European Geosciences Union General Assembly 2010
Vienna, Austria, 2-7 May 2010

Session NP4.1: Open Session on Geoscientific Time Series Analysis

Some problems in inference from time series of geophysical processes

Demetris Koutsoyiannis
Department of Water Resources and Environmental Engineering
Faculty of Civil Engineering
National Technical University of Athens, Greece
(dk@itia.ntua.gr, http://www.itia.ntua.gr/dk/)

Presentation available online: http://www.itia.ntua.gr/en/docinfo/973/

Some inequalities

- time series ≠ process
- physical process ≠ mathematical process
  - geophysical process ≠ stochastic process
- time series ≠ geophysical process
- time series ≠ stochastic process

Geophysical laws are of stochastic type

- Even deterministic Newton's laws become stochastic when applied to geophysical systems
  - Example: Navier-Stokes equations in turbulent flows
- Stochastic laws are inferred, verified or falsified from empirical data (time series) usually by induction (rather than deduction)
- Even deterministic controls in geophysical processes are detected and explored using stochastic approaches
- Unconsciousness of the stochastic character of a concept, law, approach, or tool may result in terrible mistakes
  - Example: In detecting deterministic chaos in geophysical processes it was often missed that entropy is a stochastic concept and its estimation is subject to statistical uncertainty (see Koutsoyiannis, 2006)

Note on terminology Stochastics/stochastic are used to collectively incorporate probability theory, statistics and stochastic processes

Case 1: Statistical implications of scaling in state

- Experiment: A Google search with terms *multifractal rainfall moments* was performed
- The first (highest PageRank) paper was chosen and its first figure is reproduced here

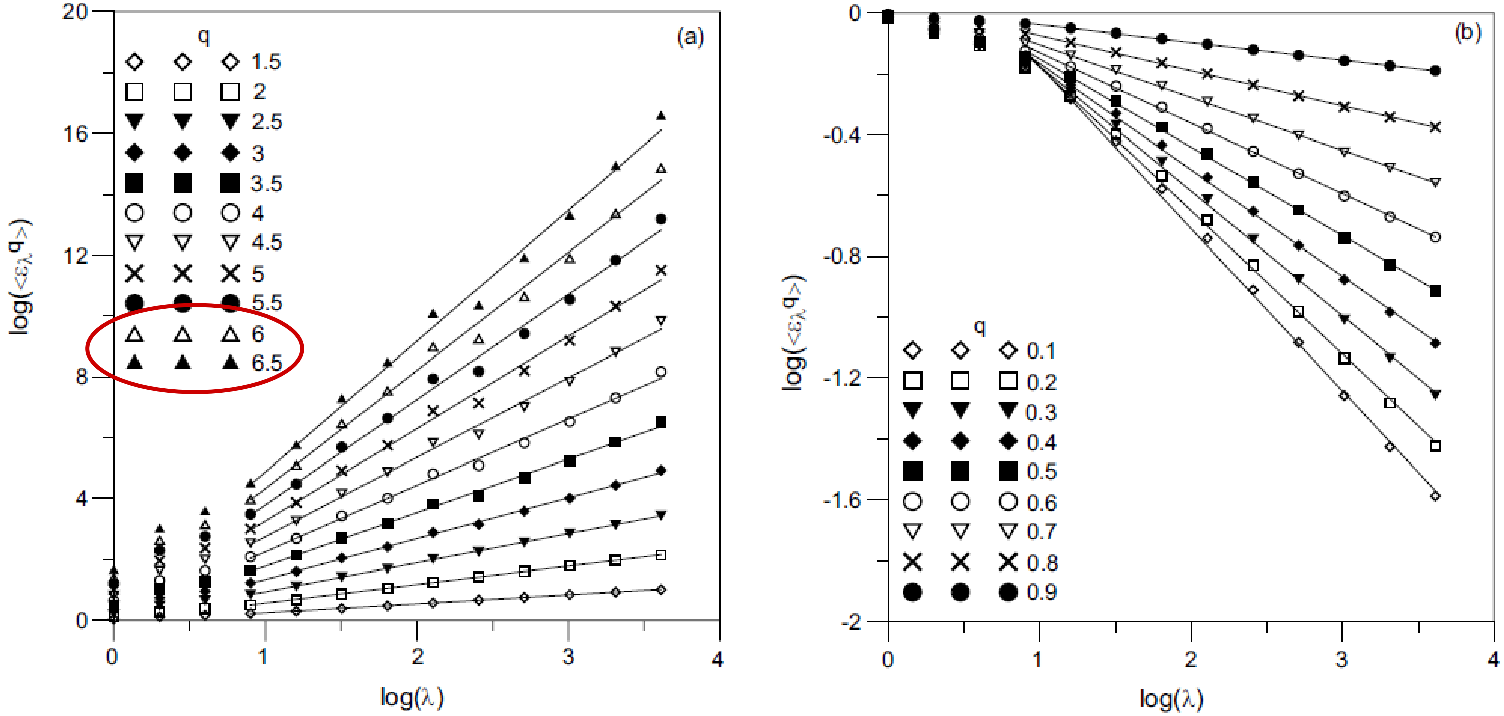


Fig. 1. Log-log plot of the qth moments of the rainfall intensity on the time scales from 1 hour to almost 6 months versus the scale ratio λ. (a) For moments larger than 1; (b) for moments smaller than 1.

