Statistical analysis of climatic time series: uncertainty and insights

Demetris Koutsoyiannis[1] and Alberto Montanari[2]

[1]Department of Water Resources, National Technical University of Athens, Greece

[2]Department DISTART, Faculty of Engineering, University of Bologna, Italy

Abstract

Several studies of instrumental and proxy climatic time series have identified long-term persistence (LTP), which may reflect a long-term variability of several factors (solar forcing, volcanic activity, etc.) and, thus, can support a more consistent physical understanding and uncertainty characterization of climate. The implications of LTP in climatic research, especially in statistical questions and problems, may be substantial, but appear to be not fully understood or recognized. To offer insights on these implications, we demonstrate using analytical methods that the characteristics of temperature series, which appear to be consistent with the LTP hypothesis, imply a dramatic increase of uncertainty in statistical estimation and reduction of significance in statistical testing, in comparison with classical statistics. Therefore, we maintain that statistical analysis in climatic research should be revisited, in order not to derive misleading results, and simultaneously that merely statistical arguments do not suffice to verify the LTP (or another) climatic hypothesis.

Introduction

"Even if we would know everything, we should still have to derive statistical information from this knowledge in order to answer what are essentially statistical problems, such as explaining gas pressure or the intensity of spectral lines". (Karl Popper to Albert Einstein)

The increasing interest on statistical analysis of climatic records in the last two decades (Bloomfield, 1992; von Storch, 1995; von Storch and Zwiers, 1999; Cohn and Lins, 2005; Rybski et al., 2006; and many others) can be attributed, above all, to the well known detection problem (whether or not climatic changes have occurred) and the attribution problem (whether or not observed changes are related to anthropogenic forcings of the climate system). In several cases, such analyses have claimed the presence in the data of Long Term Persistence (LTP, also known as Hurst phenomenon, Joseph effect, long memory, long-range dependence, scaling behavior, and multi-scale fluctuation). LTP is a behavior defined on statistical grounds (see equation 2 below) and can be easily reproduced by appropriate stochastic models. However, this does not mean that LTP implies necessarily stochastic dynamics. For instance, it has been demonstrated that a simple deterministic nonlinear model (involving no random component) can produce trajectories exhibiting LTP (Koutsoyiannis, 2006). From a practical point of view, LTP indicates that the process is consistent with the presence of fluctuations on a range of timescales, which may reflect the long term variability of several factors such as solar forcing, volcanic activity and so forth. LTP can be also conceptualized as a tendency of clustering in time of similar events (droughts, floods, etc).

In statistical terms, the presence in a time series of long term fluctuations, which are difficult to describe and quantify in a purely deterministic manner, implies dramatically increased uncertainty, especially on long timescales, in comparison to classical statistics. This is easy to understand, as the observed record could be a small portion of a longer cycle whose characteristics might be difficult to infer on the basis of the available observations. In this respect, in processes characterized by LTP, the results of the statistical analysis may be difficult to decipher. As a consequence, the application of statistical tools to climatic time series should be carefully considered and classical statistics should be carefully revisited to locate points that may produce misleading or incorrect results.

In stochastic terms, LTP is contrasted to independence of events in time or even to short-term persistence (STP, also known as short-term dependence). The most representative of the latter is the Markovian dependence (also known as autoregressive of order 1 – AR(1)), according to which the future is not influenced by the past when the present is known. On the contrary, in LTP the influence of the past (possibly through deterministic dynamics that in stochastic terms are reflected in the dependence structure) never ceases. Both Markovian dependence and LTP can result from physical principles. For example, the maximum entropy principle results in Markovian dependence if the maximization of entropy is done on a particular timescale and in LTP if the maximization is done on a range of timescales (Koutsoyiannis, 2005). Although many have considered the Markovian behavior physically more plausible for the climate system (e.g. Mann and Lees, 1996), its two aforementioned features (non influence of the past, exclusiveness of a single scale of fluctuation) and other features discussed below might make it implausible, in our opinion. Moreover, climatic records do not verify a hypothesis of Markovian behavior. Thus, its adoption has been usually combined with a decomposition of a climatic series into components, one of which is Markovian (equation (6) in Mann and Lees, 1996); this decomposition is made on stochastic grounds (by spectral methods) and its physical fundament may be disputable, in our opinion.

In contrast, the LTP behavior agrees with empirical evidence in a variety of geophysical records. In fact, the history of LTP started more than half a century ago, after its discovery in geophysics by Hurst (1951), although, in a physical context (turbulence) the concept may have been pioneered a decade earlier by A. Kolmogorov. Throughout these decades numerous studies provided indications that LTP may be omnipresent in several natural (geophysical, biological) and human-associated (social, economical and technological) processes (see Kantelhardt et al., 2002; Koutsoyiannis, 2003; Montanari, 2003). Most recently, the presence of LTP in temperature data has been considered by Cohn and Lins (2005) and Rybski et al. (2006). Both have found that instrumental records and reconstructed time series of temperature are consistent with the hypothesis of LTP and therefore suggested that this property should be taken into account in statistical tests. Earlier, Koutsoyiannis (2003) arrived at similar conclusions, arguing that there is the need in hydroclimatic research to adapt classical statistics, which is based on the Independent-Identically-Distributed (IID) paradigm, so as to become consistent with the observed LTP behavior.

