AGU FALL MEETING

San Francisco | 15–19 December 2014

Random musings on stochastics

(Lorenz Lecture)

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

Presentation available online: www.itia.ntua.gr/1500/

Introductory note on the lower part of the title page


School of Civil Engineering
National Technical University of Athens, Greece
(dk@itia.ntua.gr, http://www.itia.ntua.gr/dk/)
Presentation available online: www.itia.ntua.gr/1500/

Conditions of current operation of NTUA's School of Civil Engineering

- **Rank #28 in the world and #7 in Europe** among Civil Engineering Schools according to QS World University Rankings [*]; but other statistics are not good:
- **1850 students**—may see an obscure future: **youth unemployment rate = 58%**.[†]
- **25% fewer professors** (no appointments of young professors after retirements); may **climb to 50%** in the next 5-6 years.
- **50% dismissal** of administrative and technical personnel.
- **40% reduction** in salaries.
- **90% reduction** in School's budget.
- **50% increase** of students' admissions (as a result of government's social policy).
- A "reform" imposed by the government (opposite to the former democratic/ participatory organization of the university) contributed to chaos in our operation.

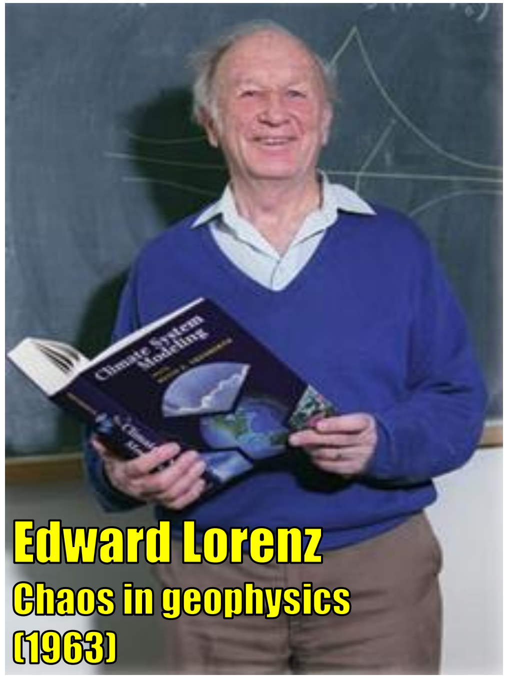
---

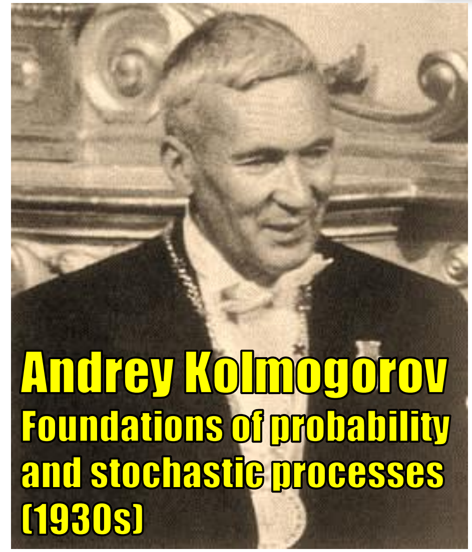
[*] www.topuniversities.com/university-rankings/university-subject-rankings/2014/engineering-civil-structural
[†] epp.eurostat.ec.europa.eu/statistics_explained/index.php/Unemployment_statistics

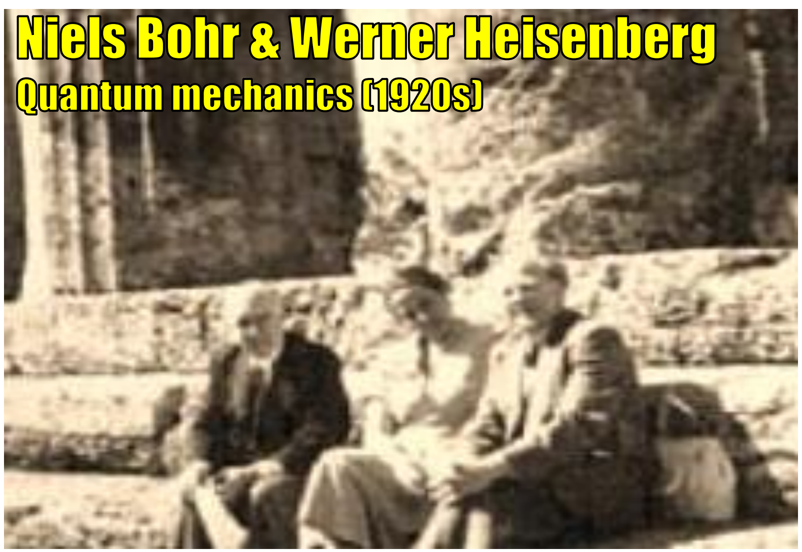
Introductory note on the subtitle

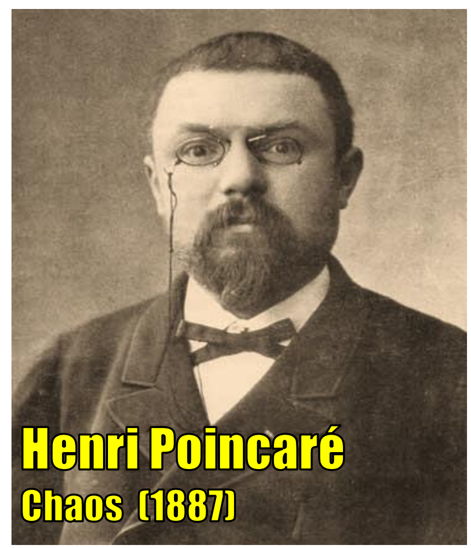
(Lorenz Lecture)

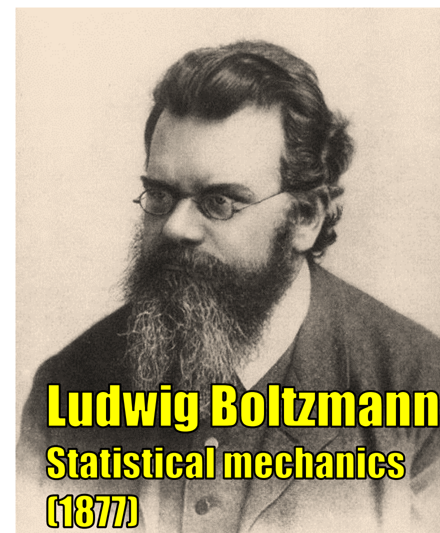
From Lorenz (2007)—A letter for the opening of the conference *“20 Years of Nonlinear Dynamics in Geosciences”* (Rhodes, Greece, 2006): *“Why 20 years, rather than something closer to 200?”*











Introductory note on the title

Almost synonymous

Random musings on stochastics

Same etymology, similar colloquial meaning

Τυχαίοι στοχασμοί στη στοχαστική

Stochastic stochastics on stochastics

The meaning of randomness and stochastics

Deterministic world view	**Indeterministic world view**
Sharp exactness	Uncertainty
	Random = unpredictable, uncertain
Regular variable x: it represents a number	Random variable, $\underline{x}$: an abstract mathematical entity whose realizations x belong to a set of possible numerical values. $\underline{x}$ is associated with a probability density (or mass) function $f(x)$. A random variable $\underline{x}$ becomes identical to a regular variable x only if $f(x) = \delta(x)$ (Dirac function).
Trajectory $x(t)$: the sequence of a system's states x as time t changes	Stochastic process $\underline{x}(t)$: A collection of (usually infinitely many) random variables $\underline{x}$ indexed by t (typically representing time). It represents the evolution of some uncertain system over time. A realization (sample) $x(t)$ of $\underline{x}(t)$ is a trajectory; if it is known at certain points t_i it is a time series.
	Stochastics: The mathematics of random variables and stochastic processes. Stochastics = probability theory + statistics + stochastic processes

Part 1: The meaning of nonlinearity: Stochastic vs. deterministic perspective

Vit Klemes checking his data for nonlinearities

“The first thing to do was to check the data for nonlinearities and get rid of them by a proper transformation. Contemplating which might be the most appropriate one in this case, the scene of my last inspiration came to mind. I saw myself sitting in what could be justly regarded as ‘log space’ — and that led me to use the log-transformation, so popular in hydrology and beyond.”

From Klemes (2007): “An unorthodox physically-based stochastic treatment of tree rings”

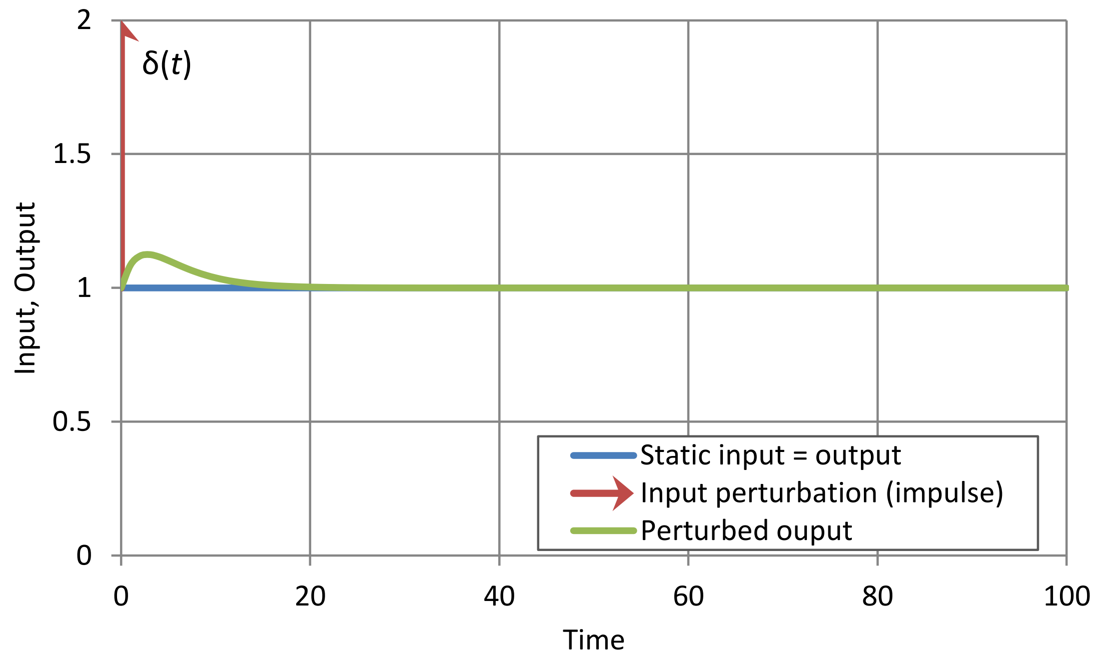
Linearity as perturbation damper

Linear dynamics (g: input, x: output)	$a_n \frac{\mathrm{d}^n x}{\mathrm{d}t^n} + \cdots + a_1 \frac{dx}{\mathrm{d}t} + a_0 x = g$
General solution (convolution, with impulse response function $h(t)$)	$x(t) = g(t) * h(t) = \int_{-\infty}^{\infty} g(t-\tau) h(\tau) \mathrm{d}\tau$
For a causal system ($h(t) = 0$ for $t < 0$)	$x(t) = \int_{0}^{\infty} g(t-\tau) h(\tau) \mathrm{d}\tau$ $= \int_{-\infty}^{t} g(\tau) h(t-\tau) \mathrm{d}\tau$
Perturbation in output, $e_x(t) = x'(t) - x(t)$ where $x'(t)$ is the output for perturbed input $g'(t) = g(t) + e_g(t)$	$e_x(t) = e_g(t) * h(t)$ $= \int_{-\infty}^{\infty} e_g(t) h(t-\tau) \mathrm{d}\tau$
Young's inequality	$\lVert x(t) \rVert_2 \le \lVert h(t) \rVert_1 \lVert g(t) \rVert_2$, $\lVert e_x(t) \rVert_2 \le \lVert h(t) \rVert_1 \lVert e_g(t) \rVert_2$
For a mass preserving transformation, $\lVert h(t) \rVert_1 = 1$	$\lVert x(t) \rVert_2 \le \lVert g(t) \rVert_2$ $\lVert e_x(t) \rVert_2 \le \lVert e_g(t) \rVert_2$

A linear system reduces the variability and uncertainty when transforming input to output.

An example with linear dynamics

Dynamics	$8\,x''(t)\; +\; 6\,x'(t)\; +\; x(t)\; = g(t)$
Impulse response function ($U(t)$ is the Heaviside step function)	$h(t) = \frac{1}{2}\left(e^{-t/4} - e^{-t/2}\right) U(t), \lVert h(t) \rVert_1 = 1$
Constant input	$g(t) = 1 \rightarrow h(t) = 1$
Perturbation in inflow	$e_g(t) = \delta(t), \lVert e_g(t) \rVert_1 = 1, \lVert e_g(t) \rVert_2 = \infty$
Resulting perturbation in outflow	$e_x(t) = \frac{1}{2}\left(e^{-t/4} - e^{-t/2}\right), \lVert e_x(t) \rVert_2 = \frac{1}{\sqrt{12}}$



The stochastic version of the example

Dynamics: $8\,\underline{x}''(t) + 6\,\underline{x}'(t) + \underline{x}(t) = \underline{g}(t)$

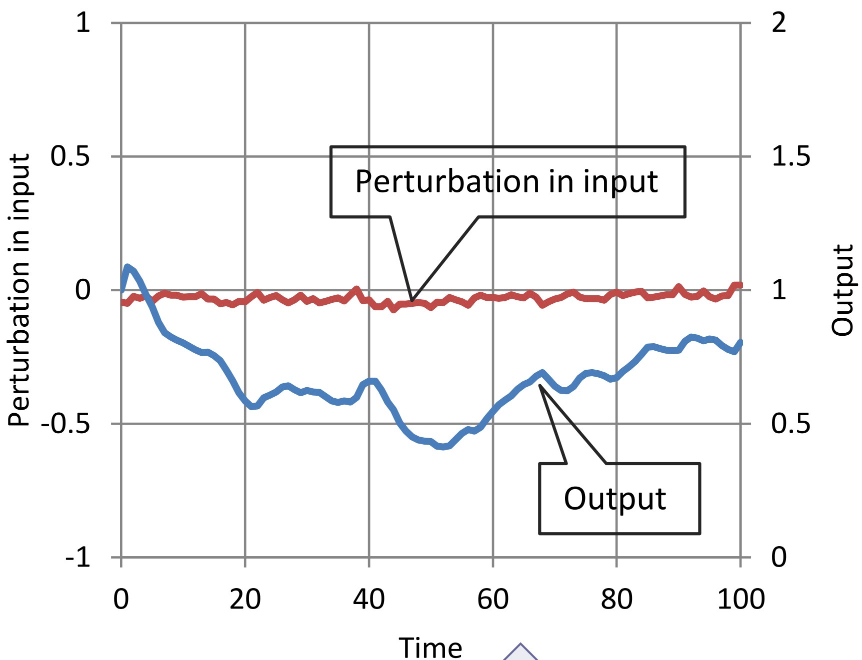
Solution: $\underline{x}(t) = a_1 \underline{x}(t-1) - a_2 \underline{x}(t-2) + \underline{w}(t)$

$a_1 := \mathrm{e}^{-1/2} + \mathrm{e}^{-1/4}$, $a_2 := -\mathrm{e}^{-3/4}$,

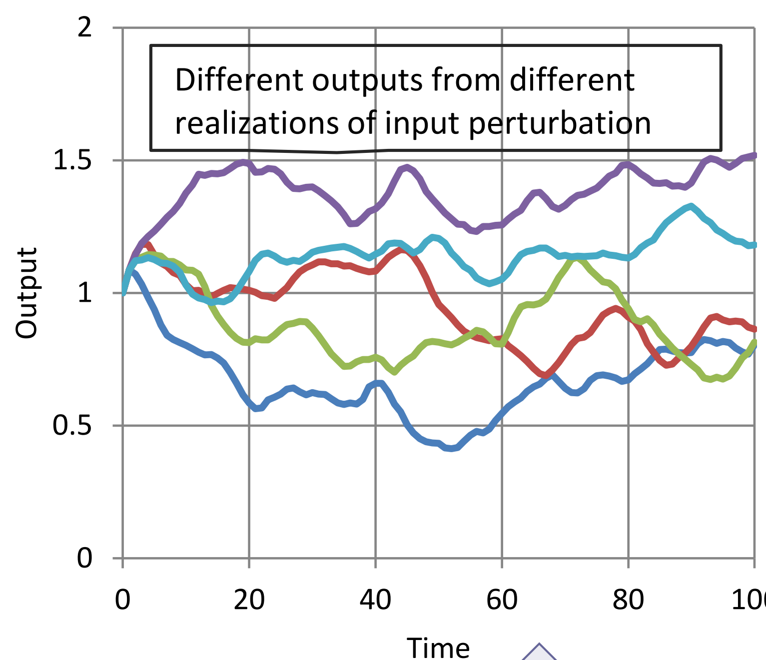
$\underline{w}(t) := \underline{u}(t) - \mathrm{e}^{-1/2}\underline{u}(t-1) - \underline{v}(t) + \mathrm{e}^{-1/4}\underline{v}(t-1)$

$\underline{u}(t) := \frac{1}{2}\int_{t-1}^{t} \mathrm{e}^{-(t-\tau)/4}\, \underline{g}(\tau)d\tau$, $\underline{v}(t) := \frac{1}{2}\int_{t-1}^{t} \mathrm{e}^{-(t-\tau)/2}\, \underline{g}(\tau)d\tau$

Linear stationary stochastic model
(Looks like an AR(2) process but it is not because $\underline{w}(t)$ is not white noise.)



