European Geosciences Union General Assembly 2005
Vienna, Austria, 24-29 April 2005
HS27 Times series analysis in hydrology

The long-range dependence of hydrological processes as a result of the maximum entropy principle

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

Some type-"why?" questions

- Why the probability for each outcome of a die is 1/6?
- Why the normal distribution is so common for variables with relatively low variation?
- Why variables with high variation tend to have asymmetric inverse-J-shaped (rather than bell-shaped) distributions?
- Why variables with high variation tend to have a scaling behaviour in state?

Because this behaviour maximizes entropy (i.e. uncertainty)

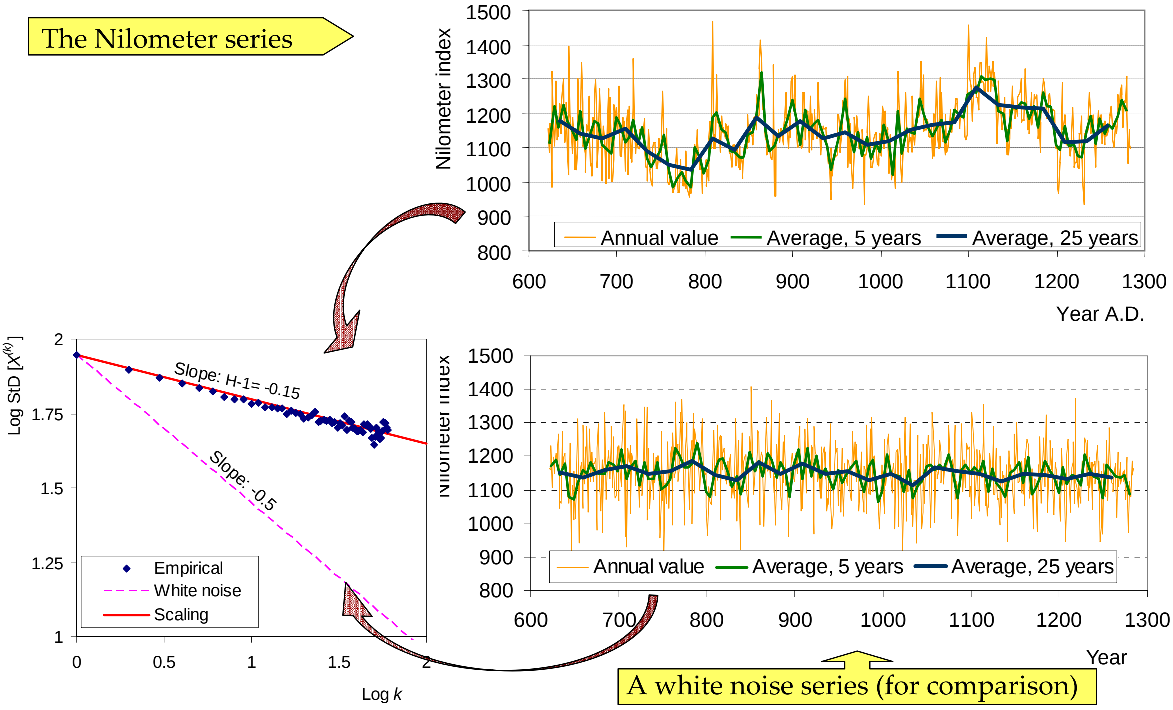
- Why the Hurst phenomenon (scaling behaviour in time) is so common in geophysical, socioeconomical and technological processes?

Same reason?

