LTP: Looking Trendy—Persistently

Demetris Koutsoyiannis, National Technical University of Athens, Greece, itia.ntua.gr/dk/

Stochastics and its importance in studying climate

Probability, *statistics* and *stochastic processes*, lately described by the collective term *stochastics*, provide essential concepts and tools to deal with uncertainty useful for all scientific disciplines. However, there is at least one scientific discipline whose very domain relies on stochastics: Climatology. To refer to a popular definition of this domain by IPCC[1] (also quoted in Wikipedia):

> *Climate in a narrow sense is usually defined as the average weather, or more rigorously, as the statistical description in terms of the mean and variability of relevant quantities over a period of time ranging from months to thousands or millions of years. The classical period for averaging these variables is 30 years, as defined by the World Meteorological Organization.*

"Average", "statistical description", "mean", "variability", are all statistical terms. Several questions related to climate also involve probability, as exemplified in question 6 of the Introduction of this Climate Dialogue theme:[2]

> *Based on your statistical model of preference what is the probability that 11 of the warmest years in a 162 year long time series (HadCrut4) all lie in the last 12 years?*

Interestingly, similar probabilistic and statistical notions are implied in a recent President Obama statement:[3]

> *Yes, it's true that no single event makes a trend. But the fact is, the 12 hottest years on record have all come in the last 15.*

The latter statement highlights how important statistical questions are for policy matters and presumes some public perception of probability and statistics, which determines how the message is received.

I have no doubt that the average human being has some understanding of probability and statistics, not only thanks to education, but because life is uncertain and each of us needs to develop understanding of uncertainty and skills to cope with it. However, common experience and perception are mostly related to too simple uncertainties, like in coin tosses, dice throws and roulette wheels. Also education is mainly based on classical statistics in which:

- Consecutive events are independent to each other: the outcome of an event does not affect that of the next one.
- As a result, time averages tend to stabilize relatively fast: their variability, expressed by the probabilistic concept of variance, is inversely proportional to the length of the averaging period.

Adhering to classical statistics, when dealing with climate and other complex geophysical processes, is not just a problem of laymen. There are numerous research publications adopting, tacitly or explicitly, the independence assumption for systems in which it is totally inappropriate. Even the very definition of climate quoted above, particularly the phrase "*The classical period is 30 years*" historically reflects a perception of a constant climate[4, 5] and a hope that 30 years would be enough for a climatic quantity to get stabilized to a constant value—and this is roughly supported by classical statistics. In this perception a constant climate would be the norm and a deviation from the norm would be something caused by an extraordinary agent. The same static-climate conviction is evident in the "weather vs. climate" dichotomy (e.g. the "climate is what you expect, weather is what you get"; see critical discussions in Refs. 6, 7, 5).

