European Geosciences Union 1st General Assembly
Nice, France, 25-30 April 2004

NP3.01 Scaling and nonlinearity in the hydrological cycle

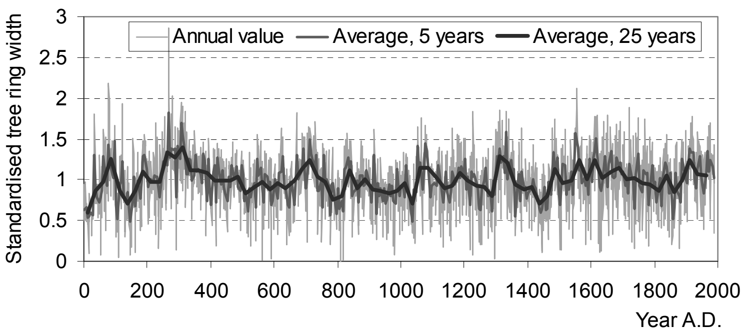
Simple methods to generate time series with scaling behaviour

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

Visual recognition of scaling

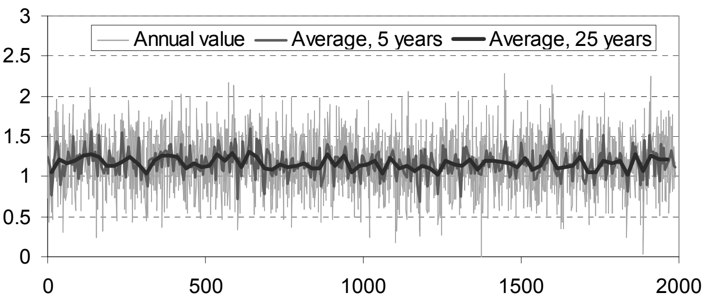
Standardised tree ring widths from a paleoclimatological study at Mammoth Creek, Utah, for the years 0-1989 (1990 years; from ftp://ftp.ngdc.noaa.gov/paleo/)

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



Scaling can be studied in terms of the behaviour of a time series aggregated on different time scales

Original formulation of the scaling behaviour

- The scaling behaviour is equivalent to the Hurst phenomenon (or long-range dependence, or long-term persistence, or the Joseph effect) and has been found to be omnipresent in several hydroclimatic and other (long) time series
- The Hurst phenomenon is typically formulated in terms of the statistical properties of a quantity called "range", (Hurst, 1951) which describes the difference of **accumulated inflows minus outflows** from a hypothetical infinite **reservoir**
- In this respect, it has been regarded that it **affects the reservoir planning, design and operation**, but only when the reservoir performs **multi-year regulation** (e.g. Klemeš et al., 1981)

Simpler formulation of the Hurst phenomenon

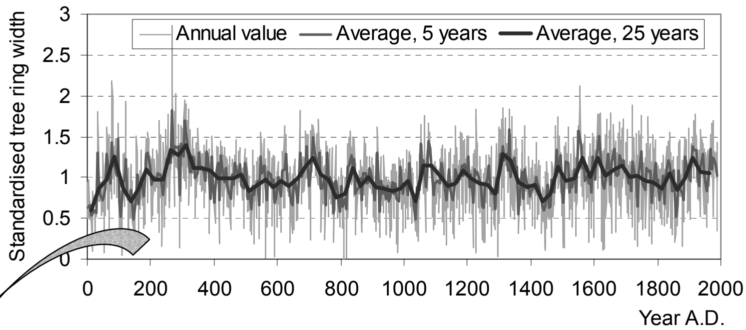
A process at the annual scale	X_i
The mean of X_i	$\mu := \mathrm{E}[X_i]$
The standard deviation of X_i	$\sigma := \sqrt{\mathrm{Var}[X_i]}$
The aggregated process at a multi-year scale $k \geq 1$	$Z_1^{(k)} := X_1 + \ldots + X_k$ $Z_2^{(k)} := X_{k+1} + \ldots + X_{2k}$ $\vdots$ $Z_i^{(k)} := X_{(i-1)k+1} + \ldots + X_{ik}$
The mean of $Z_i^{(k)}$	$E[Z_i^{(k)}] = k\,\mu$
The standard deviation of $Z_i^{(k)}$	$\sigma^{(k)} := \sqrt{\mathrm{Var}\,[Z_i^{(k)}]}$
if consecutive X_i are independent	$\sigma^{(k)} = \sqrt{k}\,\sigma$
if consecutive X_i are positively correlated	$\sigma^{(k)} > \sqrt{k}\,\sigma$
if X_i follows the Hurst phenomenon	$\sigma^{(k)} = k^H \sigma$ (0.5 < H <1)
Extension of the standard deviation scaling and definition of a simple scaling stochastic process (SSS)	$(Z_i^{(k)} - k\mu) \overset{\mathrm{d}}{=} \left(\frac{k}{l}\right)^H (Z_j^{(l)} - l\mu)$ for any scales k and l

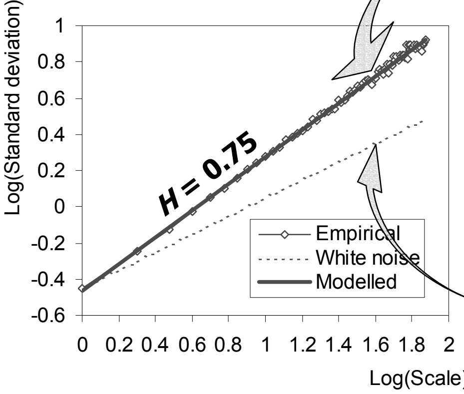
The power-laws of the second-order properties of an SSS process

