The Hurst phenomenon and climate

Demetris Koutsoyiannis
Department of Water Resources, Faculty of Civil Engineering, National Technical University of Athens, Greece
(dk@itia.ntua.gr)

Tim Cohn
U.S. Geological Survey, Reston, Virginia, USA
(tacohn@usgs.gov)
Hurst-Kolmogorov pragmaticity and climate

Demetris Koutsoyiannis
Department of Water Resources, Faculty of Civil Engineering, National Technical University of Athens, Greece
(dk@itia.ntua.gr)

Tim Cohn
U.S. Geological Survey, Reston, Virginia, USA
(tacohn@usgs.gov)
1. What is the Hurst-Kolmogorov (HK) pragmaticity?
An introductory example

Proxy data (Northern Hemisphere temperature from Moberg et al., 2005)
Definition

- Hurst-Kolmogorov pragmaticity is
  - the recognition that real world processes behave differently from an ideal roulette wheel,
  - where the differences mainly involve the behaviour of local averages (long excursions from global mean).

- Observation 1: “Real world” processes include Mother Nature’s processes (physical, geophysical, astrophysical etc.) and also human-related processes (socio-economical, technological).

- Observation 2: Even a real-world roulette wheel behaves differently from an ideal roulette wheel. To prevent exploitation by a few players who are able to see the difference and to model the behaviour of the real-world roulette wheel, the casinos monitor the performance of their wheels, and rebalance and realign them regularly to keep the result of the spins as nearly random as possible (en.wikipedia.org/wiki/Roulette).

- Observation 3: The recognition of a different behaviour of natural processes calls for quantification of the differences.
Multi-scale stochastic properties of a HK process

A natural process usually evolves in continuous time \( t : X(t) \)

... but we observe or study it in discrete time, averaging it over a fixed time scale \( k \) and using discrete time steps \( i = 1, 2, \ldots \)

\[
X_i^{(k)} := \frac{1}{k} \int_{(i-1)k}^{ik} X(t) \, dt
\]

<table>
<thead>
<tr>
<th>Properties of the HK process</th>
<th>At an arbitrary observation scale ( k = 1 ) (e.g. annual)</th>
<th>At any scale ( k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>( \sigma = \sigma^{(1)} )</td>
<td>( \sigma^{(k)} = k^{H-1} \sigma )</td>
</tr>
<tr>
<td>Autocorrelation function (for lag ( j ))</td>
<td>( \rho_j = \rho_j^{(1)} = \rho_j^{(k)} \approx H(2H - 1) j^{2H-2} )</td>
<td></td>
</tr>
<tr>
<td>Power spectrum (for frequency ( \omega ))</td>
<td>( \mathcal{S}(\omega) = \mathcal{S}^{(1)}(\omega) \approx \frac{4}{(1-H)} \sigma^2 (2\omega)^{1-2H} )</td>
<td>( \mathcal{S}^{(k)}(\omega) \approx \frac{4}{(1-H)} \sigma^2 k^{2H-2} (2\omega)^{1-2H} )</td>
</tr>
</tbody>
</table>

A long tail (or fat tail)

All equations are power laws of scale \( k \), lag \( j \), frequency \( \omega \)
Illustration of the differences between “roulette” and HK statistics

- Assume we have \( n = 100 \) years of annual observations \( x_i^{(1)} (i = 1, \ldots, 100) \) of a natural process with \( \sigma = 1 \) and \( H = 0.94 \) (= the value estimated for Moberg et al. proxy); the mean \( \mu \) is unknown.
- The value \( x_{1}^{(100)} = (x_{1}^{(1)} + \ldots + x_{1}^{(100)})/100 \) is the sample average.
- We can estimate the true mean \( \mu \) as the sample average \( x_{1}^{(100)} \).
- As we have only one value \( x_{1}^{(100)} \) there is no numerical procedure to calculate the uncertainty of the mean. However, this can be derived theoretically by \( \sigma^{(100)} \).
- According to classical statistical law \( \sigma^{(k)} = \sigma / \sqrt{k} \), so \( \sigma^{(100)} = 0.10 \).
- According to HK statistics, \( \sigma^{(k)} = \sigma / k^{1-H} \), so \( \sigma^{(100)} = 0.76 \).
- To attain HK uncertainty as low as in classical (“roulette”) statistics, i.e. \( \sigma^{(n^*)} = 0.10 \), we would need a sample size \( n^* = n^{1/2(1-H)} = 5 \times 10^{16} \) years (3 million times the age of the universe).
- The “roulette”-equivalent sample size \( n' \) of a \( n = 100 \) years sample exhibiting HK behaviour is \( n' = n^{2(1-H)} = 1.74 \) (< 2).
Can we still reconcile Nature with probability and statistics?

