Long-term persistence and uncertainty on the long term

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

After a significant delay of half a century since the discovery of the Hurst phenomenon in geophysics and after accumulation of evidence that climate may be consistent with the hypothesis of long-term persistence (LTP), discussion is currently ongoing about the implications of this hypothesis in climate research. However, recent publications admitting LTP arrive to disagreeing conclusions. This may manifest incomplete understanding of this behavior and its consequences in statistical tasks. To offer some insight on this we demonstrate that LTP implies increased uncertainty on the long term and thus entails dramatic differences in estimations and tests, in comparison to classical statistics. On these grounds, we also discuss the problem of detection and attribution of climatic change.

Introduction

In two recent Letters, Cohn and Lins (2005) and Rybski et al. (2006) raise two important issues: (a) they find that both instrumental climatic records and reconstructed time series of climate are consistent with the hypothesis of long-term persistence, and (b) they suggest that this property should be taken into account in statistical tests and they offer methods to cope with this. Earlier, Koutsoyiannis (2003) arrived to similar conclusions. Cohn and Lins (2005) state: "It is therefore surprising that nearly every assessment of trend significance in geophysical variables published during the past few decades has failed to account properly for long-term persistence." We share this opinion, as the LTP hypothesis (also known as Hurst phenomenon, Joseph effect, long memory, long-range dependence, scaling behaviour, and multi-scale fluctuation; Koutsoyiannis, 2006) has a history of more than half a century since the discovery of LTP in geophysics by Hurst (1951), and even longer in mathematics and physics since the pioneering work by Kolmogorov (1940). Throughout these decades numerous studies have accumulated indications that LTP may be omnipresent in several natural (geophysical, biological) and human-associated (social, economical and technological) processes (for references see Kantelhardt et al., 2003; Koutsoyiannis, 2003; Montanari, 2003). This behavior seems to be particularly the case in climatic processes and harmonizes with the general recognition of a perpetually changing climate.

Cohn and Lins (2005) and Rybski et al. (2006) agree on the presence of LTP and the importance of taking it into account in statistical tests, and both base their tests essentially on the same climatic record, the instrumental temperature record of the Northern Hemisphere between 1856 and 2004 (due to Climatic Research Unit – CRU – available from http://www.cru.uea.ac.uk/ftpdata/tavenh2v.dat). Interestingly, however, their conclusions on the so called detection or attribution problem 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, whereas 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, in addition to the delay in incorporating the LTP behavior in the climatic research, may indicate, in our opinion, that our understanding of this behavior and its consequences is not complete yet and that additional insights are needed.

In this respect, with this Letter we wish contribute our thoughts, views and analyses on these issues, and also put emphasis to another closely related issue: the high uncertainty involved in such exercises. Specifically, LTP entails high uncertainty on the long term, much higher than in classical statistics that are based on the hypothesis of independently identically distributed (IID) variables. The familiarity with the classical (IID) statistical paradigm or even with typical short-term persistence (STP) stochastic processes may mislead us so as fail to consider the higher uncertainty associated with LTP.

As the focus of this study is on understanding rather than on providing accurate statistical tests, we will avoid a categorical answer to the question raised in the previous two Letters regarding the detection/attribution problem. But we will study it and expose our own views. We study the same basic data set as in both earlier studies (the CRU record, now extended up to 2005) as well as all six recently reconstructed temperature records of the northern hemisphere analyzed in Rybski et al. (2006) as auxiliary information (here abbreviated as JBBT98, MBH99, B00, ECS02, MM03, MSHDK05 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). The LTP properties of some of these and some other proxy series have been also studied in other works recently (Stockwell, 2006a) and earlier (Koutsoyiannis, 2003 for JBBT98).