In this respect, Cohn and Lins (2005) and Rybski et al. (2006) have suggested a necessary rectification of the prevailing incorrect practices. Both have proposed adapted statistical tests, which they have illustrated essentially on the same climatic record, the instrumental temperature record of the Northern Hemisphere between 1856 and 2004 (due to Climatic Research Unit – CRU). Interestingly, however, their conclusions on the detection and attribution problems are opposite. Rybski et al. (2006) conclude that the hypothesis that at least part of the recent warming cannot be solely related to natural factors, can be accepted with a very low risk. Cohn and Lins (2005) state that, given what we know about the complexity, long-term persistence, and non-linearity of the climate system, this warming can be due to natural dynamics. This disagreement may indicate, in our opinion, that our understanding of the behavior of LTP and its consequences in climatic analyses and statistical testing is not complete yet and that additional insights are needed.

Such insights are sought in this study using simple analytical methods, rather than complicated numerical methods. The justification for this choice is that analytical methods are more insightful (albeit less accurate for reasons that we will discuss) than numerical ones. As an empirical basis we use the same basic data set as in the two aforementioned recent studies, the CRU record (now extended up to 2005) and, as auxiliary information, the six recently reconstructed temperature records of the northern hemisphere analyzed in Rybski et al. (2006) (here abbreviated as J98, M99, B00, E02, M03, M05 that stand respectively for Jones et al., 1998; Mann et al., 1999; Briffa, 2000; Esper et al., 2002; McIntyre and McKitrick, 2003; and Moberg et al., 2005; note that M03 was not proposed as a reconstruction but only as a modification of M99 to illustrate lack of robustness of methodology). These series are annual and therefore are not affected by seasonality. The LTP properties of some of these and some other proxy series have been also studied in other works recently (D. Stockwell, Scale invariance for dummies, http://landshape.org/enm/?p=13) and earlier (Koutsoyiannis, 2003 for J98).

Our focus is on providing insight on uncertainty rather than on proposing accurate statistical tests. In this respect, our study of the detection/attribution problem is carried out on a conceptual basis and therefore we avoid proposing categorical results. In addition, we try to locate potential pitfalls, which may appear if this uncertainty is not explicitly considered and may have also influenced previous studies.

Detecting the presence and intensity of long-term persistence

Since Hurst (1951) discovered LTP, several formalisms and conceptualizations have been used to study it, on which the algorithms to detect this behavior are based (Taqqu et al., 1995; Montanari et al. 1997; Kantelhardt et al., 2003). Among these, the most common are the original formalism by Hurst, based on the so-called rescaled range statistic (R/S) and the detrended fluctuation analysis (DFA). However, we choose to use in our analysis of climatic series the formalism based on the aggregated standard deviation (ASD). The latter has several advantages such as (a) easy understandability and transparency that enables better perception of the behavior and does not hide its implications; (b) simplicity and minimal parameterization (it does not involve any other concept than standard deviation), which enables a probabilistic description of the concepts it uses and hence a statistical framework of estimation and testing; and (c) consistency, in terms of the estimators it produces. The method is based on the analysis of the variability of the data aggregated at different time scales. Specifically, let X_i be the process of interest on discrete time i (referring to years in our case) with (true – or population) standard deviation σ and let

$$X_i^{(k)} := (X_i + \ldots + X_{i-k+1})/k \tag{1}$$

denote the aggregate (average) process at time scale k, with (true) standard deviation $\sigma^{(k)}$ (the notation implies that $X_i^{(1)} \equiv X_i$). For sufficiently large k, $X_i^{(k)}$ represents the climatic process; typically, the convention $k = 30$ is used to standardize the climatic time scale (number of years). Now, LTP is expressed by the elementary scaling property

$$\sigma^{(k)} = \frac{\sigma}{k^{1-H}} \tag{2}$$

where H is the Hurst exponent, which for stationary and positively correlated processes varies in the range (0.5, 1). $H = 0.5$ means independence and increasing values represent increasing LTP intensities.