Change



Uncertainty

Linearity: Difference in deterministic and stochastic systems

- Deterministic model + linear dynamics → No change (unless there is change in input)— Reduced uncertainty
- Deterministic model + nonlinear dynamics → Change —Uncertainty

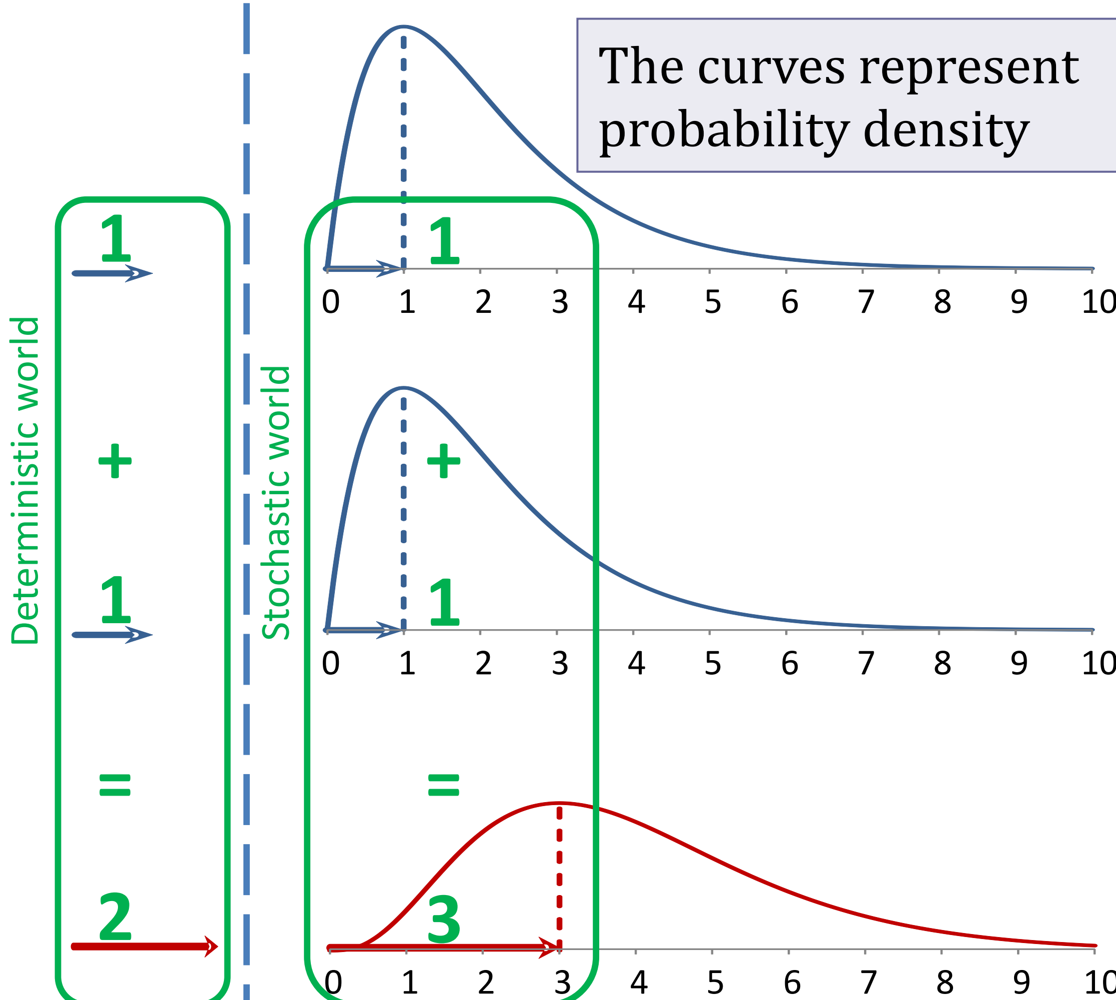
> …two states differing by imperceptible amounts may eventually evolve into two considerably different states (Lorenz, 1963).

- Stochastic model + linear or nonlinear dynamics → Change —Uncertainty (where stochastic model means either stochastic input or stochastic dynamics or both stochastic input and stochastic dynamics)

> Even if the dynamics is linear and even if the two initial states (assumed to be realizations of random variables) are identical, the later states will be always considerably different.

Linearity: different meaning of $x + y = z$ and $\underline{x} + \underline{y} = \underline{z}$

The curves represent probability density



Conversely: decomposition of 3 to, say, 1 +2 or 1 + 1, is not meaningful

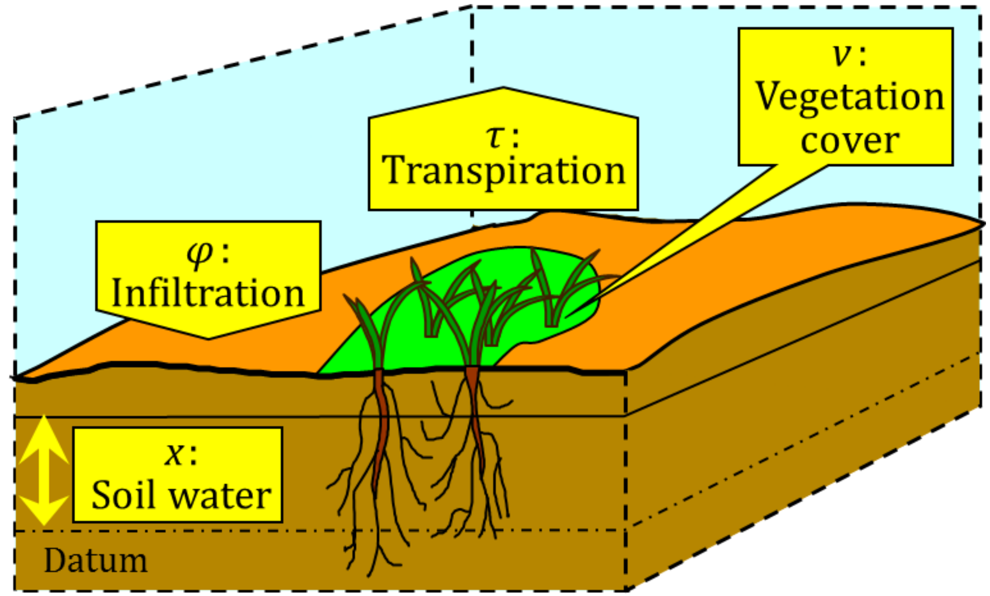
If each of two addends is most probably 1, then the sum is most probably 3.

This is a precise result: "most probably" suggests taking the mode of the distribution.

Note: $\underline{x}$ and $\underline{y}$ are independent and identically distributed with gamma distribution and shape parameter $\kappa = 2$. If they were dependent, then $\underline{x} + \underline{y}$ would have mode between 2 and ~3.5.

Emergence of linear randomness from nonlinear determinism

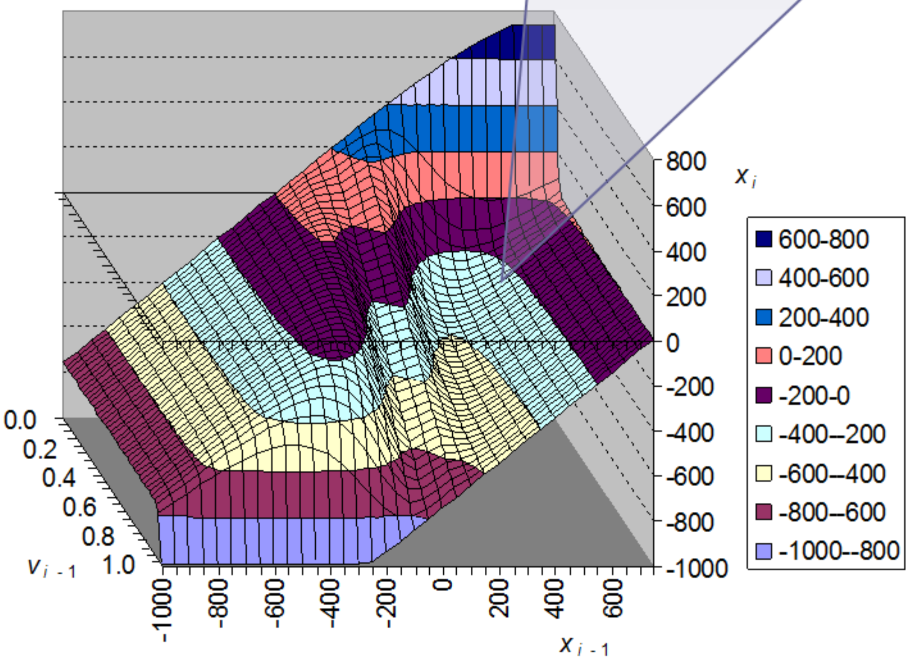
- We use a toy model for a caricature hydrological system, designed intentionally simple (Koutsoyiannis, 2010, "A random walk on water").
- Only infiltration, transpiration and soil water storage are considered.
- Discrete time: i ($t = i\Delta$ where Δ is an arbitrary time unit, Δ = 1 TU).
- The rates of infiltration φ and potential transpiration τ_p are **constant**.
  - Input: φ = 250 mm/TU;
  - Potential output: τ_p = 1000 mm/TU.
- State variables (**a 2D semidynamical system**):
  - Vegetation cover, v_i ($0 \le v_i \le 1$) ;
  - Soil water (no distinction from groundwater): x_i ($-\infty \le x_i \le \alpha = 750$ mm).
- Actual output: $\tau_i = v_i\, \tau_\mathrm{p}\, \Delta$
- Water balance: $x_i = \min(x_{i-1} + \Delta(\varphi - v_{i-1}\tau_\mathrm{p}), \alpha)$

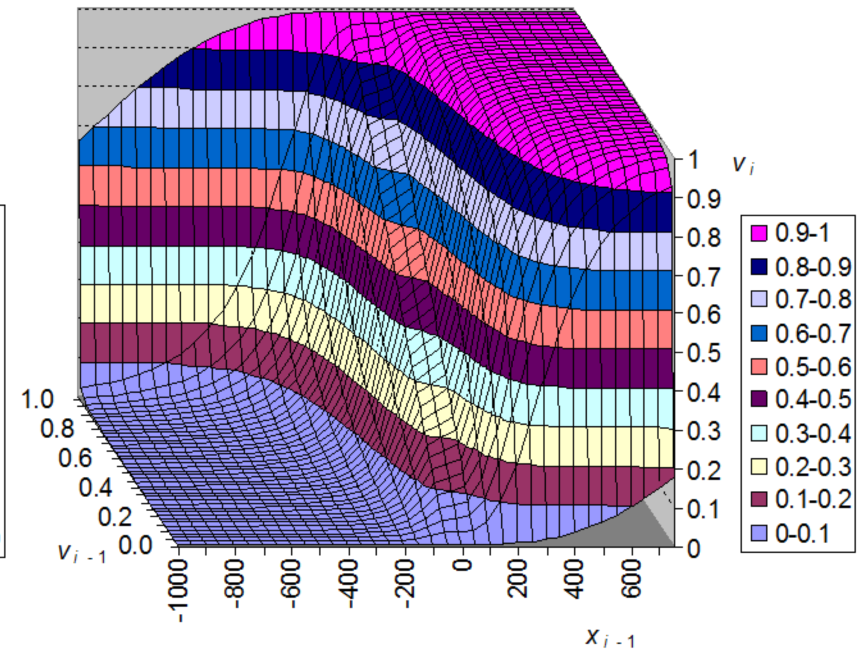


Nothing in the model is set to be random.

Toy model: system dynamics

Interesting surface—Not invertible transformation





Water balance	**Vegetation cover dynamics**
$x_i = \min(x_{i-1} + \Delta(\varphi - v_{i-1}\tau_\mathrm{p}), \alpha)$	$v_i = \dfrac{\max(1 + (x_{i-1}/\beta)^3, 1)\, v_{i-1}}{\max(1 - (x_{i-1}/\beta)^3, 1) + (x_{i-1}/\beta)^3 v_{i-1}}$

Assumed constants: φ = 250 mm/TU, τ_p = 1000 mm/TU, α = 750 mm, β = 100 mm.
Easy to program in a hand calculator or a spreadsheet: www.itia.ntua.gr/923/.

Detailed system dynamics: deterministic and stochastic

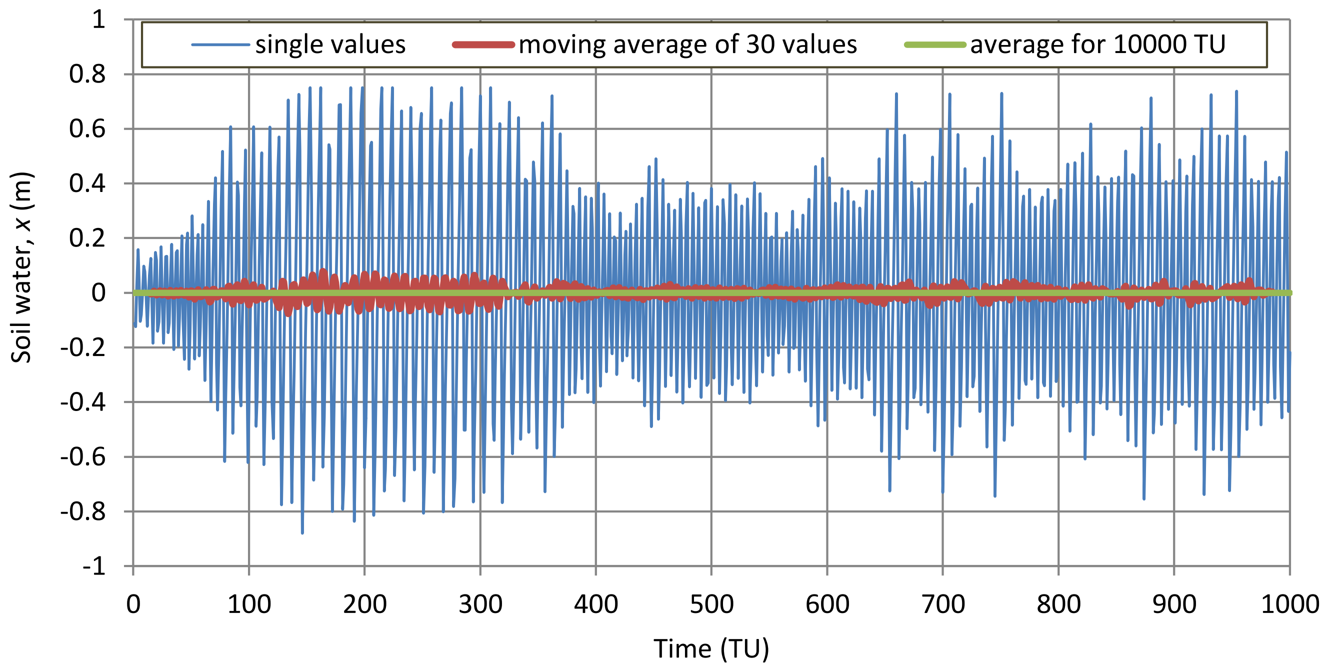
- In a deterministic description, $\boldsymbol{x}_i := (x_i, v_i)$ is the vector of the system state and $\boldsymbol{S}(\ )$ is the vector function representing the known deterministic dynamics of the system.
- Even though the deterministic description is complete, a couple of runs with slightly differing initial conditions will show that the deterministic dynamics does not allow reliable prediction except for a small time horizon.
- Therefore, we turn into a stochastic description and consider $\underline{\boldsymbol{x}}_i$ as a random variable with a probability density function $f_i(\boldsymbol{x})$.
- The stochastic representation behaves like a deterministic solution, but refers to the evolution in time of admissible sets and probability density functions, rather than to trajectories of points:

Deterministic description	**Stochastic description**
$\boldsymbol{x}_i = \boldsymbol{S}(\boldsymbol{x}_{i-1})$ where $\boldsymbol{S}$ is a vector transformation defining the system dynamics	$f_i(\boldsymbol{x}) = \dfrac{\partial^2}{\partial x \partial v} \int_{\boldsymbol{S}^{-1}(A)} f_{i-1}(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}$ where $A := \{\underline{\boldsymbol{x}} \le (x, v)\}$ and $\boldsymbol{S}^{-1}(A)$ is the counterimage of A

Interesting trajectories produced by simple deterministic dynamics

The plot of the soil water for a long period (1000 TU) indicates:

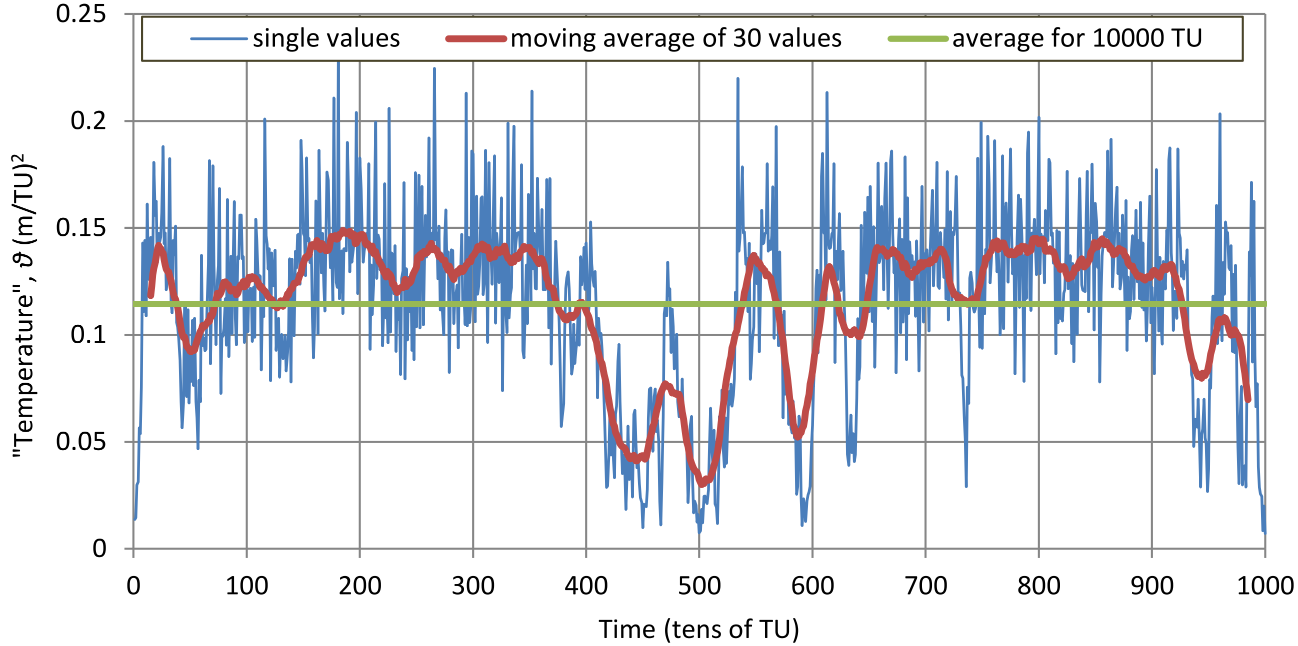
- High variability at a short (annual) scale.
- A flat time average at a 30-TU (30-year) scale (“climate”).
- Peculiar variation patterns.