Exploration of the information content in high moments of rainfall depths

- High moments, i.e. $m_q := \mathrm{E}[x^q]$ for $q = 4, 5, 6, 7, ...$, depend enormously and exclusively on the distribution tail
- Recent research results (e.g. Koutsoyiannis 2004, 2005; Papalexiou and Koutsoyiannis, 2010; and references therein) suggest power-type/Pareto tail with shape parameter κ = 0.13-0.15, almost constant worldwide
- This reflects the (imperfect) **scaling in state** of rainfall rate
- Beyond $q_{\max} = 1/\kappa = 6.67$ (for $\kappa = 0.15$) the moments are infinite
- However, their numerical estimates from a time series are always finite: an **infinite negative bias**
- Even below $q_{\max}$, the estimation of moments is problematic; this can be demonstrated by Monte Carlo simulation

Setting up the Monte Carlo simulation

- Random variable $\underline{x}$ (representing rainfall distribution tail, i.e. rainfall excess above a certain threshold)
- Pareto distribution function with parameters κ (shape) and λ (scale)

$$P\{\underline{x} > x\} =: F^*(x) = (1 + \kappa\, x/\lambda)^{-1/\kappa}$$

- Analytically calculated moments ($B(\ )$ denotes the beta function)

$$m_q = \mathrm{E}[\underline{x}^q] = q\,(\lambda/\kappa)^q\, B(1/\kappa - q, q) \text{ for } q < 1/\kappa$$

$$m_q = \mathrm{E}[\underline{x}^q] = \infty \text{ for } q \geq 1/\kappa$$

- Random sample $\underline{x}_1, \underline{x}_2, \ldots \underline{x}_n$, with size $n = 100$
- Moment estimator (a random variable)

$$\underline{\tilde{m}}_q = (1/n) \sum_{i=1}^{n} \underline{x}_i^q$$

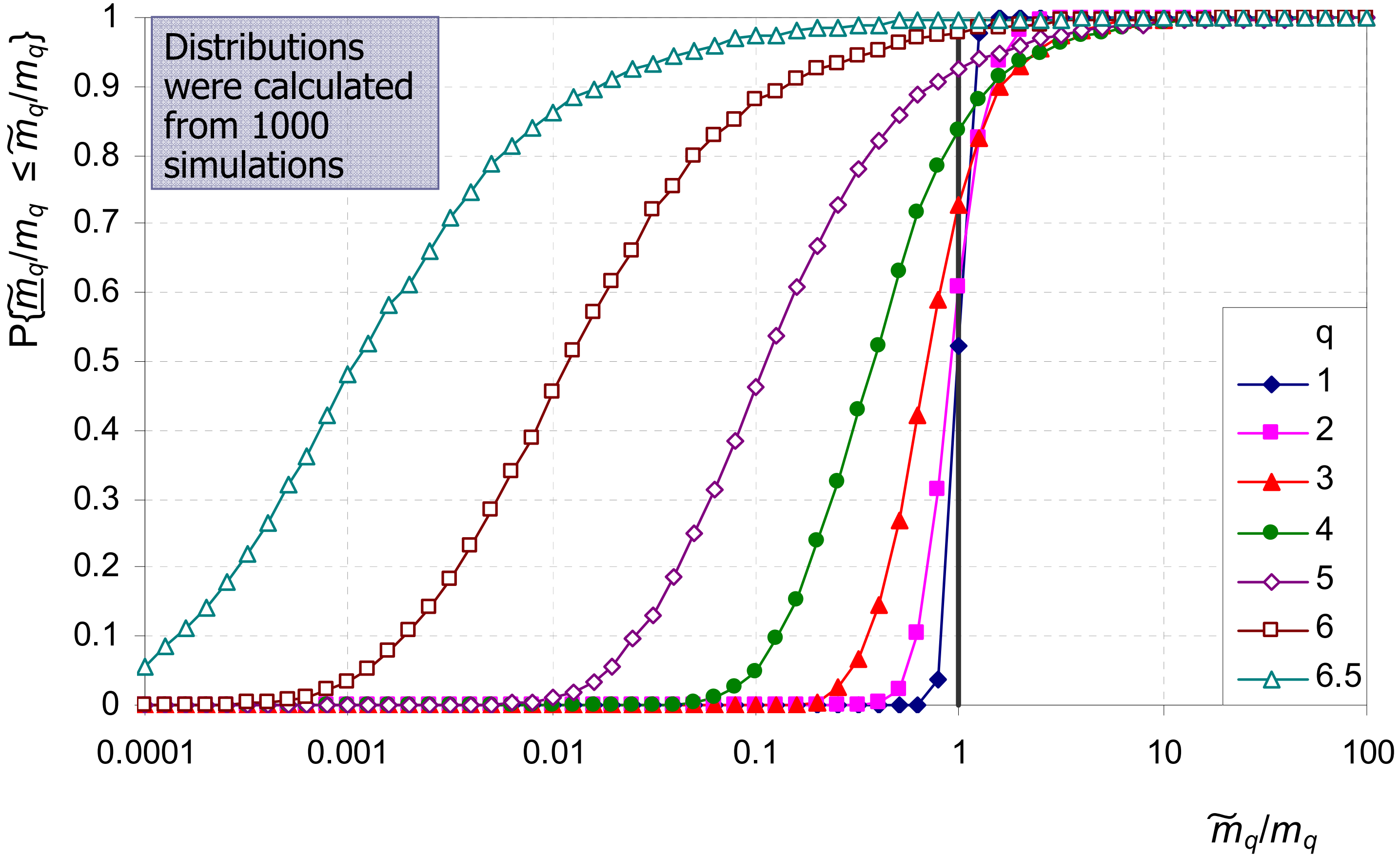
- Moment estimate (a numerical value)

$$\tilde{m}_q = (1/n) \sum_{i=1}^{n} x_i^q$$

More inequalities (notice, underlined quantities denote random variables)
$m_q \neq \underline{\tilde{m}}_q \neq \tilde{m}_q \neq m_q$ (three conceptually different mathematical objects)

