Orlob International Symposium
University California Davis, 5-6 August 2013

Entropy:
from thermodynamics to hydrology

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

Presentation available online: **itia.ntua.gr/1365/**

Φύσις κρύπτεσθαι φιλεί

(Nature loves to hide herself)

Heraclitus (ca. 540-480 BC)

Part 1

Logical and mathematical foundation

Entropy = uncertainty

- The definition of entropy relies on probability theory and follows some postulates originally set up by Shannon (1948).
- Assuming a discrete random variable[*] $\underline{z}$ taking values z_j with probability mass function $P_j \equiv P(z_j) = P\{\underline{z} = z_j\}$, $j = 1,\dots,w$ the postulates, as reformulated by Jaynes (2003, p. 347) are:
  - (a) It is possible to set up a numerical measure Φ of the *amount of uncertainty*[†] which is expressed as a real number.
  - (b) Φ is a continuous function of P_j.
  - (c) If all the P_j are equal ($P_j = 1/w$) then Φ should be a monotonic increasing function of w.
  - (d) If there is more than one way of working out the value of Φ, then we should get the same value for every possible way.[‡]
- From these general postulates about uncertainty, a unique (within a multiplicative factor) function Φ is defined.

---

[*] Following the Dutch notation (Hemelrijk, 1966), an underlined symbol denotes a random variable; the same symbol not underlined represents a value of the random variable.

[†] The notation of entropy by Φ was done deliberately to avoid confusion with the classical thermodynamic entropy S, which has some differences discussed below.

[‡] Quantification of this postulate is given by Uffink (1995; theorem 1) and Robertson (1993, p. 3) and is related to refinement of partitions to which the probabilities P_j refer.

Entropy definition

Discrete random variable $\underline{z}$ with probability mass function $P_j \equiv P(z_j)$, $j = 1,\ldots,w$	Continuous random variable $\underline{z}$ with probability density function $f(z)$
Fundamental constraint	
$\sum_{j=1}^{w} P_j = 1 \qquad (1)$	$\int_{-\infty}^{\infty} f(z)\mathrm{d}z = 1 \qquad (2)$
Definition[*]	
$\Phi[\underline{z}] := \mathrm{E}[-\ln P(\underline{z})]$ $= -\sum_{j=1}^{w} P_j \ln P_j \qquad (3)$	$\Phi[\underline{z}] := \mathrm{E}\left[-\ln \frac{f(\underline{z})}{h(\underline{z})}\right]$ $= -\int_{-\infty}^{\infty} \ln \frac{f(z)}{h(z)} f(z) dz \qquad (4)$ where $h(z)$ is a background measure[†]
Basic properties of $\Phi[\underline{z}]$	
A nonnegative dimensionless quantity	A dimensionless quantity, either positive or negative, that depends on the assumed $h(z)$

---

[*] In case of risk of ambiguity, we call $\Phi[\underline{z}]$ *probabilistic entropy*; most commonly it is referred to as *information entropy*.

[†] The function $h(z)$ can be any probability density, proper (with integral equal to 1, as in (2)) or improper (meaning that its integral does not converge); typically it is an (improper) Lebesgue density, i.e. a constant with dimensions $[h(z)] = [f(z)] = [z^{-1}]$, so that the argument of the logarithm function be dimensionless.

Is probabilistic entropy different from thermodynamic entropy?

- Typically, probabilistic and thermodynamic entropy are regarded as two distinct concepts having in common only the name.
- The classical definition of thermodynamic entropy, S, through the equation $\mathrm{d}S = \mathrm{d}Q/T$, where Q and T denote heat and temperature, respectively, does not give any hint about similarity with the probabilistic entropy.
- The fact that the probabilistic entropy Φ is a dimensionless quantity, while the thermodynamic entropy S is not (units: J/K), has been regarded as an argument that the two are not identical.
- Even Jaynes (2003), founder of the maximum entropy principle (see below), states:

> *We must warn at the outset that the major occupational disease of this field is a persistent failure to distinguish between the information entropy, which is a property of any probability distribution, and the experimental entropy of thermodynamics, which is instead a property of a thermodynamic state as defined, for example by such observed quantities as pressure, volume, temperature, magnetization, of some physical system. They should never have been called by the same name; the experimental entropy makes no reference to any probability distribution, and the information entropy makes no reference to thermodynamics. Many textbooks and research papers are flawed fatally by the author's failure to distinguish between these entirely different things, and in consequence proving nonsense theorems.*

Probabilistic entropy = thermodynamic entropy*

- The classical definition of thermodynamic entropy is not necessary; it can be abandoned and replaced by the probabilistic definition.
- The thus defined entropy is the fundamental thermodynamic quantity, which supports the definition of all other derived ones.
  - For example, the temperature is defined as the inverse of the partial derivative of entropy with respect to the internal energy (see eqn. (25)).
- The entropy retains its dimensionless character even in thermodynamics, thus rendering the unit of kelvin an energy unit, a multiple of the joule (i.e., 1 K = 0.138 065 05 yJ = 1.380 650 5×10^{-23} J).
  - The introduction of the kelvin is an historical accident (cf. Atkins, 2007).
- The entropy retains its probabilistic interpretation as a measure of uncertainty, leaving aside the obscure 'disorder' interpretation (cf. Ben-Naim, 2008).
- Two examples related to hydrology will illustrate how thermodynamic laws can be derived from probabilistic entropy: the **law of ideal gases** and the **law of phase change transition** (Clausius-Clapeyron).

---

* The equality is meant on logical grounds; on technical grounds there may be some quantitative differences as shown below.

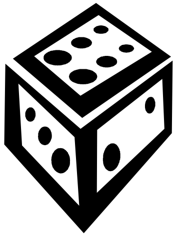
The principle of maximum entropy: Why entropy is important

- From a physical perspective, the tendency of entropy to become maximal (2nd Law of thermodynamics) is the driving force of natural change (in contrast to quantities such as mass, momentum and energy which are conserved).
- The counterpart of the physical law in logic is the principle of maximum entropy (ME; Jaynes, 1957).
- The ME principle 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.
- The rationale of the principle is very simple and almost self-evident: If uncertainty is not the maximum possible, then there must be some more information; but all information is already incorporated in the constraints.
- The ME principle can be regarded as:
  - a physical (ontological) principle obeyed by natural systems, as well as
  - a logical (epistemological) principle applicable in making inference about natural systems.*
- Compared to physical laws expressed in the form of equations, the ME principle, as a variational law, is extremely more powerful: it can determine infinitely many (or even uncountably many) unknown probabilities.

---

* This implies an optimistic view that our logic in making inference about natural systems could be consistent with the behaviour of the natural systems.

Entropy maximization: The die example



- What is the probability that the outcome of a die throw will be i?
- The entropy is:

$$\Phi = \mathrm{E}[-\ln P(\underline{z})] = -P_1 \ln P_1 - P_2 \ln P_2 - P_3 \ln P_3 - P_4 \ln P_4 - P_5 \ln P_5 - P_6 \ln P_6 \qquad (5)$$

- The equality constraint is:

$$P_1 + P_2 + P_3 + P_4 + P_5 + P_6 = 1 \qquad (6)$$

- The inequality constraint is $0 \le P_i \le 1$ (but is not necessary to include).
- Solution of the optimization problem (e.g. by the Lagrange method) yields a single maximum:

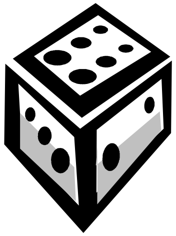
$$P_1 = P_2 = P_3 = P_4 = P_5 = P_6 = 1/6 \qquad (7)$$

- The entropy is $\Phi = -6\,(1/6)\ln(1/6) = \ln 6$. In general the entropy for w equiprobable outcomes is:

$$\Phi = \ln w \qquad (8)$$

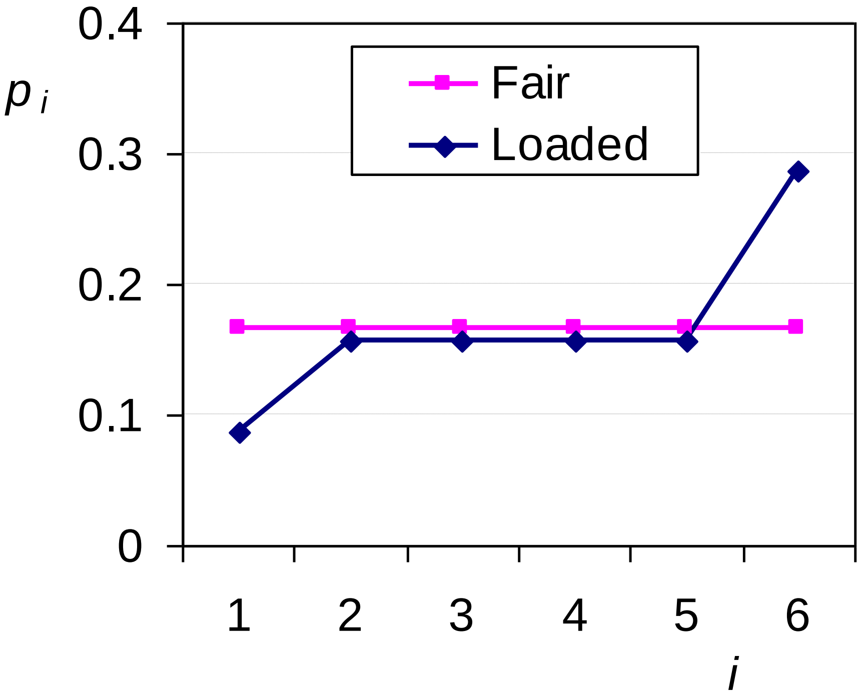
- In this case, the application of the ME principle (a variational law) is equivalent to the **principle of insufficient reason** (Bernoulli-Laplace; an “equation” form).
- Entropy and information are complementary to each other. When we know (observe) that the outcome is i ($P_i = 1$, $P_j = 0$ for $j \neq i$), the entropy is zero.

Entropy maximization: The loaded die example



- What is the probability that the outcome of a die throw will be *i* 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 ME 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 as shown in the figure.
- The entropy is $\Phi = 1.732$, smaller than in the case of equiprobability, where $\Phi = \ln 6 = 1.792$.
- The decrease of entropy in the loaded die derives from the additional information incorporated in the constraints.