What is the Hurst phenomenon? (simple scaling behaviour)

A process at the annual scale	X_i
The mean of X_i	$\mu := E[X_i]$
The standard deviation of X_i	$\sigma := \sqrt{\mathrm{Var}[X_i]}$
The aggregated process at a multi-year scale $k \geq 1$	$Y_1^{(k)} := (1/k)\,(X_1 + \ldots + X_k)$ $Y_2^{(k)} := (1/k)\,(X_{k+1} + \ldots + X_{2k})$ $\vdots$ $Y_i^{(k)} := (1/k)\,(X_{(i-1)k+1} + \ldots + X_{ik})$
The mean of $Y_i^{(k)}$	$E[Y_i^{(k)}] = \mu$
The standard deviation of $Y_i^{(k)}$	$\sigma^{(k)} := \sqrt{\mathrm{Var}\,[Y_i^{(k)}]}$
if consecutive X_i are independent	$\sigma^{(k)} = \sigma / \sqrt{k}$
if consecutive X_i are positively correlated	$\sigma^{(k)} > \sigma / \sqrt{k}$
if X_i follows the Hurst phenomenon	$\boxed{\sigma^{(k)} = k^{H-1}\,\sigma}$ $(0.5 < H < 1)$
Extension of the standard deviation scaling and definition of a simple scaling stochastic process	$(Y_i^{(k)} - \mu) \overset{\mathrm{d}}{=} \left(\frac{k}{l}\right)^H (Y_j^{(l)} - \mu)$ for any scales k and l

Tracing and quantification of the Hurst phenomenon
Example: The Nilometer data series



What is entropy?

- For a discrete random variable X taking values x_j with probability mass function $p_j \equiv p(x_j)$, the Boltzmann-Gibbs-Shannon (or extensive) entropy is defined as

$$\phi := E[-\ln p(X)] = -\sum_{j=1}^{w} p_j \ln p_j, \quad \text{where} \quad \sum_{j=1}^{w} p_j = 1$$

- For a continuous random variable X with probability density function $f(x)$, the entropy is defined as

$$\phi := E[-\ln f(X)] = -\int_{-\infty}^{\infty} f(x) \ln f(x)\, dx, \quad \text{where} \quad \int_{-\infty}^{\infty} f(x)\, dx = 1$$

- In both cases the entropy ϕ is a measure of **uncertainty** about X and equals the **information** gained when X is observed.
- In other disciplines (statistical mechanics, thermodynamics, dynamical systems, fluid mechanics), entropy is regarded as a measure of **order** or **disorder** and **complexity**.

Entropic quantities of a stochastic process

- The *order 1 entropy* (or simply *entropy* or *unconditional entropy*) refers to the marginal distribution of the process X_i:

$$\phi := E[-\ln f(X_i)] = -\int f(x) \ln f(x)\, dx$$

- The *order n entropy* refers to the joint distribution of the vector of variables $\mathbf{X}_n = (X_1, \dots, X_n)$ taking values $\mathbf{x}_n = (x_1, \dots, x_n)$:

$$\phi_n := E[-\ln f(\mathbf{X}_n)] = -\int f(\mathbf{x}_n) \ln f(\mathbf{x}_n)\, \mathrm{d}\mathbf{x}_n$$

- The *order m conditional entropy* refers to the distribution of a future variable (for one time step ahead) conditional on known m past and present variables (Papoulis, 1991):

$$\phi_{\mathrm{c},m} := E[-\ln f(X_1 \mid X_0, \dots, X_{-m+1})] = \phi_m - \phi_{m-1}$$

- The *conditional entropy* refers to the case where the entire past is observed:

$$\phi_{\mathrm{c}} := \lim_{m \to \infty} \phi_{\mathrm{c},m}$$

- The *information gain* when present and past are observed is:

$$\psi := \phi - \phi_{\mathrm{c}}$$

Note: notation assumes stationarity

What is the principle of maximum entropy (ME)?

