Dipartimento di Idraulica,Trasporti e Strade
Università di Roma "La Sapienza"
Roma, Italia, 1 Ottobre 2008

Entropy as an explanatory concept and modelling tool in hydrology

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

Presentation available online: http://www.itia.ntua.gr/en/docinfo/879/

How nature works? (my views ...)

Does it ever obey **power laws**?
Does it reflect **fractals** or **multifractals** everywhere?
Does it just reflect **chaos**?
Or is it based on a principle of **self organized criticality** (a cooperative behaviour, where the different items of large systems act together in some concerted way)?

{1, 2, 22}

Nature is parsimonious

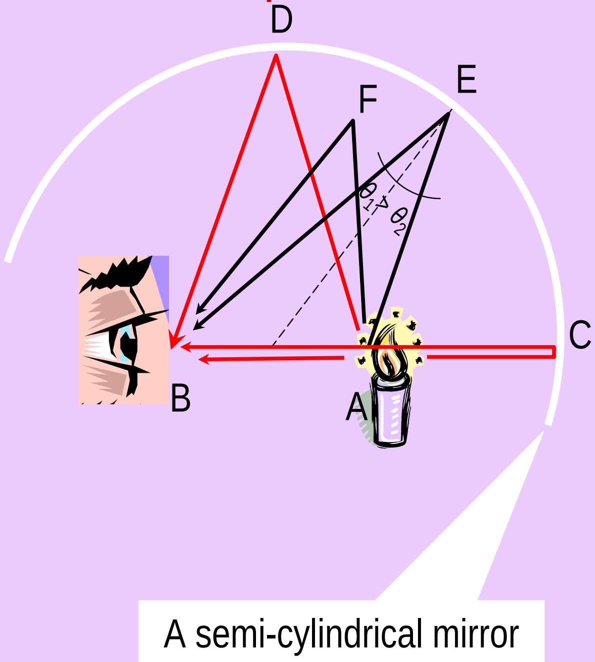
- Example of a parsimonious natural law:
  - Dogs bark.
- Examples of non-parsimonious laws:
  - Black, white and spotted dogs bark.
  - Dogs bark on Mondays, Wednesdays and Fridays.

{3}

Nature conserves a few quantities

- Conservation laws govern the following (macroscopic) quantities:
  - Mass (or matter);
  - Linear momentum;
  - Angular momentum;
  - Energy;
  - Electric charge.
- Other quantities (e.g. temperature, velocity, acceleration) are not conserved.
- Conservation laws refer to closed systems that do not exchange heat and mass with the environment (in open systems there is no conservation).
- Mathematically, the conservation laws are formulated as equations (scalar for mass, energy and charge, vector for linear and angular momentum).

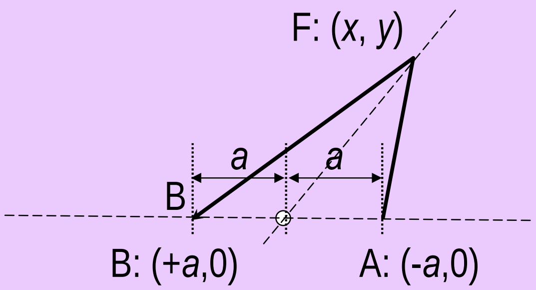
Nature loves extremes: A first example …



Why light follows the red paths from A to B (AB, ACB, ADB) and not other (the black) ones (e.g. AEB, AFB)?

- The red paths are those that (a) reach the mirror and (b) form an angle of incidence equal to the angle of reflection.
  (True for most cases; not true for AB; not general or parsimonious).
- The red paths have minimum travel time (or length).
  (Fermat's principle – Not true for ADB).
- The red paths have extreme (stationary, i.e. minimum or maximum) travel time (or length).
  (True).

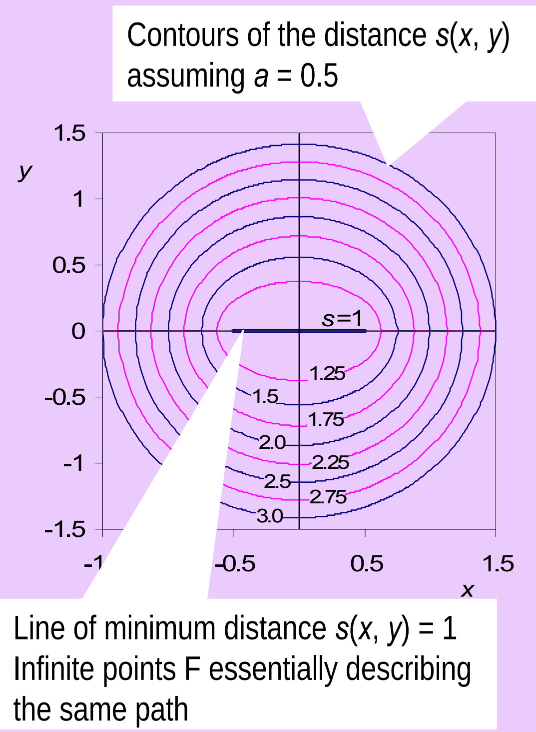
The light example – no mirror



Assume that light can travel from A to B along a broken line with a break point F with coordinates (x, y).
(This is not restrictive: later we can add a second, third, … break points).
The travel distance is $s(x, y)$ = AF + FB where

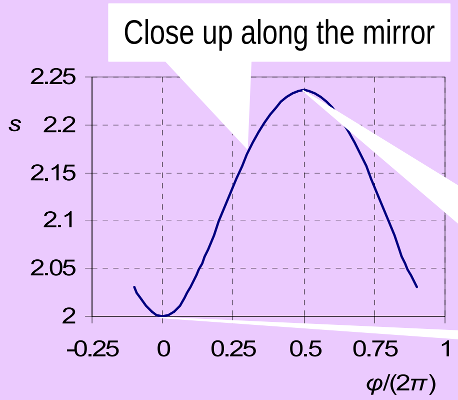
$$AF = \sqrt{(x-a)^2 + y^2}$$

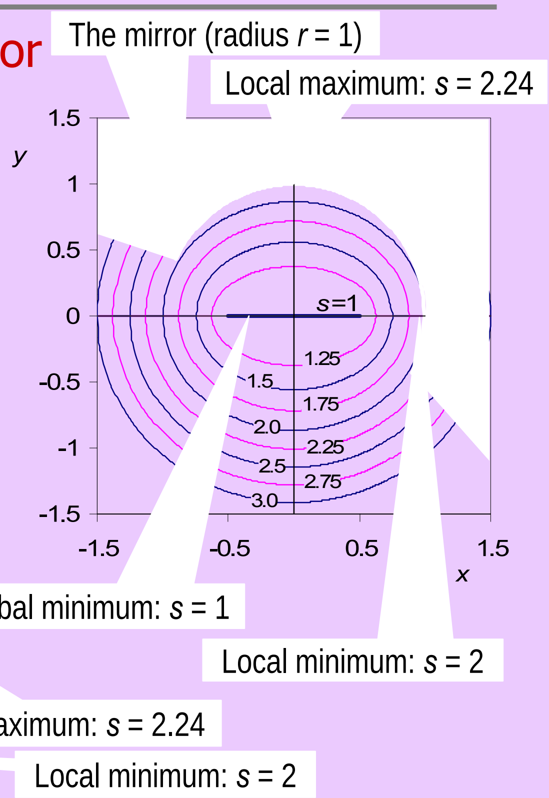
$$FB = \sqrt{(x+a)^2 + y^2}$$



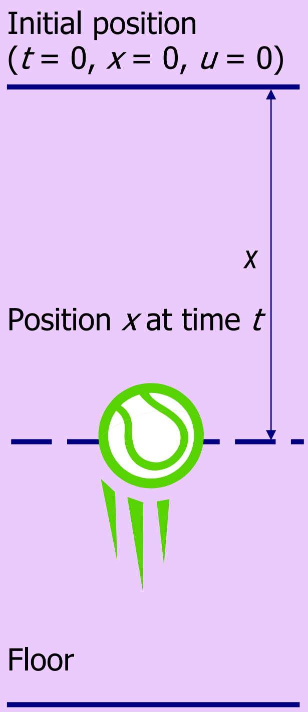
The light example with mirror

- The mirror introduces an inequality constraint in the optimization: the point F should not be behind the mirror.
- Two points of local optima emerge on the mirror surface (the curve where the constraint is binding).