Expected values as constraints: general solution

- In the most typical application of the ME principle, we wish to infer the probability density function *f*(z) of a continuous random variable $\underline{z}$ (scalar or vector) for constant background measure ($h(z) = 1$) with constraints formulated as expectations of functions $g_j(\underline{z})$.
- In other words, the given information, which is used in maximizing entropy, is expressed as a set of constraints formed as:

$$\mathrm{E}\left[g_j\left(\underline{z}\right)\right] = \int_{-\infty}^{\infty} g_j(z) f(z) \mathrm{d}z \ = \ \eta_j, \quad j = 1, \ldots, n \tag{9}$$

- The resulting maximum entropy distribution (by maximizing entropy as defined in (4) with constraints (9) and the obvious additional constraint (2)) is (Papoulis, 1991, p. 571):

$$f(z) = \exp\left(-\lambda_0 - \sum_{j=1}^{n} \lambda_j g_j(z)\right) \tag{10}$$

  where λ_0 and λ_j are constants determined such as to satisfy (2) and (9), respectively.
- The resulting maximum entropy is:

$$\Phi\left[\underline{z}\right] = \lambda_0 + \sum_{j=1}^{n} \lambda_j \eta_j \tag{11}$$

Typical results of entropy maximization

Constraints for the continuous variable $\underline{z}$	Resulting distribution $f(z)$ and entropy Φ (for $h(z) = 1$)
$\underline{z}$ bounded in $[0, w]$ no equality constraint	$f(z) = 1/w$ (uniform) $\Phi = \ln w$
$\underline{z}$ unbounded from both below and above	
No constraint or constrained mean μ	not defined
Constrained mean μ and standard deviation σ	$f(z) = \exp\{-[(z-\mu)/\sigma]^2/2\} / (\sigma\sqrt{2\pi})$ (Gaussian) $\Phi = \ln(\sigma\sqrt{2\pi e})$
Nonnegative $\underline{z}$ unbounded from above	
No equality constraint	not defined
Constrained mean μ	$f(z) = (1/\mu)\exp(-z/\mu)$ (exponential) $\Phi = \ln(\mu e)$
Constrained mean μ and standard deviation σ with $\sigma < \mu$	$f(z) = A\exp\{-[(z-\alpha)/\beta]^2/2\}$ (truncated Gaussian tending to exponential as $\sigma \to \mu$); the constants A, a and β are determined from the constraints and Φ from (11)
As above but with $\sigma > \mu$	not defined

Part 2

Application to simple physical systems

ME applied to the uncertain motion of a particle: setup

- We 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 $i = 1, 2, 3$ (**three degrees of freedom**); all are represented as random variables, forming the vector $\underline{\boldsymbol{z}} = (\underline{x}_1, \underline{x}_2, \underline{x}_3, \underline{u}_1, \underline{u}_2, \underline{u}_3)$.
- The constraints for position are:

$$0 \le \underline{x}_i \le a, \quad i = 1, 2, 3 \tag{12}$$

- The constraints for velocity are (where the integrals are over feasible space Ω, i.e. $(0, a)$ for each x_i and $(-\infty, \infty)$ for each u_i):
    - Conservation of momentum: $\mathrm{E}[m_0\, \underline{u}_i] = m_0 \int_\Omega u_i\, f(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z} = 0$ (the cube is not in motion), so that:

$$\mathrm{E}[\underline{u}_i] = 0, \quad i = 1, 2, 3 \tag{13}$$

    - Conservation of energy*: $\mathrm{E}[m_0 \|\underline{u}\|^2/2] = (m_0/2) \int_\Omega \|u\|^2 f(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z} = \varepsilon$, *where* ε is the energy per particle (known as **thermal energy**) and $\|\underline{u}\|^2 = \underline{u}_1^2 + \underline{u}_2^2 + \underline{u}_3^2$; thus, the constraint is:

$$\mathrm{E}[\|\underline{\boldsymbol{u}}\|^2] = 2\varepsilon/m_0 \tag{14}$$

---

[*] The expectation $\mathrm{E}[\underline{u}_i]$ represents a macroscopic motion, while $\underline{u}_i - \mathrm{E}[\underline{u}_i]$ represents fluctuation at a microscopic level. If $\mathrm{E}[\underline{u}_i] \neq 0$, then the macroscopic and microscopic kinetic energies should be treated separately, the latter being $\varepsilon = \mathrm{E}[m_0 (\|\underline{u} - \mathrm{E}[\underline{u}]\|)^2/2]$.

ME applied to the uncertain motion of a particle: results

- By dimensional considerations we define the background measure in terms of universal constants, i.e. the Planck constant h = 6.626×10^{-34} J·s and the proton mass m_p; thus $h(z) = (m_0/\mathrm{h})^3$ $[\mathbf{L}^{-6}\,\mathbf{T}^3]$, thereby giving the entropy as:

$$\Phi[\underline{\boldsymbol{z}}] := \mathrm{E}\left[-\ln\left((\mathrm{h}/m_\mathrm{p})^3 f(\underline{\boldsymbol{z}})\right)\right] = -\int_\Omega \ln\left((\mathrm{h}/m_\mathrm{p})^3 f(\boldsymbol{z})\right) f(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} \qquad (15)$$

- Application of the principle of maximum entropy with constraints (2), (12), (13) and (14) gives the distribution of $\underline{\boldsymbol{z}}$ (see proof in Appendix 1) as:

$$f(\boldsymbol{z}) = (1/a)^3\,(3m_0 / 4\pi\varepsilon)^{3/2} \exp(-3m_0 \|\boldsymbol{u}\|^2 / 4\varepsilon), \;\; 0 \le \underline{x}_i \le a \qquad (16)$$

- The marginal distribution of each of the location coordinates x_i is uniform in [0, a], i.e.,

$$f(x_i) = 1/a, \;\; i = 1, 2, 3 \qquad (17)$$

- The marginal distribution of each of the velocity coordinates u_i is derived as:

$$f(u_i) = (3m_0 / 4\pi\varepsilon)^{1/2} \exp(-3m_0 u_i^2 / 4\varepsilon), \;\; i = 1, 2, 3 \qquad (18)$$

This is Gaussian with mean 0 and variance $2\varepsilon / 3m_0$ = 2 × energy per unit mass per degree of freedom.

ME applied to the uncertain motion of a particle: results (2)

- The marginal distribution of the velocity magnitude $\|\boldsymbol{u}\|$ is:

$$f(\|\boldsymbol{u}\|) = (2/\pi)^{1/2}(3m_0 \,/\, 2\varepsilon)^3 \, \|\boldsymbol{u}\|^2 \exp(-3m_0\|\boldsymbol{u}\|^2/\, 4\varepsilon) \qquad (19)$$

  This is known as the Maxwell–Boltzmann distribution.

- The entropy is then calculated as follows, where e is the base of natural logarithms:

$$\Phi[\underline{\boldsymbol{z}}] = \frac{3}{2}\ln\left(\frac{4\pi\mathrm{e}}{3}\,\frac{m_\mathrm{p}^2}{\mathrm{h}^2\, m_0}\,\varepsilon\, V^{2/3}\right) = \frac{3}{2}\ln\left(\frac{4\pi\mathrm{e}}{3}\,\frac{m_\mathrm{p}^2}{\mathrm{h}^2\, m_0}\right) + \frac{3}{2}\ln\varepsilon + \ln V \qquad (20)$$

- From (16) we readily observe that the joint distribution $f(\boldsymbol{z})$ is a product of functions of $\boldsymbol{z}$'s coordinates x_1, x_2, x_3, u_1, u_2, u_3. This means that all six random variables are jointly independent. The independence results from entropy maximization.
- From (16) and (18) we also observe symmetry with respect to the three velocity coordinates, resulting in uniform distribution of the energy ε into $\varepsilon/3$ for each direction or degree of freedom. This is known as the *equipartition* principle and is again a result of entropy maximization.
- From (20) we can verify that the entropy $\Phi[\underline{\boldsymbol{z}}]$ is a dimensionless quantity.

Extension for many particles

- The coordinates of N identical monoatomic molecules which are in motion in the same cube of volume V form a vector $\underline{\boldsymbol{Z}} = (\underline{\boldsymbol{z}}_1, \ldots, \underline{\boldsymbol{z}}_N)$ with $3N$ location coordinates and $3N$ velocity coordinates.
- If E is the total kinetic energy of the N molecules and $\varepsilon = E/N$ is the energy per particle, then following a similar approach we find the entropy as:

$$\Phi[\underline{\boldsymbol{Z}}] = \frac{3N}{2} \ln\left(\frac{4\pi\mathrm{e}}{3} \frac{m_\mathrm{p}^2}{\mathrm{h}^2 m_0} \varepsilon V^{2/3} \right) = \frac{3N}{2} \ln\left(\frac{4\pi\mathrm{e}}{3} \frac{m_\mathrm{p}^2}{\mathrm{h}^2 m_0} \right) + \frac{3N}{2} \ln \varepsilon + N \ln V \qquad (21)$$

- The equation found in literature, known as the Sackur-Tetrode equation, (after H. M. Tetrode and O. Sackur, who developed it independently at about the same time in 1912) differs from (21) in the last term, which is $N \ln (V/N)$ instead of $N \ln V$.
- To derive the Sackur-Tetrode expression an assumption is made (inspired from quantum physics) that particles are indistinguishable; this assumption is problematic and here is avoided (cf. Koutsoyiannis, 2013).
- The result (21) is fully consistent with the probabilistic character of entropy and also with the thermodynamic content of entropy (Koutsoyiannis, 2013).

Extension to many degrees of freedom

- The number of microscopic degrees of freedom β that can store energy in a particle depend on the particle architecture; thus:
  - A monoatomic molecule has $\beta = 3$ translational degrees of freedom, corresponding to the 3 components of the velocity vector.
  - A diatomic molecule (e.g. of N_2 or O_2 which are the most typical in the atmosphere) has a linear structure; thus, in addition to the kinetic energy it has rotational energy at two axes perpendicular to the line defined by the two atoms; in total it has $\beta = 5$ degrees of freedom.
  - A triatomic (or more complex) molecule has 3 rotational degrees of freedom or $\beta = 6$ degrees of freedom in total.
  - In solids and liquids there are degrees of freedom associated to vibrational energy.
- Generalizing (20) and (21) for β degrees of freedom we obtain the entropy per molecule as:

$$\varphi \coloneqq \Phi[\underline{\boldsymbol{z}}] = c + \frac{\beta}{2} \ln \varepsilon + \ln V \tag{22}$$

where c incorporates all related physical and mathematical constants.

- Likewise, the total entropy of N molecules is:

$$\Phi = N\varphi = \Phi[\underline{\boldsymbol{Z}}] = Nc + \frac{\beta N}{2} \ln \varepsilon + N \ln V = Nc + \frac{\beta N}{2} \ln \frac{E}{N} + N \ln V \tag{23}$$

Definition of internal energy and temperature

- In gases, the internal energy E_I equals the thermal energy ($\varepsilon_\mathrm{I} = \varepsilon$, $E_\mathrm{I} = E$).
- In liquids and solids, the bonds between molecules are associated with dynamic energy; denoting the dynamic energy per molecule as $-\xi$, we write:

$$\varepsilon_\mathrm{I} = \varepsilon - \xi, \quad E_\mathrm{I} = E - N\xi, \tag{24}$$

where $\xi = 0$ for gases and $\xi > 0$ for liquids and solids.

- Temperature* is defined to be the inverse of the partial derivative of entropy with respect to energy†, i.e.,

$$\frac{1}{\theta} := \frac{\partial \Phi}{\partial E_\mathrm{I}} = \frac{\partial \varphi}{\partial \varepsilon_\mathrm{I}} \tag{25}$$

- From(22), (23) and (25) we obtain:

$$\theta = \frac{2\,\varepsilon}{\beta} \tag{26}$$

(temperature = 2 × particle's kinetic energy per degree of freedom).

---

* Since entropy is dimensionless and E_I has dimensions of energy, temperature has also dimensions of energy (J). This contradicts the common practice of using different units of temperature, such as K or °C. To distinguish from the common practice, we use the symbol θ (instead of T which is in K) and we call θ the *natural* temperature (instead of absolute temperature for T).

† The definition is based on the internal energy, but assuming constant ξ, there is no difference if we take the thermal energy instead.