The simple equation (2) can support: (a) a definition of LTP; (b) a definition of a stochastic process having this property, that is the simple scaling stochastic (SSS) process (also known as stationary intervals of a self similar process); and (c) the estimation of H using sample estimates of $\sigma^{(k)}$ at several scales k. (2) implies that the autocorrelation $\rho_j^{(k)}$ for scale k and lag j (defined as $\rho_j^{(k)} := \mathrm{Cov}[X_i^{(k)}, X_{i+kj}^{(k)}] / \mathrm{Var}[X_i^{(k)}]$) is independent of scale (e.g. Koutsoyiannis, 2002):

$$\rho_j^{(k)} = \rho_j = (1/2)\,[(|j+1|)^{2H} + (|j-1|)^{2H}\,] - |j|^{2H} \tag{3}$$

LTP is more precisely defined as an asymptotic property for large scales, in which case (2) should be replaced by $\sigma^{(k)} = \sigma^{(l)} / (k/l)^{1-H}$ for $k/l > 1$ and $l \to \infty$; also SSS is more precisely defined in terms of scaling properties of the distribution function. It is important to note that, even though the same equation (2) can serve as a basis for the definition of the LTP as well as the SSS process, these two are totally different notions:

LTP is a behavior that can be verified in any type of time series, such as series of observations of a natural process, output of a deterministic model, or synthetic series generated by a stochastic process. In contrast, SSS is a stochastic process.

For comparison, in the case of the simplest STP model, which is the AR(1), (2) and (3) become respectively (Koutsoyiannis, 2002):

$$\sigma^{(k)} = \frac{\sigma}{\sqrt{k}} \sqrt{\frac{(1 - \rho^2) - 2\rho\,(1 - \rho^k) / k}{(1 - \rho)^2}} \tag{4}$$

$$\rho_1^{(k)} = \frac{\rho\,(1 - \rho^k)^2}{k\,(1 - \rho^2) - 2\rho\,(1 - \rho^k)}, \quad \rho_j^{(k)} = \rho_1^{(k)}\,\rho^{k\,(j - 1)}, \quad j \geq 1 \tag{5}$$

where $\rho \equiv \rho_1^{(1)}$. These indicate that (a) for large k, $\sigma^{(k)} \sim \sigma/\sqrt{k}$; (b) $\rho_j^{(k)}$ is a decreasing function of k; and (c) only at scale $k = 1$ (annual) is the process Markovian (i.e., $\rho_j = \rho^j$); at all other scales the autocorrelation structure in (5) (i.e. $\rho_j^{(k)} = \rho_1^{(k)}\,(\rho^k)^{j-1}$) is identical to that of an autoregressive moving average (ARMA) process of order (1, 1), another classical example of STP. Note that both AR(1) and SSS involve a single parameter each and that the equations (2) and (3) of SSS are simpler than (4) and (5) of AR(1), even though the former has been regarded by many as very complicated.

Obviously, the different formalisms to LTP imply different estimates of H. This is demonstrated in Table 1 for the seven time series and for three formalisms: the DFA as derived by Rybski et al. (2006), the R/S and the ASD. In the latter we used an algorithm by Koutsoyiannis (2003), which by construction ensures consistency ($H < 1$); it can be observed that the other methods resulted in some inconsistent (> 1) values. Generally, all methods result in very high but different H values.

Statistical uncertainty

Given a sample $X_1, \ldots, X_n$ of size n and observations $x_1, \ldots, x_n$, clearly $X_1^{(n)}$ is the standard estimator of the mean μ of the process (most typically denoted as $\bar{X}$) and $x_1^{(n)}$ is the estimate of μ. The standard deviation $\mathrm{StD}[\bar{X}] \equiv \mathrm{StD}[X_1^{(n)}]$ is a convenient indicator of uncertainty, and according to the scaling property (2), $\mathrm{StD}[\bar{X}] = \sigma^{(n)} = \sigma/n^{1-H}$. (Here $\mathrm{StD}[.] := \sqrt{\mathrm{Var}[.]}$ denotes the standard deviation of a random variable). If we compare it to the classical statistical law StD[$\overline{X}$] = $\sigma/\sqrt{k}$ (also valid asymptotically for STP processes as shown in (4)), the differences are dramatic as H grows away from 0.5. To demonstrate it, for a series of length n with LTP we can calculate the "equivalent" sample size n' in the classical statistics (IID) sense, so that $\sigma/n^{1-H} = \sigma/n'^{0.5}$. Clearly,

$$n' = n^{2(1-H)} \tag{6}$$

As shown in Table 1, the equivalent sample sizes resulting by this equation for the seven time series are as low as 2-5. For instance in the SSS sense, the longest sample size (1979), is equivalent to a classical statistical sample size of about 3! Thus, a record with length of 1979 years, which certainly would be called a long record having in mind classical statistics, is a very short record in the SSS framework. Only this example suffices to demonstrate that the Hurst behavior has astonishing effects in the foundation of climatology and hydrologic statistics, provided that the LTP hypothesis is true.

Even the AR(1) model implies reduction of sample size; in this case using (4) and a similar logic, we obtain that

$$n' = n\frac{(1-\rho)^2}{(1-\rho^2) - 2\rho\,(1-\rho^n)\,/\,n} \tag{7}$$

Values estimated from (7) are also given in Table 1 and show that the reduction is not as dramatic as in the SSS case.

However, the implications are perhaps even worse than described above, because the analysis was based on the assumption that H is known a priori. In reality, H is estimated from the data, so there is additional sampling uncertainty (statistical estimation error). The sampling uncertainty applies also to all other statistics and we can anticipate that all confidence bands are wider than in classical statistics, as will be discussed below. In addition, because LTP is eventually an asymptotical property of the process (which should be detected on the tail, i.e. on the largest scales), even the detection of LTP is highly uncertain when dealing with time series with short length (Taqqu et al., 1995).

