European Geosciences Union General Assembly 2011
Vienna, Austria, 3-8 April 2011

Session NP3.8/SSS5.7
Scaling, nonlinearity, and complexity in soils and surface hydrology

Scaling as enhanced uncertainty

Demetris Koutsoyiannis & Simon-Michael Papalexiou
Department of Water Resources and Environmental Engineering
Faculty 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/1115/**

Entropy as a measure of uncertainty

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

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

  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 $f(z)$)

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

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

- Definition of **entropy production in logarithmic time** (**EPLT**)

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

- 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 t and any two processes $\underline{z}_1(t)$ and $\underline{z}_2(t)$, any inequality between entropy productions, e.g. $\Phi'[\underline{z}_1(t)] < \Phi'[\underline{z}_2(t)]$ holds also for EPLTs, e.g. $\varphi[\underline{z}_1(t)] < \varphi[\underline{z}_2(t)]$

The principle of maximum entropy (ME)

- 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
- Entropy maximization of a random variable $\underline{z}$ bounded within [0, b]:
  - uniform, $f(z) = 1/b$
- 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
- 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, $f(z) = \exp\{-[(z-\mu)/\sigma]^2/2\} / [\surd(2\pi)\sigma]$

Typical application of principle of ME to thermodynamics

- Consider a motionless cube with edge *a* (volume $V = a^3$) containing *N* identical monoatomic molecules, each one with mass *m*, of a gas in motion with total (internal) energy *U*
- Each molecule is described by 6 variables, 3 indicating its position $\underline{x}_i$ and 3 indicating its velocity $\underline{v}_i$, with $i = 1, 2, 3$; all are represented as random variables, forming the vector $\underline{\boldsymbol{z}} = (\underline{x}_1, \underline{x}_2, \underline{x}_3, \underline{v}_1, \underline{v}_2, \underline{v}_3)$
- Independence among molecules can arguably be assumed
- Constraints for position: $0 \le \underline{x}_i \le a$
- Constraints for velocity (where the integrals are over feasible ranges of variables):
  - Preservation of momentum: $E[m\,\underline{v}_i] = -\!\int v_i f(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} = 0$ (the cube is not in motion)
  - Preservation of energy: $E[m\,||\underline{\boldsymbol{v}}||^2/2] = -\!\int ||\boldsymbol{v}||^2/2\, f(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} = m\,e$, *where e* is the energy per unit mass ($e := U/(m\,N)$)
- Application of the principle of maximum entropy with the above constraints will give the distribution of $\underline{\boldsymbol{z}}$ as:

$$f(\boldsymbol{z}) = [3/(4\pi\, e)]^{3/2}\,(1/V)\,\exp[-3||\boldsymbol{v}||^2/(4e)]$$

- The marginal distributions are given by:

$f(x_i) = 1/a$ (uniform in [0, a])

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

- These equations can yield the entire framework of the behaviour of gases at equilibrium, including relationships of macroscopic quantities such as pressure and temperature with volume (the equation of state, $pV = nRT$)

Disambiguation of scaling: different types

- Scaling behaviours are typically represented as power laws of different statistical properties
  - distribution tails
  - autocorrelograms
  - periodograms
  - climacograms
- Independent variables in such power laws could be different quantities such as
  - state-related: random variates representing states of a system
  - time-related: temporal scale, spatial scale, time lag, frequency (inverse time)
  - space-related: spatial scale, spatial displacement, inverse space
- The power laws are applicable either on the entire domain of the variable of interest or asymptotically

Different (albeit often confused) types of scaling

- scaling in state: refers to marginal distributions
- scaling in time: refers to joint distributions of stochastic processes
- scaling in space: refers to joint distributions of random fields

