Lectio Inauguralis
Pontificia Universidad Javeriana
Bogotá, Colombia, 17 February 2014

Glimpsing God playing dice over water and climate

Demetris Koutsoyiannis

Department of Water Resources and Environmental Engineering Faculty of Civil Engineering
National Technical University of Athens, Greece

(dk@itia.ntua.gr – http://itia.ntua.gr/dk)

Presentation available online: http://itia.ntua.gr/1426/

Human dice games are old



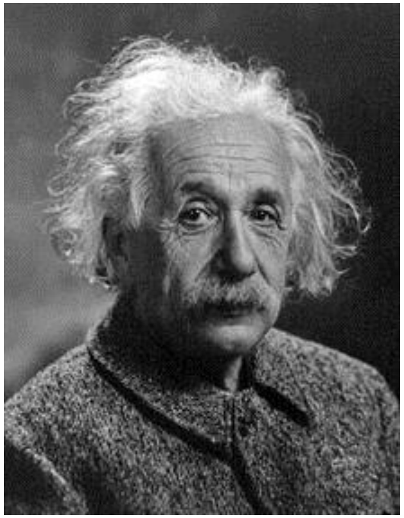
- All these dice are of the period 580-570 BC from Greek archaeological sites:
  - Left, Kerameikos Ancient Cemetery Museum, Athens, photo by author
  - Middle: Bronze die (1.6 cm), Greek National Archaeological Museum, www.namuseum.gr/object-month/2011/apr/7515.png
  - Right: Terracotta die (4 cm) from Sounion, Greek National Archaeological Museum, http://www.namuseum.gr/object-month/2011/dec/dies_b.png
- Much older dice (up to 5000 years old) have been found in Asia (Iran, India).

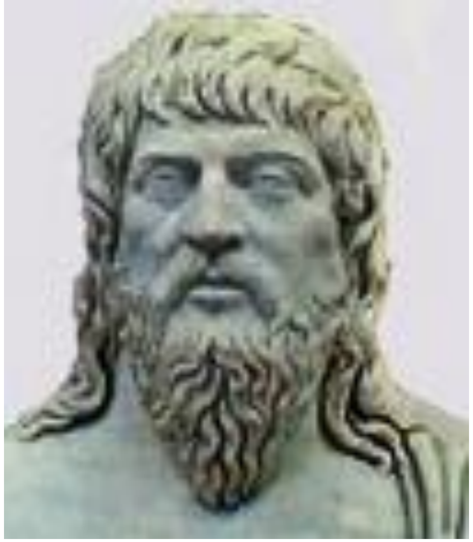
Modern Colombian dice (art objects by Obando de Pasto)

With special thanks to Prof. Efraín Antonio Dominguez Calle

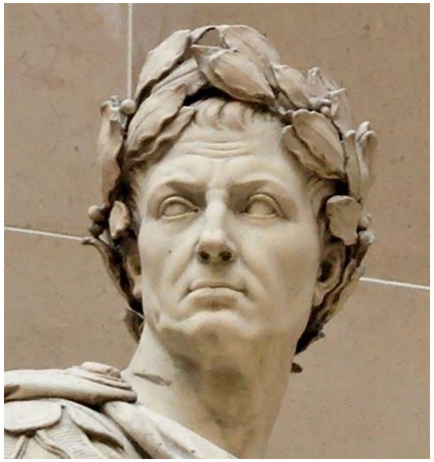
Some famous quotations about dice

Jedenfalls bin ich überzeugt, daß der nicht würfelt
I, at any rate, am convinced that He [God] does not throw dice
(Albert Einstein, in a letter to Max Born in 1926)





Αἰών παῖς ἐστι παίζων πεσσεύων
Time is a child playing, throwing dice
(Heraclitus; ca. 540-480 BC; Fragment 52)



Ἀνερρίφθω κύβος	***Iacta alea est***
Let the die be cast	The die has been cast
[Plutarch's version, in Greek]	[Suetonius's version, in Latin]

(Julius Caesar, 49 BC, when crossing Rubicon River)

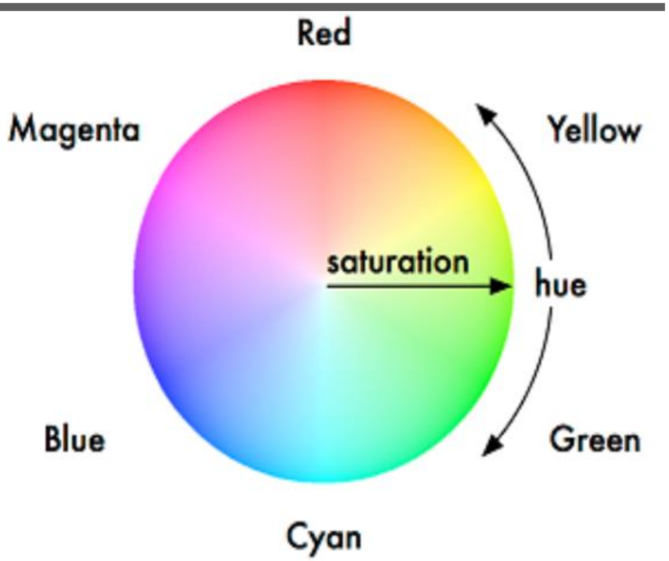
Physical setting of die motion

- The die motion is described by the laws of classical (Newtonian) mechanics and is determined by:
  - Die characteristics:
    - dimensions (incl. imperfections with respect to cubic shape),
    - density (incl. inhomogeneities).
  - Initial conditions which determine the die motion:
    - position,
    - velocity,
    - angular velocity.
  - External factors which influence the die motion:
    - acceleration due to gravity,
    - viscosity of the air,
    - friction factors of the table,
    - elasticity moduli of the dice and the table.
- Knowing all these, in principle we should be able to predict the motion and outcome solving the deterministic equations of motion.
- However the die has been the symbol of randomness (paradox?).

A die experiment: researchers and apparatus

Technical details

- Each side of the die is painted with a different colour: blue, magenta, red, yellow and green (basic primary colours) and black (highly traceable from the video as the box is white).
- The visualization is done via a camera with frame frequency of 120 Hz. The video is analyzed to frames and numerical codes are assigned to coloured pixels (based on the HSL system) and position in the box (two Cartesian coordinates).

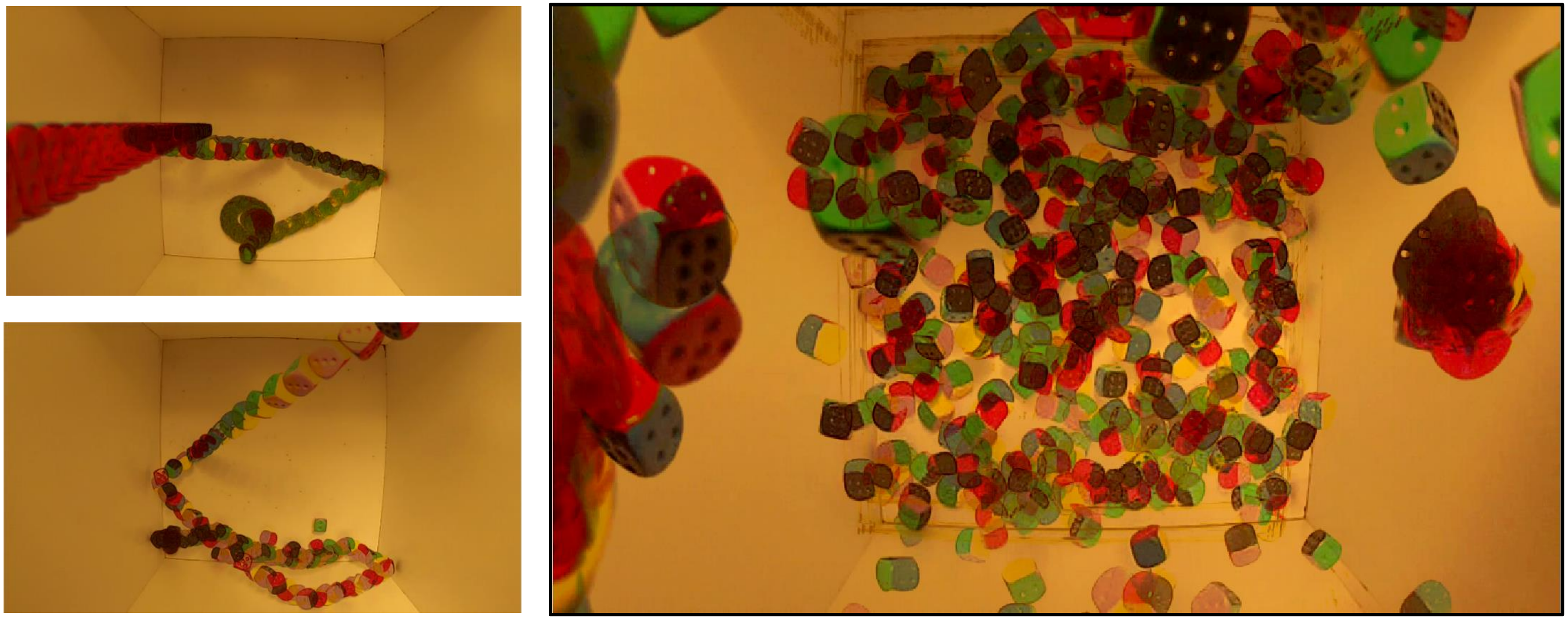


Color wheel of primary colours hue and saturation (www.highend.com/support/controllers/documents/html/en/sect-colour_matching.htm)

- The area of each colour traced by the camera is estimated and then non-dimensionalized with the total traced area of the die. Pixels not assigned to any colour (due to low camera analysis and blurriness) are typically ~30% of the total traced die area.
- In this way, the orientation of the die in each frame is known (with some observation error) through the colours seen looking from above.
- The audio is transformed to a non-dimensional index from 0 to 1 (with 1 indicating the highest noise produced in each video) and can be used to locate the times in which the die hits the bottom or the sides of the box.

Experiments made

- In total, 123 die throws were performed, 52 with initial angular momentum and 71 without.
- The height from which the die was thrown remained constant for all experiments (15 cm).
- However, the initial orientation of the die varied .
- The duration of each throw varied from 1 to 9 s.



A selection of frames from die throws 48 (upper left) and 78 (lower left) and video for 78 (right).

Representation of die orientation

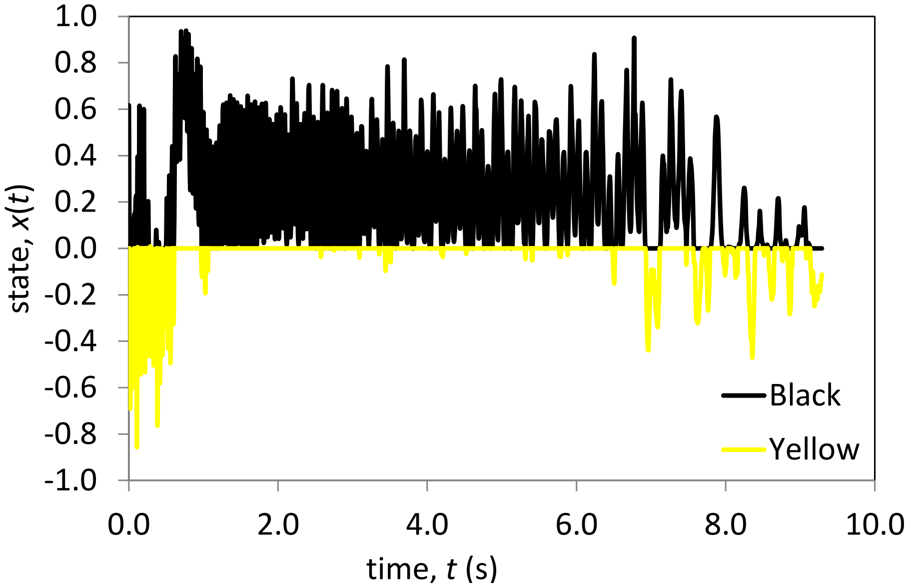
- The evolution of die orientation is most important as it determines the outcome.
- The orientation can be described by three variables representing proportions of each colour, as seen from above, each of which varies in [−1,1] (see table).

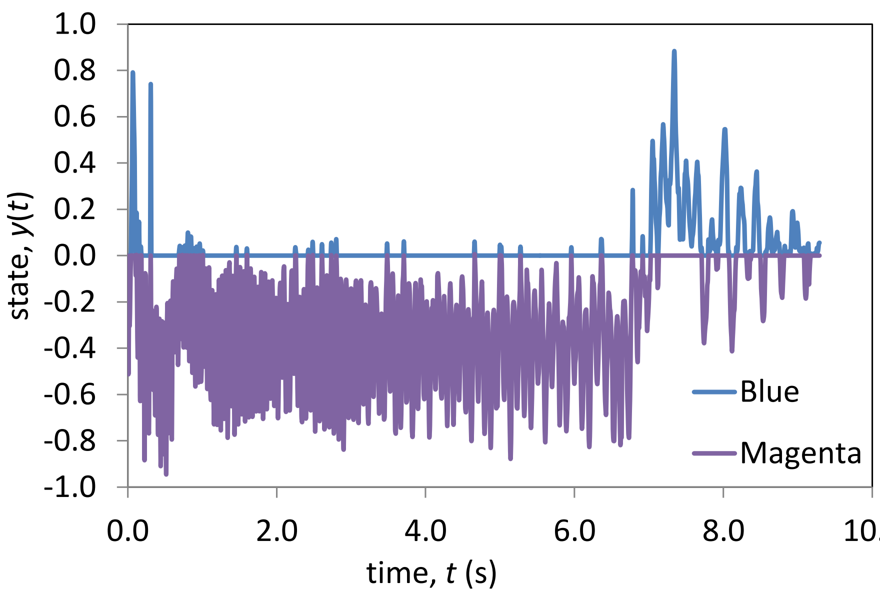
Value →	−1		+1	
Variable ↓	Colour	Pips	Colour	Pips
x	yellow	1	black	6
y	magenta	3	blue	4
z	red	5	green	2

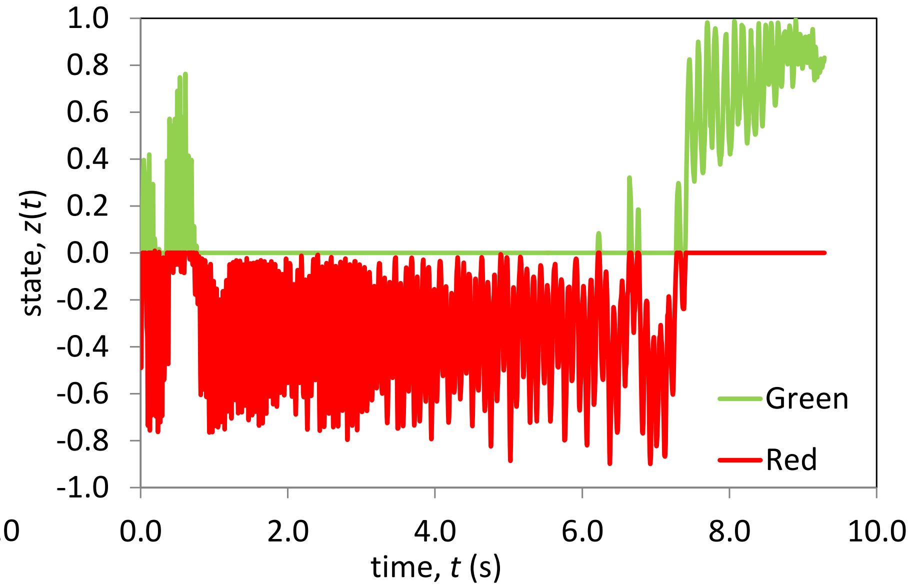
Example:
$x = -0.25$ (yellow)
$y = 0.4$ (blue)
$z = -0.35$ (red)

Persistence and change in die orientation

- The typical evolution of die orientation is far from a random pattern fluctuating around the horizontal axis.
- There is persistence of a particular orientation for relatively long times.
- At times, the dominant orientation switches to a new one.
- These are seem in the figures which refer to experiment 78.



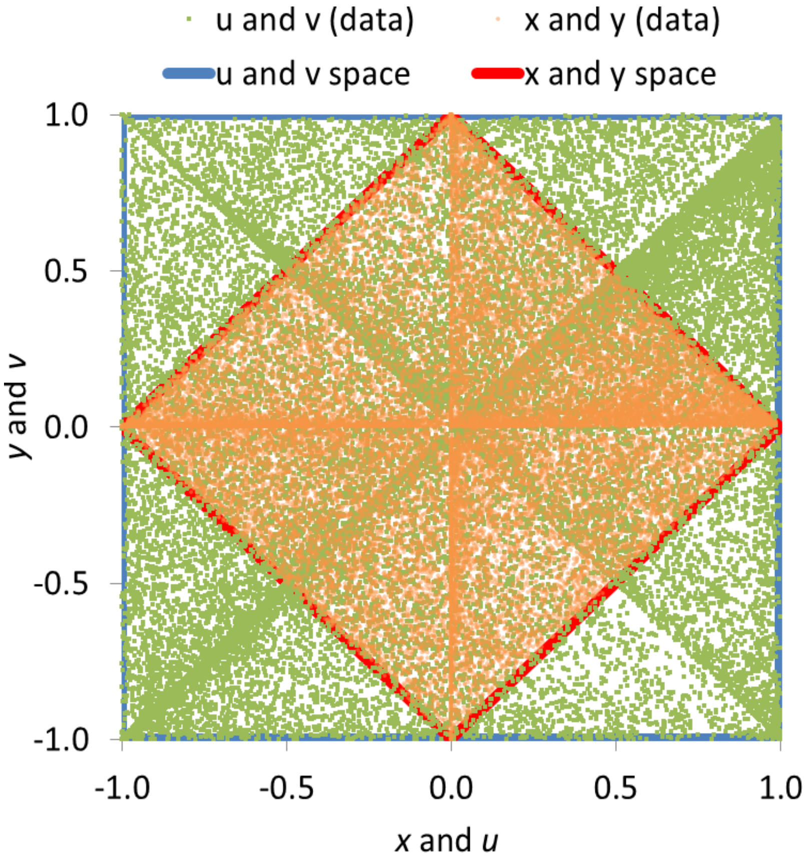


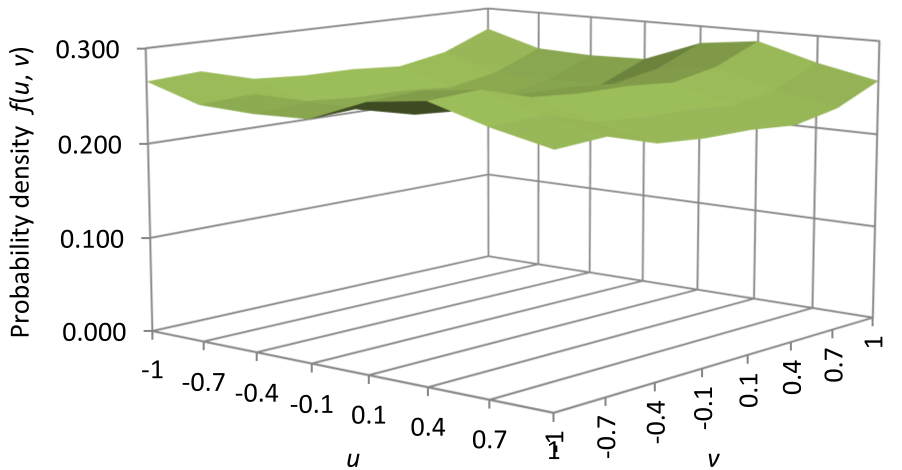


Alternative representation

- The variables x, y and z are not stochastically independent of each other because of the obvious relationship $|x| + |y| + |z| = 1$.
- The following transformation produces a set of independent variables u, v, w, where u, v vary in $[-1,1]$ and w is two-valued $(-1,1)$:

$$\begin{Bmatrix} u = x + y \\ v = x - y \\ w = \mathrm{sign}(z) \end{Bmatrix} \leftrightarrow \begin{Bmatrix} x = (u+v)\,/\,2 \\ y = (u-v)\,/\,2 \\ z = w(1 - \max(|u|, |v|)) \end{Bmatrix}$$





The plot of all experimental points and the probability density function show that u and v are independent and fairly uniformly distributed except that states for which $u \pm v = 0$ (corresponding to one of the final outcomes) are more probable.

A simple forecast model

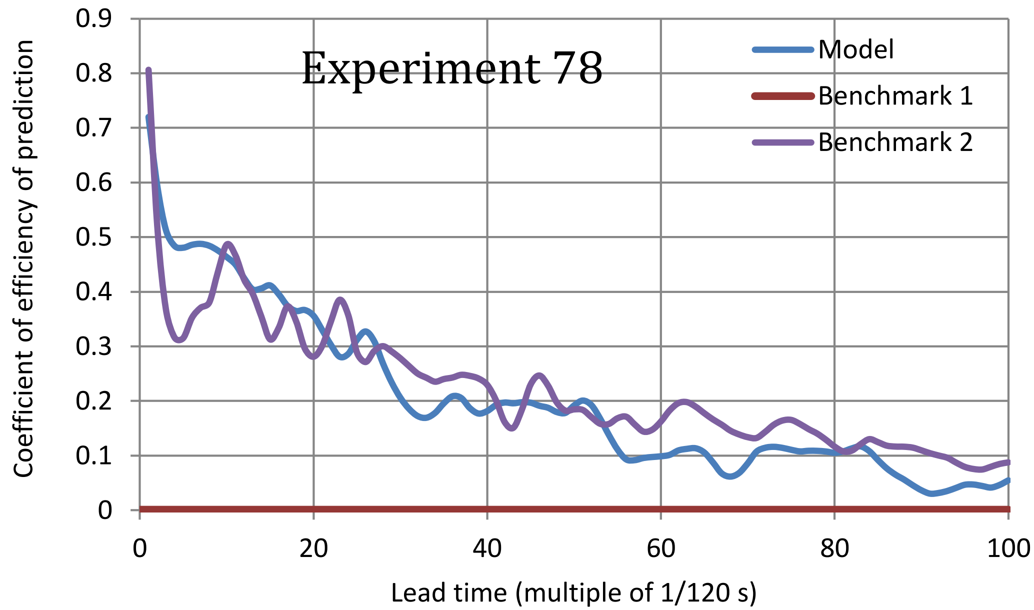
- We use a deterministic model, purely data-driven, known as the analogue model (e.g. see Koutsoyiannis et al. 2008); it does not assume any mathematical expression between variables.
- To predict $\boldsymbol{s}((t+l)\Delta)$, $l = 1, 2, \ldots$, based on past states $\boldsymbol{s}((t-p)\Delta)$, $p = 0, 1, \ldots, m$, where $\boldsymbol{s} = (x, y, z)$:
  - We search the data base of all experiments to find similar states (neighbours or analogues) $\boldsymbol{s}^i((t^i - p)\Delta)$, so that $\sum_{p=1}^{m} \left\| \boldsymbol{s}^i\left((t^i - p)\Delta\right) - \boldsymbol{s}\left((t-p)\Delta\right) \right\|^2 \le c$, where c is an error threshold.
  - Assuming that n such neighbours are found, for each one we find the state at time $(t^i + l)\Delta$, i.e. $\boldsymbol{s}^i((t^i + l)\Delta)$ and calculate an average state $\tilde{\boldsymbol{s}}((t+l)\Delta) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{s}^i\left((t^i + l)\Delta\right)$.
  - We adjust $\tilde{\boldsymbol{s}}((t+l)\Delta)$ to ensure consistency : $x((t+l)\Delta) = \tilde{x}((t+l)\Delta)/\tilde{s}$, $y((t+l)\Delta) = \tilde{y}((t+l)\Delta)/\tilde{s}$, $z((t+l)\Delta) = \tilde{z}((t+l)\Delta)/\tilde{s}$ where $\tilde{s} \coloneqq |\tilde{x}((t+l)\Delta)| + |\tilde{y}((t+l)\Delta)| + |\tilde{z}((t+l)\Delta)|$.
- After preliminary investigation, it was found that a number of past values $m = 10$ and a threshold $c = 0.5$ work relatively well.

