EGS - AGU - EUG Joint Assembly
Nice, France, 6-11 April 2003
Session HS19/ Climate change impacts on the hydrological cycle, extremes, forecasting and implications on engineering design

Hydrological statistics for engineering design in a varying climate

Demetris Koutsoyiannis
Department of Water Resources, School of Civil Engineering,
National Technical University, Athens, Greece

The role of hydrological statistics

- **Objective**: Quantification of uncertainty and risk in hydrologic processes
- **Utility:** Engineering design and management of hydrosystems
- **Mathematical basis:** Concepts of probability, statistics and stochastic processes
- **Empirical basis:** Records of hydrological measurements
- **Typical problems:**
  - Analysis and enhancement of data sets
  - Testing of hypotheses
  - Estimation of distribution quantiles and confidence intervals

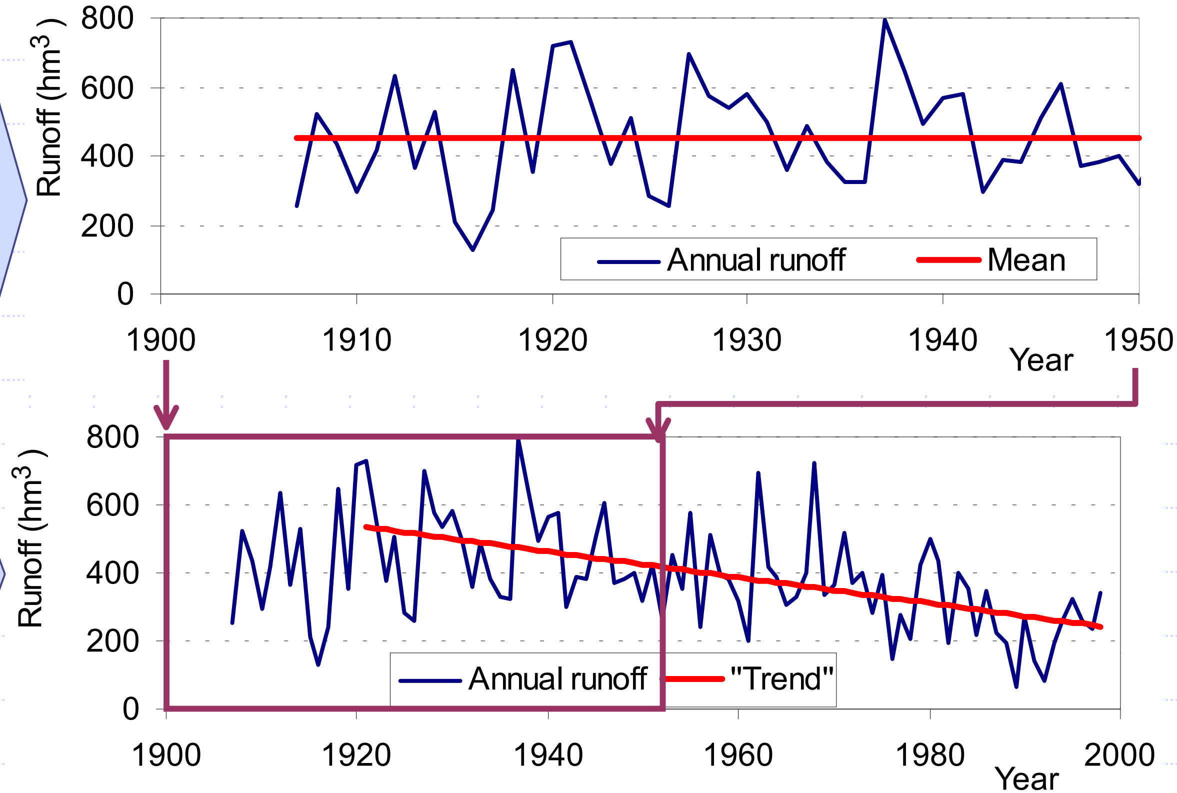
Empirical basis in hydrological statistics

A typical "short" time series: Annual runoff (expressed as equivalent depth) of the Boeoticos Kephisos River basin, Greece

Stable behaviour, annual random fluctuation around a constant mean

The same time series for a longer period

Appearance of overyear "trends"



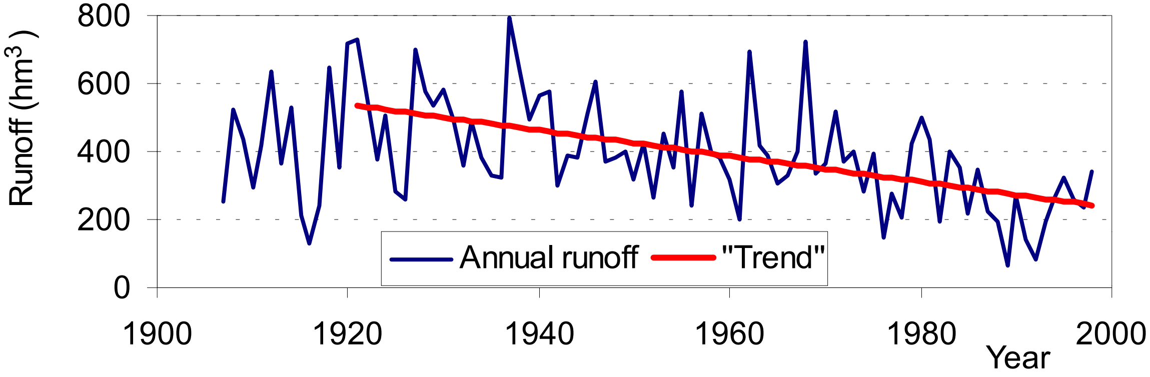
Typical processing of a time series with a "trend"