A second example: a falling weight



Quantities involved

- Potential energy: $V = -m\,g\,x$
- Kinetic energy: $T = (1/2)\ m\,u^2 = (1/2)\ m\,(dx/dt)^2$
- Total energy: $E = T + V$
- Lagrangian: $L = T - V = E - 2V$
- Action: $S = \int L\,dt$

Alternative methodologies to find equations for the movement

1. Directly by integrating $d^2x/dt^2 = g$
2. From conservation of total energy
3. From minimization of action (more difficult)

- All methodologies result in same solution ($x = g\,t^2/2$, $u = dx/dt = g\,t$)

Principle of least action (Hamilton's principle – applicable both in classical and in quantum physics)

- From all possible motions between two points, the true motion has least action.
- More correct to substitute "extreme" (or "stationary") for "least".

{23}

How nature works? (synthesis)

Property	Mathematical formulation
She conserves a few quantities (mass, momentum energy, ….).	One equation per conserved quantity: $g_i(\mathbf{s}) = c_i \quad i = 1, \dots, k$ where c_i constants; **s** the size n vector of state variables ($n \geq k$, sometimes $n = \infty$).
She optimizes a single quantity (dependent on the specific system; difficult to find what this quantity is).	A single "optimation": optimize $f(\mathbf{s})$ [i.e. maximize/minimize $f(\mathbf{s})$] **This is equivalent to many equations** (as many as required to determine **s**) Conversely, many equations can be combined into an "optimation".
She disallows some states (dependent on the specific system; maybe difficult to find).	Inequality constraints: $h_j(\mathbf{s}) \geq 0, \quad j = 1, \dots, m$

- In conclusion, we may find how nature works solving the problem:

$$\begin{aligned} &\text{optimize } f(\mathbf{s}) \\ &\text{s.t. } g_i(\mathbf{s}) = c_i \quad i = 1, \dots, k \\ &\qquad h_j(\mathbf{s}) \geq 0 \quad j = 1, \dots, m \end{aligned}$$

The typical "optimizable" quantity in complex systems …

- … is entropy – entropie – Entropie – entropia – entropía – entropi – entrópia – entroopia – entropija – энтропия – ентропія – 熵 – エントロピー – مقياس – אנטרופיה– εντροπία.
- The word is ancient Greek (εντροπία, a feminine noun meaning: turning into; turning towards someone's position; turning round and round).
- The scientific term is due to Clausius (1850).
- The entropy concept was fundamental to formulate the second law of thermodynamics.
- Boltzmann (1877), then complemented by Gibbs (1948), gave it a statistical mechanical content, showing that entropy of a macroscopical stationary state is proportional to the logarithm of the number w of possible microscopical states that correspond to this macroscopical state.
- Shannon (1948) generalized the mathematical form of entropy and also explored it further. At the same time, Kolmogorov (1957) founded the concept on more mathematical grounds on the basis of the measure theory.

What is entropy?

- Entropy is defined on grounds of probability theory.
- For a discrete random variable X taking values x_j with probability mass function $p_j \equiv p(x_j)$, $j = 1,\dots,w$, the Boltzmann-Gibbs-Shannon (or extensive) entropy is defined as

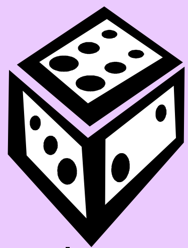
$$\varphi := E[-\ln p(X)] = -\sum_{j=1}^{w} p_j \ln p_j\,, \qquad \text{where } \sum_{j=1}^{w} p_j = 1$$

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

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

- 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/disorder** and **complexity**.
- Generalizations of the entropy definition have been introduced more recently (Renyi, Tsallis).

Entropy maximization: The die example



- What is the probability that the outcome of a toss of a die will be i? ($i = 1, \dots, 6$)
- The entropy is:

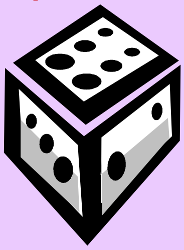
$$\varphi := E[-\ln p(X)] = -p_1 \ln p_1 - p_2 \ln p_2 - \dots - p_6 \ln p_6$$

- The equality constraint (mass preservation) is

$$p_1 + p_2 + \dots + p_6 = 1$$

- The inequality constraint is $p_i \geq 0$.
- Solution of the optimization problem (e.g. by the Lagrange method) yields a single maximum: $p_1 = p_2 = \dots = p_6 = 1/6$.
- This method, the application of the Maximum Entropy Principle (mathematically, an “optimation” form) is equivalent to the Principle of Insufficient Reason (Bernoulli-Laplace; mathematically, an “equation” form).

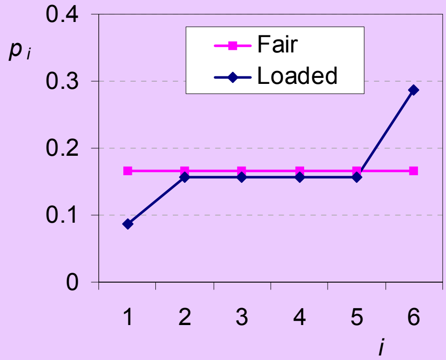
Entropy maximization: The loaded die example



- What is the probability that the outcome of a toss of a die will be i (i = 1, …, 6) if we know that it is loaded, so that $p_6 - p_1 = 0.2$?
- The principle of insufficient reason does not work in this case.
- The maximum entropy principle works. We simply pose an additional constraint:

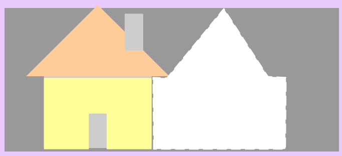
$$p_6 - p_1 = 0.2$$

- The solution of the optimization problem (e.g. by the Lagrange method) is a single maximum:



Entropy maximization: The temperature example

- What will be the temperature in my house (T_H), compared to that of the environment (T_E)? (Assume an open window and no heating equipment).
- Take a space of environment (E) in contact to the house (H) with volume equal to that of the house.



- Partition the continuous range of kinetic energy of molecules into several classes i = 1 (coldest), 2, …, k (hottest).
- Denote p_i the probability that a molecule belongs to class i, and partition it to p_{Hi} and p_{Ei}, if the molecule is in the house or the environment, respectively.
- Form the entropy in terms of p_{Hi} and p_{Ei}.
- Maximize entropy conditional on $p_{Hi} + p_{Ei} = p_i$.
- The result is $p_{Hi} = p_{Ei}$.
- Equal number of molecules of each class are in the house and the environment, so $T_H = T_E$.
- This could be obtained also from the principle of insufficient reason.

Is the principle of maximum entropy ontological or epistemological?

- In thermodynamics and statistical physics the principle of maximum entropy is clearly ontological:
  - It determines (macroscopic thermodynamical) actual states of physical systems.
- Jaynes (1957) introduced the principle of maximum entropy as an epistemological principle in a probabilistic context:
  - It is used to infer unknown probabilities from known information.
  - The (unknown) density function $f(x)$ of a random variable X is the one that maximizes the entropy φ, subject to any known constraints.
- Are these two different principles or one?
- If Nature aligns itself with the (ontological) principle, why not use the same principle in logic for inference about Nature?