Long-term persistence and its formalisms

Since Hurst's (1951) discovery of LTP, several formalisms and conceptualizations have been used to study it, on which the algorithms to detect this behaviour are based (Taqqu et al., 1995; Montanari et al. 1997). All formalisms include the so called Hurst coefficient (or exponent) denoted after Hurst as H (or as α in Rybski et al., 2006). In general, the most common is the original formalism by Hurst, based on the so-called rescaled range statistic (R/S). In contrast, in climatological studies (e.g Vjushin et al., 2001; Kantelhardt et al., 2002; including Rybski et al., 2006) the most common formalism has been the so-called detrended fluctuation analysis (DFA), introduced by Peng et al. (1994) to study long-range correlations in DNA sequences.

A particular formalism should not be viewed just as an algorithm or a recipe to determine H. In contrast, it is important in understanding, conceptualizing and modeling of the underlying behavior. In this respect, an ideal formalism should have certain characteristics such as (a) easy understandability and transparency in order to enable perception of the behavior and not to hide its implications; (b) simplicity, in order to enable a probabilistic description of the concepts it uses and hence a statistical framework of estimation and testing; (c) objectivity, in order for its application to be as free as possible from arbitrary choices; and (d) consistency, in terms of the estimators it produces. Perhaps none of the available formalisms have all these properties (which emphasizes the need for development of better ones) but the properties can serve as reference when discussing different existing formalisms.

With respect to point (d), it should be reminded that in a stationarity framework (in fact always assumed in such analyses – at least in the null hypothesis) an estimate of H should be in the interval (0, 1) or, for processes with positive temporal dependence, in the interval (0.5, 1), as also pointed out in Cohn and Lins (2005) and Rybski et al. (2006). Values $H > 1$ simply indicate inconsistency of the algorithm. This is not difficult to verify: benchmark synthetic time series can be produced by a stationary model for which many algorithms for the estimation of LTP produce $H > 1$.

With reference to the points (a)-(d) listed above, DFA (used by Rybski et al.), may be far from an ideal method. This method is based on best fit polynomials that describe local "trends" on separate intervals of the data sequence. By construction, it may hide the important uncertainty characteristics, as will be discussed later, being a rather deterministic conception (fitting of polynomials). It may be not simple enough to enable description of the probability distribution of the variables it uses (parameters of polynomials). It may be not objective as the separation into intervals and the order of the polynomial used are subjectively chosen by the user (e.g. Rybski et al. used a quadratic fitting). And it may be not consistent as it can result in H greater than 1 (this is the case for ECS02, for which Rybski et al. gave $H = 1.04$; see Table 1).

This inconsistency is common to other methods: for example the original (Hurst's) R/S algorithm often produces $H > 1$ and has several other deficiencies: it does not enable an analytical computation of statistics and the method is associated with bias and low efficiency in the estimation of H (Koutsoyiannis, 2002, 2003). Even the maximum likelihood estimation method (which still depends on an assumed model, such as the fractionally differenced autoregressive moving average process, also called fractional ARIMA) is not free of this inconsistency if the assumptions underlying the estimation are not fully satisfied. For instance, by fitting a fractional ARIMA to the ECS02 record we obtained $H > 1$– but a check of the model residuals allowed us to detect that the model did not provide a good fit.

The aggregated standard deviation (ASD) method gives perhaps the most convenient formalism (and closest to the ideal one) because it does not involve any other concept except standard deviation and thus it enables easy understandability and convenient statistical description. 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

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

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

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

This simple equation: (a) can support a definition of LTP; (b) can support a definition of the simple scaling stochastic (SSS) process (also known as stationary intervals of a self similar process); and (c) suffices to estimate 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):

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

More precisely, LTP is 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.

For comparison, in the simplest STP model, which is the Markovian or autoregressive model of order 1 (AR(1)), (2) and (3) become respectively (e.g. Koutsoyiannis, 2002):

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

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

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 $k = 1$ is the process Markovian (i.e., $\rho_j = \rho^j$).

Obviously, the different formalisms in 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 method. 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 H but the specific values obtained by the different methods differ.

Statistical uncertainty

Some believe that the distinction between STP and LTP is a pointless disputation. However, we will show in what follows that LTP implies relevant practical upshots, related to high variability and uncertainty on the long term. We discuss here several aspects of uncertainty which perhaps the most common formalisms (R/S and DFA) obscure.