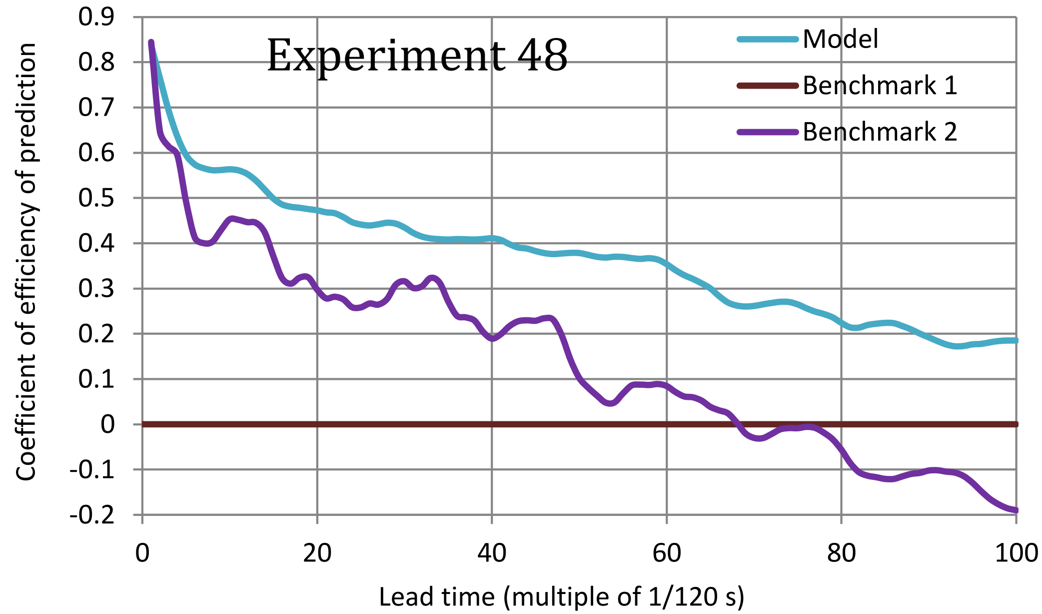
Benchmark models

- The forecast model is checked against two naïve benchmark models.
- In **Benchmark 1** the prediction is the average state, i.e. $\boldsymbol{s}((t+l)\Delta) = \mathbf{0}$. Although the zero state is not permissible per se, the Benchmark 1 is useful, as any model worse than that is totally useless.
- In **Benchmark 2** the prediction is the current state, i.e. $\boldsymbol{s}((t+l)\Delta) = \boldsymbol{s}(t\Delta)$, regardless of how long the lead time $l\Delta$ is. Because of the persistence, it is expected that the Benchmark 2 will work well for relatively small lead times.
- For the performance assessment of the forecast and benchmark models (as well as for the comparison thereof), the coefficient of efficiency is used, defined as CE = 1 – e^2/σ^2, where e^2 is the mean squared error of prediction and σ^2 is the variance of the true time series:
  - A value CE = 1 indicates perfect forecast (no error).
  - The value CE = 0 is the CE of Benchmark 1 (purely statistical).
  - A value CE < 0 indicates a useless model (worse that purely statistical).

Results

- For lead times $l\Delta \lesssim 1/10$ s, the forecast model, as well as Benchmark 2 provide relatively good predictions (efficiency $\gtrsim$ 0.5)
- Predictability is generally superior than pure statistical (Benchmark 1) for lead times $l\Delta \lesssim 1$ s.
- For longer lead times the state is unpredictable.
- The final outcome is unpredictable from just the initial state.





What is randomness?

- Given that the die outcome is random, is there an "agent of randomness"? Where and when did it act?
- Are there two mutually exclusive types of events or processes—deterministic and random (or stochastic)?
- Can we distinguish the natural events into these two types—with random being those that we do not understand or explain?
- Are natural process really composed of mixtures of these two parts or components—deterministic and random?
- Can each part be further subdivided into subparts (e.g., deterministic part = periodic + aperiodic/trend)?
- Does the deterministic part represent a cause-effect relationship, which is the subject physics and science (**the "good"**)?
- Is the random part a *noise* that has little relationship with science and no relationship with understanding (**the "evil"**)?

The replies to the above questions are commonly positive.

However, positive replies imply a naïve and incorrect view of randomness and a manichean perception of Nature

Hydrol. Earth Syst. Sci., 14, 585–601, 2010
www.hydrol-earth-syst-sci.net/14/585/2010/

Hydrology and Earth System Sciences

HESS Opinions
"A random walk on water"

D. Koutsoyiannis

Department of Water Resources and Environmental
of Athens, Greece

Invited contribution by D. Koutsoyiannis, recipient

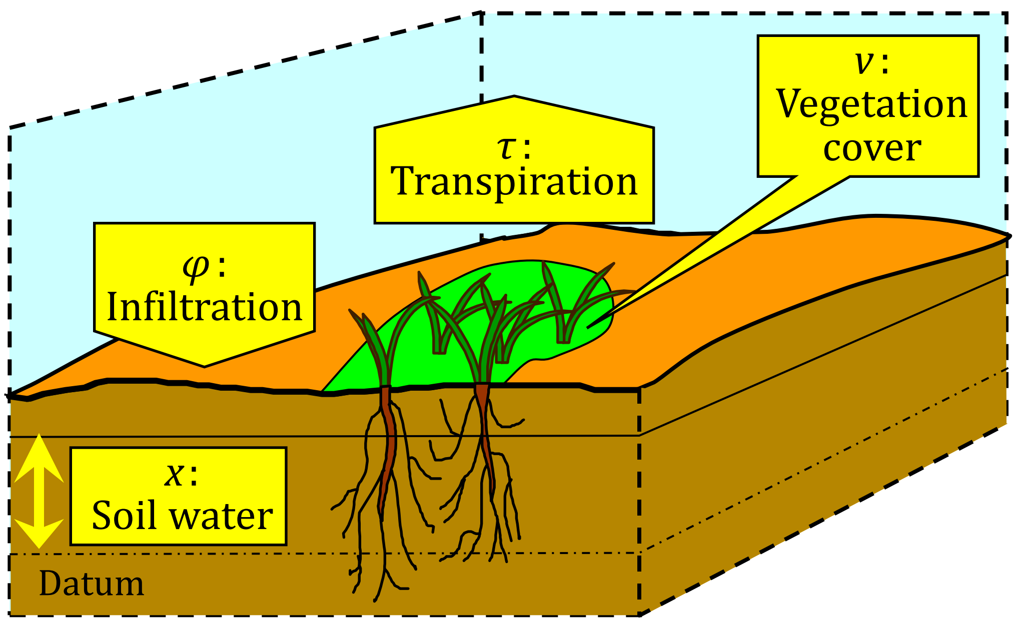
Received: 5 October 2009 – Published in Hydrol. E
Revised: 19 March 2010 – Accepted: 19 March 201

Abstract. According to the traditional notion of ran
and uncertainty, natural phenomena are separated
mutually exclusive components, random (or stocha
deterministic. Within this dichotomous logic, the d
istic part supposedly represents cause-effect relat
and, thus, is physics and science (the "good"),
randomness has little relationship with science

Emergence of randomness from determinism: A toy model of a caricature hydrological system

- The toy model is designed intentionally simple.
- Only infiltration, transpiration and water storage are considered.
- The rates of infiltration φ and potential transpiration τ_p are **constant**.

Nothing in the model is set to be random.



- Discrete time: i ("years").
- Constants (per "year")
  - Input: φ = 250 mm;
  - Potential output: τ_p = 1000 mm.
- State variables (**a 2D dynamical 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: $\tau_i = v_i \, \tau_p$
- Water balance

$$x_i = \min(x_{i-1} + \varphi - v_{i-1}\tau_p \, , \alpha)$$

The toy model at equilibrium

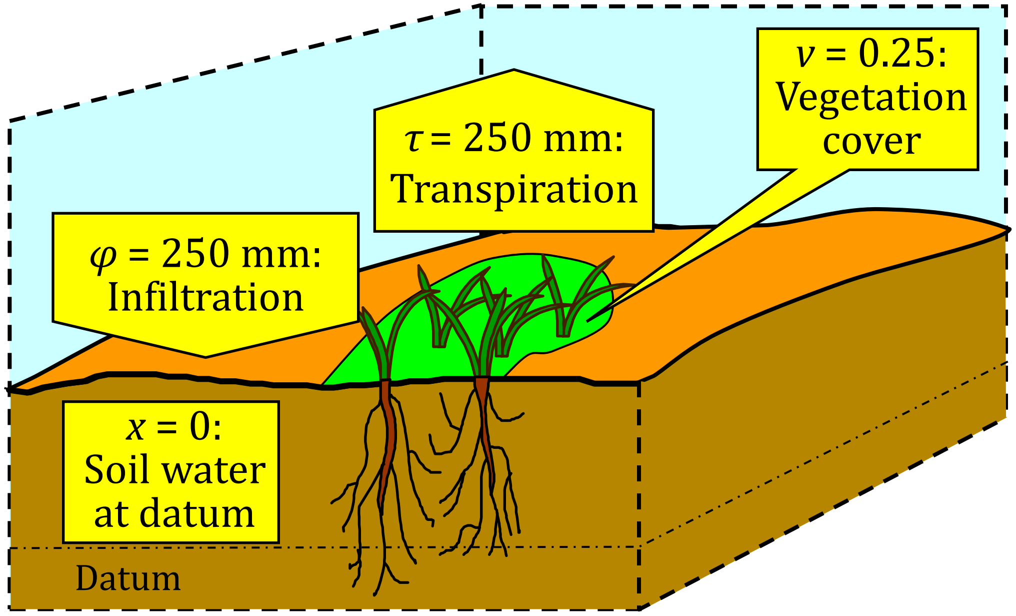
- If at some time $i - 1$:

  $$v_{i-1} = \varphi / \tau_p = 250/1000 = 0.25$$

  then the water balance results in

  $$x_i = x_{i-1} + \varphi - v_{i-1}\tau_p = x_{i-1}$$

- Continuity of system dynamics demands that for some x_{i-1}, $v_i = v_{i-1}$. Without loss of generality we set this value $x_{i-1} = 0$ (this defines a datum for soil water).



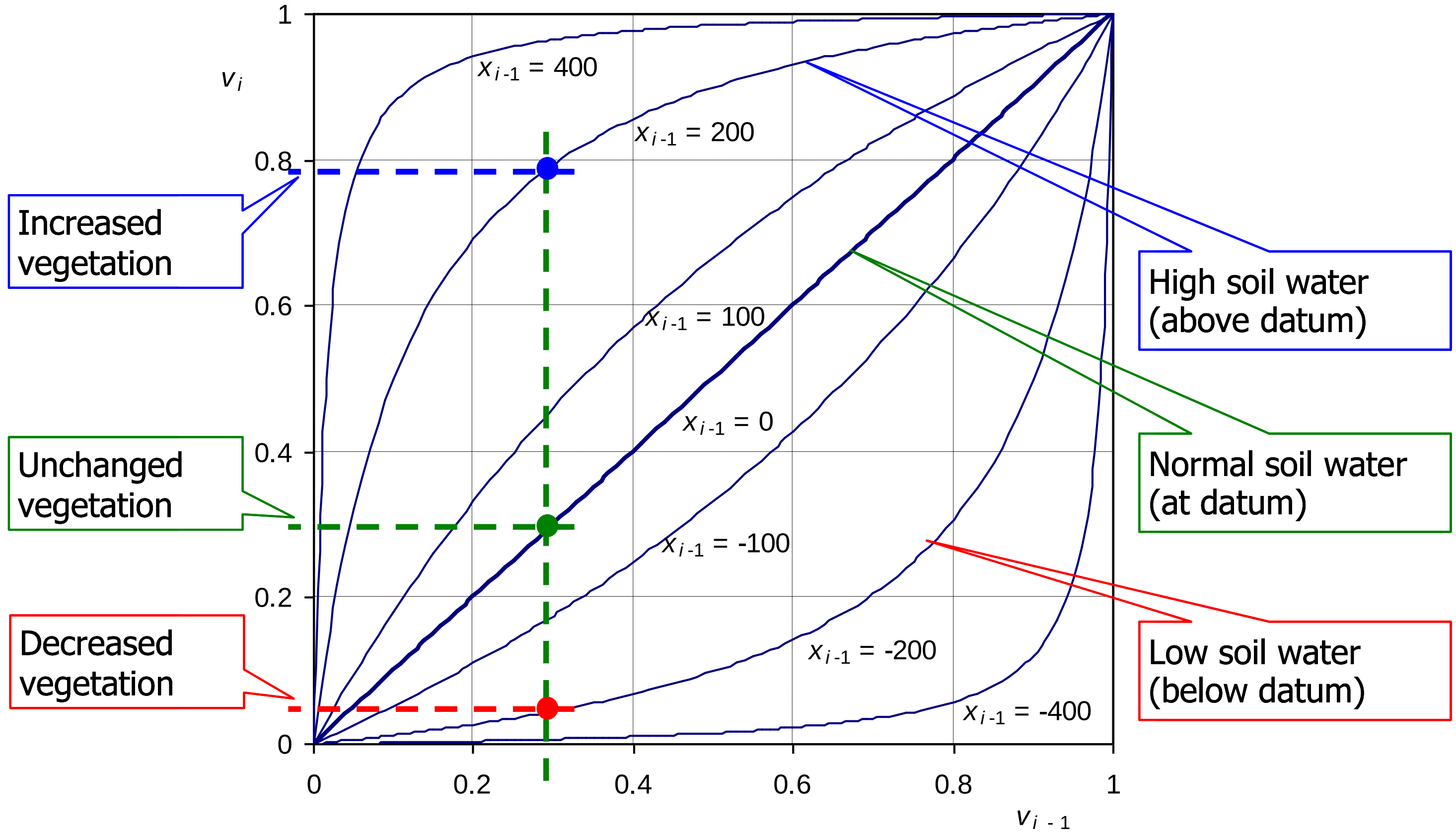
- Thus the system state:

  $$v_i = v_{i-1} = 0.25$$
  $$x_i = x_{i-1} = 0$$

  represents the equilibrium of the system.

- If the system arrives at equilibrium it will stay there for ever.
- This state can be called "the dead equilibrium".

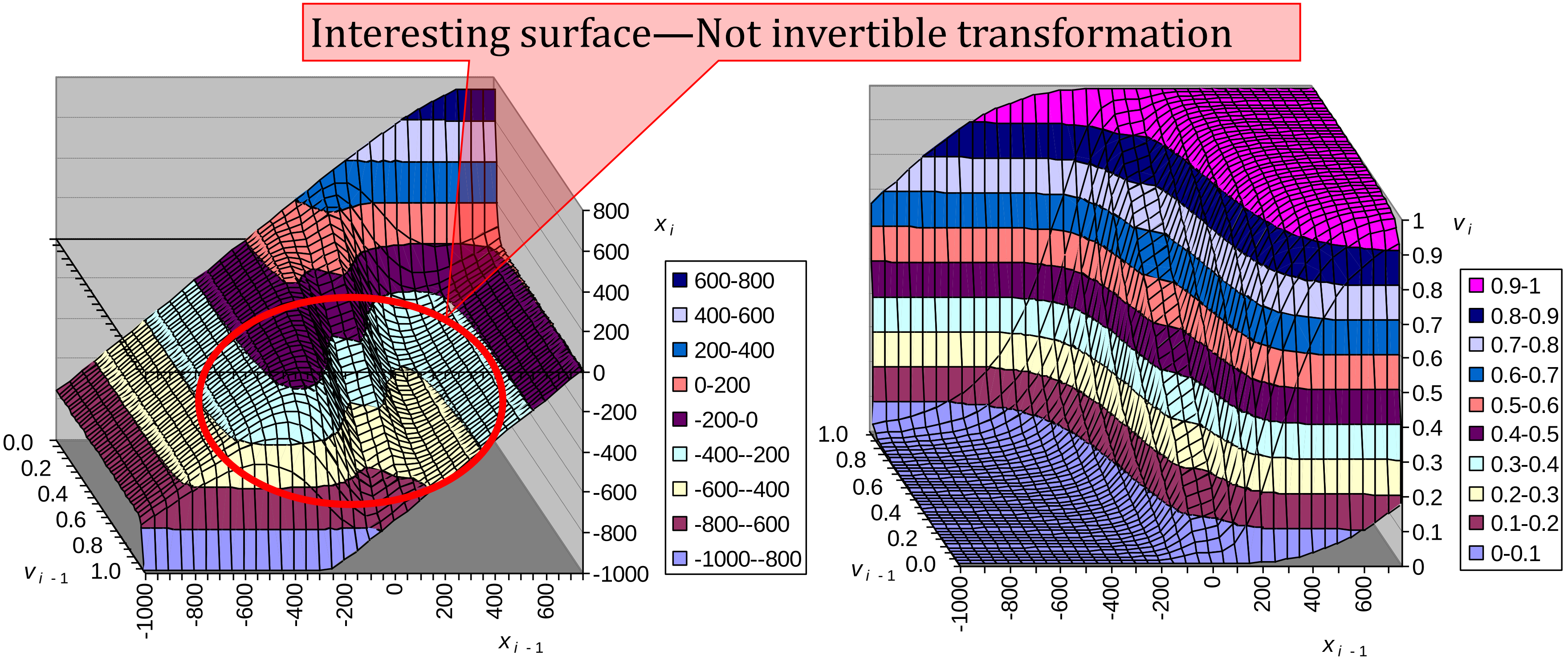
Non-equilibrium state – conceptual dynamics of vegetation



The graph is described by the following equation (with $\beta = 100$ mm—a standardizing constant):

$$v_i = \frac{\max(1 + (s_{i-1}/\beta)^3, 1)\, v_{i-1}}{\max(1 - (s_{i-1}/\beta)^3, 1) + (s_{i-1}/\beta)^3\, v_{i-1}}$$

System dynamics



Water balance + Vegetation cover dynamics

$$x_i = \min(x_{i-1} + \varphi - v_{i-1}\tau_{\mathrm{p}}, \alpha)$$

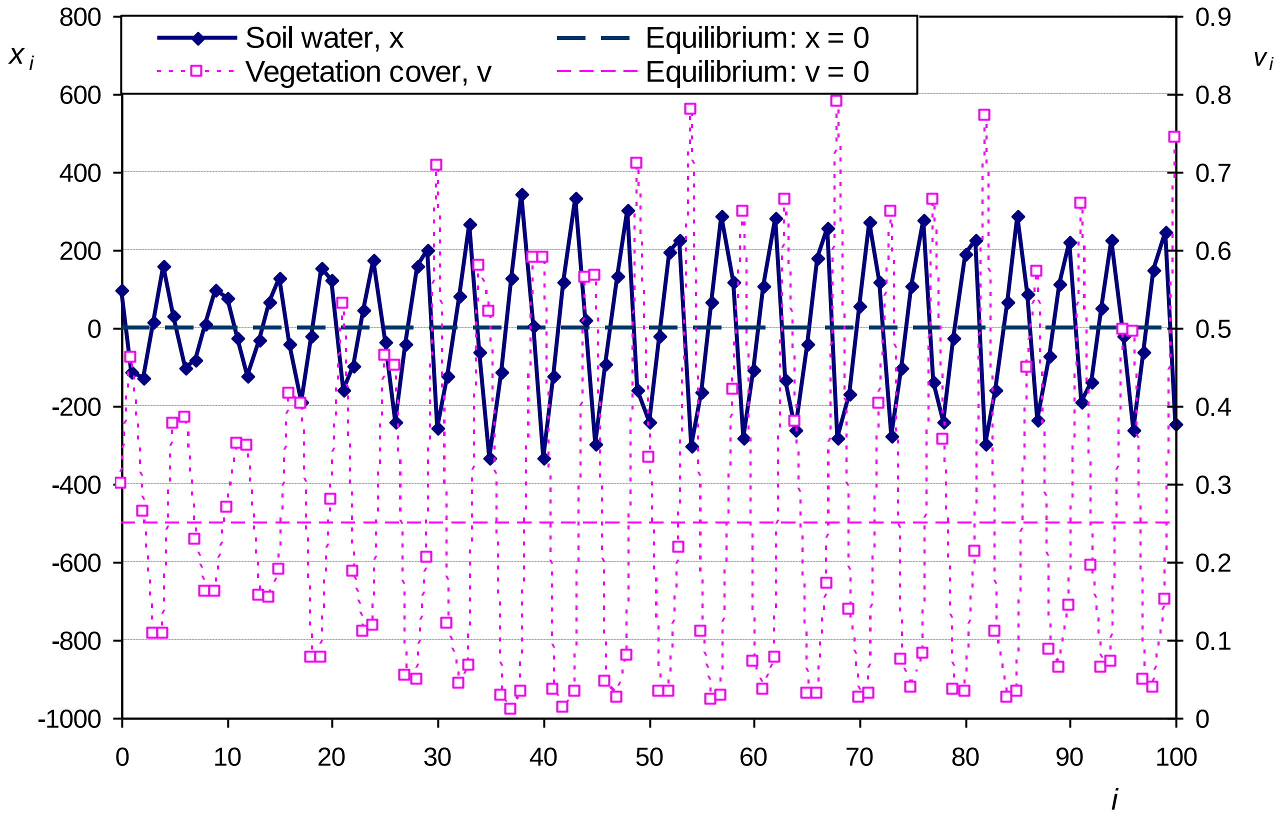
for finite storage ≤ α

$$v_i = \frac{\max(1 + (s_{i-1}/\beta)^3, 1)\, v_{i-1}}{\max(1 - (s_{i-1}/\beta)^3, 1) + (s_{i-1}/\beta)^3 v_{i-1}}$$

Assumed constants: φ = 250 mm, τ_{p} = 1000 mm, α = 750 mm, β = 100 mm.
Easy to program in a hand calculator or a spreadsheet.

Interesting trajectories produced by simple deterministic dynamics

- These trajectories of x and v, for time i = 1 to 100 were produced assuming initial conditions x_0 = 100 mm (≠ 0) and v_0 = 0.30 (≠ 0.25) using a spreadsheet (it can be downloaded from itia. ntua.gr/923/).
- The system state does not converge to the equilibrium.
- The trajectories seem periodic.
- Iterative application of the simple dynamics allows prediction for arbitrarily long time horizons (e.g., x_{100} = –244.55 mm; v_{100} = 0.7423).