This point has already been made in some studies. For example Koutsoyiannis (2002) showed that the sum of three Markovian processes (whose behavior, rigorously speaking is STP) is virtually indistinguishable from a process with LTP for lags as high as of the order of 1000. To demonstrate this point further, we fitted to the E02 series an ARMA(1, 1) process. Testing the autocorrelation function of the residuals of this, we concluded that they are indistinguishable from white noise; this means that the series is consistent with the ARMA(1, 1) process, i.e. it exhibits STP with Hurst coefficient 0.50. Furthermore, we generated with the fitted ARMA(1, 1) a synthetic series with sample size 2000, and all estimation methods we tried gave incorrect values of H in the order 0.79-0.93. Continuing this experiment, we also found that we need a series with length of about 20 000 to correctly estimate H, viz to find a value around 0.50. These examples clearly point out that even the distinction between the extreme cases $H = 0.5$ and $H \to 1$ is not statistically decidable with typical sample sizes.

Observation uncertainty

It is well known that observations of hydrometeorological processes involve several inaccuracies; even in the instrumental CRU series, some observation uncertainty exists, mainly because of spatial integration of point measurements whose number and locations differ through history. But in the case of proxy data, there is an extra source of high uncertainty because the data are not instrumental. In fact, all six proxy series considered here are supposed to represent exactly the same process, that is, the evolution of the northern hemisphere temperature. The different values assigned for the same year in the different series manifest none other than the uncertainty in reconstruction of the past climate. This is well known and is related to the subjectivity of dendroclimatology on which the given proxy series are primarily based. The subjectivity originates from sampling procedures (e.g. in picking and choosing which samples to use) and from the differing statistical calibration approaches (recall, for instance, that M03 and M99 are based on the same original data; for additional discussions see McIntyre and McKitrick, 2003, and Jones and Mann, 2004). The differences in seasonal and spatial representativeness of the various reconstructions is an additional source of uncertainty.

An interesting piece of information conveyed by all proxies is the consistency of all of them with the LTP hypothesis, even if we do not include in the analyses the years of instrumental observations (which one may argue that are affected by global warming). To make this clearer, we redid all analyses for the period 1400-1855, which is the common period of all proxy series prior to the period of instrumental records. The results, shown in Table 1, indicate that the H values obtained for this period are virtually identical to those for the complete data set and close to each other, averaging to 0.91, a value close to that of CRU (0.93). On the other hand, the standard deviations, even though they do not depart significantly from the values of the whole period of each sample, are very different to each other (ranging in 0.09-0.21°C vs. 0.27°C of the CRU series).

It is interesting to compare the above range of values with the sampling uncertainty of the standard deviation of the CRU series. Combining known results (Matalas, 1967; Koutsoyiannis, 2003), it is observed that, when there is temporal dependence in the process of interest, the standard estimator S of the standard deviation σ is not unbiased and that an approximately unbiased estimator for both the LTP and STP cases is

$$\overset{\approx}{S} := \sqrt{\frac{n'}{n'-1}}\, S \tag{8}$$

This assumes that n (the actual sample size) is large enough (for a more accurate expression for small n see Koutsoyiannis, 2003). Obviously, in the SSS case the estimate $\overset{\approx}{s}$ may differ dramatically from the standard estimate s (notice the notational convenience of lower case letters for estimates, i.e. numerical values, and upper case ones for estimators, i.e. random variables). Also, combining results from Koutsoyiannis (2003) (based on systematic Monte Carlo simulations) and using $\overset{\approx}{s}$ as an estimate of the true standard deviation σ, it can be obtained that in the SSS case,

$$\frac{\mathrm{StD}[\overset{\approx}{S}]}{\overset{\approx}{s}} = \frac{\mathrm{StD}[S]}{s} \approx \sqrt{\frac{(0.1\ n + 0.8)^{\lambda(H)}}{2\ (n-1)}}, \text{ with } \lambda(H) := 0.088\ (4\ H^2 - 1)^2 \tag{9}$$

Now using the statistics of the CRU series, it is computed that the estimate of StD[S]

is 0.033°C (vs. 0.015°C in classical statistics). Roughly speaking, this justifies a difference in standard deviation between the different series of about 0.08°C (at significance 1%; even though the distribution of StD[S] is not normal). Consequently, from the values in Table 1, we can conclude that the variability of the J98 and M05 series is compatible with the variability of the CRU record. The same result does not apply to all other series. Thus, if one accepts one of the other four series as representative of the past climate, one can readily conclude that the observed temperature variation in the last years is not a result of natural dynamics. In other words, there is a statistical significance in the change of standard deviation, so no additional statistical test is needed. This also explains why in Rybski et al. (2006) these four series resulted in "earlier detection" (to use their terminology). Furthermore, with simple statistical calculations with the standard deviation estimates shown in Table 1, we can easily classify the proxies in two groups (one is J98, M03, M05 and the other one M99, B00, E02), each of which contains series compatible to each other but the two groups are incompatible to each other. This makes unrealistic the possibility to use all series simultaneously in a global statistical approach and highlights once again the uncertainty involved in the use of proxy series.