In the Aesop’s Fable, the Fox, whose tail was cut in a trap, tries to convince all the other Foxes that being tailless would be much more attractive and convenient.
Remedy 1: The Procrustean solution

- Long tails are not good. (Manicheism?)
- Long tails are an artefact.
- Cut the long tails.
- Short tails can be convenient.

- It is infeasible to cut the tails fully, i.e. to make autocorrelation disappear.
- Therefore, a short-tail, or exponential autocorrelation has been regarded as acceptable.
- This corresponds to a Markovian dependence ($\rho_j = \rho^j$), defined as the dependence in which the future does not depend on the past if the present is known.
- This Markovian behaviour has been put by many (e.g., Mann and Lees, 1996) as a postulate for Nature (cf. the pre-Keplerian postulate that the astral bodies should follow cyclical orbits).
Cutting the long tails – Method 1

- Assume a convenient dichotomy: “Actual” = “deterministic” + “stochastic”.

- Fit a series of linear segments to the time series (using statistical methods).

- Call the fitted lines “deterministic trends” or “nonstationarities”.

- Subtract the “trends” from the “actual” time series to get a “stationary stochastic noise”.

An almost “roulettized” time series (“Almost” because it has autocorrelation – but with short tail)
Cutting the long tails – Method 2

- Use same assumptions as in Method 1 but replace “linear trends” with “shifting means”.
- The methodology is the same and the result similar.

An almost “roulettized” time series (“Almost” because it has autocorrelation – but with short tail)
Is there anything wrong with procrusteanism?

- Theoretical problem
  - “Nonstationarity” is by definition a deterministic function of time.
  - A deterministic function is a function that can be produced by deduction, independently of the data (a priori; e.g. by a model that could predict them).
  - In contrast, the “trends” and “shifts” in the means were inferred by induction based on the data (a posteriori).
  - Hence these fitted lines are not “deterministic” and do not represent nonstationarity.
  - The method implies a confusion between the observed past (hence deterministically known, so that “trends” are no more deterministic than the remaining “noise”) and the unknown future, whose prediction (deterministic or stochastic) is sought.

- Practical problem
  - Given that the “trends” or “shifts” are unpredictable, their existence implies higher uncertainty.
  - However, the “detrending” or “roulettization” gives a false message of reduced uncertainty (initial StD 0.22, final 0.16).

See also: Koutsoyiannis (2006).
Remedy 2: HK statistics

- Recognize that natural processes are not roulette processes.
- Admit the existence of long tails.
- Make “long tail” an object of science – Investigate causes and study consequences.
- Do not apply “truncation operations” to the observed time series.
- Adapt statistics to describe the natural processes.
2. Historical evolution of the (celebrated) Hurst-Kolmogorov law and its (unfortunate) terminology
Hurst (1950)

- The motivation of Hurst was the design of the High Aswan Dam on the Nile River.
- However the paper was theoretical and explored numerous data sets of diverse fields.
- Hurst did not recognize any peculiarity of Nature nor did he use terms such as “long memory”, “long-term persistence”, “long-range dependence” etc.
- However, he observed that: “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”.

Three things were unfavourable for the dissemination/prevalence of Hurst’s study:
- Its direct connection with reservoir storage.
- The use of the complicated “range” statistic.
- Its association with the Nile.
Kolmogorov (1940)

- Kolmogorov studied the stochastic process that describes the behaviour to be discovered in geophysics 10 years later by Hurst.
- The proof of the existence of this process is important, because several researches, ignorant of Kolmogorov’s work, regarded Hurst’s findings as inconsistent with stochastics and as numerical artefacts.
- It is also important that Kolmogorov’s motivation was turbulence (transfer/dissipation of energy among scales in turbulent flows), manifesting a more general applicability of the HK law.