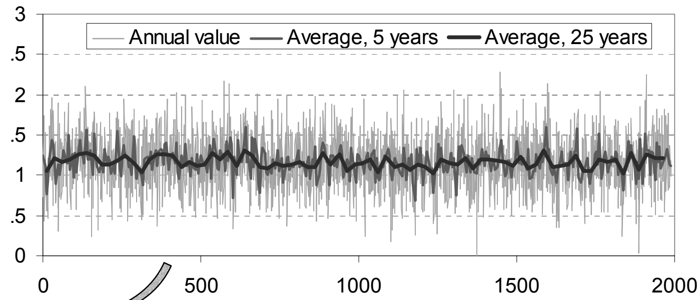
The standard deviation of $Z_i^{(k)}$ (a power law of scale k)	$\sigma^{(k)} = k^H \sigma$
The lag-j autocorrelation of $Z_i^{(k)}$ (a power law of lag j; independent of scale k)	$\rho_j^{(k)} = \rho_j \approx H(2H-1)\lvert j \rvert^{2H-2}$
The lag-j autocovariance of $Z_i^{(k)}$ (a power law of scale k and lag j)	$\gamma_j^{(k)} \approx H(2H-1)\ \gamma_0\ k^{2H} \lvert j \rvert^{2H-2}$
The power spectrum of $Z_i^{(k)}$ (a power law of scale k and frequency ω)	$s_\gamma^{(k)}(\omega) \approx 4\,(1-H)\ \gamma_0\ k^{2H}\,(2\,\omega)^{1-2H}$

Tracing and quantification of the Hurst phenomenon

The standardised tree ring width series at Utah







A white noise series (for comparison)

Why the Hurst phenomenon is important in statistics and engineering applications

- Fundamental law of classical statistics

$$\mathrm{StD}[\bar{X}] = \frac{\sigma}{\sqrt{n}}$$

$\bar{X}$ = sample mean
σ = standard deviation
n = sample size

- Modified law for SSS

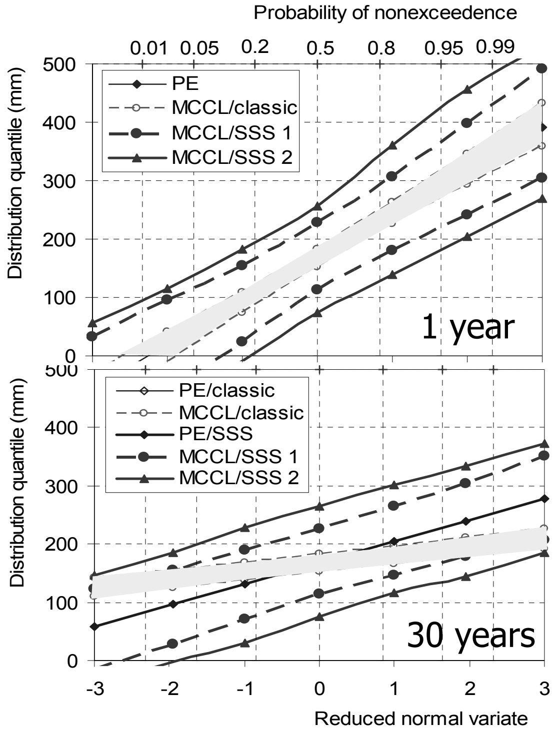
$$\mathrm{StD}[\bar{X}] = \frac{\sigma}{n^{1-H}}, \ H > 0.5$$

- Example
  To obtain $\mathrm{StD}[\bar{X}] / \sigma = 10\%$
  - n = 100 for classical statistics
  - n = 10 000 for SSS
    with H = 0.75 (as in the example)

The scaling behaviour increases uncertainty dramatically

See additional discussion in Koutsoyiannis (2003b)

Comparisons of runoff uncertainty: 1- and 30-year scales



Dependence structure	Parameters	Total uncertainty, % of mean	
		Annual scale	30-year scale
IID	m^*, s^*	174	32
IID	m, s	206	50
SSS	m^*, s^*, H^*	174	87
SSS	m, s, H^*	236	165
SSS	m, s, H	268	199

Parameters marked with * are fixed

Case study: Runoff of the Boeoticos Kephisos river, Greece
Record length: n = 96
Mean: μ = 167.7 mm
Standard deviation: σ = 74.5 mm
Hurst coefficient: H = 0.79
Normal distribution
Confidence: $a = \alpha$ = 95%

Source: Koutsoyiannis and Efstratiadis (2004)