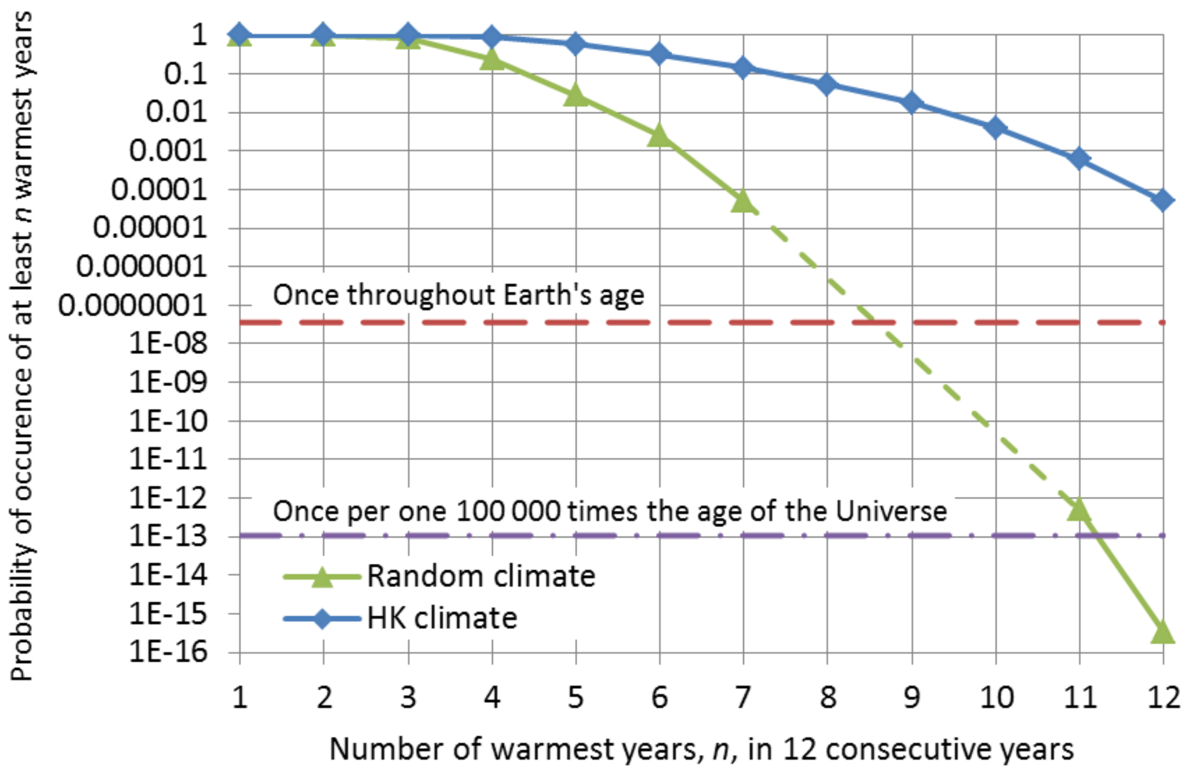

Figure 1 Probability that a 12-year period contains the specified number of warmest years (n) or more in a 162-year long period, as calculated assuming a random climate and a Hurst-Kolmogorov (HK) climate with Hurst parameter $H = 0.92$ (see text below for explanation of the latter).

Now let us pretend that, as in classical statistics, climate was synthesized by averaging random events without dependence and try to study on this basis the above question (slightly modified for reasons that will be explained later). So, what is the probability that, in a 162-year long time series, at least n (where $n = 1, 2, \ldots 12$) of the warmest years all lie in a 12-year long period? The reply is depicted in Figure 1 labelled "Random climate". The first seven points are calculated by Monte-Carlo simulation. For $n = 7$ years this probability is 0.00005 (1/20 000). The Monte Carlo simulation would require too much time to find the probability that all 12 warmest years are consecutive ($n = 12$), because this probability is really an astonishingly small number; but I was able to find it analytically and plotted it on the graph. I also approximated with analytical calculations the probability that at least 11 warmest years are clustered within 12 years. From the graph we can conclude that it is quite unlikely that more than 8-9 warmest years would cluster, even throughout the entire Earth's life (4.5 billion years separated in segments of 162 years). To have 11 warmest events clustering in a 12-year period we would need, on the average, one hundred thousand times the age of the Universe.

Is this overwhelming evidence that something extraordinary has occurred during our lives, or that the independence assumption leads to blatantly irrational results?

No one would believe that the weather this hour does not depend on that an hour ago. It is natural to assume that there is time dependence in weather. Therefore, we must study weather not on the basis of classical statistics, but we should rather use the notion of a stochastic process. Now, if we average the process to another scale, daily, monthly, annual, decadal, centennial, etc. we get other stochastic processes, not qualitatively different from the hourly one. Of course, as the scale of averaging increases the variability decreases—but not as much as implied by classical statistics. Naturally the dependence makes clustering of similar events more likely.

The first who studied clustering in natural processes was Harold Edwin Hurst, a British hydrologist who worked in the Nile for more than 60 years. In 1951 he published a seminal paper[8] in which he stated:

> *Although in random events groups of high or low values do occur, their tendency to occur in natural events is greater. This is the main difference between natural and random events.*

Herodotus said that the Egyptian land is "a gift of the Nile". Nile gave also hydrology and climatology invaluable gifts: one of them is the longest record of instrumental observations in history. Its water levels were measured in so-called Nilometers and archived for many centuries. In the 1920s Omar Toussoun, Prince of Alexandria, published a book[9] containing, among other things, annual minimum and maximum water levels of the Nile at the Roda Nilometer from AD 622 to 1921. Figure 2 depicts the time series of annual minimum levels up to 1470 (849 values; unfortunately, after 1470 there are substantial gaps). Climatic, i.e. 30-year average, values are also plotted. One may say that these values are not climatic in strict sense. But they are strongly linked to the variability of the climate of a large area, from Mediterranean to the tropics. And they are instrumental.

The clustering of similar events, more formally described as Long-Term Persistence (LTP) is obvious. For example, around AD 780 we have a group of low values producing a low climatic value, and around 1110 and 1440 we have groups of large values. Such grouping would not appear in a climate that would be the synthesis of independent random events. The latter would be more flat as illustrated by the synthetic example of Figure 3.