- In a probabilistic context, the principle of ME was introduced by Janes (1957) as a generalization of the "principle of insufficient reason" attributed to Bernoulli (1713) or to Laplace (1829).
- In a probabilistic context, the principle of ME is used to infer unknown probabilities from known information.
- In a physical context, a homonymous and relative physical principle determines thermodynamical states.
- The principle postulates that the entropy of a random variable should be at maximum, under some conditions, formulated as constraints, which incorporate the information that is given about this variable.
- Typical constraints used in a probabilistic or physical context are:

Mass — Mean/Momentum — Non-negativity

$$\int_{-\infty}^{\infty} f(x)\, dx = 1, \quad E[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx = \mu \qquad x \geq 0$$

Variance/Energy — Dependence/Stress

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \sigma^2 + \mu^2, \; E[X_i X_{i+1}] = \int_{-\infty}^{\infty} x_i x_{i+1} f(x_i, x_{i+1})\, dx_i\, dx_{i+1} = \rho\, \sigma^2 + \mu^2$$

Application of the ME principle at the basic time scale

- Maximization of either ϕ_n (for any n) or ϕ_c with the mass/mean/variance constraints results in **Gaussian white noise,** with maximized entropy

$$\phi = \phi_c = \ln(\sigma \sqrt{2\pi e}), \quad \phi_n = n\, \phi$$

and information gain $\psi = 0$. This result remains valid even with the non-negativity constraint if variation is low ($\sigma/\mu \ll 1$).

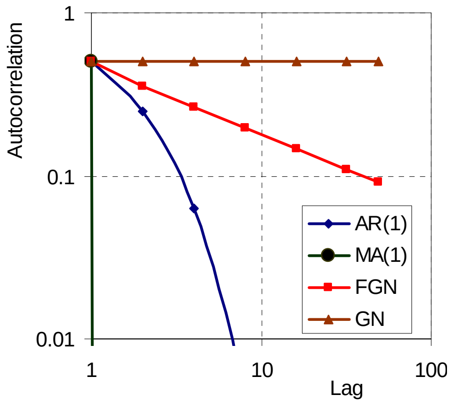
- Maximization of either ϕ_n (for any n) or ϕ_c with the additional constraint of dependence with $\rho > 0$ results in a **Gaussian Markovian process (AR(1))** with maximized entropy

$$\phi = \ln(\sigma \sqrt{2\pi e}), \quad \phi_c = \ln[\sigma \sqrt{2\, \pi\, e\, (1 - \rho^2)}], \quad \phi_n = \phi + (n - 1)\, \phi_c$$

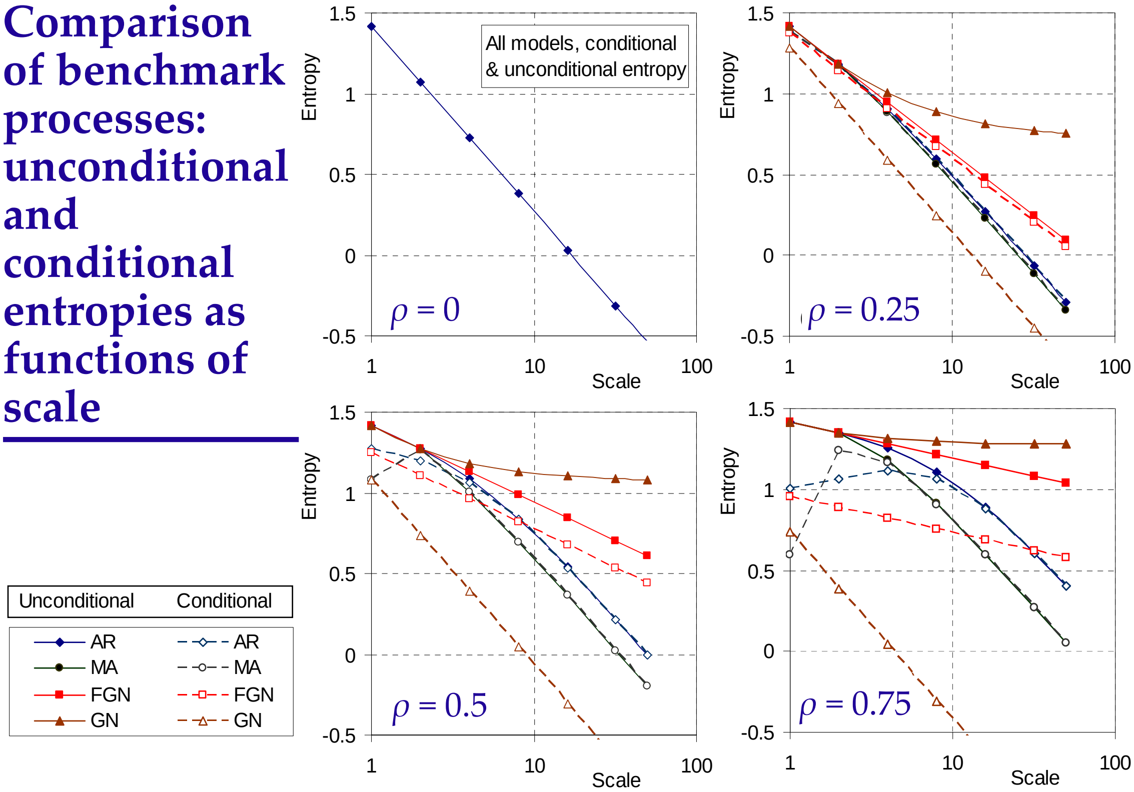
and information gain $\psi = -\ln\sqrt{1 - \rho^2}$.