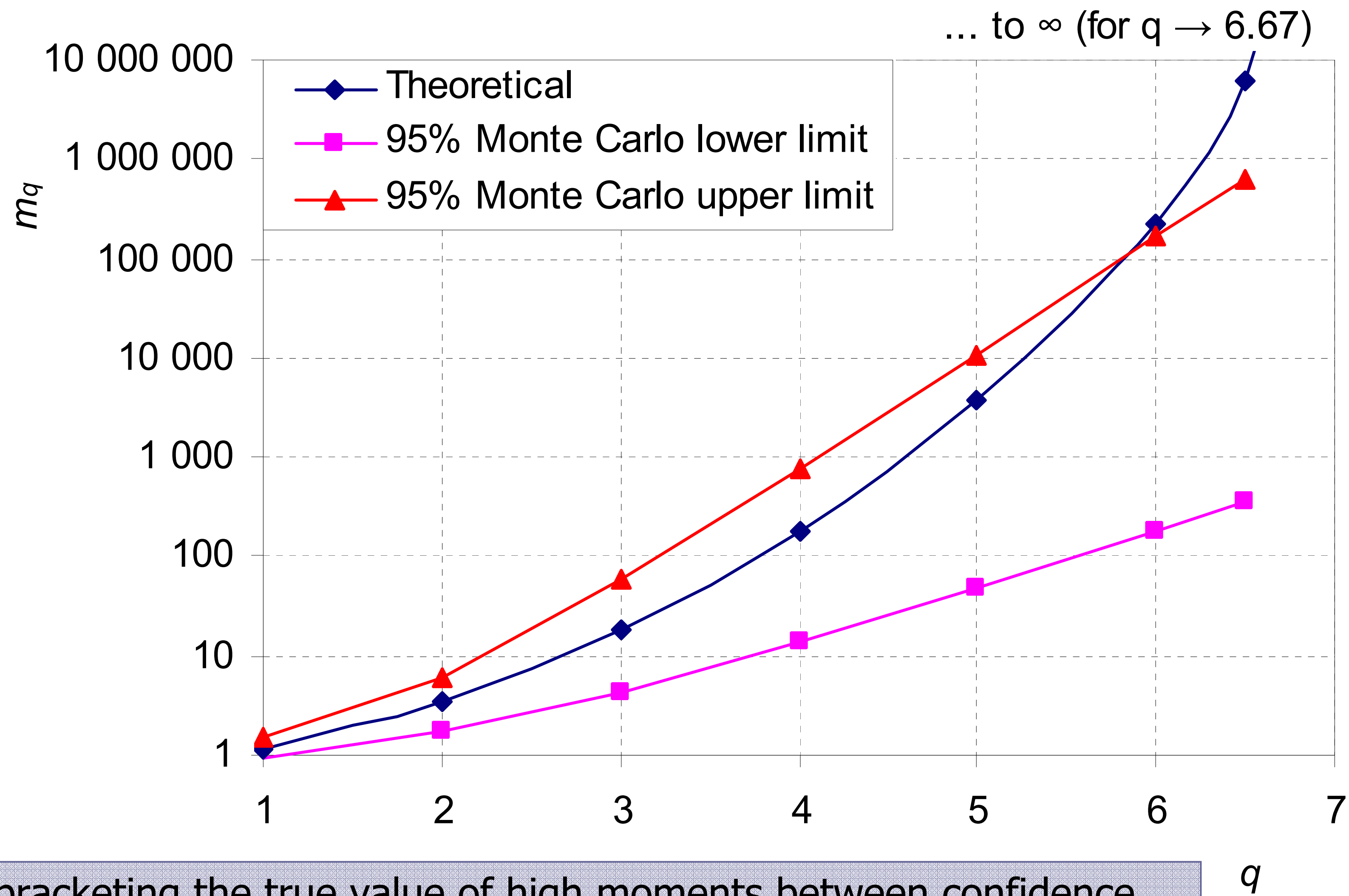
Results of Monte Carlo simulation

- The information content of the empirically estimated moments is high if the distribution of the random variable ($\tilde{m}_q / m_q$) is concentrated around 1
- Only low moments (q = 1 and 2) have reasonably low variation
- All others vary within orders of magintude
- Even the medians are by one or more orders of magnitude lower than 1 for $q > 4$



Variables whose distribution ranges over several orders of magnitude cannot support inference about a natural behaviour; they can only show the uncertainty

Results of Monte Carlo simulation (2)



Even bracketing the true value of high moments between confidence limits may be impossible for a distribution with Pareto tail

