IUGG XXV General Assembly
Earth on the Edge: Science for a Sustainable Planet
Melbourne, Australia, 27 June - 8 July 2011

Session U-09: Do We Really Know the Hydrological Cycle?

A hymn to entropy

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

Presentation available online: itia.ntua.gr/1136/

Entropy is uncertainty quantified: Definitions

- **Entropy** of a continuous random variable $\underline{z}$ (adapted from Papoulis, 1991)

  $\Phi[\underline{z}] := E[-\ln[f(\underline{z})/l(\underline{z})]] = -\int_{-\infty}^{\infty} f(z) \ln [f(z)/l(z)]\, \mathrm{d}z$ [dimensionless]

  where $f(z)$ the probability density function, with $\int_{-\infty}^{\infty} f(z)\, \mathrm{d}z = 1$, and $l(z)$ a Lebesgue density (numerically equal to 1 with dimensions same as in $f(z)$)

- **Entropy production** for the stochastic process $\underline{z}(t)$ in continuous time t (from Koutsoyiannis, 2011)

  $\Phi'[\underline{z}(t)] := \mathrm{d}\Phi[\underline{z}(t)] / \mathrm{d}t$ [units T^{-1}]

- **Entropy production in logarithmic time** (**EPLT**)

  $\varphi_{\mathrm{LT}}[\underline{z}(t)] := \mathrm{d}\Phi[\underline{z}(t)] / \mathrm{d}(\ln t) \equiv \Phi'[\underline{z}(t)]\, t$ [dimensionless]

- Note 1: Starting from a stationary stochastic process $\underline{x}(t)$, the cumulative (nonstationary) process $\underline{z}(t)$ is defined as $\underline{z}(t) := \int_0^t \underline{x}(\tau)\, \mathrm{d}\tau$; consequently, the discrete time process $\underline{x}_i^{\Delta} := \underline{z}(i\Delta) - \underline{z}((i-1)\Delta)$ represents stationary intervals (for time step Δ in discrete time i) of the cumulative process $\underline{z}(t)$

- Note 2: For any specified time t and any two processes $\underline{z}_1(t)$ and $\underline{z}_2(t)$, an inequality relationship between entropy productions, such as $\Phi'[\underline{z}_1(t)] < \Phi'[\underline{z}_2(t)]$ holds also true for EPLTs, i.e. $\varphi_{\mathrm{LT}}[\underline{z}_1(t)] < \varphi_{\mathrm{LT}}[\underline{z}_2(t)]$

The principle of maximum entropy (ME)

- The principle of maximum entropy postulates that the entropy of a random variable $\underline{z}$ should be at maximum, under some conditions, formulated as constraints, which incorporate the information that is given about this variable
- Entropy maximization of a random variable $\underline{z}$ bounded within $[0, b]$:
  - uniform, $f(z) = 1/b$
- Entropy maximization of a random variable $\underline{z}$ unbounded from both below and above:
  - No constraint or constrained mean μ: not defined
  - Constrained mean μ and standard deviation σ:
    Gaussian, $f(z) = \exp\{-[(z-\mu)/\sigma]^2/2\} / [\surd(2\pi)\sigma]$
- Entropy maximization of a nonnegative variable unbounded from above:
  - No constraint: not defined
  - Constrained mean μ: exponential, $f(z) = (1/\mu)\exp(-z/\mu)$
  - Constrained mean μ and standard deviation σ:
    - if $\sigma < \mu$: truncated Gaussian, $f(z) = \exp\{-[(z-\mu)/\sigma]^2/2\} / [\surd(2\pi)\sigma]$ tending to exponential as $\sigma \to \mu$ (or $\sigma/\mu \to 1$ from below)
    - if $\sigma > \mu$: not defined

Some general questions and answers

- Why the probability for each outcome of a **die** is 1/6?
  - Because this maximizes entropy (the uniform distribution maximizes uncertainty)
- Why in a particle's motion the energy is equally distributed among the different degrees of freedom (the **equipartition** principle)?
  - Because this maximizes entropy (see Appendix)
- Why the **temperature** in this room is fairly uniform?
  - Because this maximizes entropy (see Appendix)
- Why in the **composition of Earth's atmosphere**, nitrogen and oxygen are present and hydrogen is absent, while in planets far from the Sun hydrogen is present?
  - Because, to maximize entropy, the kinetic energy is equally distributed among different molecules; hence, hydrogen, which has molecular mass lower than oxygen and nitrogen, moves faster (~4 times) and escapes to space, while nitrogen and oxygen cannot reach the escape velocity
  - Planets far from the sun have lower temperature, which is proportional to the kinetic energy, and thus hydrogen cannot reach the escape velocity

Some general questions and answers (cont'd)

- Is the **law of ideal gases** ($pV = nRT$) an empirical relationship or can it be deduced and how?
  - It can be easily derived by maximizing entropy (see Appendix)
- Is the **specific heat** (heat capacity per unit mass) of a gas an experimental quantity or can it be derived theoretically?
  - The entropy maximization framework can derive heat capacities theoretically (see Appendix, where the theoretical value of the specific heat at constant pressure of the dry atmospheric air mixture (assuming that it is composed of diatomic molecules) is derived as $c_p = 1004.8 \text{ J K}^{-1} \text{ kg}^{-1}$, while the experimental value is $c_p = 1004 \text{ J K}^{-1} \text{ kg}^{-1}$)
- Why the temperature in a vertical cross section across the troposphere varies substantially (decreases with increasing elevation), while the entropy per unit mass is fairly uniform (an **isentropic** state)?
  - Because this maximizes entropy (see Appendix)
- What determines how much **water is evaporated** and **condensed** (cf. the Clausius-Clapeyron equation), thus providing the physical basis of the hydrological cycle?
  - A combination of entropy maximization with energy availability

Statistical *vs.* classical thermodynamics

- In classical thermodynamics, first we define temperature T, as well as a basic unit for it (kelvins) and then entropy S by $\mathrm{d}S := \mathrm{d}Q \,/\, T$, where Q denotes heat; the definition is perhaps affected by circularity, as it is valid for reversible processes, which are those in which $\mathrm{d}S = \mathrm{d}Q \,/\, T$, while irreversible are those in which $\mathrm{d}S > \mathrm{d}Q \,/\, T$
- In statistical thermodynamics, the entropy Φ is just the uncertainty, as defined in probability theory, and is dimensionless; the temperature is then defined by $1/\vartheta := \partial\Phi/\partial U$, where U is the internal energy of the system; the natural unit for temperature ϑ is then identical with that of energy (joules)
- Statistical thermodynamics provides explanations for the processes described by classical thermodynamics
- Classical thermodynamic equations can be derived from statistical thermodynamics by simple linear transformations: $S = k\,\Phi$ and $T = \vartheta/k$ where $k = 0.138\,065$ yJ/K is the Boltzmann's constant
- The essential difference is in interpretation:
  - Awareness of maximum uncertainty in statistical thermodynamics
  - Delusion of deterministic laws in classical thermodynamics

Extremizing *vs*. equating

- The traditional approach to physics is based on writing equations, which express conservation laws; these govern the following quantities only:
  - Mass (scalar equation)
  - Linear momentum (vector equation)
  - Angular momentum (vector equation)
  - Energy (scalar equation)
  - Electric charge (scalar equation)
- Conservation equations are not enough to describe processes, in which change occurs; in processes, Nature extremizes certain quantities, such as:
  - Time for the path of light (cf. Hero's and Fermat's principles)
  - Action for the motion of simple systems
  - Entropy for complex systems in equilibrium
  - Entropy production for complex systems evolving in time
- Mathematically, extremizing is much more powerful than equating:
  - A system of equations "$\boldsymbol{g}(\boldsymbol{s}) = \mathbf{0}$" can work if the number of equations equals the number of unknowns
  - A single extremizing expression like "maximize $f(\boldsymbol{s})$" can work irrespective of the number of unknowns (it is equivalent to as many equations as needed)
- Nature is an **extremizer**—not an **equalizer**

Uncertainty *vs.* certainty

- Thermodynamic equations, e.g. the ideal gas law $pV = nRT$, are statistical laws (relationships of expectations of random variables)
- While these laws are derived by maximizing entropy, i.e. uncertainty, they express near certainties and are commonly misinterpreted as deterministic laws
- The explanation of near certainty relies on these two facts:
  - Typical thermodynamic systems are composed of hugely **many identical** elements: $N \sim 10^{24}$ per kilogram of mass
  - The random motion of each of the system elements is practically **independent** to the others'
- As a consequence, a random variable $\underline{x}$ expressing a macroscopic state, will have a variation $\mathrm{std}[\underline{x}]/E[\underline{x}] \sim 1/\sqrt{N} \sim 10^{-12}$ (for a kilogram of mass)
- The fact that the macroscopic variability is practically zero should not mislead us to interpret the laws in deterministic terms

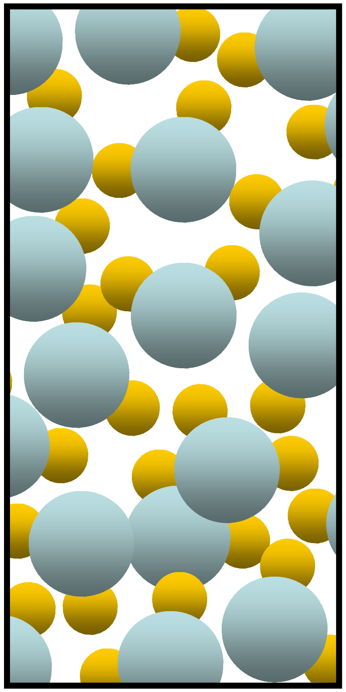
Typical thermodynamic *vs*. hydrological systems

A rain drop is a typical thermodynamic system with identical elements

However, rain drops are not identical to each other and their motion is affected by turbulence (photo from a monsoon in India)

As the macroscopization level increases, the diversity of elements becomes more prominent (notice the different colours in the confluent rivers)

Macroscopization level 1 (typical thermodynamic system)



Macroscopization level 2

Macroscopization level 3

- Identical elements
- Constraints derived from conservation principles
- Deduction: possible

- Different elements, each one being unique
- Constraints not related to conservation principles
- Deduction: impossible; Induction (based on data): necessary
- Yet maximization of entropy, i.e., uncertainty, should work

From statistical thermodynamics of systems with identical elements to hydrological systems

- Higher level of macroscopization is associated with higher complexity
- Therefore, the probabilistic description is more imperative in the more complex hydrological systems
- Maximization of entropy (i.e., uncertainty) provides the path to deal with the complex macroscopic hydrological systems
- This contrasts the recent research trend in hydrology, which invested hopes to the power of computers that would enable faithful and detailed representation of the diverse system elements and the hydrological processes, based on merely "first principles", thus resulting in "physically-based" models that tend to approach in complexity the real world systems
- The aspiration of detailed and exact modelling traces a research direction that is wrong and opposite to how Nature works