Understanding of mechanisms and system dynamics

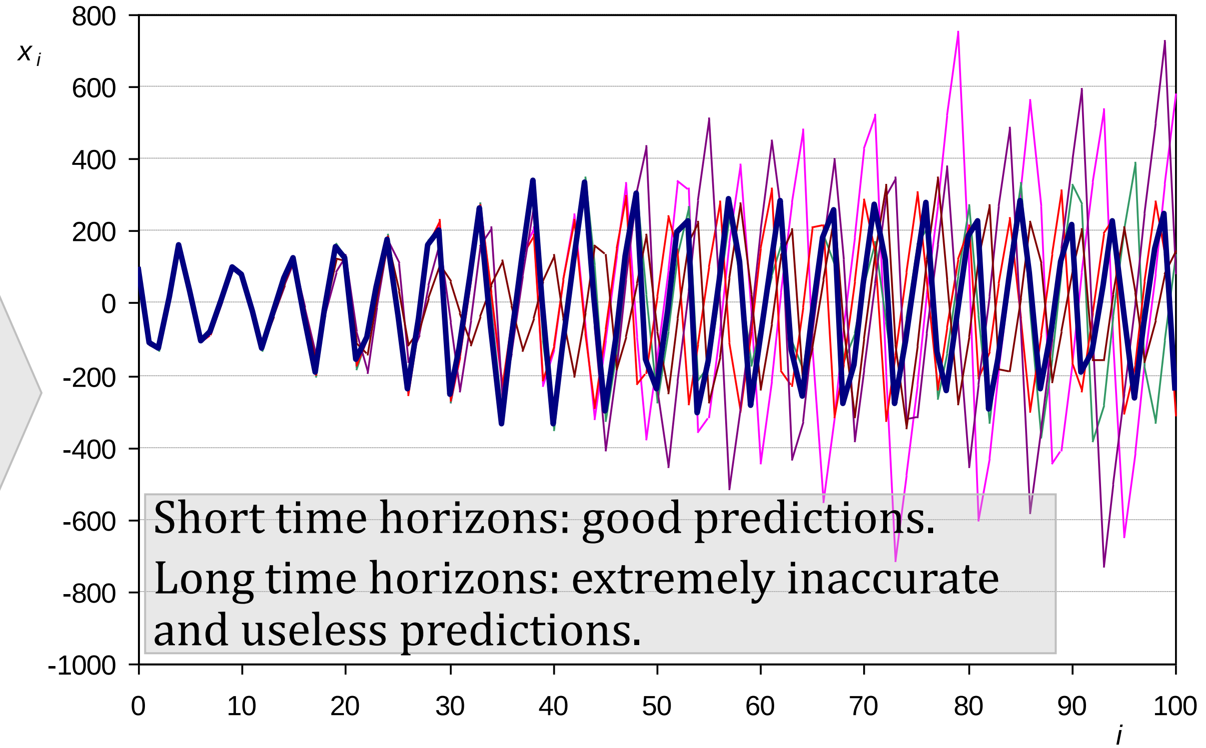
- System understanding—causative relationships:
  - There is water balance (conservation of mass);
  - Excessive soil water causes increase of vegetation;
  - Deficient soil water causes decrease of vegetation;
  - Excessive vegetation causes decrease of soil water;
  - Deficient vegetation causes increase of soil water.
- System dynamics are:
  - Fully consistent with this understanding;
  - Very simple, fully deterministic;
  - Nonlinear, chaotic.

Does deterministic dynamics allow a reliable prediction at an arbitrarily long time horizon?

- Postulate: A continuous (real) variable that varies in time cannot be ever known with full precision (infinite decimal points).
- It is reasonable then to assume that there is some small uncertainty in the initial conditions (initial values of state variables).
- Sensitivity analysis allows to see that a tiny uncertainty in initial conditions may get amplified.

Bold blue line corresponds to initial conditions x_0 = 100 mm, v_0 = 0.30.

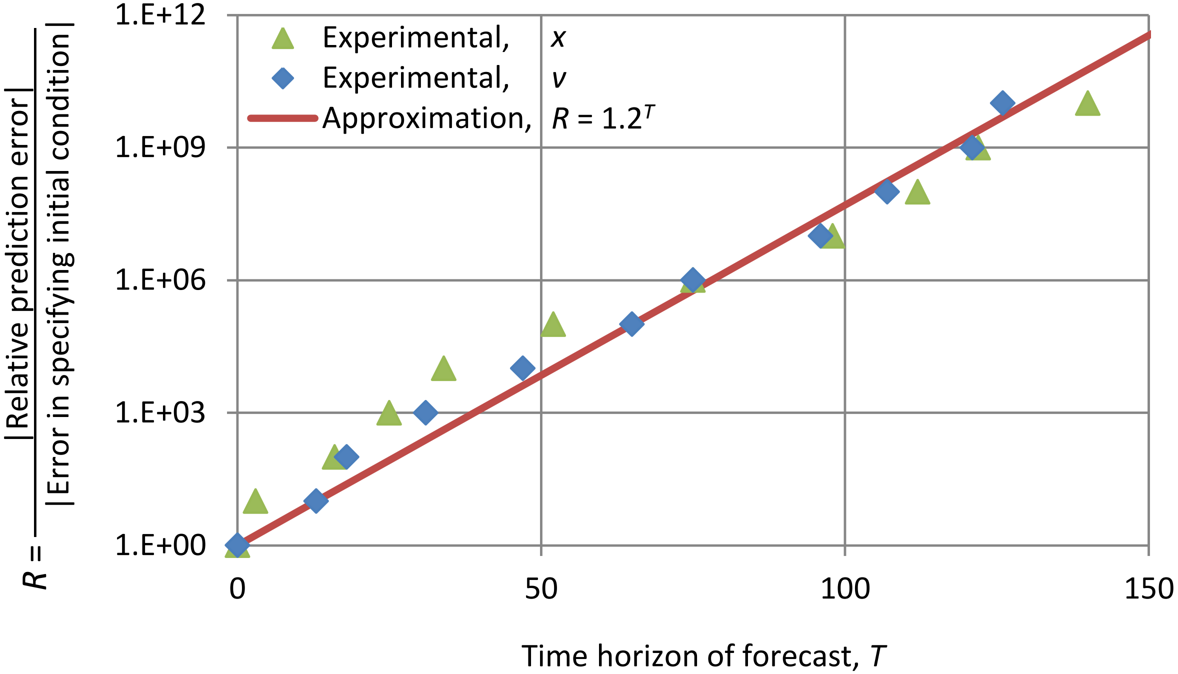
All other lines represent initial conditions slightly (< 1%) different.



Error propagation and limits of predictability

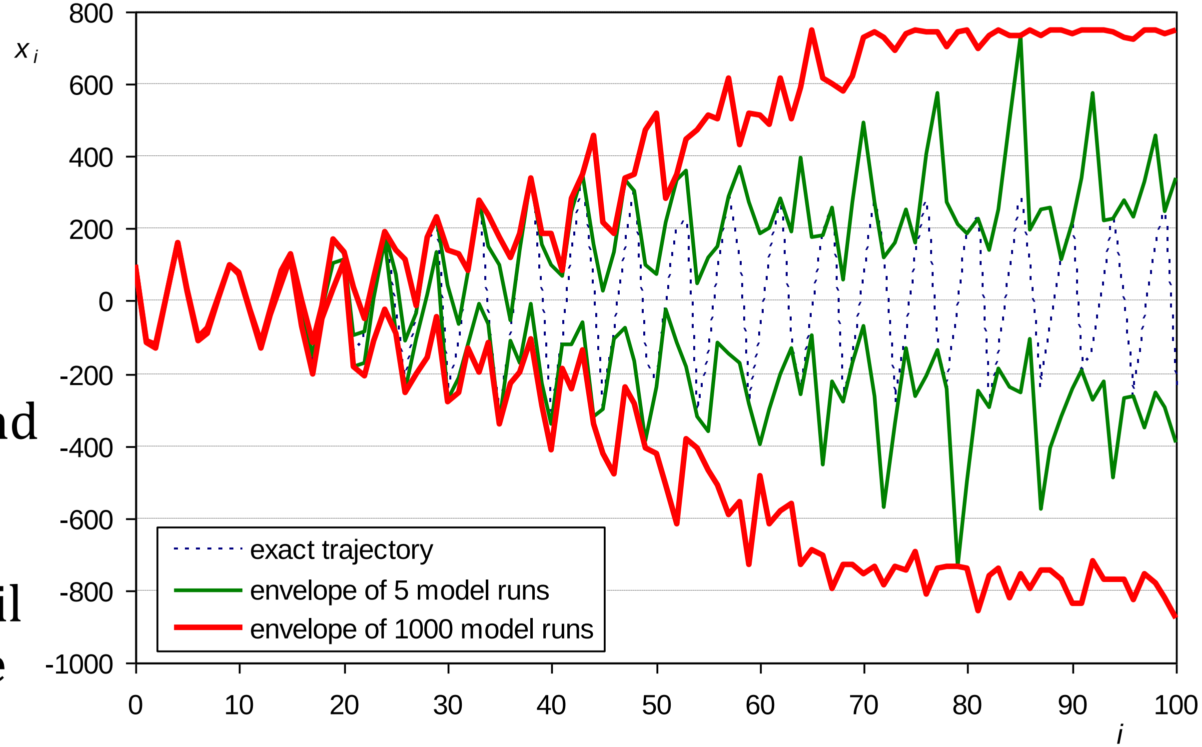
- We assume a small error in specifying the initial condition.
- How does this error propagate in time?
  - Is it kept constant?
  - Does it decrease?
  - Is it magnified?
- For the particular system it increases exponentially.
- If we are satisfied with a prediction error, say, 3 orders of magnitude higher than the initial error, then the limit of deterministic predictability is ~40 time steps.

The error in specifying the initial condition was assumed 10^{-12} . However, the graph remains unchanged for a wide range of values of the initial error.



From exact trajectories to "stream tube" representation

- Instead of a single model run with exact initial conditions, we perform multiple runs with initial conditions contaminated by a small error.
- We can then construct envelop curves of the trajectories.
- This changes our vision of the system evolution: from trajectories to *stream tubes*.
- The problem is that the stream tubes depend on the values of the initial errors and the number of model runs; they widen for more model runs until they cover all feasible space.



From determinism to stochastics

- **Probabilization of uncertainty**: axiomatic reduction from the notion of an uncertain quantity to the notion of a random variable.
- According to the system introduced by **Andrey Kolmogorov** (1933), **probability** is a normalized measure, i.e., a function that maps sets (areas where unknown quantities lie) to real numbers (in the interval [0, 1]).
- **Random variable**: a mathematical object $\underline{x}$ representing all possible outcomes x (also known as realizations) and associating each of them with a probability or a probability density $f(x)$.
- **Stochastics** (modern meaning): **probability + statistics + stochastic processes**.
- **Stochastics** (first use and definition) is the **Science of Prediction**, i.e., the science of measuring as exactly as possible the probabilities of events (**Jakob Bernoulli**, 1713—Ars Conjectandi, written 1684-1689).
- **Stochastic process**: An infinite collection of random variables (the term was introduced by **Andrey Kolmogorov** and **Aleksandr Khinchin** in 1930s).
- **Stochastics** (etymology): < Greek Stochasticos (Στοχαστικός) < Stochazesthai (Στοχάζεσθαι = (1) to aim, point, or shoot (an arrow) at a target; (2) to guess or conjecture (the target) (3) to imagine, think deeply, bethink, contemplate, cogitate, meditate) < Stochos (Στόχος= target).

If one 'stochazetai' (thinks deeply), eventually he goes 'stochastic' (with the probability-theoretical meaning) and he will hit 'stochos' (the target).

The stochastic formulation of system evolution

- We fully utilize the deterministic dynamics: $\boldsymbol{x}_i = \boldsymbol{S}(\boldsymbol{x}_{i-1})$, where $\boldsymbol{x}_i := (x_i, v_i)$ is the vector of the system state and $\boldsymbol{S}$ is the vector function representing the known deterministic dynamics of the system.
- We assume that $f(\boldsymbol{x}_0)$ is known, e.g. a uniform distribution extending 1% around the value $\boldsymbol{x}_0 = (100\text{ mm}, 0.30)$.
- Given the probability density function at time $i - 1$, $f(\boldsymbol{x}_{i-1})$, that of next time i, $f(\boldsymbol{x}_i)$, is given by the **Frobenius-Perron operator** FP, i.e. $f(\boldsymbol{x}_i) = \mathrm{FP}f(\boldsymbol{x}_{i-1})$, uniquely defined by an integral equation (e.g. Lasota and Mackey, 1991), which in our case takes the following form:

$$\mathrm{FP}f(\boldsymbol{x}) = \frac{\partial^2}{\partial x \partial v} \int_{\boldsymbol{S}^{-1}(A)} f(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}$$

  were $A := \{\underline{\boldsymbol{x}} \le (x, v)\}$ and $\boldsymbol{S}^{-1}(A)$ is the **counterimage** of A.
- The equation is easily deduced by standard probability theory.
- Iterative application of the equation can determine the density $f(\boldsymbol{x}_i)$ for any time i — but we may need to calculate a high-dimensional integral.

Stochastics does not disregard the deterministic dynamics: it is included in the counterimage $\boldsymbol{S}^{-1}(A)$.

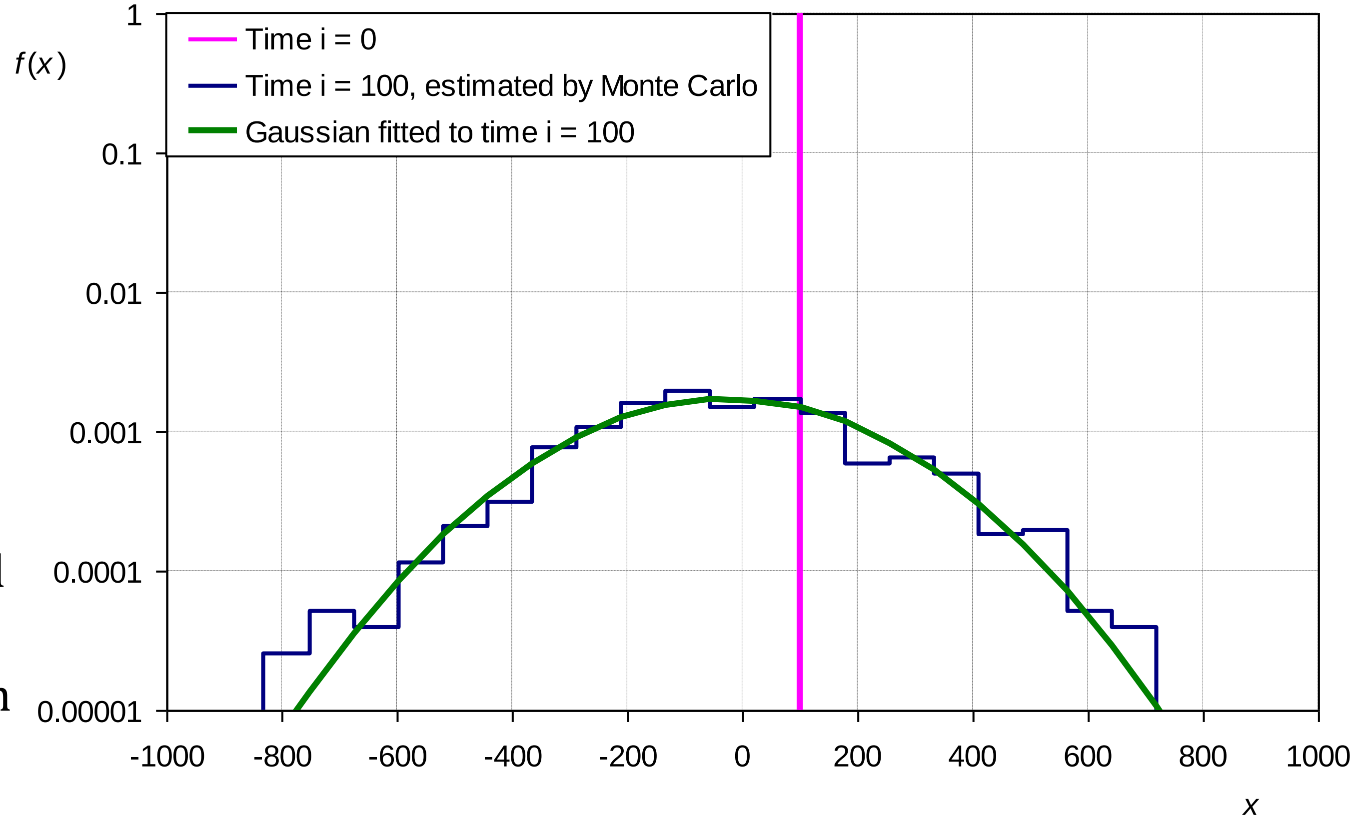
Difficulties in applying the stochastic framework and their overcoming using stochastic tools

- The stochastic representation has potentially an analytical solution that behaves like a deterministic solution, but refers to the evolution in time of admissible sets and densities, rather than to trajectories of points.
  - From $\boldsymbol{x}_i = \boldsymbol{S}(\boldsymbol{x}_{i-1})$ to $f_i(\boldsymbol{x}) = \dfrac{\partial^2}{\partial x \partial v} \int_{\boldsymbol{S}^{-1}(A)} f_{i-1}(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}$
- In the iterative application of the stochastic description of system evolution we encounter two difficulties:
  - Despite being simple, the dynamics is not invertible and the counterimage $\boldsymbol{S}^{-1}(A)$ needs to be evaluated numerically → **numerical integration.**
  - The stochastic formulation is more meaningful for long time horizons → **high dimensional numerical integration**.
- For a number of dimensions $d > 4$, a **stochastic (Monte Carlo) integration** method (evaluation points taken at random) is more accurate than classical numerical integration, based on a grid representation of the integration space (e.g., **Metropolis and Ulam**, 1949; **Niederreiter**, 1992).
- In our case the Monte Carlo method bypasses the calculation of $\boldsymbol{S}^{-1}(A)$.

Monte Carlo integration is very powerful, yet so easy that we may forget that what we are doing is numerical integration.

Results of Monte Carlo integration: Time 100

- We assume $f(\boldsymbol{x}_0)$ to be a uniform density extending 1% around the value $\boldsymbol{x}_0$ = (100 mm, 0.30).
- From 1000 simulations we are able to numerically evaluate $f(\boldsymbol{x}_{100})$.
- The figure shows the density of the soil water, x.
- Moving from time i = 0 to i = 100, the density changes:
  - from concentrated to broad;
  - from uniform to Gaussian.



This analysis provides an approximate numerical solution; an exact asymptotic solution is derived by determining the stationary density, which satisfies FP$f(\boldsymbol{x}) = f(\boldsymbol{x})$.

Why is the distribution of soil water, after a long time, Gaussian?

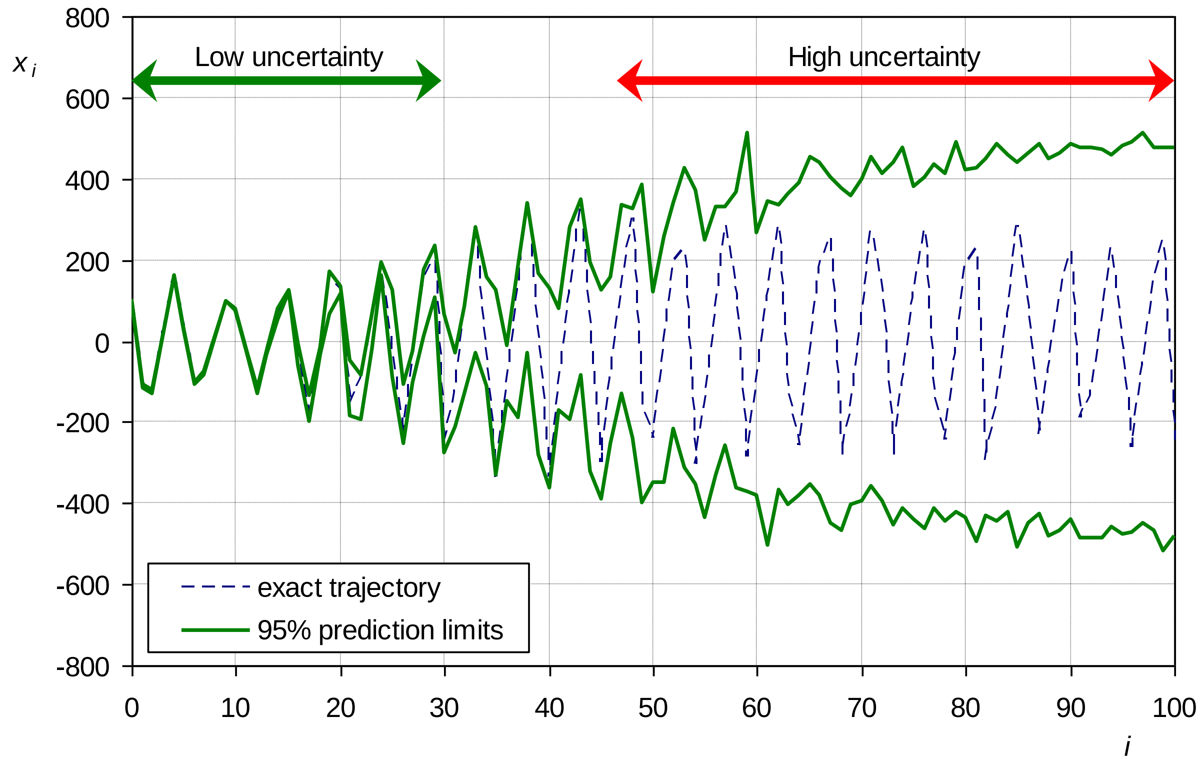
- There are a number of theoretical reasons resulting in **Gaussian** distribution; see **Jaynes** (2003).
- Among them the most widely known is the Central Limit Theorem: multiple integrals of a number of density functions tend to the Gaussian density (this is more typical in sums of variables, not appearing here).
- Another explanation is given by the **Principle of Maximum Entropy**: for fixed mean and variance the distribution that maximizes entropy is the normal distribution (or the truncated normal, if the domain of the variable is an interval in the real line).
- Entropy [< Greek εντροπία < entrepesthai (εντρέπεσθαι) = to turn into] is a probabilistic concept, which for a continuous random variable $\underline{x}$ with density f(x) and for background measure $l(x)$ (typically = 1) is defined as

$$\Phi[\underline{x}] := \mathrm{E}\left[-\ln\left(f(\underline{x})/l(\underline{x})\right)\right] = -\int_{-\infty}^{\infty} f(x)\ln\left(f(x)/l(x)\right)\mathrm{d}x$$

- Entropy is a typical measure of uncertainty, so its maximization indicates that the uncertainty spontaneously becomes as high as possible (this is the basis of the **Second Law** of thermodynamics).

Propagation of uncertainty in time

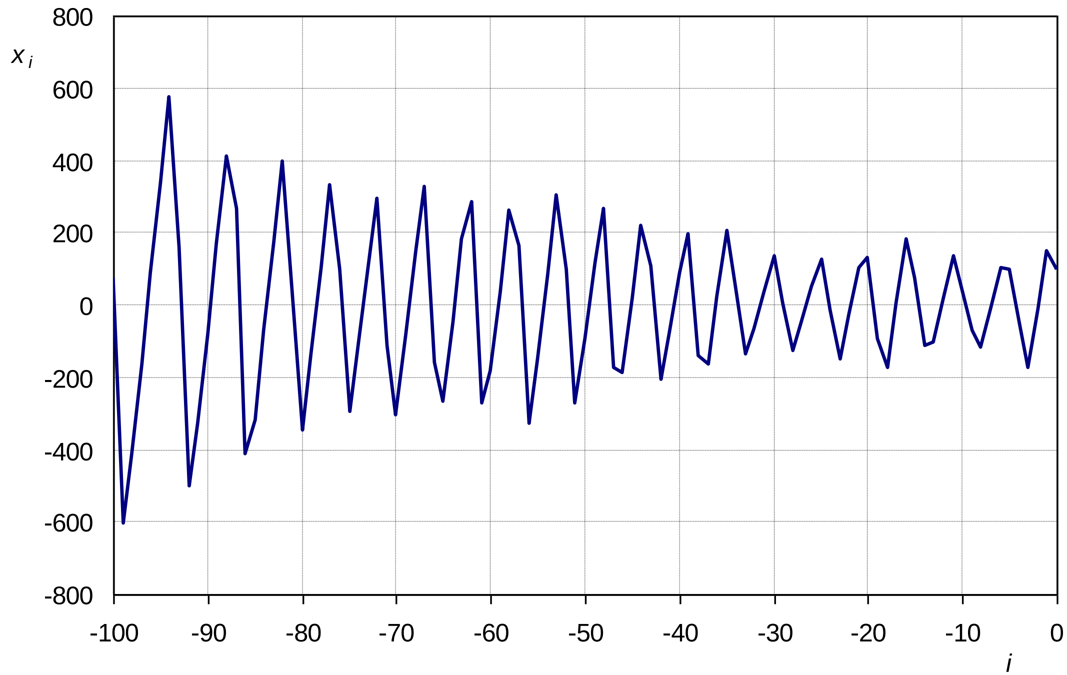
- The propagation of uncertainty is completely determined using stochastics.
- In summary, the stochastic representation:
  - incorporates the deterministic dynamics—yet describes uncertainty;
  - has a rigorous analytical expression (Frobenius-Perron);
  - is free of the defects of deterministic methods;
  - provides and utilizes a powerful numerical integration method (Monte Carlo);
  - is honest as it does not fool us with false certainties.