Demystification of scaling

- The omnipresence of scaling behaviours has often been regarded as a mystery and has been interpreted by analogous ways, e.g. by invoking a "self-organizing" power of natural systems (cf. "self-organized criticalities")
- In another view, power laws just contrast exponential laws:
  - We often encounter functions $f(x) \geq 0$ for which $f(x) \to 0$ as $x \to \infty$
  - Asymptotically exponential decay is a fast decay:
    - there exist a, b, $c > 0$, $b < 1$, so that for all $x > c$, $f(x) \leq a\, b^x$
  - Asymptotically power-law decay is a slow decay which is not exponential:
    - there exist a, b, $c > 0$, so that for all $x > c$, $f(x) \leq a\, x^{-b}$
      (note: for $f(x)$ to have finite integral over (c, ∞), b must be > 1)
- According to this view, scaling behaviours are just manifestations of enhanced uncertainty and are consistent with the principle of maximum entropy (as will be shown below)

Maximum entropy and scaling in state

- Most hydrometeorological variables are non-negative physical quantities unbounded from above; examples: precipitation, streamflow, temperature (expressed in kelvins or in joules)
- 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 e.g. in the kinetic theory of gases; rather they are estimated from data
- When $\sigma/\mu < 1$, the mean and variance can be loosely used as constraints, thus yielding distributions with exponential tails: from Gaussian to exponential → **no scaling**; example: temperature
- When $\sigma/\mu > 1$ (**highly variable processes**), the mean and variance cannot provide workable constraints and different constraints should be used → **possible scaling** (example: precipitation at fine temporal or spatial scales)
- When at some temporal or spatial scale a process exhibits scaling in state, i.e. power-law tail of its density function, $f(x) \propto x^{-b}$ with $b > 0$, then **it can be shown** that it will have the **same asymptotic scaling behaviour with same** b at **all** time scales
  (note: in aggregate scales this may be difficult to observe, except in very long records)

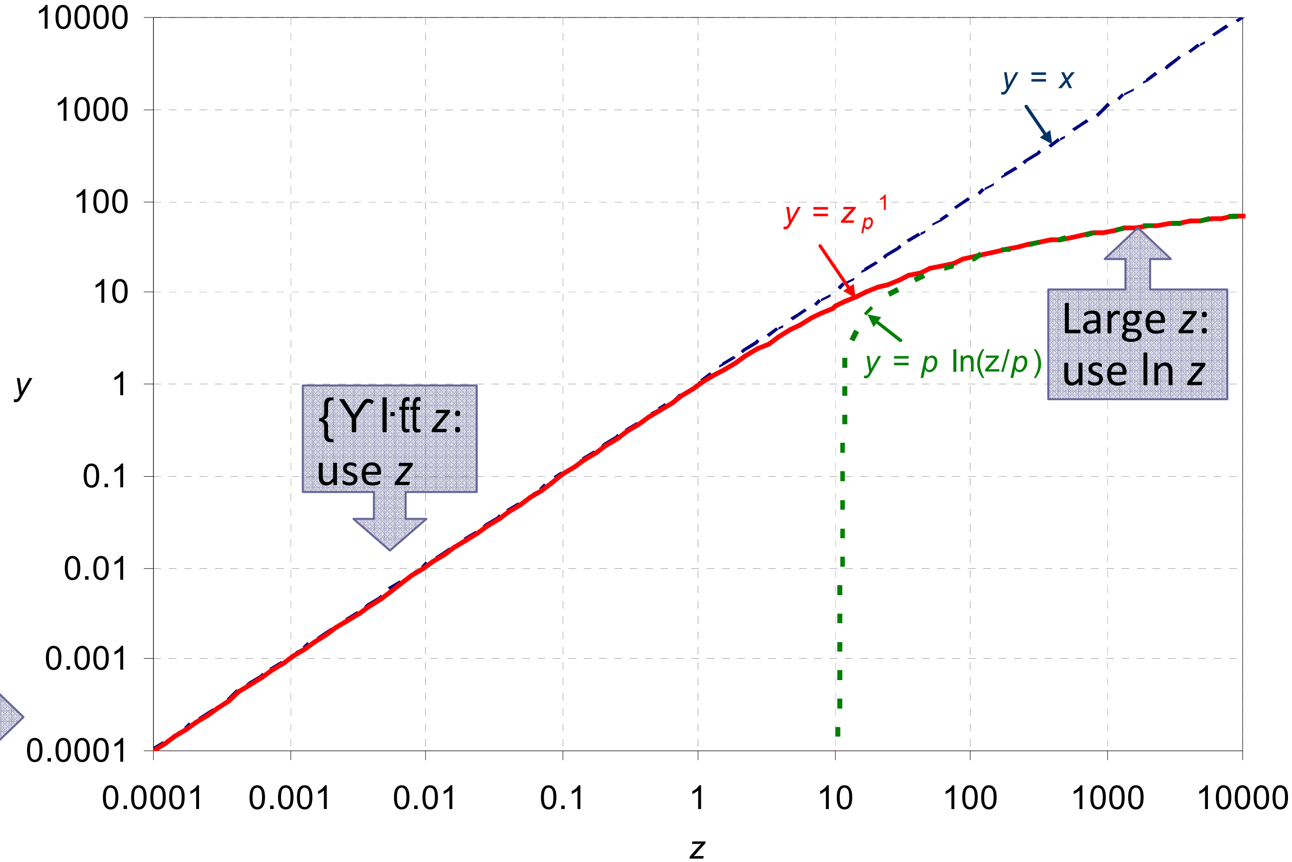
Towards workable constraints in highly variable geophysical processes