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), $\mathrm{StD}[\bar{X}] = \sigma^{(n)} = \sigma/n^{1-H}$. If we compare it to the classical statistical law $\mathrm{StD}[\overline{X}] = \sigma/\sqrt{k}$ (also valid asymptotically for STP processes as shown above), the differences are dramatic as H grows away from 0.5. To demonstrate it, for a series of length n with SSS we can calculate the "equivalent" sample size n' in the classical statistics 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 SSS, the longest data set, with size 1979, is equivalent to a classical statistical (IID) sample of size ~3! This emphasizes the fact that 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 and that the edifices have to be rebuilt on more solid grounds, provided that the LTP hypothesis is correct.

Even the STP Markovian 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 as the analysis was based on the assumption that H is known a priori. In reality, H is typically 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 zones are wider than in classical statistics, as will be discussed below.

One may argue that the concept of "equivalent sample size", as exposed above, applies only to the uncertainty of the mean and that the uncertainty for other statistical quantities, including H, might be lower. This, however, is not the case. In fact, LTP is in fact an asymptotical property of the process (which should be detected on the tail, i.e. on the largest scales) and therefore the detection of LTP is highly uncertain when dealing with time series with short length (Taqqu et al., 1995).

This point has been already done in some studies. For example Koutsoyiannis (2002) showed that the sum of three Markovian processes is virtually indistinguishable from a process with LTP for lags as high as of the order of 1000. Also, Maraun et al. (2004) argued using DFA that LTP cannot not be concluded unambiguously from typical samples. To demonstrate this point further, we fitted to the ECS02 series (for which Rybski et al., 2006, gave the highest H) an autoregressive moving average (ARMA) linear process of order (1, 1) (another classical example with STP). The resulting coefficients of the process are $\phi = 0.95$ (the autoregressive coefficient) and $\theta = 0.4$ (the moving average coefficient). Based on the series of residuals in this process, one can easily conclude that the autocorrelation function is not distinguishable from white noise and that the Hurst coefficient is around 0.50; this means that the series is consistent with the ARMA(1, 1) model, i.e. exhibits STP. Besides we generated with this process 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 even the distinction between the extreme cases $H = 0.5$ and $H \rightarrow 1$ is not statistically decidable with typical sample sizes.

Observation uncertainty

It is well known that observations of hydrometeorological processes involve several inaccuracies; this is particularly true for spatially integrated quantities, which come from point measurements whose number and locations differ through history. So, even in the "instrumental" CRU series, some observation uncertainty exists. But in the very case of proxy data, there is an extra source of high uncertainty because the data are not instrumental. In fact, all six proxy series are supposed to represent exactly the same process, 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. Interestingly, two of the series (MBH99 and MM03) are based on the same original data but they have significant differences due to different approaches of reconstruction (see discussions by McIntyre and McKitrick, 2003, 2005, and Schmidt and Amman, 2005). In addition, we must have in mind that dendroclimatology, on which the given proxy series are primarily based, cannot give accurate and objective values. For example, the subjectivity of the method (due to selection of tree samples that correlate well with temperature records and disregarding of others that do not correlate well) is expressed by Esper et al. (2003), when they say “The ability to pick and choose which samples to use is an advantage unique to dendroclimatology. That said it begs the question, how low can we go?”. (For a commentary on this see McIntyre, 2005, and for a related parody see Stockwell, 2006b).