- It is unfortunate that Kolmogorov’s work did not become widely known.
- The process named by Kolmogorov as Wiener’s Spiral (Wienersche Spiralen) is called today “Self-similar process”, one case of which is what Mandelbrot called later “fractional Brownian motion”.
- Mandelbrot was aware of Kolmogorov’s work, as he says: “we discovered (while writing our paper) that fBm’s have already been considered (implicitly) by Kolmogorov and others” (Mandelbrot and Van Ness, 1968).
This is a purely mathematical paper.

It mentions an application to diffusion processes (but not in turbulence and geophysics).

No reference to Kolmogorov (1940) is given.

The influence of the paper was minimal.

The term *semi-stable process*, standing for what we call today a self-similar process, was not widely accepted.
Mandelbrot & Wallis (1968); Mandelbrot & van Ness (1968)

Mandelbrot offered the most complete theoretical studies, which were also the most influential.

Names given by Mandelbrot that perhaps became obstacles in the widespread recognition of HK:

- Joseph effect (direct connection with the Nile; implies some mystery).
- Fractional Brownian motion, fractional Gaussian noise (too technical and peculiar, not understandable; Nature’s verses are not noises).
- Short memory/long memory (wrong message leading to incorrect understanding of the physical mechanism – a long memory is not realistic).
Other common names – with some problems

- *Hurst phenomenon* ("phenomenon" may imply that Mother Nature’s reality is something extraordinary or exceptional).
- *Long-range dependence* (better than "long memory" but perhaps misleading as it does not point to any physical mechanism).
- *Long-term persistence* (implies some mystery or peculiarity).
- *Scaling behaviour* (concise and fashionable term, expressing the equivalence of time scales in this behaviour – but scaling is not a physical mechanism but a result of one or more other physical mechanisms or principles).
- *Fractional ARIMA process (FARIMA or ARFIMA)* (too algorithmic-oriented; no information that helps in understanding).
- *Brown noise* (perhaps points to the fact that brown is not one of the colours of the visible light spectrum but a mixture of colours and also suggestive of "Brownian").
Klemes (1974)

- Klemes was the first to point out a simple conceptual explanation (long term changes) for HK.
- He pointed out that the "infinite memory" is misleading (in fact it is an effect, not the cause).

Unfortunate developments with Klemes’s paper:
- Klemes made extensive use of the term “nonstationarity” for the changes in the mean he assumed.
- Such changes do not necessarily imply nonstationarity. On the contrary, changes in the mean without being deterministic functions of time (as is the case in natural processes) result in a stationary process.
- While Klemes did point out that his final models were in fact stationary, he kept the term “nonstationary” for all changes in the mean; the readers may have overlooked the stationarity of Klemes’s final model.
- An extension of Klemes’s idea using multiple scales of changes or fluctuations within a stationary setting results in a HK process (Koutsoyiannis, 2002).

It is shown that the Hurst phenomenon is not necessarily an indicator of infinite memory of a process. It can also be caused by nonstationarity in the mean and by random walks with one absorbing barrier, which often arise in natural storage systems. Attention is drawn to the fact that inferences about physical features of a process, based on operational models, can be not only inaccurate but grossly misleading.
Leland et al. (1994, 1995)

- These are the most cited papers on HK (> 3500 citations; cf. Hurst ~1100, Mandelbrot & van Ness ~1900).
- Using long data series they find $H$ between 0.85 and 0.95 for Ethernet LAN traffic.

On the Self-Similar Nature of Ethernet Traffic

Will E. Leland, Member, IEEE, Murad S. Taqqu, Member, IEEE, Walter Willinger, and Daniel V. Wilson, Member, IEEE

Abstract—We demonstrate that Ethernet LAN traffic is statistically self-similar, that none of the commonly used traffic models is able to capture this fractal-like behavior, that such behavior has serious implications for the design, control, and analysis of high-speed, cell-based networks, and that aggregating streams of such traffic typically intensifies the self-similarity ("burstiness") instead of smoothing it. Our conclusions are supported by a rigorous statistical analysis of hundreds of high-quality Ethernet traffic measurements collected between 1989 and 1993, coupled with a discussion of the underlying mathematical and statistical properties of self-similarity and their relationships with traffic measurements presented in [14]. Moreover, we illustrate some of the most striking differences between self-similar models and the standard models for packet traffic currently considered in the literature. For example, our analysis of the Ethernet data shows that the generally accepted argument for the "Poisson-like" nature of aggregate traffic, namely, that aggregate traffic becomes smoother (less bursty) as the number of traffic sources increases, has very little to do with reality. In fact, using the degree of self-similarity (which typically