Another way of viewing the long-term variability of the Nile in Figure 2 is by using the notion of trends, irregularly changing from positive to negative and from mild to steep. The long instrumental Nile series may help those who prefer the view of variability in terms of trends to recognize "Nature's style [as] Naturally trendy" to invoke the title of a celebrated recent paper.[10]

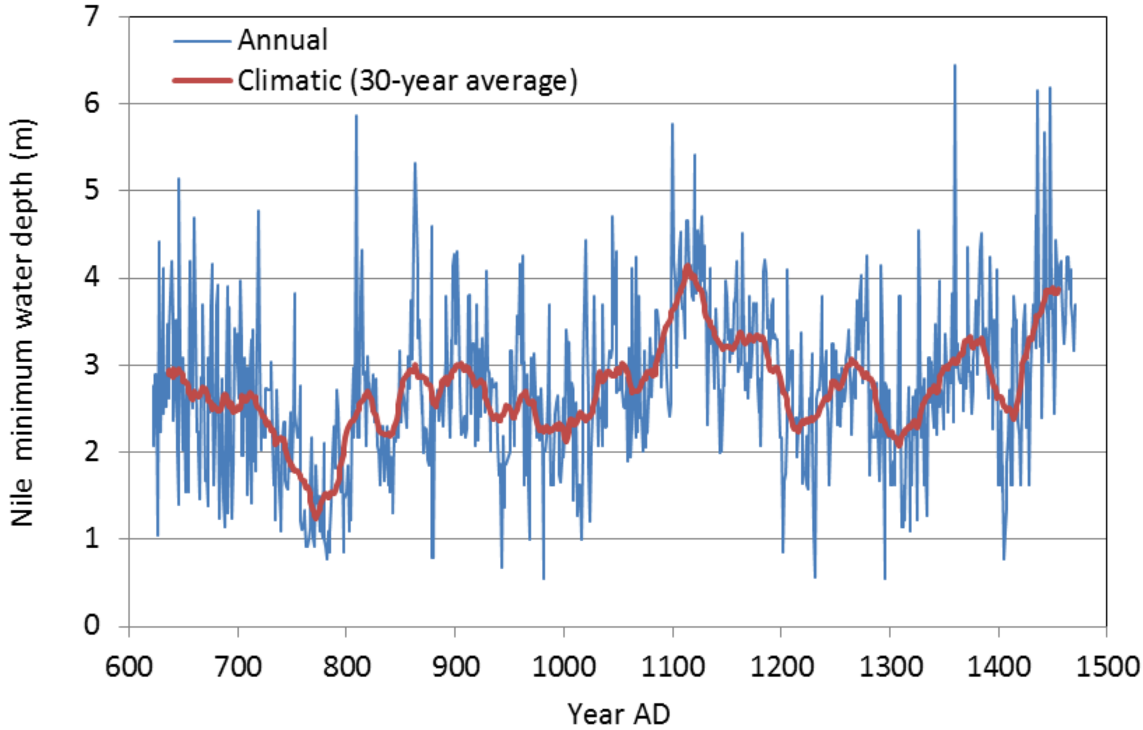

Figure 2 Nile River annual minimum water level at Roda Nilometer (from Ref. 9, here converted into water depths assuming a datum for the river bottom at 8.80 m), along with 30-year averages (centred). A few missing values at years 1285, 1297, 1303, 1310, 1319, 1363 and 1434 are filled in using a simple method from Ref. 11. The estimated statistics are mean = 2.74 m, standard deviation = 1.00 m, Hurst parameter = 0.87.