What happens at other scales? Benchmark processes

- Should maximization be based on a single time scale (annual) and not on other (e.g. multi-annual) time scales?
- How do entopic quantities behave at larger time scales if entropy maximization is done at the basic (annual) time scale?
- First step: demonstration using benchmark processes, all assuming positive autocorrelation function that is a non-increasing function of lag.
  1. Markovian (AR(1)) with exponential decay of autocorrelation, $\rho_j = \rho^j$
  2. Moving average (MA(1) or MA(q) if MA(1) is infeasible) with $\rho_j = 0$ for $j > q$: The minimum autocorrelation structure
  3. Gray noise (GN) with $\rho_j = \rho$: The maximum autocorrelation structure (non-ergodic)
  4. Fractional Gaussian Noise (FGN) with power type decay of autocorrelation, $\rho_j \approx H\,(2\,H - 1)\;|\,j\,|^{2H-2}$



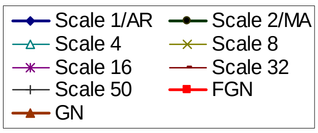
Comparison of benchmark processes: unconditional and conditional entropies as functions of scale

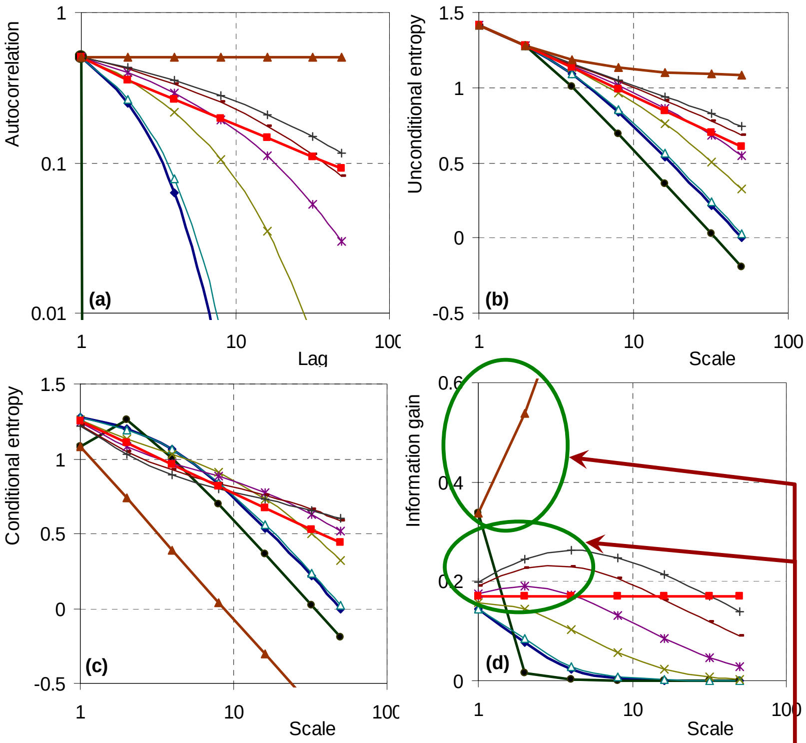


Entropy maximization at larger scales

- All five constrains are used (mass/mean/variance/dependence/non-negativity)
- The lag one autocorrelation (used in the dependence constraint) is determined for the basic (annual) scale but the entropy maximization is done on other scales