The so-called ensemble forecasting in weather and flood prediction does not differ from this stochastic framework.

Do we really need the deterministic dynamics to make a long-term prediction?

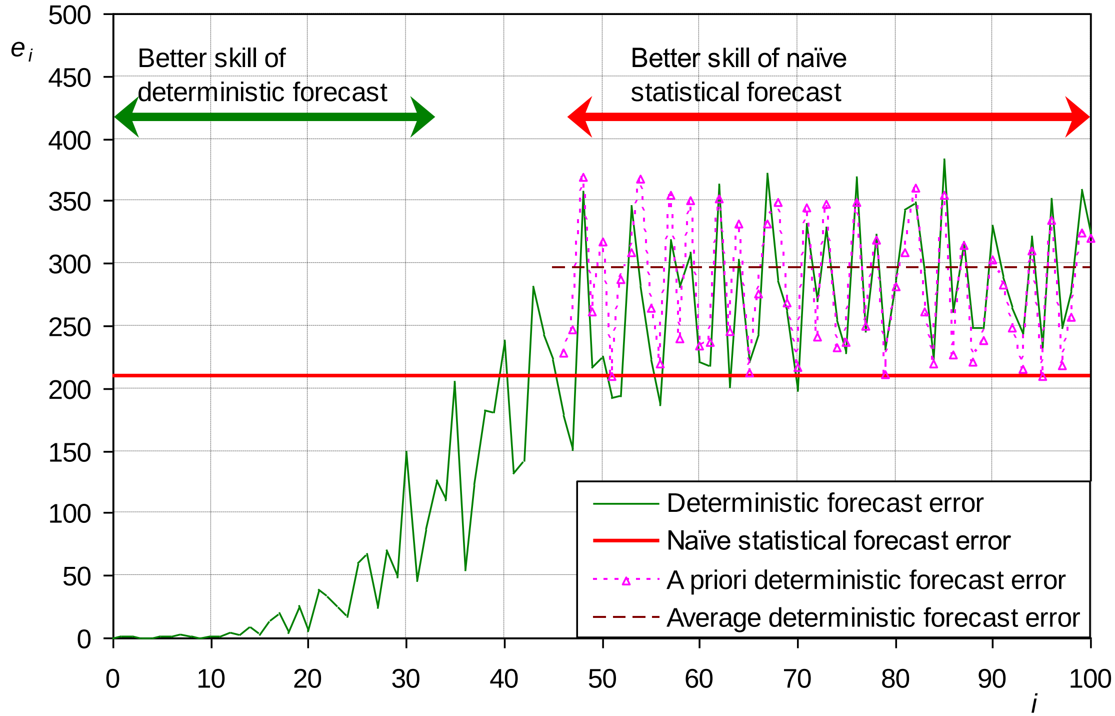
- Working hypothesis: A set of observations contains enough information, which for long horizons renders knowledge of dynamics unnecessary.
- Here we use 100 “years” of “past observations”, for times i = –100 to –1.
- Initial conditions: x_{-100} = 73.99 mm and v_{-100} = 0.904.
- At time i = 0, the resulting state is x_0 = 99.5034 ≈ 100 mm; v_0 = 0.3019 ≈ 0.30.
- Interpreting “observations” as a statistical sample, we estimate (in mm): mean = –2.52; standard deviation = 209.13.



In further investigations, we will refer to the state x_0 = 99.5034; v_0 = 0.3019 as the *exact* (or *true*) *initial state* and x_0 = 100; v_0 = 0.30 as the *rounded off initial state*.

A naïve statistical prediction vs. deterministic prediction

- We compare two different predictions:
  - That derived by immediate application of the system dynamics;
  - A naïve prediction: the future equals the **average** of past data.
- For long prediction times the naïve prediction is more skilful.
- Its error e_i is smaller than that of deterministic prediction by a factor of $\sqrt{2}$.
- This result is obtained both by Monte Carlo simulation and by probability-theoretic reasoning (assuming independence among different trajectories).



For long horizons use of deterministic dynamics gives misleading results. Unless a stochastic framework is used, neglecting deterministic dynamics is preferable.

Past data and ergodicity

- **Ergodicity** (< Greek εργοδικός < [έργον = work] + [οδός = path]) is an important concept in dynamical systems and stochastics.
- By definition (e.g. **Lasota and Mackey**, 1994, p. 59), a transformation is ergodic if all its invariant sets are trivial (have zero probability [= measure]).
- In other words, in an ergodic transformation starting from any point, a trajectory will visit all other points, without being trapped to a certain subset. (In contrast, in non-ergodic transformations there are invariant subsets, such that a trajectory starting from within a subset will never depart from it).
- An important theorem by **George David Birkhoff** (1931) (also known as Birkhoff–Khinchin theorem) says that for an ergodic transformation S and for any integrable function g the following property holds true:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} g\left(S^i(x)\right) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

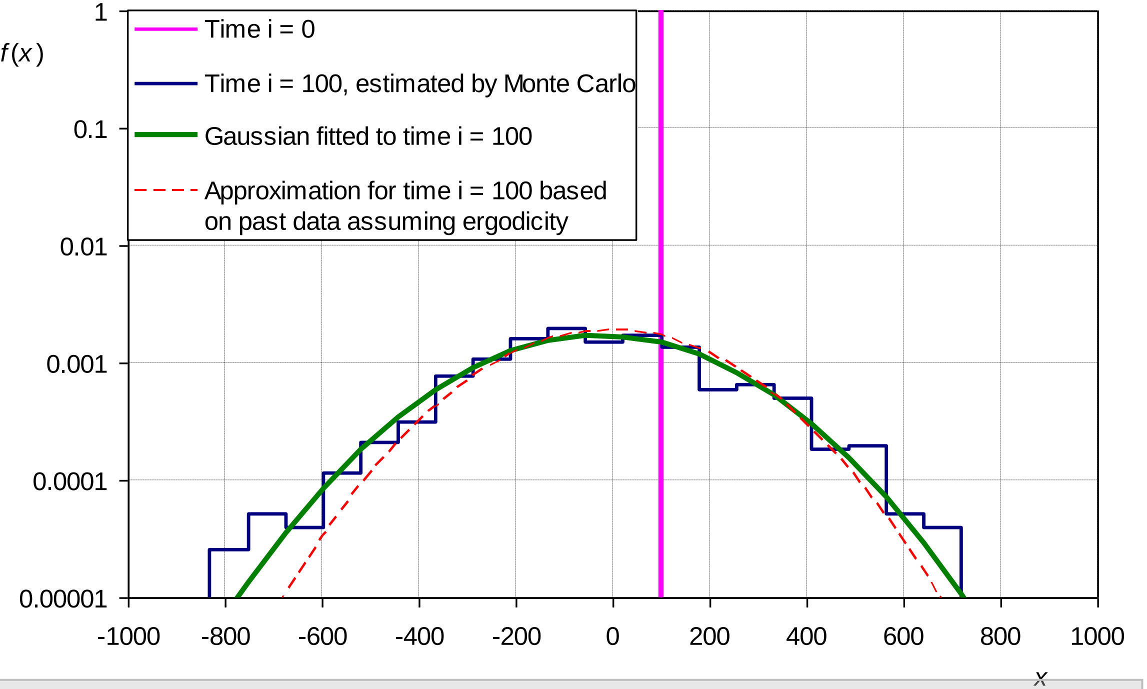
- For instance, for $g(x) = x$, setting x_0 the initial system state, observing that the sequence x_0, $x_1 = S(x_0)$, $x_2 = S^2(x_0)$, ..., n represents a trajectory of the system and taking the equality in the limit as in approximation with finite terms, we obtain that the time average equals the true (ensemble) average:

$$\frac{1}{n} \sum_{i=0}^{n-1} x_i \approx \int_{-\infty}^{\infty} x\, f(x) \mathrm{d}x$$

Ergodicity allows estimation of the system properties using past data only.

What is an informative prediction?

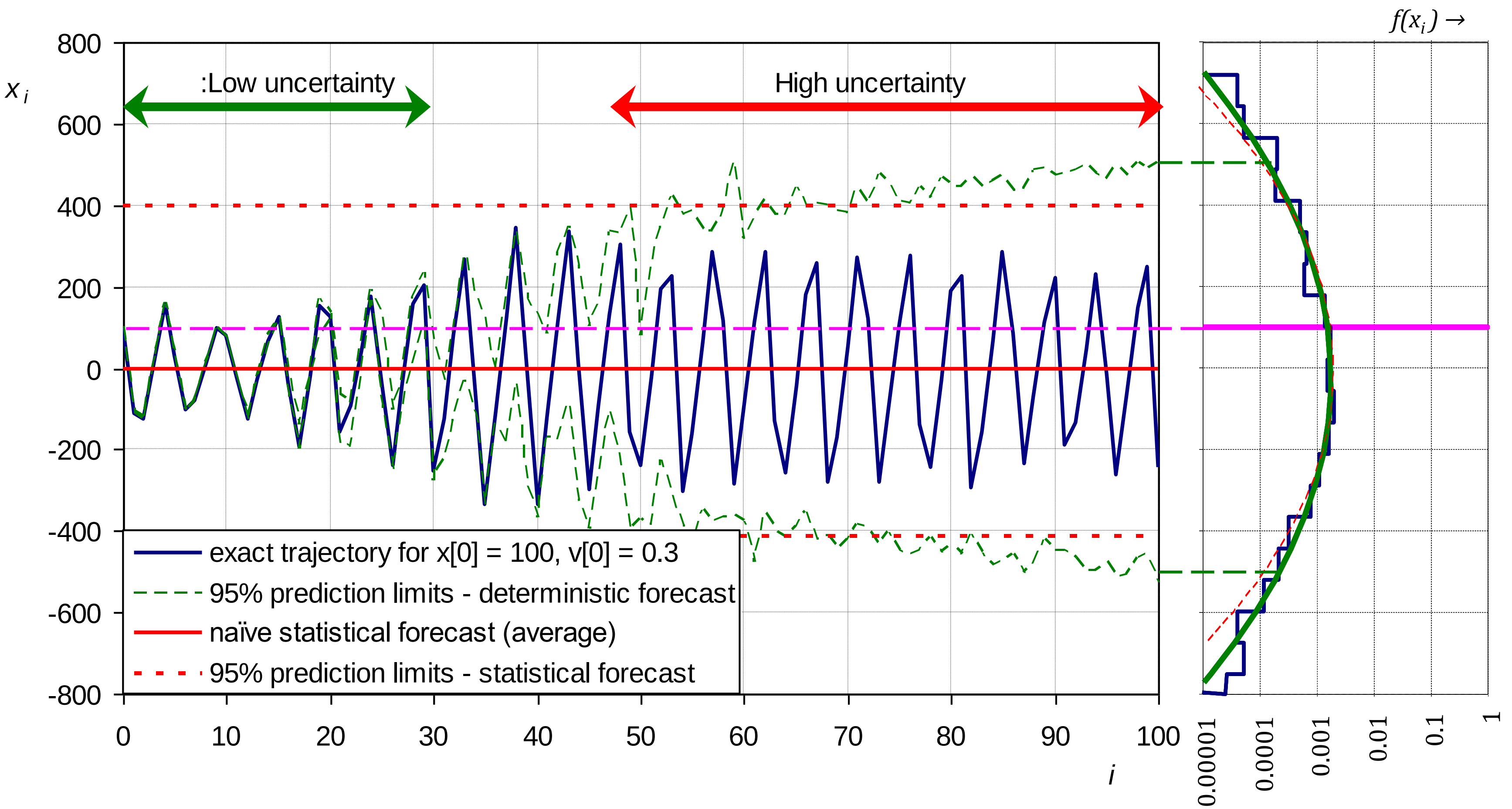
- Reduction of uncertainty for long time horizons: **No way!**
  - No margin for better knowledge of dynamics (full knowledge already).
  - Indifference of improved knowledge of initial conditions (e.g. reduction of initial uncertainty from 1% to 10^{-6} results in no reduction of final uncertainty at i = 100 (try it!).
- Informative prediction = point prediction + quantified uncertainty.
  - Past data: temporal mean & variance at times i = –100 to 0.
  - Ergodicity: ensemble mean & variance at time i = 100.
  - Principle of maximum entropy: Gaussian distribution.



Stochastic inference using (a) past data, (b) ergodicity, and (c) maximum entropy provides an informative prediction.

Knowledge of dynamics does not improve this prediction.

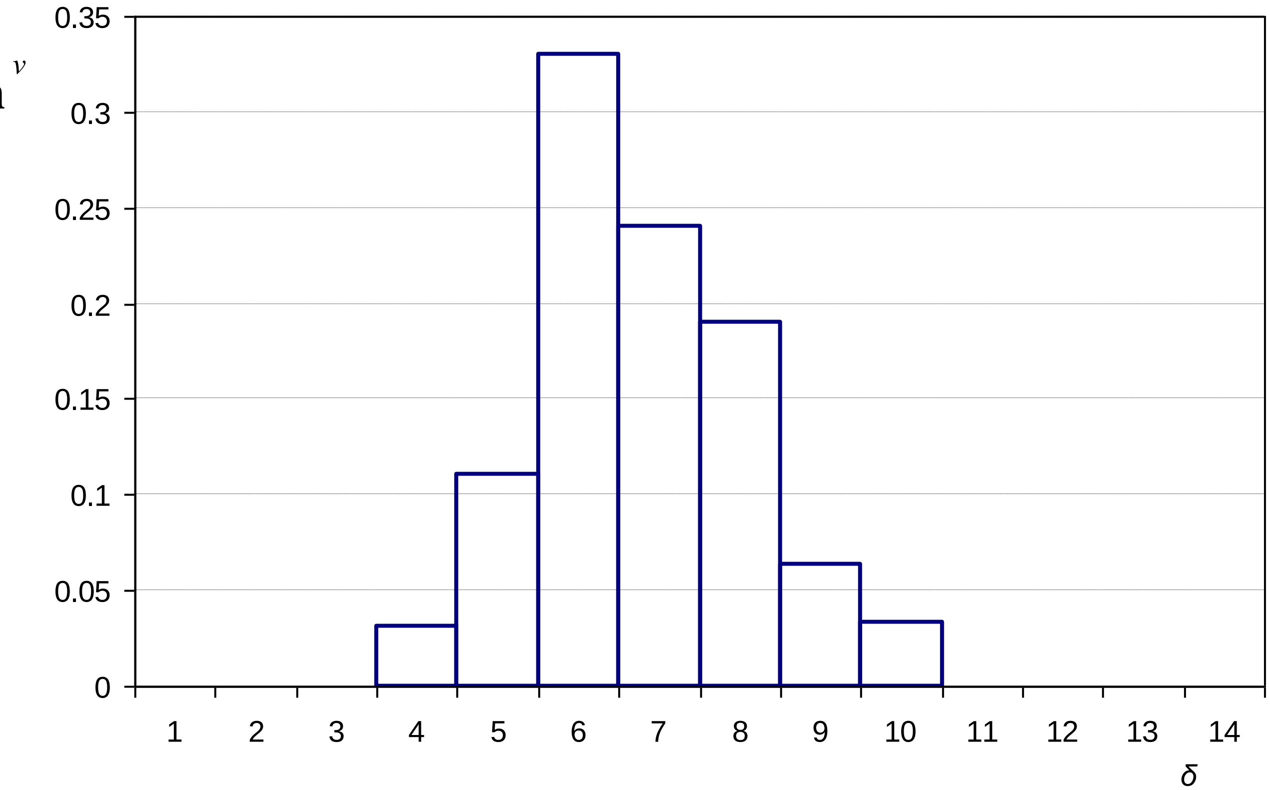
Stochastics for ever...



The stochastic representation is good for both short and long horizons, and helps figure out when the deterministic dynamics should be considered or neglected.

Further exploration of the system properties: Is the system evolution periodic?

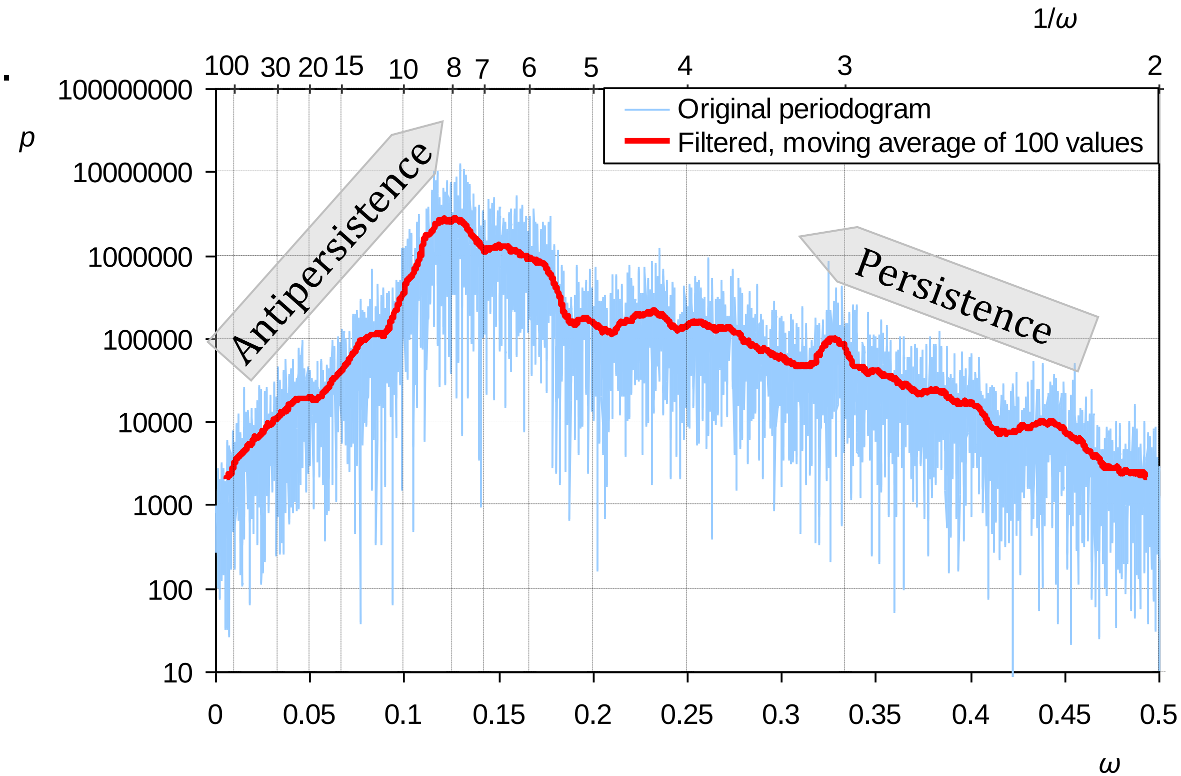
- A longer simulation of the system (10 000 terms) using the rounded-off initial conditions shows that the period δ between consecutive peaks is **not constant** but varies between 4 and 10 "years".
- The period with maximum frequency ν is 6 "years".



The trajectories of the system state do not resemble a typical periodic deterministic system—nor a purely random process.

A stochastic tool to detect periodicity: Periodogram

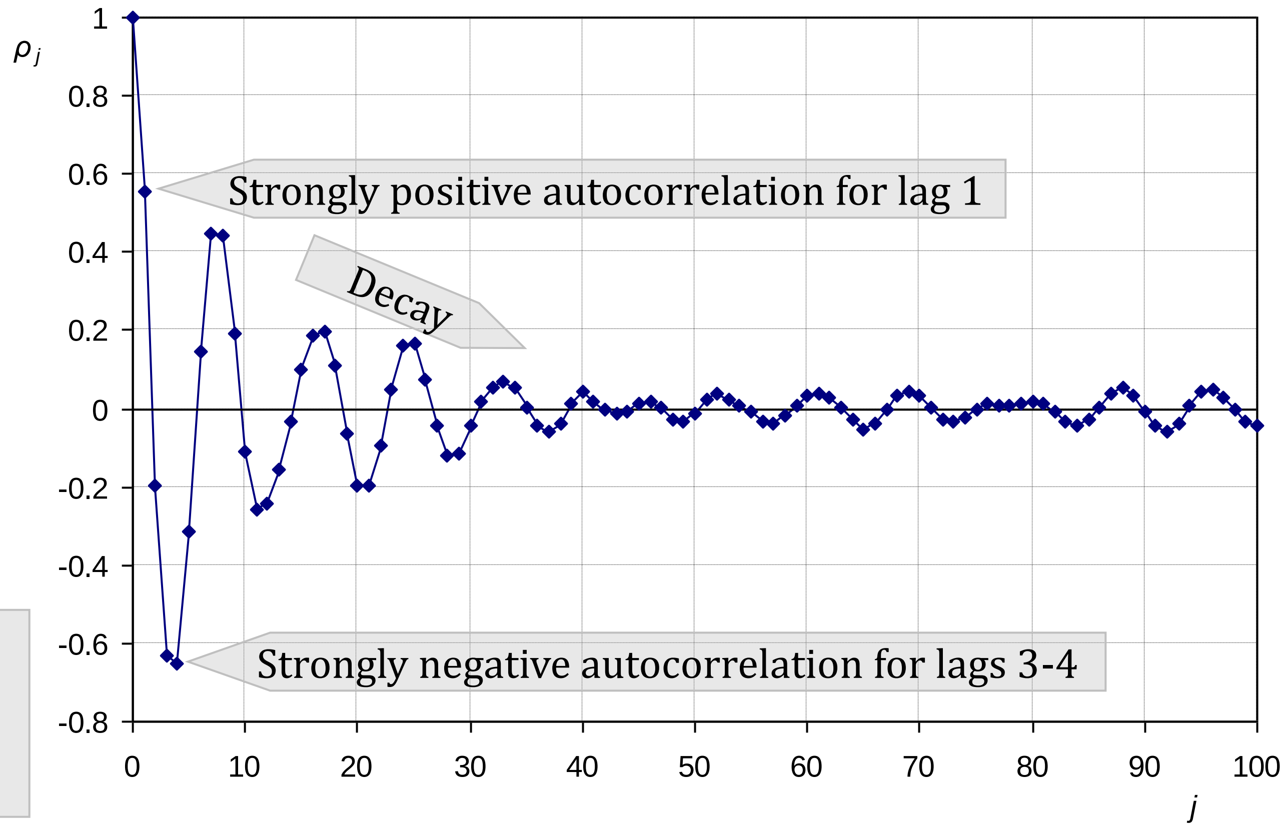
- The square absolute value of the Discrete Fourier Transform (a real function $p(\omega)$ where ω is frequency) of the time series (here 10 000 terms) is the periodogram of the time series.
- $p(\omega)\, d\omega$ is the fraction of variance explained by ω and thus excessive values of $p(\omega)$ indicate strong cycles with period $1/\omega$.
- Here we have large $p(\omega)$ at $1/\omega$ between 4 and 12 "years" without a clearly dominant frequency.
- The shape indicates a combination of **persistence** (short periods) and **antipersistence** (long periods).