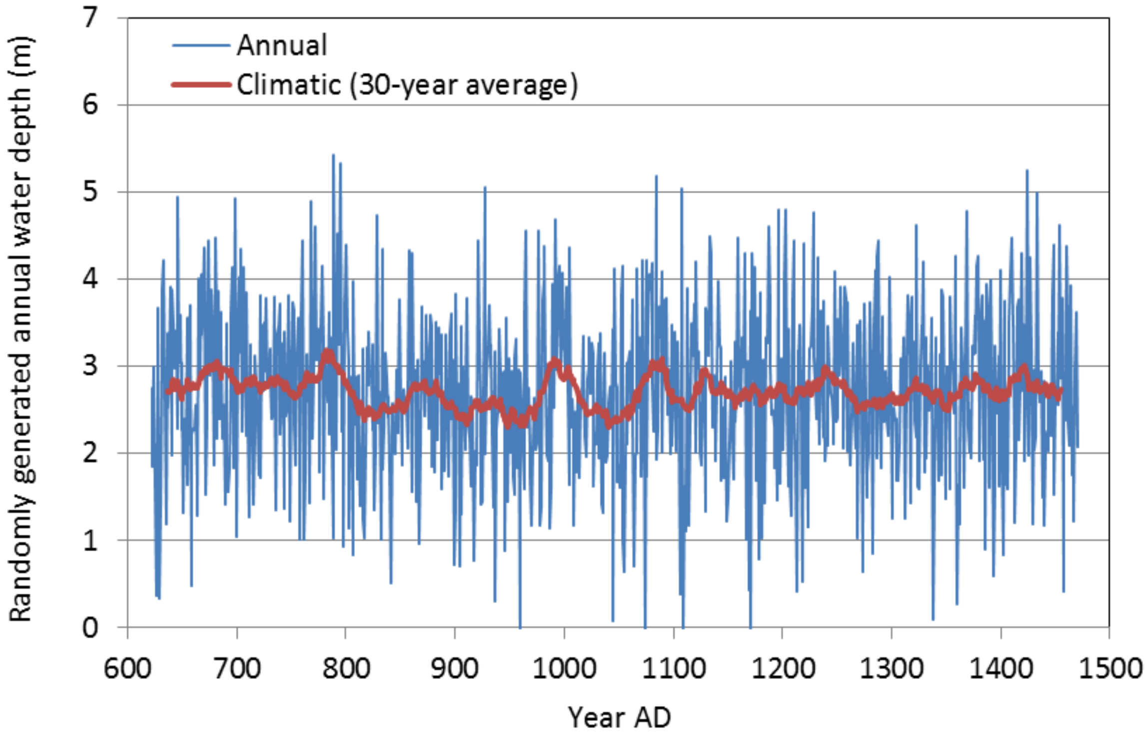

Figure 3 A synthetic time series from an independent (white noise) process with same statistics as those of the Nilometer series shown in the caption of Figure 2.

Seeking a proper stochastic model for climate

Variability over different time scales, trends, clustering and persistence are all closely linked to each other. The former is a more rigorous concept and is mathematically described by the variance (or the standard deviation) of the averaged process as a function of the averaging time scale, aka climacogram.[6,12] The variability over scale (the climacogram), is also one-to-one related (by transformation) to the stochastic concepts of dependence in time (the autocorrelation function) and the spectral properties (the power spectrum) of the process of interest.[6,12]

In white noise, i.e., the process characterized by complete independence in time, the variability is infinite at the instantaneous time scale (in technical terms its autocorrelation in continuous time is a Dirac Delta function). No variability is added at any finite time scale. Clearly, this is a mathematical construct which cannot occur in nature (the adjective "white", suggestive of the white light as a mixture of frequencies, is misleading; the spectral density of white noise is flat, while that of the white light is not).

A seemingly realistic stochastic process, which has been widely used for climate, is the Markov process, whose discrete time version is more widely known as the AR(1) process. The characteristic properties of this process are two:

- Its past has no influence on the future whenever the present is known (in other words, only the latest known value matters for the future).[13]
- It assumes a single characteristic time scale in which change or variability is created (but in contrast to the white noise, this time scale is non-zero and technically is expressed by the denominator of the exponent in an exponential function that constitutes its autocorrelation function). As a result, when the time scale of interest is fairly larger than this characteristic scale, the process behaves like white noise.

It is difficult to explain why this model has become dominant in climatology. Even these two theoretical properties should have hampered its popularity. How could the future be affected just by the latest value and not by the entire past? Could any geophysical process, including climate, be determined by just one mechanism acting on a single time scale?

The flow in a river (not necessarily the Nile) may help us understand better the multiplicity of mechanisms producing change and the multiplicity of the relevant time scales (see also Ref. 14):

- Next second: the hydraulic characteristics (water level, velocity) will change due to turbulence.
- Next day: the river discharge will change (even dramatically, in case of a flood).
- Next year: The river cross-section will change (erosion-deposition of sediments).
- Next century: The river basin characteristics (e.g. vegetation, land use) will change.
- Next millennium: All could be very different (e.g. the area could be glacialized).
- Next millions of years: The river may have disappeared (owing to tectonic processes).

Of course none of these changes will be a surprise; rather, it would be a surprise if things remained static. Despite being anticipated, all these changes are not predictable.

Does a plurality of mechanisms acting on different scales require a complex stochastic model? Not necessarily. A decade before Hurst detected LTP in natural processes, Andrey Kolmogorov,[15] devised a mathematical model which describes this behaviour using one parameter only, i.e. no more than in the Markov model. We call this model the Hurst-Kolmogorov (HK) model (aka fGn—for fractional Gaussian noise, simple scaling process etc.), while its parameter has been known as the Hurst parameter and is typically denoted by H. In this model, change is produced at all scales and thus it never becomes similar to white noise, whatever the time scale of averaging is.

Specifically, the variance will never become inversely proportional to time scale; it will decrease at a lower rate, inversely proportional to the power $(2 - 2H)$ of the time scale (nb. $0.5 \leq H < 1$, where the value $H = 0.5$ corresponds to white noise). A characteristic property of the HK process is that its autocorrelation function is independent of time scale. In other words if there is some correlation in the average temperature between a year and the next one (and in fact there is), the same correlation will be between a decade and the next one, a century and the next one, and so on to infinity. Why? Because there will always be another natural mechanism acting on a bigger scale, which will create change, and thus positive correlation at all lower scales (the relationship of change with autocorrelation is better explained in Ref. 6). The HK behaviour seems to be consistent with the principle of extremal entropy production.[16]

The Nilometer record described above is consistent with the HK model with $H = 0.87$. Are there other records of geophysical processes consistent with the HK behaviour? A recent overview paper[17] cites numerous studies where this behaviour has been verified. It also examines several instrumental and proxy climate data series related to temperature and, by superimposing the climacograms of the different series, it obtains an overview of the variability for time scales spanning almost nine orders of magnitude—from 1 month to 50 million years. The overall climacogram supports the presence of HK dynamics in climate with H at least 0.92. The orbital forcing (Milankovitch cycles) is also evident in the combined climacogram at time scales between 10 and 100 thousand years.