Formalization of the principle of maximum entropy

- In both physics and logical inference, the principle of maximum entropy 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

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

- In statistical physics, if X denotes the momentum of molecules or atoms in a gas volume, the mean and variance constraints correspond precisely to the principles of preservation of momentum and energy.

{18, 19}

Typical results of entropy maximization

Constraints	**Result**
$a \le X \le b$	Uniform distribution, $f(x) = 1 / (b - a)$
$X \ge 0$, fixed mean μ	Exponential distribution, $f(x) = \exp(-\lambda_0 - \lambda_1 x)$ or $f(x) = (1/\mu) \exp(-x/\mu)$
Fixed mean μ and standard deviation σ	Normal distribution, $f(x) = \exp(-\lambda_0 - \lambda_1 x - \lambda_2 x^2)$ or $f(x) = (2\pi\,\sigma)^{-1/2} \exp\{(-1/2)[(x - \mu)/\sigma]^2\}$
$X \ge 0$, fixed μ and σ	Truncated normal distribution $f(x) = \exp(-\lambda_0 - \lambda_1 x - \lambda_2 x^2)$
Two variables, X, Y with fixed μ_X, σ_X, μ_Y, σ_Y and ρ_{XY} (correlation)	Bivariate normal distribution $f(x, y) = \exp(-\lambda_0 - \lambda_1 x - \lambda_2 x^2 - \lambda_3 y - \lambda_4 y^2 - \lambda_5 xy)$
Many variables X_i with fixed μ, σ, and ρ_1 (lag 1 autocorrelation)	Multivariate normal distribution Markovian dependence

Note: In all cases with $X \ge 0$, the above solutions exist only if $\sigma/\mu \le 1$.

Application 1: The distribution of hydrological variables – An example for rainfall at a fine time scale

Step 1

- Let X_i denote the rainfall rate at time i discretized at a fine time scale (tending to zero).
- What we definitely know about X_i is $X_i \geq 0$.
- Maximization of entropy with only this condition is not possible.
- Now let us assume that rainfall has a specified mean μ.
- Maximization of entropy with constraints

$$X_i \geq 0, \; E[X] = \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x = \mu$$

  results in the exponential distribution: $f(x) = \exp(-x/\mu)/\mu$.

- In addition, let us assume that there is some time dependence of X_i, quantified by $E[X_i\, X_{i+1}] = \gamma$; this will introduce an additional constraint for the multivariate distribution

$$E[X_i\, X_{i+1}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_i\, x_{i+1}\; f(x_i\,, x_{i+1})\; \mathrm{d}x_i\, \mathrm{d}x_{i+1} = \gamma = \rho\,\sigma^2 + \mu^2$$

  (for the exponential distribution $\sigma = \mu$ and thus $\gamma = \rho\,\sigma^2 + \mu^2 = (\rho + 1)\,\mu^2 > \mu^2$).

- Entropy maximization in multivariate setting will result in Markovian dependence.

{14}

Step 2

- The constant mean constraint in rainfall modelling does not result from a natural principle – as for instance in the physics of an ideal gas, where it represents the preservation of momentum.
- Although it is reasonable to assume a specific mean rainfall, we can allow this to vary in time.
- In this case we can assume that the mean at time i is the realization of a random process M_i which has mean μ and lag 1 autocorrelation $\rho^M > \rho$.
- Application of the maximum entropy principle will produce that M_i is Markovian with exponential distribution.
- Then application of conditional distribution algebra results in

$$f(x) = 2\, K_0(2\,(x/\mu)^{1/2})/\mu, \quad F(x) = 1 - 2\,(x/\mu)^{1/2}\, K_1(2\,(x/\mu)^{1/2})/\mu$$

  where $K_n(x)$ is the modified Bessel function of the second kind (important observation: $f(0) = \infty$, whereas in the exponential distribution $f(0) = \mu < \infty$).

- The moments of this distribution are $E[X^n] = \mu^n\, n!^2$ (note: in exponential distribution $E[X^n] = \mu^n\, n!$) so that

$$E[X] = \mu, \quad \mathrm{Var}[X] = 3\,\mu^2 \;\; \rightarrow \;\; C_V = \sigma/\mu = \sqrt{3} > 1$$

- The dependence structure becomes more complex than Markovian (difficult to find an analytical solution).

Step 3

- Proceeding in a similar manner as in step 2, we can now replace the constant mean μ of the process M_i with a varying mean, represented by another stochastic process N_i with mean μ and lag 1 autocorrelation $\rho^N > \rho^M > \rho$.
- In this manner we can construct a chain of processes, each member of which represents the mean of the previous process.
- By construction, the lag 1 autocorrelations of these processes form a monotonically increasing sequence, i.e. …. $> \rho^N > \rho^M > \rho$.
- The scale of change or fluctuation of each process of the chain is a monotonically increasing sequence, i.e. …. $> q^N > q^M > q$, where $q := (-\ln \rho)^{-1}$; the scale of fluctuation represents the time required for the process to decorrelate down to an autocorrelation 1/e.
- The (unconditional) mean of all processes is the same, μ.
- All moments except the first form an increasing sequence as we proceed through the chain; higher moments increase more.
- Analytical handling of the marginal distribution and the dependence structure is very difficult.
- However we can easily inspect the idea using Monte Carlo simulation.

A demonstration using a chain with 3 processes

- Simulation of a Markovian process with exponential distribution is easy and precise; there are several methodologies to implement it.
- Here we implement an Exponential Markov (EM) process as

  $$X_i = \mu\,[-\ln G(Y_i)]$$

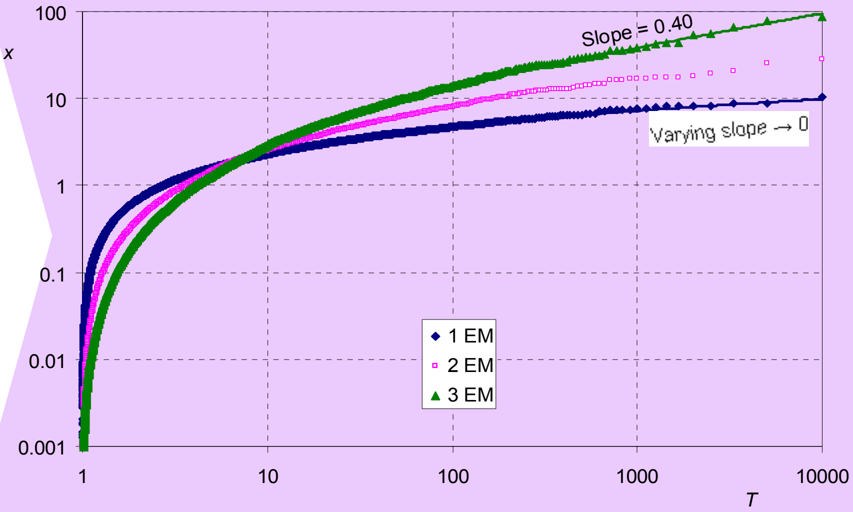
  where μ is the mean, Y_i is a standard AR(1) process with standard normal distribution and $G(\ )$ is the standard normal distribution function.
- Simulations with a length 10 000 were performed for the following cases (for comparison).

	Case	1 EM	2 EM		3 EM		
	Process	X	M	X	N	M	X
Processes in chain	Mean	1	1	-	1	-	-
	Lag 1 autocorrelation*	0.48	0.9	0.25	0.99	0.85	0.2
	Scale of fluctuation	1.37	9.5	0.72	99.5	6.2	0.62
Final process (X)	Mean	1	1		1		
	Standard deviation	1	1.73		3.30		
	Lag 1 autocorrelation	0.48	0.48		0.48		

* Autocorrelation coefficients refer to the standard AR(1) process but are approximately equal in the EM process.