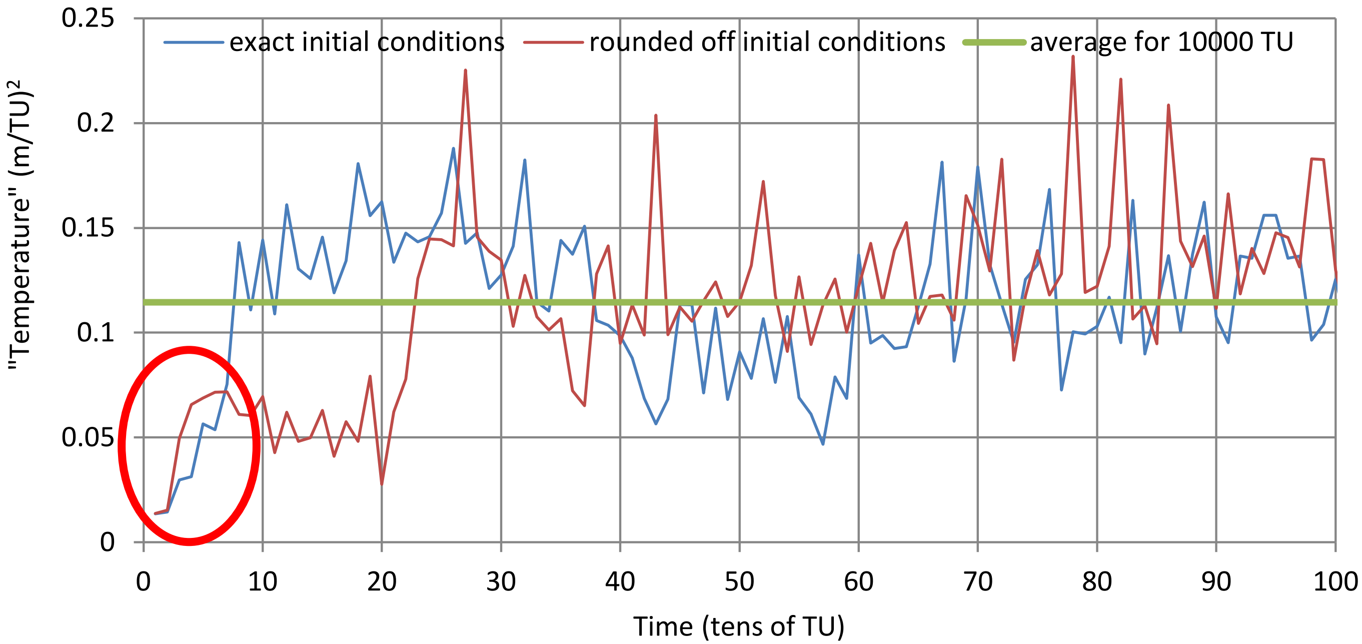
The behaviour quickly flattening the time average is known as **antipersistence** (often confused with periodicity/oscillation, which is an error).

Quantification of variability

- To study the peculiar variability of the soil water $\underline{x}_i$, we introduce the random variable $\underline{e}_i := ((\underline{x}_i - \underline{x}_{i-1})/\Delta)^2$, where $\Delta = 1$ TU; $\underline{e}_i$ is an analogue of the "kinetic energy" in the variation of the soil water.
- Furthermore we introduce a macroscopic variable $\underline{\theta}$, an analogue of "temperature", which is the average of 10 consecutive $\underline{e}_i$; high or low θ indicates high or low rates of variation of soil water.
- The plot of the time series of θ for a long period (10000 TU) indicates long and persistent excursions of the local average ("the climate") from the global average (of 10000 values).
- These remarkable changes are produced by the internal dynamics (no perturbation, no forcing).

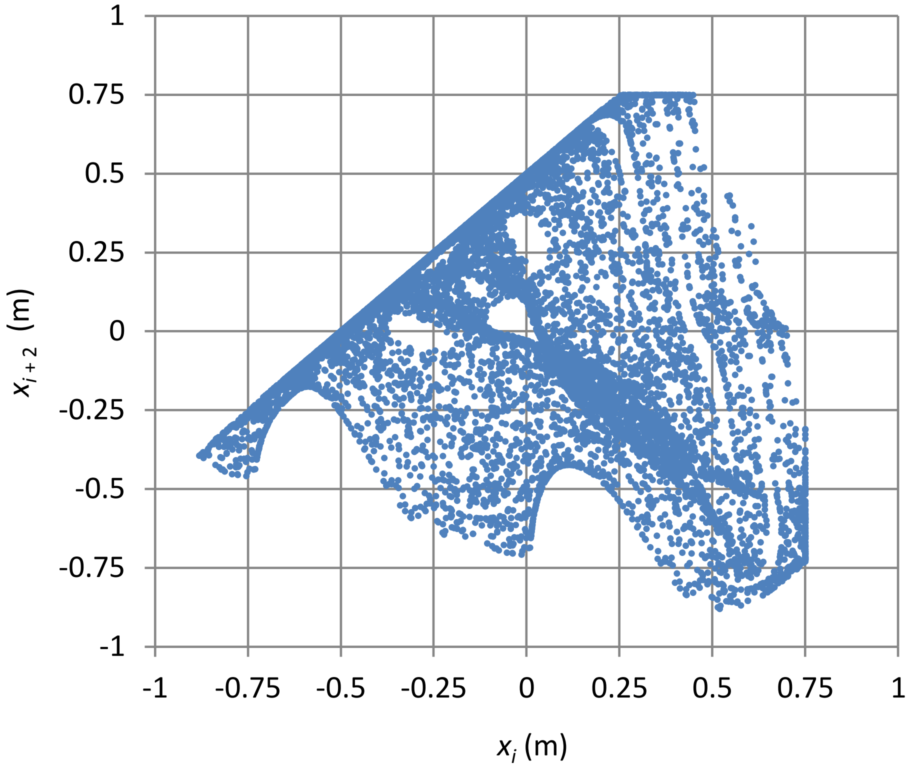


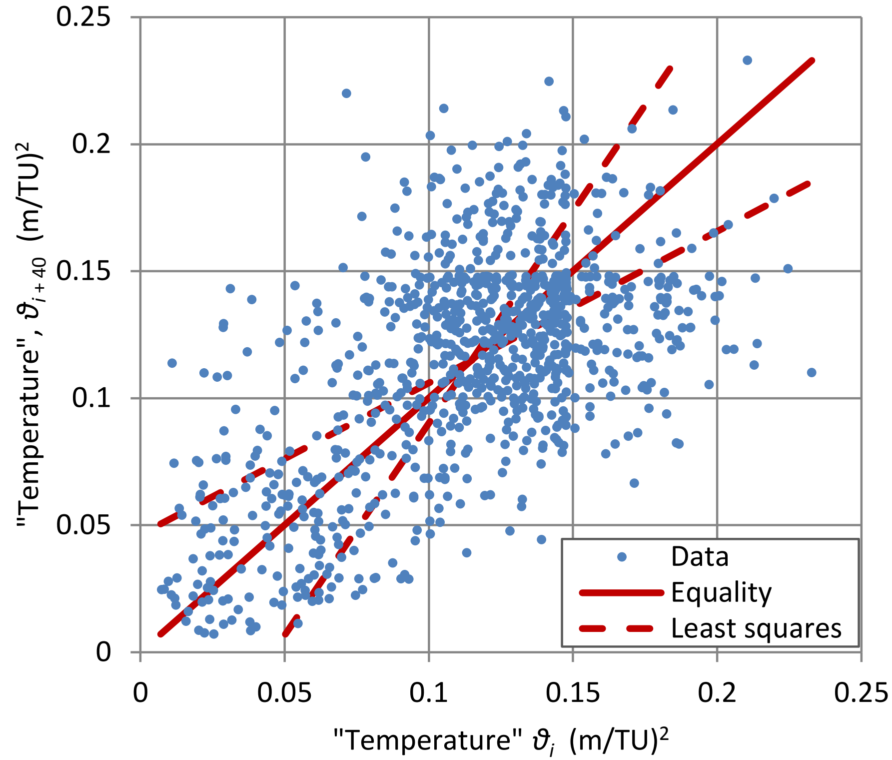
Is a fully deterministic nonlinear system predictable? [Reply: No, it is fully unpredictable in deterministic terms]



- The plot shows 100 terms of "temperature" time series produced with exact, as well as rounded off (by 10^{-2}), initial conditions.
- The departures in the two cases are striking.
- The detailed nonlinear deterministic (or stochastic) dynamics is good only for the short-term predictions (e.g. 1-5 time steps).
- For long-term predictions it is better to use macroscopic stochastic dynamics (possibly **linear**).

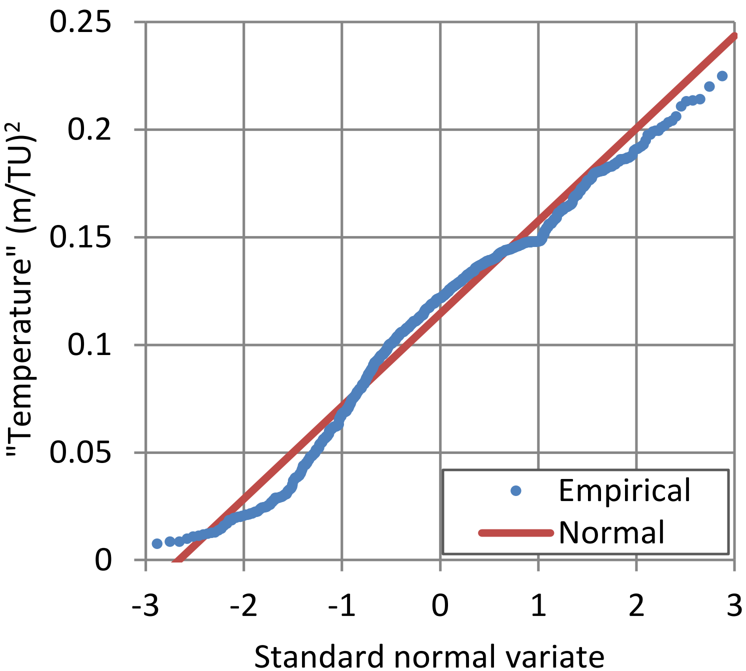
From detailed nonlinearity to macroscopic linearity

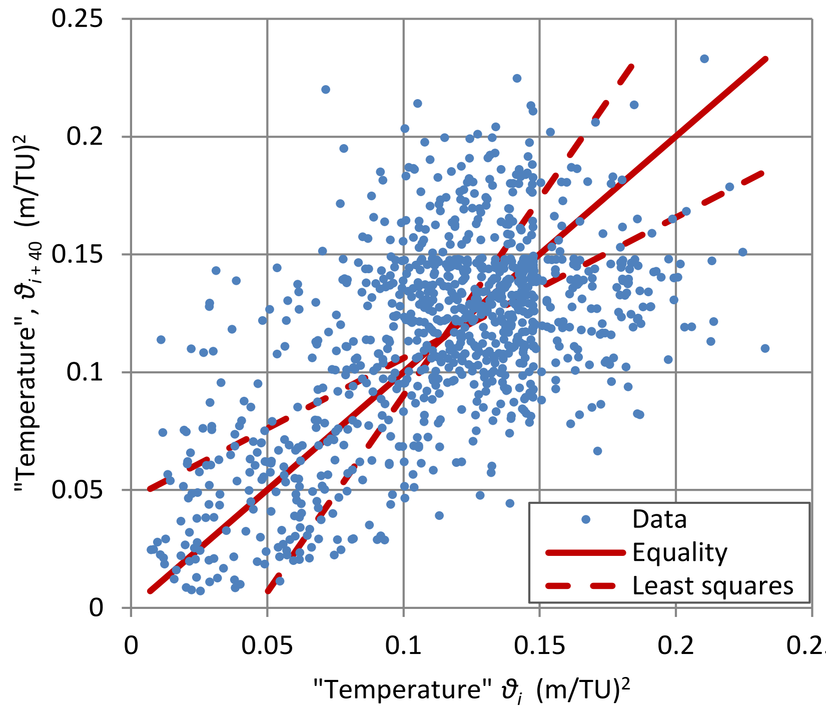




- The time lag plot of the detailed process (x_{i+2} vs. x_i) clearly reflects the nonlinear deterministic dynamics.
- The time lag plot of the macroscopic process “temperature” (θ_{i+40} vs. θ_i) reflects a linear statistical relationship.

Why macroscopization (coarse graining) is accompanied by a tendency to normality and linearity?





- Normality is a consequence of the central limit theorem.
- Both normality and linearity are consequences of the **principle of maximum entropy** for simple constraints, related to preservation of mean, variance and one or more autocovariance terms.
- The principle of maximum entropy makes macroscopic descriptions as simple and parsimonious as possible.

Part 2: Entropy and uncertainty

Definition and importance of entropy

- Historically entropy was introduced in thermodynamics but later it was given a rigorous definition within probability theory (owing to Boltzmann, Gibbs and Shannon).
- Thermodynamic and probabilistic entropy are essentially the same thing (Koutsoyiannis, 2013a, 2014; but others have different opinion).
- Entropy is a dimensionless measure of uncertainty defined as follows:

Discrete random variable $\underline{z}$	**Continuous random variable $\underline{z}$**
$\Phi[\underline{z}] := \mathrm{E}[-\ln P(\underline{z})] = -\sum_{j=1}^{w} P_j \ln P_j$ where $P_j := P\{\underline{z} = z_j\}$	$\Phi[\underline{z}] := \mathrm{E}\left[-\ln \frac{f(\underline{z})}{h(\underline{z})}\right] = -\int_{-\infty}^{\infty} \ln \frac{f(z)}{h(z)} f(z) \mathrm{d}z$ where $f(z)$ denotes probability density while $h(\mathrm{z})$ is the density of a background measure (usually $h(\mathrm{z}) = 1[z^{-1}]$)

- Entropy acquires its importance from the **principle of maximum entropy** (Jaynes, 1957), which 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.
- Its physical counterpart, the tendency of entropy to become maximal (2nd Law of thermodynamics) is the driving force of natural change.