Statistical testing for climatic change

The above findings highlight the potential lack of reliability of statistical tests performed on climatic records, especially proxy ones. Some of these critical behaviors are not known and not immediately evident. It is interesting to inspect with deeper detail the potential effects on the statistical detection of climatic change.

Cohn and Lins (2005) used as a test statistic the slope of a linear fit to the time series to test whether or not a climate variable has changed in a statistically significant sense, over the available observation period. Rybski et al. (2006) proposed essentially the statistic $D_{i,\,l}^{(k)} := X_i^{(k)} - X_{i-l}^{(k)}$ to test whether a or not a climate variable, defined on a time scale k, has changed in a statistically significant sense, over a period of l years (starting from year i). This is indeed an interesting statistic and we wish to discuss it further (noting that similar analyses apply to any type of statistical test). First, $D^{(k)}_{i,\,l}$ does not depend on a fitted model (as e.g. a linear fitting to the data). Second, it is flexible and convenient as it allows choosing the climatic time scale k and the lag l/k (defined on scale k). Third, and more important, it yields a simple, general (not dependent on the process), convenient and exact expression for the standard deviation of the test statistic, which we have obtained from (1):

$$\mathrm{StD}[D^{(k)}_{i,\,l}] = \sqrt{2}\,\sigma^{(k)}\sqrt{1-\rho^{(k)}_{l/k}} \tag{10}$$

This does not depend on the mean of the process and includes two multiplicative terms, the first ($\sigma^{(k)}$, computed by (2) or (4)) depending on the standard deviation and the autocorrelation structure of the process, and the second (computed by (3) or (5)) depending merely on the autocorrelation structure.

The variation of the two terms with ρ for both the SSS and AR(1) processes is depicted in Figure 1(a) for the assumptions indicated in the caption. The two terms have opposing effects. The first term increases with ρ, faster in the SSS than in the AR(1) case. The second term is a decreasing function of ρ but in AR(1) it practically equals 1 unless ρ takes very high values (> 0.95). The combined effect of the two terms is demonstrated in Figure 1(b) for $\sigma = 1$. In the SSS case, for relatively low ρ (or H), $\mathrm{StD}[D^{(k)}_{i,\,l}]$ is an increasing function of H but for $\rho > {\sim}0.70$ it becomes a decreasing function tending to zero as $\rho \to 1$ (because the second term dominates). The situation is similar in the AR(1) case but $\mathrm{StD}[D^{(k)}_{i,\,l}]$ becomes decreasing function of ρ only for $\rho > 0.95$.

In all this demonstration it was assumed that both σ and ρ are known. In practice, however both are unknown and estimated from the sample. The picture changes in this case. To estimate $\mathrm{StD}[D^{(k)}_{i,\,l}]$, one may be tempted to use the standard estimate s of σ that is used in classical statistics (for example, Rybski et al. do not mention this problem at all). However, as explained above (eqn. (8)), in SSS statistics, s is strongly biased and $\overset{\approx}{s}$ should be used instead; thus, if $s = 1$ then, according to (8) and (6), an approximately unbiased estimate of σ is $[n^{2(1-H)} / (n^{2(1-H)} - 1)]^{1/2}$. It can be seen in Figure 1(b) that in this case $\mathrm{StD}[D_{i,\,l}^{(k)}]$ is an increasing function for virtually the entire domain of ρ.

The effects of autocorrelation to the significance of rejecting the null hypothesis of no change in climate is demonstrated in Figure 1(c), assuming that a classical statistical test has already resulted in rejection of the null hypothesis with extremely low risk (i.e. significance level) 10^{-15}. It is observed that the significance level increases dramatically with ρ. For $\rho = 0.7$ the significance level becomes 10^{-2} in the SSS case and 10^{-3} in the AR(1) case. For higher values of ρ both the SSS and the AR(1) processes yield significance levels that are very close to each other; this may be interesting to those who do not trust the LTP hypothesis and prefer to assume an STP behavior.

Yet this modified analysis was based on the tacit assumption that the true value of H or ρ is known. But this assumption is not true and thus the above methodology does not consist a formal test, so we call it a "pseudo-test" and anticipate that it only yields a lower bound of the significance level. For unknown H, the estimate of $\mathrm{StD}[D_{i,\,l}^{(k)}]$ is anticipated to be greater but its calculation may be intractable by analytical means (given that the estimators of H and σ are dependent; Koutsoyiannis, 2003). A Monte Carlo testing framework becomes then the method of choice (such a framework was proposed in a different context by Cohn and Lins, 2005, which results in even greater escalation of orders of magnitude of significance level). However, as explained above, the focus of this Letter is on understanding so we preferred the analytical discussion, even though it yields a pseudo-test rather than a formal one.