Why generation of time series with scaling behaviour is important

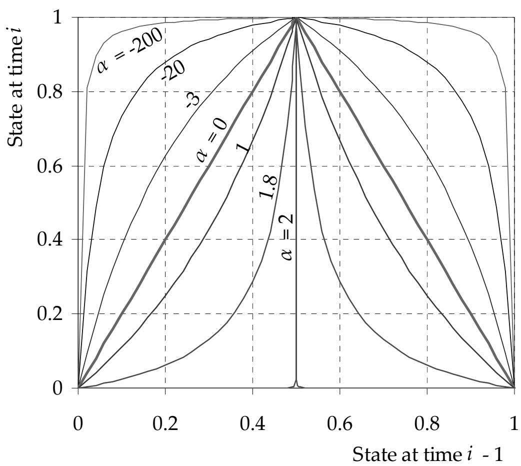
- Theoretical solutions in prediction and estimation problems may be infeasible
- Stochastic simulation is the method of choice for such problems in hydrosystems modelling and management and in climatic studies
- However, common stochastic models do not respect the scaling behaviour at present, thus ignoring a significant source of uncertainty of hydrosystems
- One of the reasons is related to the difficult handling of SSS processes
- With this motivation, four simple methods, utilising and simultaneously highlighting different aspects of SSS processes, are discussed

Algorithm 1: Deterministic with simplified dynamics

- The first method emphasises the fact that simple nonlinear dynamics may produce time series with erratic yet simple scaling behaviour
- Starting point:
  The generalised tent map

$$x_i = g(x_{i-1}; a) = \frac{(2 - a)\min(x_{i-1}, 1 - x_{i-1})}{1 - a \min(x_{i-1}, 1 - x_{i-1})}$$

with $0 \le x_i \le 1$, $a < 2$



- Example uses:
  1. Approximates the relation between successive maxima simplified climatic dynamics described by the Lorenz equations (Lasota and Mackey, 1994)
  2. Can describe the compound effect of positive and negative feedbacks in the climate system (Koutsoyiannis, 2003b)

Algorithm 1: The double tent map

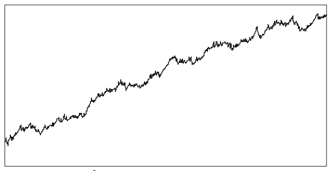
- Make parameter of the tent map time dependent using the same (tent) map, and obtain the double tent map

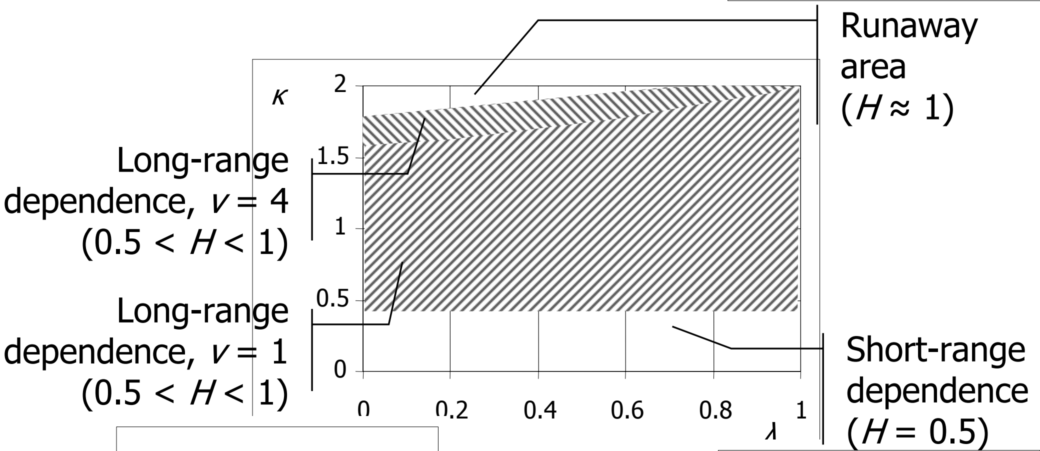
  $u_i = G(u_{i-1}, a_{i-1}; \kappa, \lambda) = g(u_{i-1}; \kappa\, a_{i-1})$ with $a_i = g(a_{i-1}; \lambda)$
- Extend the double tent map by adding $\nu - 1$ hidden terms

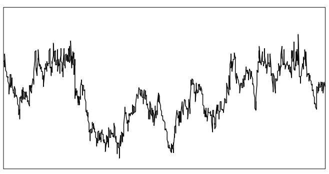
  $u_i = y_{\nu i}$ with $y_{\nu i} = G(y_{\nu i-1}, a_{\nu i-1}; \kappa, \lambda)$, $y_0 = u_0$, $i = 0, 1, 2, \ldots$
- Apply an additional transformation to shift from [0, 1] to [0, ∞) or to (–∞, ∞), e.g.

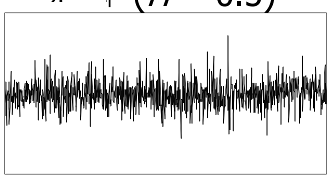
  $x_i = b + c \tan(\pi z_i / 2)^d$ $\quad$ $x_i = b + c \ln[u_i / (1 - u_i)]$
- The final model for x_i
  - is two dimensional (involves two degrees of freedom corresponding to the initial conditions a_0 and z_0)
  - contains up to five real-valued parameters (κ, λ, b, c, d) and an integral one (ν)