Antipersistence is often confused with **periodicity**—however, the two are different.

A stochastic tool to detect periodicity and dependence: Autocorrelation function

- The Finite Fourier Transform of the periodogram is the empirical autocorrelation function (autocorrelogram), which is a sequence of values ρ_j, where j is a lag. It is more easily determined as $\rho_j = \mathrm{Cov}[\underline{x}_i, \underline{x}_{i-j}] / \mathrm{Var}[\underline{x}_i]$.
- The positive ρ_1 is expected because of physical consistency.
- The existence of negative values is an indication of antipersistence.



Models with all auto-correlations negative are not physically consistent.

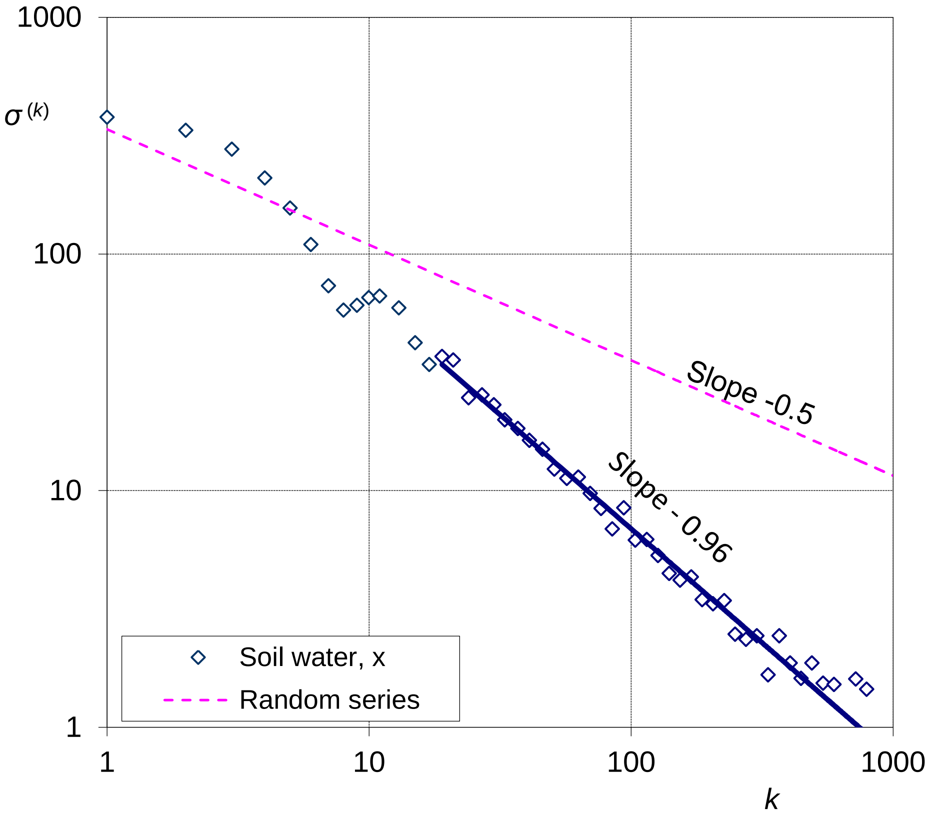
A handy stochastic tool: the climacogram

- A discrete-time random variable x_i refers to a specific time scale.
- A multi-scale stochastic representation defines a process at any scale $k \geq 1$ by:

$$\underline{x}_i^{(k)} := \frac{1}{k} \sum_{l=(i-1)k}^{ik} \underline{x}_l$$

- A key multi-scale characteristic is the standard deviation $\sigma^{(k)}$ of $\underline{x}_i^{(k)}$.
- $\sigma^{(k)}$ is a function of the scale k, termed the climacogram (< Greek Κλιμακόγραμμα < [climax (κλίμαξ) = scale] + [gramma (γράμμα) = written]) and typically depicted on a double logarithmic plot.
- The climacogram is a transformation of the autocorrelation function:

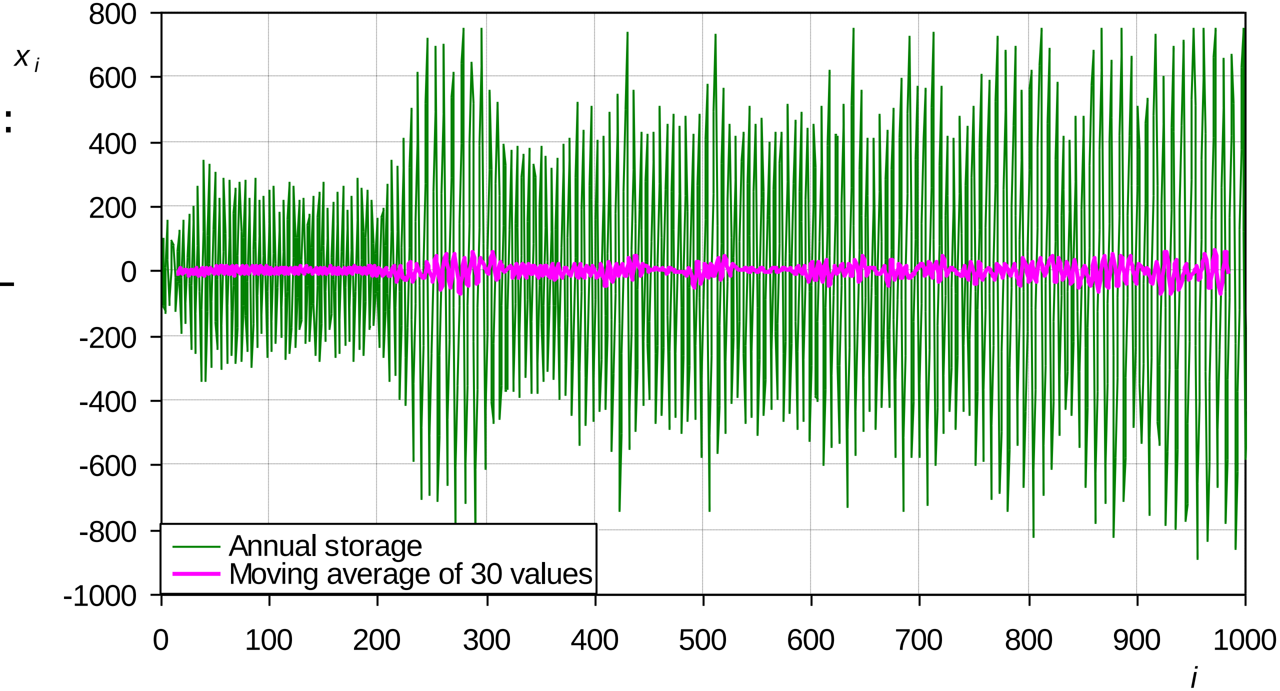
$$\sigma^{(k)} = \frac{\sigma}{\sqrt{k}} \sqrt{1 + 2\sum_{j=1}^{k-1} \left(1 - \frac{j}{k}\right) \rho_j}$$



The asymptotic slope of the climacogram for large time scales characterizes the long-term properties of the process. Steep slopes indicate antipersistence.

A different perspective of long-term predictability and the key consequence of antipersistence

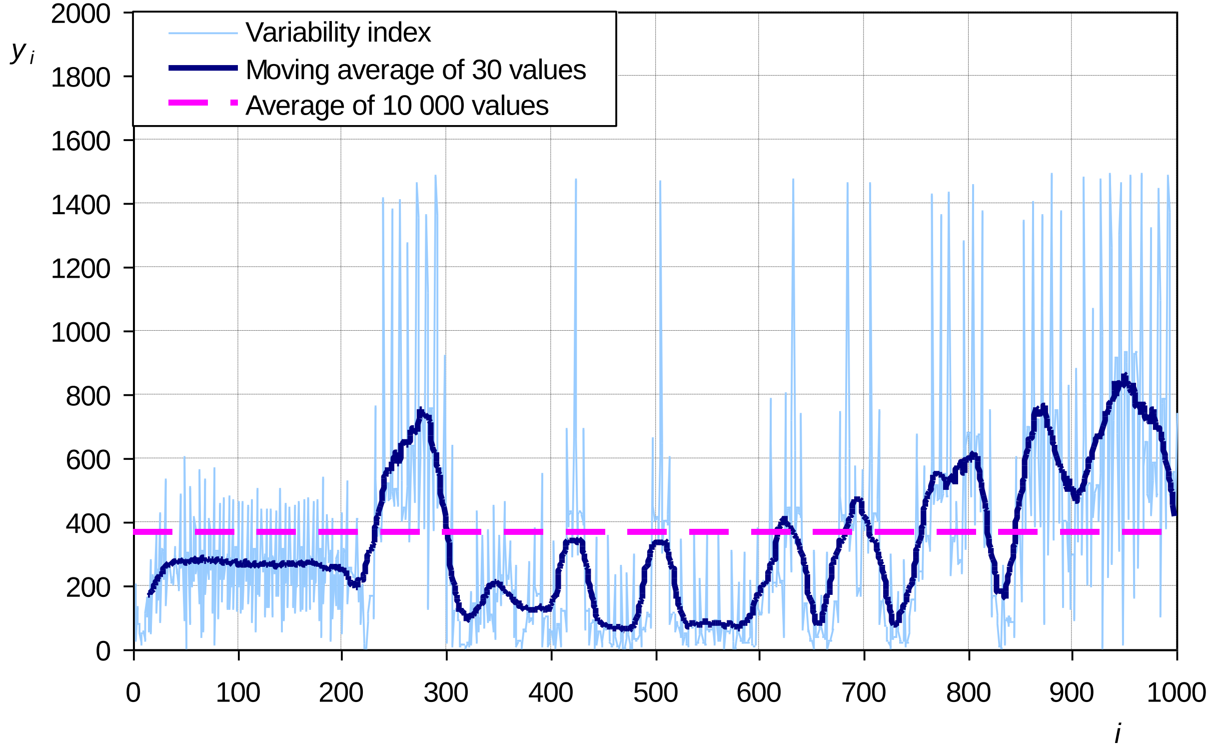
- Arguably, when we are interested for a prediction for a long time horizon, we do not demand to know the exact value at a specified time but an average behaviour around that time (the "**climate**" rather that the "**weather**").
- The plot of the soil water for a long period (1000 "years") indicates:
  - High variability at a short (annual) scale—with peculiar variation patterns;
  - A flat time average at a 30-year scale ("climate").



Antipersistence enhances **climatic-type predictability** (prediction of average).

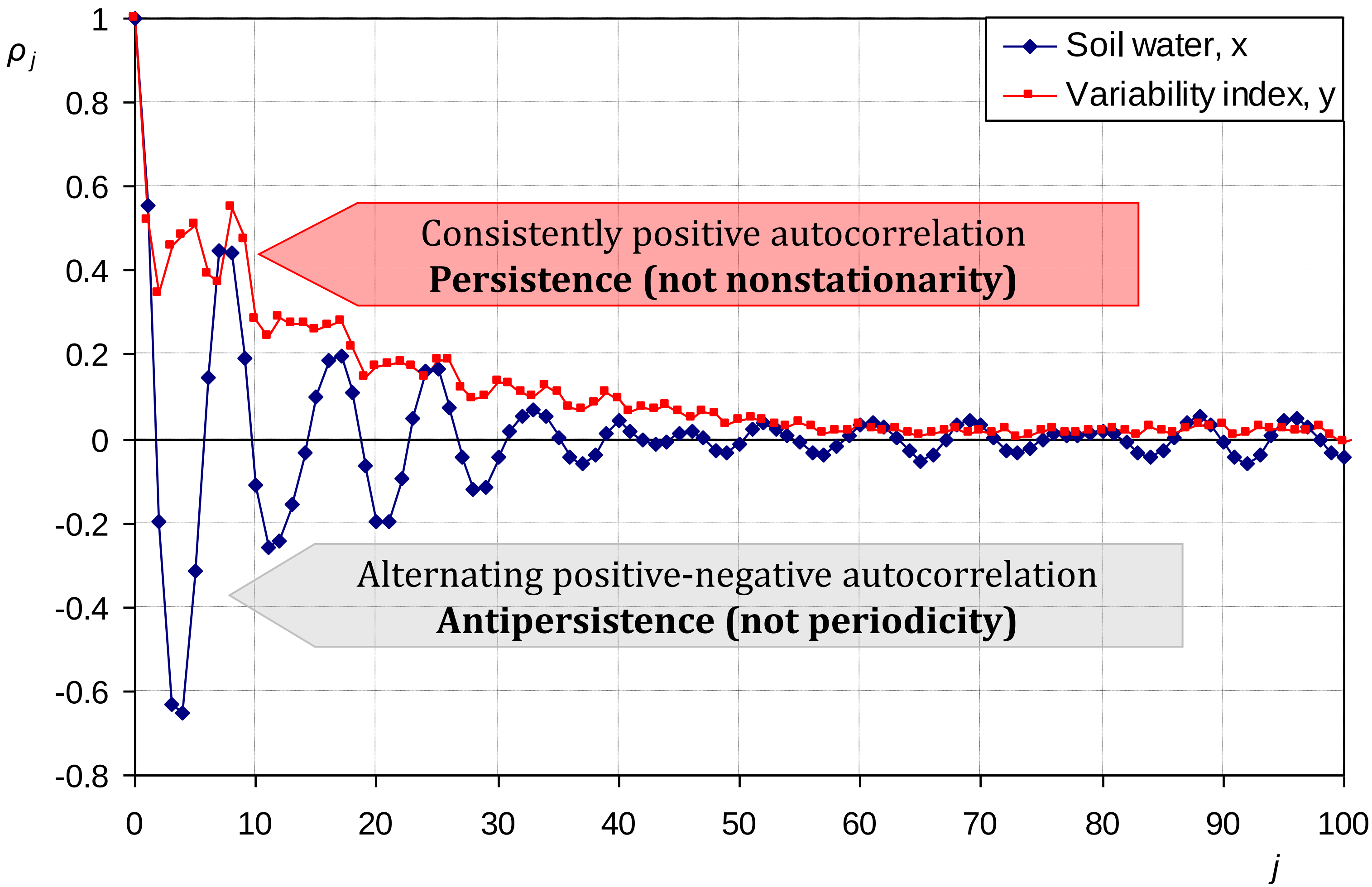
An index of variability

- To study the peculiar variability of the soil water $\underline{x}_i$ we introduce the random variable $\underline{y}_i := |\underline{x}_i - \underline{x}_{i-6}|$ where the lag 6 was chosen to be equal to the most frequent period appearing in the time series of x_i.
- We call $\underline{y}_i$ the **variability index**.
- The plot of the time series of y_i for a long period (1000 "years") indicates:
  - High variability at a short ("annual") scale;
  - Long excursions of the 30"year" average ("the climate") from the global average (of 10000 values).



The frequent and long excursions of the local average from the global average indicate **long-term persistence**, or **long-term change** (not static conditions)
Persistence/change are often confused with **nonstationarity**—but this is an error.

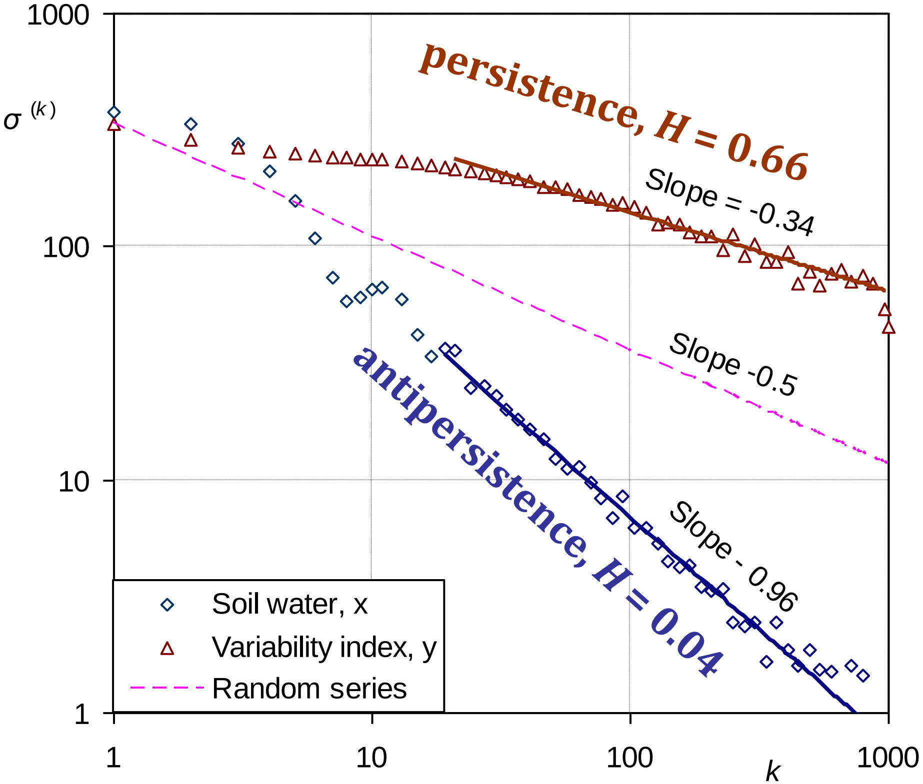
The autocorrelation of the variability index



The consistently positive autocorrelations ρ_j for high lags j indicate **long-term persistence**.

Multi-scale stochastics and the Hurst-Kolmogorov dynamics

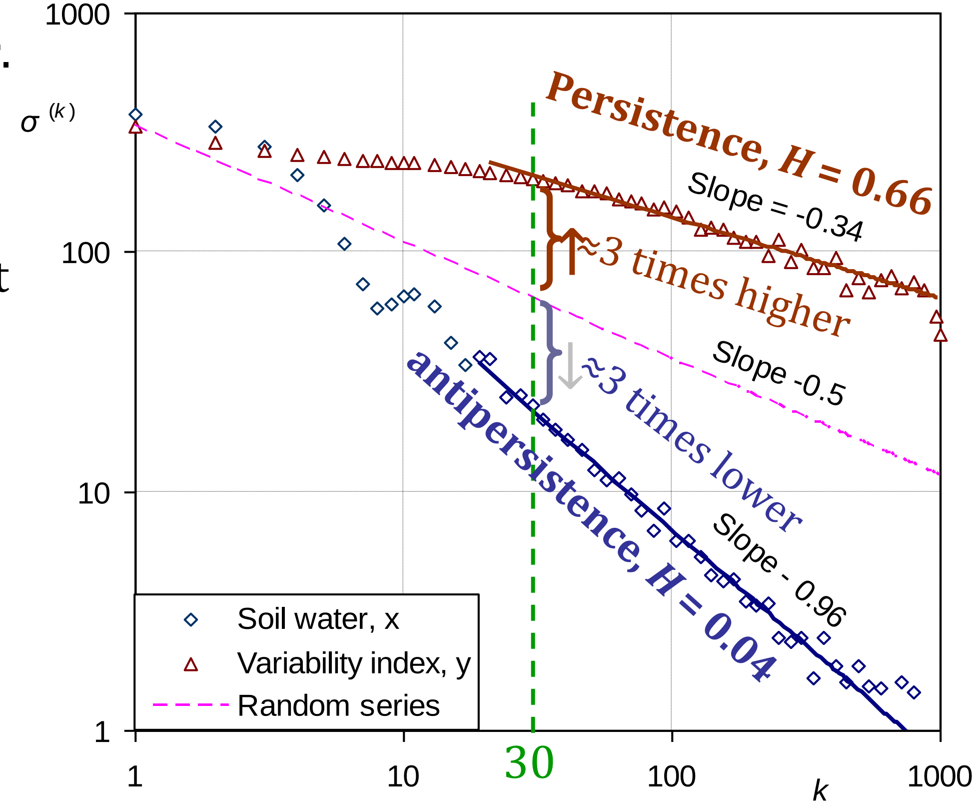
- The long term persistence and antipersistence are better visualized using the climacogram.
- The visualization is based on the slope of the double logarithmic plot of the climacogram ($\sigma^{(k)}$ as a function of the scale k) for large time scales k.
- The quantity H = 1 + slope in this plot is termed the Hurst coefficient.
- H = 0.5 indicates pure randomness.
- H between 0 and 0.5 indicates antipersistence.
- H between 0.5 and 1 indicates persistence.



A process with constant slope and H between 0.5 and 1 is a **Hurst-Kolmogorov** process (after **Hurst**, 1951, and **Kolmogorov**, 1940) with long term persistence.

Multi-scale stochastics and predictability

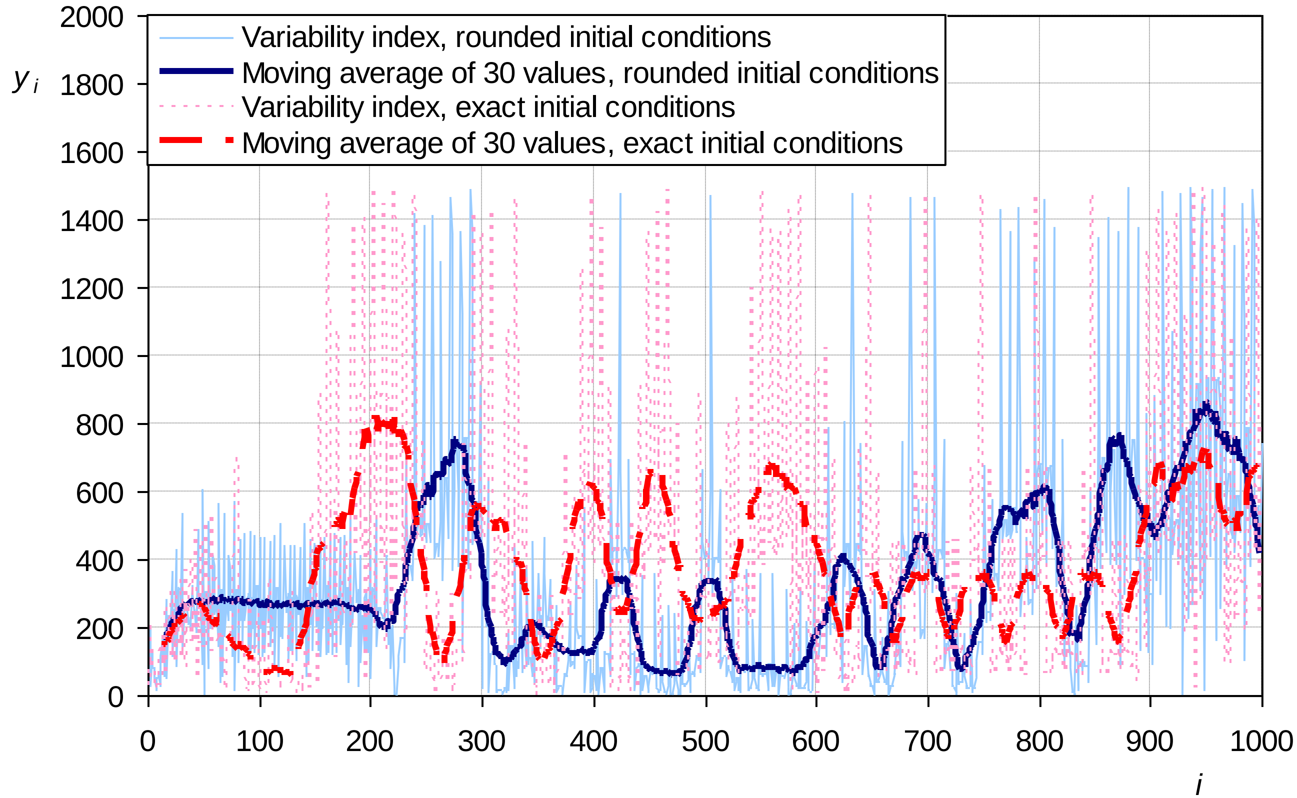
- For an one-step ahead prediction, a purely random process x_i is the most unpredictable.
- Dependence enhances one-step ahead predictability; e.g. in a Markovian process with $\rho_1 = 0.5$ (comparable to that of our series x_i and y_i) the conditional standard deviation is $\sqrt{(1 - \rho_1^2)}$ times the unconditional, i.e. by 13% smaller.
- However, in the climatic-type predictions, for which we are interested about the average behaviour rather than about exact values, the situation is different.
- In the example shown, at the 30-“year” climatic scale, predictability is deteriorated by a factor of 3 for the persistent process y_i (thus annihilating the 13% reduction due to conditioning on the past).