Thus, an approach isolating the series and examining each one separately, as done in Rybski et al. (2006), may not be consistent. Perhaps an approach that would combine the different series as an ensemble in an uncertainty framework (e.g. in a manner similar as in Monte Carlo simulations), which should also combine the instrumental series and include sampling uncertainty, is worth trying. But even if we fail to assemble such an approach, certainly we should identify and exploit the useful information contained on the proxy series. The most important information is that all series are consistent with the LTP hypothesis, even if we do not include in the analyses the years of instrumental observations (which one may argue that are already affected by global warming). This allows avoiding a circular logic, in which detection of a change would be based on data which were already used to infer the model. This remedy is none other than the splitting technique described by von Storch (1995) and also applied for JBBT98 by Koutsoyiannis (2003). 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 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 of the proxy series with the sampling uncertainty of the standard deviation of the CRU series. Combining known results (Matalas, 1967; Salas, 1993, p. 19.11; Beran, 1994, p. 156; Koutsoyiannis, 2003), it is observed that 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

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

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 $\tilde{\tilde{s}}$ 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 $\tilde{\tilde{s}}$ as an estimate of the true standard deviation σ, it can be obtained that in the SSS case

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

where $\mathrm{StD}[.] := \sqrt{\mathrm{Var}[.]}$ denotes the standard deviation of a random variable.

Now using the statistics of the CRU series, it is computed that the estimate of StD[S] (which could be spelled out as "the estimate of the standard deviation of the standard estimator of standard deviation"!) 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 series JBBT98 and MSHDK05 are "compatible" with CRU, whereas all other series are not compatible even in an SSS framework. Thus, if one accepts one of the other four series as representative of the past climate, one can readily conclude that the warming in the last years is not a result of natural dynamics; no additional statistical test is needed. This also explains why these series in Rybski et al. (2006) resulted in "earlier detection" (to use their terminology). Had a sample size of 456 years been used in the above calculations (for the period 1400-1855) the estimate of StD[S] would be 0.025°C and would justify a difference among standard deviations thereof of about 0.06°C. Consequently, from Table 1 we can observe that two groups (one is JBBT98, MM03, MSHDK05 and the other one MBH99, B00, ECS02) are formed, each of which contain 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 changes

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. Firstly, it does not depend on a fitted model (as e.g. a linear fitting to the data). Secondly, it is flexible and convenient as it allows choosing the climatic time scale k and the lag l/k (defined on scale k). Thirdly, and more importantly, it yields a simple, general (not dependent on the process), convenient and exact expression for the standard deviation of the test statistic, which can be readily obtained from (1):

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

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)) dependent merely on the autocorrelation structure. It can be noticed that Rybski et al. (2006) did not use this expression but instead they derived an approximate yet more complex expression (denoting $\mathrm{StD}[D_{i,\,l}^{(k)}]$ as $\sigma(k, l)$), which in addition involves explicitly the lag one autocorrelation of the process at scale 1.

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) in practice 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, $\mathrm{StD}[D_{i,\,l}^{(k)}]$ is an increasing function of H but for $\rho >$ ~0.70 it becomes a decreasing function tending to zero as $\rho \rightarrow 1$ (because the second term dominates). More specifically, for high ρ $\mathrm{StD}[D_{i,\,l}^{(k)}]$ becomes smaller than in the case of the classical statistics (the value corresponding to $\rho = 0$). The situation is similar in the AR(1) case but $\mathrm{StD}[D_{i,\,l}^{(k)}]$ becomes decreasing function of ρ only for $\rho > 0.95$. Interestingly (and perhaps contrary to intuition), for $\rho >$ ~0.75, $\mathrm{StD}[D_{i,\,l}^{(k)}]$ is larger in the AR(1) case than in the SSS case.

In all this demonstration it was assumed that both σ and ρ are known. In practice, however both are unknown estimated from the sample. The picture changes drastically in this case. To estimate $\mathrm{StD}[D_{i,\,l}^{(k)}]$, one may be tempted to use the standard estimate s of σ that is used in classical statistics (this perhaps has been done in Rybski et al., as there is no mention of an alternative method). However, as explained above (eqn. (8)), in SSS statistics, s is strongly biased and $\tilde{\tilde{s}}$ should be used instead; so, 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 StD$[D_{i,\,l}^{(k)}]$ is an increasing function for virtually all domain of ρ. StD$[D_{i,\,l}^{(k)}]$ estimated with SSS statistics is greater than that estimated by classical statistics for $\rho > 0.3$ and the difference becomes dramatic for ρ approaching 1. The situation is different in the AR(1) process where, as seen in Figure 1(b), the use of classical statistical estimate practically has no effect unless $\rho > 0.95$.