1992 on several Ethernet LANs at the Bellcore Morristown Research and Engineering Center (MR). Leland and Wilson (1991) present a preliminary statistical analysis of this unique high-quality data and comment in detail on the presence of "burstiness" across an extremely wide range of time scales: traffic "spikes" ride on longer-term "ripples", that in turn ride on still longer term "swells", etc. This self-similar or apparently fractal-like behavior of aggregate Ethernet LAN traffic is very different both from conventional telephone traffic and from currently considered formal models for packet traffic (e.g., pure Poisson or Poisson-related models such as Poisson-batch or Markov-Modulated Poisson processes (Heffes and Lucantoni (1986)), packet-train models (Jain and Routhier (1986)), and

Koutsoyiannis & Cohn, The Hurst phenomenon and climate
Terminology proposed

- **Hurst-Kolmogorov pragmaticity** (see definition in the beginning)
  - Alternative/related names: *Hurst-Kolmogorov reality, Hurst-Kolmogorov behaviour, Multi-scale fluctuation* (pointing to the fact that this behaviour is none other than fluctuation of processes on multiple scales, including very large scales).

- **Hurst-Kolmogorov law** (the power law defining the dependence of standard deviation on scale; this produces also the power laws of the autocorrelation function and power spectrum).
  - Related name: *Asymptotic Hurst-Kolmogorov law* (when the law holds asymptotically for large scales).

- **Hurst-Kolmogorov process** (the stochastic process in discrete time defined by the Hurst-Kolmogorov law with properties as described in the relevant table).
  - Alternative names: *Simple Scaling Stochastic (SSS) Process, Stationary intervals of a self-similar process* (precise but complicated); *fractional Gaussian noise* (if the process is Gaussian; also complicated).

- **Hurst-Kolmogorov cumulative process** (the cumulative of the Hurst-Kolmogorov process; can be defined in continuous time).
  - Alternative name: *Self-similar process.*
3. Applications to climate
Nilometer annual minimum water levels

- The longest available instrumental hydroclimatic data set (813 years).
- Hurst coefficient $H = 0.84$.
- The same $H$ is estimated from the simultaneous record of maximum water levels and from the modern record (131 years) of the Nile flows at Aswan.

The classical statistical estimator of standard deviation was used, which however is biased for HK processes. Bias was determined by Monte Carlo simulation.
Moberg et al. proxy series

Suggests an HK behaviour with a very high Hurst coefficient: $H = 0.94$

The classical statistical estimator of standard deviation was used, which however is biased for HK processes. Bias was determined by Monte Carlo simulation.
Global temperature: Global warming?

From Climatic Research Unit (CRU) Information Sheets
Déjà vu: the global cooling hypothesis

Global temperature: Spectral properties

Power law, suggesting an HK process

Spectral densities higher than annual at periods ~5, 20, 50 years

Monthly average (°C)

Power spectrum, \( s(\omega) \)

Frequency, \( \omega \) (years\(^{-1}\))

\[ \frac{1}{\omega} = \begin{cases} \approx 50 & \text{~months} \\ \approx 20 & \text{~years} \\ \approx 5 & \text{~years} \\ \approx 3 & \text{~years} \\ \approx 2 & \text{~years} \\ 1 & \text{~year} \\ 0.5 & \text{~year} \end{cases} \]
Global temperature: HK properties

- Suggests an HK behaviour (see also Cohn and Lins, 2005, and Koutsoyiannis and Montanari, 2007).
- The same (very high) Hurst coefficient as in Moberg et al. proxy: $H = 0.94$
Evolution of the global temperature in the last ten years (from CRU; combined land and marine)

Data from www.cru.uea.ac.uk/cru/data/temperature/hadcrut3gl.txt

Even merely these 10 years indicate a HK behaviour with $H = 0.92$!
Global temperature: How probable are eight or nine warmest years in a decade?