Contrary to what is believed, positive dependence/persistence substantially deteriorates predictability over long time scales—but antipersistent improves it.

Demonstration of unpredictability of processes with persistence

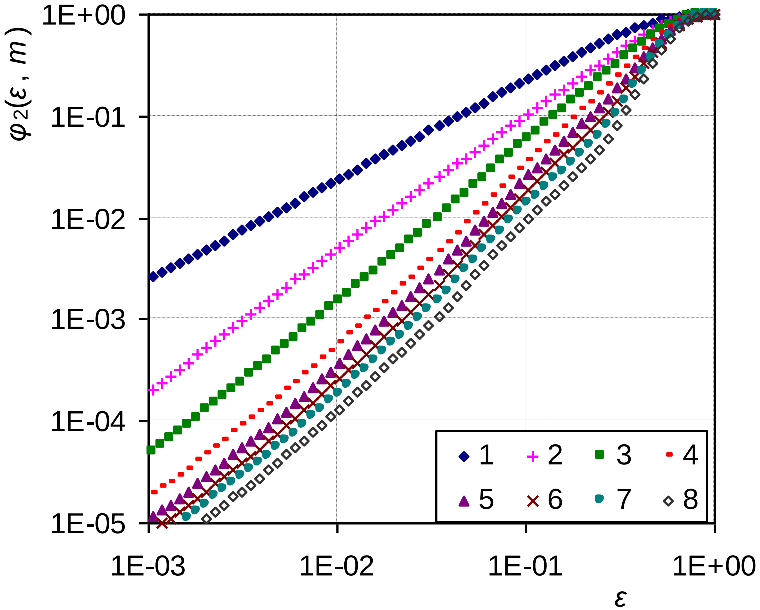
- The plot shows 1000 "years" of the time series y_i (variability index) at the annual and the climatic, 30-"year" scale and for initial conditions
  - exact, and
  - rounded off.
- The departures in the two cases are evident.

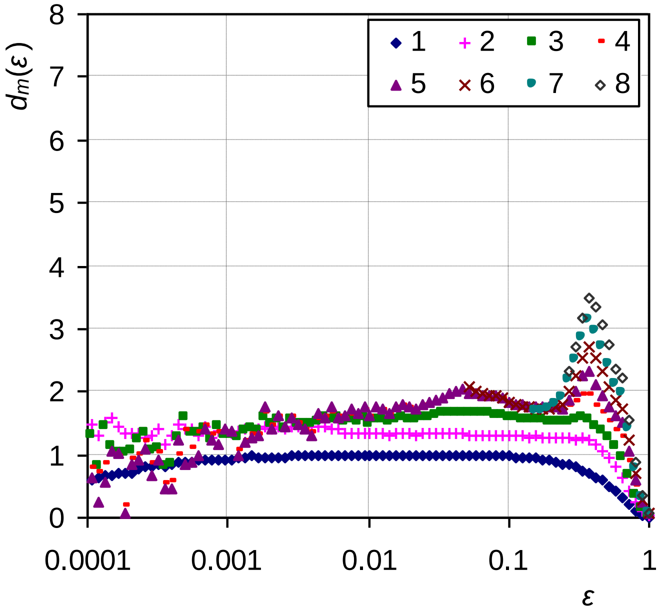


Even a **fully deterministic system** is **fully unpredictable** at a long (climatic) time scale when there is **persistence**.

Recovery of dynamics from time series

- Stochastics—the concept of entropy in particular—provides a way to recover the dynamics of a system, if the dynamics is deterministic and unknown and if a long time series is available.
- Forming time delayed vectors with trial dimensions *m* and calculating the multidimensional entropy of vector trajectories we are able to recover the unknown dynamics (employing **Takens**, 1981, theorem).
- In the example we find that the dimensionality of our toy system is 2.





Recovering of unknown deterministic dynamics does not enhance long-term predictability.

What is randomness? (Alternative reply)

- **Random** means none other than **unpredictable** (in deterministic terms), **uncertain** or **unknown**; we may understand it, we may explain it, but we cannot predict it.
- There is no “virus of randomness” that affects natural systems (including dice).
- Randomness and determinism:
  - coexist in the same process;
  - are not separable or additive components; and
  - it is a matter of specifying the time horizon and time scale of prediction to decide which of the two dominates.
- Dichotomies such as *deterministic* vs. *random* and *aleatory* vs. *epistemic uncertainty* are false dichotomies.
- Almost all physical systems, including the motions of dice and planets, are predictable for short horizons and unpredictable for long horizons.
- The difference of dice from other common physical systems is that they *enable unpredictability* very quickly, at times < 1 s.

“*Prediction is difficult, especially of the future*” (**Niels Bohr**).

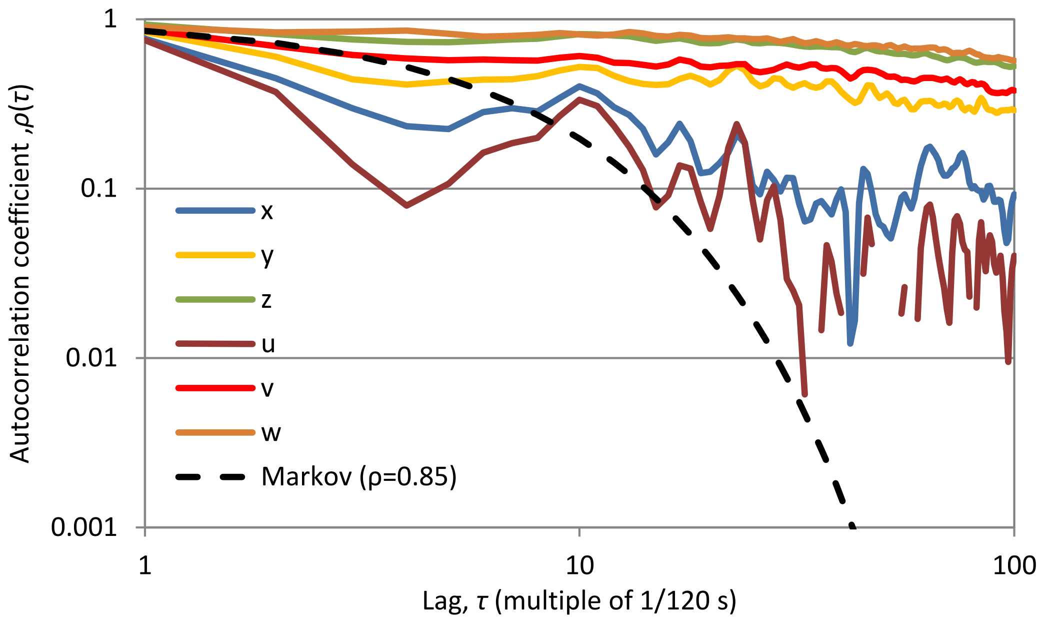
From the toy model to the real world

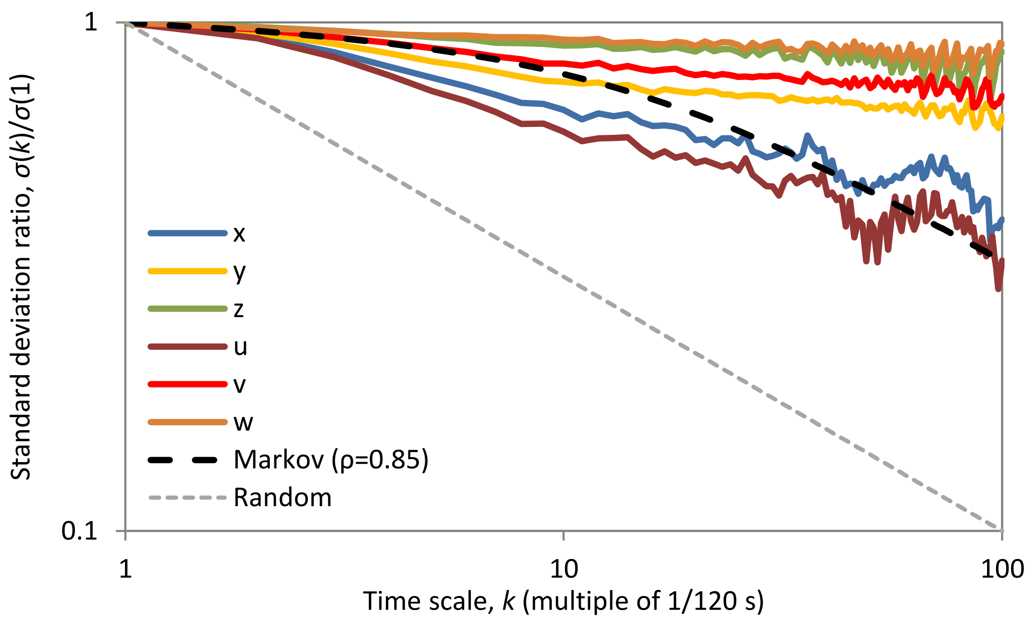
- In comparison to our simple toy model, a natural system (e.g., the atmosphere, a river basin, etc.):
  - is extremely more complex;
  - has time-varying inputs and outputs;
  - has spatial extent, variability and dependence (in addition to temporal);
  - has greater dimensionality (virtually infinite);
  - has dynamics that to a large extent is unknown and difficult or impossible to express deterministically; and
  - has parameters that are unknown.
- Hence uncertainty and unpredictability are naturally even more prominent in a natural system.
- The role of stochastics is even more crucial:
  - to infer dynamics (laws) from past data;
  - to formulate the system equations;
  - to estimate the involved parameters;
  - to test any hypothesis about the dynamics.

Data offer the only solid grounds for all these tasks, and failure in founding on, and testing against, evidence from data renders the hypothesized dynamics worthless.

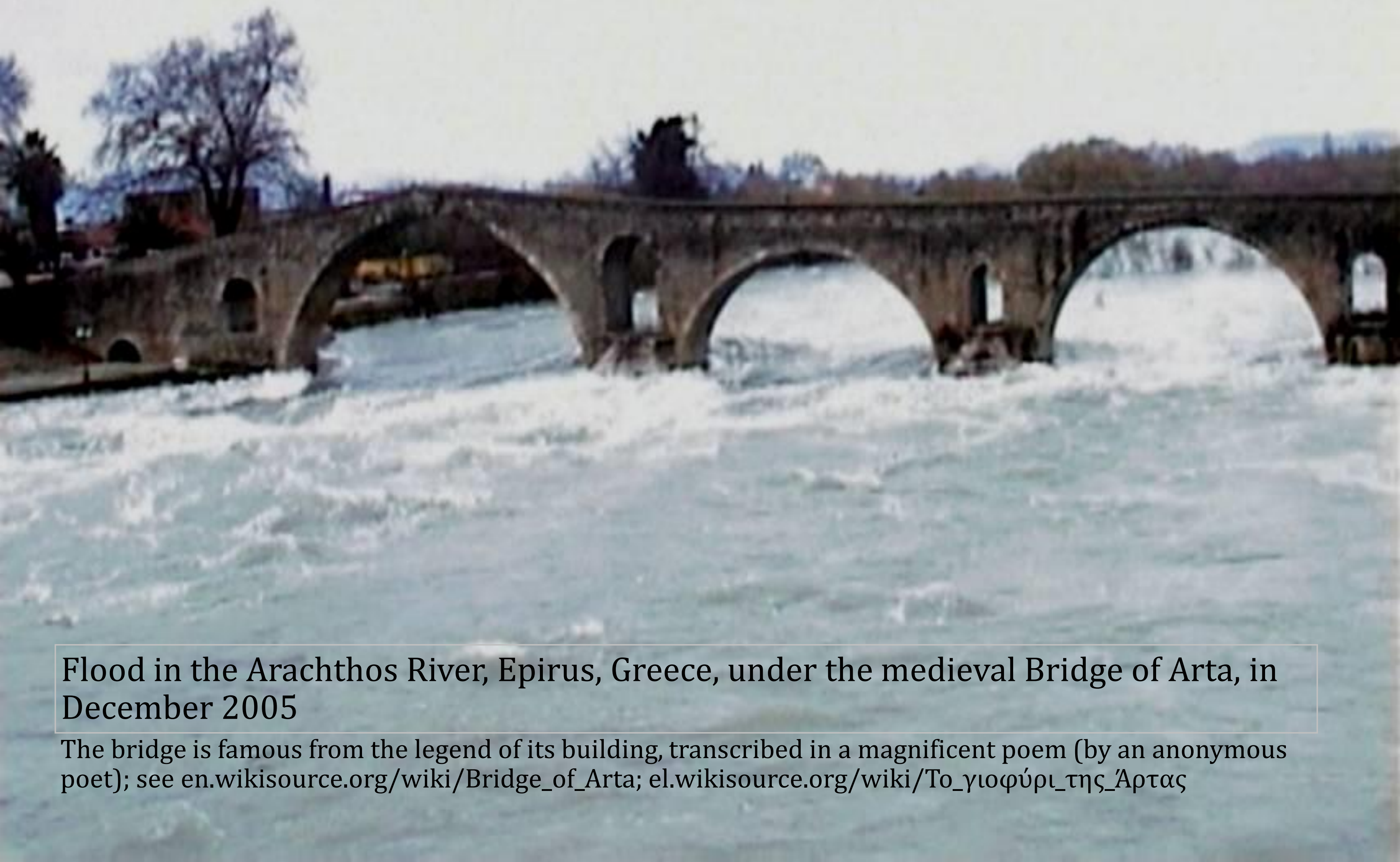
Back to real world applications: Stochastic behaviour of dice

- Autocorrelograms and climacograms (here those for experiment 78 are shown) indicate:
  - Strong dependence in time, which enables stochastic predictability for (very) short time;
  - Long-term, rather than short-term persistence.





Glimpsing God's dice in rivers:
From mixing and turbulence to floods and droughts



Flood in the Arachthos River, Epirus, Greece, under the medieval Bridge of Arta, in December 2005

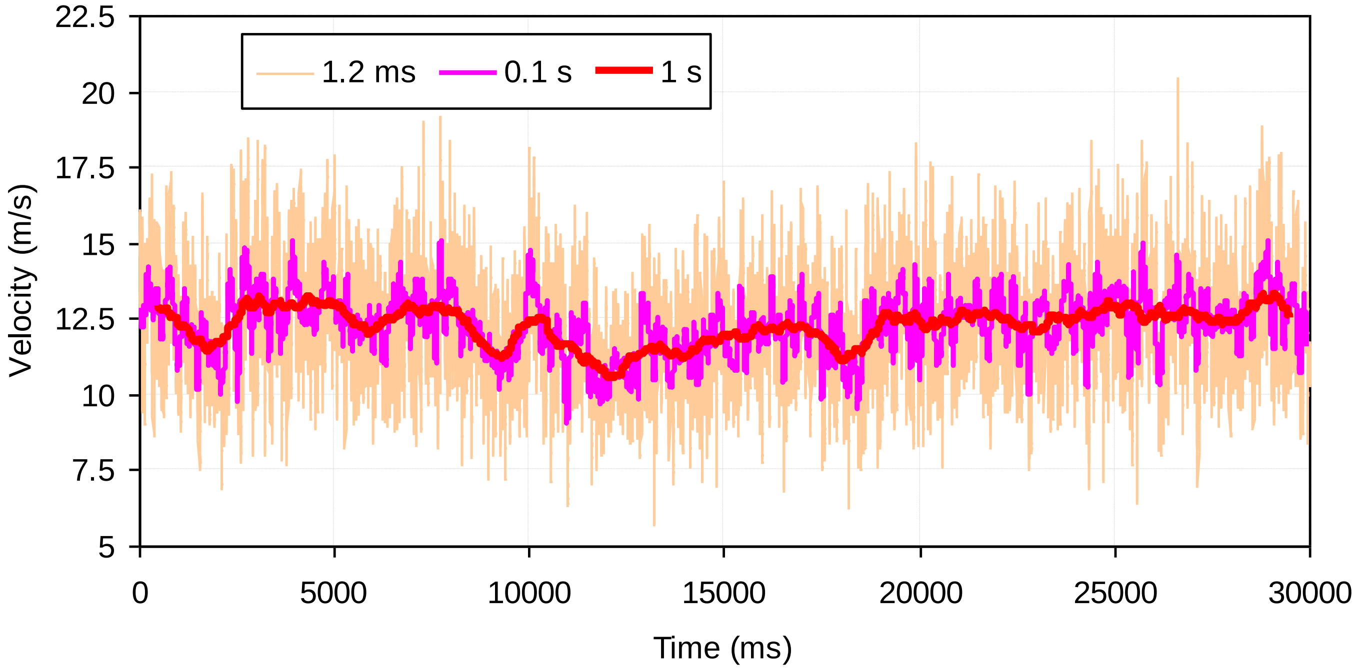
The bridge is famous from the legend of its building, transcribed in a magnificent poem (by an anonymous poet); see en.wikisource.org/wiki/Bridge_of_Arta; el.wikisource.org/wiki/Το_γιοφύρι_της_Άρτας

Turbulence: macroscopic motion at millisecond scale

- Laboratory measurements of nearly isotropic turbulence in Corrsin Wind Tunnel (section length 10 m; cross-section 1.22 m by 0.91 m) at a high-Reynolds-number (Kang *et al.*, 2003).
- Measurements by X-wire probes; Sampling rate of 40 kHz, here aggregated at 0.833 kHz—each point is the average of 48 original values.

When I meet God, I am going to ask Him two questions: Why relativity? And why turbulence? I really believe He will have an answer for the first.

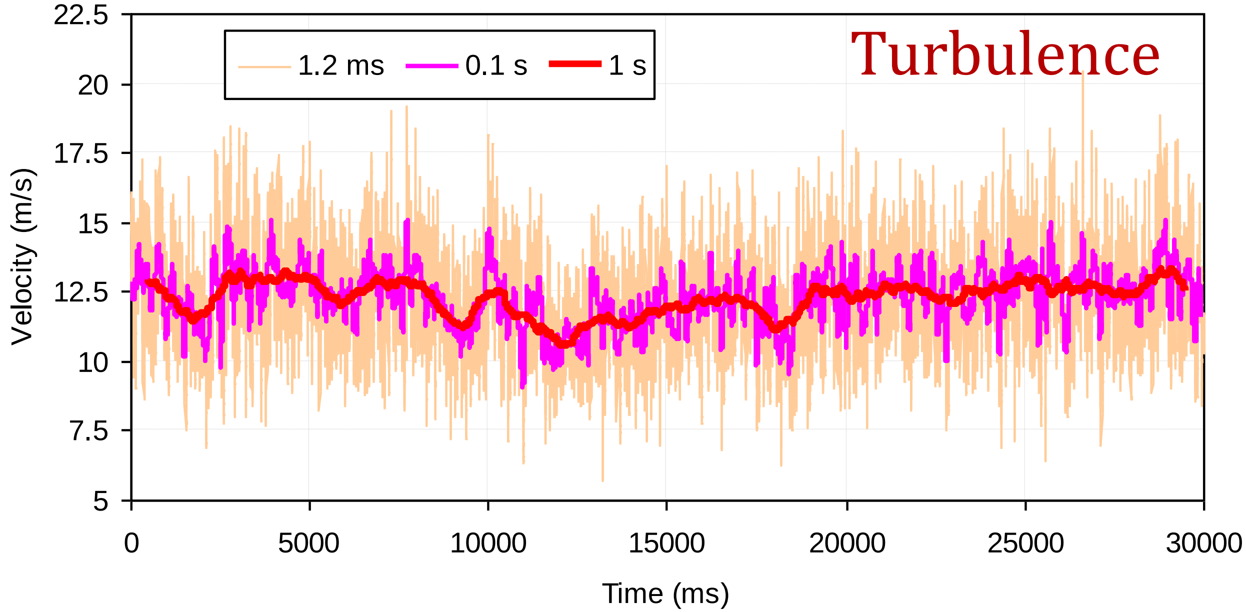
(attributed to **Werner Heisenberg** or, in different versions, to **Albert Einstein** or to **Horace Lamb**)

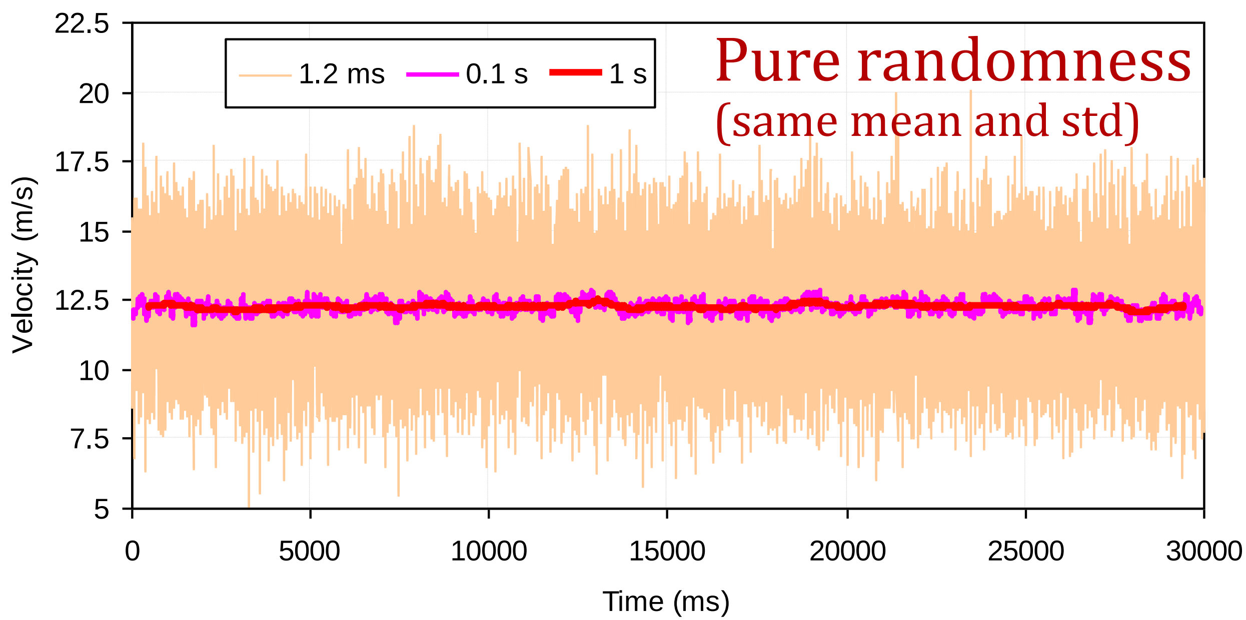


Data downloaded from www.me.jhu.edu/meneveau/datasets/Activegrid/M20/H1/m20h1-01.zip

Turbulence *vs.* pure randomness

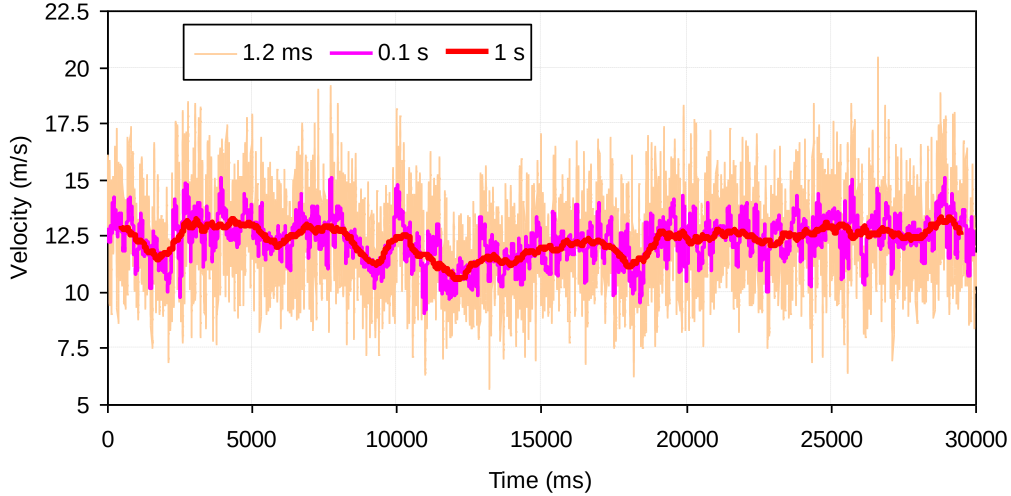
- Pure random processes (white noise), have been effective in modelling microscopic motion (e.g. in statistical thermodynamics).
- Macroscopic random motion is more complex.
- In pure randomness, **change** and **uncertainty** vanish at large scales.
- In turbulence, change occurs at all scales.