Case 2: Statistical implications of scaling in time

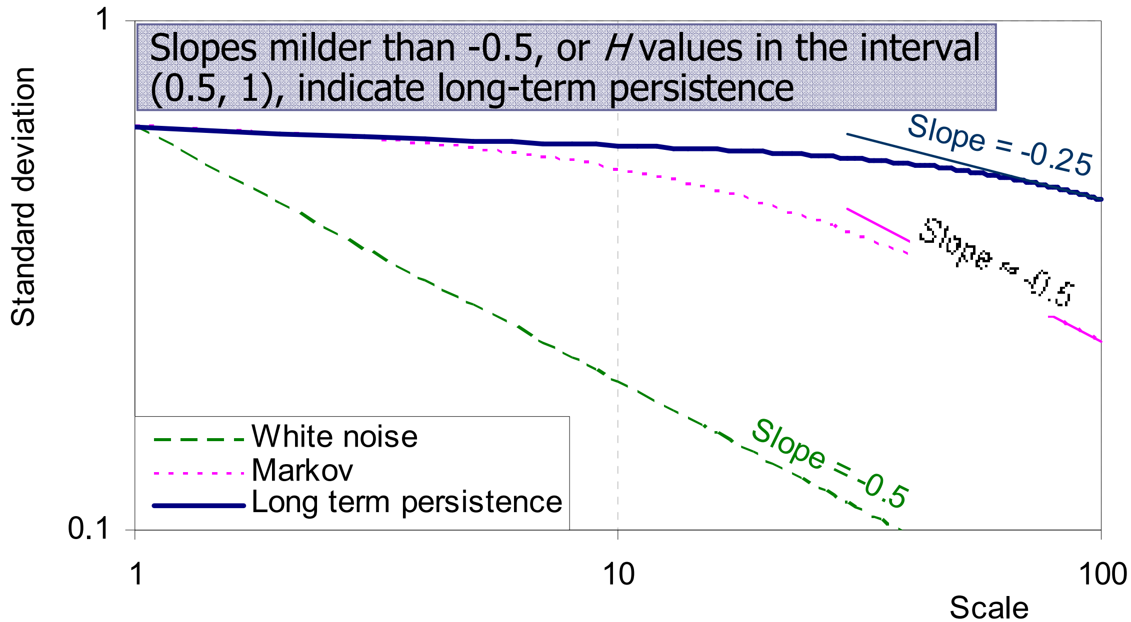
- Scaling in time is best viewed through the time averaged process

$$\underline{x}_i^{(k)} := \frac{1}{k} \sum_{l=(i-1)k}^{ik} \underline{x}_l$$

- Its standard deviation $\sigma^{(k)}$ at scale k is related to the autocorrelogram ρ_j (where j is lag), by a simple transformation, i.e.,

$$\sigma^{(k)} = \frac{\sigma}{\sqrt{k}}\sqrt{\alpha_k}\,, \qquad \alpha_k = 1 + 2\sum_{j=1}^{k-1}\left(1-\frac{j}{k}\right)\rho_j \leftrightarrow \rho_j = \frac{j+1}{2}\alpha_{j+1} - j\alpha_j + \frac{j-1}{2}\alpha_{j-1}$$

- The plot of $\sigma^{(k)}$ vs. k has been termed the climacogram
- The asymptotic slope (high k) in a logarithmic plot is a characteristic of scaling defining the so-called Hurst coefficient: $H = 1 + \text{slope}$



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

The simplest process with scaling in time (or long-term persistence), the Hurst-Kolmogorov process, has constant slope of climacogram throughout all scales (power-law climacogram or **perfect time scaling**)

Also its autocorrelogram and power spectrum are power laws of lag j, frequency ω and scale k

Properties of the HK process	At an arbitrary observation scale $k = 1$ (e.g. annual)	At any scale k
Standard deviation	$\sigma \equiv \sigma^{(1)}$	$\sigma^{(k)} = k^{H-1}\sigma$ (can serve as a definition of the HK process; H is the Hurst coefficient; $0.5 < H < 1$)
Autocorrelation function (for lag j)	$\rho_j \equiv \rho_j^{(1)} = \rho_j^{(k)} \approx H(2H-1)\,\lvert j\rvert^{2H-2}$	
Power spectrum (for frequency ω)	$s(\omega) \equiv s^{(1)}(\omega) \approx 4(1-H)\,\sigma^2\,(2\omega)^{1-2H}$	$s^{(k)}(\omega) \approx 4(1-H)\,\sigma^2\,k^{2H-2}\,(2\omega)^{1-2H}$

A historical note: Hurst & Kolmogorov

We owe the discovery and first study of scaling behaviour in time of natural processes to Hurst and Kolmogorov (see Koutsoyiannis and Cohn, 2008)

AMERICAN SOCIETY OF CIVIL ENGINEERS

Founded November 5, 1852

TRANSACTIONS

Paper No. 2447

LONG-TERM STORAGE CAPACITY OF RESERVOIRS

By H. E. Hurst[1]

With Discussion by Ven Te Chow, Henri Milleret, Louis M. Laushey, and H. E. Hurst.

Synopsis

A solution of the problem of determining the reservoir storage required on a given stream, to guarantee a given draft, is presented in this paper. For example, if a long-time record of annual total discharges from the stream is available, the storage required to yield the average flow, each year, is obtained by

Hurst (1951) studied numerous geophysical time series and 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*"

Kolmogorov (1940) studied the stochastic process that describes this behaviour 10 years earlier than Hurst

Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS
1940. Volume XXVI, № 2

MATHEMATIK

WIENERSCHE SPIRALEN UND EINIGE ANDERE INTERESSANTE KURVEN IM HILBERTSCHEN RAUM

Von A. N. KOLMOGOROFF, Mitglied der Akademie

Wir werden hier einige Sonderfälle von Kurven betrachten, denen meine vorhergehende Note «Kurven im Hilbertschen Raum, die gegenüber einer einparametrigen Gruppe von Bewegungen invariant sind» (¹) gewidmet ist.