Expected difficulties in entropy maximization in high-level macroscopic hydrological systems

- The constraints in entropy extremization do not necessarily coincide with those in classical statistical thermophysics
  - In particular, the mean μ and variance σ^2 are important indices of the statistical behaviour (see Koutsoyiannis, 2005) with a intuitive conceptual meaning, but they are not constrained by physical laws as in the kinetic theory of gases—rather they are estimated from data
- Independence among different elements and across time is most often invalidated
  - This, combined with the diverse elements entails that all laws should remain probabilistic
  - High-level macroscopic quantities in hydrological systems will **never** approach the near certainty of low-level macroscopic quantities in typical thermodynamic systems—regardless of progress in computers and algorithms
- **Physically-based hydrological models are necessarily stochastic models**

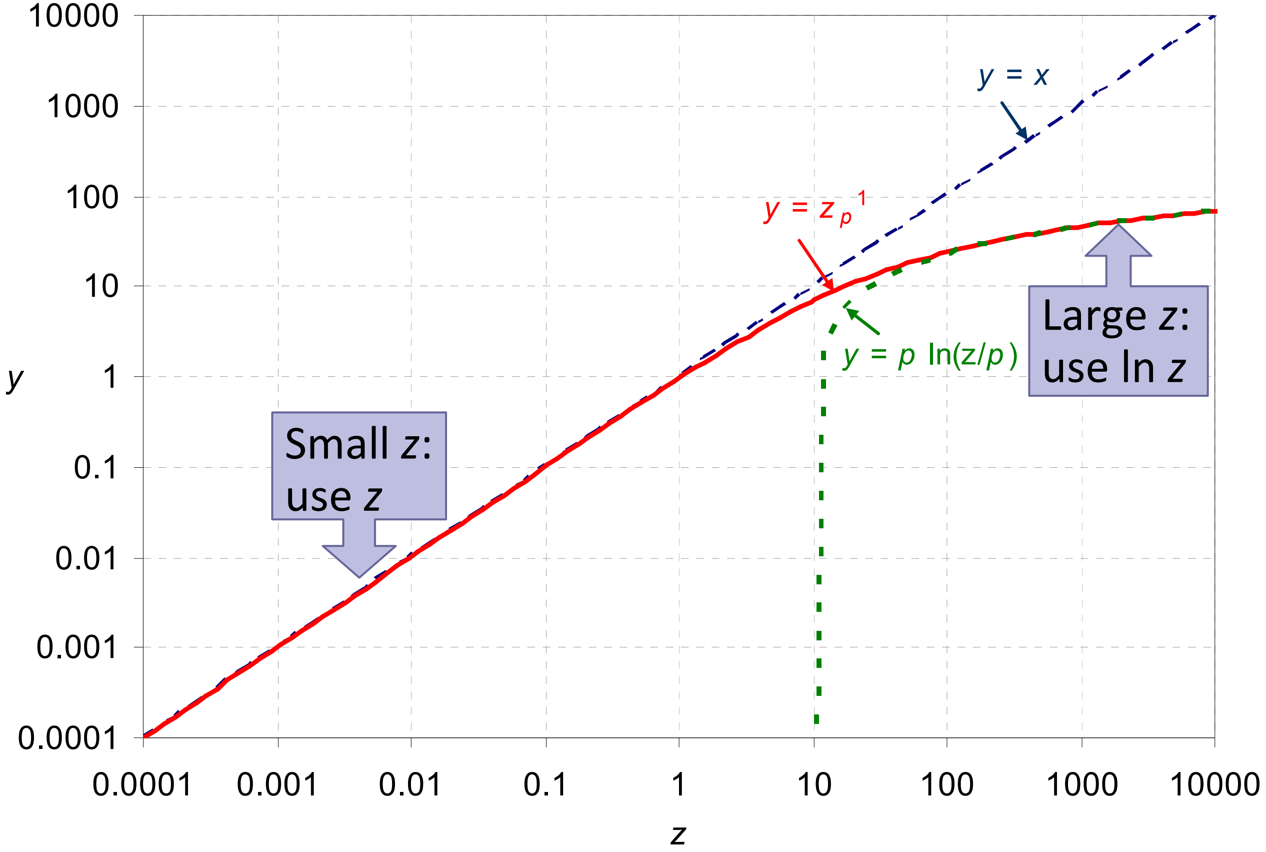
Towards reasonable constraints for positive random variables

- Most hydrometeorological variables are non-negative physical quantities unbounded from above; examples: precipitation, streamflow, temperature (expressed in kelvins or in joules)
- What is a meaningful *distance* for a positive physical quantity, e.g. rainfall depth? (Note: a distance is typically used to measure model errors)
    - Should 0.1 and 0.2 mm, on the one hand, be equidistant with 100.0 and 100.1 mm, on the other hand (having a Euclidean distance of 0.1 mm)?
    - Should 0.1 and 0.2 mm, on the one hand, be equidistant with 100 and 200 mm, on the other hand (having a logarithmic distance of ln 2)?
- We heuristically define the function $h(z) := p \ln(1 + z/p)$ where p is a characteristic scale parameter and we define a meaningful distance between any two points z and z' as $d(z, z') := |h(z) - h(z')|$
    - For small values, i.e., $z < z' << p$, $d(x, x') = p \ln [1 + (z' - z)/(p + z)] \approx z' - z$ (Euclidean distance)
    - For large values, $p << z < z'$, $d(z, z') \approx p \ln (z'/z)$ (logarithmic distance)
    - Note that $h(z)$ and $d(z, z')$ have always the same units as z (physical consistency)
    - Such distance has been sometimes used to describe model errors, e.g. Koutsoyiannis and Onof (2002)

Towards workable constraints in highly variable geophysical processes

- The function $h(z)$ enables a generalization of the classical power moments, thus defining the p-moments (Papalexiou and Koutsoyiannis, 2011)
- A p-moment is defined to be the expectation $E[\underline{z}_p^q]$, where $z_p^q := p^q \ln[1 + (z/p)^q]$ and p is a scale parameter with units identical to those of z
- For $p \to \infty$, or for finite p and small z, $z_p^q \approx z^q$ (the classical raise to power q)
- For $p \to 0$ or for finite p and large z, $z_p^q \propto [\ln(z/p)]^q$

An example plot of z_p^q for p = 10 and q = 1



Small z: use z

Large z: use ln z

Simplest case—a single constraint

- The simplest constraint is formed by setting the exponent $q = 1$, so that we get a "generalized mean", i.e.:

  $$E[\underline{z}_p^{\,1}] = E[p \ln(1 + z/p)] = m_p$$

- The entropy maximizing distribution (derived by the general methodology in Papoulis, 1991, p. 571) is

  $$f(z) = A \exp\,[-\lambda_1 p \ln(1 + z/p)] = A\,(1 + z/p)^{-\lambda_1 p}$$

  where λ_1 is a Lagrange multiplier and A is such that $\int_{-\infty}^{\infty} f(z)\,\mathrm{d}z = 1$

- By renaming parameters ($p = \lambda/\kappa$, $\lambda_1 = (1 + \kappa)/\lambda$) we obtain the typical expression of the 2-parameter Pareto distribution

  $$f(z) = (1/\lambda)\,(1 + \kappa\, z/\lambda)^{-1 - 1/\kappa}$$

  with mean $\mu = \lambda/(1 - \kappa)$, standard deviation $\sigma = \lambda/[(1 - \kappa)\sqrt{(1 - 2\kappa)}]$, generalized mean $m_p = \lambda$ and entropy $\Phi[\underline{z}] = 1 + \kappa + \ln \lambda$

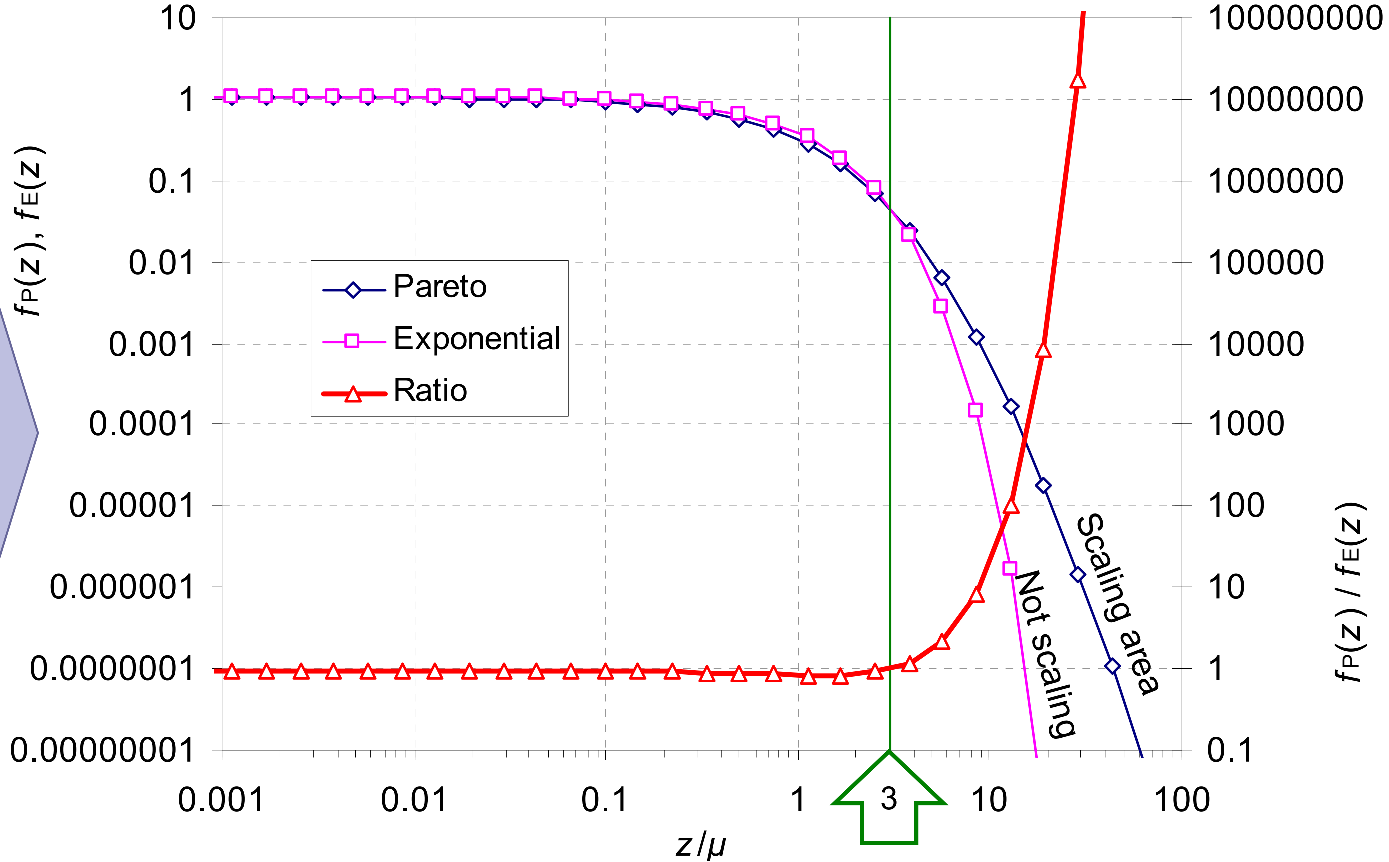
- The exponential distribution is fully recovered by setting $\kappa = 0$ ($p = \infty$); its statistics are $\mu = \sigma = \lambda$, $m_p = p \exp(\lambda p)\, \Gamma_{\lambda p}(0)/\lambda^2$, and $\Phi[\underline{z}] = 1 + \ln \lambda$