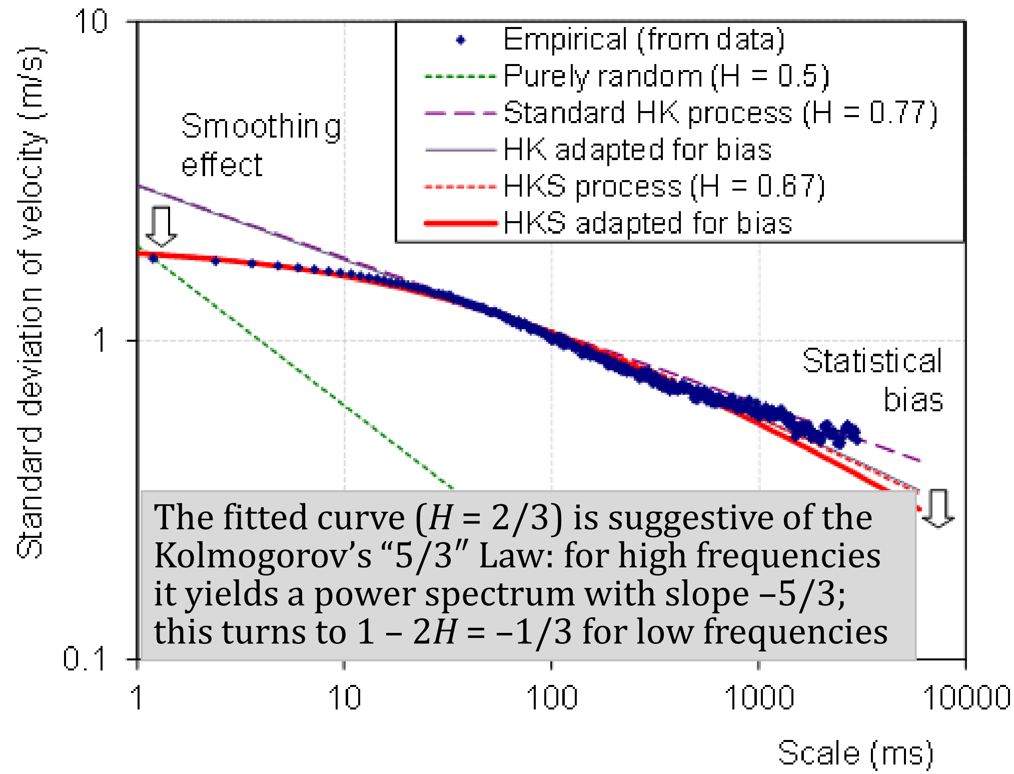




The climacogram of the turbulent velocity time series

- The simple-scaling HK process is appropriate for scales > 50 ms, but not for scales smaller than that.
- For small scales, a smoothing effect reduces variability (in comparison to that of the HK process).
- A Hurst-Kolmogorov process with Smoothing (HKS) is consistent with turbulence measurements at the entire range of scales;, in addition to the Hurst coefficient, it involves a smoothing parameter (α):
$\sigma^{(k)} = \sigma^{(\alpha)}\ (2/\ (1+(k/\alpha)^{2\,-\,2H}))^{1/2}$





A river viewed at different time scales—from seconds to millions of years

- Next second: the hydraulic characteristics (water level, velocity) will change due to turbulence.
- Next day: the river discharge will change (even dramatically, in case of a flood).
- Next year: The river bed will change (erosion-deposition of sediments).
- Next century: The climate and the river basin characteristics (e.g. vegetation, land use) will change.
- Next millennia: All could be very different (e.g. the area could be glacialized).
- Next millions of years: The river may have disappeared.

- None of these changes will be a surprise.
- Rather, it would be a surprise if things remained static.
- These changes are not predictable—**change** and **uncertainty** are tightly connected.
- Most of these changes can be mathematically modelled in a stochastic framework admitting stationarity!

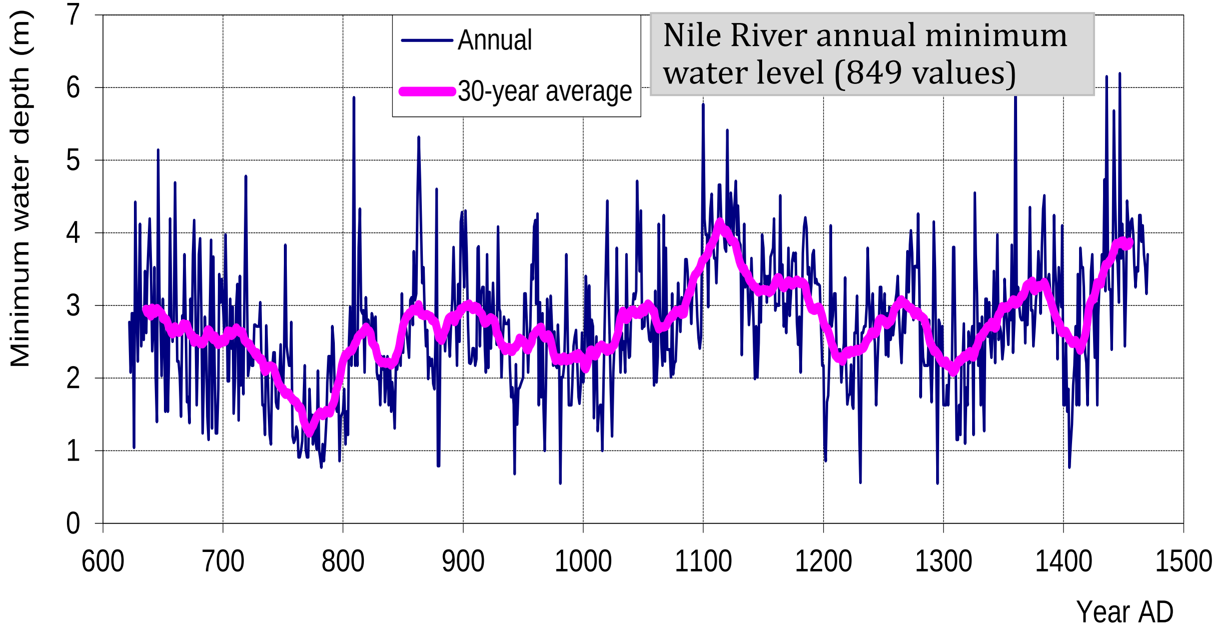
The Roda Nilometer and over-centennial change

(Credit: Aris Georgakakos)

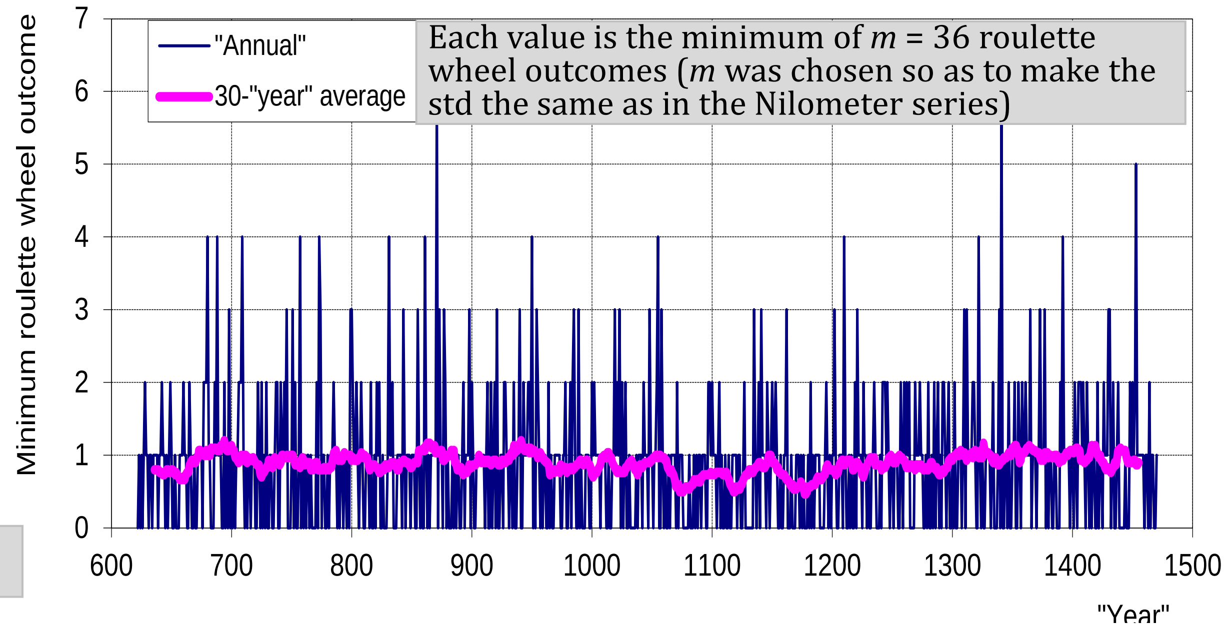
- The Roda Nilometer as it stands today; water entered and filled the Nilometer chamber up to river level through three tunnels.
- In the centre of the chamber stands a marble octagonal column with a Corinthian crown; the column is graded and divided into 19 “cubits” (1 cubit ~ 0.5 m) and could measure floods up to about 9.2 m.
- A maximum level below the 16th mark could portend drought and famine; a level above the 19th mark meant catastrophic flood.

The Nilometer record *vs*. a pure random process

A real-world process



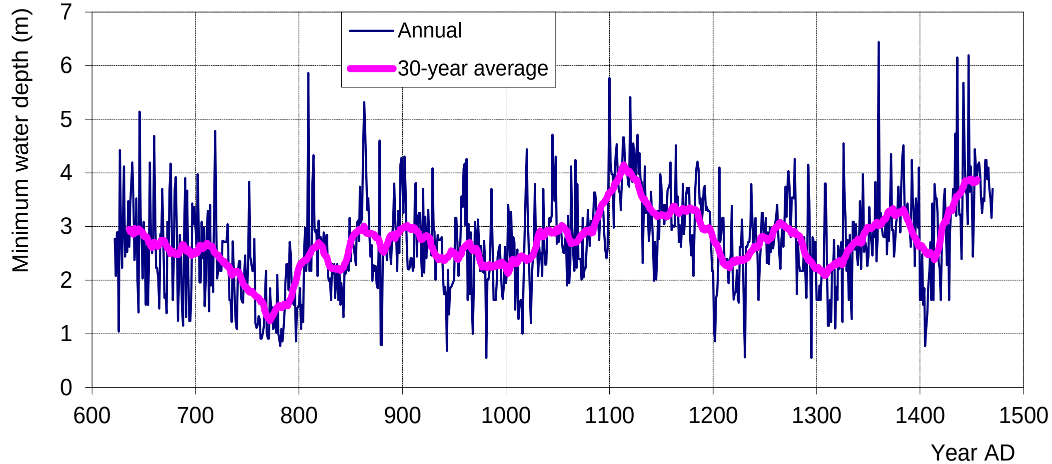
A "roulette" process

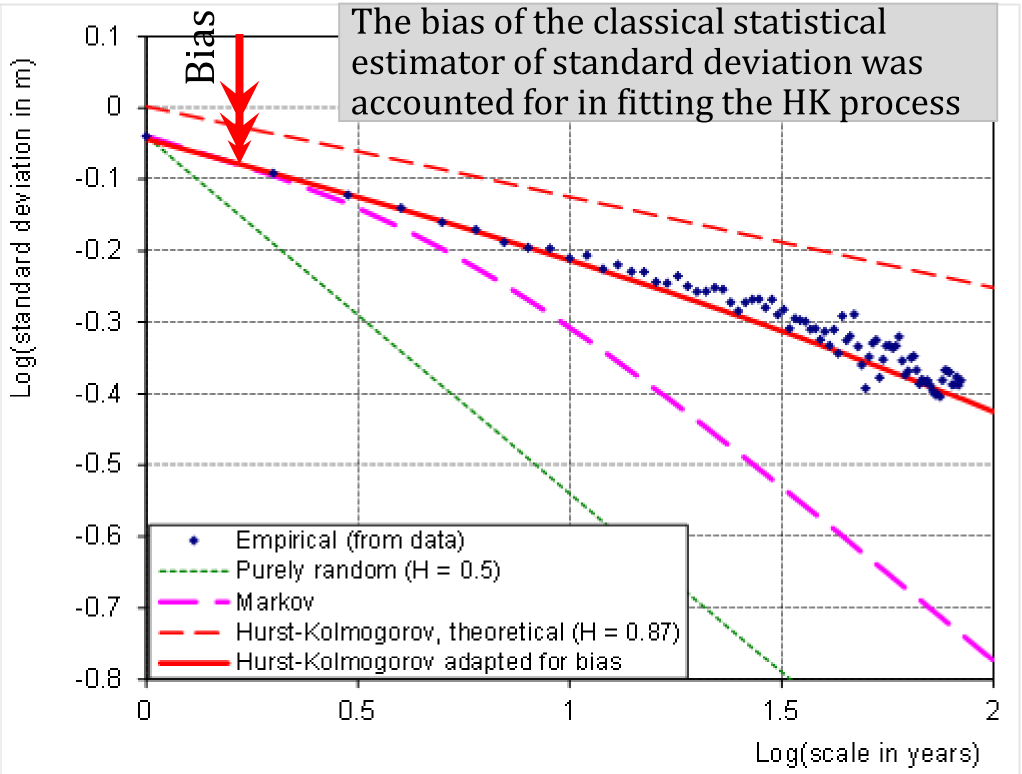


Nilometer data: Toussoun (1925); see also Koutsoyiannis (2013)

The climacogram of the Nilometer time series

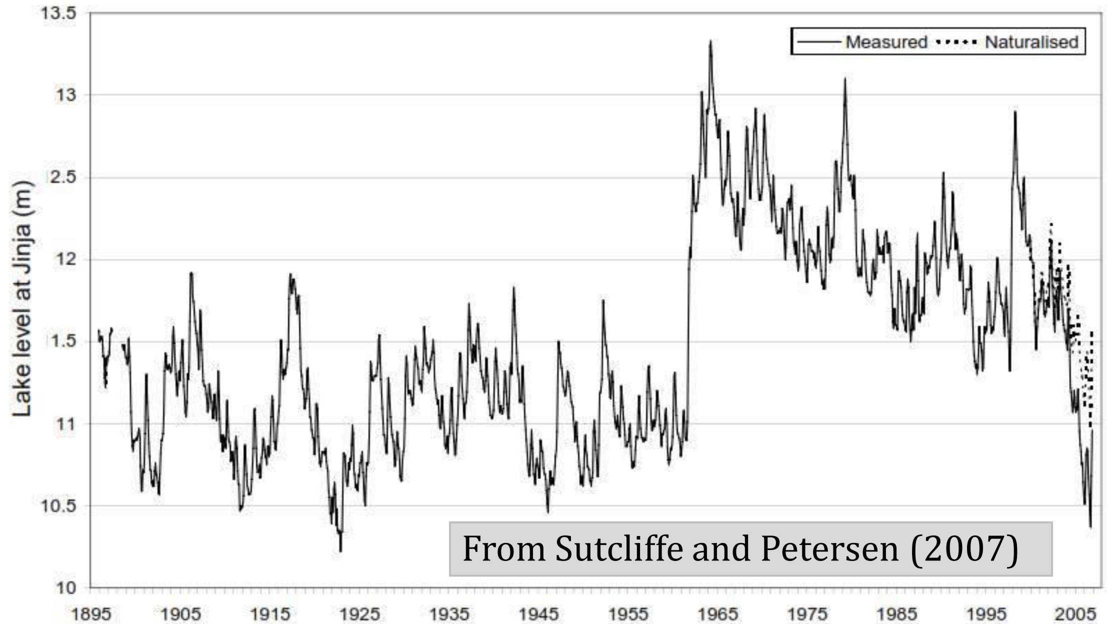
- The Hurst-Kolmogorov process seems consistent with reality.
- The Hurst coefficient is H = 0.87 (Similar H values are estimated from the simultaneous record of maximum water levels and from the modern, 131-year, flow record of the Nile flows at Aswan).
- Essentially, the Hurst-Kolmogorov behaviour manifests that long-term changes are much more frequent and intense than commonly perceived and, simultaneously, that the future states are much more uncertain and unpredictable on long time horizons than implied by pure randomness.



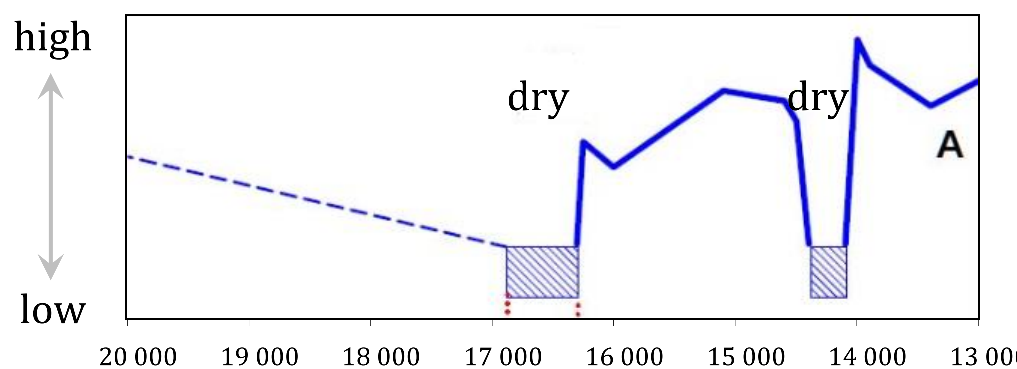


Are reconstructions of past hydroclimatic behaviours consistent with the perception of enhanced change?

- Lake Victoria is the largest tropical lake in the world (68 800 km^2) and is the headwater of the White Nile.
- The contemporary record of water level (covering a period of more than a century) indicates huge changes.
- Reconstructions of water level for past millennia from sediment cores (Stager *et al.*, 2011) suggest that the lake was even dried for several centuries.



From Sutcliffe and Petersen (2007)



Adapted from Stager *et al.* (2011)

Years before present

From hydrology to climate

- Even the very definition of climate relies on stochastics:
  - *Climate in a narrow sense is usually defined as the **average** weather, or more rigorously, as the **statistical description** in terms of the **mean** and **variability** of relevant quantities over a period of time ranging from months to thousands or millions of years. The classical period for **averaging** these variables is **30 years**, as defined by the World Meteorological Organization. The relevant quantities are most often surface variables such as temperature, precipitation and wind. Climate in a wider sense is the state, including a **statistical description**, of the climate system* (IPCC, 2013).
- Most questions related to climate are statistical questions.
- However, statistical perception of climate is typically based on too simple uncertainties, like dice throws and roulette wheels. Also, analyses are based on classical statistics in which variables are independent.
- Even the very definition of climate, particularly the phrase "*The classical period for averaging these variables is 30 years*" historically reflects a perception of a constant climate and a hope that 30 years would be enough for a climatic quantity to get stabilized to a constant value—as would indeed happen if events were independent.
- Classical statistics is inappropriate for climate, which has never been static.