Algorithm 1: Model behaviour with respect to parameters κ and λ



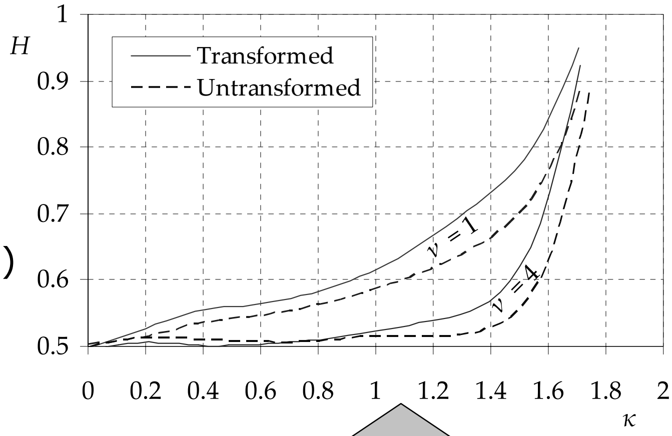






Algorithm 1: Parameter estimation and generation procedure

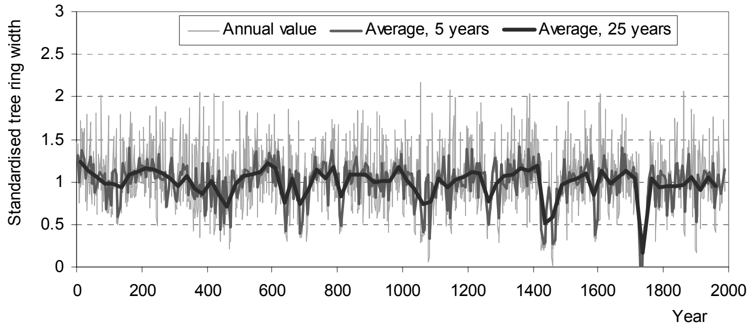
- Assume fixed values of parameters λ = 0.001 and ν = 4
- From the known H estimate parameter κ from the figure (constructed by simulation)
- Generate a series u_i from the double tent map
- Transform the series by $z_i = \ln[u_i / (1 - u_i)]$
- Estimate parameters b and c so that the final series $x_i = b + c\, z_i$ has the desired mean and standard deviation

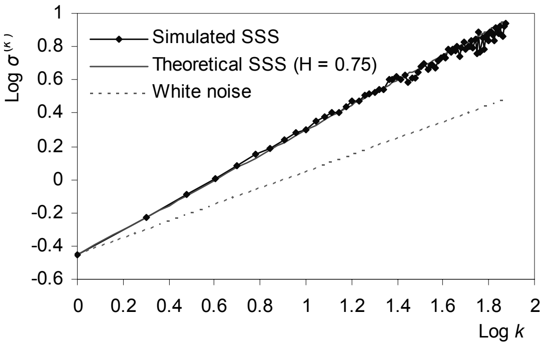


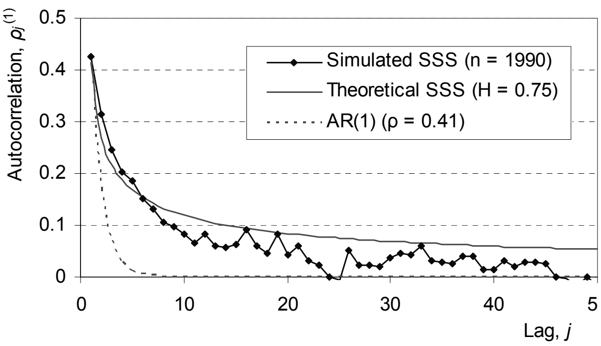
Hurst coefficient of a time series generated from the double tent map for parameter values λ = 0.001, κ ranging from 0 to 2 and ν = 1 and 4

Algorithm 1: Results

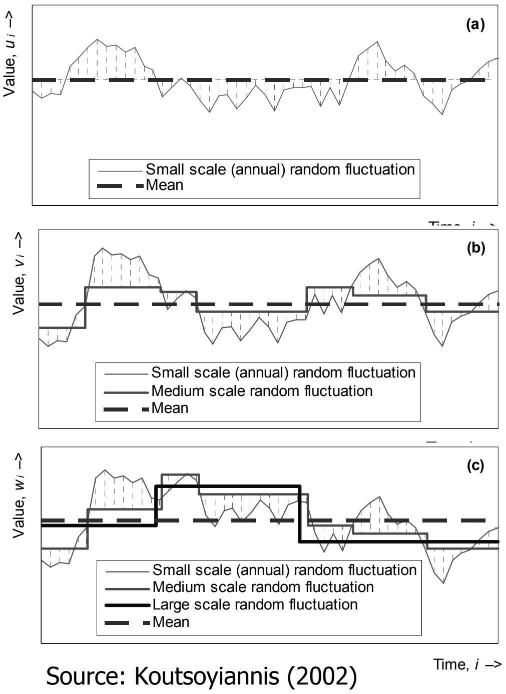
A synthetic time series with length and statistics equal to those of the tree rings at Utah, generated by the double tent map algorithm

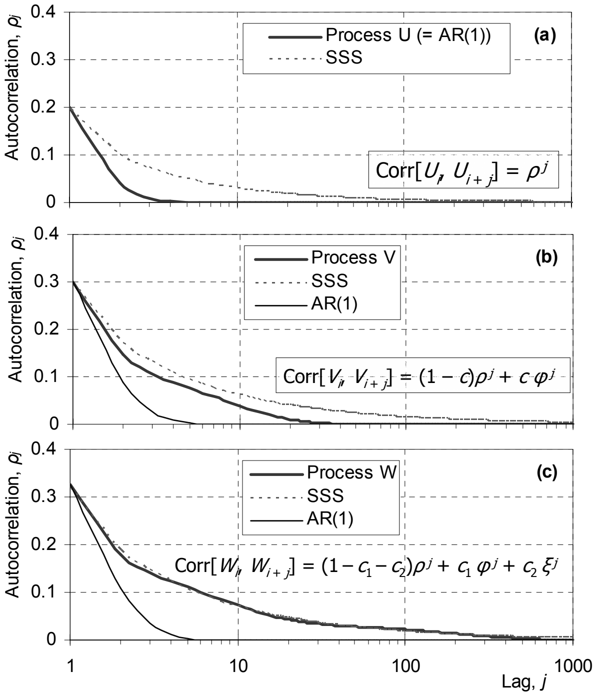






Algorithm 2: An SSS process as the result of random fluctuations at many time scales