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 significantly with ρ. For $\rho = 0.7$ the significance level becomes10^{-2} in the SSS case and 10^{-3} in the AR(1) case. For $\rho > 0.8$ both the SSS and the AR(1) processes give 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 behaviour.

Yet this modified analysis was based on the tacit assumption that the true value of H is known. But since this assumption is not true, the above methodology does not form 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 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). This justifies why Cohn and Lins (2005) used a Monte Carlo testing framework, which resulted 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; in this respect, the construction of an exact Monte-Carlo test based on $D_{i,\,l}^{(k)}$ is out of our current scope.

The detection problem

The above discussion shows that the detection and attribution problem should be studied in a framework admitting LTP, as also pointed out both by Cohn and Lins (2005) and Rybski et al. (2006), or at least a framework incorporating a high dependence structure, and that the classical IID framework should be abandoned. However, the whole problem may be more difficult than suggested in the latter work. Specifically, it requires more attention in avoiding classical statistical results that are not valid in an SSS framework (or even in processes with high autocorrelations), and an accurate description of all sources of uncertainty.

It may have some interest to apply the above “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 uncertainty bands, as discussed above.

Certainly, the detection and attribution problem will continue to be an attracting one in the years to come, as newer data will be accumulated. Before concrete conclusions can be drawn, a rigorous methodological framework should be built, in which SSS statistics should play a role. 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 as, no doubt, are both studies that motivated our Letter.

In our opinion the application of such a framework would give more reliable results on point and regional basis, rather than on global or “hemispheric” basis, thus avoiding the additional uncertainty resulting from integration over the globe. This is particularly the case for hydrological variables such as rainfall and runoff, for which the integration over the globe may not have any meaning. Obviously, the framework we are discussing will also be applicable for those hydrological processes, given that they are also characterized by LTP.

The building of such a framework cannot be based on merely statistical arguments, because, as we demonstrated above, even the presence of LTP can be disputable on purely statistical grounds. The fact that the recently reconstructed proxy time series of climate are consistent with the hypothesis of LTP does not necessarily mean that the hypothesis is true. Thus, we believe that better understanding and theoretical arguments are strongly needed to illustrate and justify the hypothesis. In this respect, perhaps the concepts of multiple-scale fluctuation (Koutsoyiannis, 2002), which is an extension of the common single-scale fluctuation perception, and the application of the physical and mathematical principle of maximum entropy, viz. maximum uncertainty (Koutsoyiannis, 2005) on a range of scales, may have some interest as they attempt to provide theoretical justification of LTP. Until a concrete framework can be established, we strongly endorse the following quotation by Cohn and Lins (2005): “From a practical standpoint, however, it may be preferable to acknowledge that the concept of statistical significance is meaningless when discussing poorly understood systems.”

References

Beran, J. (1994), *Statistics for Long-Memory Processes*, Volume 61 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, New York.

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

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

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 general circulation model control-run temperatures, Holocene, 8(4), 455– 471.

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

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.

Esper, J., E.R. Cook, P.J. Krusic, K. Peters, and F.H. Schweingruber (2003), Tests of the RCS method for preserving low-frequency variability in long tree-ring chronologies, Tree Ring Res., 59(2), 81-98.

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.

Kolmogorov, A. N. (1940), Wienersche Spiralen und einige andere interessante Kurven in Hilbertschen Raum, Comptes Rendus (Doklady) Acad. Sci. USSR (N.S.) 26, 115–118.

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

Koutsoyiannis, D. (2002), 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 and time scaling, *Hydrol. Sci. J.*, 50(3), 405-426.

Koutsoyiannis, D. (2006), Hurst, Joseph, colours and noises: The importance of names in an important natural behaviour, http://landshape.org/enm/?p=25.

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.

Maraun, D., H.W. Rust, and J. Timmer (2004), Tempting long-memory – on the interpretation of DFA results, Nonlinear Processes in Geophysics, 11: 495–503.