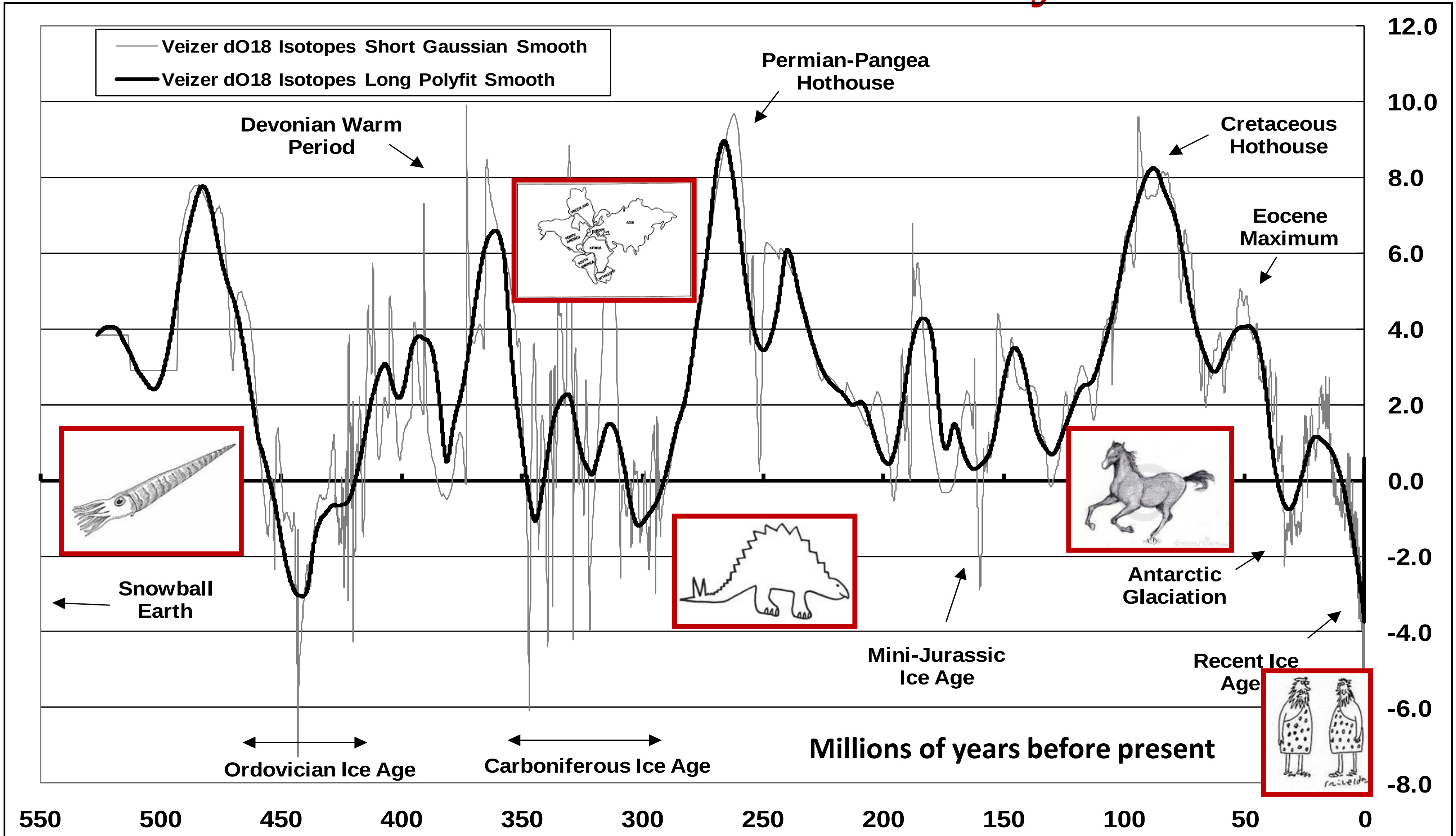
Assessing climate uncertainty from data (instrumental and proxy)

Surv Geophys (2013) 34:181–207
DOI 10.1007/s10712-012-9208-9

Climatic Variability Over Time Scales Spanning Nine Orders of Magnitude: Connecting Milankovitch Cycles with Hurst–Kolmogorov Dynamics

Yannis Markonis · Demetris Koutsoyiannis

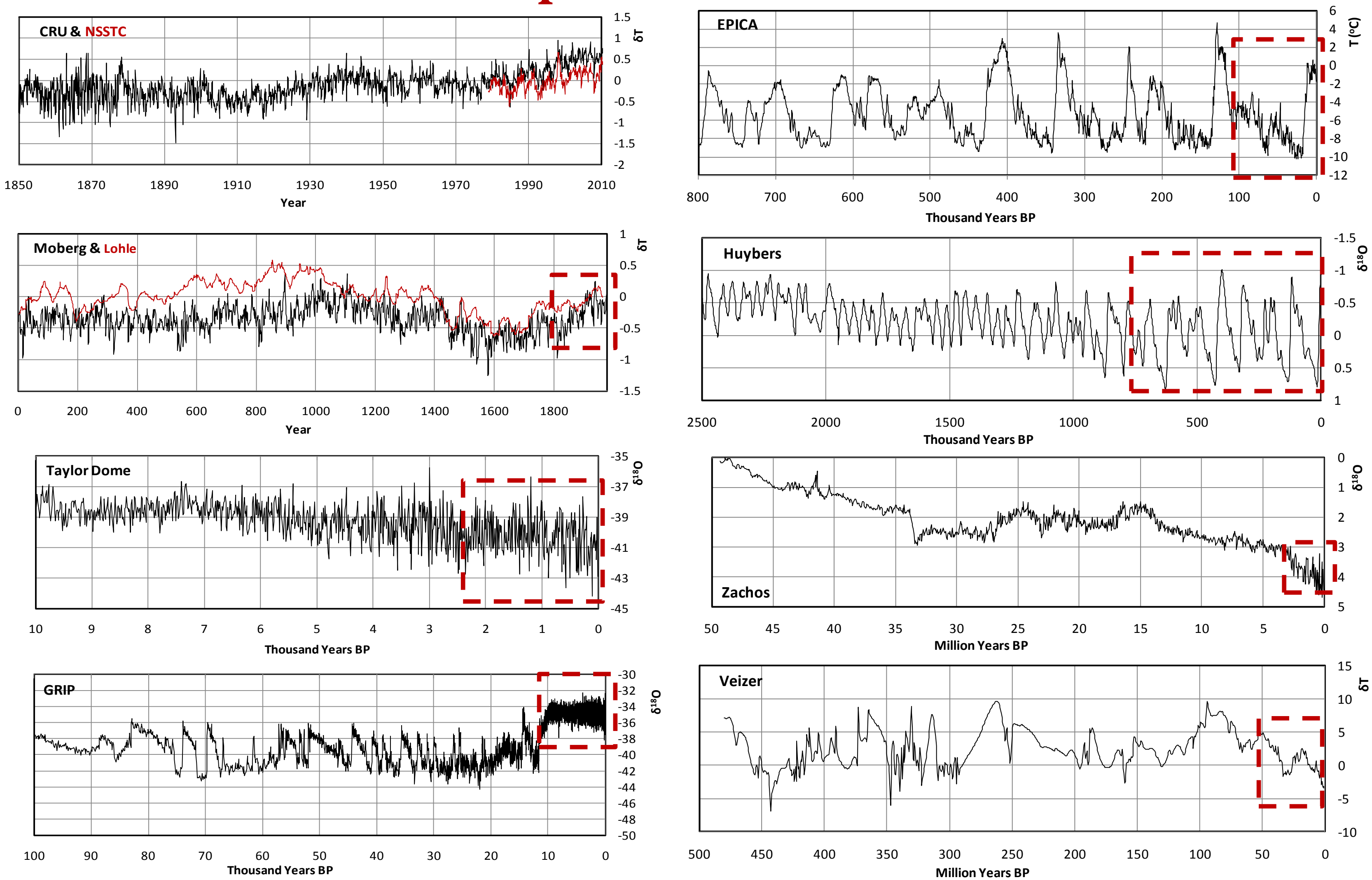
Co-evolution of climate with tectonics and life on Earth over the last half billion years



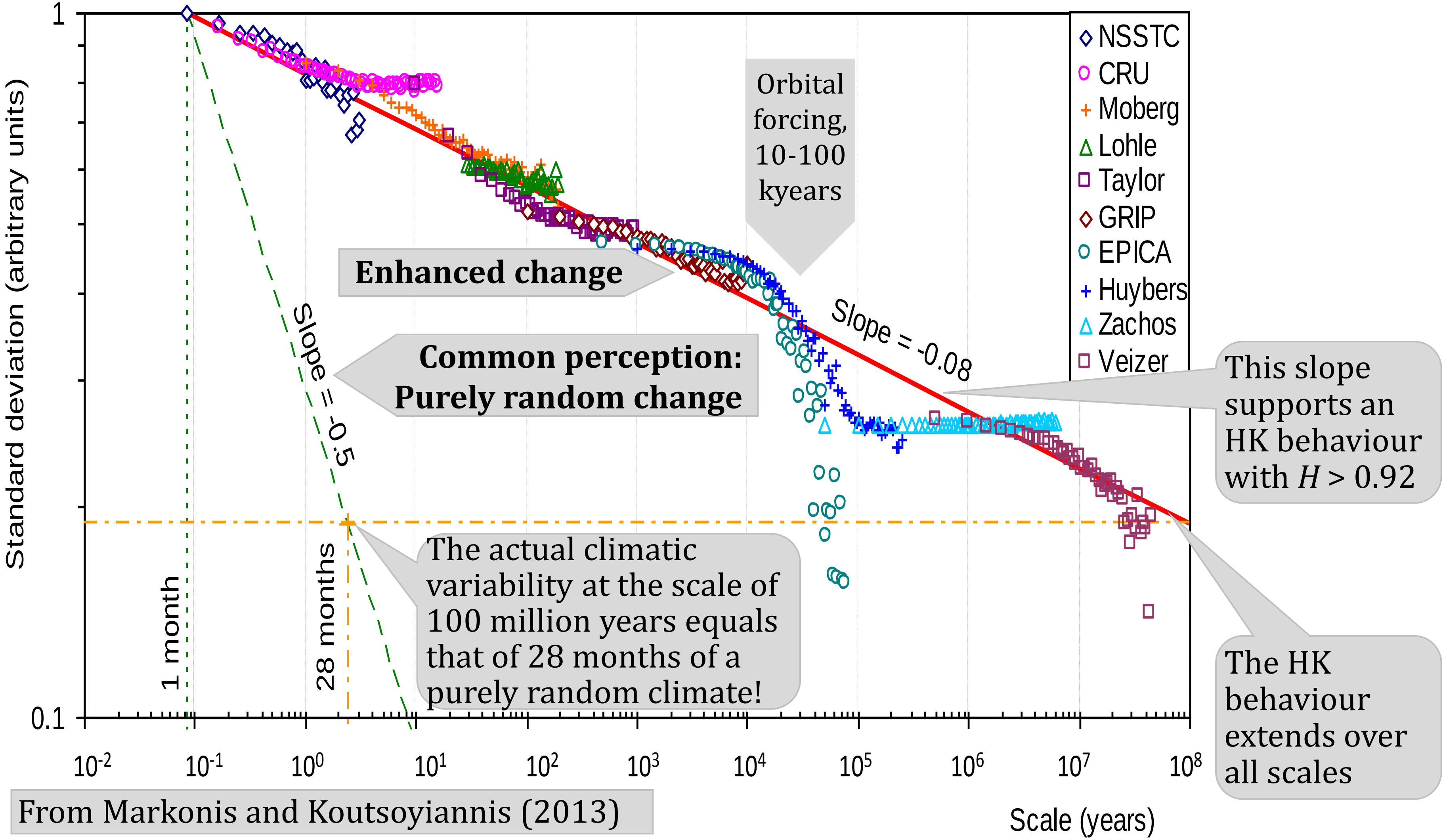
Palaeoclimate temperature estimates based on $\delta^{18}O$; adapted from Veizer *et al.* (2000)

Temperature change on Earth based on observations and proxies

Details on data sets: see Markonis and Koutsoyiannis (2013)

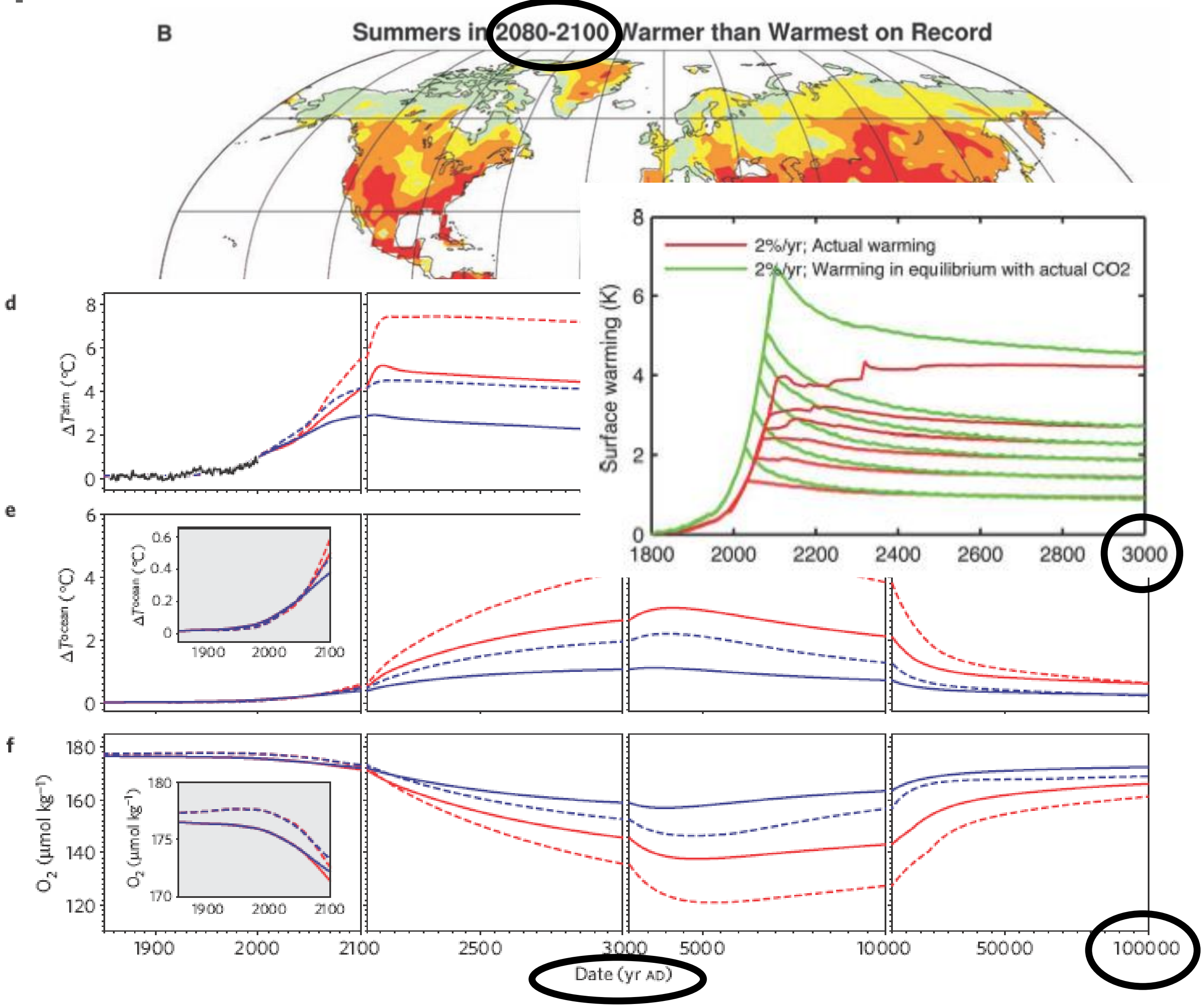


A combined climacogram of all 10 temperature observation sets and proxies



From Markonis and Koutsoyiannis (2013)

Future climate: Long term predictions are trendy...

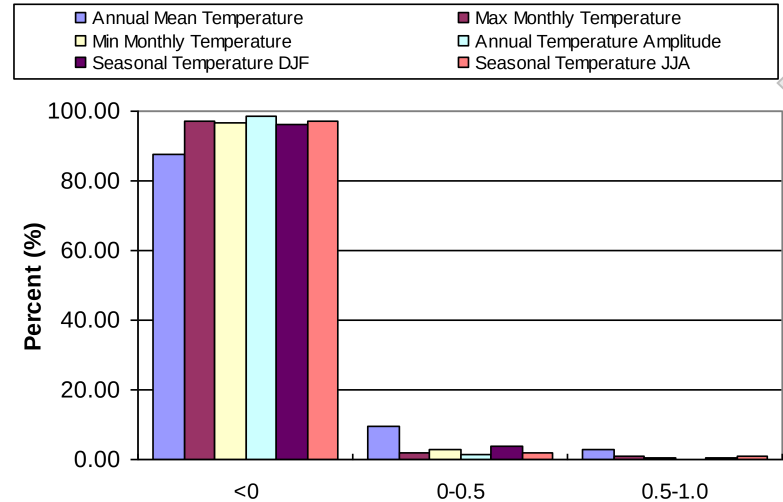


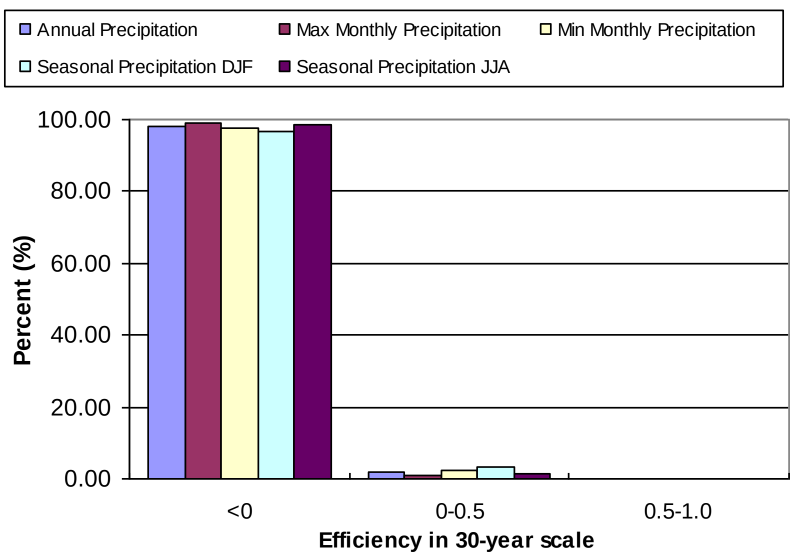
From 2100 AD (Battisti and Naylor, *Science*, 2009)...

... to 3000 AD (Solomon et al., *Nature Geoscience*, 2009)

...to 100 000 AD (Shaffer et al., *PNAS*, 2009)

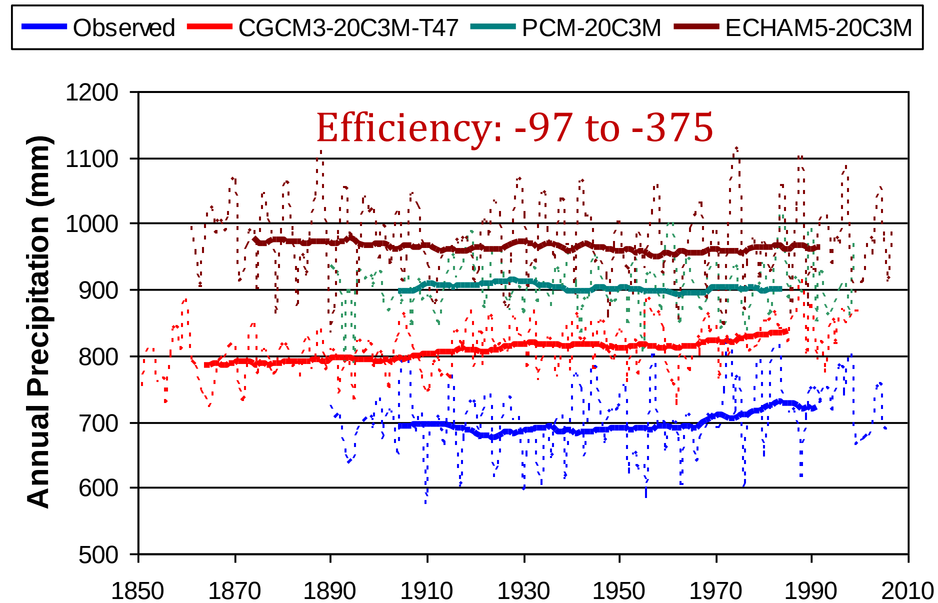
Is there any indication that climate is predictable in deterministic terms?





Comparison of 3 IPCC TAR and 3 IPCC AR4 climate models with historical series of more than 100 years length in 55 stations worldwide

Comparison of 3 IPCC AR4 climate models with reality in sub-continental scale (contiguous USA)



Source: Anagnostopoulos, *et al.* (2010);
See also Koutsoyiannis *et al.* (2008).

Concluding remarks

- The natural world obeys the laws of physics; there are no particular agents of randomness (at the visible, macroscopic world).
- However, there are agents of change, acting on several scales; the world exists only in change.
- Change and uncertainty are tightly connected; only dead systems are certain.
- Uncertainty is quantified by entropy, whose tendency to become maximum drives change.
- Physical laws support predictability—but only at short time scales; the distant future is unpredictable in terms of both exact state and average behaviour.
- Humans are part of the changing Nature—but change is hardly controllable by humans (fortunately).
- Stochastics is the tool to study complex natural systems.
- Hurst-Kolmogorov stochastic dynamics is the key to perceive multi-scale change and model the implied uncertainty and risk.

Both classical physics and quantum physics are indeterministic

Karl Popper (in his book "Quantum Theory and the Schism in Physics")

The future is not contained in the present or the past

W. W. Bartley III (in Editor's Foreward to the same book)

References

- Anagnostopoulos, G. G., D. Koutsoyiannis, A. Christofides, A. Efstratiadis, and N. Mamassis, A comparison of local and aggregated climate model outputs with observed data, *Hydrological Sciences Journal*, 55 (7), 1094–1110, 2010.
- Battisti, D.S. and R.L. Naylor, Historical warnings of future food insecurity with unprecedented seasonal heat, *Science* 323, 240-244, 2009.
- Bernoulli, J., *Ars Conjectandi*, Thurnisii fratres, 306+35 pp., Basel, 1713.
- Birkhoff, G.D., Proof of the ergodic theorem, *Proceedings of the National Academy of Sciences*, 17, 656-660, 1931.
- Hurst, H.E., Long term storage capacities of reservoirs, *Trans. Am. Soc. Civil Engrs.*, 116, 776–808, 1951.
- Jaynes, E.T., *Probability Theory: The Logic of Science*, Cambridge Univ. Press, 2003.
- IPCC, Climate Change 2013: The Physical Science Basis, Working Group I contribution to the IPCC 5th Assessment Report, 2013 (final draft Report, dated 7 June 2013 approved on 26 September 2013).
- 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
- Kolmogorov, A. N., *Grundbegrijfe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Math. (2), Berlin, 1933; 2nd English Edition: *Foundations of the Theory of Probability*, 84 pp. Chelsea Publishing Company, New York, 1956.
- Kolmogorov, A. N., Wienersche Spiralen und einige andere interessante Kurven in Hilbertschen Raum, *Dokl. Akad. Nauk URSS* 26, 115–118, 1940.
- Koutsoyiannis, D., Hydrology and Change, *Hydrological Sciences Journal*, 58 (6), 1177–1197, 2013.
- 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., A. Efstratiadis, N. Mamassis, and A. Christofides, On the credibility of climate predictions, *Hydrological Sciences Journal*, 53 (4), 671–684, 2008.
- 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.
- 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.
- Stager, J. C., D. B. Ryves, B. M. Chase and F. S. R. Pausata, Catastrophic Drought in the Afro-Asian Monsoon Region During Heinrich Event, *Science*, DOI: 10.1126/science.1198322, 2011
- Sutcliffe, J. V., and G. Petersen, Lake Victoria: derivation of a corrected natural water level series, *Hydrological Sciences Journal*, 52 (6), 1316-1321, 2007
- Dimitriadis, P., K. Tzouka, and D. Koutsoyiannis, Windows of predictability in dice motion, *Facets of Uncertainty: 5th EGU Leonardo Conference – Hydrofractals 2013 – STAHY 2013*, Kos Island, Greece, European Geosciences Union, International Association of Hydrological Sciences, International Union of Geodesy and Geophysics, 2013.
- Lasota, A., and M.C. Mackey, *Chaos, Fractals and Noise*, Springer-Verlag, 1994.
- Metropolis, N., and S. Ulam, The Monte Carlo method, *Journal of the American Statistical Association*, 44(247), 335-341, 1949.
- Niederreiter, H., *Random Number Generation and Quasi-Monte Carlo Methods*, Society for Industrial and Applied Mathematics, Philadelphia, 1992.
- Shaffer, G., S.M. Olsen and J.O.P. Pedersen, Long-term ocean oxygen depletion in response to carbon dioxide emissions from fossil fuels, *Nature Geoscience*, DOI: 10.1038/NGEO420, 2009.
- Solomon, S., G.-K. Plattner, R. Knutti and P. Friedlingstein, Irreversible climate change due to carbon dioxide emissions, *Proceedings of the National Academy of Sciences*, 106(6), 1704–1709, 2009.
- Takens, F., Detecting strange attractors in turbulence. In: *Dynamical Systems and Turbulence*, D.A. Rand and L.-S. Young (eds.), 336–381, Lecture Notes in Mathematics no. 898, Springer-Verlag, New York, USA, 1981.
- Toussoun, O., 1925.Mémoire sur l'histoire du Nil. Mémoire de l'Institut d'Egypte, 18, 366–404.
- Veizer, J., D. Ala, K. Azmy, P. Bruckschen, D. Buhl, F. Bruhn, G. A. F. Carden, A. Diener, S. Ebneth, Y. Godderis, T. Jasper, C. Korte, F. Pawellek, O. Podlaha and H. Strauss, 87Sr/86Sr, d13C and d18O evolution of Phanerozoic seawater, *Chemical Geology*, 161, 59-88, 2000