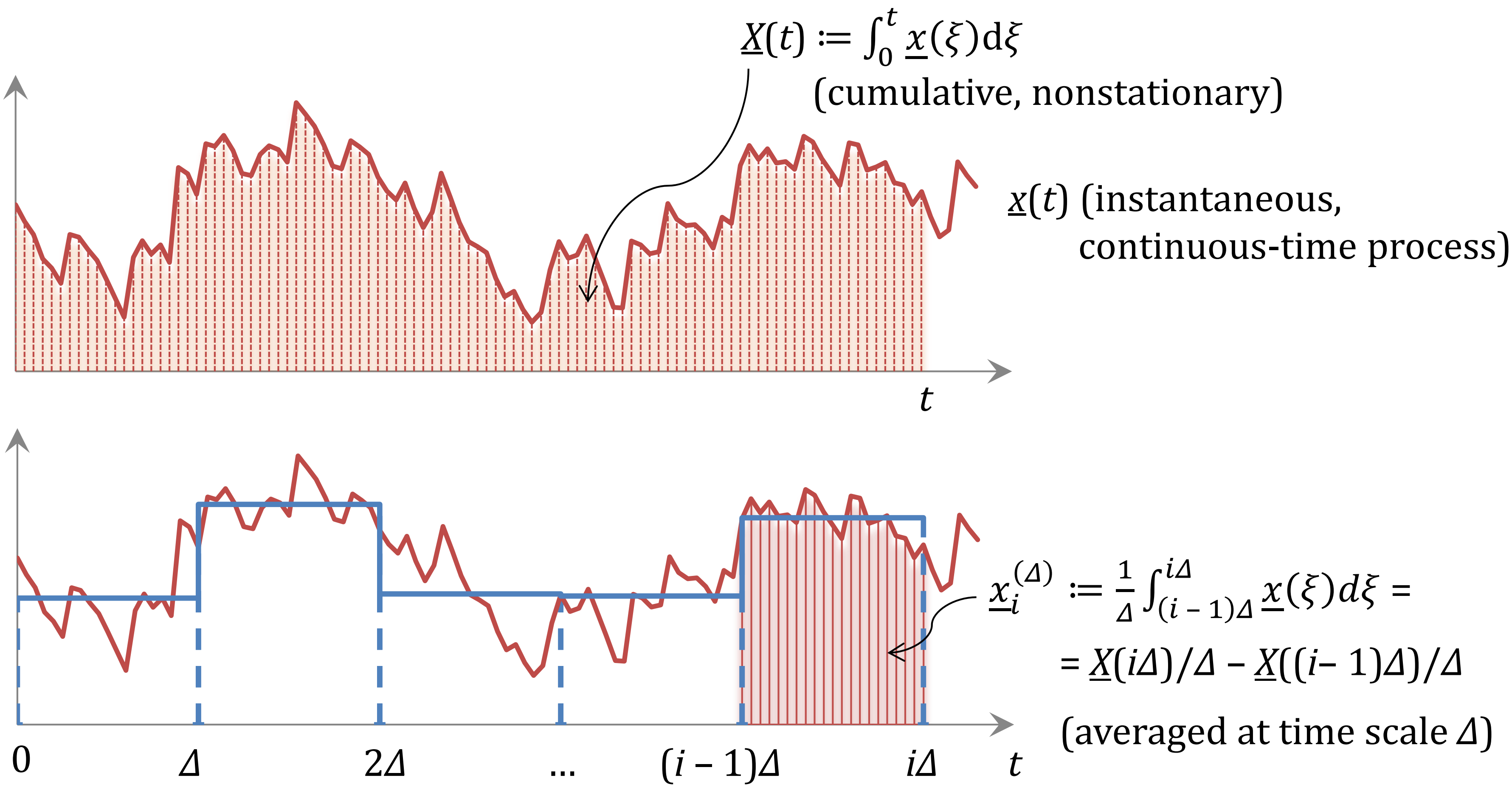
Entropy maximization: only in a stochastic macroscopic world

- A **dynamical law** S_t maps a system's state y at time $t = 0$ into new states $S_t(y)$ as time t changes.
- A **dynamical system** is, by definition, time invertible (reversible): $S_t(S_{t'}(y)) = S_{t+t'}(y)$ for $t, t' \in R$ (positive or negative), so that $S_t(S_{-t}(y)) = y$.
- A **semidynamical system** is, by definition, noninvertible (irreversible) in time: the relationship $S_t(S_{t'}(y)) = S_{t+t'}(y)$ holds only for $t, t' \in R^+$ (only positive), so that $S_t(S_{-t}(y)) \neq y$.
- In a **dynamical system** (time invertible) system the **entropy is constant** (Mackey, 2003, p. 31).
- In a **semidynamical system** (noninvertible in time) the **entropy is nondecreasing** reaching a limit (maximum) as $t \to \infty$ (Mackey, 2003, p. 30).
- **God theorem** (name given by Mackey, 2003, p. 111): Every continuous trajectory $x(t)$ in a space X is the trace (projection) of a single dynamical system $S_t(y)$ operating in a higher dimensional phase space Y.

As elementary physical laws are time invertible, the entropy increase is inherent with macroscopization:

- in a detailed (high-dimensional) system description (Y), the entropy should be constant, but
- in a macroscopic (lower-dimensional) description (X) it may increase in time.

Toward entropy metrics for stochastic processes: The time scale in the description of a stochastic process



Second-order properties of a stochastic process

Instantaneous process	$\underline{x}(\xi)$ [stationary with variance γ_0]
Cumulative process	$\underline{X}(t) := \int_0^t \underline{x}(\xi)\mathrm{d}\xi$ [nonstationary]
Autocovariance	$c(\tau) := \mathrm{Cov}[\underline{x}(t), \underline{x}(t+\tau)]$
Power spectrum	$s(w) := 4\int_0^\infty c(\tau)\,cos(2\pi w\tau)\,d\tau$
Structure function (aka semivariogram or variogram)	$h(\tau) := \frac{1}{2}\mathrm{Var}\left[\underline{x}(t) - \underline{x}(t+\tau)\right] = \gamma_0 - c(\tau)$
Cumulative climacogram	$\Gamma(t) := \mathrm{Var}[\underline{X}(t)]$
Climacogram	$\gamma(\Delta) := \mathrm{Var}[\underline{X}(\Delta)/\Delta] = \Gamma(\Delta)/\Delta^2$

Every second-order property of the process can be obtained from any other, e.g.

$$c(\tau) = \int_0^\infty s(w)\cos(2\pi w\tau)\,\mathrm{d}w$$

$$c(\tau) = \frac{1}{2}\frac{\mathrm{d}^2\Gamma(\tau)}{\mathrm{d}\tau^2} = \frac{1}{2}\frac{\mathrm{d}^2\left(\tau^2\gamma(\tau)\right)}{\mathrm{d}\tau^2}$$

$$\gamma(\Delta) = \frac{\Gamma(\Delta)}{\Delta^2} = \frac{2}{\Delta^2}\int_0^\Delta(\Delta-\tau)c(\tau)\mathrm{d}\tau = 2\int_0^1(1-\xi)c(\xi\Delta)\mathrm{d}\xi$$

Entropy production in stochastic processes

- In a stochastic process the change of uncertainty in time can be quantified by the **entropy production**, i.e. the time derivative (Koutsoyiannis, 2011):

$$\Phi'[\underline{X}(t)] := \mathrm{d}\Phi[\underline{X}(t)]/\mathrm{d}t$$

- A more convenient (and dimensionless) measure is the entropy production (i.e. the derivative) in logarithmic time (EPLT):

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

- For a Gaussian process, the entropy depends on its variance $\Gamma(t)$ only and is given as (Papoulis, 1991):

$$\Phi[\underline{X}(t)] = (1/2)\ln(2\pi \mathrm{e}\, \Gamma(t))$$

- The EPLT of a Gaussian process is thus easily shown to be:

$$\varphi(t) = \Gamma'(t)\, t \,/\, 2\Gamma(t)$$

- When the past and the present are observed, instead of the unconditional variance $\Gamma(t)$ we should use a variance $\Gamma_{\mathrm{C}}(t)$ conditional on the known past and present. This turns out to be:

$$\Gamma_{\mathrm{C}}(t) \approx 2\Gamma(t) - \Gamma(2t)/2$$

Three processes extremizing entropy production

Process	Definition (through is autocovariance $c(t)$ or its climacogram $\gamma(\Delta)$)
Markov	$c(\tau) = \lambda \mathrm{e}^{-\tau/\alpha} \rightarrow \gamma(\Delta) = \frac{2\lambda}{\Delta/\alpha}\left(1 - \frac{1-\mathrm{e}^{-\Delta/\alpha}}{\Delta/\alpha}\right)$
Hurst-Kolmogorov (HK)	$\gamma(\Delta) = \lambda(\alpha/\Delta)^{2-2H}$
Hybrid Hurst-Kolmogorov (HHK)	$\gamma(\Delta) = \lambda\left(1 + (\Delta/\alpha)^{2\kappa}\right)^{\frac{H-1}{\kappa}}$

See details in Koutsoyiannis (2011, 2015).

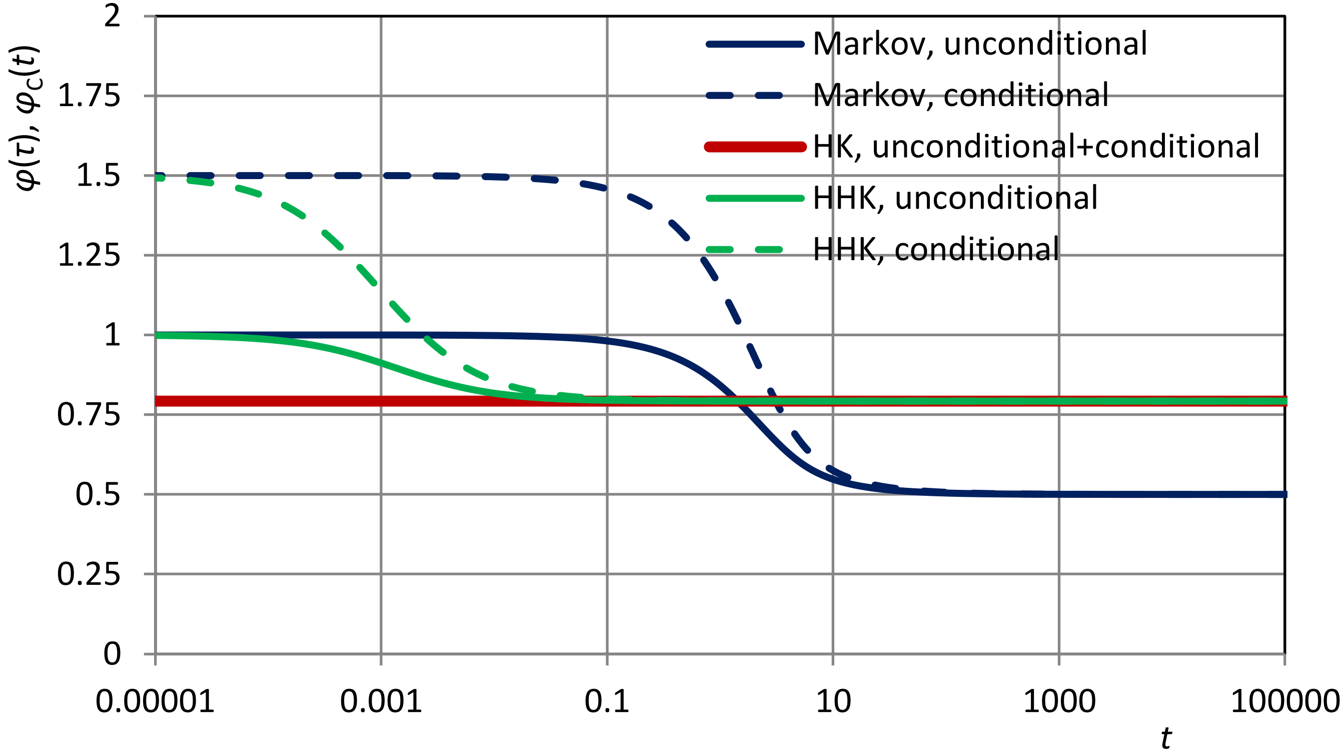
Parameters:

- λ: state-scale parameter, $[x]^2$
- α: time-scale parameter, $[t]$
- H: scaling parameter ($0 < H < 1$; Hurst parameter)
- κ: scaling parameter ($0 < \kappa < 1$; fractal parameter; fractal dimension = $2 - \kappa$)

Note: In general, the fractal and Hurst parameters are two different things (Gneiting and Schlather, 2004):

- The **fractal parameter** determines the **local properties** of the process (as $t \rightarrow 0$)
- The **Hurst parameter** determines the **global properties** of the process (as $t \rightarrow \infty$)

Entropy production in the three processes

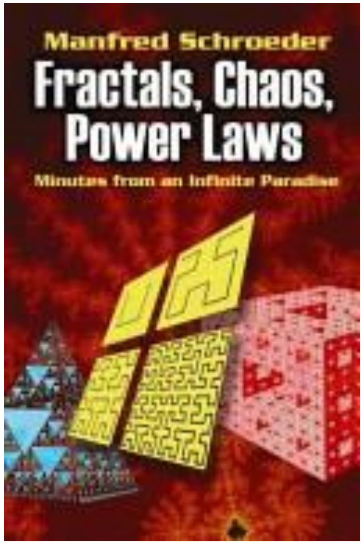


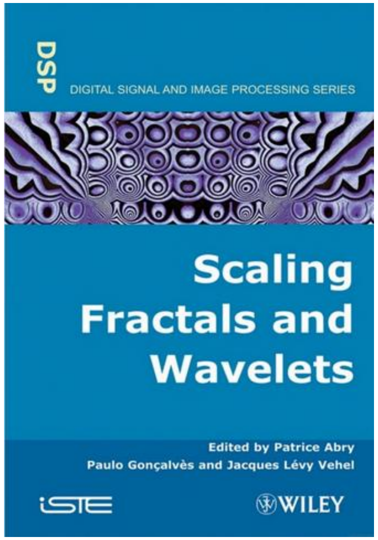
Conditions for EPLT maximization:

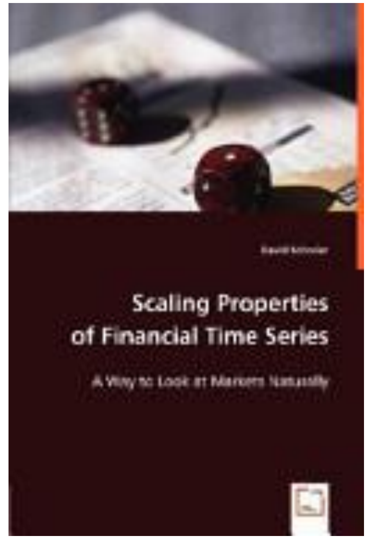
At time scale $\Delta = 1$ all three processes have the same variance $\gamma(1) = 1$ and the same autocovariance for lag 1, $c_1^{(1)} = 0.5$; for the HHK process, $\kappa = 0.5$ was assumed.

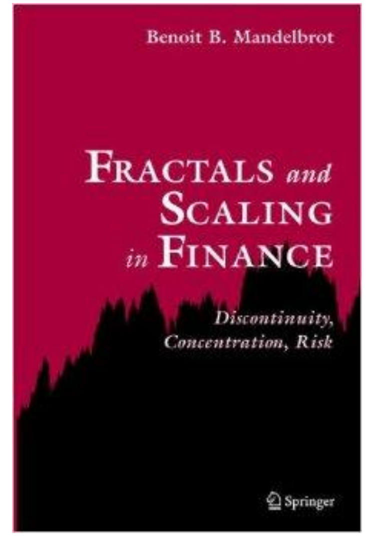
(Koutsoyiannis, 2011, 2015)

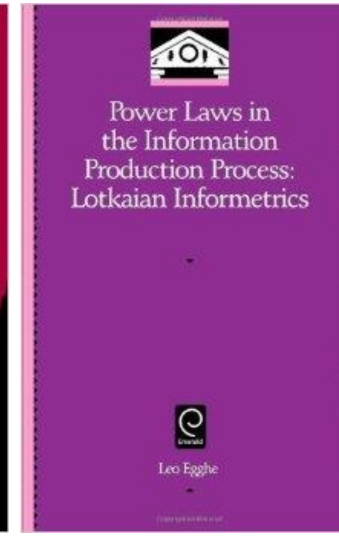
- The Markov process maximizes local entropy production (as $t \to 0$) and minimizes global entropy production (as $t \to \infty$).
- The HK process minimizes local entropy production (as $t \to 0$) and maximizes global entropy production (as $t \to \infty$).
- The HHK process maximizes both local (as $t \to 0$) and global (as $t \to \infty$) entropy production.



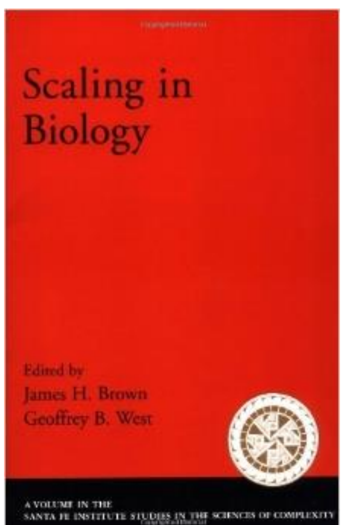


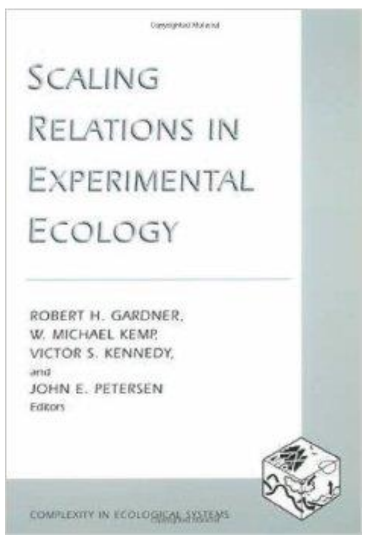


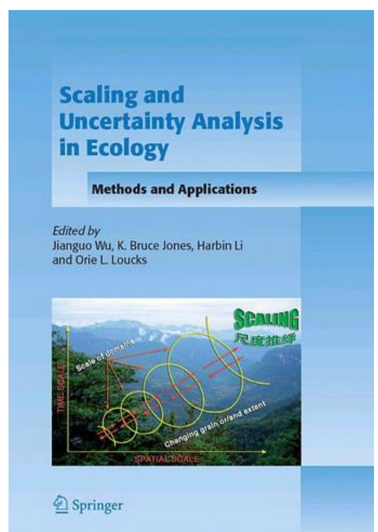


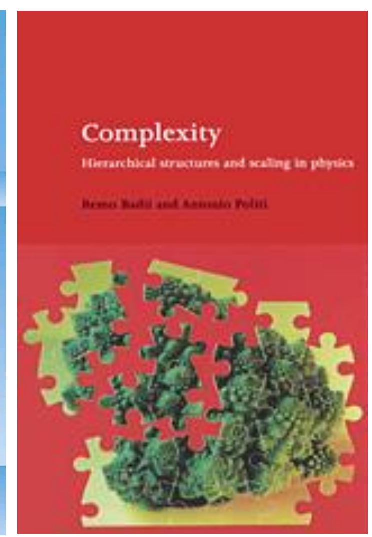


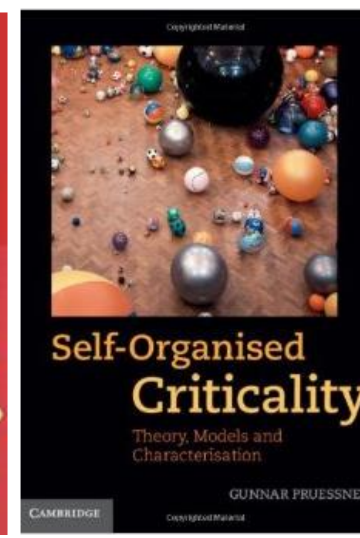
Part 3: Scaling and power laws











How to identify TWO power laws in (almost) ANY function

Lemma: Every nonzero continuous function $g(x)$ defined in $(0, \infty)$, whose limits at 0 and ∞ exist, is associated with two asymptotic power laws.

