1. What is memory, short memory and long memory?
What do books say about “memory”?

- Papoulis (1991; Probability, Random Variables, and Stochastic Processes ≡ my gospel) does not use the term memory for a stochastic process
  - He only defines memoryless systems, as those whose outputs $y(t)$ are given as a function $g(x(t))$ of the input $x(t)$ at the same time $t$
  - This involves two stochastic processes and cannot be applied to a single stochastic process
- Beran (1994, Statistics for Long-Memory Processes) defines long and short memory intuitively and relates it to autocorrelation $\rho(j)$ for lag $j$:
  - “The intuitive interpretation of $[\rho(j) \approx c_\rho |j|^{-\alpha}]$ is that the process has long memory”
  - “in contrast ... processes with summable [exponentially decaying] correlations ... are also called processes with short memory or short-range correlations or weak dependence”

How intuitive is the term “memory”?

- The panels on the left show different time windows of the same time series
- The panel above shows how autocorrelation increases with window length, starting from about 0 at the smallest window
- Does the high autocorrelation reflect memory or amnesia?
Notes on the construction of the previous example

- The time series was constructed by superposition of:
  - A stochastic process with values $m_j \sim N(2, 0.5)$ each lasting a period $\tau_j$ exponentially distributed with $E[\tau_j] = 50$ (red line);
  - White noise $N(0, 0.2)$
- The process of our example is completely stationary
- Thus, the autocorrelation function is independent of time location and the autocorrelogram is fully meaningful
- However, an estimate thereof from a few terms of a time series is inaccurate
- Nothing in the process reminds us of any type of memory
- The high autocorrelation is a result of the changing local average

Slow decaying autocorrelation does not reflect long-term memory but long-term change

- The first who pointed this out was Vit Klemes (1974)
  - He wrote “It is shown that the Hurst phenomenon is not necessarily an indicator of infinite memory of a process”
  - He compared the memory-based interpretation to the Ptolemaic planetary model, which worked well but hampered progress in astronomy for centuries
  - Unfortunately, he used the term “nonstationarity” for change, while in fact his final explanatory model was stationary (and he noted it)
- The Hurst phenomenon, named after Hurst (1951) who studied several geophysical time series, essentially reflects long-term changes, even though Hurst formulated it in a slightly different manner:
  - “Although in random events groups of high or low values do occur, their tendency to occur in natural events is greater”
- Earlier than Hurst, Kolmogorov (1940) had proposed a mathematical model that can describe stochastically this natural behaviour
- Here the name Hurst-Kolmogorov (HK) dynamics is used instead of the term “long-term memory” and the mathematical model is termed the HK process (from Koutsoyiannis and Cohn, 2008)
2. What is climate?

A typical definition and some remarks

- “Climate [is] the long-term average of conditions in the atmosphere, ocean, and ice sheets and sea ice described by statistics, such as means and extremes” (U.S. Global Change Research Program, 2009)
- Remark 1: The definition may have missed the importance of the occurrence, circulation, and phase transition of water in the land and the atmosphere, as well the fact that water is the regulator of the entire climate system; hence, it may misrepresent the links of Climate with Hydrology
- Remark 2: To study “statistics, such as means and extremes” one may need some knowledge of Statistics
- Remark 3: To study “long-term average conditions” one may need to shape multi-scale representations of processes and, hence, one may need some more advanced knowledge of Statistics
- Remark 4: Inappropriate representations of hydroclimatic processes, based on Classical Statistics (assuming independence in time), abound; these have severe consequences in the perception of climate, as well as in hypothesis testing, estimation and prediction
How do we expect climate to look like?

- Both time series are synthetic and have the same marginal properties:
  - Size $n = 2^{16} = 65536$
  - Mean $\mu = 5$
  - Standard deviation $\sigma = 1$
  - Normal distribution
- In both time series “climatic” averages for scales $2^5$ and $2^{10}$ are also given in addition to the “annual” plot

**HKC is a non-static, all-scales-changing climate**

- Upper time series:
  - Pure randomness (no dependence in time)
  - Change occurs only at scale 1
  - Climatic (i.e. long-term) averages are flat
- Lower time series:
  - HK dependence with Hurst coefficient $H = 0.99$ (see below)
  - Change occurs at all scales
- Real world evidence (see below) supports HKC rather than SC

If climate is ever changing, why coin a term “climate change”?
3. What is the HK process, how it describes the multi-scale properties of a process, and which are its consequences in Statistics?