- In the Pareto distribution $\sigma/\mu = 1/\sqrt{(1 - 2\kappa)} > 1$, while in the exponential distribution $\sigma/\mu = 1$

Enhanced uncertainty in comparison to classical thermodynamics

- In classical thermodynamics, a constrained mean results in exponential distribution
- The two density functions plotted, Pareto, $f_{\mathrm{P}}(z)$, with $\kappa = 0.15$ and $\lambda_{\mathrm{P}} = 0.9$ and exponential, $f_{\mathrm{E}}(z)$, with $\lambda_{\mathrm{E}} = 0.953$ have same $m_p = 0.9$ for $p = \lambda_{\mathrm{P}}/\kappa = 6$
- Their means are $\mu_{\mathrm{P}} = 1.059 > \mu_{\mathrm{E}} = 0.953$ and their entropies are $\Phi_{\mathrm{P}} = 1.045 > \Phi_{\mathrm{E}} = 0.952$

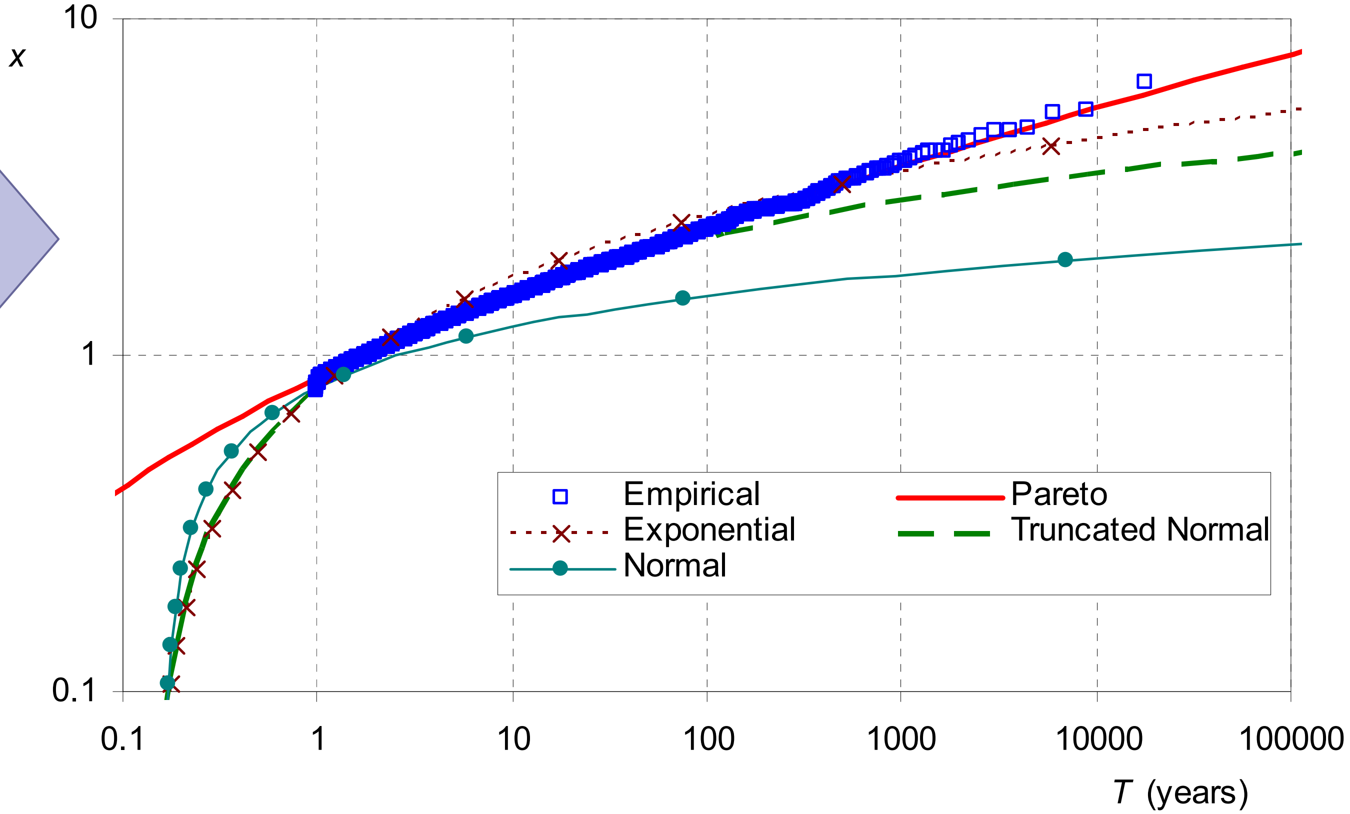
While the two distributions are almost indistinguishable for $z < 3\mu$, the Pareto distribution gives extremes orders of magnitude more often than the non-scaling exponential distribution



Validation based on intense daily rainfall worldwide

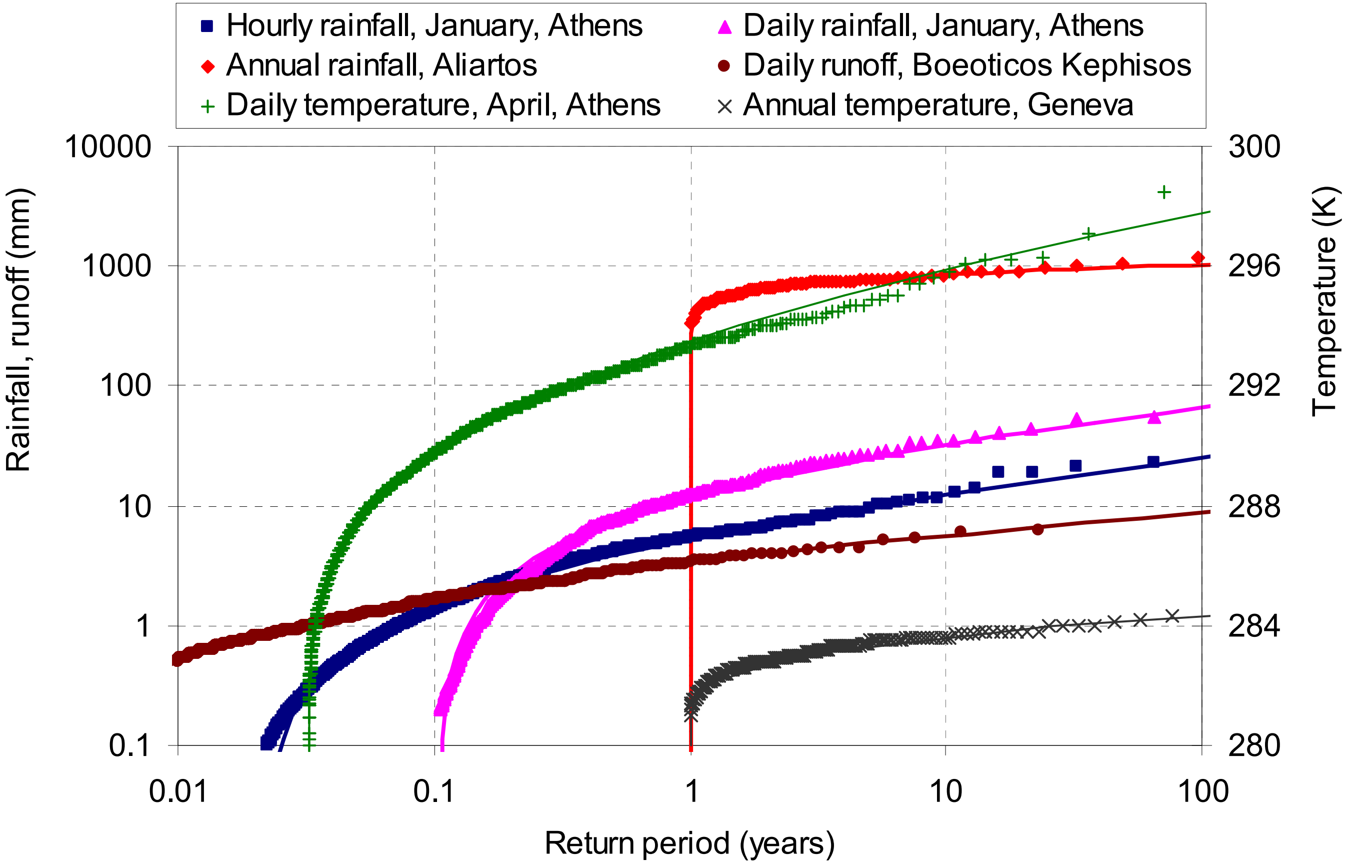
Data set: Daily rainfall from 168 stations worldwide each having at least 100 years of measurements; series above threshold, standardized by mean and unified; period 1822-2002; 17922 station-years of data (Koutsoyiannis, 2005)

$\mu = 0.28$
(mean minus threshold)
$\sigma/\mu = 1.19 > 1$
ME distribution: Pareto, $\kappa = 0.15$



Generalization for the marginal distribution of any hydrometeorological variable

- When the variation is small ($\sigma/\mu < 1$, where μ is the mean and σ the standard deviation), the typical ME framework with constraints μ and σ, gives satisfactory results (although p-moment constraints are again superior)
- Generally, the ME framework can give the shape of the distribution function based on a single metric, σ/μ (Koutsoyiannis, 2005)