The law of ideal gases

- We consider again the cube of edge a containing N identical molecules of a gas, each with mass m_0 and β degrees of freedom.
- We consider a time interval dt; any particle at distance from the bottom edge $\mathrm{d}x_3 \le -u_3\mathrm{d}t$ will collide with the cube edge ($x_3 = 0$).
- From (19), generalized for β degrees of freedom, the joint distribution function of (x_3, u_3) of a single particle is:

$$f(x_3, u_3) = (1/a)(\beta\, m_0 / 4\pi\, \varepsilon)^{1/2} \exp(-\beta m_0 u_3^2 / 4\varepsilon) \qquad (27)$$

- Thus, the expected value of the momentum $\underline{q}(dt)$ of molecules colliding at the cube edge ($x_3 = 0$) within time interval dt is:

$$\mathrm{E}[\underline{q}(dt)] = N \int_{-\infty}^{\infty} \mathrm{d}x_3 \int_{-\infty}^{-x_3/dt} m_0 u_3 f(x_3, u_3) \mathrm{d}u_3 = N\, \varepsilon\, \mathrm{d}t / \beta a \qquad (28)$$

- According to Newton's 2nd law, the force exerted on the edge is $F = 2\mathrm{E}[\underline{q}(dt)]/\mathrm{d}t$ and the pressure is $p = F / a^2 = 2\, N\, \varepsilon / (\beta\, V)$, or finally (by using (26)),

$$p = N\, \theta / V = \theta / v \Leftrightarrow p\, V = N\, \theta \Leftrightarrow p\, v = \theta \qquad (29)$$

This is the well-known law of ideal gasses written for natural temperature.

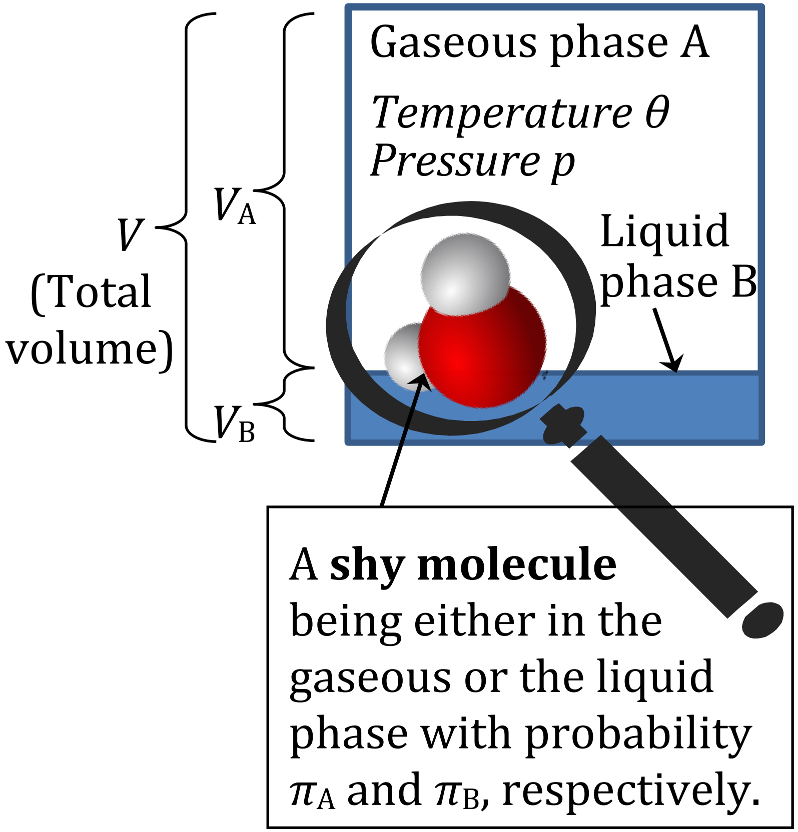
Phase change and the shy molecule

- The law determining the equilibrium of liquid and gaseous phase of water, known as the Clausius-Clapeyron equation is important in hydrology.
- As will be shown next, this law can be derived by maximizing probabilistic entropy, i.e. uncertainty.
- In particular, the law is derived by studying a single molecule which "wishes to hide itself" and, to this aim, it maximizes the combined uncertainty related to:

  (a) its **phase** (whether liquid or gaseous);

  (b) its **position in space**; and

  (c) its **kinetic state**, i.e. its velocity and other coordinates corresponding to its degrees of freedom and making up its thermal energy.



A **shy molecule** being either in the gaseous or the liquid phase with probability π_A and π_B, respectively.

Phase change: setup

- For a molecule to move from the liquid to gaseous phase, an amount of energy ξ to break its bonds with other molecules needs to be supplied (*phase change energy*).
- The partial entropies of the two phases, i.e. the entropies conditional on the particle being in the gaseous (A) or liquid (B) phase, are:

$$\varphi_A = c_A + (\beta_A/2)\ln \varepsilon_A + \ln V_A, \quad \varphi_B = c_B + (\beta_B/2)\ln \varepsilon_B + \ln V_B \tag{30}$$

- The total entropy is (Koutsoyiannis, 2013):

$$\varphi = \pi_A\,\varphi_A + \pi_B\,\varphi_B + \varphi_\pi, \text{ where } \varphi_\pi := -\pi_A \ln \pi_A - \pi_B \ln \pi_B \tag{31}$$

or

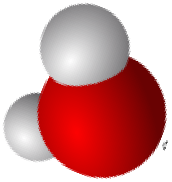
$$\varphi = \pi_A\,(\varphi_A - \ln \pi_A) + \pi_B\,(\varphi_B - \ln \pi_B) \tag{32}$$

- The two phases are in open interaction and the constraints are:

$$\pi_A + \pi_B = 1 \tag{33}$$

$$\pi_A\,\varepsilon_A + \pi_B\,(\varepsilon_B - \xi) = \varepsilon \tag{34}$$

Phase change: assumptions



- **Water vapour** behaves as a perfect gas.
- As its molecule has a 3D (not linear) structure, the rotational energy is distributed into three directions, so that the total number of degrees of freedom (translational and rotational) are:

$$\beta_A = 6 \tag{35}$$

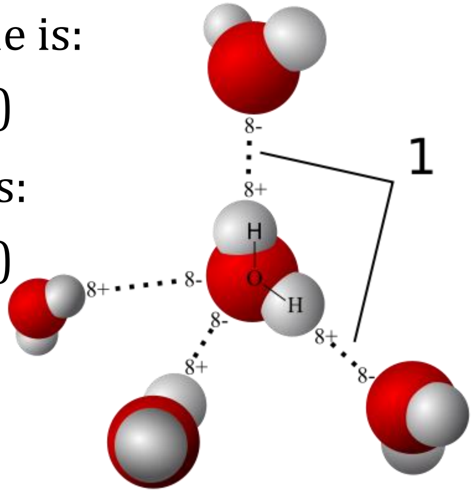
- **Liquid water** is incompressible, so that the volume per particle is:

$$v_B := V_B/N_B = \text{constant} \tag{36}$$

- Hence, if V is the total volume, then that of the gaseous phase is:

$$V_A = V - V_B = V - \pi_B v_B N \tag{37}$$

- The number of degrees of freedom in the liquid phase is greater because of the “social behaviour” of water molecules.
- Specifically, in addition to the translational and rotational degrees of freedom of individual molecules, there are local clusters with low energy vibrational modes that can be thermally excited.



http://en.wikipedia.org/wiki/Water

- The average number of degrees of freedom per molecule (individual and collective involving more than one water molecules) is very high (e.g. Fraundorf, 2003):

$$\beta_B = 18 \tag{38}$$

Phase change: results

- The calculations of entropy maximization are shown in Appendix 2. Their result is:

$$p = \text{constant} \times \mathrm{e}^{-\xi/\theta}\, \theta^{-(\beta_\mathrm{B}/2 - \beta_\mathrm{A}/2 - 1)} \tag{39}$$

- Assuming that at some temperature θ_0, $p(\theta_0) = p_0$, we write (39) in a more convenient and dimensionally consistent manner as:

$$p = p_0\, \mathrm{e}^{\xi/\theta_0 (1 - \theta_0/\theta)} (\theta_0/\theta)^{(\beta_\mathrm{B}/2 - \beta_\mathrm{A}/2 - 1)} \tag{40}$$

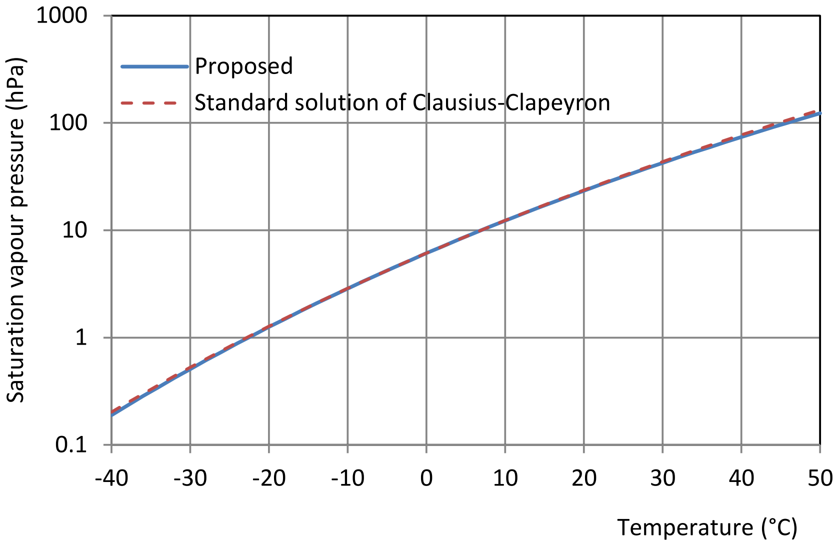
- This is the final form of the proposed equation quantifying phase change.
- The Newton-Raphson method gives the approximation:

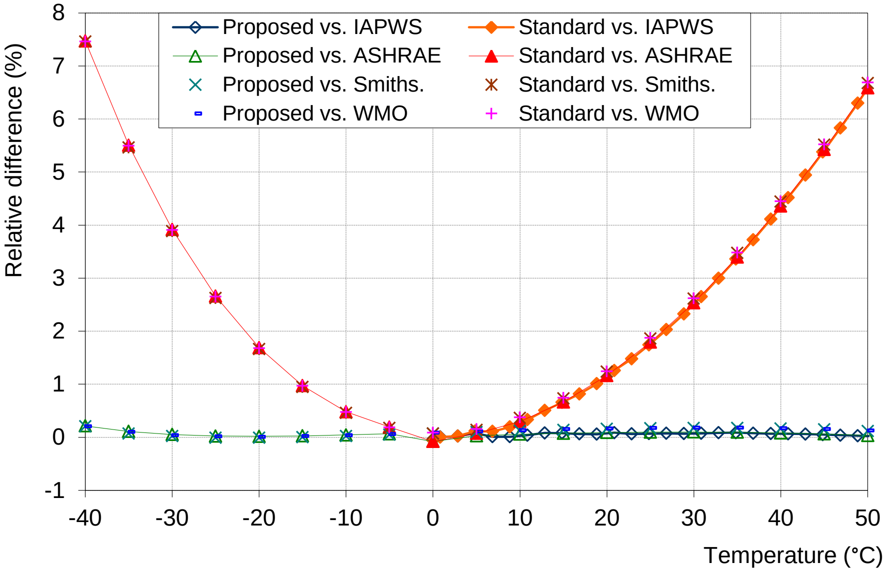
$$p = p_0\, \mathrm{e}^{(\xi/\theta_0 - \beta_\mathrm{B}/2 + \beta_\mathrm{A}/2 + 1)(1 - \theta_0/\theta)} \tag{41}$$

- The latter is the standard solution of the Clausius-Clapeyron equation appearing in books, which however is an inconsistent approximate description of the phenomenon (Koutsoyiannis, 2012).
- Equation (40) can be anchored at the triple point of water, in which θ_0 = 37.714 yJ = 273.16 K, p_0 = 6.11657 hPa (Wagner and Pruss, 2002), while an optimized value of the constant ξ/θ_0 based on accurate measurements is $\xi/\theta_0 = 24.861$.