It may be of some interest to apply this pseudo-test to the CRU data series. The application is shown graphically in Figure 2, for a double-sided test for significance level 10^{-2} and for the SSS case, using all possible integer lags l/k from 1 ($l = 30$) to 4 ($l = 120$). In neither case the pseudo-test resulted in rejection of the null hypothesis (no change), although it comes close to rejection for 2005 for $l/k = 3$. As noted above, a real test would be even more tolerant in rejecting the null hypothesis. This result agrees with Cohn and Lins (2005) rather than with Rybski et al. (2006) who perhaps underestimated some uncertainty factors, as discussed above.

Conclusion and discussion

The above analysis shows that the detection and attribution problem should be studied in a framework properly recognizing and characterizing the dependence structure of the climatic records, and that the classical IID framework should be abandoned. It also shows that the statistical uncertainty is dramatically increased in the presence of dependence, especially if this dependence is LTP.

Certainly, the detection and attribution problem will continue attracting attention in the years to come, as newer data accumulate. Before concrete conclusions can be drawn, a rigorous methodological framework, based on both physical and statistical arguments, should be built. Obviously, the aim of this Letter was neither to provide such a framework nor to give an answer to the detection/attribution problem. We hope, however, that our remarks may be useful in building this framework.

The answer to the very important question whether the dependence structure of climatic processes is LTP or STP is very relevant to the detection and attribution problems. However, a categorical answer to this question cannot be based on merely statistical arguments, because, as we demonstrated above, even the presence of LTP can be disputable on purely statistical grounds. Certainly, better physical understanding and theoretical analyses are strongly needed to illustrate and verify or falsify the LTP hypothesis or other climatic hypotheses.

This emphasizes the need of a theory, in addition to statistical tools, to assess the natural behaviors. Without a concrete theoretical framework the situation can be summarized by the following quotation from Cohn and Lins (2005): “From a practical standpoint … it may be preferable to acknowledge that the concept of statistical significance is meaningless when discussing poorly understood systems.”

References

Briffa, K. R. (2000), Annual climate variability in the Holocene: Interpreting the message of ancient trees, *Quat. Sci. Rev.*, 19(1– 5), 87–105.

Bloomfield, P. (1992), Trends in global temperature, *Climatic Change*, 21, 1-16, 1992.

Cohn, T. A., and H. F. Lins (2005), Nature's style: Naturally trendy, *Geophys. Res. Lett.*, 32(23), L23402, doi:10.1029/2005GL024476.

Esper, J., E. R. Cook, and F. H. Schweingruber (2002), Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability, *Science*, 295(5563), 2250–2253.

Hurst, H. E. (1951) Long term storage capacities of reservoirs. *Trans. Am. Soc. Civil Engrs*, 116, 776–808.

Jones, P. D., K. R. Briffa, T. P. Barnett, and S. F. B. Tett (1998), High resolution palaeoclimatic records for the last millennium: Interpretation, integration and comparison with GCM control-run temperatures, *Holocene*, 8(4), 455– 471.

Jones, P.D. and Mann, M.E., Climate Over Past Millennia, *Reviews of Geophysics*, 42, RG2002, doi: 10.1029/2003RG000143, 2004.

Kantelhardt, J. W, S. A. Zschiegnera, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley (2002), Multifractal detrended fluctuation analysis of nonstationary time series, *Physica A*, 316, 87 – 114.

Koutsoyiannis, D. (2002), The Hurst phenomenon and fractional Gaussian noise made easy, *Hydrol. Sci. J.*, 47(4), 573–595.

Koutsoyiannis, D. (2003), Climate change, the Hurst phenomenon, and hydrological statistics,. *Hydrol. Sci. J.* 48(1), 3-24.

Koutsoyiannis, D. (2005), Uncertainty, entropy, scaling and hydrological stochastics, 2, Time dependence of hydrological processes, *Hydrol. Sci. J.*, 50(3), 405-426.

Koutsoyiannis, D. (2006), A toy model of climatic variability with scaling behaviour, *J. Hydrol.*, 322, 25-48.

Mann, M. E., R. S. Bradley, and M. K. Hughes (1999), Northern Hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations, *Geophys. Res. Lett.*, 26(6), 759– 762.

Mann, M.E., and J.M. Lees (1996), Robust estimation of background noise and signal detection in climatic time series, *Climatic Change*, 33, 409-445.

Matalas, N. C. (1967), Mathematical assessment of synthetic hydrology, *Water Resour. Res.*, 3(4), 937-945.

McIntyre, S., and R. McKitrick (2003), Corrections to the Mann et al. (1998) proxy data base and northern hemispheric average temperature series, *Energy Environ.*, 14(6), 751– 771.

Moberg, A., D. M. Sonechkin, K. Holmgren, N. M. Datsenko, and W. Karlen (2005), Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data, *Nature*, 433(7026), 613– 617.