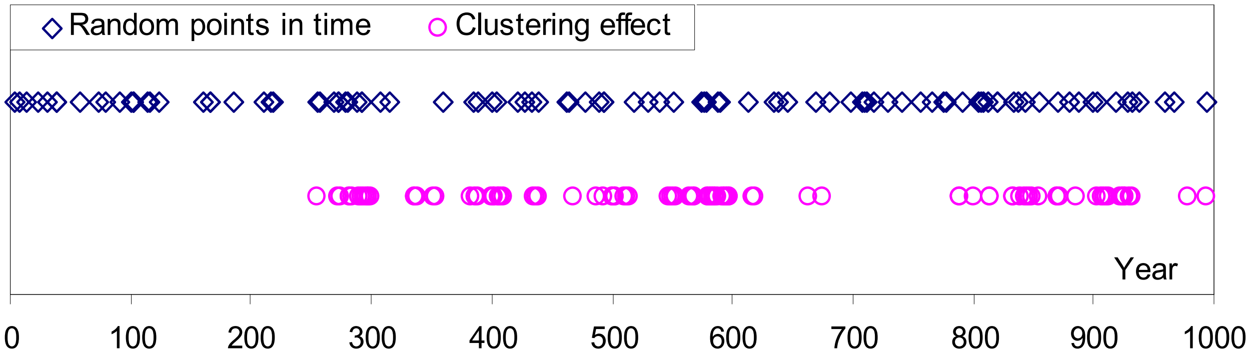


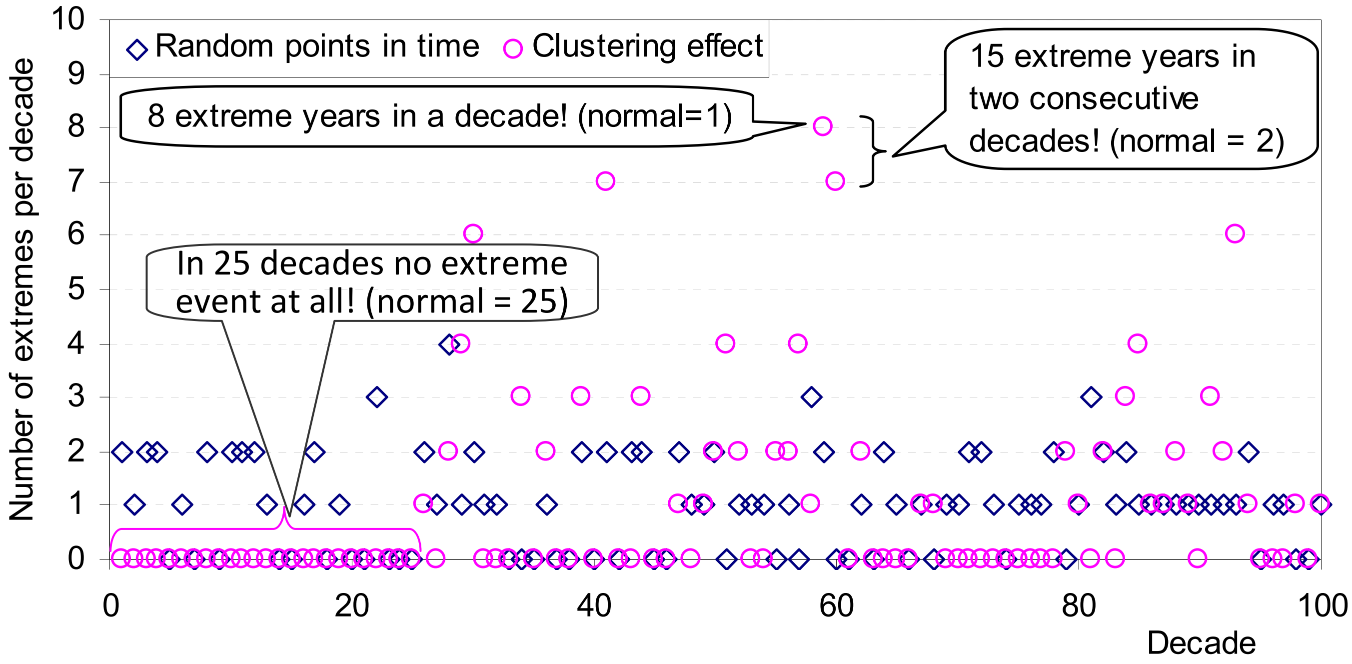
Points: empirical distributions; lines: maximum entropy distributions; see details in Koutsoyiannis (2005)

Involving time: Entropy and clustering

Two simulated series of 100 “extreme” events each, in a period of 1000 “years”; the probability of an extreme event is p =1/10 and the entropies of the two series are equal:
$\Phi_C = \Phi_R = \ln(10)/10 + (9/10)\ln(10/9) = 0.33$

However, if we view the series at a decadal time scale, the entropy of the clustered series is higher: the entropy estimates, considering the probabilities of all possible numbers of extreme events (from 0 to 10), are $\Phi_C = 1.29 > \Phi_R = 1.23$





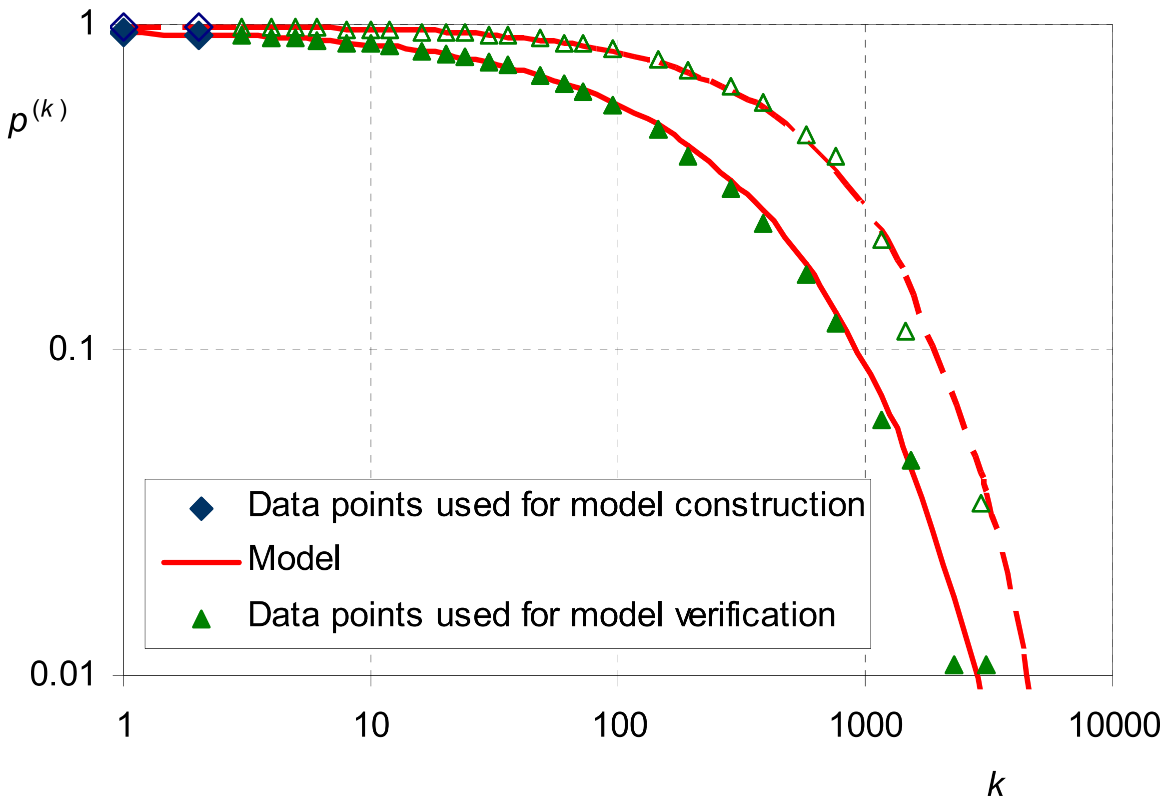
8 extreme years in a decade! (normal=1)

15 extreme years in two consecutive decades! (normal = 2)

In 25 decades no extreme event at all! (normal = 25)

Maximum entropy and clustering of rainfall occurrence

- Rainfall occurrence is characterized by a clustering behaviour (also known as overdispersion)
- The observed behaviour can be explained by maximizing, for a range of scales, the entropy of the binary-state rainfall process using two constraints representing the observed occurrence probabilities at two time scales (1 and 2 hours)
- Entropy maximization with only two parameters determined from the data (those defining the constraints) give very good predictions for all time scales
- The ME solution is not Markovian neither scaling but in between



Probability $p^{(k)}$ that an interval of k hours is dry, as estimated from the Athens rainfall data set (70 years) 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); see details in Koutsoyiannis (2006)

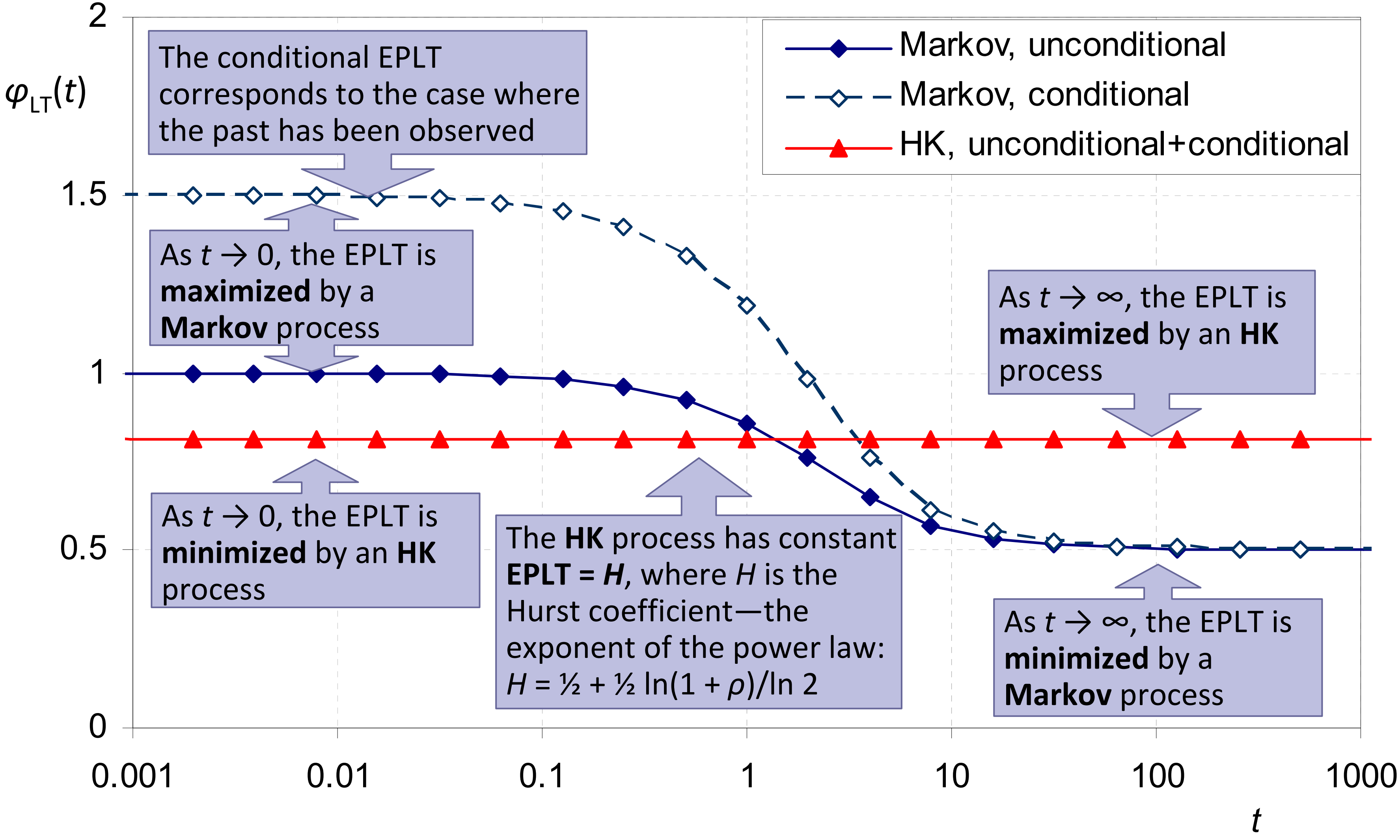
Maximum entropy production and scaling in time