Saturation vapour pressure: comparisons

- The figure on the left compares the proposed equation (40) with the standard (41); they seem indistinguishable.
- However, the figure on the right, which compares relative differences from measurements, clearly indicates the superiority of (40) derived here.*





- This is an amazing example of how we can derive a deterministic law by maximizing entropy; the key is the huge number of identical elements.

---

[*] A slightly more accurate version, based on experimental values of specific heats, instead of using integral degrees of freedom, can be found in Koutsoyiannis (2012).

Part 3

Application to complex hydrological systems

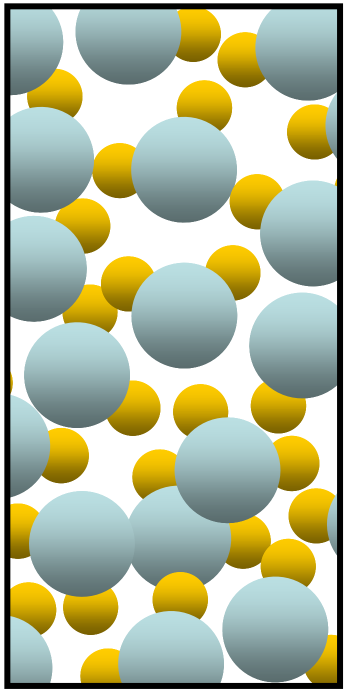
Typical thermodynamic vs. hydrological systems

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 rainfall in India).

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

Macroscopization level 1
(typical thermodynamic system)



Macroscopization level 2

Macroscopization level 3

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 way to deal with the complex macroscopic hydrological systems.
- This contrasts the recent research trend in hydrology (and other disciplines), which invested hopes to the power of computers that would enable faithful and detailed representation of the diverse system elements.
- This research trend was based on the idea that the hydrological processes could be modelled using merely “first principles”, thus resulting in (deterministic) “physically-based” models that would tend to approach in complexity the real world systems.
- The aspiration of detailed and exact deterministic modelling traces a research direction that is wrong and **opposite to the parsimonious way that Nature works**.

Expected difficulties 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 an 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 diversity of 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 inevitably stochastic models** (cf. Montanari and Koutsoyiannis, 2012).

Towards adaptation of the ME framework for nonnegative random variables

- Typically for continuous random variables ranging in $(-\infty,+\infty)$ the Lebesgue measure is used in the entropy function, so that $h(z)$ = constant = $1/[z]$ where $[z]$ denotes the physical unit in which the quantity z is expressed.
- The background measure $h(z)$ determines the way of measuring distances d between values of z; the Lebesgue measure corresponds to the Euclidean distance, $d(z, z') = |z' - z|$.
- Most hydrometeorological variables are non-negative physical quantities unbounded from above (examples: precipitation, streamflow, temperature).
- In positive physical quantities (e.g. rainfall depth) often the Euclidean distance is not a proper metric; sometimes we use a logarithmic distance $d(z, z') = |\ln(z'/ z)|$, as shown in the example.

	Euclidean distance	Logarithmic distance
z = 0.1 mm, z' = 0.2 mm	0.1 mm	ln 2
z = 100 mm, z' = 100.1 mm	0.1 mm	ln 1.001
z = 100 mm, z' = 200 mm	100 mm	ln 2

- Can we merge/unify the Euclidean and logarithmic distance?

Adaptation of the ME framework for nonnegative random variables

- For nonnegative variables we heuristically introduce the generalized background measure:

$$h(z) = \frac{1}{p+z} \qquad (42)$$

  where *p* is a characteristic scale parameter, which also serves as a physical unit for *z*; for $p \to \infty$, $h(x)$ tends to the Lebesgue measure.

- According to this generalized measure, the distance of any point *z* from 0 is:

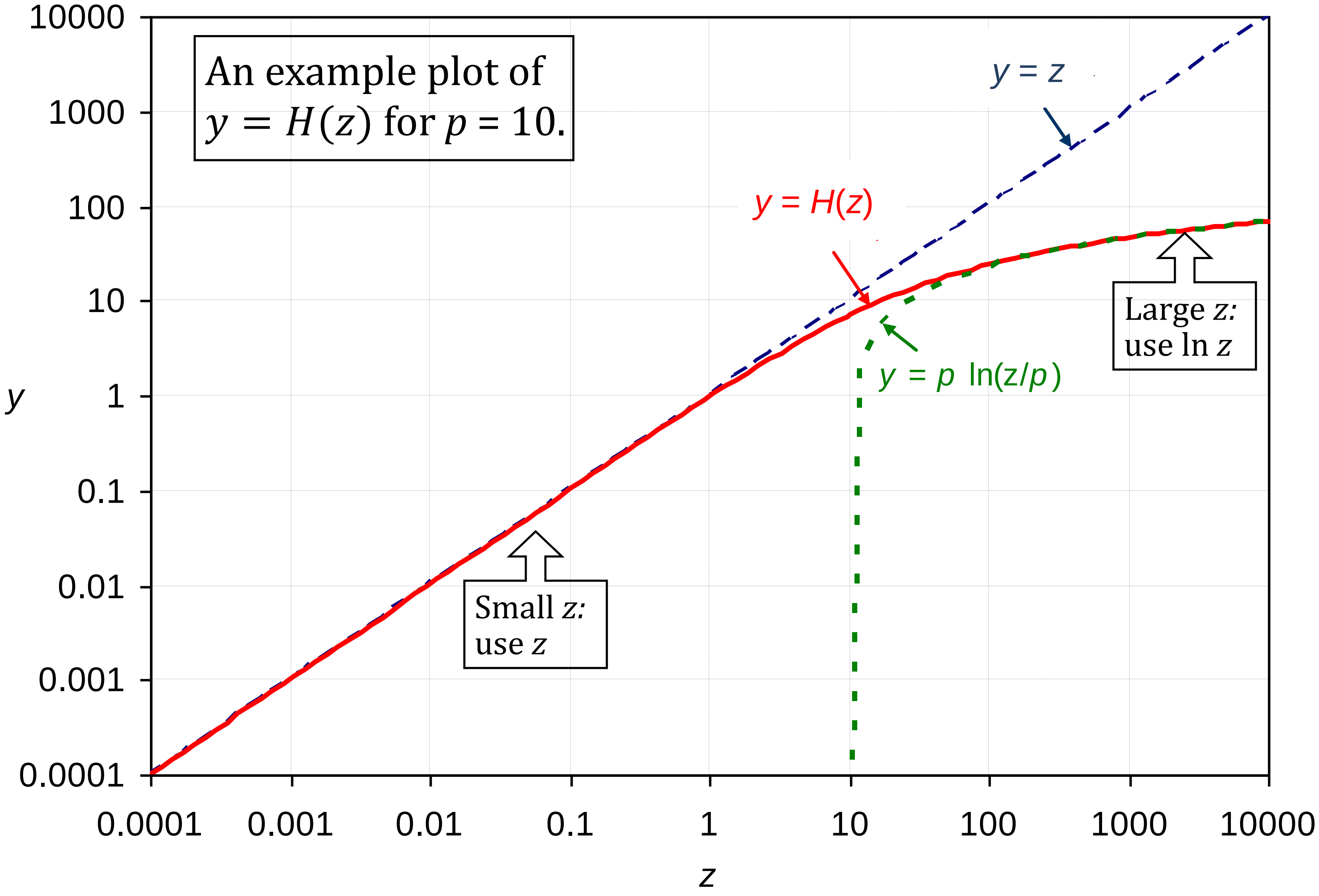
$$H(z) = \int_0^z p\, h(\zeta) \mathrm{d}\zeta = p \ln(1 + z/p) \qquad (43)$$

- Hence, the distance between any two points *z* and *z*′ is:

$$d(z, z') = |H(z') - H(z)| = p \left| \ln \left(\frac{1+z'/p}{1+z/p} \right) \right| \qquad (44)$$

  - For small values, i.e., $z < z' << p$, $d(z, z') = p \ln [1 + (z' - z)/(p + z)] \approx z' - z$ (Euclidean distance).
  - For large values, $p << z < z'$, $d(z, z') \approx p \ln (z'/z)$ (logarithmic distance).
  - Note: $H(z)$ and $d(z, z')$ have the same units as *z* (physical consistency).

Illustration of the distance function $H(z)$



An example plot of $y = H(z)$ for p = 10.

Small z: use z

Large z: use $\ln z$

Simplest case—a single constraint

- It is reasonable to replace constraints of raw moments with those of generalized moments (cf. Papalexiou and Koutsoyiannis, 2012).
- The simplest constraint is the preservation of a "generalized mean", i.e.:

$$\mathrm{E}[H(\underline{z})] = \mathrm{E}[p \ln(1 + z/p)] = m_p \tag{45}$$

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

$$f(z) = A \exp\big((1 - \lambda_1 p) \ln(1 + z/p)\big) = A\,(1 + z/p)^{1-\lambda_1 p} \tag{46}$$

where λ_1 is a Lagrange multiplier and A is such that (2) holds.

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

$$f(z) = (1/\lambda)\,(1 + \kappa\, z/\lambda)^{-1-1/\kappa} \tag{47}$$

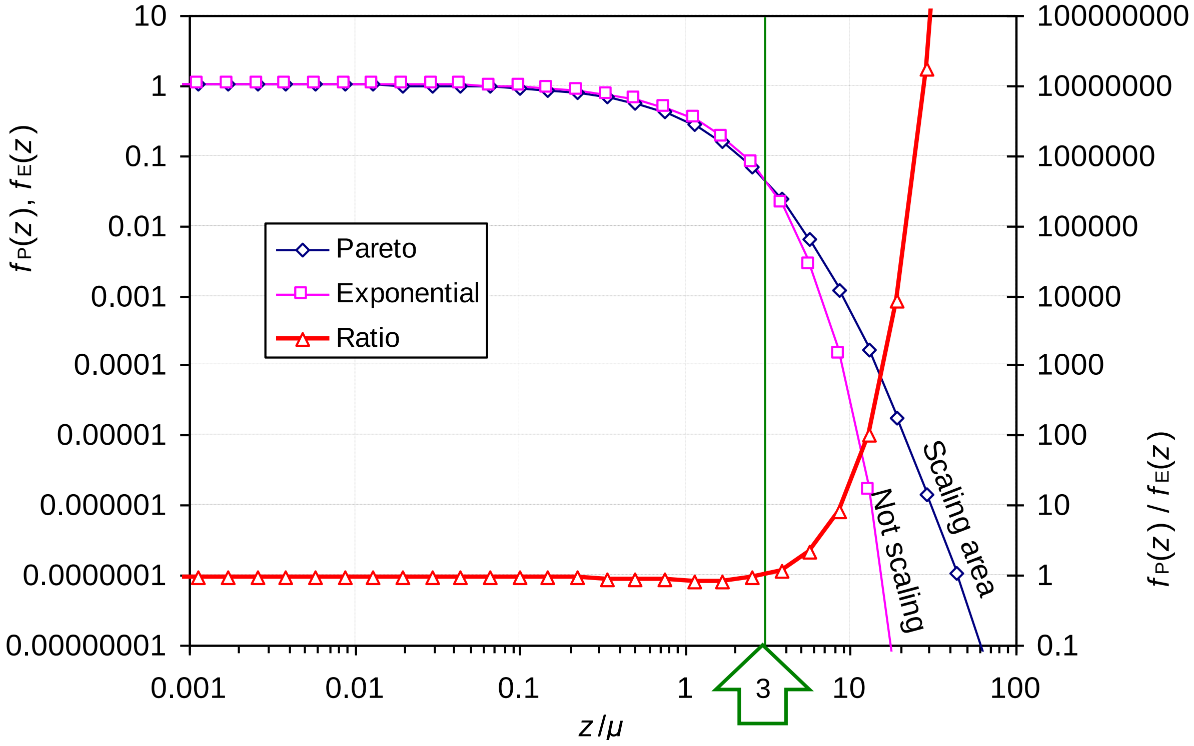
with mean $\mu = \lambda/(1 - \kappa)$, standard deviation $\sigma = \lambda/[(1 - \kappa)\sqrt{1-2\kappa}]$, generalized mean $m_p = \lambda$ and entropy $\Phi[\underline{z}] = \mathrm{E}[-\ln(f(\underline{z})\,(\lambda/\kappa + \underline{z})] = \ln(\mathrm{e}\kappa)$.