Simulation results – distribution function

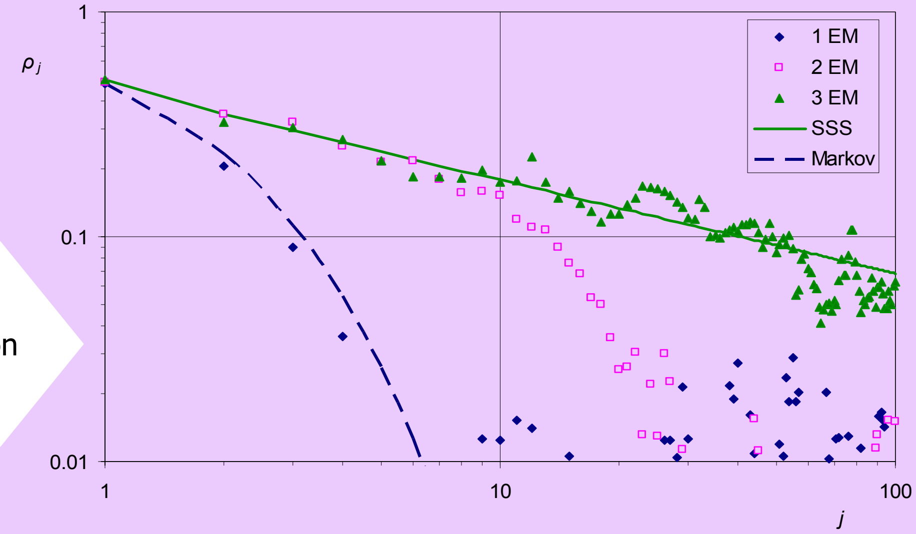
Logarithmic plot of "rainfall intensity" (x) vs. empirically estimated return period $T(x) := 1/F^*(x) = 1/[1 - F(x)]$ where F(x) is the distribution function and $F^*(x)$ the exceedence probability



As the number of processes in the chain increases, the right tail of the distribution moves toward higher "rainfall intensity" values and its shape changes; simultaneously the probability density becomes infinite for $x = 0$.

The probability plot of the "3 EM" case seems to suggest a long tail (a power type law, instead of the exponential type of model "1 EM"), which in a double logarithmic plot is depicted as a constant nonzero slope ($\kappa = 0.40$) of the empirical distribution (or an asymptotic relationship of the form $x \sim T^\kappa$ for large x).

Simulation results – dependence structure



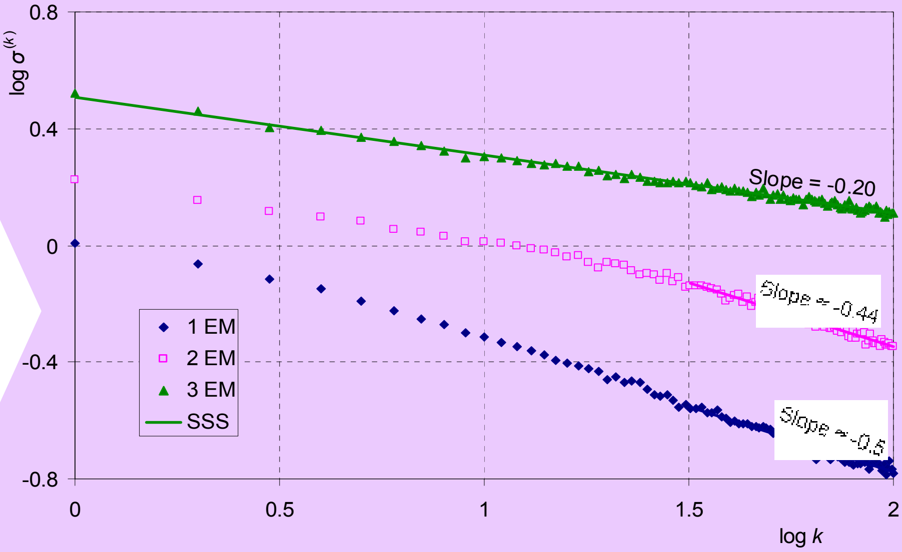
Logarithmic plot of autocorrelation coefficient ρ_j vs. lag j

As the number of processes in the chain increases, the shape of the autocorrelation function changes from Markovian (exponential decay – short range dependence) to power type (long range dependence).

The latter type is characteristic of the Hurst-Kolmogorov behaviour, which can be represented by a simple scaling stochastic process (SSS process).

{6, 7, 12, 15}

Simulation results – variation of the aggregated process



Logarithmic plot of standard deviation $\sigma^{(k)}$ of the process aggregated at scale k, vs. scale k

The slope of the logarithmic plot (as $k\to\infty$) is $H-1$ where H is the Hurst coefficient.

The slope in the "1 EM" case is –0.5, i.e. $H = 0.5$, meaning no Hurst-Kolmogorov behaviour.

The slope in "3 EM" is –0.20, i.e. $H = 0.80$, suggesting a Hurst-Kolmogorov behaviour.

{6, 7, 12, 15}

Simulation results – visual assessment

Plots of parts of the generated time series (selected so as to include the maximum over 10 000 generated values)

As the number of processes in the chain increases the general shape changes:

- From monotony to rich patterns
- From steadiness to intermittency

Can entropy maximization be performed in a single step? (The Tsallis entropy)

- A generalization of the Boltzmann-Gibbs-Shannon entropy has been proposed by Tsallis (1998, 2004)

$$\varphi_q = \frac{1 - \int_{-\infty}^{\infty} [f(x)]^q}{q - 1}$$

with $q = 1$ corresponding to the Boltzmann-Gibbs-Shannon entropy.

- Maximization of Tsallis entropy with known μ yields

$$f(x) = [1 + \kappa(\lambda_0 + \lambda_1 x)]^{-1 - 1/\kappa}, \quad x \geq 0$$

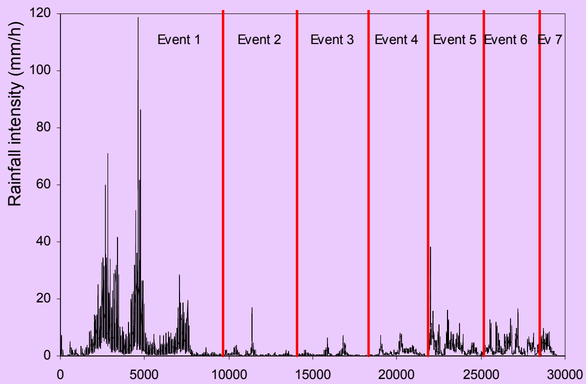
where $\kappa := (1 - q)/q$ and λ_0, λ_1, λ_2 and are parameters.

- Clearly, this is the Pareto distribution and has an over-exponential (power-type) distribution tail.
- Whilst this approach succeeds in producing a long tail to the right, it fails in reproducing the tail to the left (it underpredicts the probability of very low values).
- Furthermore, a single-step approach based on the Tsallis entropy cannot reproduce the Hurst-Kolmogorov behaviour.