Source: Koutsoyiannis (2002)

Algorithm 2: The weighted sum of three Markovian processes

An SSS process X_i can be approximated by the sum of three AR(1) processes:

$$X_i = A_i + B_i + C_i$$

with lag one autocorrelations respectively

$$\rho = 1.52\,(H - 0.5)^{1.32},$$

$$\varphi = 0.953 - 7.69\,(1 - H)^{3.85},$$

$$\xi = \begin{cases} 0.932 + 0.087\,H, & H \leq 0.76, \\ 0.993 + 0.007\,H, & H > 0.76 \end{cases}$$

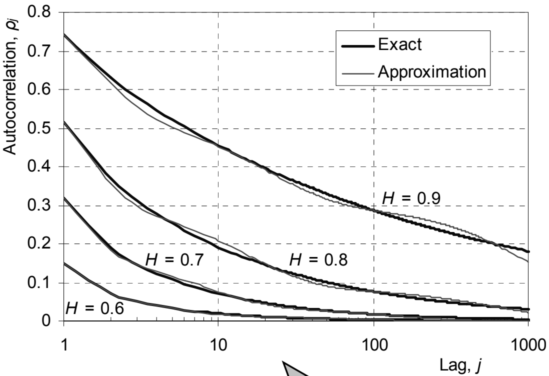
and variances respectively,

$$(1 - c_1 - c_2)\,\gamma_0, \quad c_1\,\gamma_0, \quad c_2\,\gamma_0$$

where c_1 and c_2 are estimated so that the autocorrelation of the sum of the three processes

$$\rho_j = (1 - c_1 - c_2)\rho^j + c_1\,\varphi^j + c_2\,\xi^j$$

match the theoretical SSS autocorrelation for lags 1 and 100

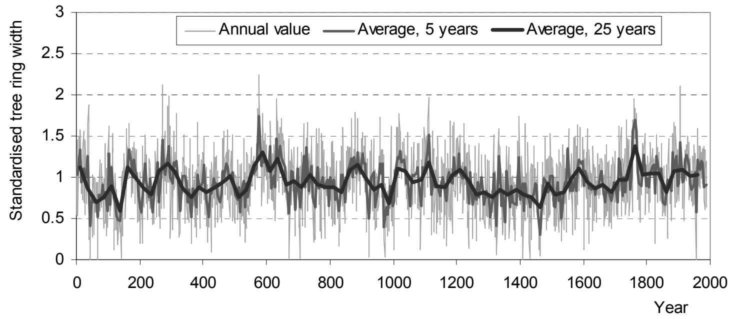


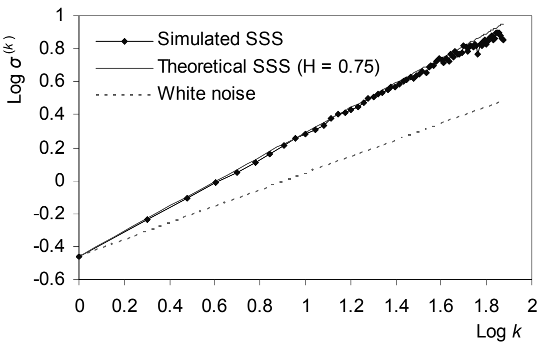
Degree of approximation of the SSS autocorrelation attained by the use of three AR(1) processes

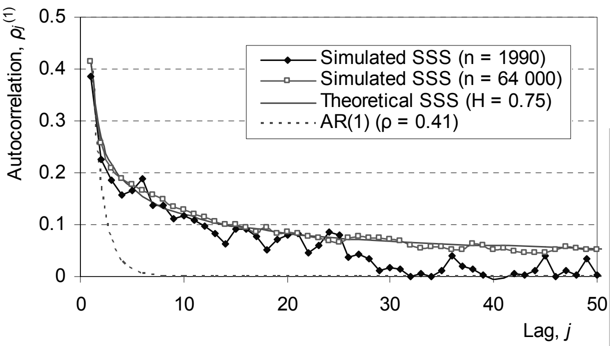
Source: Koutsoyiannis (2002)

Algorithm 2: Results

A synthetic time series with length and statistics equal to those of the tree rings at Utah, generated by the three Markovian processes algorithm