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

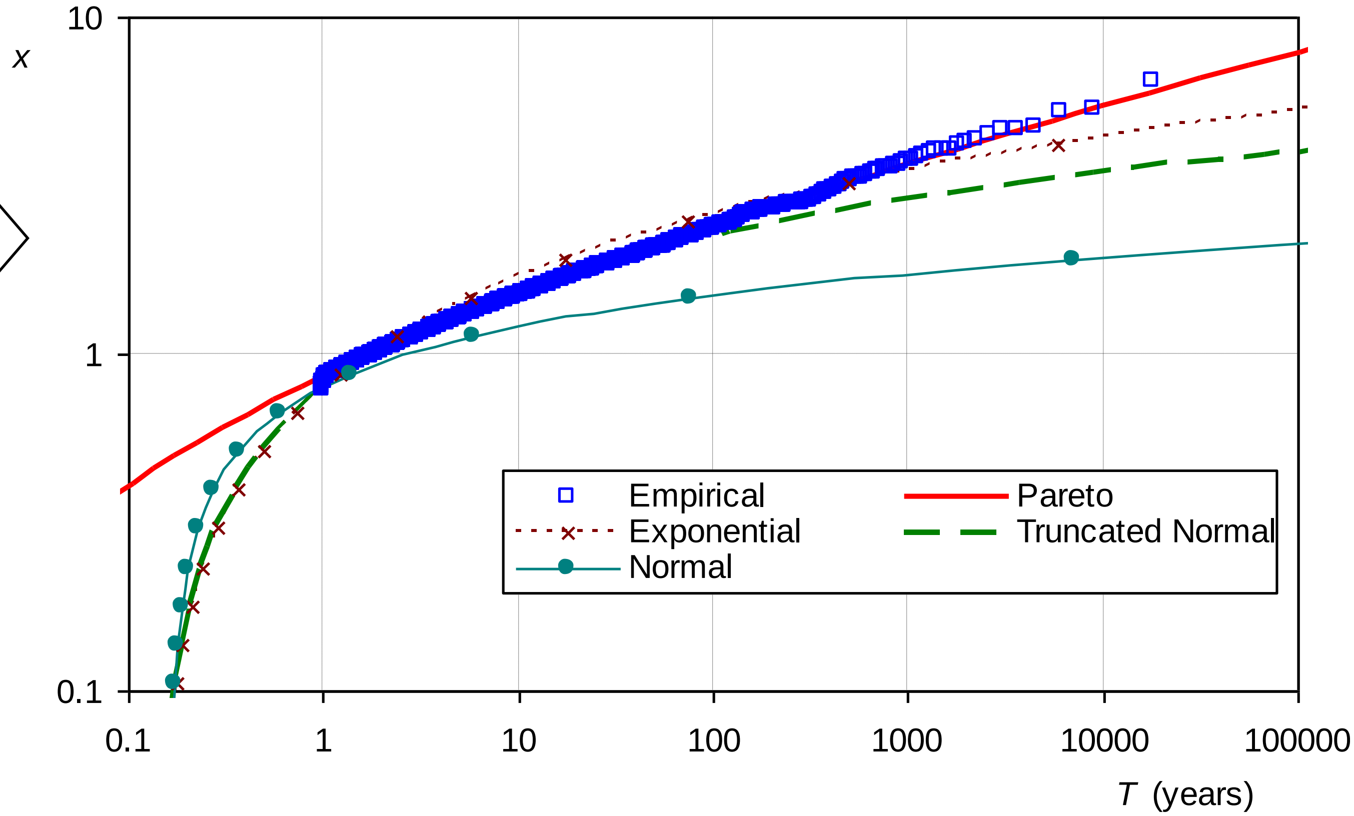
- In classical thermodynamics, a constrained mean results in exponential distribution.
- The two density functions plotted, Pareto, $f_{\mathrm{P}}(z)$, with $\kappa = 0.15$ and $\lambda_{\mathrm{P}} = 0.9$ and exponential, $f_{\mathrm{E}}(z)$, with $\lambda_{\mathrm{E}} = 0.953$ have same $m_p = 0.9$ for $p = \lambda_{\mathrm{P}}/\kappa = 6$.
- Their means are $\mu_{\mathrm{P}} = 1.059 > \mu_{\mathrm{E}} = 0.953$ and their entropies are $\Phi_{\mathrm{P}} = -0.897 > \Phi_{\mathrm{E}} = -\infty$.

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

μ = 0.28
(mean minus
threshold)
σ/μ = 1.19 > 1
ME distribution:
Pareto, κ = 0.15.

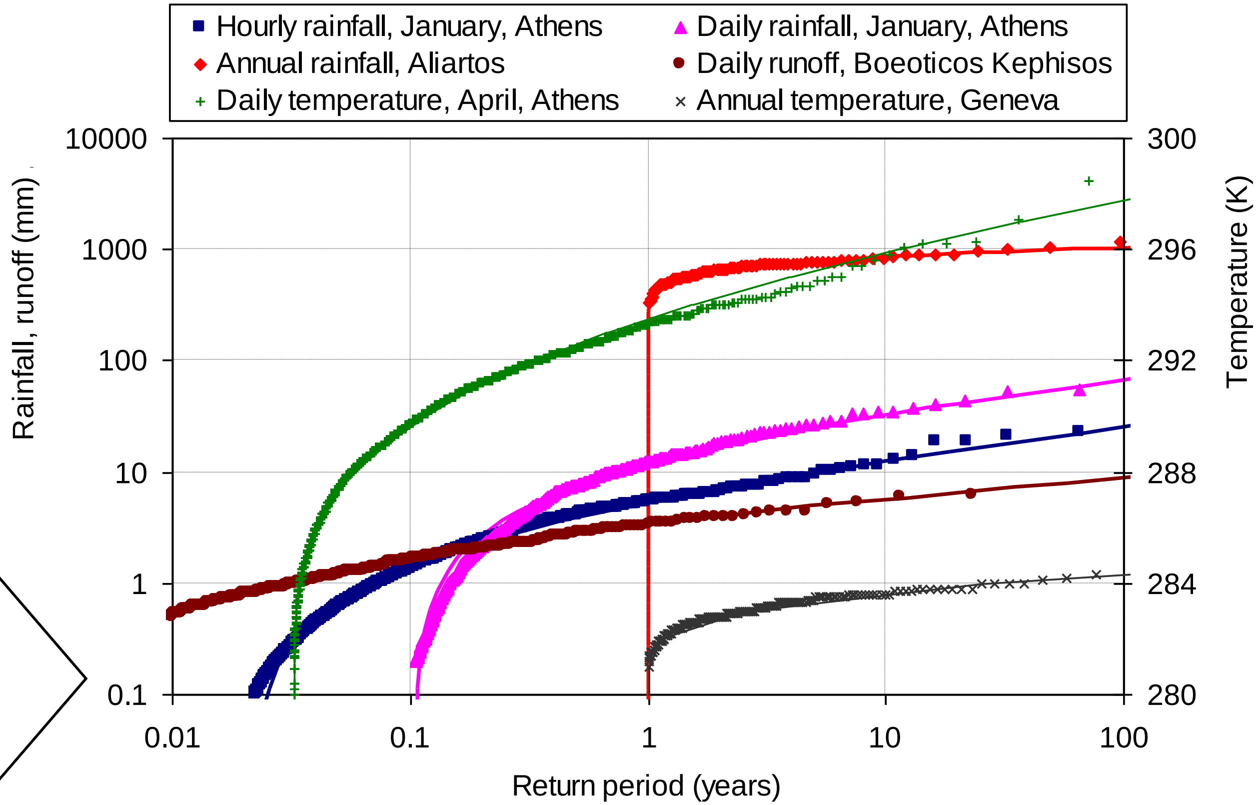


Data set: Daily rainfall from 168 stations worldwide each having at least 100 years of measurements; series above threshold, standardized by mean and merged; period 1822-2002; 17922 station-years of data (Koutsoyiannis, 2005).

Generalization for the marginal distribution of any hydrometeorological variable

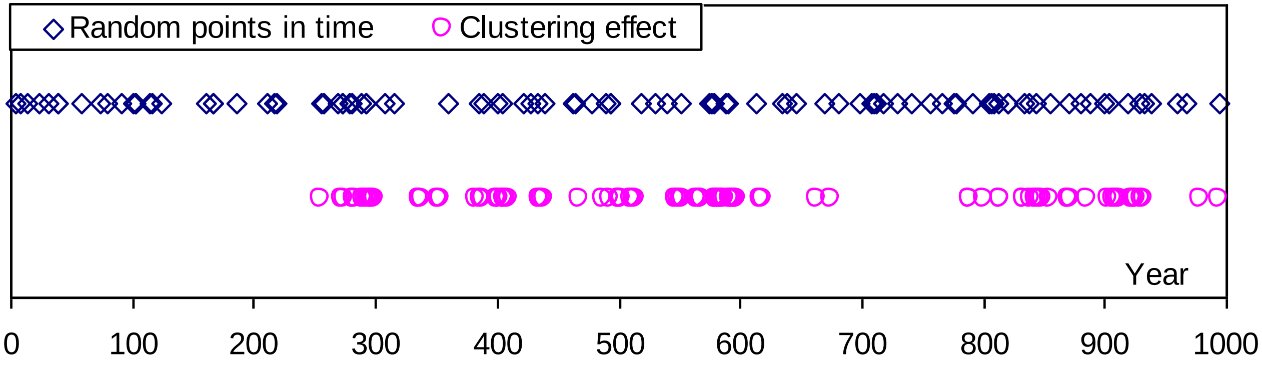
- When the variation is small ($\sigma/\mu < 1$, where μ is the mean and σ the standard deviation), the typical ME framework with constraints μ and σ, gives satisfactory results (although generalized constraints are again superior).
- Generally, the ME framework can give the shape of the distribution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