Statistical assessment of current climate evolution

Re-examining the statistical problem of 11 warmest years in 12 within 162 year period, now within an HK framework with $H = 0.92$, we will find spectacularly different results from those of the random climate, as shown in Figure 1. We may see, for example, that what, according to the classical statistical perception, would require the entire age of the Earth to occur once (i.e. clustering of 8-9 events) is a regular event for an HK climate, with probability on the order of 1-10%.

This dramatic difference can help us understand why the choice of a proper stochastic model is relevant for the detection of changes in climate. It may also help us realize how easy it is to fool ourselves, given that our perception of probability may heavily rely on classical statistics.

Figure 4 gives a close-up of the results, excluding the very low probabilities and also generalizing the "12-year period" to "N-year period" so that it can host, in addition to the Climate Dialogue statistical question 6, the results for the "Obama version" thereof as quoted above. In addition, Figure 4 is based on a slightly higher value of the Hurst coefficient, $H = 0.94$, as estimated by the *Least Squares* *based on Standard Deviation* method[18] for the HadCrut4 record. Both versions result in about the same answer: the probability of having 11 warmest years in 12, or 12 warmest years in 15, is 0.1%.

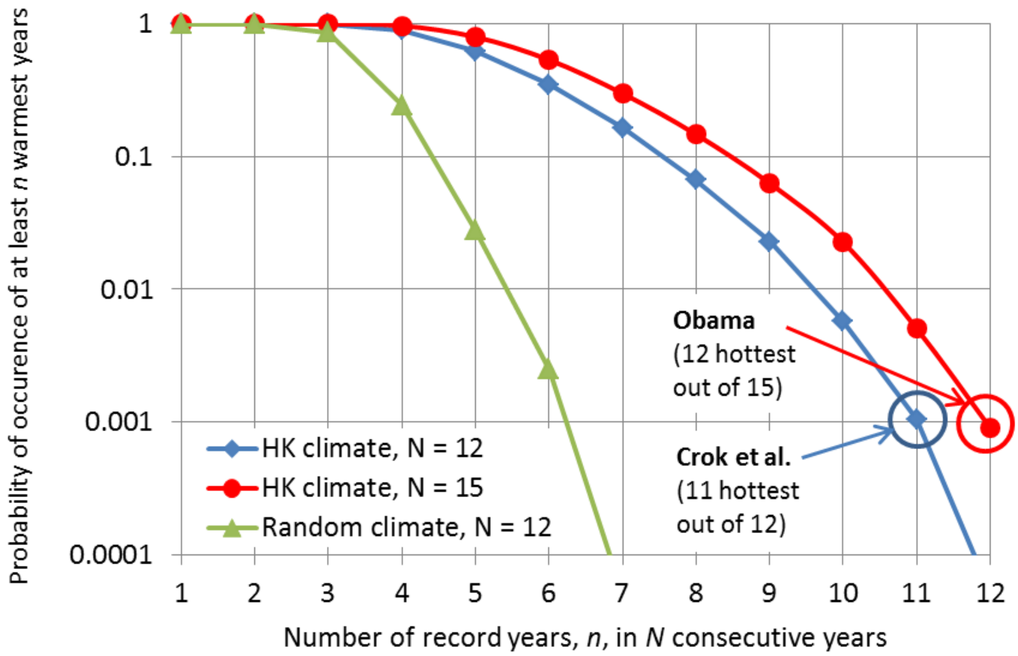

Figure 4 Probability that a *N*-year period, where *N* = 12 or 15, contains the specified number, *n*, of warmest years or more in a 162-year long period, calculated in the same manner as in Figure 1 with *H* = 0.94.

If we used the IPCC AR4 terminology[19] we would say that either of these events is *exceptionally unlikely* to have a natural cause. Interestingly, the present results do not contradict those of a recent study of Zorita, Stocker and von Storch,[20] who examined a similar question and concluded that:

> *Under two statistical null-hypotheses, autoregressive and long-memory, this probability turns to be very low: for the global records lower than p = 0.001…*

I note, though, that there are differences in the methodology followed here and that in Zorita *et al.*; for example, the analysis here did not consider whether the *N*-year period (where the *n* warmest years are clustered) is located in the end of the examined observation period or anywhere else in it (the reason will be explained below).