Asymptotic behaviour as $x \to \infty$

1. Assuming that $\lim_{x \to \infty} g(x) = \beta$, we define a function $f(x)$ as follows:

$$f(x) := \begin{cases} g(x) - \beta, & \beta \neq \pm\infty \\ 1/g(x), & \beta = \pm\infty \end{cases}$$

   Clearly, $\lim_{x \to \infty} f(x) = 0$.

2. If $\lim_{c \to \infty}(\lim_{x \to \infty} x^c f(x)) = 0$, then we replace $f(x)$ with $-1/\ln|f(x)|$ (which preserves the property $\lim_{x \to \infty} f(x) = 0$); if necessary, we make iterations so that eventually $\lim_{x \to \infty} f(x) = 0$ and $\lim_{c \to \infty}(\lim_{x \to \infty} x^c f(x)) = \infty$.

3. Given the properties in 2, there exists a unique b, $0 \le b < \infty$, satisfying

$$\lim_{x \to \infty} x^b f(x) < \infty$$

   so that for any $b' \neq b$,

$$\lim_{x \to \infty} x^{b'} f(x) = \begin{cases} 0, & \forall b' < b \\ \infty, & \forall b' > b \end{cases}$$

The constant b defines an asymptotic power law with exponent $-b$ (cf. Hausdorff dimension; the case $b = 0$ signifies an improper scaling).

How to identify TWO power laws in ANY function (contd.)

Asymptotic behaviour as $x \to 0$

We define $\tilde{g}(x) \coloneqq g(1/x)$ and we proceed in the same manner to construct a function $\tilde{f}(x) \equiv f(1/x)$ and determine the unique a for which relationships similar those of the previous slide apply (i.e. $\lim_{x\to\infty} x^{-a}\tilde{f}(x) < \infty$). This determines an asymptotic power law with exponent $-a$ for $f(x)$ as $x \to 0$.

Remarks

- The two power laws refer to the same function $g(x)$ but may correspond to different functions $f(x)$, say, $f_a(x)$ and $f_b(x)$ for the asymptotic behaviours as $x \to 0$ (local or fractal behaviour) and $x \to \infty$ (global behaviour), respectively.
- However, it is easy to construct a single function that combines both, e.g. $f_a(x) f_b(x)$—but many of them can actually be constructed.
- As well as any object has a dimension, any continuous function entails asymptotic power laws; generally not one but two, which could in special cases be identical.
- There is no magic in power laws (sorry about that!), except that they are, logically and mathematically, a necessity.
- No assumption of criticality, self-organization, fractal or multi-fractal generating mechanisms, is necessary to justify their emergence.

The log-log derivative

A power law is visualized in a graph of $f(x)$ plotted in logarithmic axis vs. the logarithm of x. Formally, this slope is expressed by the log-log derivative:

$$f^{\#}(x) := \frac{\mathrm{d}(\ln f(x))}{\mathrm{d}\,(\ln x)} = \frac{x f'(x)}{f(x)}$$

Of particular interest are the asymptotic values for $x \to 0$ and ∞, symbolically $f^{\#}(0)$ and $f^{\#}(\infty)$. These are:

$$f^{\#}(\infty) = -b, \; f^{\#}(0) = -a$$

Proof for the former case (the latter can be handled in the same manner):

$$\lim_{x\to\infty} x^b f(x) = \lim_{x\to\infty} \frac{f(x)}{x^{-b}} = \lim_{x\to\infty} \frac{f'(x)}{-b\, x^{-b-1}} = \lim_{x\to\infty} \frac{f'(x)}{-b\, x^{-b}} = \lim_{x\to\infty} \frac{x f'(x)}{-b\, f(x)} \frac{f(x)}{x^{-b}} = \lim_{x\to\infty} \frac{-f^{\#}(x)}{b} \frac{f(x)}{x^{-b}}$$

This implies that $\lim_{x\to\infty} -f^{\#}(x)/b = 1$ or $f^{\#}(\infty) = -b$.

Metrics of asymptotic behaviour of stochastic processes

Metric	Definition	Comments
For the global asymptotic behaviour ($\varDelta \to \infty$): **Climacogram**	$\gamma(\varDelta) \coloneqq \mathrm{Var}[\underline{X}(\varDelta)/\varDelta] = \Gamma(\varDelta)/\varDelta^2$ where $\underline{X}(\varDelta)$ is the cumulative process in the interval $[0, \varDelta]$	For an ergodic process for $\varDelta \to \infty$ $\gamma(\varDelta) \to 0$ necessarily
For the local asymptotic behaviour ($\varDelta \to 0$): **Climacogram-based structure function** (CBSF)	$g(\varDelta) \coloneqq \gamma_0 - \gamma(\varDelta)$ where $\gamma_0 = \gamma(0)$ is the variance of the instantaneous process $\underline{x}(t)$	The definition presupposes that the variance γ_0 is finite
For both the global and local asymptotic behaviour: **Climacogram-based spectrum** (CBS)	$\psi(w) \coloneqq \frac{2}{w\gamma_0}\gamma(1/w)g(1/w) = \frac{2\,\gamma(1/w)}{w}\left(1 - \frac{\gamma(1/w)}{\gamma_0}\right)$ where $w \equiv 1/\varDelta$ is frequency (as in the power spectrum)	It combines the climacogram and the CBSF; it is valid for both finite and infinite variance

Note: The CBSF is related to the structure function $h(\tau)$ by the same way as the climacogram is related to the autocovariance function $c(\tau)$:

$$c(\tau) = \frac{1}{2}\frac{\mathrm{d}^2\big(\tau^2\gamma(\tau)\big)}{\mathrm{d}\tau^2},\; h(\tau) = \frac{1}{2}\frac{\mathrm{d}^2\big(\tau^2 g(\tau)\big)}{\mathrm{d}\tau^2}$$

Relationship of the climacogram and the climacogram-based metrics with more standard metrics

- The asymptotic behaviour of the climacogram is the same with that of the autocovariance function (under some general conditions and with the exception where $\gamma^{\#}(\infty) = 1$ or 2; see proof below).
- The asymptotic behaviour of the CBSF is the same with that of the structure function (under similar conditions and with the exception where $g^{\#}(0) = 1$ or 2; the proof is similar to that below, given the last equation of the previous slide).
- The asymptotic behaviour of the CBS is the same with that of the power spectrum (under some general conditions and with some exceptions not fully investigated yet; cf. Stein, 1999).

Proof for the first claim: We assume that $\gamma(\Delta)$ has first and second derivative which $\to 0$ as $\Delta \to \infty$. We use l'Hopital's rule to find:

$$\lim_{\tau\to\infty} \tau^b c(\tau) = \lim_{\tau\to\infty} \frac{c(\tau)}{\tau^{-b}} = \lim_{\tau\to\infty} \frac{1}{2}\,\frac{\mathrm{d}^2\big(\tau^2\gamma(\tau)\big)/\mathrm{d}\tau^2}{\tau^{-b}} = \lim_{\tau\to\infty}\left(\frac{\gamma(\tau)}{\tau^{-b}} + 2\,\frac{\gamma'(\tau)}{\tau^{-b-1}} + \frac{1}{2}\frac{\gamma''(\tau)}{\tau^{-b-2}}\right) = \frac{1}{2}(b-1)(b-2)\lim_{\tau\to\infty}\left(\frac{\gamma(\tau)}{\tau^{-b}}\right) = \frac{1}{2}(b-1)(b-2)\lim_{\tau\to\infty}\tau^b\gamma(\tau)$$

Unless $b = 1$ or $b = 2$, the limit $\lim_{\tau\to\infty} \tau^b c(\tau)$ is 0, finite or ∞, if and only if $\lim_{\tau\to\infty} \tau^b \gamma(\tau)$ is 0, finite or ∞, respectively. Note that a Markov process belongs to the exceptions because $b = 1$.

Why prefer the climacogram and the climacogram-based metrics over more standard ones?

- In stochastic processes, almost all classical statistical estimators are biased and uncertain; in processes with LTP bias and uncertainty are very high.
- In the climacogram (variance), bias and uncertainty are easy to control as they can be calculated analytically (and a priori known).
- The autocovariance function is the second derivative of the climacogram.
  - Estimation of the second derivative from data is too uncertain and makes a very rough graph.
  - Estimation of autocovariance is too biased in processes with LTP.
- The power spectrum is the Fourier transform of the autocovariance and entails an even rougher shape and more uncertain estimation than in the autocovariance (see also Dimitriadis and Koutsoyiannis, 2015).
- An additional advantage of the climacogram is its close relationship with EPLT. Specifically, combining the equations of slides 24, 25 and 31 we conclude that for Gaussian processes the EPLT is

$$\varphi(t) = 1 + \tfrac{1}{2}\, \gamma^{\#}(t)$$

- This entails that the Hurst coefficient H equals the global EPLT, $\varphi(\infty)$.

Asymptotic properties of the EPLT extremizing processes

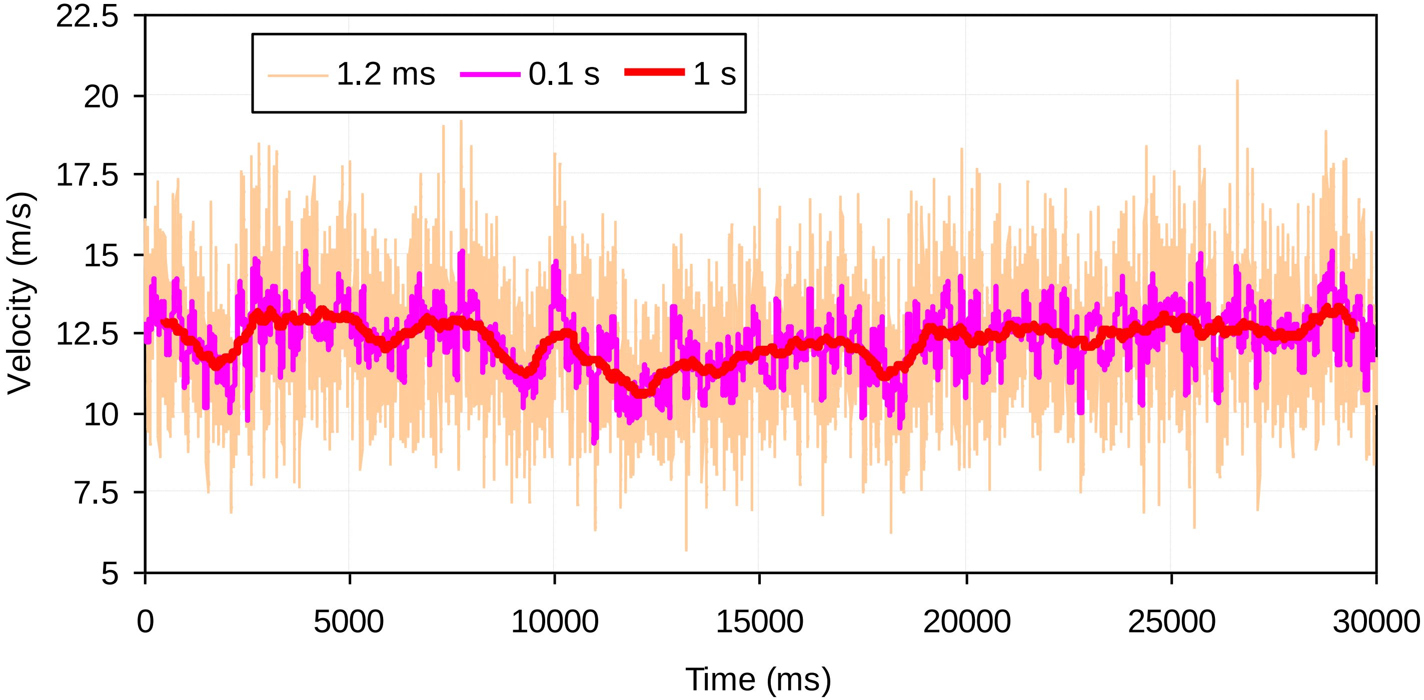
	Markov	Hurst-Kolmogorov (HK)	Hybrid Hurst-Kolmogorov (HHK)
Climacogram	$\gamma(\Delta) = \frac{2\lambda}{\Delta/\alpha}\left(1 - \frac{1-\mathrm{e}^{-\Delta/\alpha}}{\Delta/\alpha}\right)$	$\gamma(\Delta) = \lambda(\alpha/\Delta)^{2-2H}$	$\gamma(\Delta) = \lambda(1 + (\Delta/\alpha)^{2\kappa})^{\frac{H-1}{\kappa}}$
Global behaviour	$\gamma^{\#}(\infty) = -1$ ($c^{\#}(\infty) = -\infty$) $\psi^{\#}(0) = s^{\#}(0) = 0$	$\gamma^{\#}(\infty) = c^{\#}(\infty) = 2H - 2$ $\psi^{\#}(0) = s^{\#}(0) = 1-2H$	$\gamma^{\#}(\infty) = c^{\#}(\infty) = 2H - 2$ $\psi^{\#}(0) = s^{\#}(0) = 1-2H$
Hurst coefficient = EPLT $\varphi(\infty)$	0.5	H	H
Local behaviour	$g^{\#}(0) = h^{\#}(0) = 1$ $\psi^{\#}(\infty) = s^{\#}(\infty) = -2$	$\gamma^{\#}(0) = c^{\#}(0) = 2H - 2$ $\psi^{\#}(\infty) = s^{\#}(\infty) = 1-2H$	$g^{\#}(0) = h^{\#}(0) = 2\kappa$ $\psi^{\#}(\infty) = s^{\#}(\infty) = -2\kappa - 1$
Fractal dimension	1.5 (?)	$2 - H$	$2 - \kappa$
EPLT $\varphi(0)$	1	H	1
Conditional EPLT $\varphi_C(0)$	1.5	H	$1 + \kappa$

Parameters:

$\lambda > 0$ $[x]^2$ (state-scale parameter); $\alpha > 0$ $[t]$ (time-scale parameter);

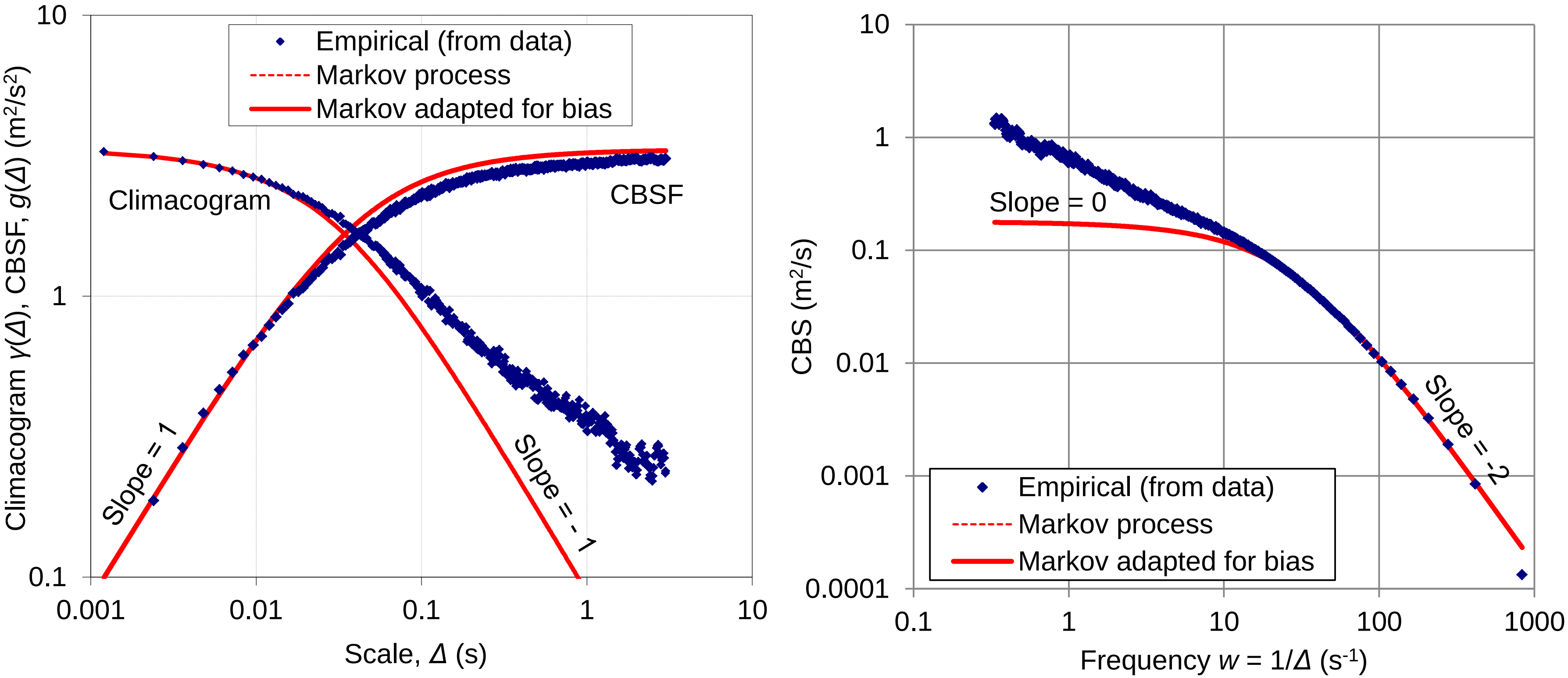
$0 < H < 1$ [–] (Hurst parameter); $0 < \kappa < 1$ [–] (fractal parameter)