- Assume that "trend" is deterministic and fit an equation
- Detrend the series (e.g. σ_{init} = 153 hm^3, σ_{detr} = 127 hm^3)
- Consequence: "Trends" decrease uncertainty

Dangerous in engineering design

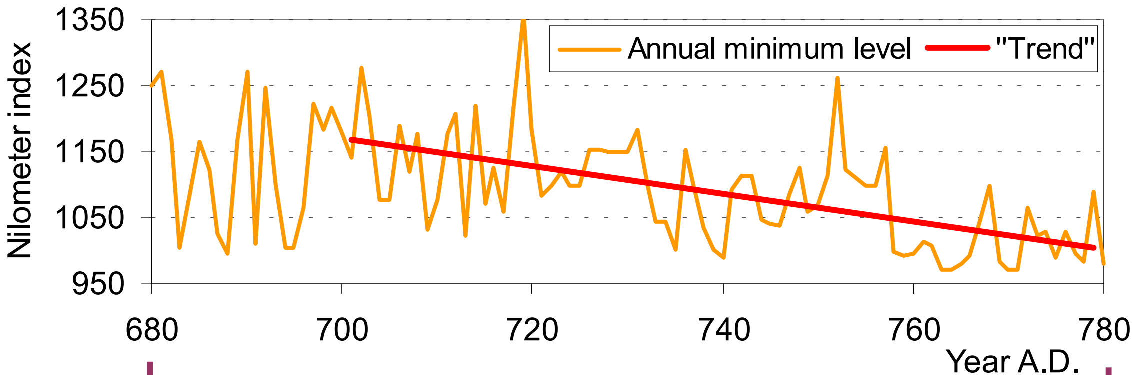
Behaviour of long series

The full Boeoticos Kephisos runoff time series



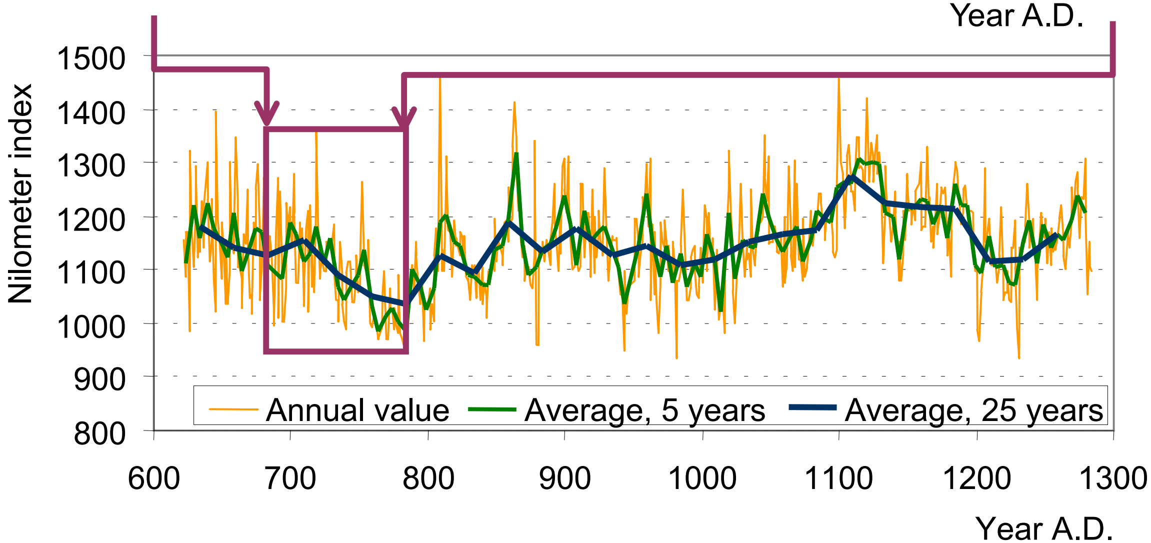
Part of the annual minimum water level of the Nile river (Nilometer)

A similar "trend"



The full Nilometer series for the years 622 to 1284 A.D. (663 years; Beran, 1994)