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

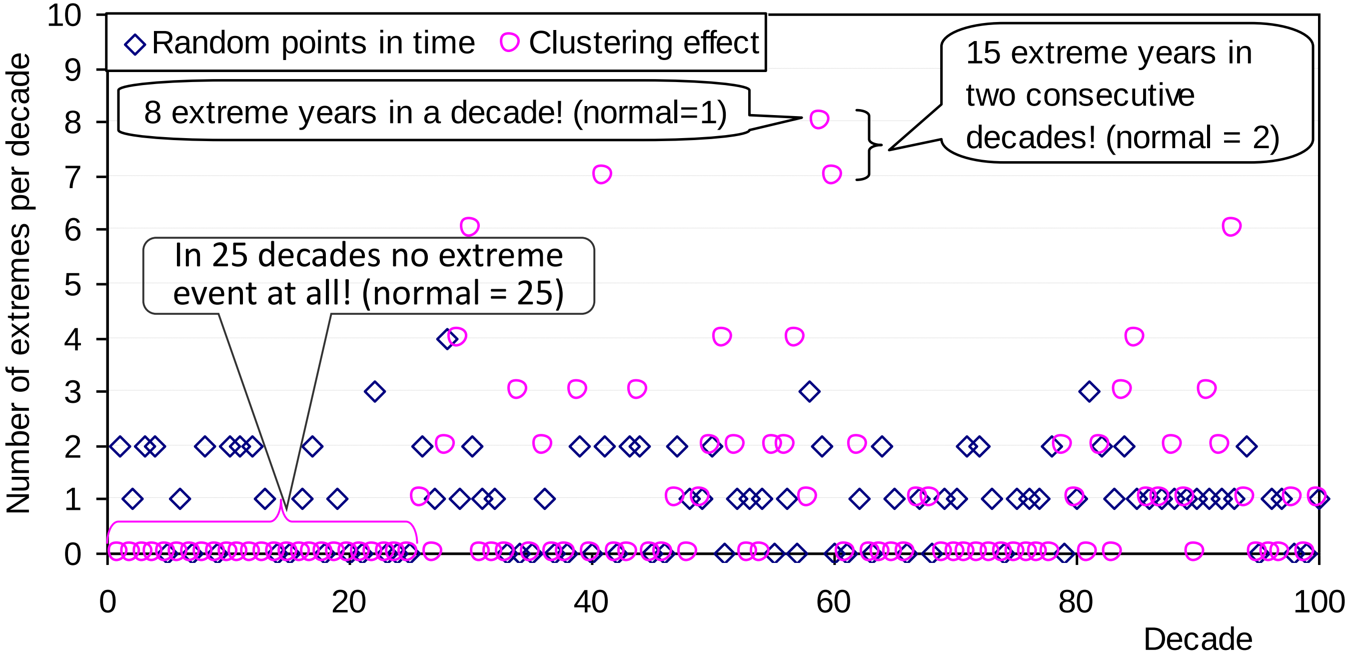


Systems evolving in time: Entropy and clustering

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 $\Phi_C = 1.29 > \Phi_R = 1.23$.



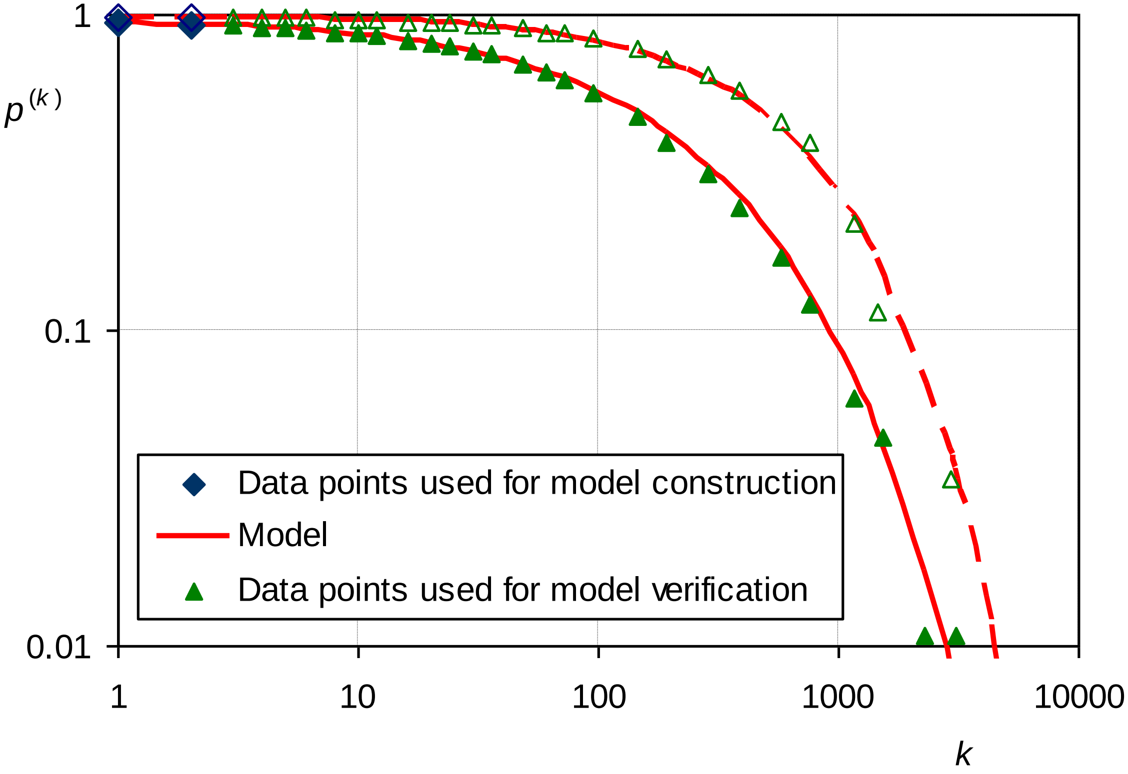
8 extreme years in a decade! (normal=1)

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.
- 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 (necessary for the constraints) give very good predictions for all time scales.



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).

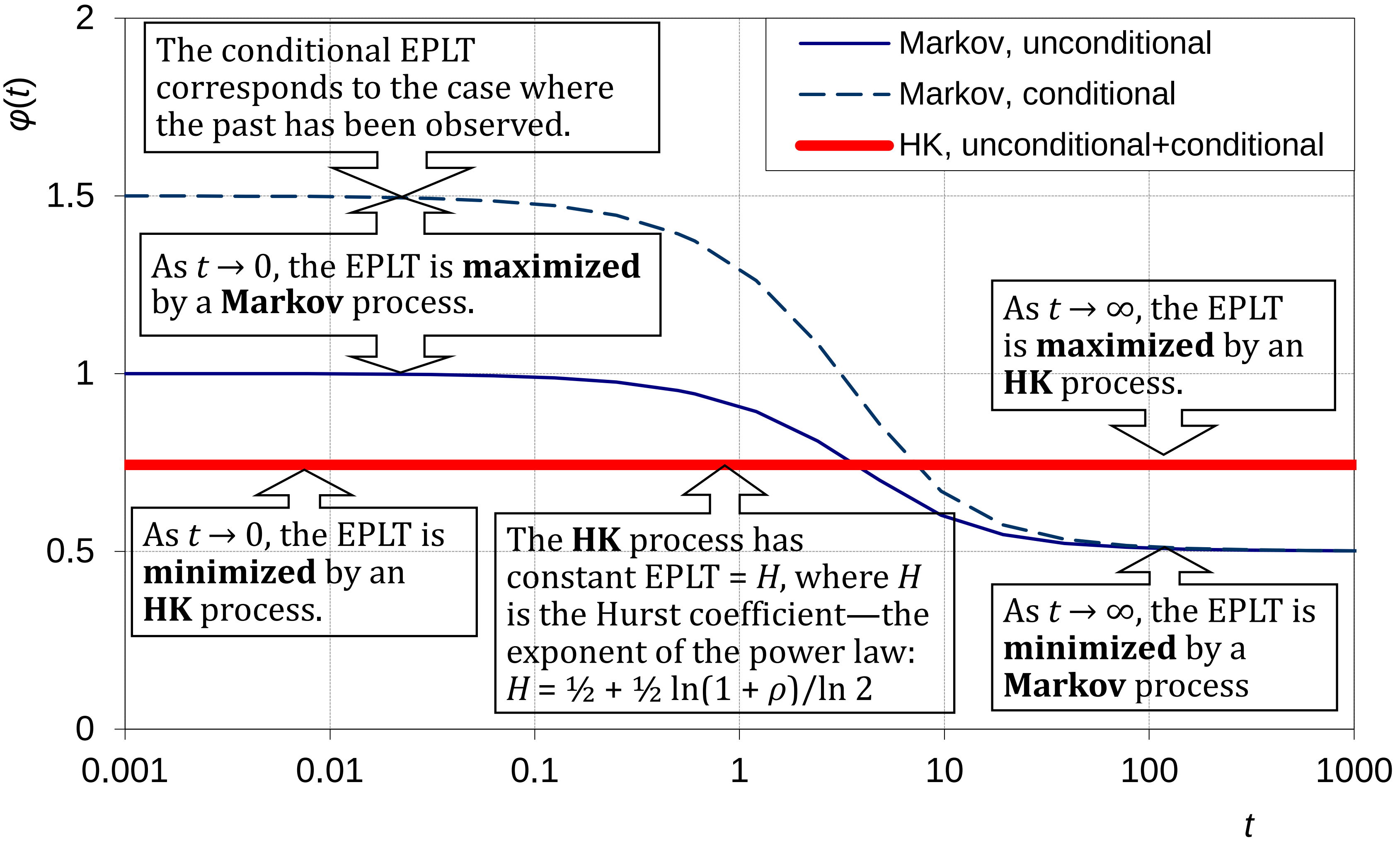
Maximum entropy production and scaling in time

- The dependence structure of processes evolving in time (expressed in terms of autocorrelogram, periodogram or climacogram, which are transformations of each other) can be determined by entropy extremization.
- Koutsoyiannis (2011) suggested the use of entropy production in logarithmic time (EPLT) defined as $\varphi[\underline{z}(t)] \coloneqq \mathrm{d}\Phi[\underline{z}(t)]/\mathrm{d}(\ln t)$, with $\underline{z}(t)$ being a cumulative stochastic process.
- The specific assumptions are:
  - Lebesgue background measure (assumption good for $\sigma/\mu << 1$) and constrained mean μ and variance σ^2; these result in Gaussian marginal distribution, hence (see slide 11):

    $\Phi[\underline{z}(t)] = (1/2)\ln[2\pi\mathrm{e}\,\gamma(t)]$ where $\gamma(t) := \mathrm{Var}\,[\underline{z}(t)]$;
  - 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).

Maximum entropy production and scaling in time (contd.)

The solutions depicted are generic, valid for any Gaussian process, independent of μ and σ, and depended on ρ only (the example is for $\rho = 0.4$)—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.