Montanari, A. (2003), Long-range dependence in hydrology, in *Theory and Applications of Long-Range Dependence*, ed. P. Doukhan et al., 461– 472, Springer, New York.

Montanari, A., Rosso, R. & Taqqu, M. S. (1997) Fractionally differenced ARIMA models applied to hydrologic time series. *Water Resour. Res.* 33(5), 1035–1044.

Rybski, D., A. Bunde, S. Havlin, and H. von Storch (2006), Long-term persistence in climate and the detection problem, *Geophys. Res. Lett.*, 33, L06718, doi:10.1029/2005GL025591.

Taqqu, M.S., V. Teverovsky, and W. Willinger (1995), Estimators for long-range dependence: an empirical study, *Fractals*, 3, 785-798.

von Storch, H., and F. Zwiers (1999), *Statistical analysis in climate research*, Cambridge University Press, Cambridge.

von Storch, H. (1995) Misuses of statistical analysis in climate research. In: *Analysis of Climate Variability: Applications of Statistical Techniques*, edited by von Storch H. and Navarra, A., pp. 11–26. Springer-Verlag, Berlin.

Table 1 Comparisons of estimates of statistics for different methods and data sets.

Data series		CRU	J98	M99	B00	E02	M03	M05
				All data				
Sample size		150	992	981	994	1162	581	1979
s, standard estimate		0.27	0.23	0.13	0.14	0.14	0.17	0.22
H by DFA*		1.09	0.82	0.97	0.93	1.04	0.83	0.86
H by R/S		1.07	0.90	0.89	0.89	0.93	0.97	0.92
H by ASD		0.93	0.88	0.91	0.91	0.94	0.92	0.94
ρ		0.84	0.53	0.65	0.64	0.81	0.66	0.91
Equivalent	SSS	1.9	5.0	3.4	3.3	2.5	2.8	2.7
sample size	AR(1)	13.8	307.5	205.0	221.1	120.8	119.3	95.3
				Period 1400-1855				
Sample size			456	456	456	456	456	456
s, standard estimate			0.20	0.10	0.13	0.09	0.16	0.21
H by ASD			0.86	0.88	0.91	0.93	0.92	0.93
ρ			0.54	0.62	0.59	0.77	0.65	0.88

* Values from Rybski et al. (2006), except in the CRU series, which was estimated in this study.

List of Figure captions

Figure 1 Variation with ρ of **(a)** the two multiplicative terms comprising StD[D] assuming σ = 1, **(b)** StD[D] assuming σ = 1 or s = 1 as indicated, and **(c)** the implied significance in rejecting the null hypothesis assuming that s = 1 and that in classical IID statistics this significance level is 10^{-15}; assumptions: k = 30, l/k = 3, n = 150.

Figure 2 Graphical depiction of the pseudo-test based on StD[D] with known H. The continuous solid curve represents the CRU time series averaged over climatic scale k = 30. The series of points represent values of D for the indicated lags l/k. Horizontal lines represent the critical values of the pseudo-test, which are the estimates of StD[D] times a factor 2.58 corresponding to a double-sided test with significance level 1% and assuming normality (only the positive critical values are plotted).