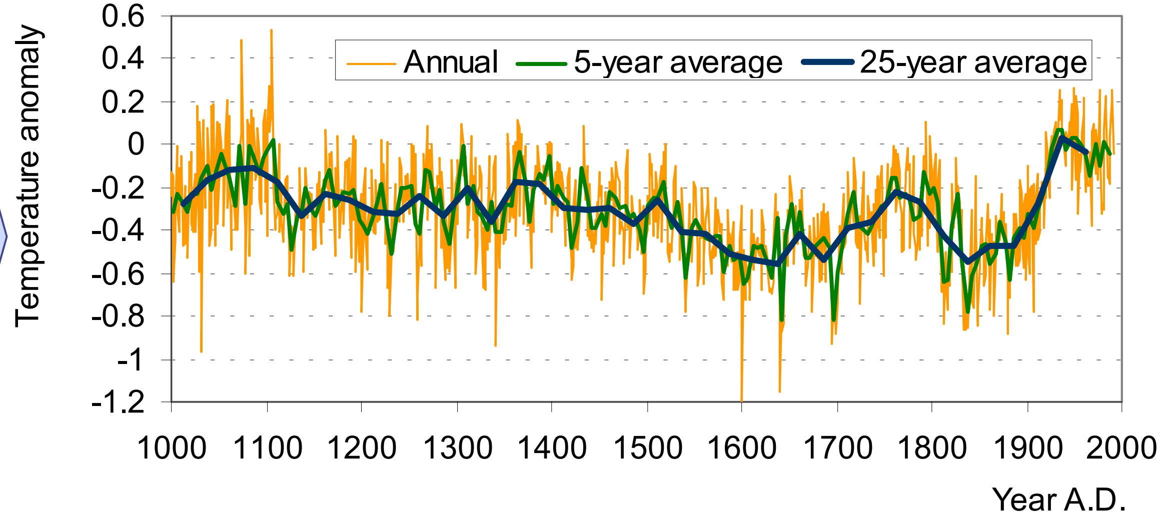
Upward and downward irregular fluctuations at all time scales



More long series

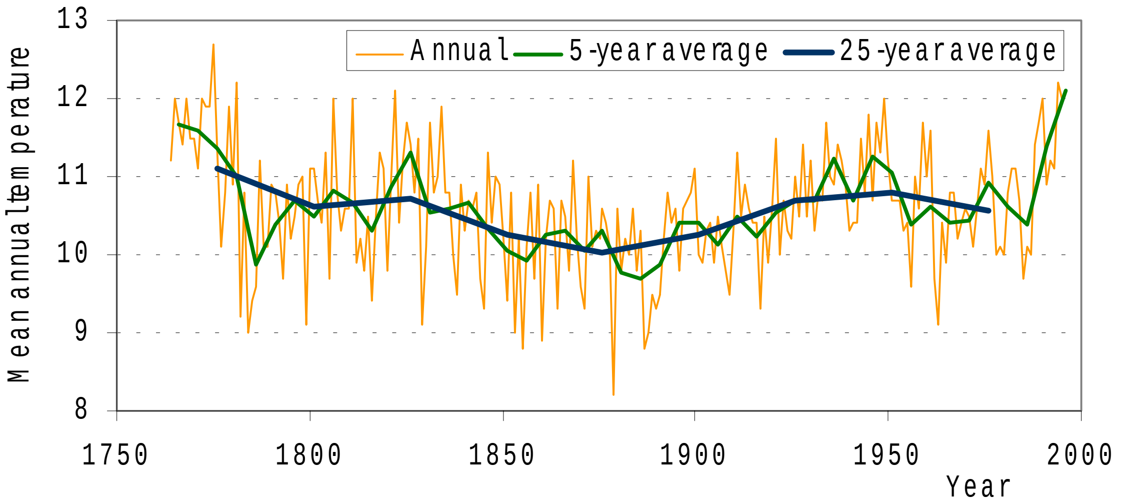
Northern Hemisphere temperature anomalies in °C vs 1961–1990 mean (992 years, reconstructed from multi-proxy data by Jones et al., 1998)

Irregular fluctuations at all time scales



Mean annual temperature at Paris/Le Bourget (instrumental meteorological observations extending through 1764–1995; from ftp.cru.uea.ac.uk).

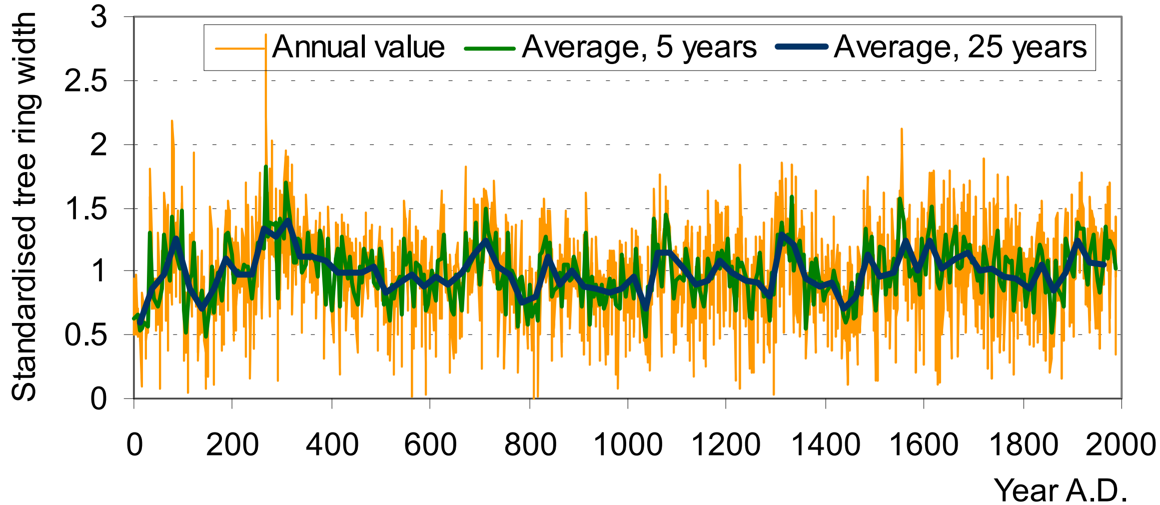
Irregular fluctuations at all time scales