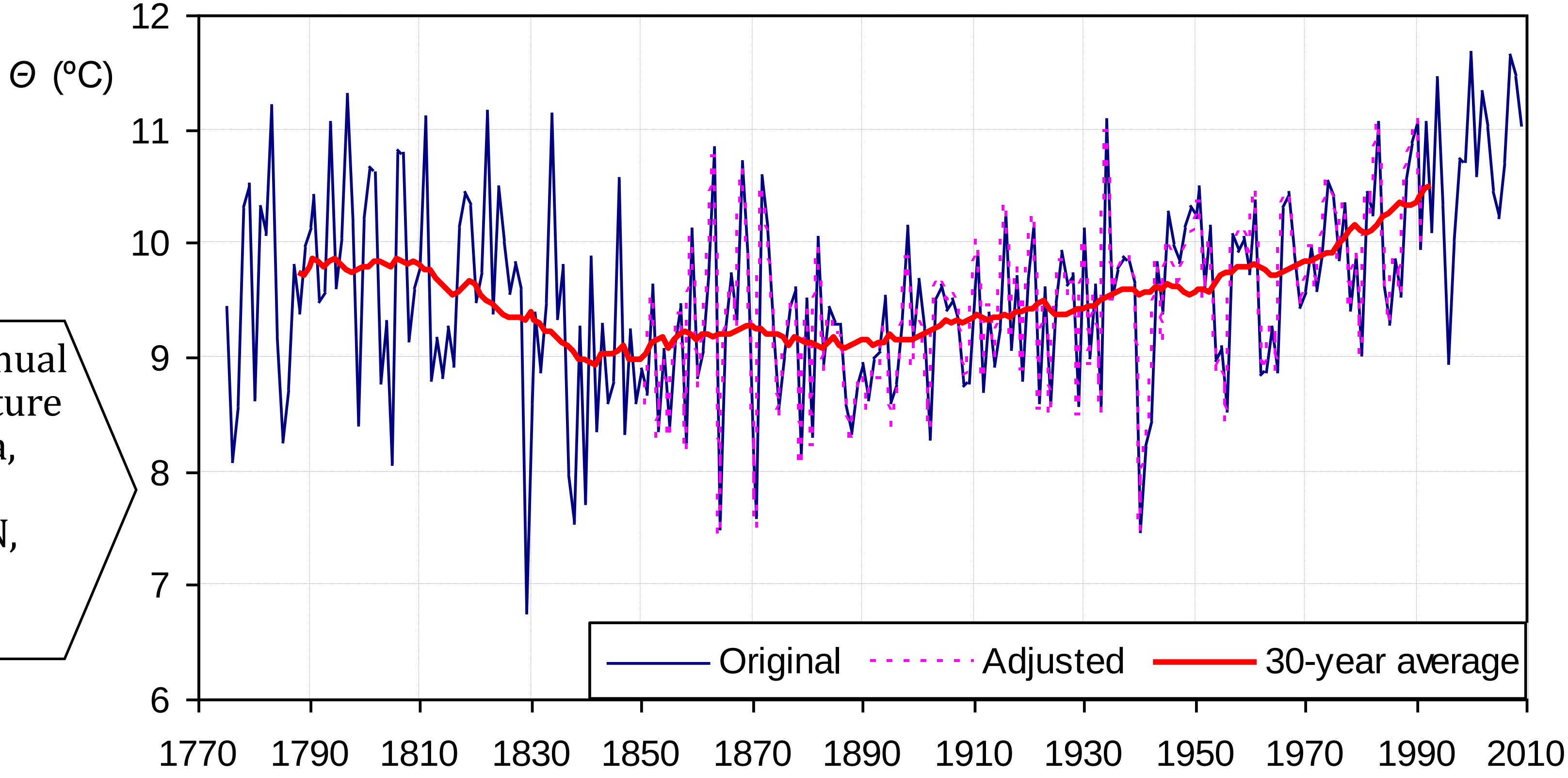
As $t \to 0$, the EPLT is **minimized** by an **HK** process.

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).



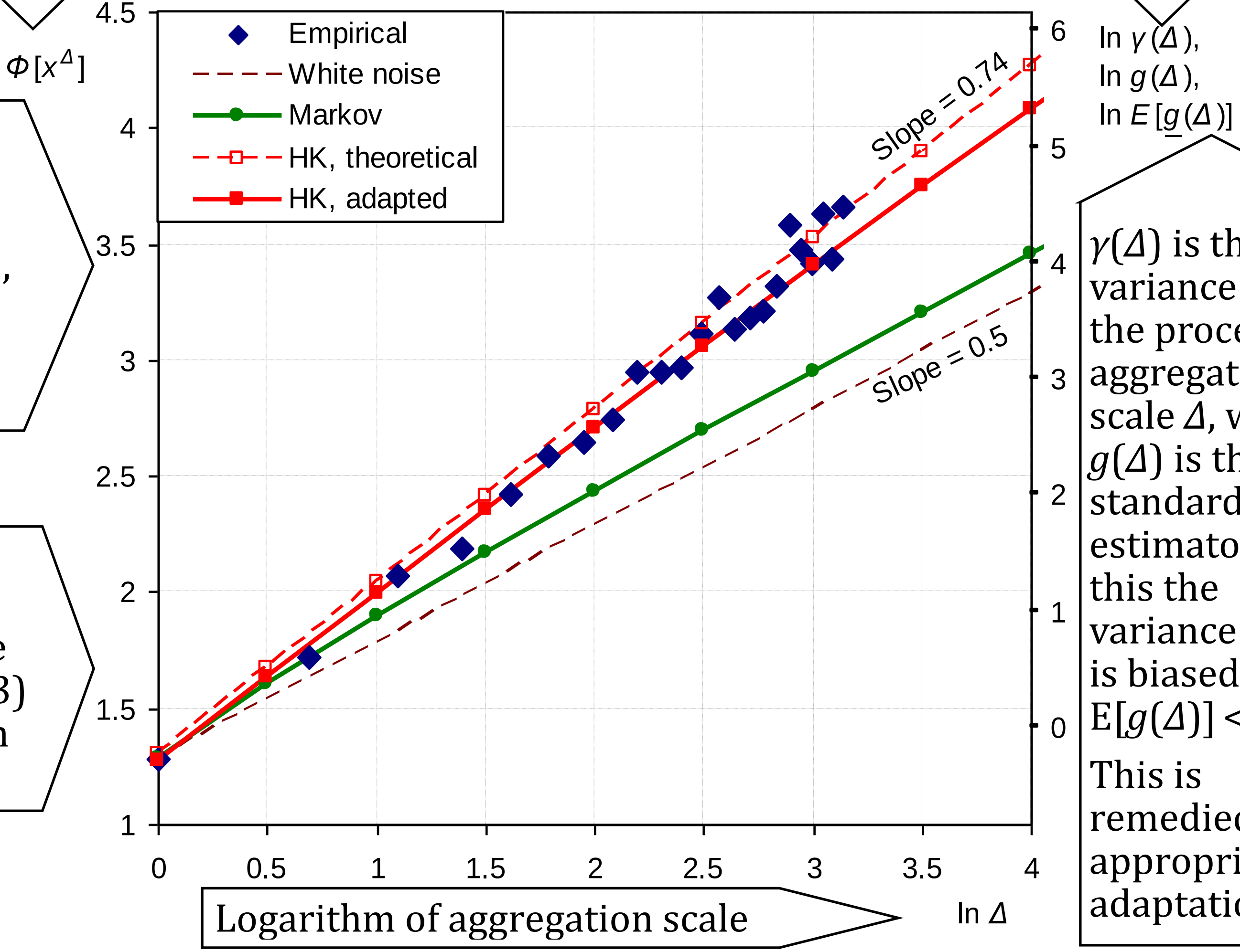
Data set: 235 years of annual temperature (1775–2009; one of the longest available instrumental geophysical records) available from the climexp.knmi.nl, partly included in the Global Historical Climatology Network (GHCN; 1851–1991); from Koutsoyiannis (2011).

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 or J), 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).



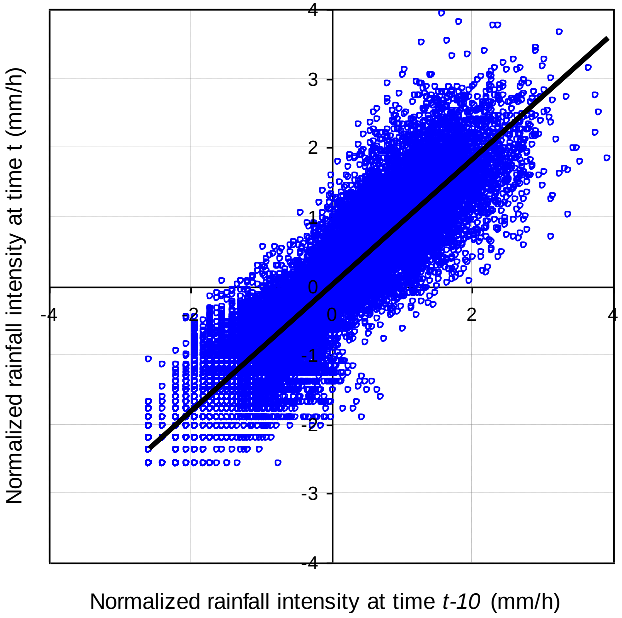
Logarithm of aggregation scale

$\gamma(\Delta)$ is the variance of the process aggregated at scale Δ, while $g(\Delta)$ is the standard estimator of this the variance and is biased: $\mathrm{E}[g(\Delta)] < \gamma(\Delta)$

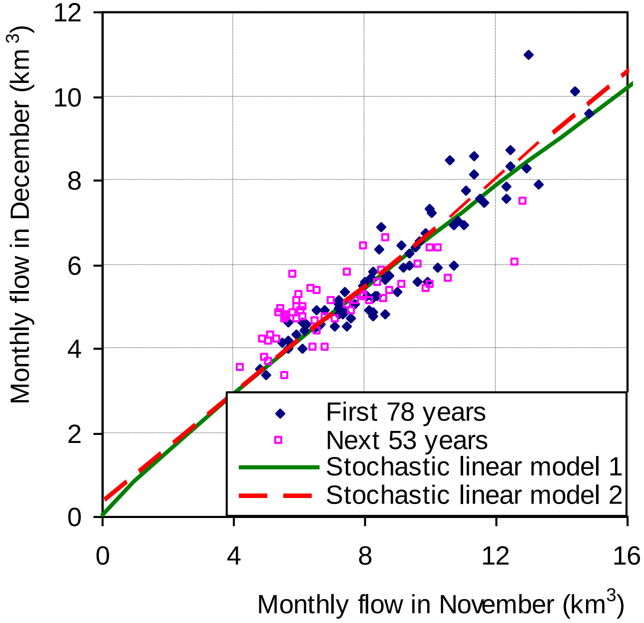
This is remedied by appropriate adaptation.

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:

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

  where $\boldsymbol{a}$ is a vector of parameters (T denotes transpose) and $\underline{v}$ is the prediction error, assumed to be 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:

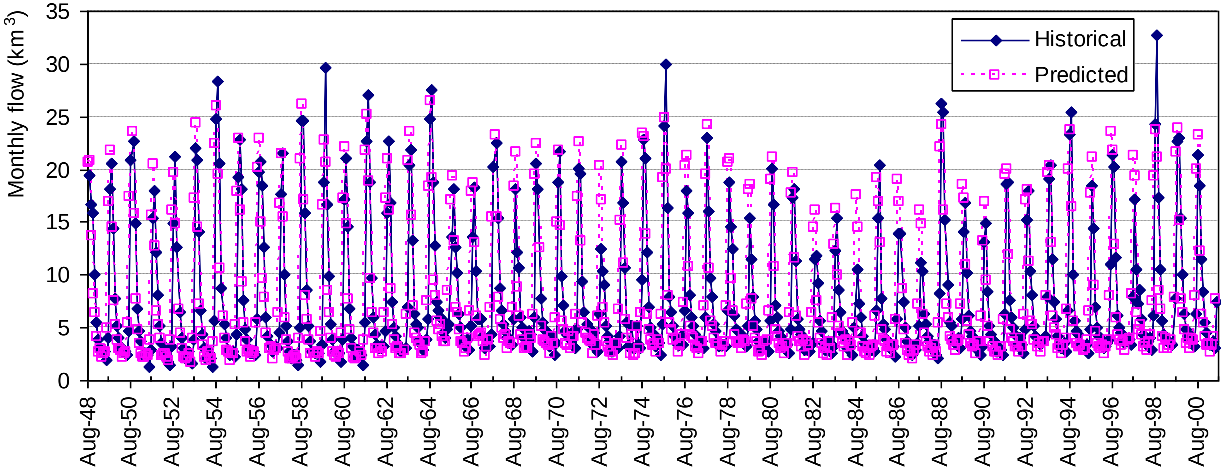
  $$\boldsymbol{a}^{\mathrm{T}} = \boldsymbol{\eta}^{\mathrm{T}}\, \boldsymbol{h}^{-1}, \quad \mathrm{Var}[\underline{v}] = 1 - \boldsymbol{\eta}^{\mathrm{T}}\, \boldsymbol{h}^{-1} \boldsymbol{\eta} = 1 - \mathbf{a}^{T}\, \boldsymbol{\eta}$$

  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 (contd.)