Data analysis and compliance with theory



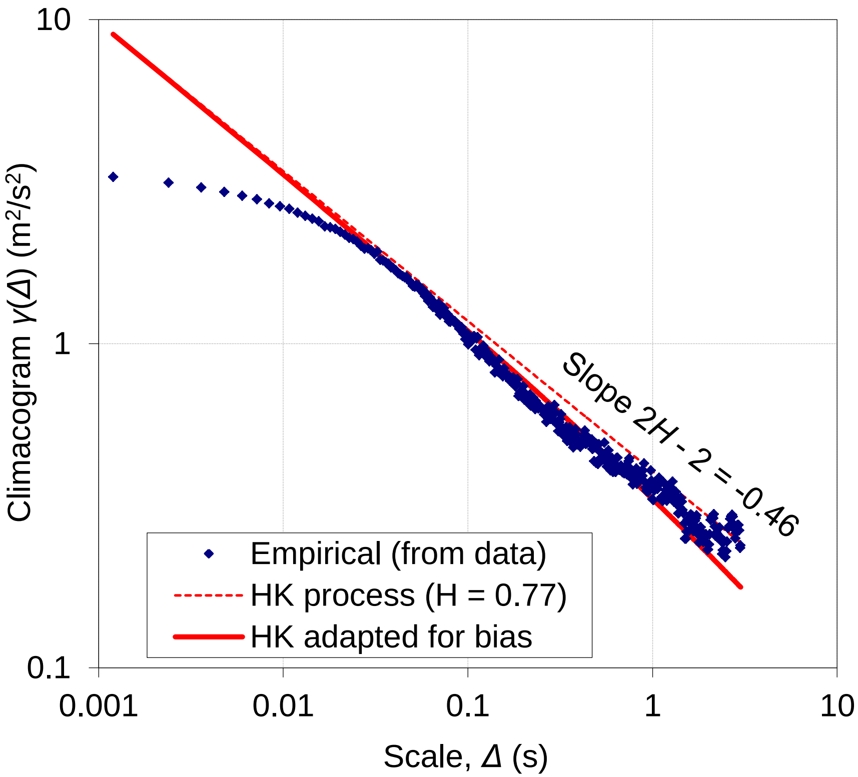
- Measurements of turbulent velocity offer the best way to sound out how nature works because they enable views on a wide range of scales, including very short ones.
- The graph shows laboratory measurements (by X-wire probes) of nearly isotropic turbulence in Corrsin Wind Tunnel at a high-Reynolds-number (Kang et al., 2003); the sampling rate was 40 kHz (one per 25 ns; here aggregated at the three scales shown).

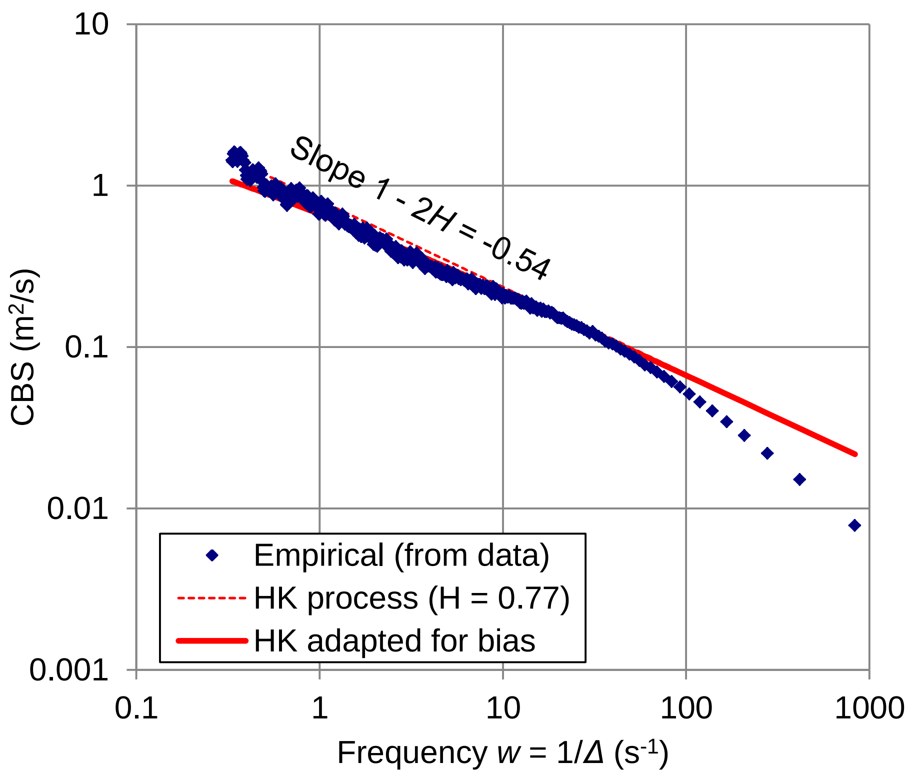
Turbulence is not a Markov process



- Here the climacogram, the climacogram-based structure function (CBSF) and the climacogram-based spectrum (CBS) are used to compare the properties estimated from measurements with the theoretical ones of a Markov process.
- The Markov process is good for the small time scales but not for the large ones.

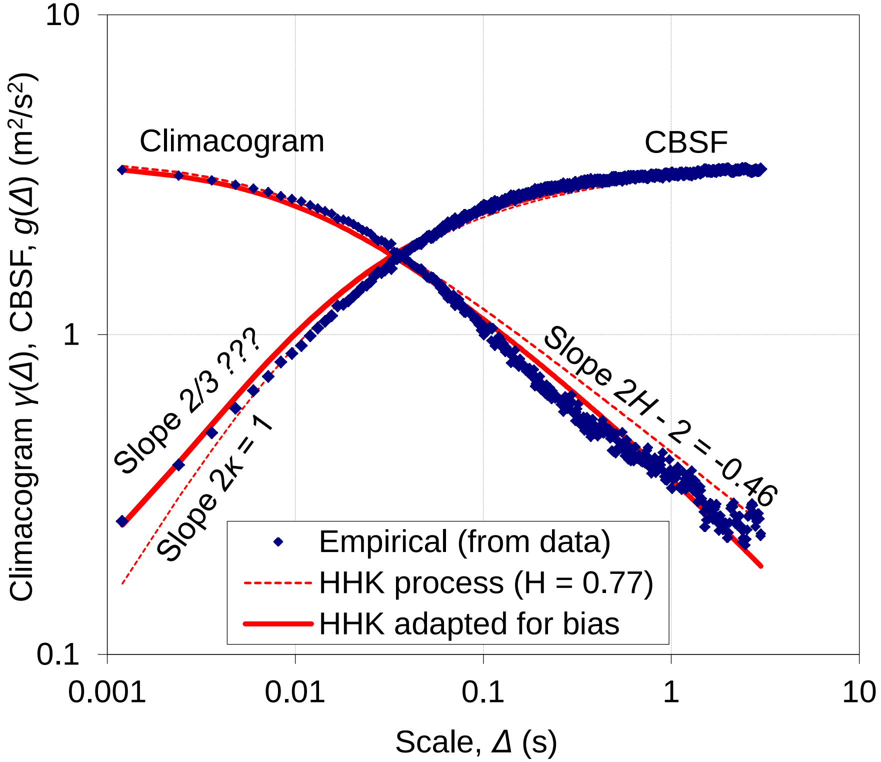
Turbulence is not a standard Hurst-Kolmogorov process

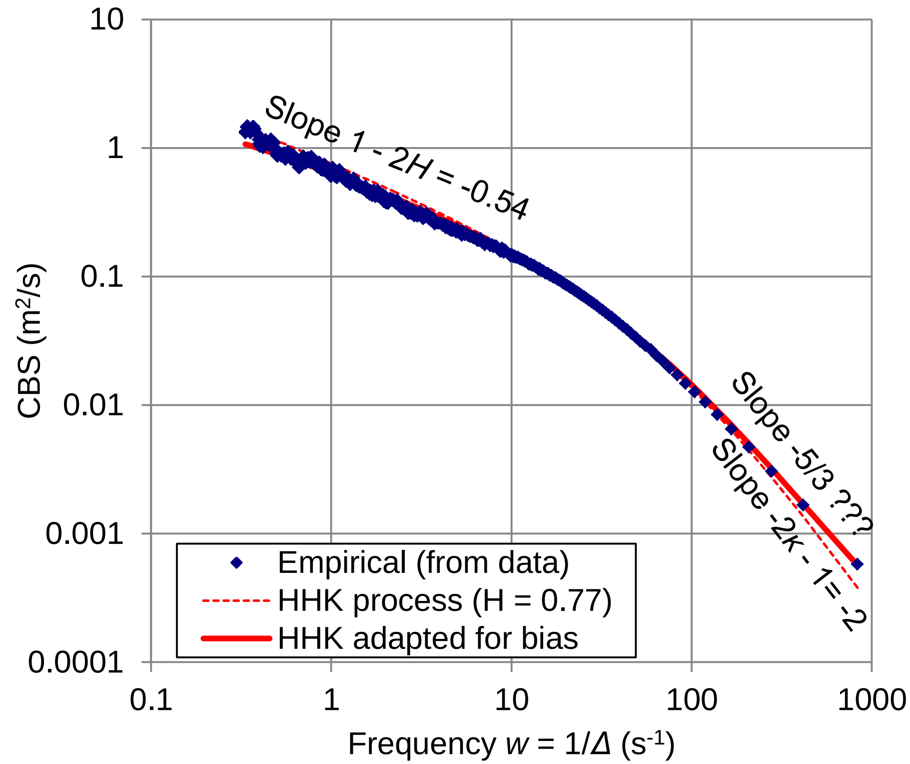




- The standard Hurst-Kolmogorov process is good for the large time scales but not for the small ones.

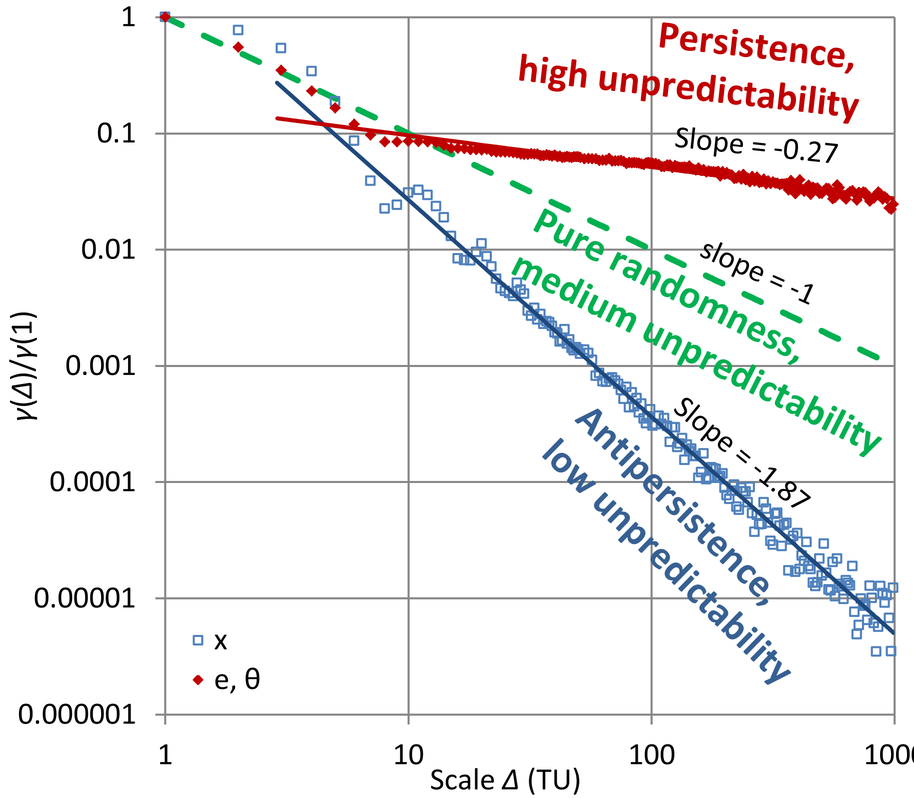
The Hybrid Hurst-Kolmogorov process for turbulence





- The Hybrid Hurst-Kolmogorov process is good for the entire range of time scales.
- It behaves like a Markov process for small scales and as a HK process for large ones.
- It indicates high entropy production both at small and large time scales, thus making the theory consistent with observation.

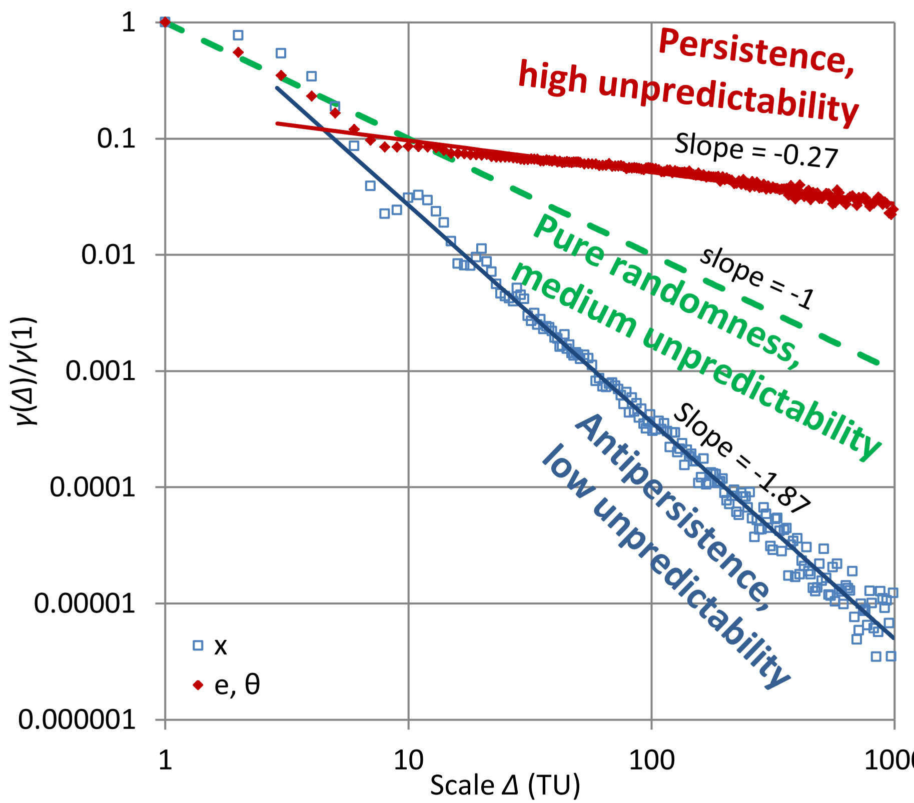
Antipersistence and persistence emerging from simple deterministic dynamics (toy model)



- The climacograms on the right refer to states x (soil water) and e or θ ("temperature") obtained from the toy model; they are compared to that of a purely random process (white noise).
- For an one-step ahead prediction, a purely random process x_i is the most unpredictable.
- Dependence and conditioning on observations enhances one-step ahead predictability.
- However, in the climatic-type predictions, which concern the local average rather than the exact value, the situation is different.
- For such prediction, most important is the variance at an aggregate scale, $\gamma(\Delta)$, while reduction due to conditioning on the past is usually annihilated.

Antipersistence and persistence emerging from simple dynamics (contd.)

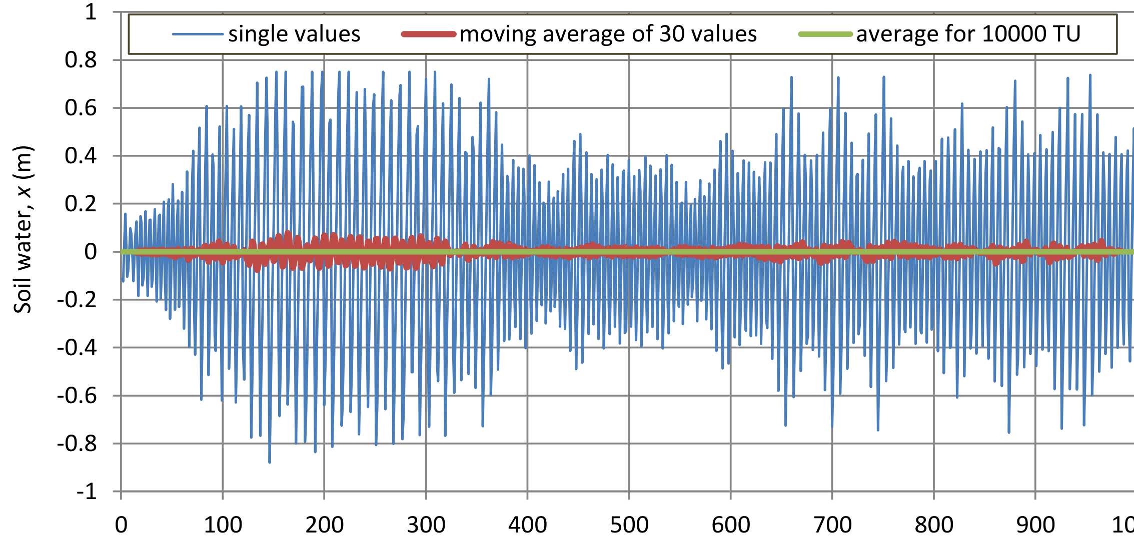
- From the climacogram of the process $\underline{x}_i$ it becomes clear that antipersistence reduces the long-term variance and thus enhances climatic-type predictability.
- Conversely, persistence (as in the processes $\underline{e}_i$ and $\underline{\theta}_i$) increases the long-term variance and thus enhances climatic-type unpredictability.
- Persistence is associated with positive dependence in time, while in antipersistent processes the dependence is negative; note though that for small scales/lags the autocorrelations should be positive even in antipersistent processes (see leftmost points on the climacogram).