Yet another long series vs. a random series

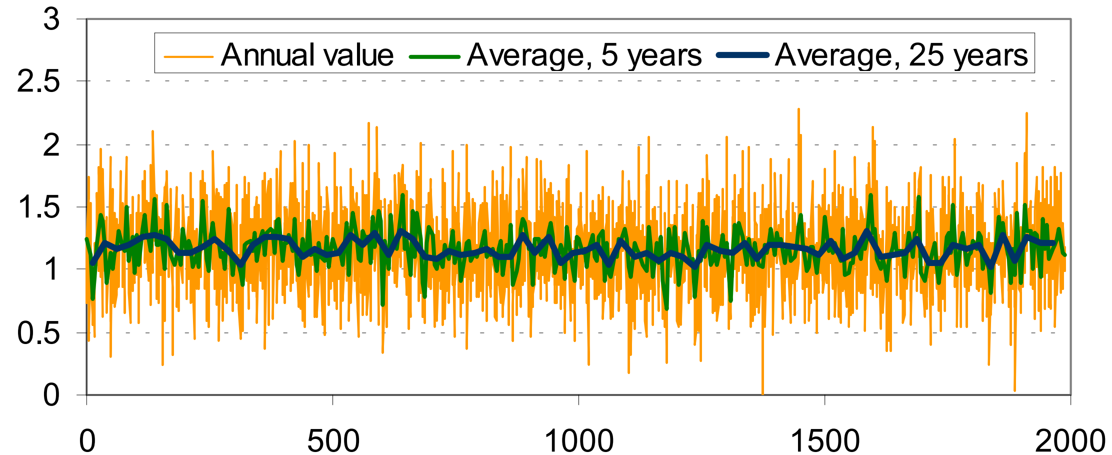
Standardised tree ring widths from a paleoclimatological study at Mammoth Creek, Utah, for the years 0-1989 (1990 years; from ftp:// ftp.ngdc.noaa.gov/ paleo/)

Irregular fluctuations at all time scales



A synthetic series of independent random variates (white noise) with marginal statistics equal to those of the tree ring series (1990 values)

Random fluctuations at the annual scale; tend to smooth out as time scales become larger



Climatic fluctuations and the Hurst phenomenon

- "Climate changes irregularly, for unknown reasons, on all timescales" (National Research Council, 1991, p. 21).
- All examined long time series confirm this motto.
- Irregular changes in time series are better modelled as stochastic fluctuations on many time scales rather than deterministic components.
- Equivalently, these fluctuations can be regarded as a manifestation of the *Hurst phenomenon* quantified through the *Hurst exponent, H* (Hurst, 1951).
- The relationship of *climatic fluctuations on many scales* and the *Hurst phenomenon* has been conjectured by Mesa & Poveda (1993) and studied by Koutsoyiannis (2002).

A basis for fluctuating climate: The SSS process

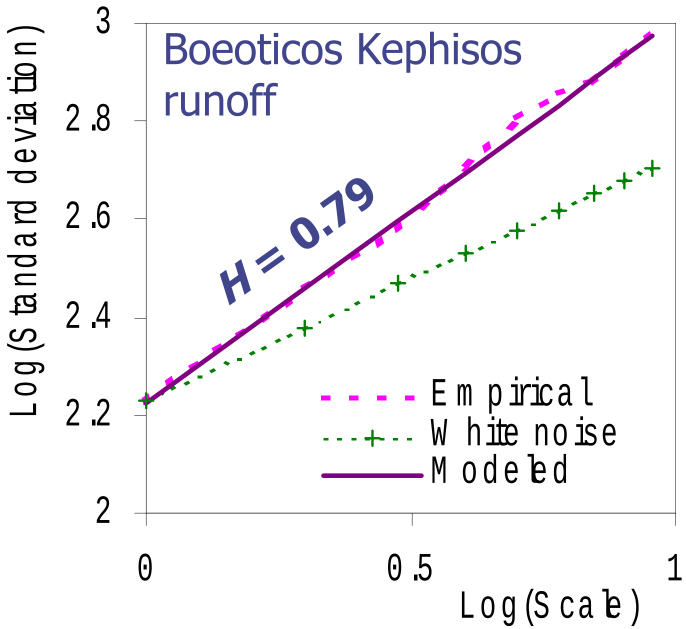
A stochastic process at the annual scale	X_i
The mean of X_i	$\mu := \mathrm{E}[X_i]$
The standard deviation of X_i	$\sigma := \sqrt{\mathrm{Var}[X_i]}$
The lag-j autocorrelation of X_i	$\rho_j := \mathrm{Corr}[X_i, X_{i-j}]$
The aggregated stochastic process at scale $k \geq 1$	$Z_i^{(k)} := \sum_{l=(i-1)k+1}^{ik} X_l$
The mean of $Z_i^{(k)}$	$E[Z_i^{(k)}] = k\,\mu$
The standard deviation of $Z_i^{(k)}$	$\sigma^{(k)} := \sqrt{\mathrm{Var}\,[Z_i^{(k)}]}$
Definition of a simple scaling stochastic process or a simple scaling signal (SSS; also known as (a) stationary increments of self-similar process (b) Fractional Gaussian noise – FGN)	$(Z_i^{(k)} - k\mu) \stackrel{\mathrm{d}}{=} \left(\frac{k}{l}\right)^H (Z_j^{(l)} - l\mu)$ for any scales k and l and for a specified H ($0 < H < 1$) known as the Hurst coefficient
The standard deviation of an SSS $Z_i^{(k)}$ (a power law of scale k)	$\sigma^{(k)} = k^H\,\sigma$
The lag-j autocorrelation of an SSS $Z_i^{(k)}$ (a power law of lag j; independent of scale k)	$\rho_j^{(k)} = \rho_j \approx H(2H-1)j^{2H-2}$ for $j > 0$