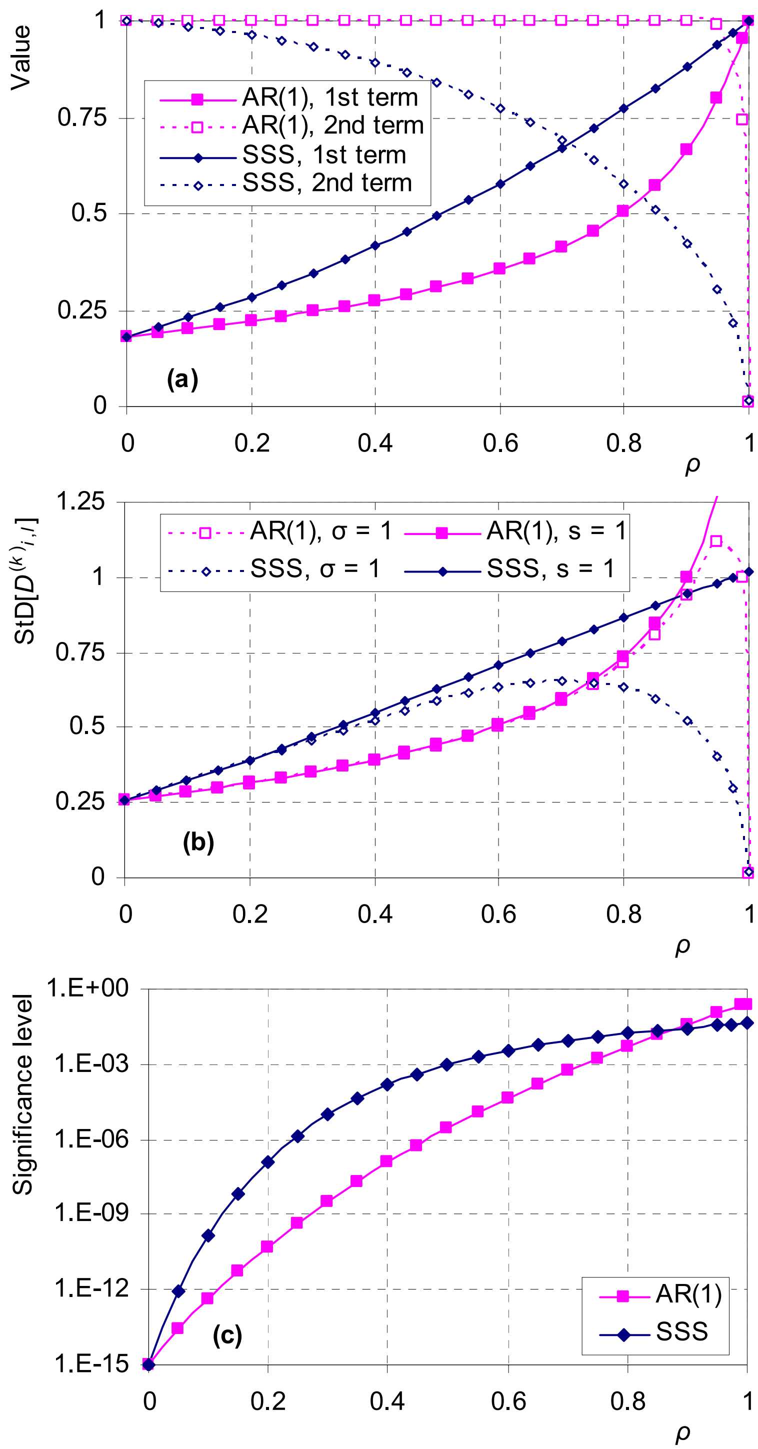

Figure 1 Variation with ρ of **(a)** the two multiplicative terms comprising StD[$D_{i,\,l}^{(k)}$] assuming $\sigma = 1$, **(b)** StD[$D_{i,\,l}^{(k)}$] assuming $\sigma = 1$ or $s = 1$ as indicated, and **(c)** the implied significance in rejecting the null hypothesis assuming that $s = 1$ and that in classical IID statistics this significance level is 10^{-15}; assumptions: $k = 30$, $l/k = 3$, $n = 150$.

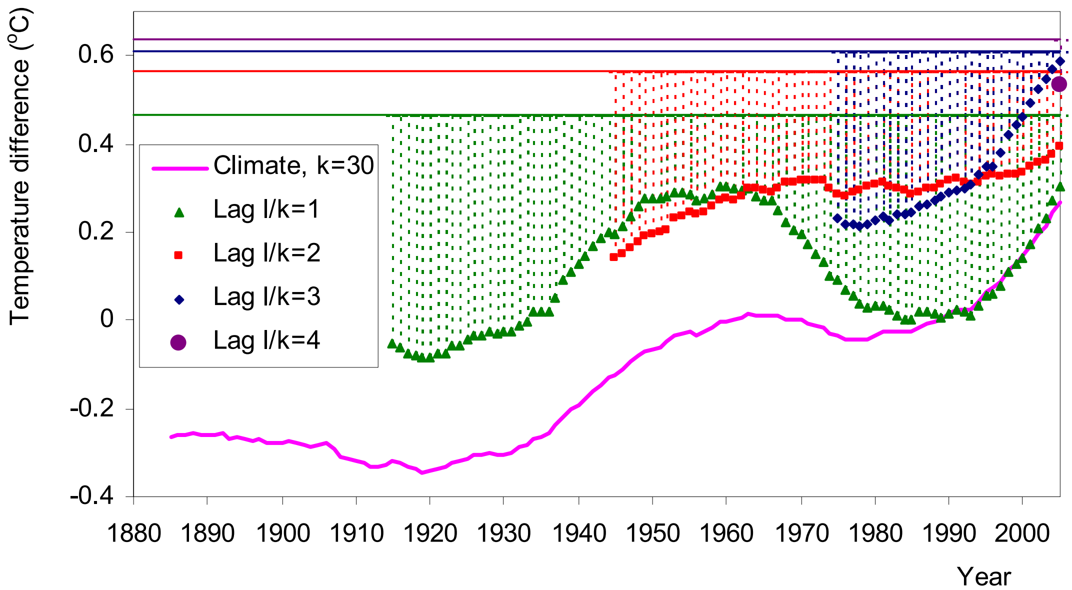

Figure 2 Graphical depiction of the pseudo-test based on StD[$D_{i,\,l}^{(k)}$] with known H. The continuous solid curve represents the CRU time series averaged over climatic scale k = 30. The series of points represent values of $D_{i,\,l}^{(k)}$ for the indicated lags l/k. Horizontal lines represent the critical values of the pseudo-test, which are the estimates of StD[$D_{i,\,l}^{(k)}$] times a factor 2.58 corresponding to a double-sided test with significance level 1% and assuming normality (only the positive critical values are plotted).