The Hurst-Kolmogorov (HK) process and its multi-scale stochastic properties

- The HK process does not provide a “perfect” and “detailed” mathematical tool for geophysical processes.
- Rather it is the most parsimonious and simplest alternative to the classical, independence-based, statistical model:
  - It can represent a non-static climate using a single parameter, the Hurst coefficient $H$, additional to those of classical statistics.
  - It produces very simple expressions for all scales.
- Also, it is ideal for the perception and intuition development of a non-static climate.

<table>
<thead>
<tr>
<th>Properties of the HK process</th>
<th>At an arbitrary observation scale $k = 1$ (e.g. annual) Process: $X_k$</th>
<th>At any scale $k$ Process: $X_k^{(k)} := \frac{1}{k} \sum_{j=-k}^{k} X_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>$\sigma \equiv \sigma^{(1)}$</td>
<td>$\sigma^{(k)} = k^{H-1} \sigma$ ($H$ is the Hurst coefficient; $0.5 &lt; H &lt; 1$)</td>
</tr>
<tr>
<td>Autocorrelation function (for lag $j$)</td>
<td>$\rho_j \equiv \rho_j^{(1)} = \rho_j^{(k)} \approx H(2H-1)</td>
<td>j</td>
</tr>
<tr>
<td>Power spectrum (for frequency $\omega$)</td>
<td>$S(\omega) \equiv S^{(1)}(\omega) \approx \frac{\sigma^2}{4(1-H)} \omega^{1-2H}$</td>
<td>$S^{(k)}(\omega) \approx \frac{\sigma^2}{4(1-H)} \omega^{1-2H} k^{2H-2} (2\omega)^{1-2H}$</td>
</tr>
</tbody>
</table>

Can serve as a definition of the HK process

All properties are power laws of scale $k$, lag $j$, frequency $\omega$. 

D. Koutsoyiannis, Memory in climate and things not to be forgotten 12
Impacts on statistical estimation: Hurst-Kolmogorov Statistics (HKS) vs. Classical Statistics (CS)

<table>
<thead>
<tr>
<th>True values →</th>
<th>Mean, $\mu$</th>
<th>Standard deviation, $\sigma$</th>
<th>Autocorrelation $\rho_l$ for lag $l$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard estimator</td>
<td>$\bar{x} := \frac{1}{n} \sum_{i=1}^{n} x_i$</td>
<td>$\sigma := \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$</td>
<td>$\rho_l := \frac{1}{(n-1)s^2} \sum_{i=1}^{n-l} (x_i - \bar{x})(x_{i+l} - \bar{x})$</td>
</tr>
<tr>
<td>Relative bias of estimation, CS</td>
<td>0</td>
<td>$\approx 0$</td>
<td>$\approx 0$</td>
</tr>
<tr>
<td>Relative bias of estimation, HKS</td>
<td>0</td>
<td>$\approx \sqrt{1 - \frac{1}{n}} / \sqrt{\frac{1}{n} - 1} = - \frac{1}{2n}$</td>
<td>$\approx - \frac{1}{n - 1}$</td>
</tr>
<tr>
<td>Standard deviation of estimator, CS</td>
<td>$\sigma \sqrt{n}$</td>
<td>$\approx \frac{\sigma}{\sqrt{2(n-1)}}$</td>
<td></td>
</tr>
<tr>
<td>Standard deviation of estimator, HKS</td>
<td>$\sigma \sqrt{n'}$</td>
<td>$\approx \frac{\sigma}{\sqrt{2(n-1)}} \sqrt{\frac{0.1 n + 0.8}{\lambda(H) (1-n^{2H-2})}}$</td>
<td>where $\lambda(H) := 0.088 (4H^2-1)^2$</td>
</tr>
</tbody>
</table>

Note: $n' := \sigma^2 / \text{Var}[x] = n^{2-2H}$ is the “equivalent” or “effective” sample size: a sample with size $n'$ in CS results in the same uncertainty of the mean as a sample with size $n$ in HKS.

See Koutsoyiannis (2003) for derivations and for more accurate expressions.

The variance of the estimator of mean

Definition

$n' := \sigma^2 / \text{Var}[x] = n^{2-2H}$

An example for $H = 0.9$ and $n = 100$

- The equivalent (in a CS sense) sample size $n' = 100^{0.72} = 2.51$
- The uncertainty in the estimation of mean is $n/n' = 40$ times greater than in CS
- To make the uncertainty equal to that in CS for sample size 100 (i.e. to make $n' = 100$) we need a real sample size $n = 100^{1/0.72} = 10^{10}$
The bias of the classical estimator of standard deviation

Classical statistics (CS) is here ($H = 0.5$)

An example for $H = 0.9$ and $n = 100$:
- There is a -22% bias
- In addition, the uncertainty in the estimation is 31% greater than in CS (not shown in figure)