Unter einer Ä h n l i c h k e i t s t r a n s f o r m a t i o n im Hilbert... Raum H werden wir eine beliebige ...

... $\neq x'$ der Punkte, die auf derselben Kurve liegen, übergeht.

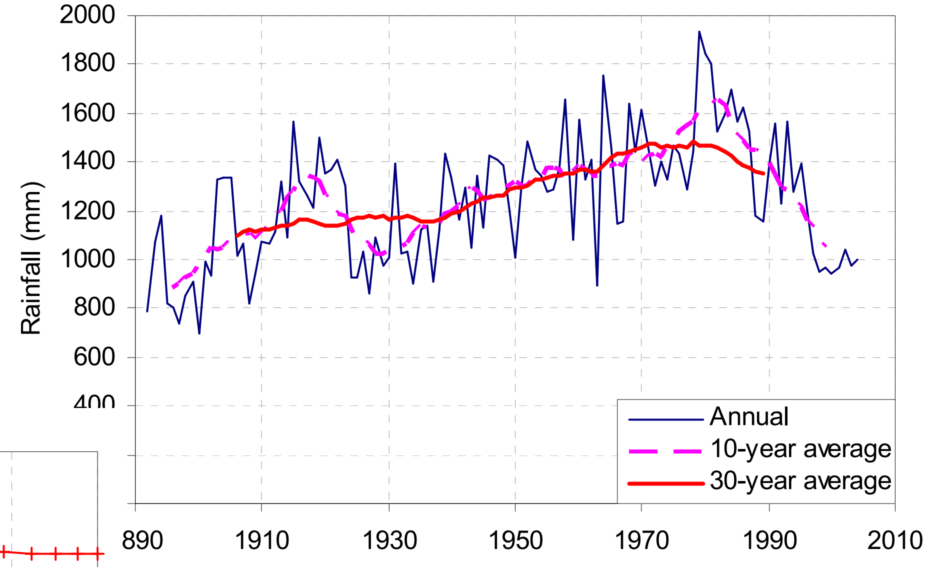
S a t z 6. *Die Funktion* $B_\xi(\tau_1, \tau_2)$, *die der Funktion* $\xi(t)$ *der Klasse* $\mathfrak{A}$ *entspricht, kann in der Form*

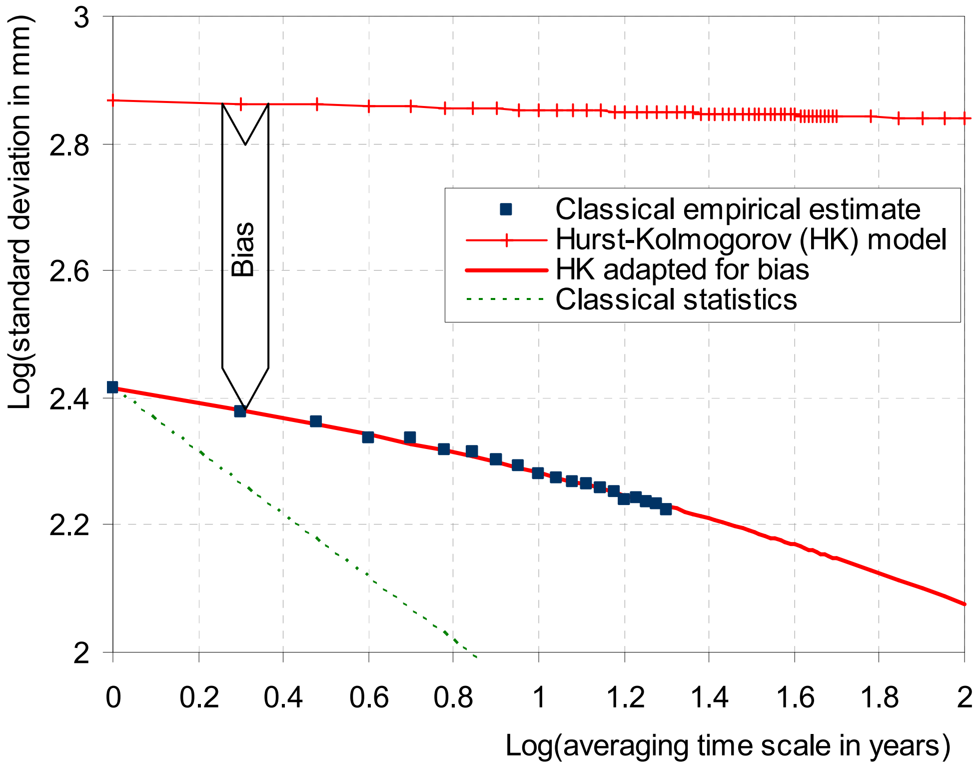
$$B_\xi(\tau_1, \tau_2) = c\,[\,|\tau_1|^\gamma + |\tau_2|^\gamma - |\tau_1 - \tau_2|^\gamma\,]$$

115

Example 1: The annual rainfall in Maatsuyker Island (Australia)