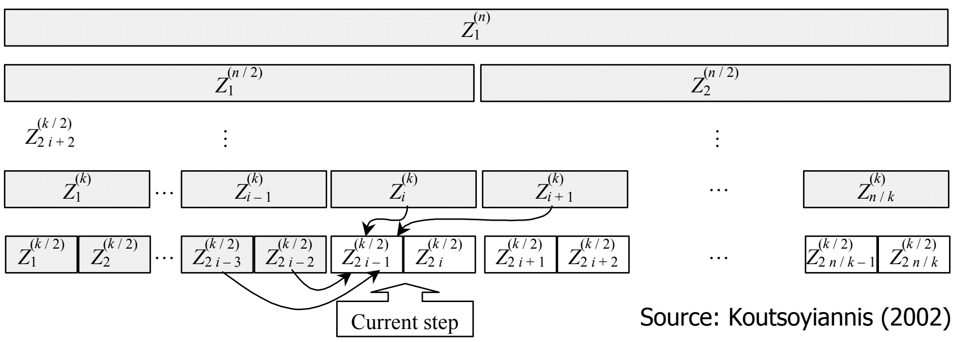






Source: Koutsoyiannis (2002)

Algorithm 3: Disaggregation based on the invariant properties of an SSS process at different time scales



Source: Koutsoyiannis (2002)

The process X_i (i = 1, …, n, where n is assumed a power of 2) is generated in consecutive steps. In step one, the sum $Z_1^{(n)}$ for the total period n is generated. In the second step, this is disaggregated in two components $Z_1^{(n/2)}$ and $Z_2^{(n/2)}$ etc. In each disaggregation step,

$$Z_{2i-1}^{(k/2)} + Z_{2i}^{(k/2)} = Z_i^{(k)}$$

whereas the autocorrelations with earlier lower-level variables (scale k / 2) and later higher-lever variables (scale k) are preserved.

Algorithm 3: Parameter estimation and generation procedure

In each disaggregation step the first lower-level variable, $Z_{2i-1}^{(k/2)}$, is generated from

$$Z_{2i-1}^{(k/2)} = a_2 Z_{2i-3}^{(k/2)} + a_1 Z_{2i-2}^{(k/2)} + b_0\, Z_i^{(k)} + b_1\, Z_{i+1}^{(k)} + V$$

and the second one, $Z_{2i}^{(k/2)}$, from

$$Z_{2i-1}^{(k/2)} + Z_{2i}^{(k/2)} = Z_i^{(k)}$$

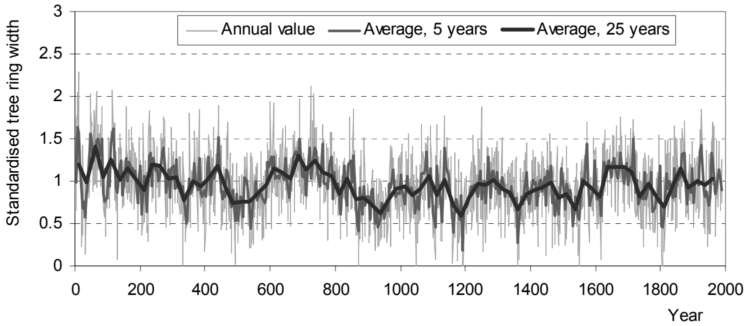
where parameters a_2, a_1, b_0 και b_1 and the variance of the random variable V are estimated in terms of correlations Corr$[Z_{2i-1}^{(k/2)}, Z_{2i-1+j}^{(k/2)}] = \rho_j$ and the variance $\gamma_0^{(k/2)}$ according to

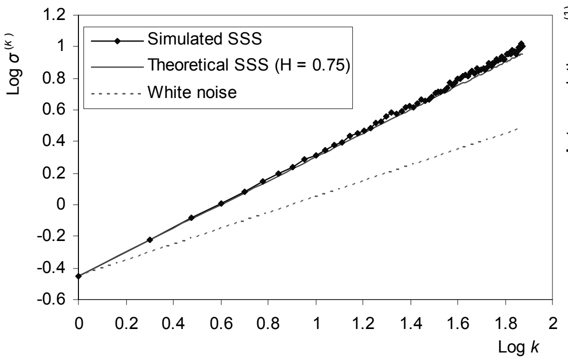
$$\begin{bmatrix} a_2 \\ a_1 \\ b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 + \rho_3 & \rho_4 + \rho_5 \\ \rho_1 & 1 & \rho_1 + \rho_2 & \rho_3 + \rho_4 \\ \rho_2 + \rho_3 & \rho_1 + \rho_2 & 2(1 + \rho_1) & \rho_1 + 2\rho_2 + \rho_3 \\ \rho_4 + \rho_5 & \rho_3 + \rho_4 & \rho_1 + 2\rho_2 + \rho_3 & 2(1 + \rho_1) \end{bmatrix}^{-1} \begin{bmatrix} \rho_2 \\ \rho_1 \\ 1 + \rho_1 \\ \rho_2 + \rho_3 \end{bmatrix}$$

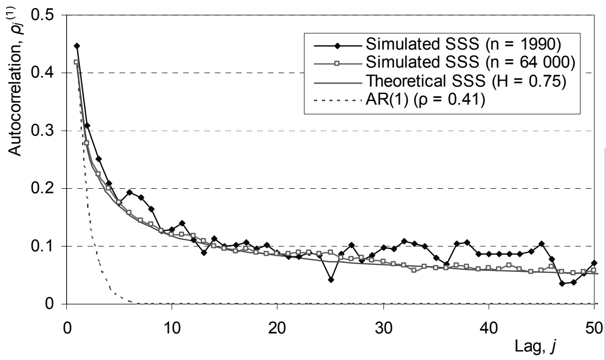
$$\mathrm{Var}[V] = \gamma_0^{(k/2)}(1 - [\rho_2, \rho_1, 1 + \rho_1, \rho_2 + \rho_3]\, [a_2, a_1, b_0, b_1]^T)$$

