given length scale ℓ, are related to the correlation function $C(\ell)$. For the relevant case of long-range power-law correlations, $C(\ell) \sim \ell^{-\gamma}$, $0 < \gamma < 1$, the fluctuations $F(\ell)$ increase by a power law [9],

$$F(\ell) \sim \ell^{\alpha}, \quad \alpha = 1 - \gamma/2. \tag{4}$$

For uncorrelated data (as well as for short-range correlations $\gamma \geq 1$), we have $\alpha = 1/2$.

To find how the fluctuations scale with ℓ, we divide the profile into nonoverlapping segments of size ℓ. We calculate the fluctuations $F_\nu(\ell)$ in each segment ν and obtain $F(\ell)$ by averaging over all segments. The DFA and the wavelet methods differ in the way the fluctuations are measured and possible nonstationarities are eliminated. We begin with the DFA.

2.1 Detrended Fluctuation Analysis (DFA)

We first divide each series of N successive daily temperatures into $K_\ell = [N/\ell]$ nonoverlapping subsequences of size ℓ starting from the beginning and K_ℓ nonoverlapping subsequences of size ℓ starting from the end of the considered temperature series. For each subsequence ν we determine the best linear fit of the profile, and calculate the standard deviation $F_\nu^2(\ell)$ of the profile from this straight line. This way, we eliminate the influence of possible linear trends on scales larger than the segment [3]. Then we average $F_\nu(\ell)$ over the K_ℓ subsequences obtained by starting from one end of the series, and over the K_ℓ subsequences obtained by starting from the other end.

2.2 Wavelets Techniques

The wavelets methods are more advanced methods, which are based on the determination of the mean values $\bar{Y}_\nu(\ell)$ of the profile in each segment ν (of length ℓ), and the calculation of the fluctuations between neighboring segments. First we divide, as above, the temperature series in $2 \times K_\ell$ subsequences. Then we determine in each segment ν the mean values $\bar{Y}_\nu(\ell)$ of the profile. The various techniques we have used here differ in the way the fluctuations between the average profiles are treated and possible nonstationarities are eliminated.

(i) In the first-order wavelets method (WL1), we simply determine the fluctuations from the first derivative

$$F_\nu^2(\ell) = (\bar{Y}_\nu(\ell) - \bar{Y}_{\nu+1}(\ell))^2. \tag{5}$$

This way, trends in the profile of a weather station originating, e.g., from an approximately linear increase of temperature due to urban development around the station, are not eliminated.

(ii) In the second-order wavelets method (WL2), we determine the fluctuations from the second derivative

$$F_\nu^2(\ell) = (\bar{Y}_\nu(\ell) - 2\bar{Y}_{\nu+1}(\ell) + \bar{Y}_{\nu+2}(\ell))^2. \tag{6}$$

So, if the profile consists of a trend term linear in ℓ and a fluctuating term, the trend term is eliminated.

(iii) In the third-order wavelets method (WL3), we determine the fluctuations from the third derivative

$$F_\nu^2(\ell) = (\bar{Y}_\nu(\ell) - 3\bar{Y}_{\nu+1}(\ell) + 3\bar{Y}_{\nu+2}(\ell) - \bar{Y}_{\nu+3}(\ell))^2. \tag{7}$$

By definition, WL3 eliminates linear and parabolic trend terms in the profile.

Finally, we average in each case the quantity $F_\nu(\ell)$ over the $2 \times K_\ell$ subsequences of the temperature series considered.

Methods (i-iii) are called wavelets methods, since they can be interpreted as transforming the profile by discrete wavelets representing first-, second-, and third-order cumulative derivatives of the profile [2]. The first-order wavelets are known in the literature as Haar wavelets (see e.g. [8]). In principle, one could also use different shapes of the wavelets (e.g. Gaussian wavelets with width ℓ), which have been used by Arneodo et al. [5] to study long-range correlations in DNA. We believe that discrete wavelets are more suitable to study temperature fluctuations for the following reason: Instead of studing the daily temperature deviations, one could start with the annual temperature deviations from the average temperature value, thus neglecting correlations on scales lower than one year. It can be easily seen that the discrete-wavelets method, applied to the annual temperature departures, is identical to the discrete-wavelets method, applied to the daily temperature, for $\ell = n \times 365$, where n is the number of years.

Finally, we discuss the fast Fourier transformation method (FFT) in the form applied by Buldyrev et al. [10] to DNA sequences, by which long-range correlations in the temperature fluctuations can also be investigated.