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

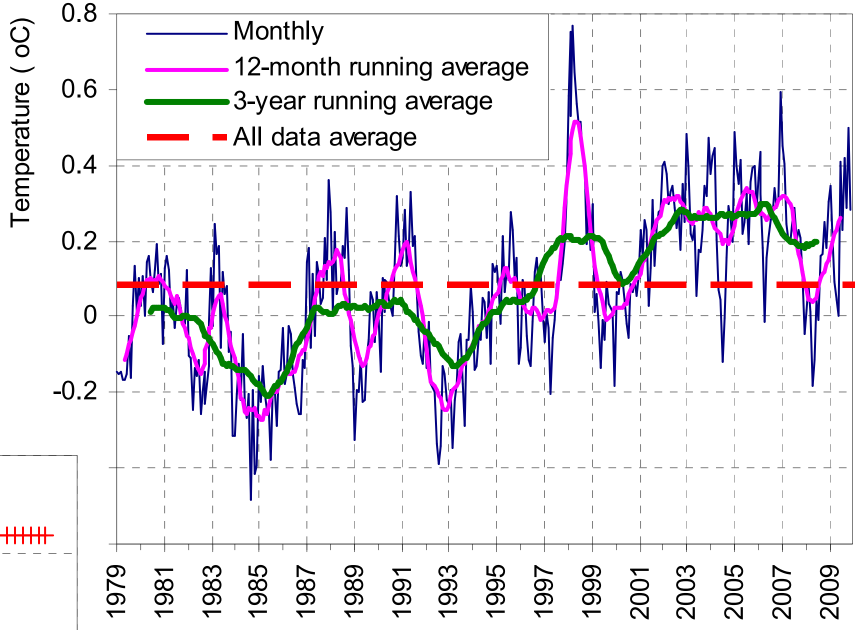


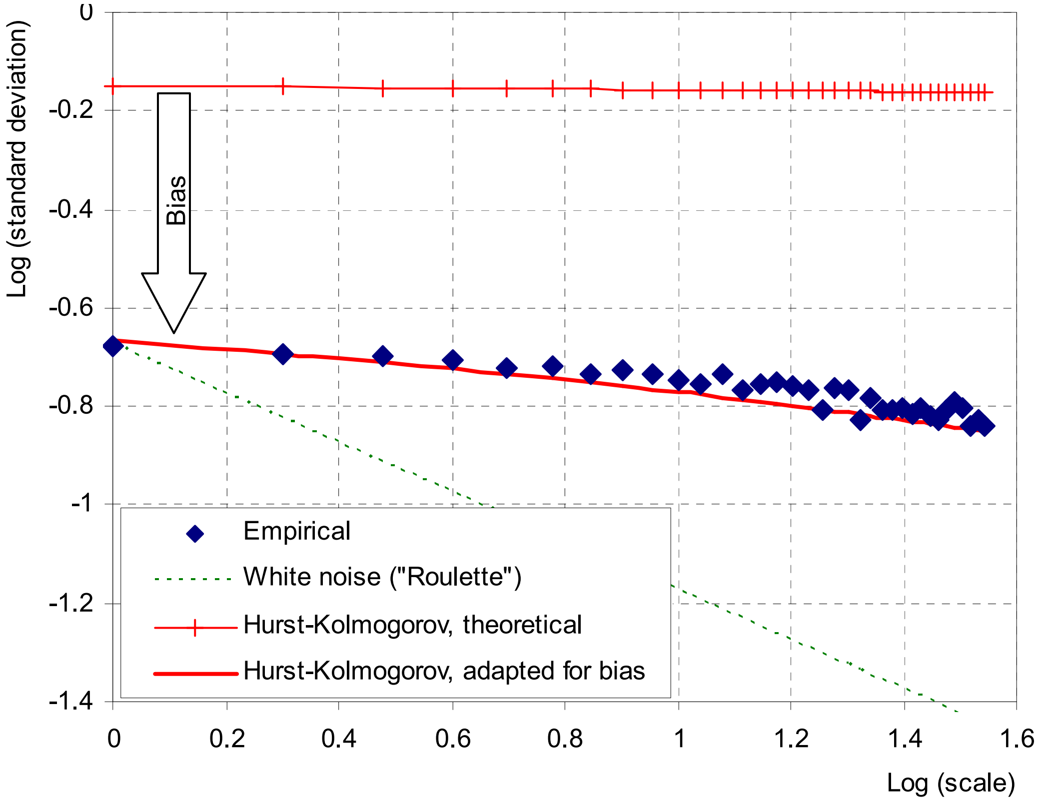


Maatsuyker Island Lighthouse (Australia),
coordinates: -43.65N, 146.27E, 147 m,
WMO station code: 94962
Data: 1892-2004, from
http://climexp.knmi.nl/getprcpall.cgi?someone@somewhere+94962+MAATSUYKER_ISLAND_LIGHTHOUSE+

Example 4: The lower tropospheric temperature

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





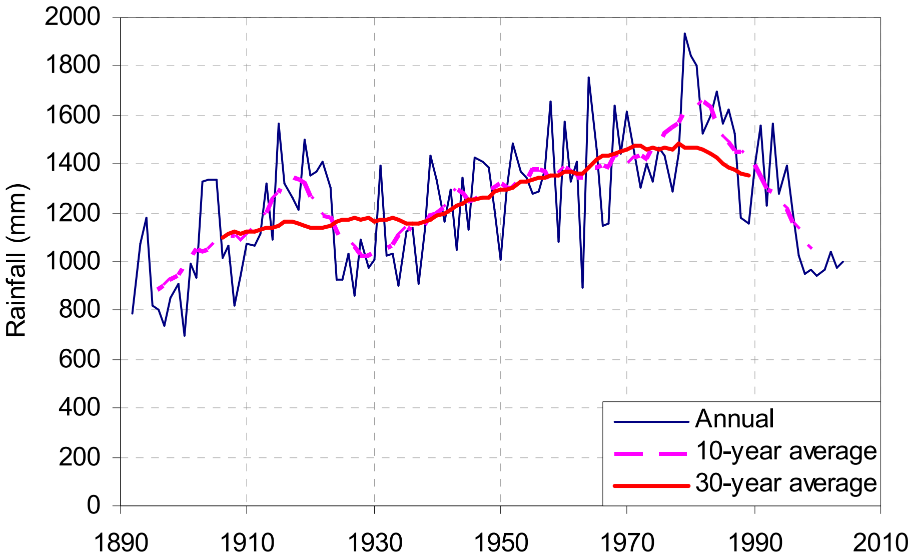
Data: 1979-2010, from http://vortex.nsstc.uah.edu/public/msu/t2lt/tltglhmam_5.2