- Generalization of the classical power moments with p-moments (Papalexiou and Koutsoyiannis, 2011)
- A p-moment is defined to be the expectation $E[\underline{z}_p^q]$, where

  $$z_p^q := p^q \ln[1 + (z/p)^q]$$

  whereas p is a scale parameter with units identical to those of z
- For $p \to \infty$, $z_p^q \to z^q$ (the classical raise to power q)
- For finite p and for small z, $z_p^q \approx z^q$, while for large z, $z_p^q \propto \ln(z/p)$

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



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 $\Phi[\underline{z}] = 1 + \kappa + \ln \lambda$

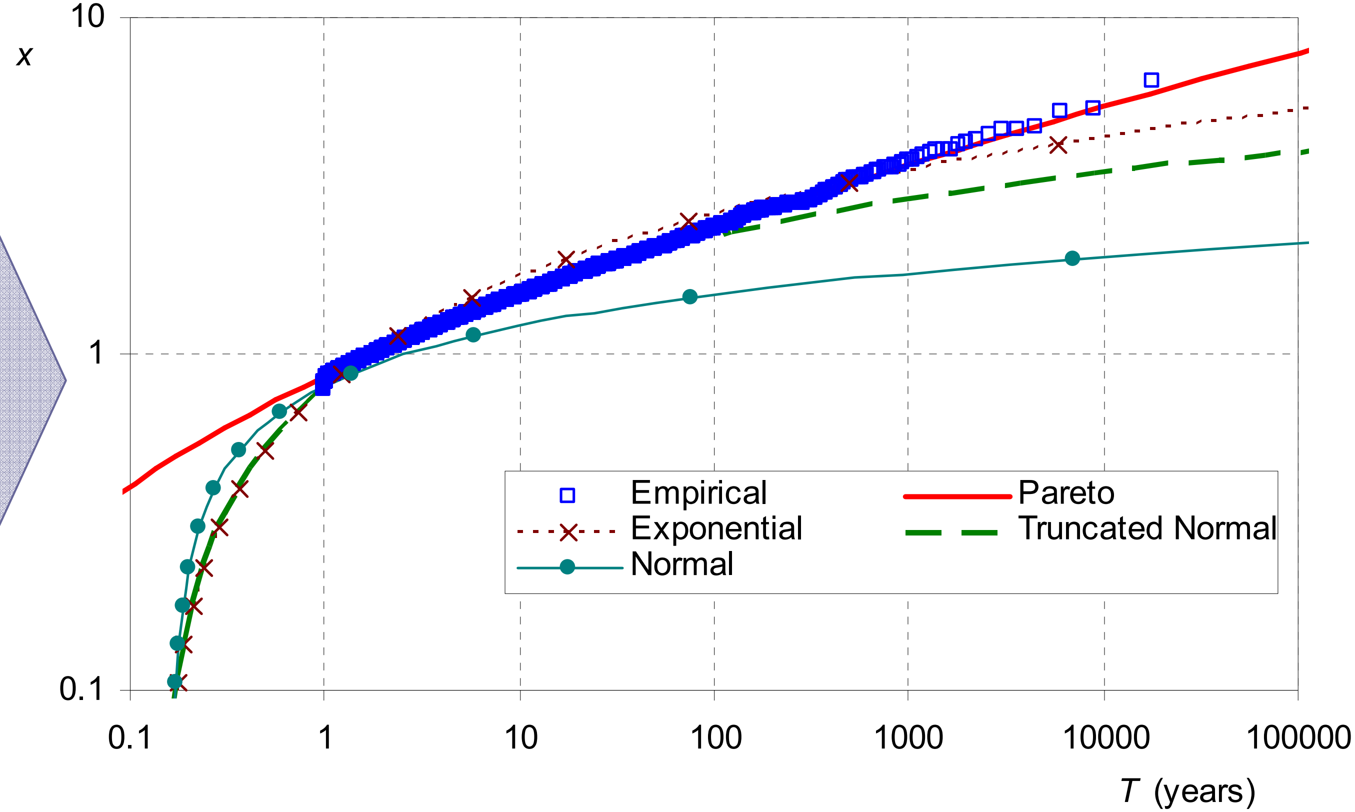
- The exponential distribution is fully recovered by setting $\kappa = 0$ ($p = \infty$); its statistics are $\mu = \sigma = \lambda$, $m_p = p \exp(\lambda p)\, \Gamma_{\lambda p}(0)/\lambda^2$, and $\Phi[\underline{z}] = 1 + \ln \lambda$
- In Pareto $\sigma/\mu = 1/\sqrt{(1 - 2\kappa)} > 1$, while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