{10, 24, 25}

Verification of results based on real world, high resolution rainfall data

Plot of a high resolution (10 s) data set consisting of seven storms occurred in Iowa in 1990-91



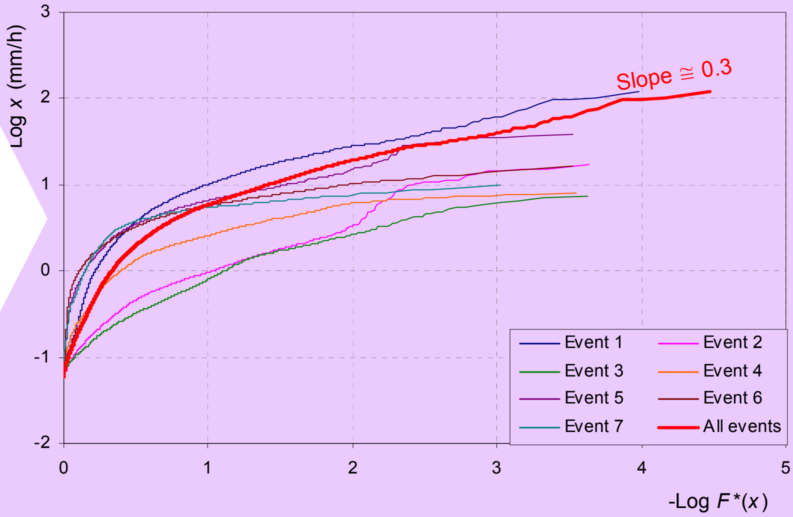
Statistics of the seven storms and the compound record of all storms

Event #	1	2	3	4	5	6	7	All
Sample size	9697	4379	4211	3539	3345	3331	1034	29536
Average (mm/h)	3.89	0.50	0.38	1.14	3.03	2.74	2.70	2.29
St. deviation (mm/h)	6.16	0.97	0.55	1.19	3.39	2.20	2.00	4.11
Coefficient of variation	1.58	1.95	1.45	1.04	1.12	0.81	0.74	1.79
Skewness	4.84	9.23	5.01	2.07	3.95	1.47	0.52	6.54
Kurtosis	47.12	110.24	37.38	5.52	27.34	2.91	-0.59	91.00
Hurst coefficient	0.94	0.76	0.92	0.95	0.90	0.87	0.97	0.94

{16}

Scaling in state

Logarithmic plots of rainfall intensity (x) vs. empirically estimated (by the Weibull formula) exceedence probability, $F^*(x) := 1 - F(x)$, for the seven events

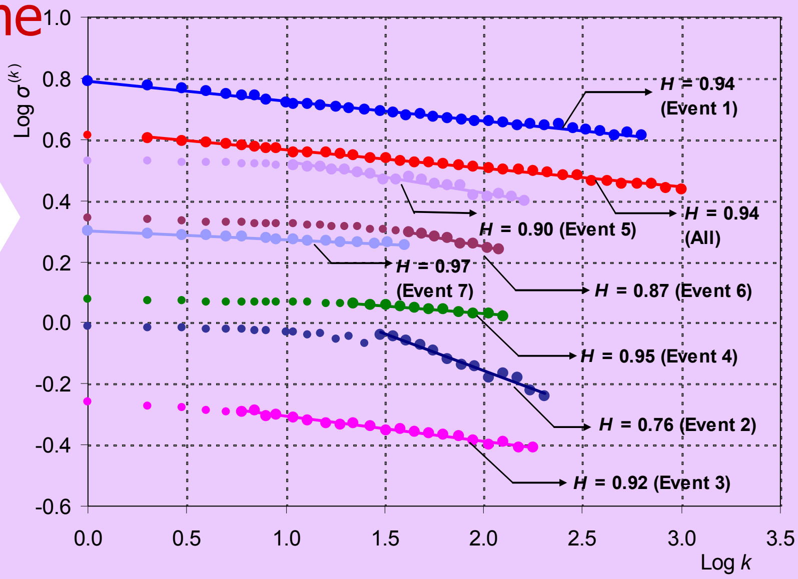


The probability plot of the compound record of all events seems to suggest a long distribution tail.
In five of the seven events (1 to 5 – those with the largest durations) the variation σ/μ is higher than 1, which suggests non applicability of standard entropy theory.

{10, 16}

Scaling in time

Logarithmic plots of standard deviation $\sigma^{(\kappa)}$ of rainfall intensity *vs.* time scale k for the seven events and the compound record



Each of the events separately indicates a Hurst-Kolmogorov behaviour with Hurst coefficient ranging form 0.76 to 0.97.
The Hurst-Kolmogorov behaviour is very clear in the compound record of all events, with Hurst coefficient 0.94.

Verification based on extreme daily rainfall worldwide

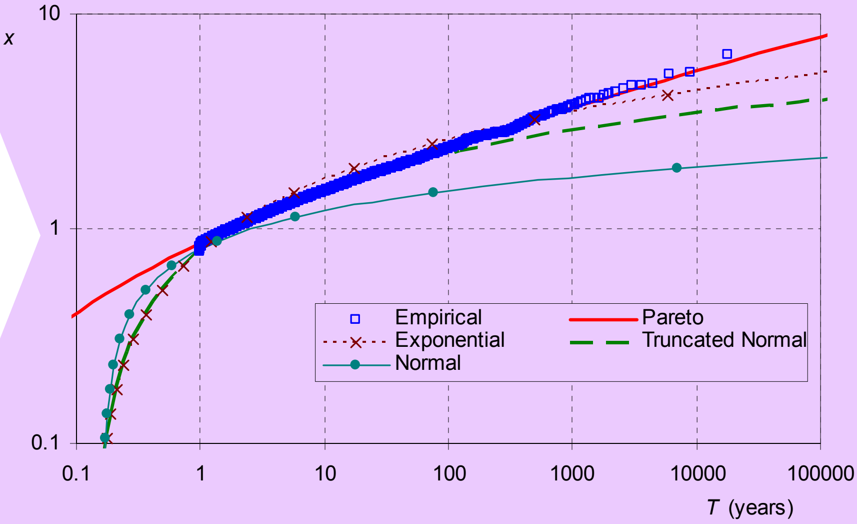
Data set: Daily rainfall from 168 stations worldwide each having at least 100 years of measurements; series above threshold, standardized by mean and unified; period 1822-2002; 17922 station-years of data.

μ = 0.28 (mean minus threshold)

σ/μ = 1.19 > 1

Maximum entropy distribution tail: Pareto, κ = 0.15

Scaling behaviour exist only for T > ~50 yr



{8, 9, 10}

Application 2: Normalizing transformations

General framework

- The normal distribution is very convenient in building a stochastic model.
- The maximum entropy framework can help establish a normalizing transformation which could preserve the distribution behaviour at its tails.
- When only the right tail is of interest, the following transformation (1) can result from application of the result of the Tsallis entropy maximization:

$$z = g(x) - g(0), \quad g(x) = c + \operatorname{sgn}(x-c)\, \lambda \sqrt{\left(1+\frac{1}{\kappa}\right) \ln\left[1+\kappa\left(\frac{x-c}{\lambda}\right)^2\right]}$$