- The variation is low ($\sigma/\mu \ll 1$) and thus the process is virtually Gaussian. This is valid for the examined annual and over-annual time scales.
- For a Gaussian process the nth order entropy is given as $\varphi_n = \ln\sqrt{(2\,\pi\, e)^n\,\delta_n}$ where δ_n is the determinant of the autocovariance matrix $c_n := \mathrm{Cov}[X_n, \mathbf{X}_n]$.
- The autocovariance function is assumed unknown to be determined by application of the ME principle. Additional constraints for this are:
  - Mathematical feasibility, i.e. positive definiteness of c_n (positive δ_n)
  - Physical feasibility, i.e. (a) positive autocorrelation function and (b) information gain that is a non-increasing function of time scale (Note: periodicity that may result in negative autocorrelations is not considered here due to annual and over-annual time scales)
- To avoid an extremely large number of unknown autocovariance terms, a parametric expression is used at an initial step, i.e., $\mathrm{Cov}[X_i, X_{i+j}] = \gamma_j = \gamma_0\,(1 + \kappa\,\beta\,|j|^{\alpha})^{-1/\beta}$ with parameters κ, α and β (see details in Koutsoyiannis, 2005b).

Maximization of conditional entropy without the constraint of non-increasing information gain



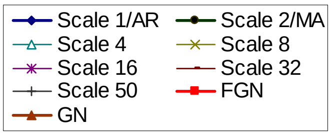


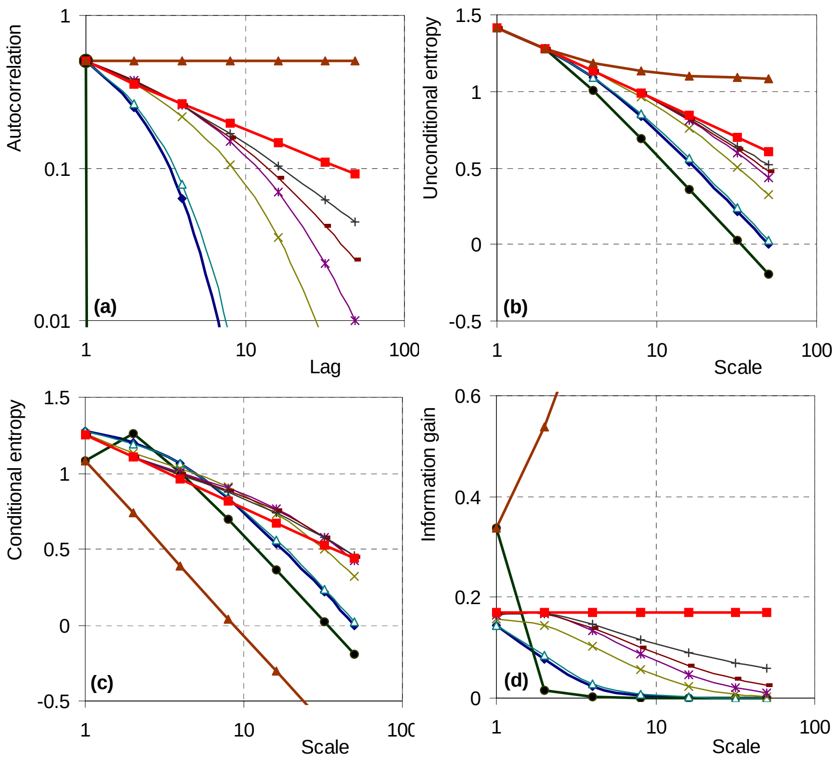
Conclusion:
As time scale increases, the dependence becomes Hurst-like