One may note that the above results, as well as those by Zorita *et al.*, are affected by uncertainty, associated with the parameter estimation but also with the data set itself. The data are altered all the time as a result of continuous adaptations and adjustments. Even the ranks of the different years are changing: for example in the CRU data examined by Koutsoyiannis and Montanari (2007)[21], 1998 was rank 1 (the warmest of all) and 2005 was rank 2, while now the ranking of these two years was inverted. But most importantly, the analysis is affected by the Mexican Hat Fallacy (MHF), if I am allowed to use this name to describe a neat example of statistical misuse offered by von Storch,[22] in which the conclusion is drawn that:

> *The Mexican Hat is not of natural origin but man-made.*

Von Storch[22] aptly describes the fallacy in these words:

> *The fundamental error is that the null hypothesis is not independent of the data which are used to conduct the test. We know a-priori that the Mexican Hat is a rare event, therefore the impossibility of finding such a combination of stones cannot be used as evidence against its natural origin. The same trick can of course be used to "prove" that any rare event is "non-natural", be it a heat wave or a particularly violent storm - the probability of observing a rare event is small.*

I believe that by rephrasing "*11 of the warmest years … all lie in the last 12 years*" into "*11 of the warmest years … all lie in a 12-year long period* " reduces the MHF effect, but I do not think it eliminates it. That is why I prefer other statistical methods of detecting changes[23], such as the tests proposed by Hamed[24] and by Cohn and Lins[10]. The former relies on a test statistic based on the ranks of all data, rather than a few of them, while the second considers also the magnitude of the actual change, not that of the change in the ranks.

Another test statistic was proposed by Rybski et al.,[25] and was modified to include the uncertainty in the estimation of standard deviation by Koutsoyiannis and Montanari[21], who also applied it for the CRU temperature data up to 2005. Note that, to make the test simple, the uncertainty in the estimation of *H* was not considered even in the latter version (thus it could rather be called a pseudo-test). Here I updated the application of this test and I present the results in Figure 5.

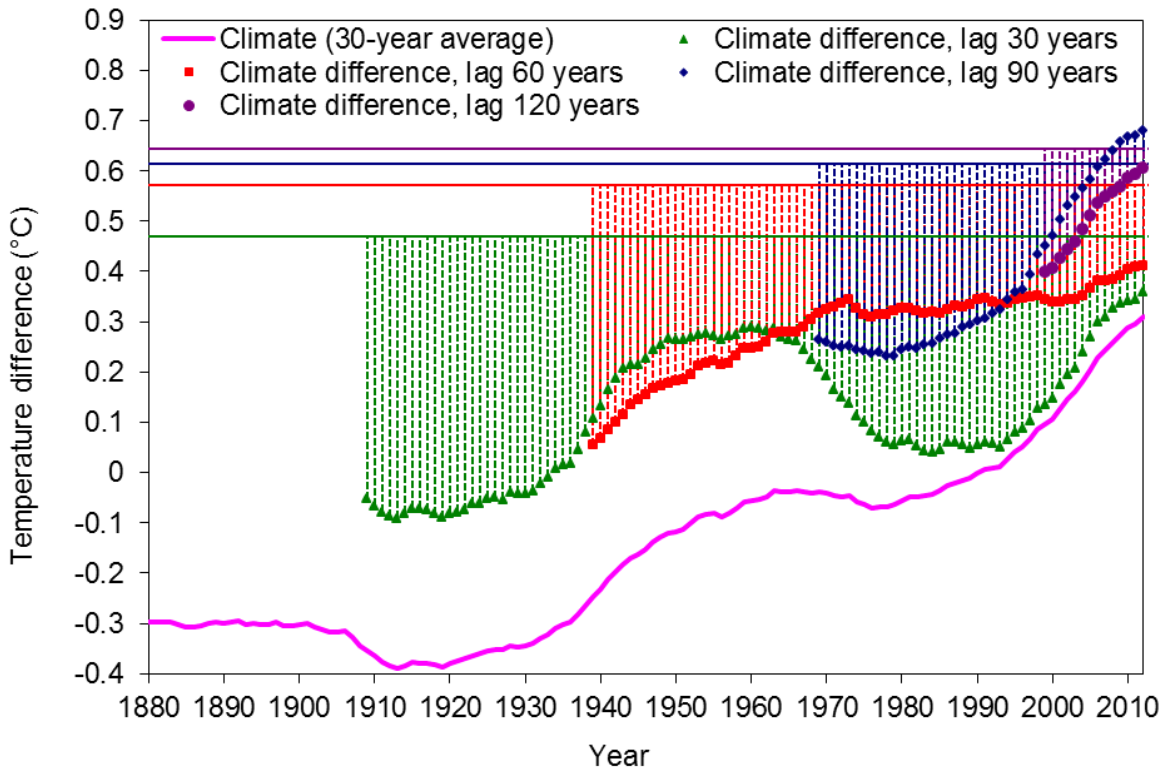

Figure 5 Updated Fig. 2 in Koutsoyiannis and Montanari[21] testing lagged climatic differences based on the HadCrut4 data set (1850-2012; see explanation in text).