How do the series of the examples behave?

- Like white noise?
  - $H = 0.5$
  - Classic statistics adequate

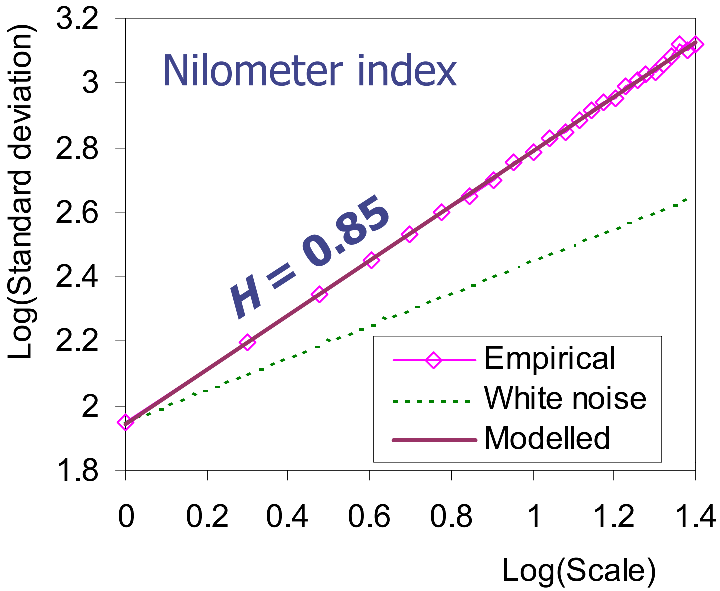
Or

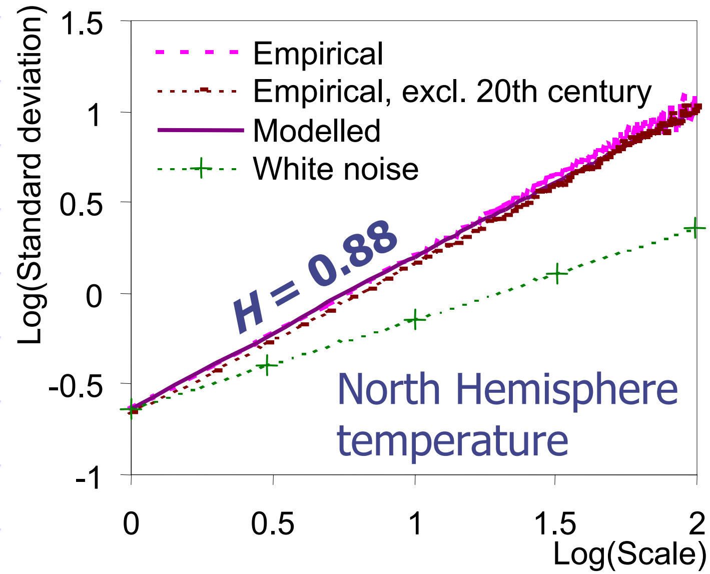
- Like SSS?
  - $H > 0.5$
  - Classic statistics inadequate