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

McIntyre, S., (2005), A Quote from Esper et al. [2003], http://www.climateaudit.org/?p=365.

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.

McIntyre, S., and R. McKitrick (2005), Hockey sticks, principal components, and spurious significance, Geophys. Res. Lett., 32, L03710, doi:10.1029/2004GL021750.

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*, edited by P. Doukhan et al., 1181 pp. 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.

Peng, C.-K., S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, and A.L. Goldberger, (1994), Mosaic organization of DNA nucleotides, Phys. Rev. E, 49, 1685-1689.

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.

Salas, J. D. (1993), Analysis and modeling of hydrologic time series, Handbook of Hydrology, edited by D. Maidment, Chapter 19, pp. 19.1-19.72, McGraw-Hill, New York.

Schmidt, G. and C. Amman (2005), Dummies guide to the latest “Hockey Stick” controversy, http://www.realclimate.org/index.php?p=121.

Stockwell, D. (2006a), Scale invariance for dummies, http://landshape.org/enm/?p=13.

Stockwell, D. (2006b), A new temperature reconstruction, http://landshape.org/enm/?p=15.

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

Vjushin, D., R.B. Govindan, R. A. Monetti, S. Havlin, and A. Bunde (2001), Scaling analysis of trends using DFA, Physica A, 302, 234–243.

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	JBBT98	MBH99	B00	ECS02	MM03	MSHDK05
				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 given by Rybski et al. (2006) except in the CRU series.