The method has the advantages that it uses the entire series (not a few values), it considers the actual climatic values (not their ranks) and it avoids specifying a mathematical form of trend (e.g. linear). Furthermore, it is simple: First we calculate the climatic value of each year as the average of the present and the previous 29 years. This is plotted as a pink continuous line in Figure 5, where we can see, among other things, that the latest climatic value is 0.31°C (at 2012, being the average of HadCrut4 data values for 1983-2012), while the earliest one was –0.30°C (at 1879, being the average of 1850-79). Thus, during the last 134 years the climate has warmed by 0.61°C. Note that no subjective smoothing is made here (in contrast to the graphs by CRU), and thus the climatic series has length 134 years (but with only 5 non-overlapping values), while the annual series has length 163.

Our (pseudo)test relies on climatic differences for different time lags (not just that of the latest and earliest values). For example, assuming a lag of 30 years (equal to the period for which we defined a climatic value), the climate difference between 2012 and 1982 is 0.31°C – (–0.05°C) = 0.36°C, where the value –0.05°C is the average of years 1953-82. The value 0.36°C is plotted as a green triangle in Figure 5 at year 2012. Likewise, we find climatic differences for years 2011, 2010, …, 1909, all for lag 30. Plotting all these we get the series of green triangles shown in Figure 5. We repeat the same procedure for time lags that are multiples of 30 years, namely 60 years (red points), 90 years (blue points) and 120 years (purple points).

Finally, we calculate, in a way described in Ref. 21, the critical values of the test statistic, which is none other than the lagged climate difference. The critical values are different for each lag and are plotted as flat lines with the same colour as the corresponding points. Technically, the (pseudo)test was made as two-sided for significance level 1% but only the high critical values are plotted in the graph. Practically, as long as the points lie below the corresponding flat lines, nothing significant has happened. This is observed for the entire length of the lag-30 and lag-60 differences. A few of the last points of the lag-90 series exceed the critical value; this corresponds to the combination of high temperatures in the 2000s and low temperatures in the 1910s. But then all points of the lag-120 series lie again below the critical value, indicating no significant change.

Concluding remarks

Assuming that the data set we used is representative and does not contain substantial errors, the only result that we can present as fact is that in the last 134 years the climate has warmed by 0.6°C (nb., this is a difference of climatic—30-year average—values while other, often higher, values that appear in the literature refer to trends based on annual values). Whether this change is statistically significant or not depends on assumptions. If we assume a 90-year lag and 1% significance, it perhaps is; again I cannot be certain as the pseudo-test did not consider the uncertainty in H. Note, the 1% significance corresponds to ±2.58 standard deviations away from the mean; if we made it ±3 everything would become insignificant.

Irrespective of statistical significance, paleoclimate and instrumental data provide evidence that the natural climatic variability, the natural potential for change, is high and concerns all time scales. The mechanisms producing change are many and, in practice, it is more important to quantify their combined effects, rather than try to describe and model each one separately.

From a practical point of view, it could be dangerous to pretend that we are able to provide credible quantitative description of the mechanisms, their causes and effects, and their combined consequences: We know that the mechanisms and their interactions are nonlinear, as well as that the climate model hindcasts are poor.[26,27] Indeed, it has been demonstrated that, particularly for runoff that is mostly relevant for water availability and flood risk, deterministically projected future traces can be too flat in comparison to changes that can be expected (and stochastically generated) admitting stationarity and natural variability characterized by HK dynamics[28].