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
ME distribution:
Pareto, κ = 0.15

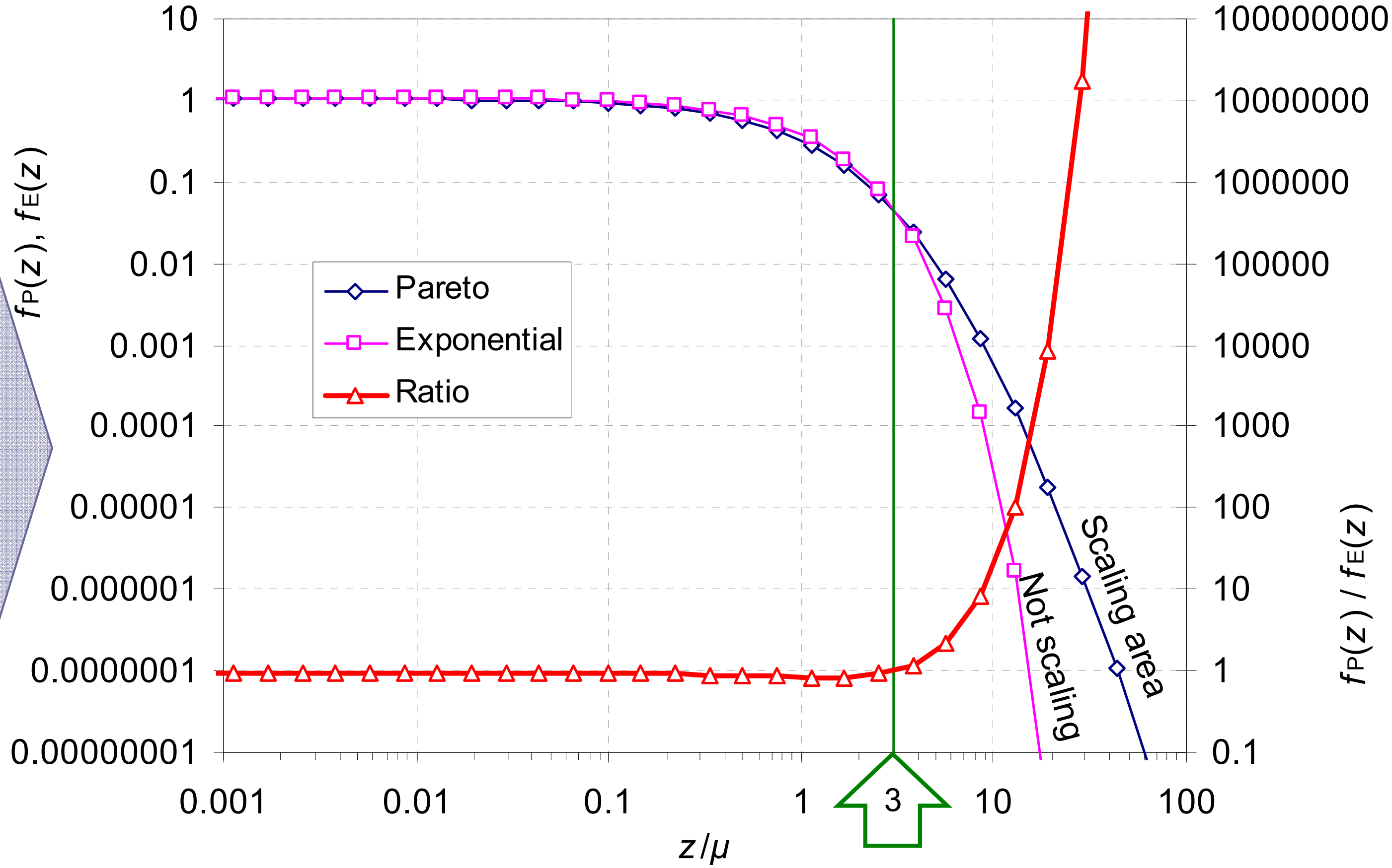
Scaling behaviour appears for T > ~50 yr (Koutsoyiannis, 2005)