- "Roulette" climate
- Hurst-Kolmogorov climate

Full colours: Simulation results; Light colours: extrapolations using probability theory

Regular behaviour at significance 1%

Extremely improbable

Simulation length: 158 years; Average simulated $H$: 0.93

Probability of $n$ record years in a decade

Number of record years in a decade, $n$
4. Concluding remarks
Does HK imply higher predictability? (in comparison to “roulette” climate)

- Definitely **no**, particularly on climatic scales.
- On the contrary, it implies much lower predictability on scales longer than 2-3 years.
- Dependence improves predictability only at small scales and short lead times.
- Particularly HK dependence dramatically reduces predictability on climatic (e.g. 30-year) scales.

![Entropy vs Scale](image)

- Unconditional entropy: No past observation available
- Conditional entropy: All past is observed
- “Roulette” climate is indifferent to past observations
- See details in Koutsoyiannis (2005), K. et al. (2007)

Moberg et al. proxy series

\[ H = 0.94, \sigma_{HK} = 0.38^\circ C - \sigma_{\text{classical}} = 0.28^\circ C \]
Do we have enough data points to estimate $H$?

- HK behaviour is hidden if samples are small ($n < \sim 100$; see e.g. Koutsoyiannis, 2003).
- The estimation of $H$ is uncertain even for $n > 100$.
- However the HK behaviour seems to be a generalized pragmaticity.
  - Not only in the Nile but in all rivers.
  - Not only in hydrology but also in climatology.
  - Not only in climatology but also in all geophysical processes.
  - Not only in geophysical processes but also in physical processes.
  - Not only in physical processes but also in human-related (e.g. technological and economical) processes.
- Those who wish to have enough data points, they can experiment with other fields: Turbulence and electronics can provide arbitrary long data sets.
Can we prove statistically that $H$ is significantly different from 0.5?

- Definitely, **yes**, in all of the cases, in which the null hypothesis involves pure randomness (“roulette” behaviour); see Cohn & Lins (2005).
  - In fact this is indirectly done in **all abundant publications that detect “trends”** in hydroclimatic time series (usually locating alternating trends in different time periods or different sites of a region; see Hamed, 2008).
  - Just **replace the fallacious alternative hypothesis** “deterministic trend” with “$H > 0.5$”.
- Perhaps, **no**, if the null hypothesis contains another type of dependence.
  - In the latter case, neither can one reject the hypothesis that $H$ is significantly different from, say, 0.9 (an estimated value).
  - Suggestion: Instead of significance testing, take **Jaynes’s** (2003) **approach**: calculate the odds of the two alternative hypotheses and choose the one with higher probability.
Is HK supported by a climatological explanation rooted in the physics?

- Which physics?
  - Is physics just Newton’s laws and mechanistic analogues?
  - Does physics include thermodynamics and entropy?
  - Does entropy have a conceptual definition (Clausius, ~1850, \( d\varphi = dQ/T \)) or a probabilistic definition (Boltzmann-Gibbs-Shannon-Jaynes, ~1870-1960, \( \varphi = E[-\ln(f(x))] \))?

- Explanation 1 (more conceptual): **Multi-scale fluctuation**
  - No **conservation laws** hold for natural quantities other than mass, momentum, energy.
  - So any other process need not have a constant mean.
  - This is true even for mass, momentum and energy in diffusive processes (open systems with exchange of matter).
  - See demonstration in Koutsoyiannis (2008; this EGU Assembly).

- Explanation 2 (more mathematical): **Maximum entropy**
  - (= maximum uncertainty)
  - Maximize entropy – but not on a single arbitrary time scale.
Do climate models reproduce HK?

None of three IPCC/AR4 climatic models was able to reproduce the long term climatic (30-year) variability of temperature at Albany (Florida, USA; sub-tropical climate).

None of three IPCC/AR4 and three of IPCC/TAR was able to reproduce the long term (30-year) change of precipitation at six out of eight examined locations worldwide.

For more information: Visit the poster by Koutsoyiannis et al. (afternoon session).
Epilogue

- “Φύσις κρύπτεσθαι φιλεῖ”
  (Nature loves to be hidden; Heraclitus)

- “Ἁρμονίη ἁφανῆς φανερῆς κρέσσων”
  (Hidden harmony is better than shown; Heraclitus)

- “He who loves practice without theory is like the sailor who boards ship without a rudder and compass and never knows where he may cast”
  (Leonardo da Vinci)
References


For more information: Visit the poster by Koutsoyiannis et al. (afternoon session)