The bias tends to -100% as $H$ tends to 1

The bias of the classical estimator of autocorrelation

An example for $H = 0.9$ and $n = 100$: Huge bias at all lags

At lag 1 the bias is -23%
At lag 4 the bias is -55%
At lag 20 and beyond the sample estimate becomes negative!
At lag 50 the bias is -135%
Alternative tools to explore the HK behaviour in time series

- Autocorrelogram?
  - Problematic, because of too high bias
- Periodogram?
  - Very problematic, because it is difficult to obtain analytical expressions of bias and uncertainty, and it may be too rough/scattered (unless the sample size is very high)
- Rescaled range? (the original Hurst’s method)
  - Most inappropriate (too much dispersion, high biases, too uncertain estimates, lack of analytical expressions of statistics; see Koutsoyiannis 2002, 2003)
- Residuals of variance (Taqqu et al., 1995) also termed detrended fluctuation analysis (DFA; Peng et al., 1994)
  - Problematic, as it treats biased and uncertain quantities as if they were unbiased and certain
- Climacogram? (term coined in Koutsoyiannis, 2010)
  - Better, provided that we are aware of (and adapt for) biases and uncertainties, for which there exist analytical expressions

A typical example of misuse in studies detecting HK behaviour

- Several studies based on exploration of the periodogram, the rescaled range, or the DFA, detect a Hurst coefficient $H > 1$
- This is mathematically absurd, because in a (stationary) HK process (also known as fractional Gaussian noise), $0 \leq H \leq 1$
- Even in a cumulative (and thus nonstationary) HK process (also known as fractional Brownian noise), again $0 \leq H \leq 1$ (the cumulative nonstationary process is characterized by the same $H$ as its corresponding stationary process)
- The reported $H > 1$ values are obtained because of inappropriate estimators (based e.g. on slopes of graphs of the explored quantities); their typical interpretation should be that $H$ must be close to 1

See additional discussion in Tyralis and Koutsoyiannis (2010), which includes:
- Two versions of climacogram-based estimation of the process parameters
- Streamlining of a precise Maximum Likelihood method for parameter estimation
- Comparisons of the three proposed methods as well as intercomparisons with other popular methods
Change and scale: The climacogram

- This is simply a (logarithmic) plot of the standard deviation $\sigma^{(k)}$ at scale $k$ vs. the scale $k$
- $\sigma^{(k)}$ can be calculated directly from the time averaged process $x_{(k)}$
- $\sigma^{(k)}$ is a simple transformation of the autocorrelogram $\rho_j$ (where $j$ is lag), i.e.:
  \[ \sigma^{(k)} = \frac{\sigma}{\sqrt{k}} \alpha_k \]
  \[ \alpha_k = 1 + 2 \sum_{j=1}^{k-1} \left( 1 - \frac{j}{k} \right) \rho_j \]

In classical statistics, $\sigma^{(k)} = \sigma / \sqrt{k}$, which indicates a straight line with slope -0.5 in the logarithmic plot

In general, the asymptotic slope (for high $k$) in the logarithmic plot determines the Hurst coefficient:

\[ H = 1 + \text{slope} \]

Slopes milder than -0.5, that is, $H$ values in the interval $(0.5, 1)$, indicate long-range dependence

An important note on the use of the climacogram

- In the HK process, the standard deviation $\sigma^{(k)}$ at scale $k$ is related to that at scale 1, i.e. $\sigma$, by
  \[ \sigma^{(k)} = \frac{\sigma}{k^{1-H}} \]

However, due to bias, the sample standard deviations are not directly comparable to $\sigma^{(k)}$ but to the quantities
  \[ \sigma_n^{(k)} = \frac{\sigma}{k^{1-H}} \sqrt{\frac{1-(k/n)^{2-2H}}{1-(k/n)}} \]

Explanation: if $(s^{(k)})^2$ is the standard estimator of the variance at scale $k$, then it can be shown that $E[(s^{(k)})^2] = (\sigma_n^{(k)})^2$

The example above was constructed assuming a sample with size $n = 100$ from an HK process with $H = 0.9$

Notice that the slope of the curve “adapted for bias” at scale 10 is -0.21 instead of -0.1 (that of the theoretical climacogram)
4. Does empirical evidence support the hypothesis of HKC?