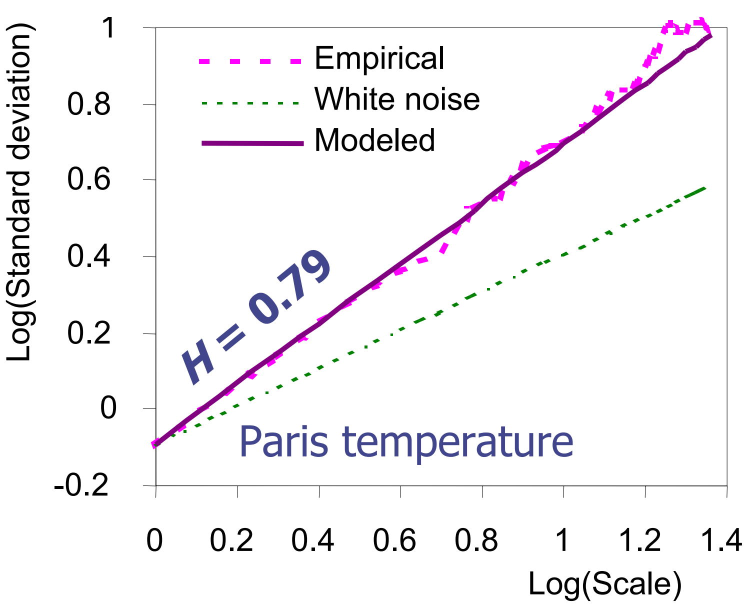


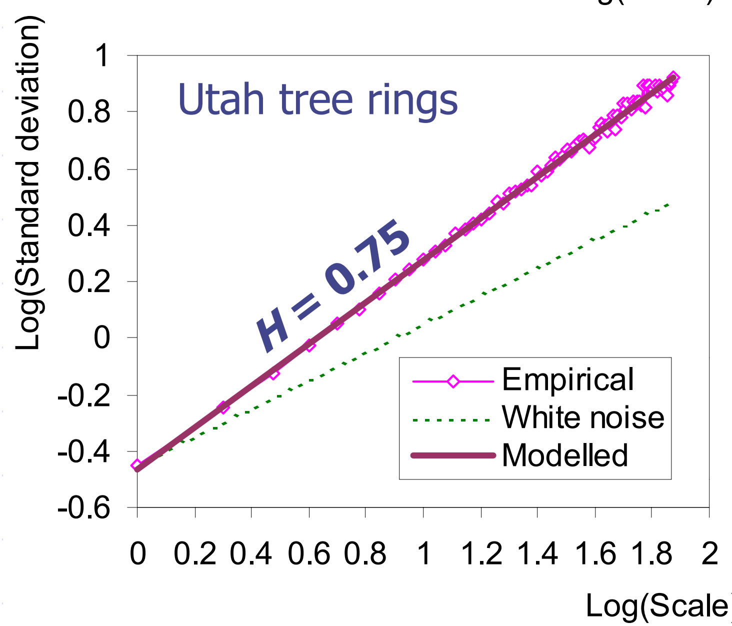
Note: Traditionally, in hydrological statistics the Hurst exponent has been defined and estimated in terms of the quantity called “range”. This in not necessary at all, as it can be much more conveniently determined in terms of the standard deviation of the aggregated process on many temporal scales.

How do the series of the examples behave? (2)









Do classical statistics apply to SSS processes?

Statistic	Classical formula	Effect in SSS processes	SSS formula
Sample average	$\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$	Unbiased	$\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$
Variance of sample average	$\mathrm{var}[\bar{X}] = \frac{\sigma^2}{n}$	Dramatic underestimation	$\mathrm{var}[\bar{X}] = \frac{\sigma^2}{n^{2-2H}}$
Sample standard deviation	$S := \sqrt{\frac{1}{(n-1)}} \times \sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2}$	Underestimation	$\tilde{S} := \sqrt{\frac{n-1/2}{(n-1)(n-n^{2H-1})}} \times \sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2}$
Variance of sample standard deviation	$\mathrm{var}[S] \approx \frac{\sigma^2}{2(n-c)}$	Underestimation	$\mathrm{var}[\tilde{S}] \approx \frac{(0.1n + 0.8)^{\lambda(H)}\sigma^2}{2(n-1)}$ $[\lambda(H) := 0.088(4H^2-1)^2]$
Hurst coefficient	Based on $S^{(k)} = k^H S$ and using regression [The algorithm based on the *range* concept is inappropriate]	Underestimation	Based on $\tilde{S}^{(k)} = k^H \tilde{S}$ and using regression and iteration [Note: $\tilde{S}$ depends on H]

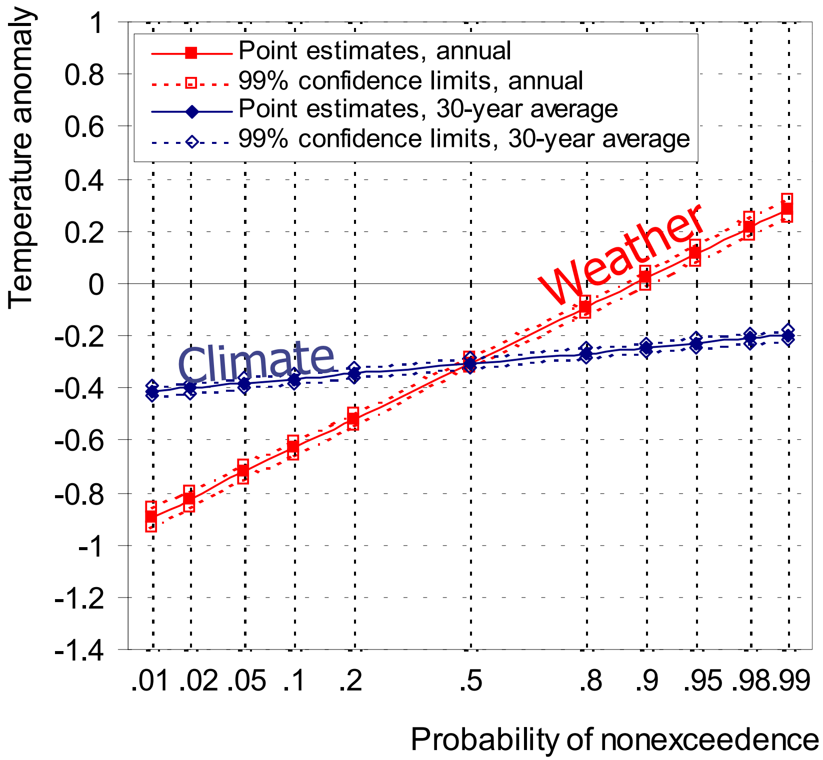
Do classical statistics apply to SSS processes? (2)