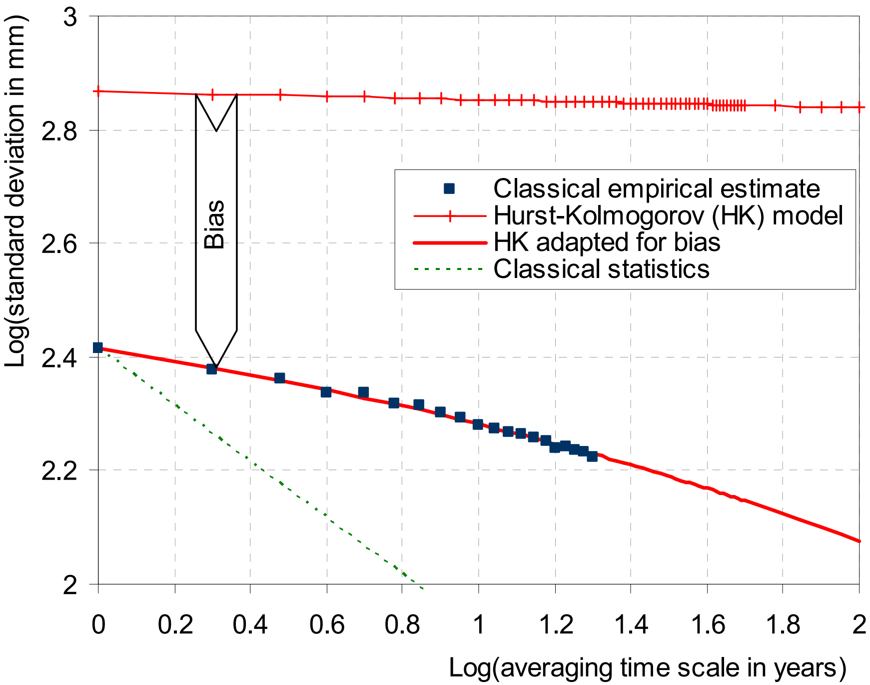
Impacts on statistical estimation: Hurst-Kolmogorov statistics (HKS) vs. classical statistics (CS)

True values ⟶	Mean, μ	Standard deviation, σ	Autocorrelation ρ_l for lag l
Standard estimator	$\bar{\underline{x}} := \frac{1}{n}\sum_{i=1}^{n}\underline{x}_i$	$\underline{s} := \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\underline{x}_i - \bar{\underline{x}})^2}$	$\underline{r}_l := \frac{1}{(n-1)\underline{s}^2}\sum_{i=1}^{n-l}(\underline{x}_i - \bar{\underline{x}})(\underline{x}_{i+l} - \bar{\underline{x}})$
Relative bias of estimation, CS	0	≈ 0	≈ 0
Relative bias of estimation, HKS	0	$\approx \sqrt{1 - \frac{1}{n'}} - 1 \approx -\frac{1}{2n'}$	$\approx -\frac{1/\rho_l - 1}{n' - 1}$
Standard deviation of estimator, CS	$\frac{\sigma}{\sqrt{n}}$	$\approx \frac{\sigma}{\sqrt{2(n-1)}}$	
Standard deviation of estimator, HKS	$\frac{\sigma}{\sqrt{n'}}$	$\approx \frac{\sigma\sqrt{(0.1\,n + 0.8)^{\lambda(H)}(1 - n^{2H-2})}}{\sqrt{2(n-1)}}$ where $\lambda(H) := 0.088\,(4H^2 - 1)^2$	

Note: $n' := n^{2-2H}$ 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 (Koutsoyiannis, 2003; Koutsoyiannis & Montanari, 2007).

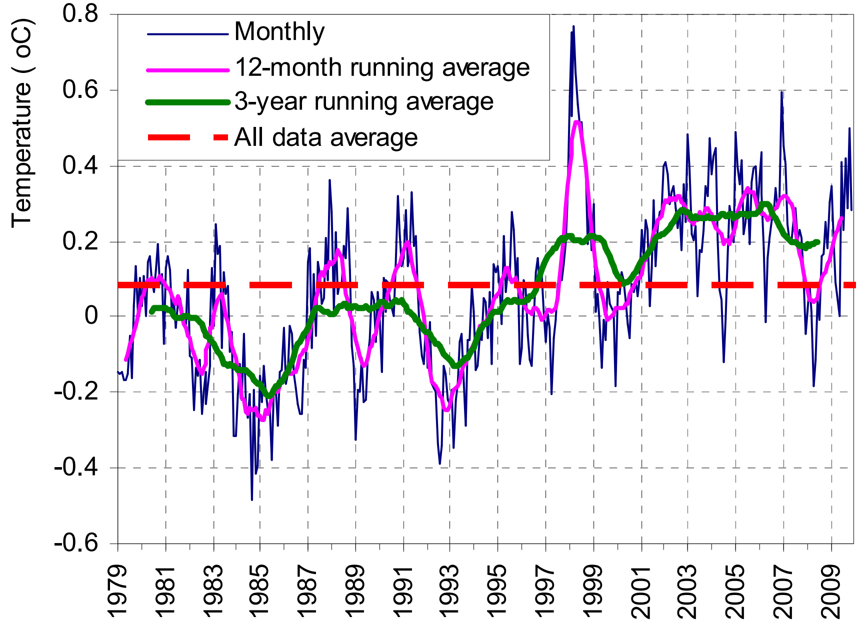
Example 1: The annual rainfall in Maatsuyker Island (Australia)