- The dependence structure of processes evolving in time (expressed in terms of autocorrelogram, periodogram or climacogram, which are transformations of one another), can be determined by entropy extremization
- Koutsoyiannis (2011) suggested the use of entropy production (the dimensionless EPLT in particular) with the assumptions of:
    - constrained mean μ and variance σ^2, which result in Gaussian marginal distribution (assumption good for $\sigma/\mu \ll 1$); in this case we have:

      $\Phi[\underline{z}(t)] = (1/2)\ln[2\pi e\,\gamma(t)]$ where $\gamma(t) := \mathrm{Var}\,[\underline{z}(t)]$ (see slide 2)
    - constrained lag-one autocorrelation ρ
- These constraints are formulated for a single observation time scale but the extremization of entropy production is made at asymptotic time scales, i.e., $t \to 0$ and $t \to \infty$
- Such extremization of entropy production yields two simple solutions:
    - A (non-scaling) Markov process (the AR(1) process in discrete time, or the Ornstein–Uhlenbeck process in continuous time)
    - A (scaling) Hurst-Kolmogorov (HK) process (due to Hurst, 1951, and Kolmogorov, 1940)

The two EPLT extremizing solutions

The solutions depicted are generic, valid for any Gaussian process, independent of μ and σ, and depended on ρ only (the example is for $\rho = 0.543$)—see Koutsoyiannis (2011)



The conditional EPLT corresponds to the case where the past has been observed

As $t \to 0$, the EPLT is **maximized** by a **Markov** process

As $t \to \infty$, the EPLT is **maximized** by an **HK** process

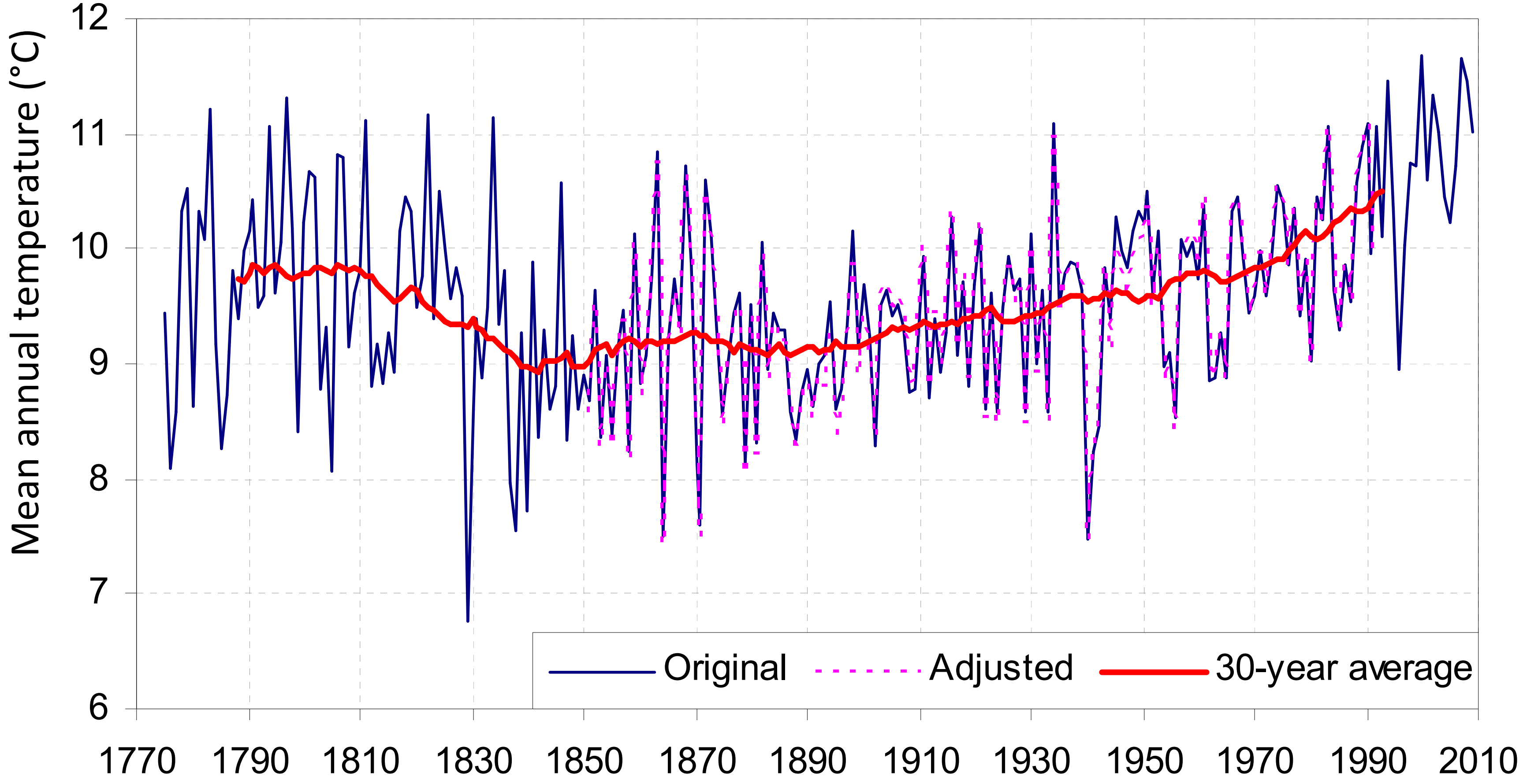
As $t \to 0$, the EPLT is **minimized** by an **HK** process

The **HK** process has constant **EPLT = *H***, where H is the Hurst coefficient—the exponent of the power law: $H = ½ + ½ \ln(1 + \rho)/\ln 2$

As $t \to \infty$, the EPLT is **minimized** by a **Markov** process

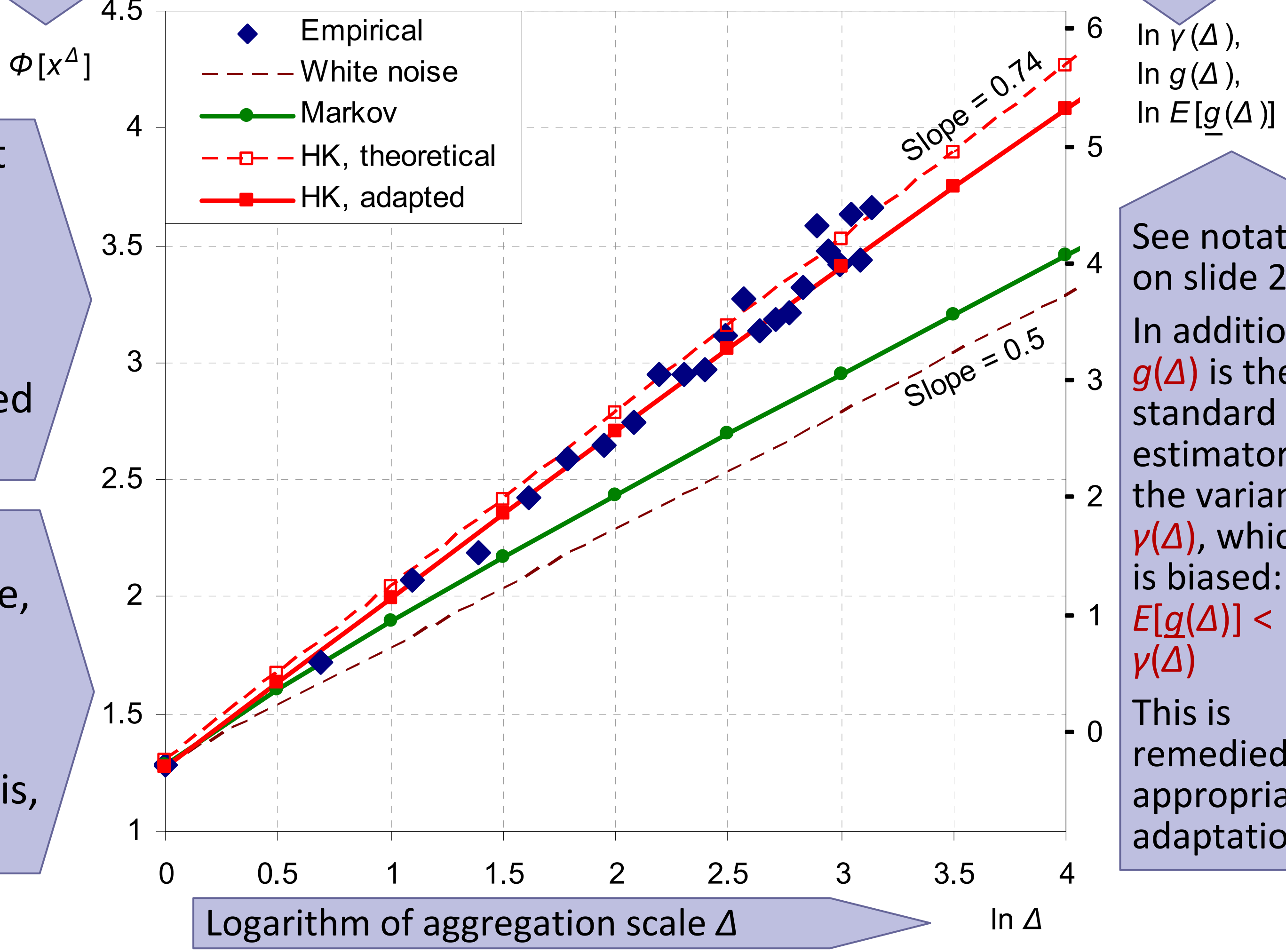
Application to the annual temperature of Vienna

Mean annual temperature of Vienna, Austria (48.25° N, 16.37° E, 209 m): one of the longest available instrumental geophysical records—235 years of data (1775–2009) available from the climexp.knmi.nl, partly included in the Global Historical Climatology Network (GHCN; 1851–1991); from Koutsoyiannis (2011)



Comparison of the Markov and HK models: Vienna temperature

One-to-one correspondence (linear relationship) between entropy $\Phi[x^{\Delta}]$ and logarithm of variance $\gamma(\Delta)$