Enhanced uncertainty with respect to extremes

- 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 scaling Pareto distribution gives extremes orders of magnitude more often than the non-scaling exponential distribution



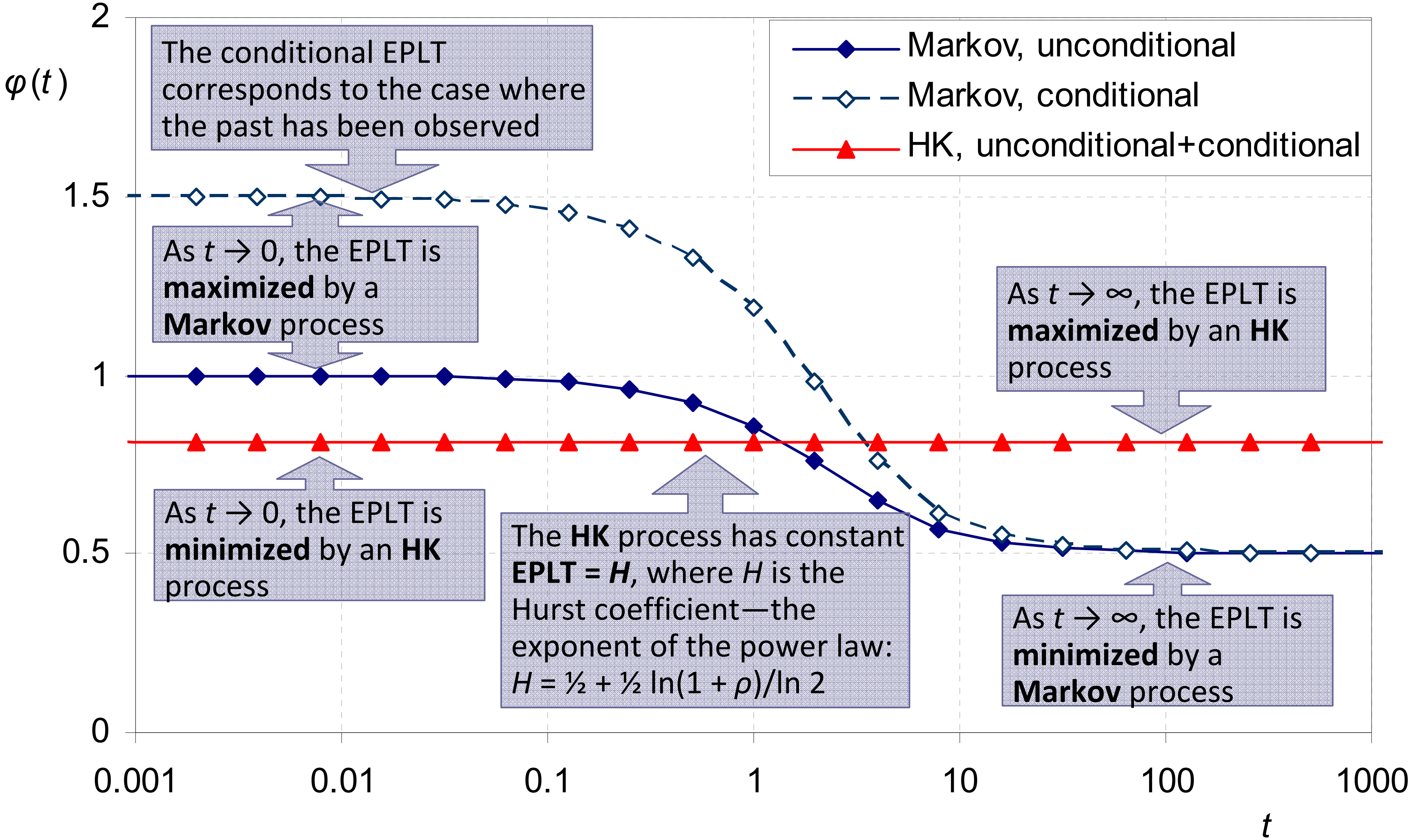
Maximum entropy and scaling in time

- Scaling in time refers to the dependence structure of a process rather than to marginal distributional properties
- The dependence structure can be expressed in terms of autocorrelogram, periodogram or climacogram, which are transformations of one another
- If any one of these is expressed as a power law, then all are power laws
- The simplest process with scaling properties in time is the Hurst-Kolmogorov (HK) process (due to Hurst, 1951, and Kolmogorov, 1940), while the simplest non-scaling process is the Markov process (AR(1) process in discrete time, Ornstein–Uhlenbeck process in continuous time)
- For determining the dependence structure by entropy extremization, because time is involved, 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
- Scaling in space is very similar to scaling in time, derived by extending the latter in higher dimensions and substituting space for time (cf. Koutsoyiannis *et al.*, 2011)