Example 1: Annual minimum water levels of the Nile

- 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
Example 2: The lower tropospheric temperature

Suggests an HK behaviour with a very high Hurst coefficient: $H \approx 0.99$

Example 3: The Monthly Atlantic Multidecadal Oscillation (AMO) Index

Suggests an HK behaviour with a very high Hurst coefficient: $H \approx 0.99$
Example 4: The annual rainfall in Maatsuyker Island (Australia)

Suggests an HK behaviour with a very high Hurst coefficient: $H \approx 0.99$

Maatsuyker Island Lighthouse (Australia), coordinates: -43.65N, 146.27E, 147 m, WMO station code: 94962

Example 5: The mean annual temperature at Svalbard (Norway)

Suggests an HK behaviour with a high Hurst coefficient: $H \approx 0.95$

Svalbard Luft (Norway), coordinates: 78.25N, 15.47E, 29 m, WMO station code: 1008 SVALBARD_LUFT
Example 6: Iowa fine resolution rainfall

Seven storm events of high temporal resolution, recorded by the Hydrometeorology Laboratory at the Iowa University (Georgakakos et al., 1994)

Suggests an HK behaviour with a very high Hurst coefficient: $H \approx 0.99$

Example 7: The Greenland temperature proxy during the Holocene

Suggests an HK behaviour with a high Hurst coefficient: $H \approx 0.94$
5. Does long-range dependence in the HKC imply lower predictive uncertainty?

The graph was constructed using analytical equations by Koutsoyiannis (2005) and Koutsoyiannis et al. (2007).

- The following assumptions were made:
  - $H = 0.90$ (known a priori)
  - Sample size $n = 100$
  - Sample standard deviation $s = 1$
  - Normal distribution

- For scale $k = 1$ (say annual), the unconditional uncertainty of a HKC prediction is higher than in RC, but that conditional on the known past is lower.

- For scale $k = 2$, even the conditional uncertainty of a HKC prediction is higher than in RC.

- For a climatic scale $k = 30$, the unconditional and conditional standard deviations in HKC are respectively 5 and 3 times higher than in RC.
Generalized comparison of HKC with RC predictions

- The graph was constructed with same assumptions as the previous one but with varying sample size $n$ and Hurst coefficient $H$, which is assumed a priori known.
- In reality, the results are worse because $H$ is estimated from data.
- Practical conclusion: except for scale 1 (e.g. annual) the HK behaviour increases predictive uncertainty.

The curves show the scale $k$ in which the HKC uncertainty (conditional on known past) becomes greater than in RC.

Does switching from stochastic to deterministic description of climate enable long term predictability?

- The Figure on the right shows “Long-term Earth system projections for different CO$_2$ emission and storage scenarios” (from Shaffer, 2010)
  - They refer to several climatic variables including “mean atmosphere warming” and “mean ocean warming”.
  - They extend up to 100 000 A.D. (!)
  - The information provided is supposed to support “geoengineering” options, e.g. the “effectiveness and consequences of carbon dioxide sequestration”
- Question 1: Do deterministic models reproduce the past? (see my answer in Koutsoyiannis et al., 2008, Anagnostopoulos et al., 2010)
- Question 2: Does knowledge of deterministic dynamics enable reliable future predictions? (see my answer in Koutsoyiannis, 2010)
Concluding remarks

- Change is Nature’s style and occurs at all time scales
- Climate is no exception: it is ever changing
- By definition, the very notion of climate relies on Statistics—in particular, on long-term statistical properties of natural processes
- Classical Statistics is inconsistent with real-world climate; rather, it corresponds to a static climate
- Hurst-Kolmogorov Statistics is the key to perceive a multi-scale changing climate and model the implied uncertainty and risk
- Hurst-Kolmogorov Statistics has marked differences from Classical Statistics, which are often missed; thus, it implies:
  - Dramatically higher uncertainty of statistical parameters of location
  - High negative bias of statistical parameters of dispersion and dependence
  - Dramatically higher predictive uncertainty

More than a fire, one needs to extinguish hubris (Heraclitus, fragment 43)
References

- Alley, R. B., The Younger Dryas cold interval as viewed from central Greenland, Quaternary Science Reviews, 19, 213-226, 2000
- Alley, R. B., GISP2 Ice Core Temperature and Accumulation Data. IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series #2004-013, NOAA/NGDC Paleoclimatology Program, Boulder CO, USA, 2004
- Koutsoyiannis, D., The Hurst phenomenon and fractional Gaussian noise made easy, Hydrological Sciences Journal, 47 (4), 573–595, 2002
- Koutsoyiannis, D., Climate change, the Hurst phenomenon, and hydrological statistics, Hydrological Sciences Journal, 48 (1), 3–24, 2003
- Koutsoyiannis, D., Uncertainty, entropy, scaling and hydrological stochastics, 2, Time dependence of hydrological processes and time scaling. Hydrological Sciences Journal, 50 (3), 405–426, 2005
- Koutsoyiannis, D., A random walk on water, Hydrology and Earth System Sciences, 14, 585–601, 2010