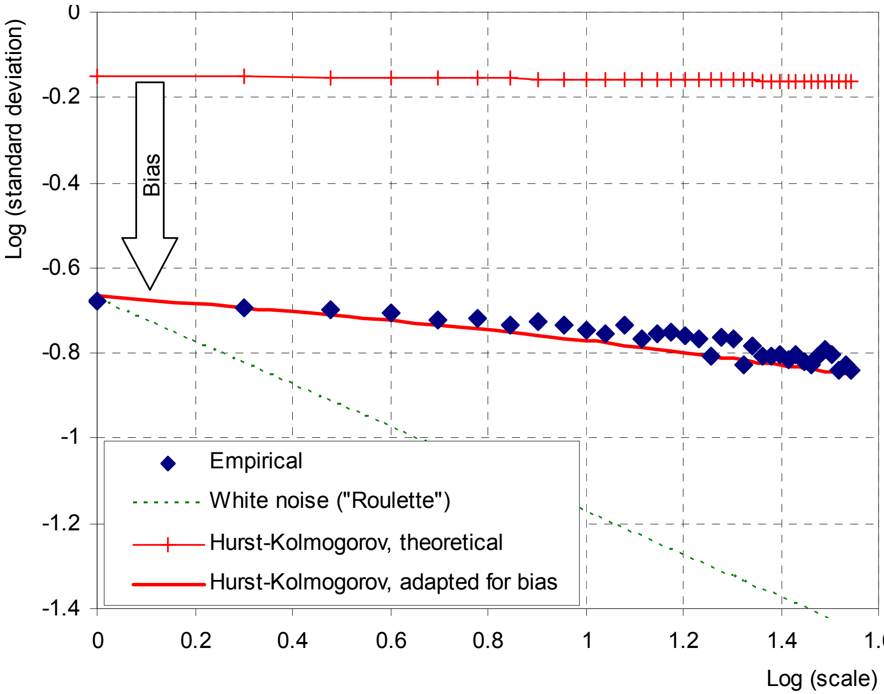




	CS	HKS	ratio
Hurst coefficient		0.986	
Sample size (n, n')	113	1.1	0.010
Mean			
Point estimate	1237.7	1237.7	1
Standard error	69.2	688.7	9.9
Standard deviation			
Point estimate	260.5	735.9	2.8
Standard error	17.4	43.3	2.5

Example 4: The lower tropospheric temperature





	CS	HKS	ratio
Hurst coefficient		0.992	
Sample size (n, n')	373	1.1	0.003
Mean			
Point estimate	0.0712	0.0712	1
Standard error	0.0369	0.6796	18.4
Standard deviation			
Point estimate	0.211	0.713	3.4
Standard error	0.0077	0.0315	4.1

Concluding remarks

- Modelling of geophysical processes heavily relies on available data series and their statistical processing
- The classical statistical approaches, often used in geophysical modelling, are based upon several simplifying assumptions, tacit or explicit, such as independence in time and exponential distribution tails, which are invalidated in natural processes
- Moreover, the perception of the general behaviour of the natural processes and the implied uncertainty is heavily affected by the classical statistical paradigm
- However, the study of natural processes reveals scaling behaviours in state (departure from exponential distribution tails) and in time (departure from independence)
- Both types of scaling result in enormous biases and/or enormously increased uncertainty in all properties of processes
- Ignorance of increased uncertainty results in inappropriate modelling, wrong inferences and false claims about the properties of the processes

knowledge of uncertainty ≠ ignorance

References

- Hurst, H.E., Long term storage capacities of reservoirs, *Trans. Am. Soc. Civil Engrs.*, 116, 776–808, 1951.
- Kolmogorov, A. N., Wienersche Spiralen und einige andere interessante Kurven in Hilbertschen Raum, *Dokl. Akad. Nauk URSS*, 26, 115–118, 1940.
- Koutsoyiannis, D., Statistics of extremes and estimation of extreme rainfall, 2, Empirical investigation of long rainfall records, *Hydrological Sciences Journal*, 49 (4), 591–610, 2004.
- Koutsoyiannis, D., Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling, *Hydrological Sciences Journal*, 50 (3), 381–404, 2005.
- Koutsoyiannis, D., On the quest for chaotic attractors in hydrological processes, *Hydrological Sciences Journal*, 51 (6), 1065–1091, 2006.
- Koutsoyiannis, D., and T.A. Cohn, The Hurst phenomenon and climate (solicited), *European Geosciences Union General Assembly 2008, Geophysical Research Abstracts, Vol. 10*, Vienna, 11804, European Geosciences Union, 2008.
- Papalexiou, S.-M., and D. Koutsoyiannis, On the tail of the daily rainfall probability distribution: Exponential-type, power-type or something else?, *European Geosciences Union General Assembly 2009, Geophysical Research Abstracts, Vol. 12*, Vienna, EGU2010-111769, European Geosciences Union, 2010.