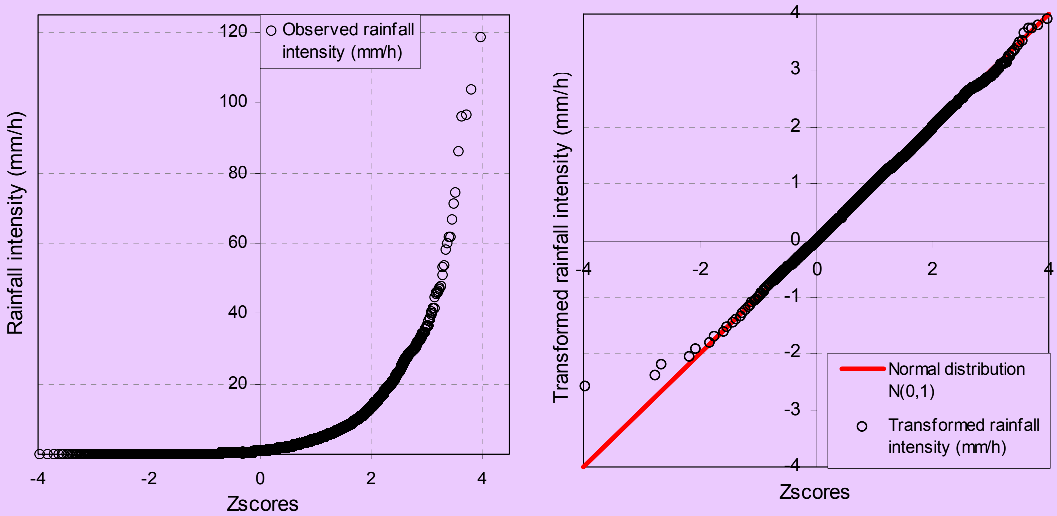
- Here c is a translation parameter with same units as x, κ the tail-determining dimensionless parameter, and λ a scale parameter with same units as x, which enables physical consistency of the transformation. It is easily seen that: (a) z has the same units as x; (b) for x/λ ranging in $[0, \infty)$, z/λ also ranges in $[0, \infty)$; and (c) for $\kappa = 0$, z is identical to x.
- When the right tail (for $x \to 0$) is also of interest, the following modification (2) with additional parameter a (with same unit as x) and v (dimensionless) yields a power-type right tail for $f(x)$ simultaneously infinitizing it for $x \to 0$:

$$g(x) = \left[\left(\frac{x}{a}\right)^{-v} + 1\right]\left\{c + \operatorname{sgn}(x-c)\, \lambda \sqrt{\left(1+\frac{1}{\kappa}\right) \ln\left[1+\kappa\left(\frac{x-c}{\lambda}\right)^2\right]}\right\}$$

{17, 20}

Application to high resolution rainfall

- The transformation (2) effectively transforms the observed data to normal, also implying a power type tail for X.
- The parameters of the transformation can be estimated by minimizing the square error of the model and the empirical distribution function.

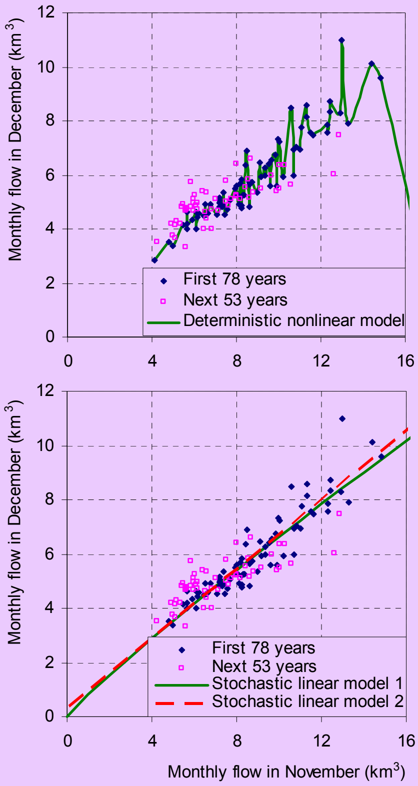


{20}

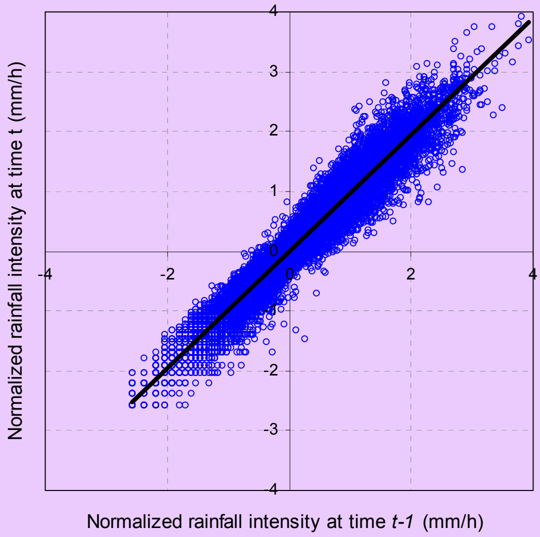
Application 3: Type of dependence of hydrological processes

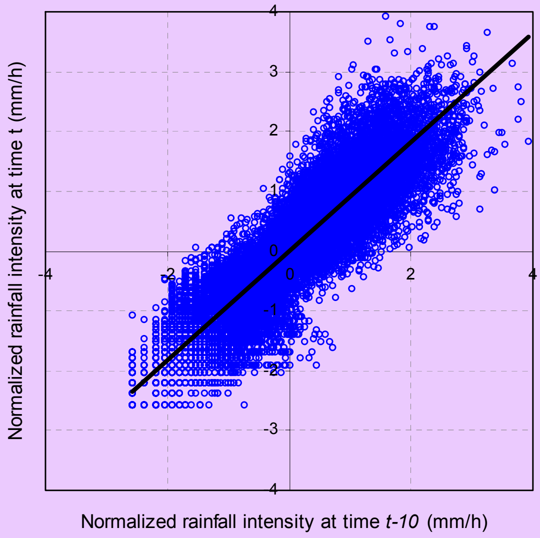
Modelling approaches and underlying concepts

- In a **deterministic approach**, a relationship of any two variables (see the example in figure, referring to lagged flows of the Nile) should be described by an "exact" function which should be a non-intersecting **nonlinear** curve (as in the caricature case shown in the figure) passing from all points (e.g. the 78 points of the 'fitting' period – but the points of the 'validation' period lie outside the curve).
- In a **stochastic approach:**
  - The variables are modelled as random variables.
  - There is no need to assume an "exact" relationship.
  - To each variable a normalizing transformation could be applied.
  - Entropy maximization for the transformed two variables simultaneously will result in bivariate normal distribution.
  - Bivariate normal distribution entails a **linear** relationship between the two variables (Linear model 1 in figure).
  - This explains why stochastic linear relationships are so common.
  - Even without normalizing transformation, the dependence between two variables is virtually linear (Linear model 2 in figure).



Verification of linearity in high resolution rainfall





The figures refer to the Iowa high temporal resolution rainfall data set after normalization by transformation 2.

Application 4: Autocorrelation structure of hydrological processes

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:

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

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

$$\varphi_n := E[-\ln f(\mathbf{X}_n)] = -\int_{D_n} f(\mathbf{x}_n) \ln f(\mathbf{x}_n)\, d\mathbf{x}_n$$

- 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):

$$\varphi_{c,m} := E[-\ln f(X_1 | X_0, \ldots, X_{-m+1})] = \varphi_m - \varphi_{m-1}$$