2.3 Fast Fourier transformation method (FFT)

Similar to DFA and wavelets methods, we divide each series of N daily temperatures into $K = [N/L]$ nonoverlapping subsequences of size L starting from the beginning and K nonoverlapping subsequences of size L starting from the end. In the analysis, we have chosen $L = 2048$. In this way, only correlations well below the length L of the subsequences can be detected. For each subsequence, we compute the Fourier transform

$$q_f = \sum_{k=1}^{N} \Delta T_k \ \exp[i(k-1)2\pi f/N] \tag{8}$$

and the power spectrum

$$S(f) = |q_f|^2 + |q_{N-f}|^2. \tag{9}$$

Then we average $S(f)$ over the K subsequences of the temperature series considered, obtained by starting from one end, and the K subsequences obtained by starting from the other end. If the temperature series has long-range power-law correlations, then

$$S(f) \sim f^{-(1-\gamma)}. \tag{10}$$

Similar to the direct determination of $C(\ell)$ from (1) and (2), the FFT method cannot eliminate trends.

In this paper, we have used both the trend-eliminating methods described above as well as the FFT and the direct calculation of $C(\ell)$ to learn about the laws of persistence governing atmospheric variability.

3. Analysis of the daily temperature fluctuations

For an illustration of the methods, we begin with the analysis of an artificial temperature series $\{\Delta T_i\}$ consisting of 40 000 data, which is long-range correlated according to Eq. (1) with a correlation exponent $\gamma = 0.7$. Figure 1 shows the results of the fluctuation analysis. For DFA and the wavelet methods, we obtain straight lines (in the double-logarithmic plots) with the predicted slope $\alpha = 0.65$. Note that for ℓ-values above 4 000 (10 percent of the sequence length) the fluctuation functions start to scatter due to lack of statistics. For the power spectrum in the Fourier-transform analysis, we obtain, in the double logarithmic representation, a straight line with the predicted slope $-(1-\gamma) = 2\alpha - 1 = -0.3$, and the direct correlation function $C(\ell)$ shows also the expected behavior $C(\ell) \sim \ell^{-\gamma}$. We study next real temperature sequences.

We have considered temperature data from 14 meteorological stations. For each station we have analysed the daily temperature variations $\Delta T_i = T_i - \langle T_i \rangle$, where T_i is the maximum temperature of day i and $\langle T_i \rangle$ is the mean maximum temperature of this particular date of the year averaged over all years in the considered temperature series. We show representative results for Prague (Fig. 2), Luling (Fig. 3) and Spokane (Fig. 4). We begin the analysis with the temperature series $\{\Delta T_i\}$ for Prague which is the largest series (218 y) in this study.

Figure 2 shows the fluctuation analysis for Prague obtained from the three methods. In the log-log plot, the DFA and wavelets curves are approximately straight lines for $\ell > 10$ days, with a slope $\alpha \cong 0.65$. For ℓ of the order of few days, the slope is a little larger. This result suggests, that there exists long-range persistence expressed by the power-law decay of the correlation function, with an exponent $\gamma \cong 0.7$. A closer look at these curves indicates that the effects of trends and correlations can be, to a certain extent, distinguished by the available methods. At about 10^3 days, the curves of FA and WL1 show a crossover towards a slightly larger exponent α. This behavior can be interpreted as the effect of the warming of Prague due to urban development. In contrast, DFA, WL2, and WL3 yield approximate straight lines until about 10^4 days above which the data start to scatter. The systematic crossover at about 10^3 days does not occur here, since DFA, WL2, and WL3 eliminate the (roughly) linear trend of warming. For the Fourier-transform analysis, we obtain, in the double logarithmic representation, a straight line with the slope $-(1-\gamma) = 2\alpha - 1 = -0.3$, consistent with the other methods. For f above $f \cong 100$, corresponding to ℓ smaller than roughly 10 days, we see a crossover towards a larger exponent, in agreement with the previous analysis. Since the power spectrum analysis is limited to 2048 days, we cannot see the influence of trends involved in WL1. The direct evaluation of the autocorrelation function (Fig. 2d) yields a consistent picture, $C(\ell) \sim \ell^{-\gamma}$. At very large time scales, scattering becomes dominant and hides the power-law behavior.

Figures 3 and 4 show the analogous results for the fluctuation functions for two cities in North America, from two different climatological zones, Luling from Texas (Fig. 3) and Spokane from Washington (Fig. 4). The curves have the same features as the curves for Prague, and the exponents α and γ seem to have almost the same values as for Prague.