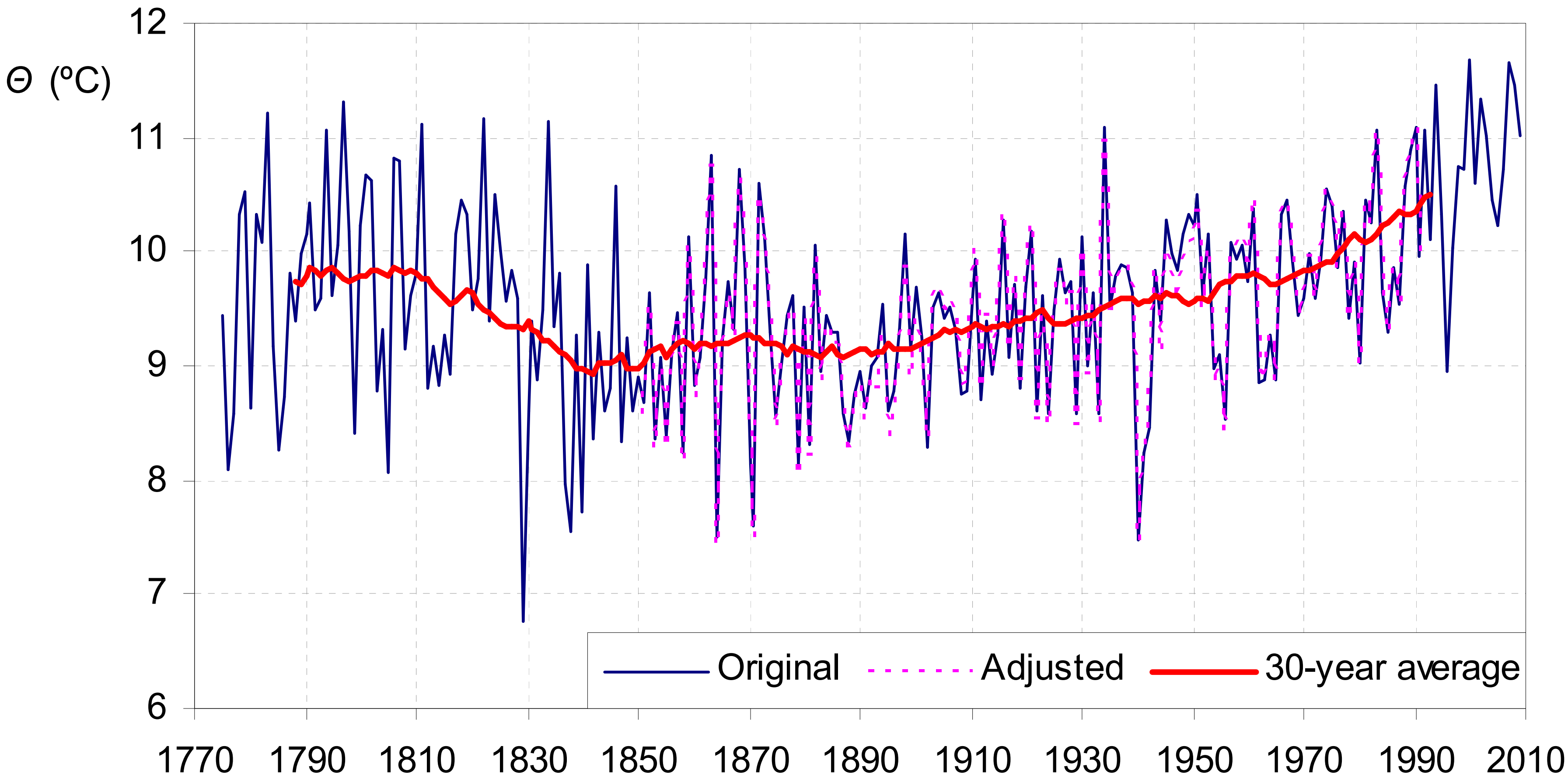
The two EPLT extremizing solutions

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



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)