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

$$\varphi_c := \lim_{m \to \infty} \varphi_{c,m}$$

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

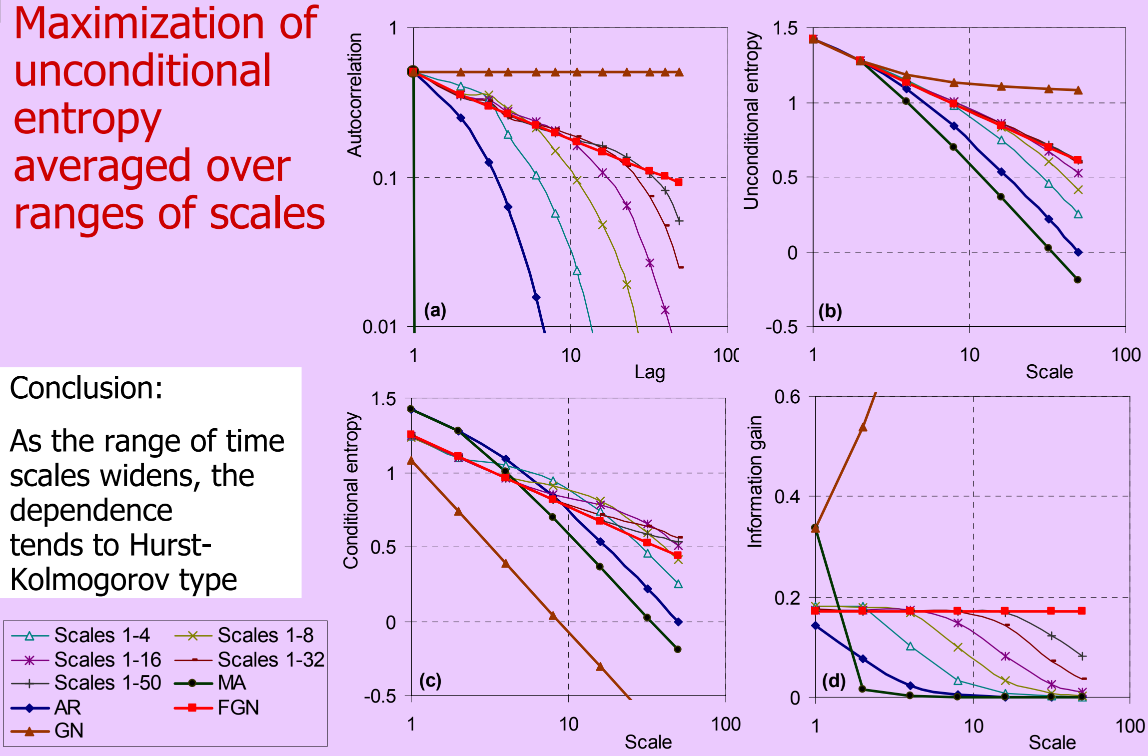
$$\psi := \varphi - \varphi_c$$

Note: notation assumes stationarity.

Entropy maximization for a stochastic process

- The minimum time scale considered is annual (to avoid periodicity).
- The purpose is to determine the dependence structure.
- The typical 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 as well.
- The variation on annual and over-annual scales is low ($\sigma/\mu << 1$) and thus the process can be approximated as Gaussian (except in tails).
- 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 $c_n := \mathrm{Cov}[\mathbf{X}_n, \mathbf{X}_n]$.
- The autocovariance function is assumed unknown to be determined by application of the maximum entropy principle. Additional constraints for this are:
  - Mathematical feasibility, i.e. positive definiteness of c_n (positive δ_n);
  - Physical feasibility, i.e. (a) autocorrelation function positive and (b) information gain not increasing with time scale.
    (Note: periodicity that may result in negative autocorrelations is not considered here due to annual and over-annual time scales).

Demonstration: Maximization of unconditional entropy averaged over ranges of scales



Conclusion:

As the range of time scales widens, the dependence tends to Hurst-Kolmogorov type

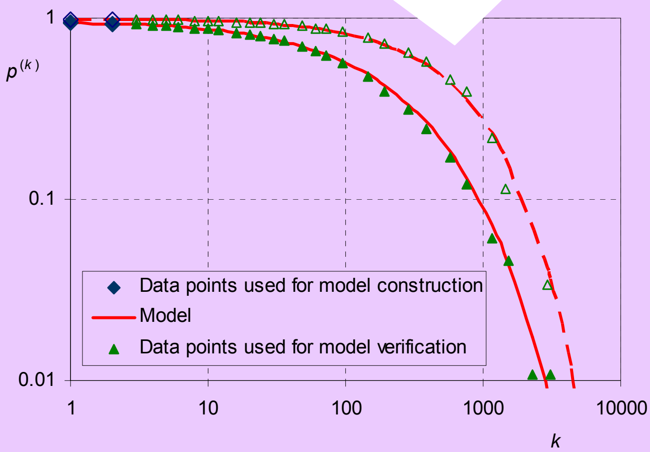
{11}

Application 5: Clustering of rainfall occurrences

Clustering in rainfall explained by maximum entropy

- Rainfall at small scales is intermittent.
- The dependence of the rainfall occurrence process is not Markovian neither scaling but in between; it has been known as clustering or overdispersion.
- The models used for the rainfall occurrence process (point processes) are essentially those describing clustering of stars and galaxies.
- The maximum entropy principle, applied to the binary state rainfall process in more or less the same way as in the continuous state process, explains this dependence.

Probability $p^{(k)}$ that an interval of k hours is dry, as estimated from the Athens rainfall data set and predicted by the model of maximum entropy for the entire year (full triangles and full line) and the dry season (empty triangles and dashed line)



Application 6: Parsimonious stochastic modelling for stochastic prediction

Stochastic model formalism

- The problem of the prediction of the monthly Nile flow is studied.
- The prediction W of the monthly flow one month ahead, conditional on a number s of other variables with known values that compose the vector $\boldsymbol{Z}$, is based on the linear model:

$$W = \boldsymbol{a}^T \boldsymbol{Z} + V$$

  where $\boldsymbol{a}$ is a vector of parameters (the superscript T denotes the transpose of a vector or matrix) and V is the prediction error, assumed independent of $\boldsymbol{Z}$; for simplicity, all elements of $\boldsymbol{Z}$ are assumed normalized and standardized with zero mean and unit variance.
- For the model to take account of both long-range and short-range dependence, an optimal composition of $\boldsymbol{Z}$ was found to be the following:
  - All available flow measurements of the same month on previous years (78 variables = monthly flows for each of the 78 years of the calibration period).
  - The flows of the two previous months of the same year (2 variables).
- The model parameters are estimated from (Koutsoyiannis, 2000)

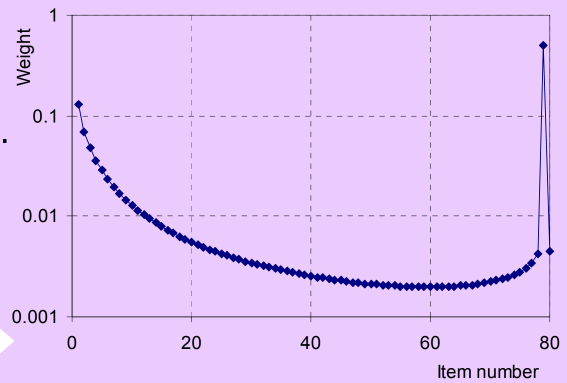
$$\boldsymbol{a}^T = \boldsymbol{\eta}^T \boldsymbol{h}^{-1}, \quad \mathrm{Var}[V] = 1 - \boldsymbol{\eta}^T \boldsymbol{h}^{-1} \boldsymbol{\eta} = 1 - \mathbf{a}^T \boldsymbol{\eta}$$

  where $\boldsymbol{\eta} := \mathrm{Cov}[W, \boldsymbol{Z}]$ and $\boldsymbol{h} := \mathrm{Cov}[\boldsymbol{Z}, \boldsymbol{Z}]$.
- In forecast mode, $V = 0$ (to obtain the expected value of W conditional on $\boldsymbol{Z} = \boldsymbol{z}$); in simulation mode V is generated from the normal distribution independently of $\boldsymbol{Z}$.