References

[1] IPCC (2007), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996 pp (Annex 1, Glossary: http://www.ipcc.ch/pdf/assessment-report/ar4/wg1/ar4-wg1-annexes.pdf).
[2] Climate Dialogue Editorial Staff (2013), How persistent is the climate system and what is its implication for the significance of observed trends and for internal variability? (This blog post).
[3] Obama's 2013 State of the Union Address, The New York Times, http://www.nytimes.com/2013/02/13/us/politics/obamas-2013-state-of-the-union-address.html?pagewanted=3&_r=0
[4] Lamb, H. H. (1972), *Climate: Past, Present, and Future*, Vol. 1, Fundamentals and Climate Now, Methuen and Co.
[5] Lovejoy, S., and D. Schertzer (2013), The climate is not what you expect, *Bull. Amer. Meteor. Soc.*, doi: 10.1175/BAMS-D-12-00094.
[6] Koutsoyiannis, D. (2011), Hurst-Kolmogorov dynamics and uncertainty, *Journal of the American Water Resources Association*, 47 (3), 481–495.
[7] Koutsoyiannis, D. (2010), A random walk on water, *Hydrology and Earth System Sciences*, 14, 585–601.
[8] Hurst, H.E. (1951), Long term storage capacities of reservoirs, Trans. Am. Soc. Civil Engrs., 116, 776–808.
[9] Toussoun, O. (1925). Mémoire sur l'histoire du Nil, in Mémoires a l'Institut d'Egypte, vol. 18, pp. 366-404.
[10] Cohn, T. A., and H. F. Lins (2005), Nature's style: Naturally trendy, *Geophys. Res. Lett.*, 32, L23402, doi: 10.1029/2005GL024476.
[11] Koutsoyiannis, D., and A. Langousis (2011), Precipitation, *Treatise on Water Science*, edited by P. Wilderer and S. Uhlenbrook, 2, 27–78, Academic Press, Oxford (p. 57).
[12] Koutsoyiannis, D. (2013), Encolpion of stochastics: Fundamentals of stochastic processes, 30 pages, National Technical University of Athens, Athens, http://itia.ntua.gr/1317/, accessed 2013-04-17.
[13] Papoulis, A. (1991), Probability, *Random Variables and Stochastic Processes*, 3rd edn., McGraw-Hill, New York (p. 635).
[14] Koutsoyiannis, D. (2013), Hydrology and Change, *Hydrological Sciences Journal* (accepted with minor revisions; currently available in the form of an IUGG Plenary lecture, Melbourne 2011, http://itia.ntua.gr/1135/, accessed 2013-04-17).
[15] Kolmogorov, A. N. (1940). Wiener spirals and some other interesting curves in a Hilbert space, *Dokl. Akad. Nauk SSSR*, 26, 115-118. English translation in: Tikhomirov, V.M. (ed.) 1991. *Selected Works of A. N. Kolmogorov: Mathematics and mechanics*, Vol. 1, Springer, 324-326.
[16] Koutsoyiannis, D. (2011), Hurst-Kolmogorov dynamics as a result of extremal entropy production, *Physica A: Statistical Mechanics and its Applications*, 390 (8), 1424–1432.
[17] Markonis, Y., and D. Koutsoyiannis (2013), Climatic variability over time scales spanning nine orders of magnitude: Connecting Milankovitch cycles with Hurst–Kolmogorov dynamics, *Surveys in Geophysics*, 34 (2), 181–207.
[18] Tyralis, H., and D. Koutsoyiannis (2011), Simultaneous estimation of the parameters of the Hurst-Kolmogorov stochastic process, *Stochastic Environmental Research & Risk Assessment*, 25 (1), 21–33.
[19] As Ref. 1, p. 23.
[20] Zorita, E., T. F. Stocker, and H. von Storch (2008), How unusual is the recent series of warm years?, *Geophys. Res. Lett.*, 35, L24706, doi: 10.1029/2008GL036228.
[21] Koutsoyiannis, D., and A. Montanari (2007), Statistical analysis of hydroclimatic time series: Uncertainty and insights, *Water Resources Research*, 43 (5), W05429, doi: 10.1029/2006WR005592.
[22] von Storch, H. (1999), Misuses of statistical analysis in climate research, in *Analysis of Climate Variability, Applications of Statistical Techniques*, Proceedings of an Autumn School organized by the Commission of the European Community, Edited by H. von Storch and A. Navarra, 2nd updated and extended edition, http://www.hvonstorch.de/klima/books/SNBOOK/springer.pdf, accessed 2013-04.

[23] Koutsoyiannis, D. (2003), Climate change, the Hurst phenomenon, and hydrological statistics, *Hydrological Sciences Journal*, 48 (1), 3–24.
[24] Hamed, K. H. (2008), Trend detection in hydrologic data: The Mann-Kendall trend test under the scaling hypothesis, Journal of Hydrology, 349(3-4), 350-363.
[25] 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.
[26] Anagnostopoulos, G. G., D. Koutsoyiannis, A. Christofides, A. Efstratiadis and N. Mamassis (2010), A comparison of local and aggregated climate model outputs with observed data, *Hydrological Sciences Journal*, 55 (7), 1094–1110.
[27] Koutsoyiannis, D., A. Christofides, A. Efstratiadis, G. G. Anagnostopoulos, and N. Mamassis (2011), Scientific dialogue on climate: is it giving black eyes or opening closed eyes? Reply to “A black eye for the Hydrological Sciences Journal” by D. Huard, *Hydrological Sciences Journal*, 56 (7), 1334–1339.
[28] Koutsoyiannis, D., A. Efstratiadis, and K. Georgakakos (2007), Uncertainty assessment of future hydroclimatic predictions: A comparison of probabilistic and scenario-based approaches, *Journal of Hydrometeorology*, 8 (3), 261–281.