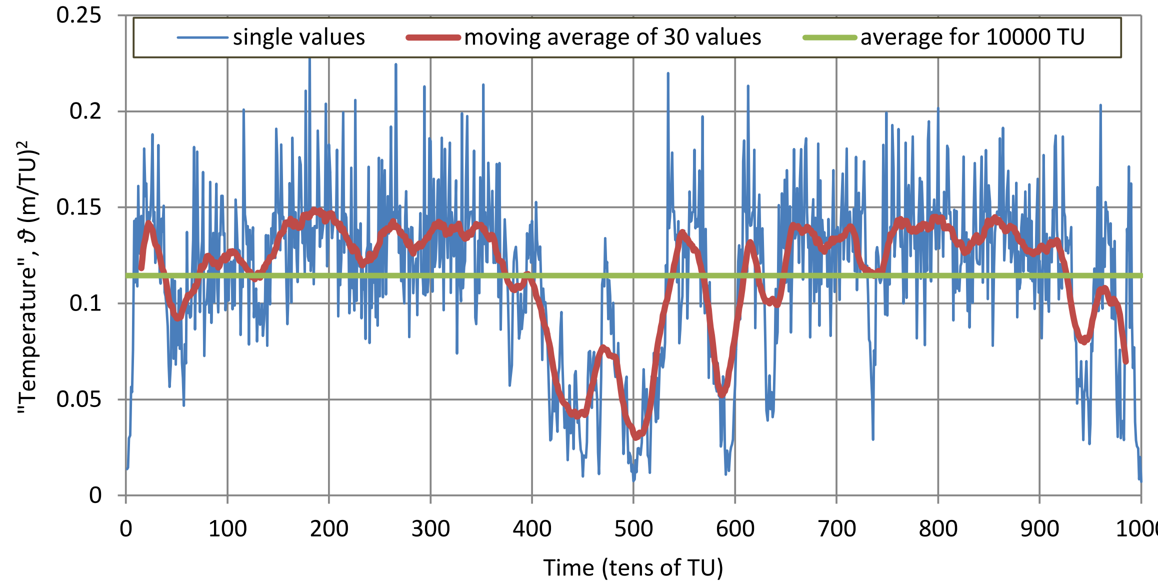


Contrary to the common perception, positive dependence/persistence substantially deteriorates predictability over long time scales—but antipersistent improves it.

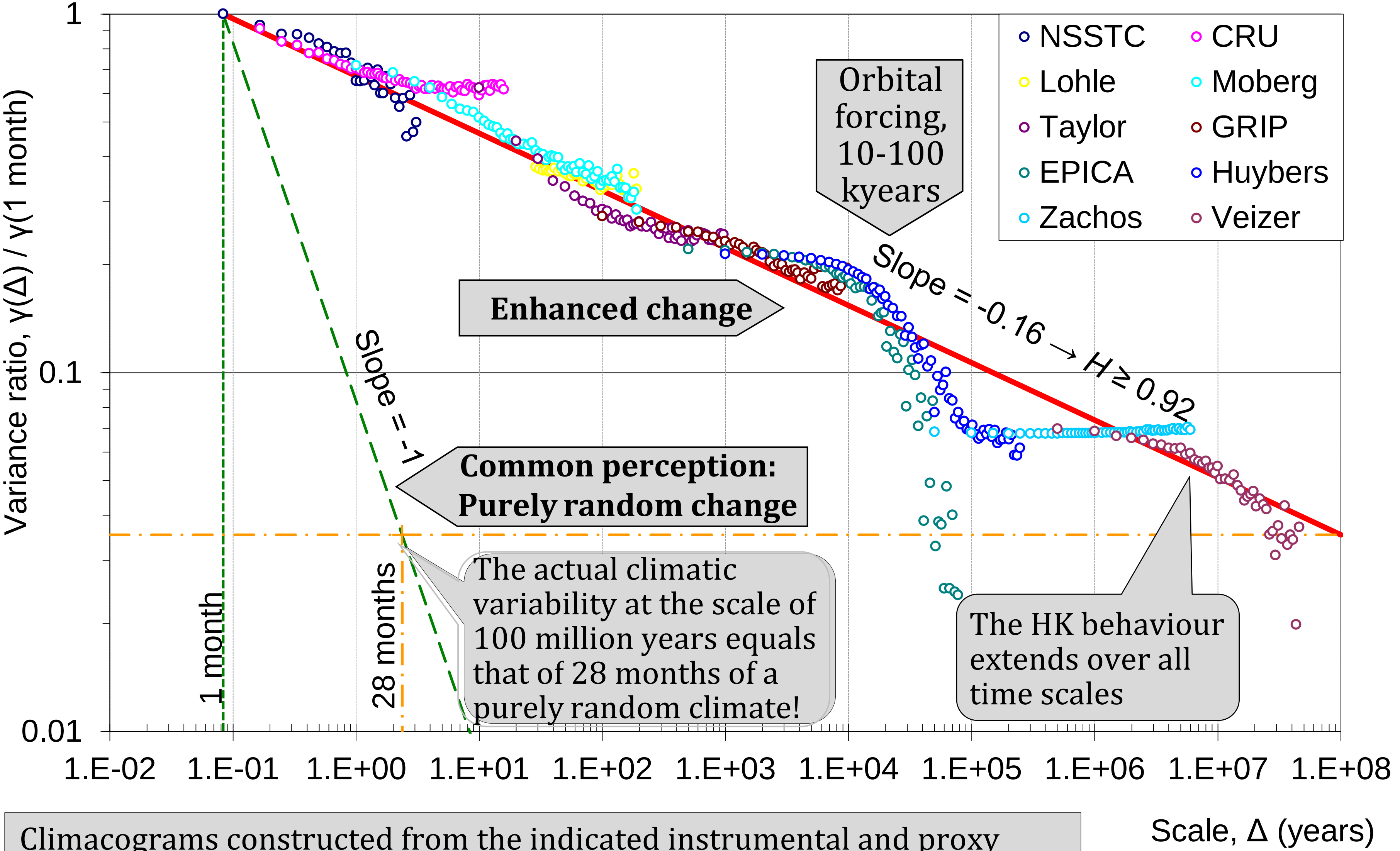
Persistence is not memory — it is change

- The "memory" interpretation of persistence, while being the most common, may be a reflection of linear deterministic thinking.
- An antipersistent process (upper graph) is characterized by (anti-) dependence in time, but primarily by **resistance to long-term change**.
- A persistent process (lower graph) is also characterized by dependence in time, but primarily by occurrence of **long-term change**.
- There is no long-term memory mechanism in the toy model.





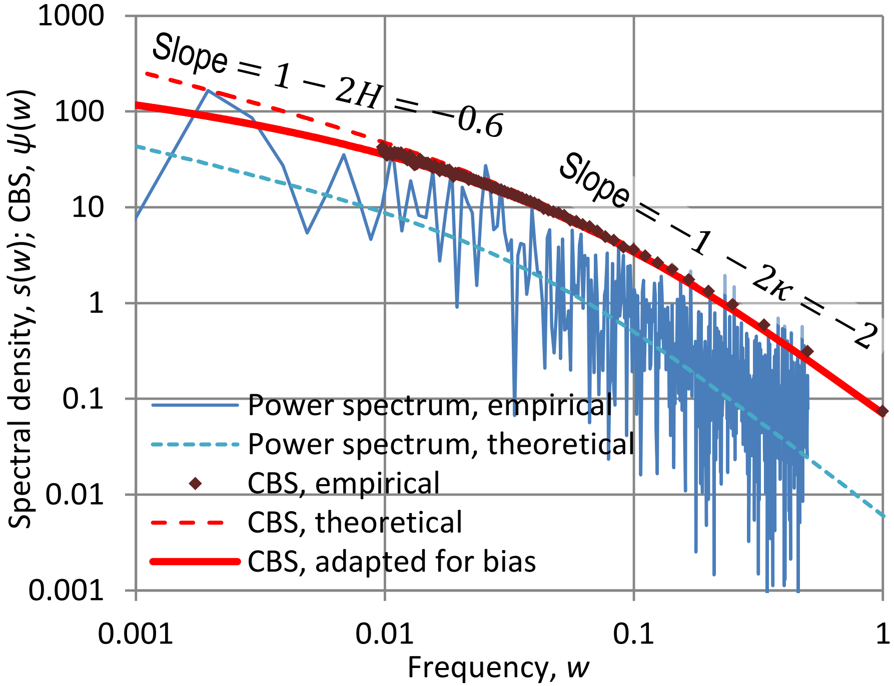
Change and persistence are the rule (antipersistence is an exception)

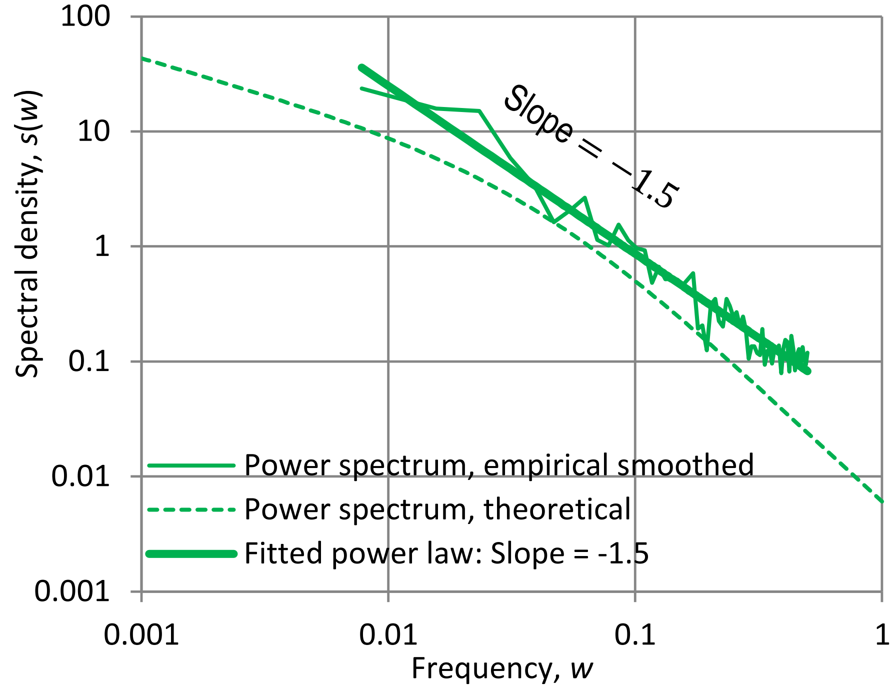


Climacograms constructed from the indicated instrumental and proxy data series (Markonis and Koutsoyiannis, 2013)

Part 4: Some practical suggestions

Tip 1: Data analysis should be consistent with theory

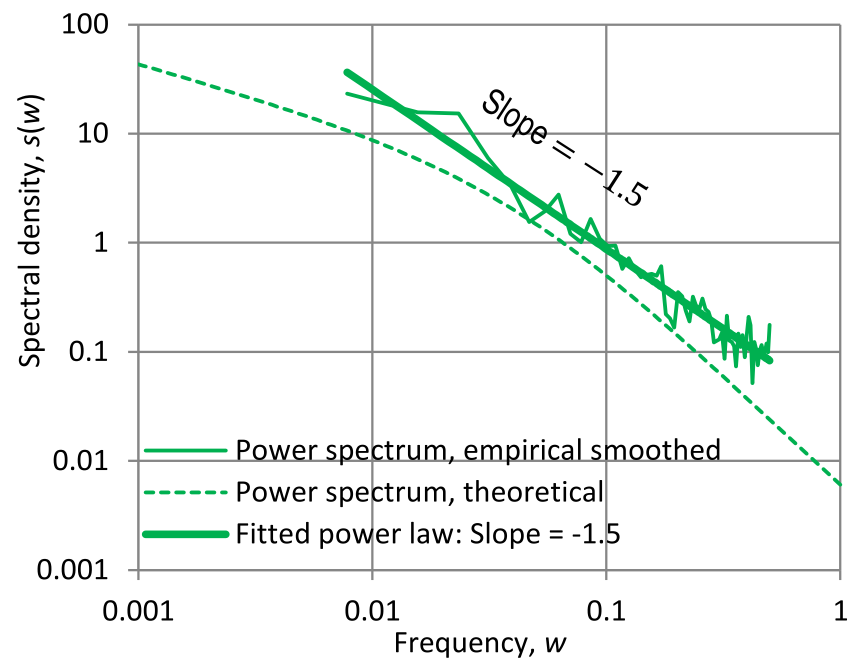




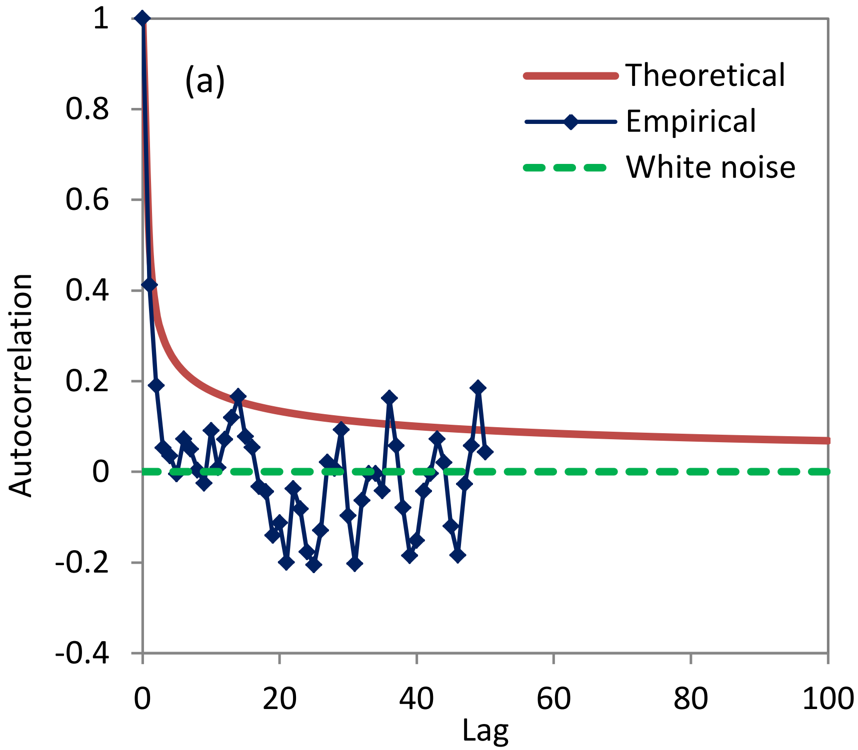
- In the example, 1024 data points have been generated from the HHK process with $\kappa = 0.5$ and $H = 0.8$ (and $\alpha = \lambda = 1$).
- The standard power spectrum (left graph) is too rough to make inference (to recover the underlying model and its parameters).
- Smoothing the power spectrum (by averaging from 8 segments; right graph) makes things even worse in terms of high bias and estimated slope.

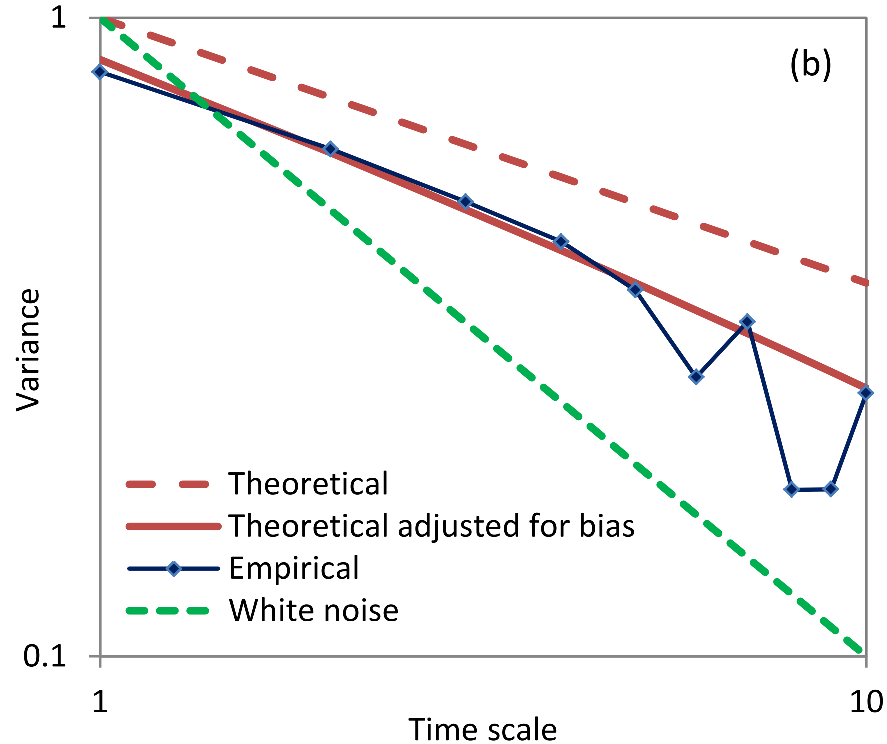
Data analysis should be consistent with theory (contd.)