When comparing Figs. 2-4 with the artificial curves from Fig. 1, one notes that apart from very small ℓ-values, the artificial curves look similar to the realistic detrended ones. Above an ℓ-value that is about 10 percent of the length N of the artificial series, the data start to scatter due to lack of statistics and the true range of correlations cannot be detected. Accordingly, the range of correlations cannot be detected from the data, but it is clear that the power-law correlations range at least over one decade. Since we obtain the same behavior for all stations considered, our results point to the possibility that there may be a universal law of persistence, with a universal exponent $\gamma \cong 0.7$.

Finally, we have studied the normalized distribution function $\mathcal{H}(\Delta T)$ of the temperature variations ΔT_i for the various meteorological stations. The distribution represents the number of days with ΔT_i in the interval $(\Delta T, \Delta T + \epsilon)$, with $\epsilon = 1\ ^o$C, divided by the total number of days. Figure 5 shows the result for four stations from two different climatological zones, for Moscow and St. Petersburg (continental climate), and Tucson (Arizona), and Luling (Texas) (arid climate). The figure also shows the corresponding Gaussian fits to the numerical data. Apparently, there is no universal behavior for the distribution function. While the distributions for Moscow and St. Petersburg, belonging to the same climatological zone according to Köppen's classification scheme, are reasonably well approximated by a Gaussian, strong deviations from a Gaussian are observed for Tucson and Luling. But this feature does not have consequences on the long-range behavior of the correlations, as we have demonstrated above.

4. Conclusions.

Our finding of long-range power-law persistence with a unique exponent α for different weather stations in different climate zones suggests that atmospheric variability may be governed by rather fundamental mechanisms. We name just a few candidates:

(i) A few years ago there were speculations about the existence of a low-dimensional (strange) attractor organizing the atmosphere's dynamics [11]. Such a dissipative structure might induce universal laws for fluctuation persistence.

(ii) The value for $\alpha \cong 0.65 \cong 2/3$ invokes reminiscences of Kolmogorov's famous 5/3 law. A first-principles explanation of our findings based on the dimensional analysis of fluid dynamics [12] would be a most intriguing but perhaps an unfeasible exercise. As a prerequisite, the spatial correlation of atmospheric patterns has to be taken into account.

(iii) The extremely long persistence of meteorological fluctuations implies that the coupling of atmospheric and oceanic processes has to be involved, as the latter rule the long-term dynamics of the system. Here we touch core questions dealt with in the CLIVAR programme of the World Climate Research Programme [13]. Our analysis may provide useful hints, if it is combined with a systematic investigation of the relation between fluctuation persistence at a geometric site and the characteristic of the relevant oceanic regime.

Furthermore, our results may be useful in the context of the current debate on atmospheric global warming as reviewed in the latest IPCC report [14]. While most scientists involved in the IPCC process hold that there is already empirical evidence for human interference with the climate, a few others strongly disagree [15,16]. Given the operable bulk of meteorological observations, the crucial point is to distinguish the “atmospheric signal” from the noise generated by the natural variability of the geophysical system. Climatologists try to circumvent this problem, e.g., by complementing empirical data with simulated ones obtained from coupled ocean-atmosphere circulation models [17].

The quantitative empirical laws we found might serve as a test for validating the quality of state-of-the-art computer simulations of short- to long-term atmospheric dynamics. If these simulations fail to reproduce the persistence characteristics detected here, then we should be concerned about their predictive power.

Acknowledgements:

We like to thank Professor Schönwiese for bringing several relevant references to our attention. This work has been supported by the German-Israeli Foundation.

LIST OF REFERENCES:

[1] O.M. Essenwanger, *General Climatology. Elements of Statistical Analysis*, Vol.1B of H.E. Landsberg (Ed.), World Survey of Climatology (Elsevier, Amsterdam, 1986); S.B. Newman, Weather and Forecasting **6**, 111 (1991)

[2] E. Koscielny-Bunde, A. Bunde, S. Havlin, H.E. Roman, Y. Goldreich, and H.J. Schellnhuber, Phys. Rev. Lett. 81, 729 (1998).

[3] C.K. Peng et al., Phys. Rev. E**49**, 1685 (1994)

[4] For a general review on wavelets, see M. Vetterli and J. Kovacevic, *Wavelets and Subbane Coding* (Prentice Hall, Englewood Cliffs NJ, 1995)

[5] A. Arneodo et al, Phys. Rev. Lett. **74**, 3293 (1995)

[6] P.Ch. Ivanov et al., Nature **383**, 323 (1996)

[7] J.W. Kantelhardt, H.E. Roman and M. Greiner, Physica A **220**, 219 (1995).

[8] see, e.g., C.D. Schönwiese and J. Rapp, *Climate Trend Atlas of Europe Based on Observations 1891-1990* (Kluwer, Dordrecht, 1997)

[9] A.-L. Barabasi and H.E. Stanley, *Fractal Concepts in Surface Growth* (Cambridge University Press, Cambridge, 1995)

[10] S.V. Buldyrev et al., Phys. Rev. E **51**, 5084 (1995).