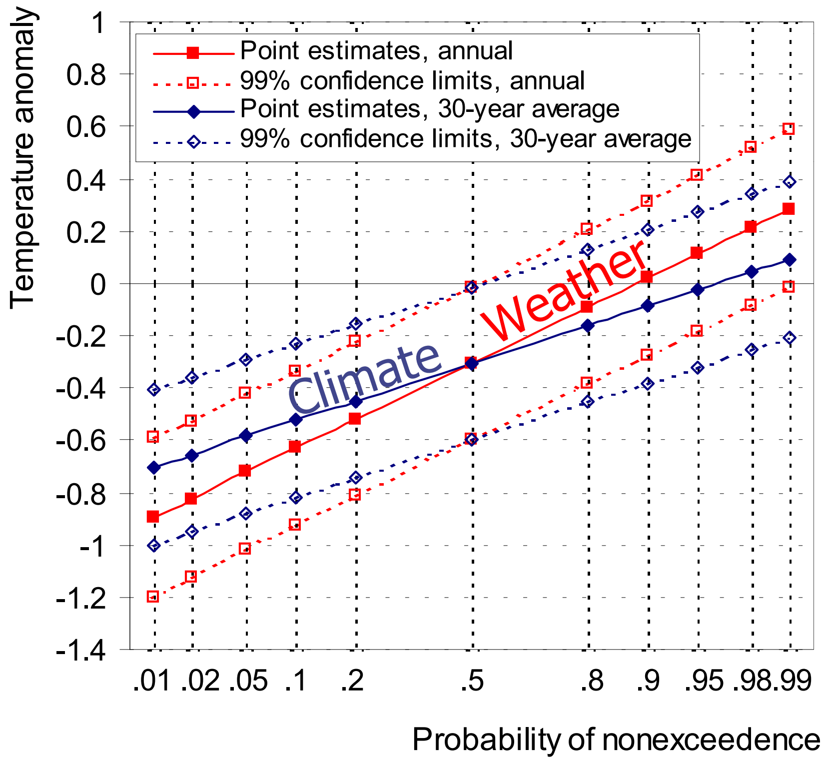
Statistic	Classical formula	Effect in SSS processes	SSS formula
Confidence intervals of distribution quantiles (for normal distribution)	$\hat{x}_{u1,2} = \hat{x}_u \pm \zeta_{(1+\gamma/2)}\varepsilon_u$ with $\varepsilon_u = \frac{s}{\sqrt{n}}\sqrt{1 + \frac{\zeta_u^2}{2}}$	Dramatic underestimation of interval length	$\hat{z}^{(k)}_{u1,2} = \hat{z}^{(k)}_u \pm \zeta_{(1+\gamma/2)}\varepsilon^{(k)}_u$ with $\varepsilon^{(k)}_u = k\frac{\tilde{\tilde{s}}}{n^{1-H}} \times \sqrt{1 + \frac{\zeta_u^2\,(0.1n + 0.8)^{\lambda(H)}}{2(k/n)^{2-2H}\,(n-1)}}$
Cross-correlation	$R_{XY} := \frac{S_{XY}}{S_X S_Y}$ with $S_{XY} := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$	Approximately unbiased	$R_{XY} := \frac{S_{XY}}{S_X S_Y}$
Auto-correlation	$R_l := \frac{n}{n-1}\frac{G_l}{S^2}$	Dramatic underestimation	$\tilde{R}_l := R_l\left(1 - \frac{1}{n^{2-2H}}\right) + \frac{1}{n^{2-2H}}$

Application 1: A simple calculation to demonstrate the difference between classical and SSS statistics

- From the Boeoticos Kephisos runoff series for $n = 91$ (years), the sample mean is $\bar{x} = 392.8$ hm^3 and the classical sample standard deviation $s = 157.3$ hm^3.
- For the same series, the SSS estimate of $H = 0.79$ and thus the sample standard deviation becomes $\tilde{\tilde{s}} = 170.2$ hm^3 (8% greater than s).
- The classical 95% confidence limits of the mean μ are 425.1 hm^3 and 360.5 hm^3 (confidence interval = 64.7 hm^3).
- The SSS 95% confidence limits of the mean μ for $H = 0.79$ are 522.1 and 263.4 hm^3 (confidence interval = 258.8 = 3.0 × 64.7 hm^3).
- To obtain a confidence interval as small as that given by the classical statistics, the required number of years of observations is $n = 67\,175$. That is, we must … wait 67 084 years (!) most probably seeing our experiment interrupted much earlier by a new glacial period.