Increasing information gain for increasing scale → Increased predictability for increasing lead time → Physically unrealistic

Maximization of conditional entropy constrained for non-increasing information gain

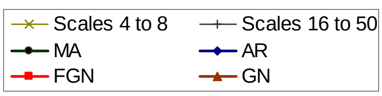
Conclusion:
As time scale increases, the dependence tends to FGN

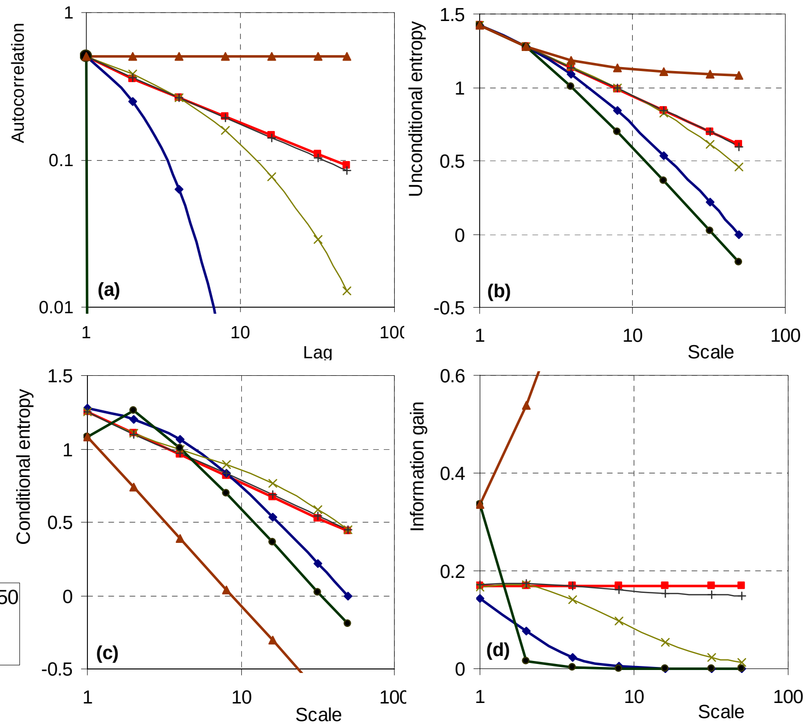




Maximization of unconditional entropy constrained for non-increasing information gain

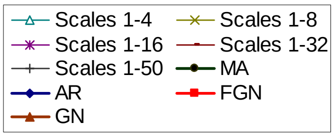
Conclusion:
As time scale increases, the dependence tends to FGN