[11] A series of contributions to Nature (vol. 311, (1984)), starting with C. Nicolis and G. Nicolis, Nature **311**, 529 (1984).

[12] see, e.g., U. Frisch, *Turbulence: The Legacy of A.N. Kolmogorov* (Cambridge Univ. Press, 1995).

[13] World Climate Research Programme, Newsletter No. 1 (Geneva, 1996)

[14] J.T. Houghton et al. (Eds.), *Climate Change 1995: The Science of Climate Change* (Cambridge Univ. Press, 1996).

[15] see, e.g., the articles by E. Masood in Nature **381**, 539, 546, 639 (1996)

[16] K. Hasselmann, Science **276**, 914 (1997)

[17] G.C. Hegerl et al., Journal of Climate **9**, 2281 (1996)

FIGURE CAPTIONS:

Fig. 1: Analysis of the sequence of 40 000 artificial random data. (a) Fluctuation function $F(\ell)$ versus ℓ obtained from DFA. (b) Fluctuation function $F(\ell)$ versus ℓ obtained from the wavelets methods: WL1 (circles), WL2 (squares), and WL3 (diamonds). In (a,b), the straight lines show the predicted slope $\alpha = 0.65$. For ℓ-values above 4 000 the obtained fluctuation functions start to scatter due to lack of statistics. (c) Power spectrum $S(f)$ versus f for $L = 2048$. The straight line has the predicted slope $-(1-\gamma) = 2\alpha - 1 = -0.3$. (d) Autocorrelation function $C(\ell)$ versus ℓ. The straight line has the predicted slope $-\gamma = 2\alpha - 2 = -0.7$.

Fig. 2: Analysis of the daily temperature variations ΔT_i (218 y) for Prague. (a) Fluctuation function $F(\ell)$ versus ℓ obtained from DFA. (b) Fluctuation function $F(\ell)$ versus ℓ obtained from the wavelets methods: WL1 (circles), WL2 (squares), and WL3 (diamonds). In (a,b), the straight lines have slope $\alpha = 0.65$. (c) Power spectrum $S(f)$ versus f for $L = 2048$. The straight line has the predicted slope -0.3. (d) Autocorrelation function $C(\ell)$ versus ℓ. The straight line has the predicted slope $-\gamma = 2\alpha - 2 = -0.7$.

Fig. 3: Analysis of the daily temperature variations ΔT_i (99 y) for Luling (Texas). (a) Fluctuation function $F(\ell)$ versus ℓ obtained from DFA. (b) Fluctuation function $F(\ell)$ versus ℓ obtained from the wavelets methods: WL1 (circles), WL2 (squares), and WL3 (diamonds). In (a,b), the straight lines have slope $\alpha = 0.65$. (c) Power spectrum $S(f)$ versus f for $L = 2048$. The straight line has the predicted slope -0.3. (d) Autocorrelation function $C(\ell)$ versus ℓ. The straight line has the predicted slope $-\gamma = 2\alpha - 2 = -0.7$.

Fig. 4: Analysis of the daily temperature variations ΔT_i (102 y) for Spokane (Washington). (a) Fluctuation function $F(\ell)$ versus ℓ obtained from DFA. (b) Fluctuation function $F(\ell)$ versus ℓ obtained from the wavelets methods: WL1 (circles), WL2 (squares), and WL3 (diamonds). In (a,b), the straight lines have slope $\alpha = 0.65$. (c) Power spectrum $S(f)$ versus f for $L = 2048$. The straight line has the predicted slope -0.3. (d) Autocorrelation function $C(\ell)$ versus ℓ. The straight line has the predicted slope $-\gamma = 2\alpha - 2 = -0.7$.

Fig. 5: The distribution $\mathcal{H}(\Delta T)$ of the temperature variations ΔT [^{o}C] for four meteorological stations: (a) Moscow (1880-1994, 115 years), (b) St. Petersburg (1884-1994, 111 years), (c) Tucson (Arizona) (1895-1991, 97 years) and (d) Luling (Texas) (1902-1991, 90 years). The lines are Gaussian fits, $P(\Delta T) = (2\pi\sigma^2)^{-1/2}\ \exp[-(\Delta T)^2/(2\sigma^2)]$ with: (a) $\sigma = 5.05$ ^{o}C, (b) $\sigma = 4.62$ ^{o}C, (c) $\sigma = 3.99$ ^{o}C and (d) $\sigma = 4.72$ ^{o}C.