Application 2: Distribution quantiles of the North Hemisphere temperature at two time scales, annual (mean annual weather) and 30 year (climate)





Climate is what you expect
Weather is what you get

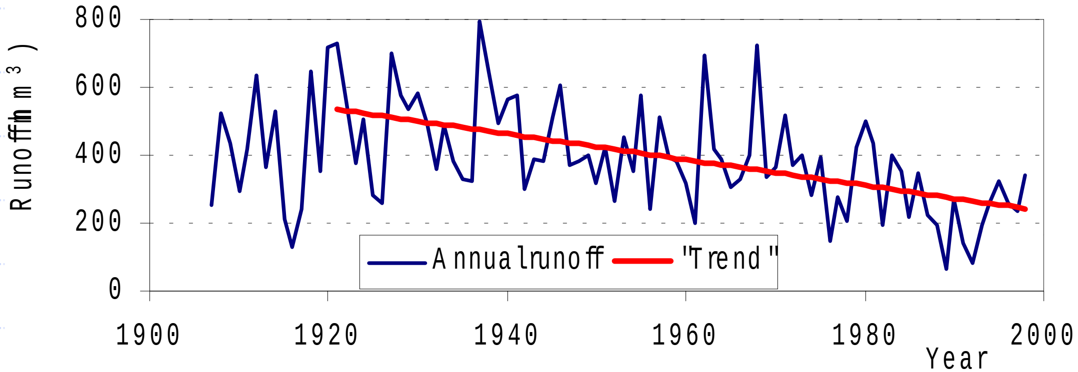
Weather is what you get
Climate is what you get
… if you expect for many years

Application 3: Statistical test of a trend

Kendall's τ statistic:

$$\tau := \frac{4p}{n(n-1)} - 1$$

where p is the number of pairs $(x_j, x_i;\ j > i,\ x_j < x_i)$.
In a random series: $E[\tau] = 0$, $\text{var}[\tau] = 2(2n+5)/9n(n-1)$, normal distribution.



- Classical procedure
  - Null hypothesis: random series; alternative hypothesis: trend
  - For n = 78, var[τ] = 0.077, τ = 0.40 = 5.2 var[τ]
  - Reject the null hypothesis; attained significance level 8.8×10^{-8}
- Modified procedure
  - Null hypothesis: SSS series, H = 0.79; alternative hypothesis: trend
  - Generate an ensemble of 100 time series, each with n = 91
  - In each series locate the 78-year period with the maximum τ
  - Estimate var[τ] = 0.252 and Pr[$\tau \geq 0.40$] = 0.055
  - Do not reject the null hypothesis at significance level 5%.

Discussion and conclusions

- It is known that anthropogenic climate change (CO_2 emissions etc.) increases uncertainty.
- Even without anthropogenic forcings, the climate varies on all time scales.
- Hydrological statistics, in its current status, has been based on the implicit assumption of a stable climate.
- In hydrological applications, classical statistics:
  - Describes only a portion of natural uncertainty;
  - Underestimates seriously the risk;
  - May characterize a regular behaviour of hydroclimatic processes as an unusual phenomenon;
  - In short series, hides the scaling behaviour of processes.

Discussion and conclusions (2)

- The Hurst phenomenon and the SSS processes offer a solid and convenient basis to adapt hydrological statistics so as to be consistent with a varying climate.
- It is feasible to derive estimators applicable to SSS processes for most statistics.
- In cases where analytical solutions are not feasible, stochastic simulation using SSS processes offers a convenient alternative.
- The SSS statistical framework is a feasible step towards making analyses closer to reality.

Discussion and conclusions (3)

- Application of the SSS statistical framework demonstrates the much higher uncertainty, especially in:
  - Confidence interval estimates at all time scales;
  - Point or interval estimates at overyear timescales (climatic indicators).
- Application of the SSS statistical framework demonstrates that observed overyear "trends" or "shifts" may not be "changes" but regular hydroclimatic behaviour.
- If the simple scaling behaviour hypothesis is correct, then the detection of anthropogenic effects in hydroclimatic time series:
  - Should be done using SSS rather than classical statistics;
  - Is much more unlikely to result in statistically significant changes.

This presentation is available on line at http://www.itia.ntua.gr/e/docinfo/565/

References

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

Hurst, H. E. (1951) Long term storage capacities of reservoirs. *Trans. ASCE* **116**, 776–808.

Jones, P. D., Briffa, K. R., Barnett, T. P. & Tett, S. F. B. (1998) High-resolution paleoclimatic records for the last millennium: interpretation, integration and comparison with General Circulation Model control-run temperatures. *Holocene* **8**(4), 455–471.

Koutsoyiannis, D. (2002) The Hurst phenomenon and fractional Gaussian noise made easy, Hydrological Sciences Journal, 47(4), 573-595.

Koutsoyiannis, D. (2003) Climate change, the Hurst phenomenon, and hydrological statistics, Hydrological Sciences Journal, 48(1), 3-24.

Mesa, O. J. & Poveda, G. (1993) The Hurst effect: the scale of fluctuation approach. *Wat. Resour. Res.* **29**(12), 3995–4002.

National Research Council (1991) *Opportunities in the Hydrologic Sciences*, 21. National Academy Press, Washington DC, USA.