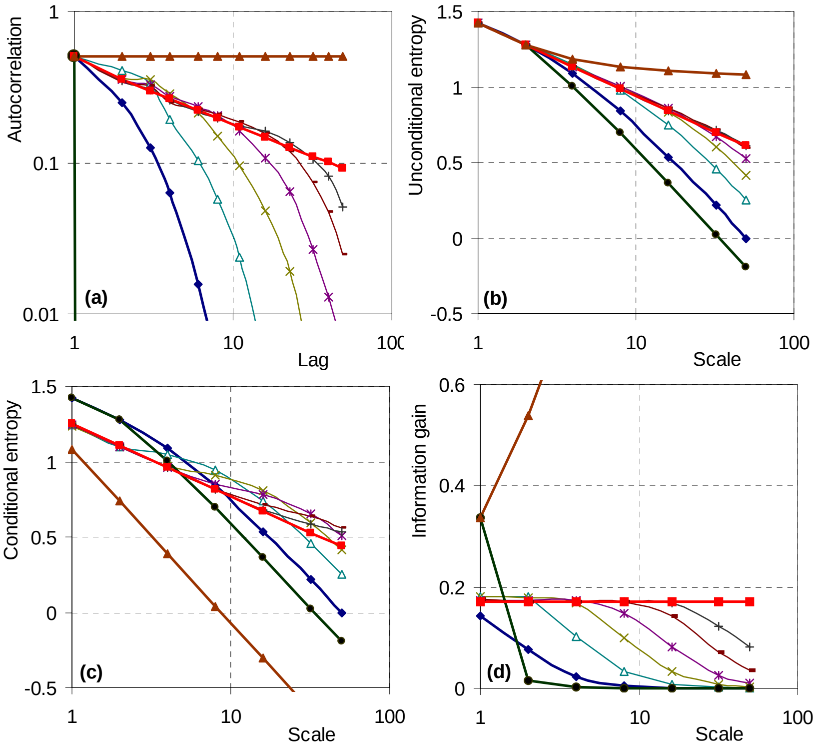




Final step: Maximization of unconditional entropy averaged over ranges of scales, with nonparametric autocovariance

Conclusion:
As the range of time scales widens, the dependence tends to FGN





Conclusions

- Maximum entropy + Low variation → Normal distribution + Time independence
- Maximum entropy + Low variation + Time dependence + Dominance of a single time scale → Normal distribution + Markovian (short-range) time dependence
- Maximum entropy + Low variation + Time dependence + Equal importance of time scales → Normal distribution + Time scaling (long-range dependence / Hurst phenomenon)
- The omnipresence of the time scaling behaviour in numerous long hydrological time series, validates the applicability of the ME principle
- This can be interpreted as dominance of uncertainty in nature.

Discussion

- The ME principle applied at **fine time scales**, where hydrological processes (rainfall, runoff) exhibit **high variation**, explains the power law tails of distribution functions and the **state scaling** at high return periods.
  (See paper in Session P3.01, Scaling and nonlinearity in the hydrological cycle and Koutsoyiannis, 2005a, b)
- It is shown (Papoulis, 1991) that **conditional entropy** equals **entropy rate**, i.e. $\lim_{n\to\infty} \phi_n/n$. Thus, **maximum conditional entropy** could be intuitively related to the physical principle of **maximum entropy production** (according to which the rate of entropy production at thermodynamical systems is at a maximum).
- The latter principle explains the long-term mean properties of the global climate system and those of turbulent fluid systems [*Ozawa et al.*, 2003].
- Specifically, this principle explains
  - the latitudinal distributions of mean air temperature and cloud cover;
  - and the meridional heat transport in the Earth;
  - the behaviour of the planetary atmospheres of Mars and Titan;
  - perhaps, the mantle convection in planets;
  - a variety of aspects of fluid turbulence, including thermal convection and shear turbulence.

References

- Bernoulli, J. (1713), *Ars Conjectandi*, Reprinted in 1968 by Culture et Civilisation, Brussels.
- Jaynes, E.T. (1957), Information Theory and Statistical Mechanics I, *Physical Review*, 106, 620-630.
- Koutsoyiannis, D., (2005a), Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling, *Hydrol. Sci. J.* (in press).
- Koutsoyiannis, D. (2005b), Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling, *Hydrol. Sci. J.*, (in press).
- Laplace, P.S. de, (1829), *Essai philosophique sur les Probabilités*, H. Remy, 5th édition, Bruxelles.
- Ozawa, H., A. Ohmura, R. D. Lorenz, and T. Pujol, The second law of thermodynamics and the global climate system: A review of the maximum entropy production principle, *Rev. Geophys.*, 41(4), 1018, doi:10.1029/2002RG000113, 2003.
- Papoulis, A. (1991) *Probability, Random Variables, and Stochastic Processes* (third edn.), McGraw-Hill, New York, USA.

This presentation is available on line at http://www.itia.ntua.gr/e/docinfo/652/