- Error 1: Model misspecification: a unique power law instead of a law with varying slope with different asymptotic slopes.
- Error 2: Parameter misrepresentation: the power law slope –1.5 does not represent anything.
- Error 3: Total theoretical failure: the slope on the left tail of the power spectrum (here the unique one) cannot be steeper than –1.
  - In making inference from data, the assumption of ergodicity is tacitly made.
  - A stochastic process with slope steeper than –1 on the left tail of the power spectrum is nonergodic (see proof in Koutsoyiannis, 2013b,c)
  - Thus the slope –1.5 estimated from the data is absurd.
- Remedy:
  - (a) awareness of theory;
  - (b) use of algorithms consistent with theory;
  - (c) use of proper stochastic tools (in this case CBS rather than power spectrum).



Tip 2: Proper stochastic tools should be used in model identification and fitting

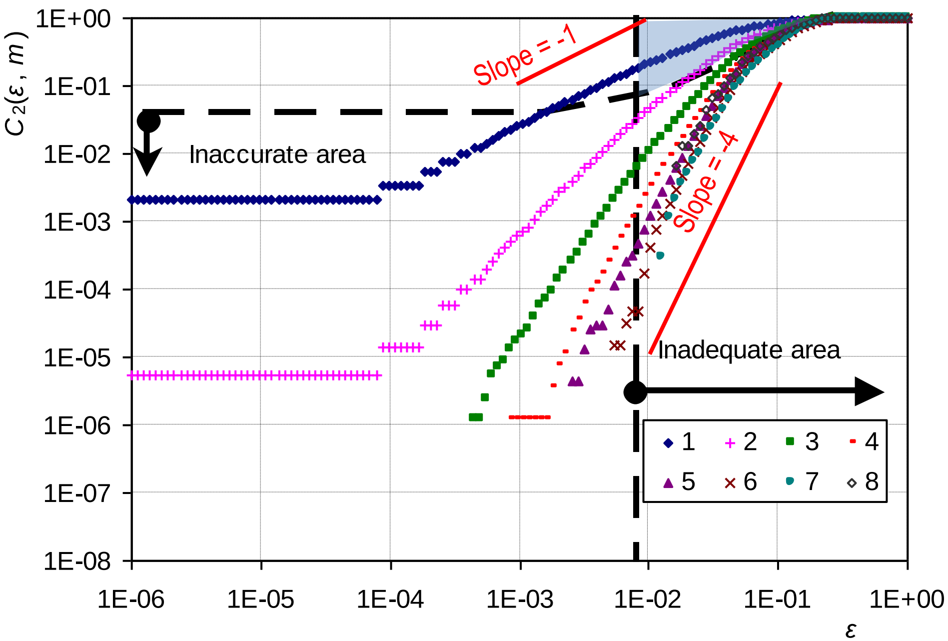




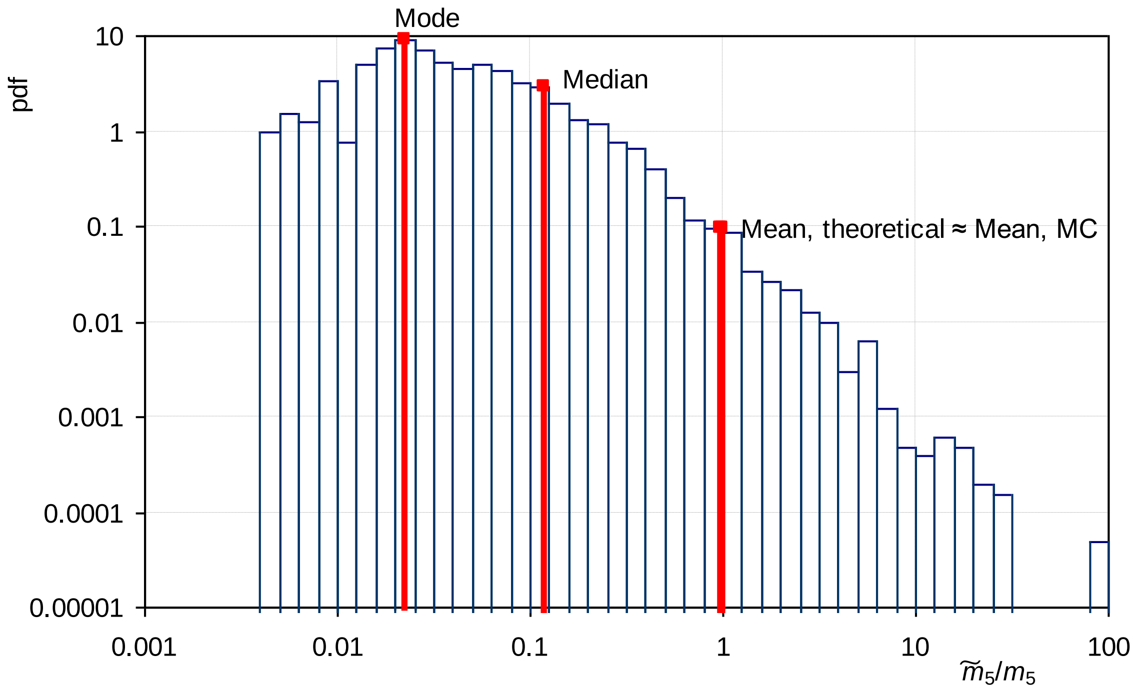
- The autocorrelogram and the climacogram were constructed from a time series of 100 terms generated from the HHK model with $H = 0.79$.
- The empirical autocorrelation does not give any hint that the time series stems from a process with long-term persistence.
- The climacogram unveils the underlying LTP process.

Tip 3: Most quantities calculated from data are statistical quantities

- Many studies have identified low-dimensional deterministic chaos in hydrological and other geophysical processes.
- Typically, they have used the so-called "correlation sum" or "correlation integral" $C_2(\varepsilon, m)$ (and its log-log slope).
- In spite its name, the correlation sum is just the probability that the distance of any two points sampled in an m-dimensional space is smaller than ε.
- Estimation of probability from data is a statistical task; because this probability for small ε turns out to be very small, the reliability of estimates is too low.
- Inattentive interpretation of the graph referring to rainfall data (Koutsoyiannis, 2006) would conclude that rainfall is a deterministic process with dimension <4.
- However, only the shaded area corresponds to statistically reliable estimations, and the only reliable conclusion is trivial: that the dimensionality is > 1 (e.g. ∞).



Tip 4: Inference from data requires awareness of the properties of statistical quantities



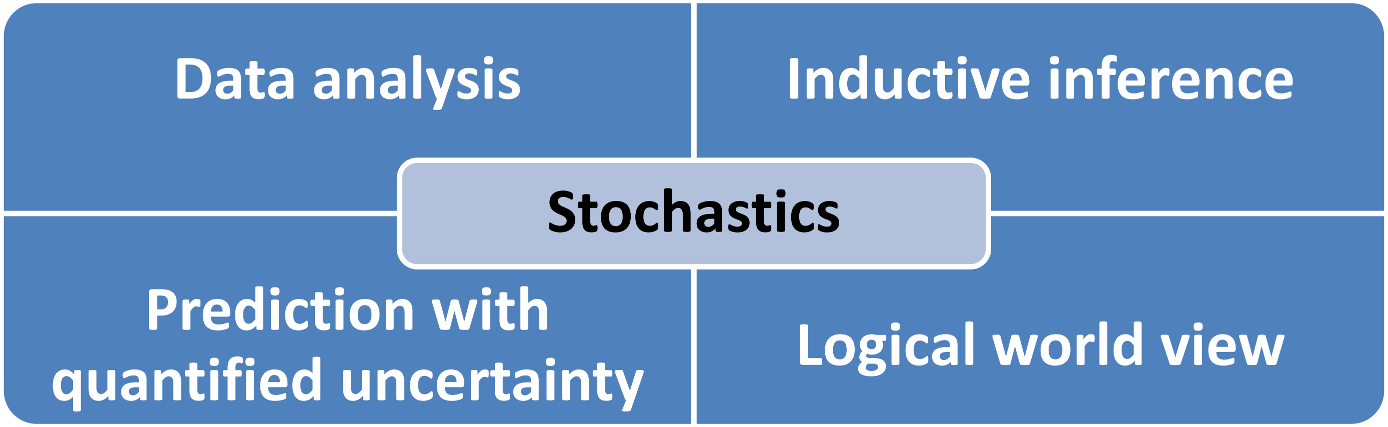
- High-order statistical moments have been very popular in multifractal studies.
- However, the example illustrates that high-order moments have no information content.
- The graph presents results of Monte Carlo simulation for the fifth moment of a Pareto distribution with shape parameter 0.15 for sample size n = 100 (Papalexiou et al. 2010).
- Here the theory guaranties that there is no estimation bias; however the distribution function is enormously skewed.
- The mode is nearly two orders of magnitude less than the mean and the probability that a calculation, based on data, will reach the mean is two orders of magnitude lower than the probability of obtaining the mode.

Tip 5: Attentive use of concepts and notation is extremely important

- Random variables should be distinguished from regular variables both conceptually and notation-wise:
  - Example 1: What is the probability of a certain ordering of two like quantities? **Reply**: We need to specify the nature of the two quantities.
    Illustration: Assume that $\underline{x}$ and $\underline{y}$ are independent random variables uniformly distributed in [0, 1]. Then $P\{\underline{x} \le \underline{y}\} = 0.5$ but $P\{\underline{x} \le y\} = y$ (assuming that y is a realization of $\underline{y}$).
  - Example 2: Does conditioning on available information decrease uncertainty (i.e., entropy)? **Reply**: **YES** but only if we are aware of the concepts; namely: $\Phi[\underline{x}|\underline{y}] \le \Phi[\underline{x}]$ (but $\Phi[\underline{x}|y] \nleq \Phi[\underline{x}]$).
    Illustration: Assume that $\underline{x}$ and $\underline{y}$ denote the dry ($x, y = 0$) or wet ($x, y = 1$) state of today ($\underline{x}$) and yesterday ($\underline{y}$) and that $P\{\underline{x} = 1\} = 0.2$, $P\{\underline{x} = 1|\underline{y} = 1\} = 0.3$, $P\{\underline{x} = 1|\underline{y} = 0\} = 0.1$. Then the entropy is:
    - unconditionally: $\Phi[\underline{x}] = -0.8 \ln 0.8 - 0.2 \ln 0.2 = 0.5$;
    - conditionally on yesterday being wet $\Phi[\underline{x}|1] = \Phi[\underline{x}|\underline{y} = 1] = -0.7 \ln 0.7 - 0.3 \ln 0.3 = 0.61$; so $\Phi[\underline{x}|y] > \Phi[\underline{x}]$ (likewise, $\Phi[\underline{x}|0]=0.33$);
    - conditionally on information about yesterday $\Phi[\underline{x}|\underline{y}]=0.8 \times 0.33 + 0.2 \times 0.61 = 0.38$ (Papoulis, 1991, pp. 172, 564); thus $\Phi[\underline{x}|\underline{y}] \le \Phi[\underline{x}]$.
- Concepts defined within stochastics should be interpreted within stochastics (failure to follow this rule may lead to statements like "*stationarity is dead*").

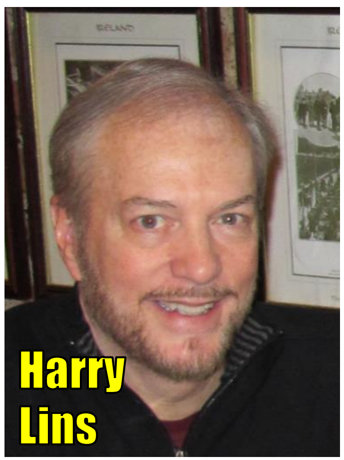
Epilogue

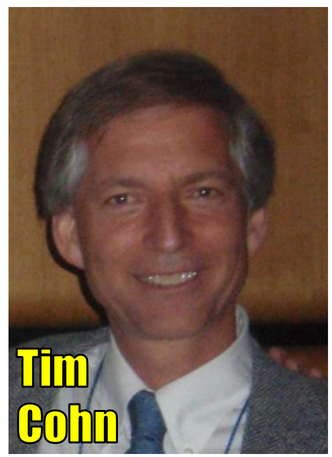
- Thanks to Ludwig Boltzmann, statistics has become a vital part of physics.
- Thanks to Niels Bohr, Werner Heisenberg, and others giants of quantum mechanics, we know that uncertainty is an intrinsic property of the world.
- Thanks to Henri Poincaré, we know that uncertainty dominates also in the macroscopic world.
- Thanks to Edward Lorenz, we know that this is particularly the case in geophysics (1963).
- Thanks to Kurt Gödel we know that solving all problems by deduction is infeasible, and thus we have to theorize inductive reasoning.
- Thanks to Andrey Kolmogorov, we have a well-founded mathematical theory of stochastics.



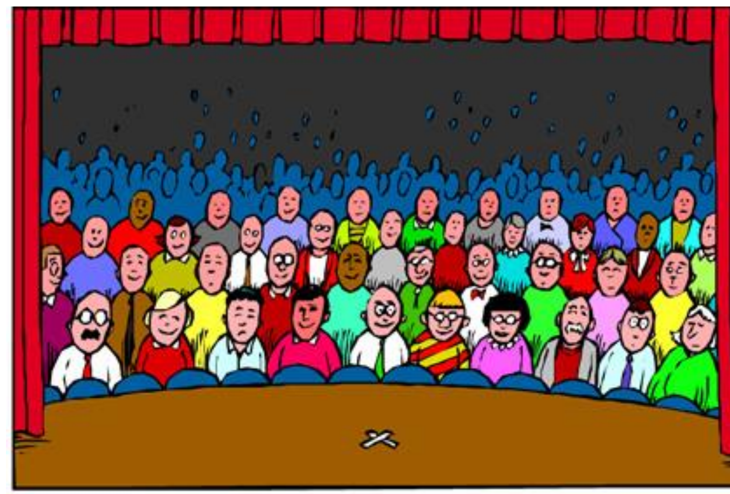
Thanks…

… to the AGU Honors Committee for selecting me as the 2014 Lorenz Lecturer





… to my colleagues and friends who inspire me to keep studying stochastics and who conspired to nominate me for the Lorenz Lecture



… to all of you who kindly attended the Lecture

… to the Greek General Secretariat for Research and Technology for funding (CRESSENDO project; grant #5145)

References

- Dimitriadis, P., and D. Koutsoyiannis, Climacogram vs. autocovariance and power spectrum in stochastic modelling for Markovian and Hurst-Kolmogorov processes, *Stochastic Environmental Research and Risk Assessment*, 2015 (in review).
- Gneiting, T., and M. Schlather, Stochastic models that separate fractal dimension and the Hurst effect, *Society for Industrial and Applied Mathematics Review*, 46 (2), 269-282, 2004.
- Jaynes, E.T. Information theory and statistical mechanics, *Physical Review*, 106 (4), 620-630, 1957.
- Kang, H. S., S. Chester and C. Meneveau, Decaying turbulence in an active-grid-generated flow and comparisons with large-eddy simulation, *J. Fluid Mech.*, 480, 129-160, 2003.
- Klemes, V., An unorthodox physically-based stochastic treatment of tree rings, *XXIV General Assembly of the International Union of Geodesy and Geophysics,* Perugia, International Union of Geodesy and Geophysics, 2007 (www.itia.ntua.gr/723/).
- Koutsoyiannis, D., On the quest for chaotic attractors in hydrological processes, *Hydrological Sciences Journal*, 51 (6), 1065–1091, 2006.
- Koutsoyiannis, D., A random walk on water, *Hydrology and Earth System Sciences*, 14, 585–601, 2010.
- 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., Physics of uncertainty, the Gibbs paradox and indistinguishable particles, *Studies in History and Philosophy of Modern Physics*, 44, 480–489, 2013a.
- Koutsoyiannis, D., *Encolpion of stochastics: Fundamentals of stochastic processes*, Department of Water Resources and Environmental Engineering – National Technical University of Athens, Athens, 2013b (www.itia.ntua.gr/1317/).
- Koutsoyiannis, D. F. Lombardo, E. Volpi and S.M. Papalexiou. Is consistency a limitation? — Reply to "Further (monofractal) limitations of climactograms" by Lovejoy et al., *Hydrol. Earth Syst. Sci. Discuss.*, 10, C5397, 2013c (www.itia.ntua.gr/1343/).
- Koutsoyiannis, D., Entropy: from thermodynamics to hydrology, *Entropy*, 16 (3), 1287–1314, 2014.
- Koutsoyiannis Generic and parsimonious stochastic modelling for hydrology and beyond, *Hydrological Sciences Journal*, 2015 (accepted with minor revisions).
- Lorenz, E.N., Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20 (2), 130–141, 1963.
- Lorenz, E.N., Letter from Edward N. Lorenz, *Nonlinear Dynamics in Geosciences* (ed by A. Tsonis and J. B. Elsner), p. vi, Springer, 2007.
- Mackey, M.C., *Time's Arrow: The Origins of Thermodynamic Behavior*, Dover, Mineola, NY, USA, 175 pp., 2003.
- Markonis, Y., and D. Koutsoyiannis, Climatic variability over time scales spanning nine orders of magnitude: Connecting Milankovitch cycles with Hurst–Kolmogorov dynamics, *Surveys in Geophysics*, 34 (2), 181–207, 2013.
- Papalexiou, S.M., D. Koutsoyiannis, and A. Montanari, Mind the bias!, *STAHY Official Workshop: Advances in statistical hydrology*, Taormina, Italy, International Association of Hydrological Sciences, 2010 (www.itia.ntua.gr/985/).
- Papoulis, A., *Probability, Random Variables and Stochastic Processes*, 3rd ed., McGraw-Hill, New York, NY, USA, 1991.
- Stein, M.L., *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, 1999.