The low coefficient of variation (σ/μ = 0.0031 for temperature in K), suggests Gaussian distribution (verified by the data)

The HK model (H = 0.74) is appropriate, while the Markov model (ρ = 0.3) is inappropriate (from Koutsoyiannis, 2011)

Logarithm of aggregation scale Δ

See notation on slide 2

In addition, $g(\Delta)$ is the standard estimator of the variance $\gamma(\Delta)$, which is biased: $E[\underline{g}(\Delta)] < \gamma(\Delta)$

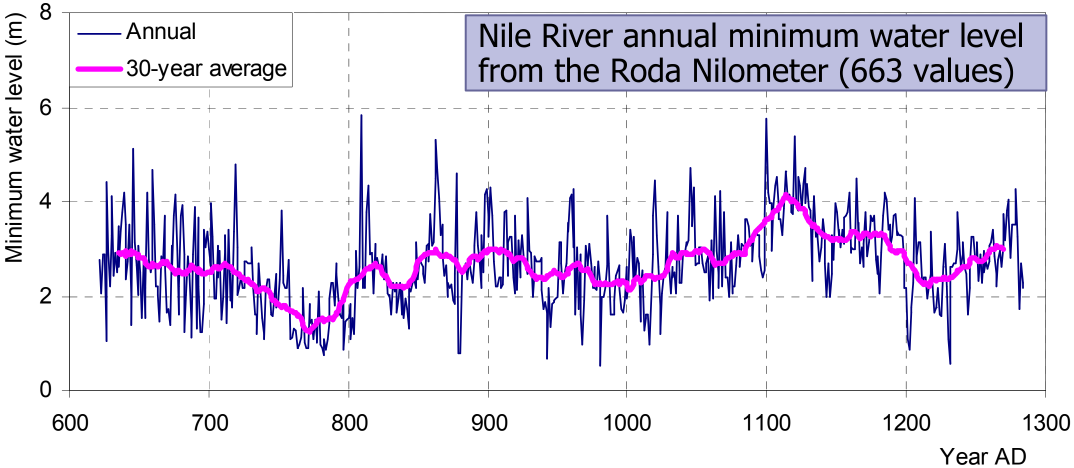
This is remedied by appropriate adaptation

Comparison of the Markov and HK models: Annual minimum water level of the Nile

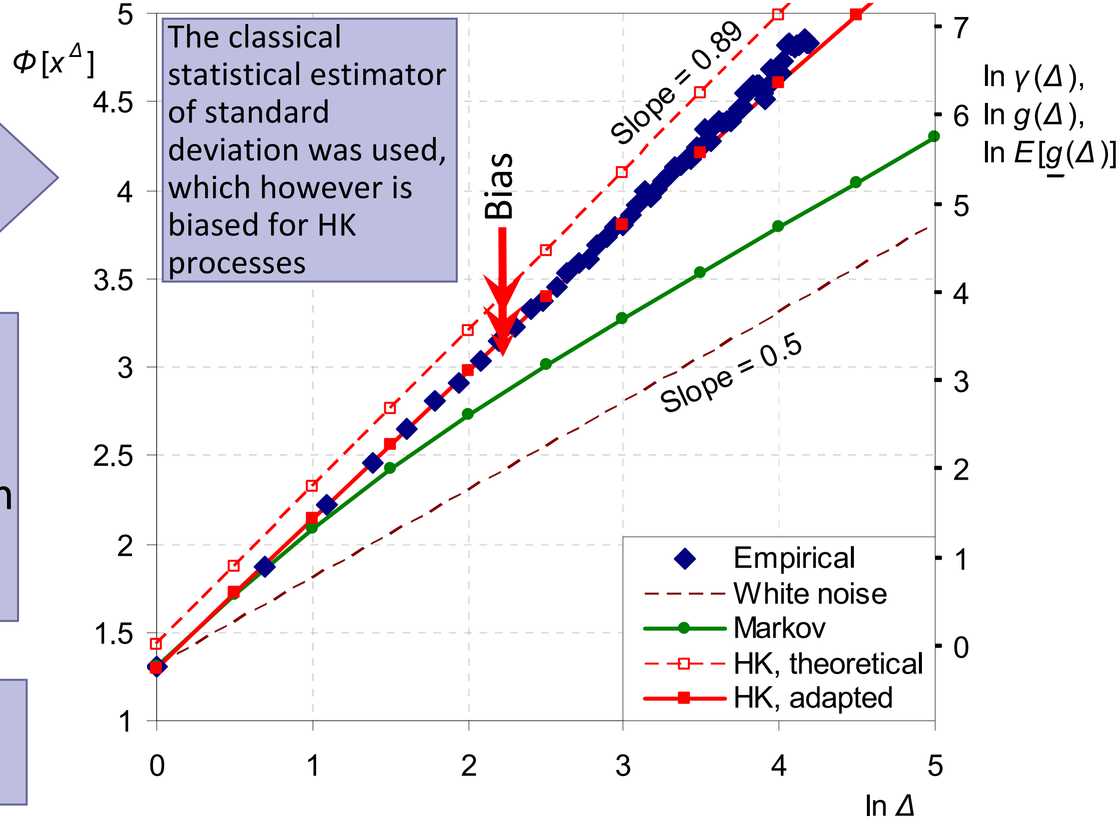
Again the Markov model is inappropriate , while the HK model (*H* = 0.89) is appropriate

Similar *H* values are estimated from the simultaneous record of maximum water levels and a similar value and from the modern record (131 years) of the Nile flows at Aswan

Nilometer data: Beran (1994) and lib.stat.cmu.edu/S/beran (here converted into meters)



Nile River annual minimum water level from the Roda Nilometer (663 values)

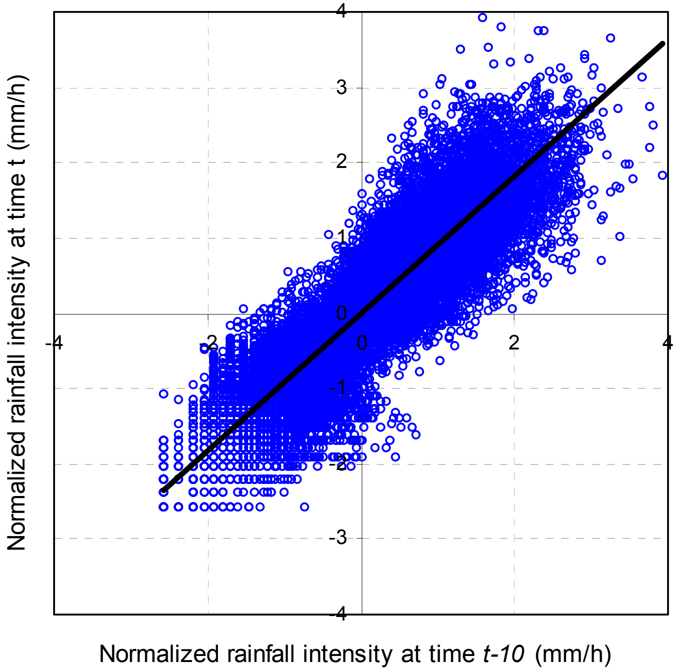


The classical statistical estimator of standard deviation was used, which however is biased for HK processes

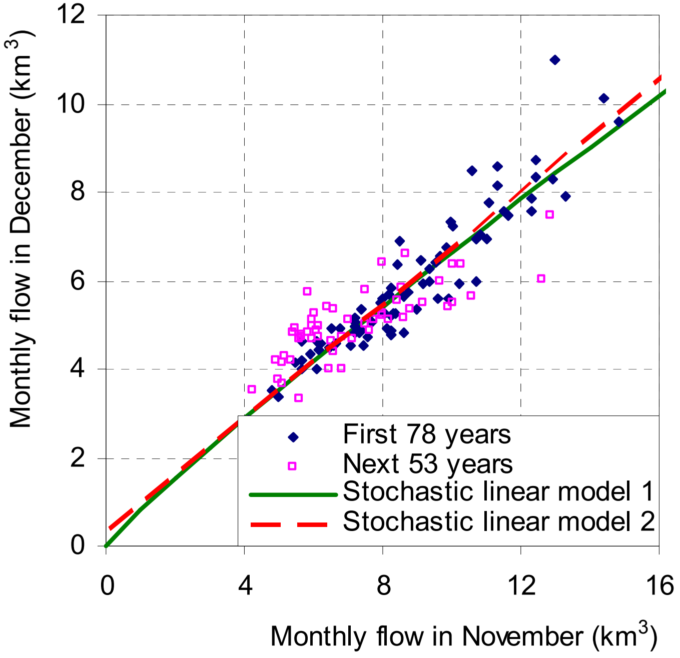
Bias

Maximum entropy and the emergence of linearity in highly nonlinear systems

- Hydrological processes (e.g. rainfall, runoff) are highly nonlinear if modelled using deterministic dynamical systems methods
- The same processes, if approached macroscopically in stochastic terms, exhibit impressively linear behaviour (after the processes are transformed to Gaussian)
- Linearity in stochastic terms is a result of the principle of maximum entropy and makes our macroscopic descriptions as simple and parsimonious as possible



Rainfall at Iowa measured at temporal resolution of 10 s (Papalexiou *et al.*, 2011)



Monthly Nile flows (Koutsoyiannis *et al.*, 2008)

Maximum entropy and parsimonious stochastic modelling

- Multivariate stochastic modelling involves vectors and matrices of parameters with very many elements
- As an example, we consider the prediction $\underline{w}$ of the monthly flow one month ahead, conditional on a number s of other variables with known values that compose the vector $\underline{\boldsymbol{z}}$, using the linear model:

  $$\underline{w} = \boldsymbol{a}^T \underline{\boldsymbol{z}} + \underline{v}$$

  where $\boldsymbol{a}$ is a vector of parameters (the superscript T denotes the transpose of a vector or matrix) and $\underline{v}$ is the prediction error, assumed independent of $\underline{\boldsymbol{z}}$; for simplicity, all elements of $\underline{\boldsymbol{z}}$ are assumed normalized and with zero mean and unit variance
- For the model to take account of both short-range and long-range dependence (HK behaviour), a possible composition of $\underline{\boldsymbol{z}}$ may include the following:
  - The flows of a few previous months of the same year
  - All available flow measurements of the same month on previous years
- The model parameters are estimated from:

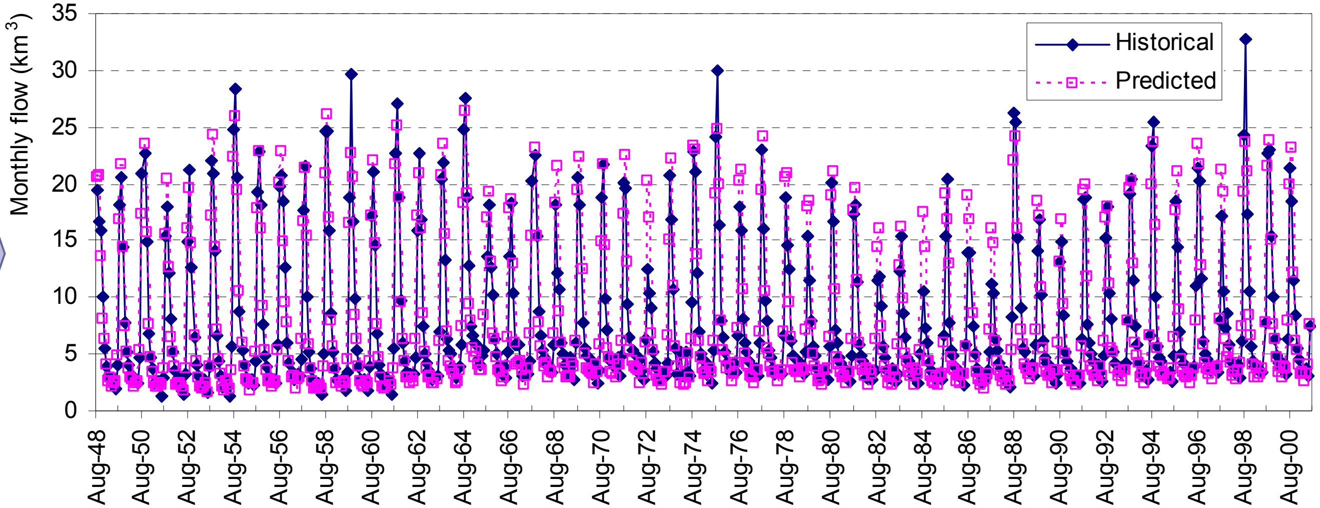
  $$\boldsymbol{a}^T = \boldsymbol{\eta}^T \boldsymbol{h}^{-1}, \quad \mathrm{Var}[\underline{v}] = 1 - \boldsymbol{\eta}^T \boldsymbol{h}^{-1} \boldsymbol{\eta} = 1 - \mathbf{a}^T \boldsymbol{\eta}$$

  where $\boldsymbol{\eta} := \mathrm{Cov}[\underline{w}, \underline{\boldsymbol{z}}]$ and $\boldsymbol{h} := \mathrm{Cov}[\underline{\boldsymbol{z}}, \underline{\boldsymbol{z}}]$ (see Koutsoyiannis, 2000)

ME and parsimonious stochastic modelling (cont'd)

- Both the vector $\boldsymbol{\eta} := \mathrm{Cov}[\underline{w}, \underline{\boldsymbol{z}}]$ and the matrix $\boldsymbol{h} := \mathrm{Cov}[\underline{\boldsymbol{z}}, \underline{\boldsymbol{z}}]$ may contain numerous items, typically of the order of 10^3-10^4 (e.g. for a dimensionality 100, if we have 100 years of observations: 100 + 100 × 100 = 10 100 items—albeit reduced due to symmetry)
- Traditionally, the items of such covariance matrices and vectors have been estimated directly from data; this is totally illogical (100 years of data cannot support the statistical estimation of 1000-10 000 parameters)
- An alternative approach is to use data to estimate a couple of parameters per month and derive all other 'unestimated' parameters by maximizing entropy
- Such entropy maximization is in fact very simple (a generalized matrix decomposition)

Example: One month ahead predictions of Nile flow in comparison to historical values for the validation period (Efficiency = 91%; see details in Koutsoyiannis *et al.*, 2008)



Finale: Entropy beyond hydrology

- **Entropy and parsimony**
  - The tendency of entropy, i.e. uncertainty, to become maximum makes possible a parsimonious description of complex natural systems
- **Entropy and change**
  - Reversible transformations, in which entropy remains constant, are trivial
  - Mass, energy and momentum are conserved in all transformations
  - It is the entropy that changes in irreversible processes and its tendency to become maximum drives the change
  - *The second law ... provides a foundation for understanding why* any *change occurs* (Atkins, 2007, p. 49)
- **Entropy, evolution and creativity**
  - Entropy production in irreversible processes causes a system to evolve
  - Without evolution and creativity the uncertainty of future would be low
- **Entropy and value of change**
  - Actions of change are not costless; the cost paid is entropy and this cost represents a value of the action taken

Entropy beyond hydrology (cont'd)

- **Entropy and life**
  - The second law allows for local decrease of entropy—at a larger cost paid to the environment
  - Living organisms, if isolated, are clusters with low entropy—but making the entropy of the big picture higher
  - Life is fully consistent with maximum uncertainty—dead systems are certain
- **Entropy and freedom**
  - *No other scientific law has contributed more to the liberation of the human spirit than the second law of thermodynamics* (Atkins, 2007, p. 49)
  - Would freedom have a meaning without uncertainty?
  - If the future was certain, and thus controllable (by some), would the world be livable?

Entropy and laughter: We predict, God laughs... (we laugh, too, but after years: these are from 1970)

- ***Civilization will end within 15 or 30 years unless immediate action is taken against problems facing mankind***
  George Wald, Harvard Biologist, share of the 1967 Nobel Prize in Physiology or Medicine, quoted in Climate Change and the Emergence of Civilization By Carl Grant Looney, Ph.D. , Xlibris, 2011, books.google.com/books?id=cMlBE3umGzMC p. 390, as well as in p. 26 of The environment: opposing viewpoints, William Dudley, Greenhaven Press, 2001, books.google.com/books?id=vwAKAQAAMAAJ
- *Demographers agree almost unanimously on the following grim timetable: by 1975 widespread famines will begin in India; these will spread by 1990 to include all of India, Pakistan, China and the Near East, Africa. By the year 2000, or conceivably sooner, South and Central America will exist under famine conditions…* ***By the year 2000, thirty years from now, the entire world, with the exception of Western Europe, North America, and Australia, will be in famine***
  Peter Gunter, professor, North Texas State University, The Living wilderness, Volumes 34-35, Wilderness Society, 1970, books.google.com/books?id=ogv0AAAAMAAJ (also quoted in Climate Change and the Emergence of Civilization By Carl Grant Looney, Ph.D. , p. 389)
- *The world has been chilling sharply for about twenty years... If present trends continue, the world will be about four degrees colder for the global mean temperature in 1990, but* ***eleven degrees colder in the year 2000. This is about twice what it would take to put us into an ice age***
  Kenneth E. W. Watt, Ecologist and Professor of University of California, Davis, Earth Day—the beginning: a guide for survival (pp. 14-15; books.google.com/books?id=1yE9AAAAIAAJ)

Concluding remarks

- Entropy is none other than uncertainty properly quantified
- The tendency of entropy to become maximum is not a curse—it is an eulogia
- It offers the basis to understand and describe Nature, but it also constitutes the driving force of change and evolution
- Extremal entropy considerations provide a theoretical basis in stochastic modelling of hydrological processes
- They also provide an important connection with statistical thermophysics, which suggests that physical modelling of complex systems can only be done in stochastic terms

Epilogue

- In 1927, **Werner Heisenberg** published his **uncertainty principle**, expressing the limitations of predictability and even observability of the quantum world
- In 1930, **David Hilbert** pronounced his aphorism “Wir müssen wissen, wir werden wissen” (“**We must know, we will know**”); Hilbert did not know that the day before, **Kurt Gödel** had announced his **incompleteness theorem** thus killing Hilbert’s dogma
- 80 years after, hydrology and other geophysical sciences still embrace the dogma “We must know, we will know”
- Wanted: Recognition of the intrinsic uncertainty in Nature and awareness of the limitations in predictability
- My view: We must live, we can live with incomplete knowledge and with uncertainty

Θέλω να ζω,
Τις προβλέψεις ν’ αναιρώ

I like to live,
The predictions to deceive

From the theater play “Τι θα κάνουμε τώρα;” (“What shall we do now?”) by D. Kouroumbalis and the youngsters’ theater team So7 (Ερευνητικό Θέατρο, Athens, Greece, 2011); transliteration into English by D. Koutsoyiannis

References

- Atkins, P., *Four Laws that Drive the Universe*, Oxford Univ. Press, 131 pp, 2007
- Beran, J., *Statistics for Long-memory Processes*, Chapman and Hall, New York, 1994
- Hurst, H. E., Long term storage capacities of reservoirs, *Trans. Am. Soc. Civil Engrs.*, 116, 776–808, 1951
- Kolmogorov, A. N., Wienersche Spiralen und einige andere interessante Kurven in Hilbertschen Raum, *Dokl. Akad. Nauk URSS*, 26, 115–118, 1940
- Koutsoyiannis, D., A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, *Water Resources Research*, 36 (6), 1519–1533, 2000
- Koutsoyiannis, D., Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling, *Hydrological Sciences Journal*, 50 (3), 381–404, 2005
- Koutsoyiannis, D., An entropic-stochastic representation of rainfall intermittency: The origin of clustering and persistence, *Water Resources Research*, 42 (1), W01401, doi:10.1029/2005WR004175, 2006
- Koutsoyiannis, D., Hurst-Kolmogorov dynamics as a result of extremal entropy production, *Physica A: Statistical Mechanics and its Applications*, 390 (8), 1424–1432, 2011
- Koutsoyiannis, D., and C. Onof, Rainfall disaggregation using adjusting procedures on a Poisson cluster model, *Journal of Hydrology*, 246, 109–122, 2001
- 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
- Papalexiou, S.-M., and D. Koutsoyiannis, Entropy based derivation of probability distributions: A case study to daily rainfall, *Advances in Water Resources* [in review], 2011
- Papalexiou, S.-M., D. Koutsoyiannis and A. Montanari, Can a simple stochastic model generate rich patterns of rainfall events?, *Journal of Hydrology* [in review], 2011
- Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw-Hill, New York, 1991

Appendix: Some mathematical derivations

Caution:
(1) Results not cross-checked; errors may be present (please report them to dk@itia.ntua.gr)
(2) No quantum effects have been taken into account

A first application of principle of ME to uncertain motion

- Consider a motionless cube with edge a (volume $V = a^3$) containing spherical particles of mass m_0 (e.g. monoatomic molecules) in fast motion, in which we cannot observe the exact position and velocity
- A particle's state is described by 6 variables, 3 indicating its position $\underline{x}_i$ and 3 indicating its velocity $\underline{u}_i$, with $i = 1, 2, 3$; all are represented as random variables, forming the vector $\underline{\boldsymbol{z}} = (\underline{x}_1, \underline{x}_2, \underline{x}_3, \underline{u}_1, \underline{u}_2, \underline{u}_3)$
- Constraints for position: $0 \le \underline{x}_i \le a$
- Constraints for velocity (where the integrals are over feasible ranges of variables):
  - ❑ Conservation of momentum: $E[\underline{u}_i] = \int u_i f(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z} = 0$ (the cube is not in motion)
  - ❑ Conservation of energy: $E[\|\underline{\boldsymbol{u}}\|^2/2] = \int \|\underline{\boldsymbol{u}}\|^2/2\, f(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z} = \varepsilon$, *where* ε is the energy
- Application of the principle of maximum entropy with the above constraints will give the distribution of $\underline{\boldsymbol{z}}$ as:

  $f(\boldsymbol{z}) = [3/(4\pi\, \varepsilon)]^{3/2}\, (1/V)\, \exp[-3\|\boldsymbol{u}\|^2/\, (4\varepsilon)]$
- The marginal distributions are given by:

  $f(x_i) = 1/a$ (uniform in $[0, \mathrm{a}]$)

  $f(u_i) = [3/(4\pi\, \varepsilon)]^{1/2} \exp[-3u_i^2/(4\varepsilon)]$ (Gaussian with mean 0 and variance $2e/3$ = 2 × energy per degree of freedom)

  $f(\|\underline{\boldsymbol{u}}\|) = (2/\pi)^{1/2}[3/(2\varepsilon)]^3\ \|\boldsymbol{u}\|^2 \exp[-3\|\boldsymbol{u}\|^2/\, (4\varepsilon)]$ (Maxwell–Boltzmann)
- The entropy is then calculated as follows, where e is the base of natural logarithms:

  $\Phi[\underline{\boldsymbol{z}}] = (3/2) \ln[(4/3)\, \pi\, \mathrm{e}\, \varepsilon)] + \ln V$

Typical application of principle of ME to the atmosphere

- The dominant atmospheric gases are diatomic (N_2, O_2); in a diatomic gas, in addition to kinetic energy we have rotational energy at two axes x and y perpendicular to the axes defined by the two molecules; these are $L_x^2 / 2I_x$ and $L_y^2 / 2I_y$, where L denotes angular momentum and I denotes rotational inertia
- We consider again a motionless cube with edge a (volume $V = a^3$) containing N identical diatomic molecules, each one with mass m_0, of a gas in motion with total (internal) energy U
- Each molecule is described by 8 variables, 3 indicating its position $\underline{x}_i$, 3 indicating its velocity $\underline{u}_i$, with $i = 1, 2, 3$, and two indicating its rotation, $\underline{u}_4 := L_x / \surd(I_x\, m_0)$ and $\underline{u}_5 := L_y / \surd(I_y\, m_0)$; all are represented as random variables, forming the vector $\underline{\boldsymbol{z}} = (\underline{x}_1, \underline{x}_2, \underline{x}_3, \underline{u}_1, \underline{u}_2, \underline{u}_3, \underline{u}_4, \underline{u}_5)$
- Independence among molecules can arguably be assumed
- Constraints are same as before
  - position: $0 \le \underline{x}_i \le a$
  - momentum/angular momentum: $E[\underline{u}_i] = 0$ (the cube is not in motion)
  - energy: $E[||\underline{\boldsymbol{u}}||^2/2] = \varepsilon$, *where* $\varepsilon = U/(N\, m_0)$ is the energy per unit mass
- Application of the principle of maximum entropy with the above constraints will give the distribution of $\underline{\boldsymbol{z}}$ (one molecule) and $\boldsymbol{z}_1,\ldots, \boldsymbol{z}_N$ (all molecules) as:

$$f(\boldsymbol{z}) = [5/(4\pi\, \varepsilon)]^{5/2}\, (1/V)\, \exp[-5||\boldsymbol{u}||^2/ (4\varepsilon)]$$

$$f(\boldsymbol{z}_1,\ldots, \boldsymbol{z}_N) = [5/(4\pi\, \varepsilon)]^{5N/2}\, (1/V)^N \prod_{j=1}^{N} \exp[-5||\boldsymbol{u}_j||^2/ (4e)]$$

- The entropy of the N molecules is then calculated as:

$$\Phi[\underline{\boldsymbol{z}}_1,\ldots, \underline{\boldsymbol{z}}_N] = (5N/2)\, \ln[(4/5)\, \pi\, e\, \varepsilon)] + N \ln V$$

Macroscopic view of entropy and related quantities

- The entropy per unit mass is

  $\varphi := \Phi/(m_0\, N) = \{(5/2)\ \ln[(4/5)\ \pi\ \mathrm{e}\ \varepsilon] + \ln V\}\ /m_0$

- The temperature ϑ is defined as

  $1/\vartheta := \partial\Phi/\partial U = \partial\varphi/\partial\varepsilon$, so that $\vartheta = (2/5)\varepsilon\ m_0 = (2/5)\ \varepsilon_\mathrm{p}$ (Note: units of energy, i.e. J)

- The entropy per unit mass can be expressed in terms of the temperature as

  $\varphi = [(5/2)\ \ln(2\ \pi\ \mathrm{e}\ /m_0) + (5/2)\ \ln \vartheta + \ln V]\ /m_0$

- Consider a time interval $\mathrm{d}t$; any particle at distance $\mathrm{d}x_3 \le -u_3\ \mathrm{d}t$ will collide with the cube edge $(x_3 = 0)$
- The distribution function of (x_3, u_3) is

  $f(x_3, u_3) = (1/a)[5/(4\pi\ \varepsilon)]^{1/2}\ \exp[-5u_3^2/(4\varepsilon)]$

- The expected value of the momentum of molecules colliding at the cube edge $(x_3 = 0)$ within time interval $\mathrm{d}t$ is

  $E[\underline{p}(\mathrm{d}t)] = N \int_0^\infty \mathrm{d}x_3 \int_{-\infty}^{-x_3/\mathrm{d}t} m_0\, u_3\, f(x_3, u_3)\ \mathrm{d}u_3 = N\ \mathrm{e}\ m_0\ \mathrm{d}t/(5a)$

- According to Newton’s 2nd law, the force exerted on the edge is $F = 2\ E[\underline{p}(\mathrm{d}t)]/\mathrm{d}t$ and the pressure is

  $p = F\ /\ a^2 = 2\ N\ \mathrm{e}\ m_0/(5V) = N\ \vartheta\ /\ V \Leftrightarrow p\ V = n\ N_\mathrm{A}\ \vartheta$ (**Law of ideal gases**)

  where n is the number of moles and N_A $(=6.022\times10^{23}\ \mathrm{mol}^{-1})$ is the Avogadro constant

- Using p instead of V, and the molecular mass $M_0 = N_\mathrm{A}\ m_0$ instead of m_0, denoting $c := \ln(2\ \pi\ \mathrm{e}\ /m_0)$ and $c' := c + \ln N_\mathrm{A}$, and finally generalizing for ν degrees of freedom per particle, the entropic quantities become:

  $\varphi = (N_\mathrm{A}/M_0)\ [c + (1+ \nu/2)\ \ln \vartheta + \ln N - \ln p] = (N_\mathrm{A}/M_0)\ [c' + (1+ \nu/2)\ \ln \vartheta + \ln n - \ln p]$

  $\Phi = N\ [c + (1+ \nu/2)\ \ln \vartheta + \ln N - \ln p] = n\ N_\mathrm{A}\ [c' + (1+ \nu/2)\ \ln \vartheta + \ln n - \ln p]$

Standard (Boltzmann) formalism

- In Boltzmann's formalism the temperature is (unnecessarily) regarded as an independent fundamental unit (kelvin, K) and the (unnecessary) constants k = 0.138 065 yJ/K (yocto-joule per kelvin) and $R_* = k\,N_A$ = 8 314.472 J K^{-1} $kmol^{-1}$ are used
- Accordingly, the so-called absolute (or thermodynamic) temperature is $T := \vartheta/k$ [units: K], while Boltzmann's entropy is $S = k\,\Phi$ [units: J/K], so that the relationship $1/T = \partial S/\partial E = \partial s/\partial \varepsilon$ is retained
- The entropy per unit mass becomes:

  $$s = [(7/2)\ \ln(T/T_0) - \ln(p/p_0)]\ (R_*/M_0)$$

  where T_0 and p_0 designate an arbitrary macroscopic state to which we assign zero entropy (typically T_0 = 200 K and p_0 = 1000 hPa)
- By setting $R_*/M_0 =: R$ and $(7/2)\ R =: c_p$ (the specific heat of the gas), we obtain the final entropic formula:

  $$s = c_p \ln(T/T_0) - R\ \ln(p/p_0)$$
- The law of ideal gases becomes

  $$p\ V = n\ R_*\ T \Leftrightarrow p\ v = R\ T$$

  where $v := V\ /\ m = V\ /\ (n\ M_0)$ is the volume per unit mass (= 1/density)
- Note that in the dry atmospheric air mixture, M = 28.96 kg/kmol, R = 8314.472/28.96 = 287.1 J K^{-1} kg^{-1}, so that c_p = (7/2) × 287.1 = 1004.8 J K^{-1} kg^{-1}, while the experimental value is c_p = 1004 J K^{-1} kg^{-1}!

Compound entropy from different states

- Suppose that a system described by a vector random variable $\underline{z}$ can alternate between two states 1 and 2 with conditional entropies $\Phi_1 := \Phi[\underline{z}|1]$ and $\Phi_2 := \Phi[\underline{z}|2]$; further, suppose that the probabilities of states 1 and 2 are π and $1 - \pi$, thus defining an entropic quantity:

  $\Phi_\pi := -\pi \ln \pi - (1 - \pi) \ln (1 - \pi)$
- It is easy to show that the total entropy is $\Phi = \pi\Phi_1 + (1 - \pi)\, \Phi_2 + \Phi_\pi$
- It is easy to show that the value of π that maximizes Φ is such that

  $\exp(\Phi_1)/\pi = \exp(\Phi_2)/(1 - \pi)$
- For a single particle with v degrees of freedom, $\Phi[\underline{z}|1] = c + (v/2) \ln \varepsilon_1 + \ln V_1$; $\Phi[\underline{z}|2] = c + (v/2) \ln \varepsilon_2 + \ln V_1$ so that for maximum entropy,

  $\varepsilon_1^{v/2} V_1/\pi = \varepsilon_2^{v/2} V_2/(1 - \pi)$
- Extending to N particles, of which $N_1 = \pi N$ and $N_2 = (1 - \pi) N$ will belong to states 1 and 2, respectively, and noting that energies ε are proportional to temperatures ϑ, we obtain:

  $\vartheta_1^{v/2} V_1/N_1 = \vartheta_2^{v/2} V_2/N_2$
- From the law of ideal gases, $V_1/N_1 = \vartheta_1/p_1$, $V_2/N_2 = \vartheta_2/p_2$, so that

  $\vartheta_1^{1 + v/2}/p_1 = \vartheta_2^{1 + v/2}/p_2$
- In Boltzmann's formalism $T_1^{c_p/R}/p_1 = T_2^{c_p/R}/p_2$ (**law of adiabatic change**)
- It is easy to show that the entropy per unit mass is the same in both states, i.e. $\varphi_1 = \varphi_2$ and $s_1 = s_2$