- 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 (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).



Concluding remarks

- Entropy is none other than uncertainty quantified.
- The tendency of entropy to become maximal is not a curse—it is blessing.
- This tendency constitutes the driving force of change and evolution; also, it offers the basis to understand and describe Nature.
- By maximizing entropy, i.e. uncertainty, we can describe the behaviour of physical systems; such description is essentially probabilistic.
- However, if a system is composed of numerous identical elements, the uncertainty, despite being maximal at the microscopic level, in the macroscopic system it becomes as low as to yield a physical law that is in effect deterministic; this is the case in the equilibrium of liquid water and water vapour (Clausius-Clapeyron equation).
- Extremal entropy considerations provide a theoretical basis in modelling hydrological processes; however, at high macroscopization levels there is no hope to derive deterministic laws; only stochastic modelling is feasible.
- Linking statistical thermophysics with hydrology with a unifying view of entropy as uncertainty is a promising scientific direction.
- Uncertainty and entropy are not enemies of science that should be eliminated; they are just important objects to be studied and understood.

Appendix 1: Proof of equations (16)-(20)

- According to (10) and taking into account the equality constraints (2), (13) and (14), the ME distribution will have density:

$$f(\boldsymbol{z}) = \exp\left(-\lambda_0 - \lambda_1 u_1 - \lambda_2 u_2 - \lambda_3 u_3 - \lambda_4(u_1^2 + u_2^2 + u_3^2)\right) \qquad (48)$$

- This proves that the density will be an exponential function of a second order polynomial of (u_1, u_2, u_3) involving no products of different u_i. The $f(\boldsymbol{z})$ in (16) is of this type, and thus it suffices to show that it satisfies the constraints.
- Note that the inequality constraint (12) is not considered at this phase but only in the integration to evaluate the constraints. That is, the integration domain will be $\Omega := \{(0 \le x_1 \le a, 0 \le x_2 \le a, 0 \le x_3 \le a, -\infty < u_1 < \infty, -\infty < u_2 < \infty, -\infty < u_3 < \infty)\}$. We denote by $\int_\Omega \mathrm{d}\boldsymbol{z}$ the integral over this domain. It is easy then to show (the calculation of integrals is trivial) that:

$$\int_\Omega f(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} = 1;\ \int_\Omega u_1 f(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} = 0;\ \int_\Omega u_2 f(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} = 0;\ \int_\Omega u_3 f(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} = 0;\ \int_\Omega (u_1^2 + u_2^2 + u_3^2) f(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} = 2\varepsilon/m_0 \qquad (49)$$

- Thus, all constraints are satisfied. To find the marginal distribution of each of the variables we integrate over the entire domain of the remaining variables; due to independence this is very easy and the results are given in (17) and (18). To find the marginal distribution of $\|\boldsymbol{u}\|$ (eqn. (19)), we recall that the sum of squares of n independent $N(0, 1)$ random variables has a $\chi^2(n)$ distribution (Papoulis, 1990, p. 219, 221); then we use known results for the density of a transformation of a random variable (Papoulis, 1990, p. 118) to obtain the distribution of the square root, thus obtaining (19).
- To calculate the entropy, we observe that $-\ln[f(\boldsymbol{z})] = (3/2)\ln(4\pi\varepsilon/3m_0) + \ln a^3 + 3m_0(u_1^2 + u_2^2 + u_3^2) / 4\varepsilon$ and $\ln h(z) = 3 \ln (m_0/\mathrm{h})$. Thus, the entropy, whose final form is given in (20), is derived as follows:

$$\Phi[\underline{\boldsymbol{z}}] = \int_\Omega (-\ln f(\boldsymbol{z}) + \ln h(z))\, f(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} = (3/2)\ln\left((4\pi\varepsilon/3m_0\,(m_0/\mathrm{h})^2)\right) + \ln a^3 + (3m_0/4\varepsilon)(2\varepsilon/m_0) = (3/2)\ln(4\pi m_0/3\,\mathrm{h}^2) + (3/2)\ln\varepsilon + \ln V + 3/2 \qquad (50)$$

Appendix 2: Proof of (39)

- Based on assumptions (36) and using (37), (30) becomes:

$$\varphi_\mathrm{A} = c_\mathrm{A} + (\beta_\mathrm{A}/2)\ln \varepsilon_\mathrm{A} + \ln (V - \pi_\mathrm{B} v_\mathrm{B} N), \quad \varphi_\mathrm{B} = c_\mathrm{B} + (\beta_\mathrm{B}/2)\ln \varepsilon_\mathrm{B} + \ln (\pi_\mathrm{B} v_\mathrm{B} N) \tag{51}$$

- We wish to find the conditions which maximize the entropy φ in (31) under constraints (34) and (33) with unknowns ε_A, ε_B, π_A, π_B. We form the function ψ incorporating the total entropy φ as well as the two constraints with Langrage multipliers κ and λ:

$$\psi = \pi_\mathrm{A}(\varphi_\mathrm{A} - \ln \pi_\mathrm{A}) + \pi_\mathrm{B}(\varphi_\mathrm{B} - \ln \pi_\mathrm{B}) + \kappa\,(\pi_\mathrm{A}\,\varepsilon_\mathrm{A} + \pi_\mathrm{B}\,(\varepsilon_\mathrm{B} - \xi) - \varepsilon) + \lambda\,(\pi_\mathrm{A} + \pi_\mathrm{B} - 1) \tag{52}$$

- To maximize ψ, equating to 0 the derivatives with respect to ε_A and ε_B, we obtain:

$$\frac{\partial\psi}{\partial\varepsilon_\mathrm{A}} = \frac{\pi_\mathrm{A}\beta_\mathrm{A}}{2\varepsilon_\mathrm{A}} + \kappa\,\pi_\mathrm{A} = 0, \quad \frac{\partial\psi}{\partial\varepsilon_\mathrm{B}} = \frac{\pi_\mathrm{B}\beta_\mathrm{B}}{2\varepsilon_\mathrm{B}} + \kappa\,\pi_\mathrm{B} = 0 \tag{53}$$

- By virtue of (26), this obviously results in equal temperature θ in the two phases, i.e.,

$$\kappa = -\frac{\beta_\mathrm{A}}{2\varepsilon_\mathrm{A}} = -\frac{\beta_\mathrm{B}}{2\varepsilon_\mathrm{B}} = -\frac{1}{\theta_\mathrm{A}} = -\frac{1}{\theta_\mathrm{B}} = -\frac{1}{\theta} \tag{54}$$

- Equating to 0 the derivatives with respect to π_A and π_B, we obtain:

$$\begin{gathered}\frac{\partial\psi}{\partial\pi_\mathrm{A}} = \varphi_\mathrm{A} - \ln \pi_\mathrm{A} - 1 + \kappa\,\varepsilon_\mathrm{A} + \lambda = 0, \\ \frac{\partial\psi}{\partial\pi_\mathrm{B}} = \frac{-\pi_\mathrm{A} v_\mathrm{B} N}{V - \pi_\mathrm{B} v_\mathrm{B} N} + \varphi_\mathrm{B} - \ln \pi_\mathrm{B} - 1 + 1 + \kappa\,(\varepsilon_\mathrm{B} - \xi) + \lambda = 0\end{gathered} \tag{55}$$

- It can be seen that the first term of $\partial\psi/\partial\pi_\mathrm{B}$ equals $-v_\mathrm{B}/v_\mathrm{A}$ and is negligible since $v_\mathrm{B} \ll v_\mathrm{A}$.

Appendix 2: Proof of (39) (contd.)

- After eliminating λ, substituting κ from (54), and making algebraic manipulations, we get:

$$(\varphi_A - \ln \pi_A) - (\varphi_B - \ln \pi_B) = \xi/\theta - (\beta_B/2 - \beta_A/2 - 1) \qquad (56)$$

- On the other hand, from (30), the entropy difference is:

$$(\varphi_A - \ln \pi_A) - (\varphi_B - \ln \pi_B) = c_A + (\beta_A/2) \ln \varepsilon_A + \ln v_A - c_B - (\beta_B/2) \ln \varepsilon_B - \ln v_B \qquad (57)$$

- Substituting θ for ε_A and ε_B from (54) and using the ideal gas to express v_A in terms of θ and p we obtain:

$$(\varphi_A - \ln \pi_A) - (\varphi_B - \ln \pi_B) = -(\beta_B/2 - \beta_A/2 - 1) \ln \theta - \ln p + \text{constant} \qquad (58)$$

- Combining (56) and (58), and eliminating $(\varphi_A - \ln \pi_A) - (\varphi_B - \ln \pi_B)$, we find (39).

References

- Atkins, P., *Four Laws that Drive the Universe*, Oxford Univ. Press, 131 pp., 2007.
- Ben-Naim, A., *A Farewell to Entropy: Statistical Thermodynamics Based on Information*, World Scientific Pub. Singapore, 384 pp., 2008.
- Fraundorf, P., Heat capacity in bits, *American Journal of Physics*, 71 (11), 1142-1151, 2003.
- Hemelrijk, J., Underlining random variables, *Statistica Neerlandica*, 20 (1), 1966.
- Hurst, H. E., Long term storage capacities of reservoirs, *Trans. Am. Soc. Civil Engrs.*, 116, 776–808, 1951.
- Jaynes, E. T., Information theory and statistical mechanics, *Physical Review*, 106 (4), 620-630, 1957.
- Jaynes, E. T., *Probability Theory: The Logic of Science*, Cambridge Univ. Press, 728 pp., 2003.
- Kolmogorov, A. N., Wiener spirals and some other interesting curves in a Hilbert space [Wienersche Spiralen und einige andere interessante Kurven in Hilbertschen Raum], *Dokl. Akad. Nauk URSS*, 26, 115–118, 1940 (English translation in: Tikhomirov, V.M. (ed.) *Selected Works of A. N. Kolmogorov: Mathematics and mechanics*, Vol. 1, Springer, 324-326, 1991).
- 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., Clausius-Clapeyron equation and saturation vapour pressure: simple theory reconciled with practice, *European Journal of Physics*, 33 (2), 295–305, 2012.
- Koutsoyiannis, D., Physics of uncertainty, the Gibbs paradox and indistinguishable particles, *Studies in History and Philosophy of Modern Physics*, 2013 [in review].
- 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.
- Montanari, A., and D. Koutsoyiannis, A blueprint for process-based modeling of uncertain hydrological systems, Water Resources Research, 48, W09555, doi:10.1029/2011WR011412, 2012.
- Papalexiou, S. M., and D. Koutsoyiannis, Entropy based derivation of probability distributions: A case study to daily rainfall, *Advances in Water Resources*, 45, 51–57, 2012.
- Papalexiou, S.M., D. Koutsoyiannis, and A. Montanari, Can a simple stochastic model generate rich patterns of rainfall events?, *Journal of Hydrology*, 411 (3-4), 279–289, 2011.
- Papoulis, A., *Probability and Statistics*, Prentice-Hall, 1990.
- Papoulis, A., *Probability, Random Variables and Stochastic Processes*, 3rd edn., McGraw-Hill, New York, 1991.
- Shannon, C. E., The mathematical theory of communication, *Bell System Technical Journal*, 27 (3), 379–423, 1948.
- Uffink, J., Can the maximum entropy principle be explained as a consistency requirement?, *Studies In History and Philosophy of Modern Physics*, 26 (3), 223-261, 1995.
- Wagner, W., and A. Pruss, The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use, *J. Phys. Chem. Ref. Data*, 31, 387-535, 2002.