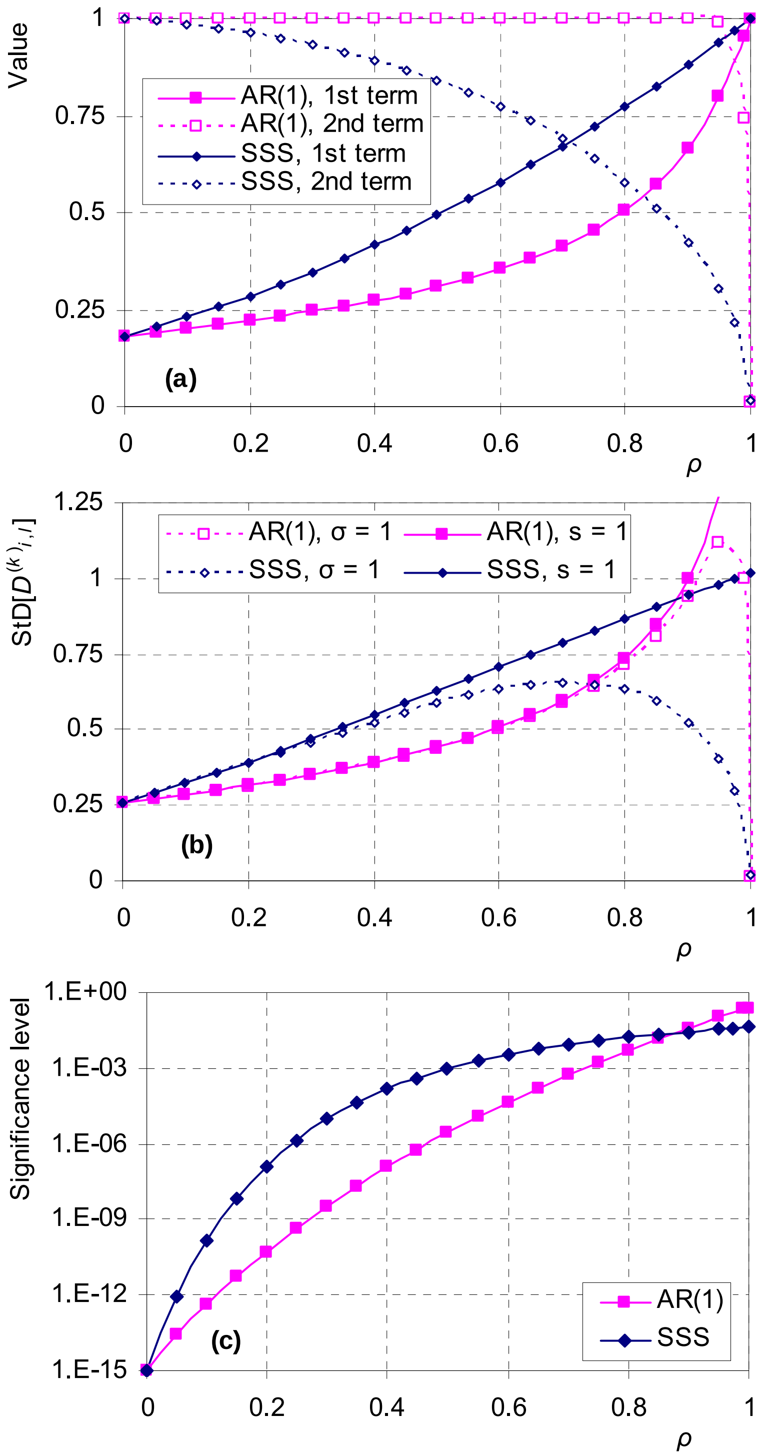

Figure 1 Variation with ρ of **(a)** the two multiplicative terms comprising StD[$D^{(k)}_{i,\,l}$] assuming $\sigma = 1$, **(b)** StD[$D^{(k)}_{i,\,l}$] assuming $\sigma = 1$ or $s = 1$ as indicated, and **(c)** the implied significance in rejecting the null hypothesis assuming that $s = 1$ and that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

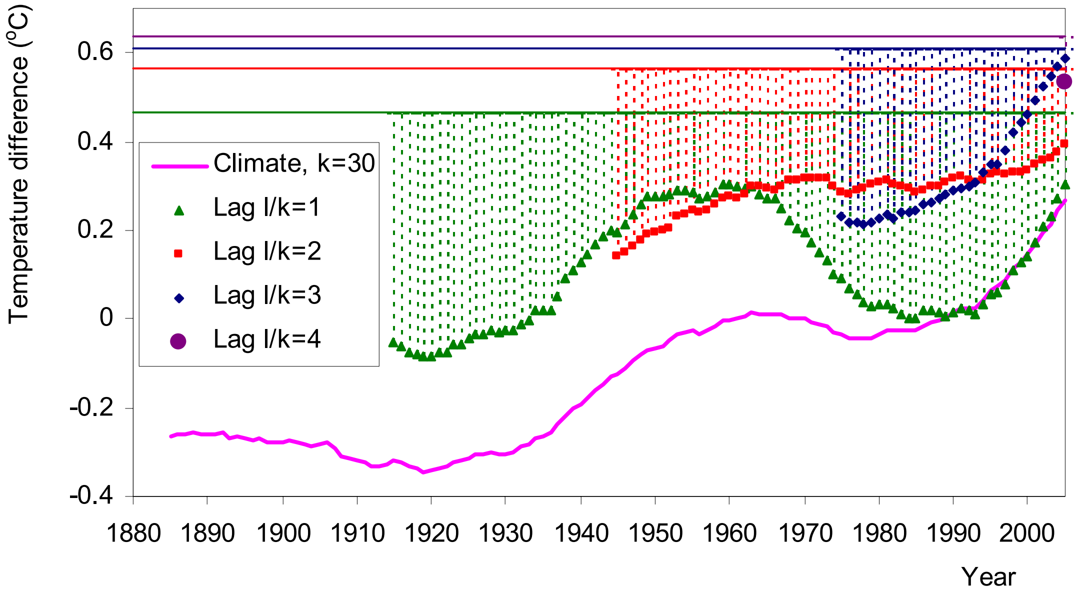

Figure 2 Graphical depiction of the pseudo-test based on StD[$D_{i,\,l}^{(k)}$] with known H. The continuous solid curve represents the CRU time series averaged over climatic scale k = 30. The series of points represent values of $D_{i,\,l}^{(k)}$ for the indicated lags l/k. Horizontal lines represent 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).