Algorithm 3: Results

A synthetic time series with length and statistics equal to those of the tree rings at Utah, generated by the disaggregation algorithm







Source: Koutsoyiannis (2002)

Algorithm 4: Filtering white noise through a symmetric moving average filter

The symmetric moving average (SMA) generating scheme (Koutsoyiannis, 2000) transforms a white noise sequence V_i into a process X_i with given autocorrelation function by

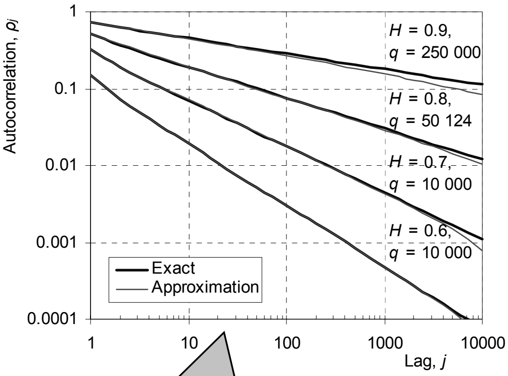
$$X_i = \sum_{j=-q}^{q} a_{|j|} \, V_{i+j}$$

where a_j are weighting factors whose number q is theoretically infinite but in practice can take a finite value. For an SSS process:

$$a_j \approx \frac{\sqrt{(2-2H)\,\gamma_0}}{3-2H} \times \left(|j+1|^{H+0.5} + |j-1|^{H+0.5} - 2\,|j|^{H+0.5}\right)$$

The method can also preserve the skewness ξ_X of X_i assuming that the white noise has skewness ξ_V determined from

$$\left(a_0^3 + 2\sum_{j=1}^{q} a_j^3\right)\xi_V = \xi_X \, \gamma_0^{3/2}$$

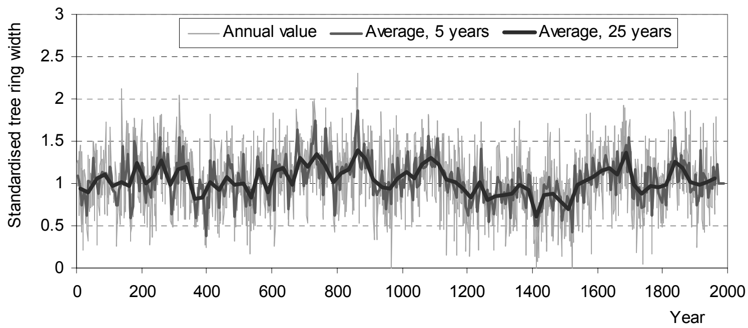


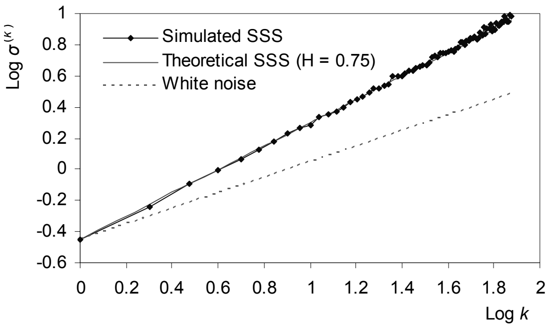
Degree of approximation of the SSS autocorrelation attained by the SMA filter method

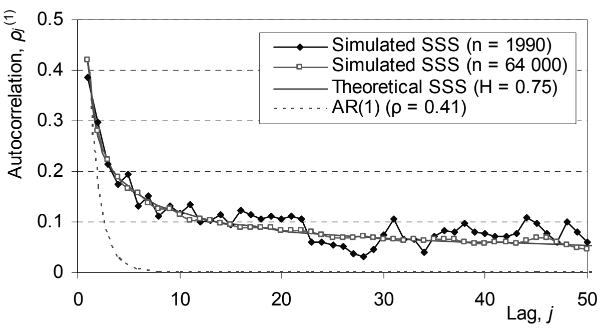
Source: Koutsoyiannis (2002)

Algorithm 4: Results

A synthetic time series with length and statistics equal to those of the tree rings at Utah, generated by the disaggregation algorithm







Source: Koutsoyiannis (2002)