11. Parameter estimation

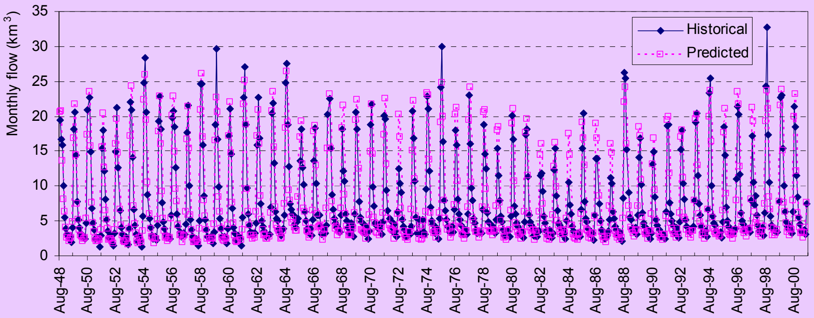
- Both $\boldsymbol{a}$ and Var[V] are estimated from the vector $\boldsymbol{\eta} := \mathrm{Cov}[W, \boldsymbol{Z}]$ and the matrix $\boldsymbol{h} := \mathrm{Cov}[\boldsymbol{Z}, \boldsymbol{Z}]$ that contain numerous items (in our case 80 + 80 × 80 = 6480 for each month; such a number of parameters cannot be estimated from 78 monthly data values).
- However, most covariances in $\boldsymbol{\eta}$ and $\boldsymbol{h}$ depend on:
  - 2-3 parameters (same for all months) expressing the long-range dependence, as estimated by application on the maximum entropy principle on a multi-time scale setting (a stationary component);
  - 2 parameters (per month) expressing the monthly autocovariances at the monthly scale (a cyclostationary component).
- All other covariances that cannot be derived from these parameters are left 'unestimated' (in terms of statistics) and are calculated by the maximum entropy principle, applied on a single scale.
- The entropy maximization in this case has an easy analytical solution that can be formulated as a generalized Cholesky decomposition (assuming that $\boldsymbol{h} = \boldsymbol{b}\,\boldsymbol{b}^T$).

Graphical depiction of the vector of weights $\boldsymbol{a}$ estimated by the maximum entropy principle for the month of July



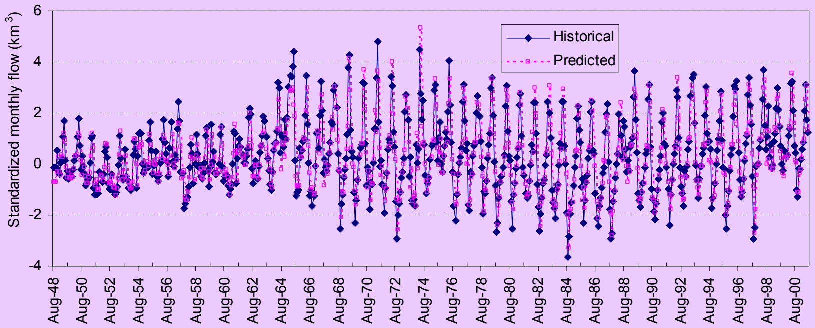
12. Results of stochastic model (validation period)

Natural values



The graphical depiction of monthly predictions in comparison to historical values, indicates good performance of the model.

Standardized values



{17}

Conclusions

- The successful application of the maximum entropy principle in nature offers an explanation for of a plethora of phenomena (e.g. in thermodynamics) and statistical behaviours including:
  - the emergence of normal distribution (independently of the central limit theorem) in some cases;
  - the emergence of the exponential distribution in other cases;
  - the linearity of most stochastic laws (including the time dependence of natural processes);
  - the scaling behaviour in state in cases with high variation (which is only a consequence of the maximum entropy principle for special cases and just an approximation, good for high return periods);
  - the scaling behaviour in time, i.e. the Hurst-Kolmogorov behaviour;
  - the clustering behaviour in rainfall occurrence.
- All these can be interpreted as dominance of uncertainty in nature.
- They harmonize with the Socratic view: «Ἓν οἶδα, ὅτι οὐδέν οἶδα» (One I know, that I know nothing).
- This view was not a confession of modesty – Socrates regarded the knowledge of ignorance as a matter of supremacy.
- In this respect, the knowledge of the dominance of uncertainty can assist to better (stochastic) prediction of natural processes as well as in safer design and management of hydrosystems.

References

1. Bak, P., *How Nature Works, The Science of Self-Organized Criticality*. Copernicus, Springer-Verlag, New York, USA, 1996.
2. Buchanan, M., *Ubiquity, The Science of History, or Why the World is Simpler Than We Think,* Weidenfeld & Nicolson, London, 2000.
3. Gauch, H. G. Jr., *Scientific Method in Practice*, Cambridge University Press, Cambridge, 2003.
4. Koutsoyiannis, D., Optimal decomposition of covariance matrices for multivariate stochastic models in hydrology, *Water Resources Research*, 35(4), 1219-1229, 1999.
5. Koutsoyiannis, D., A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, *Water Resources Research*, 36(6), 1519-1533, 2000.
6. Koutsoyiannis, D., The Hurst phenomenon and fractional Gaussian noise made easy, *Hydrological Sciences Journal*, 47(4), 573-595, 2002.
7. Koutsoyiannis, D., Climate change, the Hurst phenomenon, and hydrological statistics, *Hydrological Sciences Journal*, 48(1), 3-24, 2003.
8. Koutsoyiannis, D., Statistics of extremes and estimation of extreme rainfall, 1, Theoretical investigation, *Hydrological Sciences Journal*, 49(4), 575-590, 2004.
9. 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.
10. 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.
11. 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.
12. Koutsoyiannis, D., Nonstationarity versus scaling in hydrology, *Journal of Hydrology*, 324, 239–254, 2006.
13. Koutsoyiannis, D., An entropic-stochastic representation of rainfall intermittency: The origin of clustering and persistence, *Water Resources Research*, 42 (1), W01401, 2006.
14. Koutsoyiannis, D., Long tails of marginal distribution and autocorrelation function of rainfall produced by the maximum entropy principle, *European Geosciences Union General Assembly 2008, Geophysical Research Abstracts, Vol. 10*, Vienna, 10751, European Geosciences Union, 2008
15. 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.
16. Koutsoyiannis, D., S.M. Papalexiou, and A. Montanari, Can a simple stochastic model generate a plethora of rainfall patterns? (invited), *The Ultimate Rainmap: Rainmap Achievements and the Future in Broad-Scale Rain Modelling*, Oxford, Engineering and Physical Sciences Research Council, 2007.
17. Koutsoyiannis, D., H. Yao, and A. Georgakakos, Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods, *Hydrological Sciences Journal*, 53 (1), 142–164, 2008.
18. Müller, I., Entropy: a subtle concept in thermodynamics, in *Entropy* (edited by A. Greven, G. Keller & G. Warnecke), Princeton University Press, Princeton, NJ, 2003.
19. Müller, I., Entropy in nonequilibrium, in *Entropy* (edited by A. Greven, G. Keller & G. Warnecke), Princeton University Press, Princeton, NJ 2003.
20. Papalexiou, S.M., A. Montanari, and D. Koutsoyiannis, Scaling properties of fine resolution point rainfall and inferences for its stochastic modelling, *European Geosciences Union General Assembly 2007, Geophysical Research Abstracts, Vol. 9*, Vienna, 11253, European Geosciences Union, 2007.
21. Papoulis, A. (1991) *Probability, Random Variables, and Stochastic Processes* (third edn.), McGraw-Hill, New York, USA.
22. Schroeder, M. R., *Fractals, Chaos, Power Laws: Minutes From an Infinite Paradise*, FreeMan, New York, NY, 1991.
23. Taylor, J. C., Hidden Unity in Nature's Laws, Cambridge, 2001.
24. Tsallis, C., Possible generalization of Boltzmann-Gibbs statistics, *J. Stat. Phys.* 52, 479-487, 1988.
25. Tsallis, C., Nonextensive statistical mechanics: construction and physical interpretation, in *Nonextensive Entropy*, *Interdisciplinary Applications* (edited by M. Gell-Mann & C. Tsallis), Oxford University Press, New York, NY, 2004.