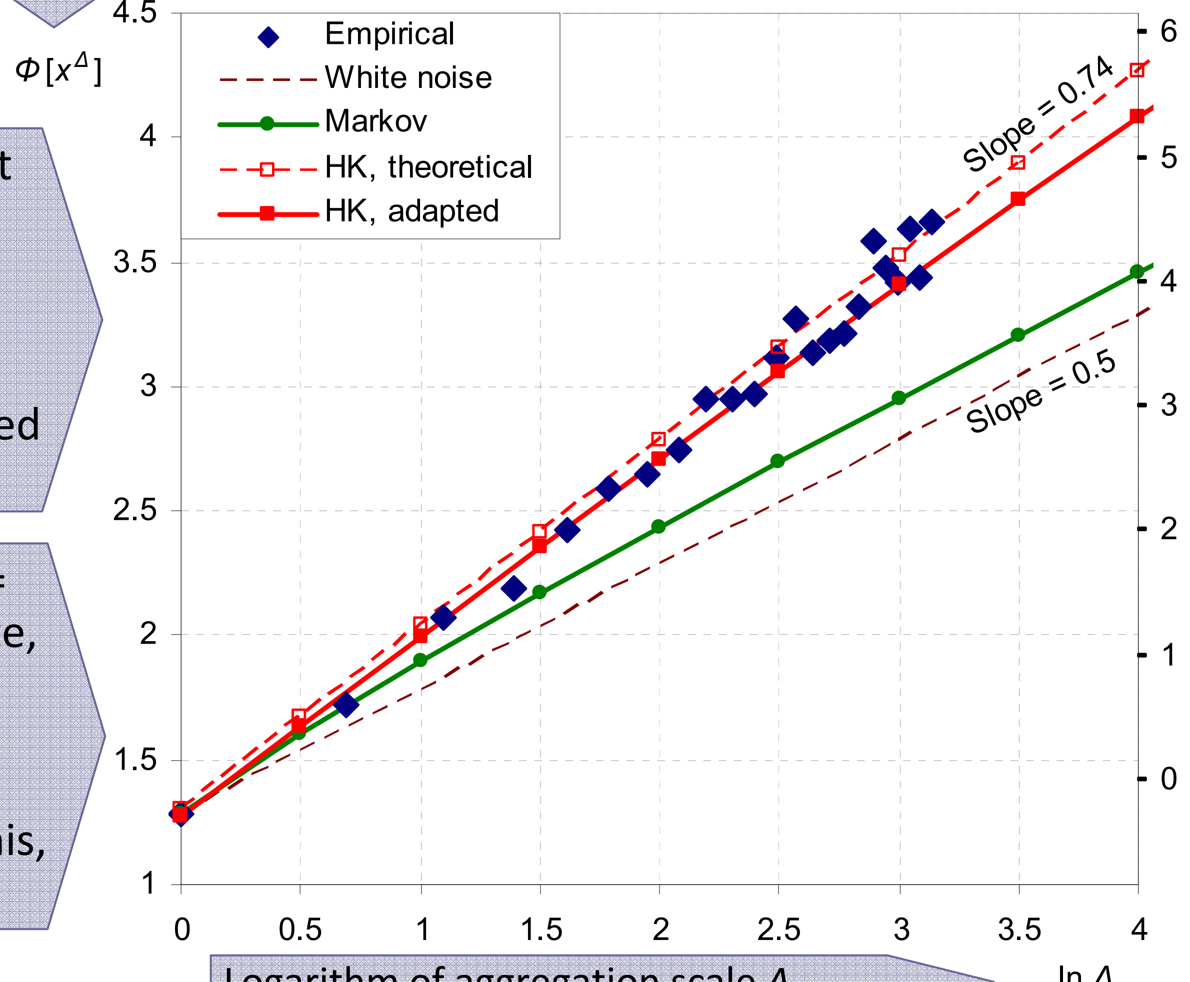


Comparison of the Markov and HK models (Vienna temp.)

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

Summary and conclusions

- Scaling behaviours are typically represented as power laws, as contrasted to exponential laws, and can be classified in different types: scaling in state, in time and in space
- Scaling behaviours are manifestations of enhanced uncertainty and are consistent with the principle of maximum entropy
- The connection of scaling with maximum entropy constitutes also a connection of stochastic representations of natural processes with statistical physics, in which, notably, maximum entropy considerations provide a basis for the Second Law of thermodynamics
- Extremal entropy considerations may thus provide theoretical background in modelling complex natural processes, which otherwise is heuristic and data-driven
- The examples given demonstrate:
  - the emergence of scaling from maximum entropy considerations
  - the consistency of scaling with real world behaviours, and
  - the enhancement of uncertainty due to scaling

> Uncertainty is the only certainty there is,
> and knowing how to live with insecurity is the only security
>
> (Paulos, 2003, p. v, quoting his father)

References

- 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., 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., Hurst-Kolmogorov dynamics as a result of extremal entropy production, *Physica A: Statistical Mechanics and its Applications*, 390 (8), 1424–1432, 2011.
- Koutsoyiannis, D., A. Paschalis, and N. Theodoratos, Two-dimensional Hurst-Kolmogorov process and its application to rainfall fields, *Journal of Hydrology*, 398 (1-2), 91–100, 2011.
- Papalexiou, S.-M., and D. Koutsoyiannis, An entropy based model of the probability distribution of daily rainfall [tentative title], Advances in Water Resources [in review] 2011.
- Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw-Hill, New York, 1991
- Paulos, J. A., *A Mathematician Plays the Stock Market*, Perseus Books Group, 2003

Presentation available online: itia.ntua.gr/1115/