Generalisation of the SMA method and the **Castalia** software package

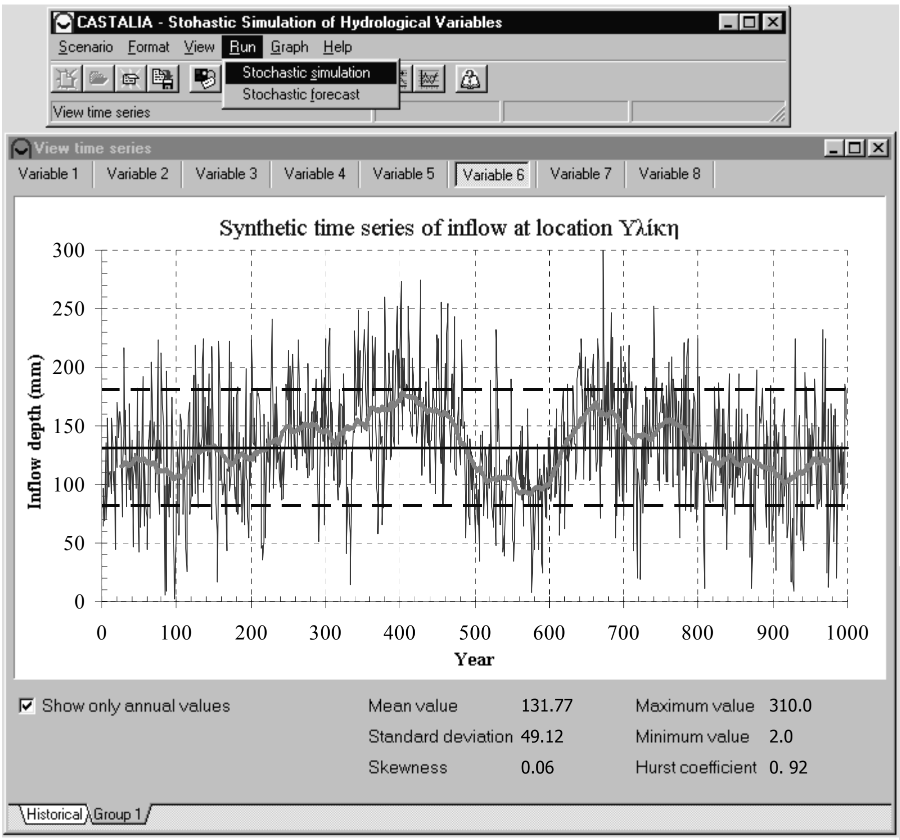
- The SMA method has been generalised so that
  - It can perform with any autocorrelation function γ_j of the process X_i; in this case the weights a_j are determined from

    $$s_a(\omega) = [2\ s_\gamma(\omega)]^{1/2}$$

    where $s_a(\omega)$ and $s_\gamma(\omega)$ the DFTs of the series a_j and γ_j, respectively
  - It can simulate several hydrological variables at multiple sites preserving joint second order statistics (cross-correlations)
  - It can preserve essential marginal statistics up to third order (skewness)
  - It is combined with a disaggregation model so that it generates processes at sub-annual scales
- The generalised method has been coded into the **Castalia** software, written in **Delphi** and utilising the **Oracle** data base

See additional information in Koutsoyiannis (2000); Koutsoyiannis and Efstratiadis (2001); Langousis and Koutsoyiannis (2003)

Castalia: Stochastic simulation with long-term persistence



Conclusion

- The scaling behaviour seems to be an omnipresent characteristic of hydroclimatic time series
- This behaviour manifests the great uncertainty and unpredictability of the hydroclimatic processes
- In simulations of hydrosystems it is important to preserve the scaling behaviour (the Hurst phenomenon should not be regarded as "a ghost to be conjured away"; Klemeš, 1974)
- This is not a difficult task and can be achieved with simple algorithms even in a spreadsheet environment
- Even a simple two-dimensional deterministic toy model can generate series respecting the scaling behaviour of hydroclimatic processes
- In a stochastic context, the scaling behaviour can be represented by the weighted sum of three Markovian processes
- Stepwise disaggregation can yield another simple method to generate time series with scaling behaviour
- Symmetric moving average filtering of white noise yields another simple method to generate simple scaling time series

This presentation is available on line at http://www.itia.ntua.gr/e/docinfo/607/

References

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

Klemeš, V. (1974), The Hurst phenomenon: a puzzle? *Wat. Resour. Res.* 10(4), 675–688.

Klemeš, V., Sricanthan, R. and McMahon T. A. (1981), Long-memory flow models in reservoir analysis: What is their practical value? *Water Resources Research*, 17(3), 737-751

Koutsoyiannis, D. (2000), A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, *Water Resources Research*, 36(6), 1519-1533

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

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

Koutsoyiannis, D. (2003b), A toy model of climatic variability with scaling behaviour, *Hydrofractals '03, An international conference on fractals in hydrosciences*, Monte Verita, Ascona, Switzerland, August 2003, ETH Zurich, MIT, Université Pierre et Marie Curie (http://www.itia.ntua.gr/g/docinfo/585/)

Koutsoyiannis, D., and Efstratiadis, A. (2001), A stochastic hydrology framework for the management of multiple reservoir systems, *26th General Assembly of the European Geophysical Society, Geophysical Research Abstracts*, Vol. 3, Nice, March 2001, European Geophysical Society, (http://www.itia.ntua.gr/e/docinfo/54/).

Koutsoyiannis, D., and Efstratiadis, A. (2004), Climate change certainty versus climate uncertainty and inferences in hydrological studies and water resources management, *1st General Assembly of the European Geosciences Union, Geophysical Research Abstracts, Vol. 6*, Nice, April 2004 (http://www.itia.ntua.gr/g/docinfo/606/)

Langousis, A., and Koutsoyiannis, D. (2003), A stochastic methodology for generation of seasonal time series reproducing overyear scaling, *Hydrofractals '03, An international conference on fractals in hydrosciences*, Monte Verita, Ascona, Switzerland, August 2003, ETH Zurich, MIT, Université Pierre et Marie Curie (http://www.itia.ntua.gr/g/docinfo/586/)

Lasota, A., and Mackey, M.C. (1994), *Chaos, Fractals and Noise, Stochastic Aspects of Dynamics*, Springer-Verlag