30 Years of Nonlinear Dynamics in Geosciences
Rhodes, Greece,
03-08 July 2016

From fractals to stochastics:

Seeking theoretical consistency in analysis of geophysical data

Demetris Koutsoyiannis[1*], Federico Lombardo[2], Panayiotis Dimitriadis[1], Yannis Markonis[1] and Spencer Stevens[3]

[1]Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Greece

[2] Independent researcher, Rome, Italy.

[3] Independent researcher, London, UK.

(*dk@itia.ntua.gr, http://www.itia.ntua.gr/dk/)

Presentation available online: http://www.itia.ntua.gr/1627/

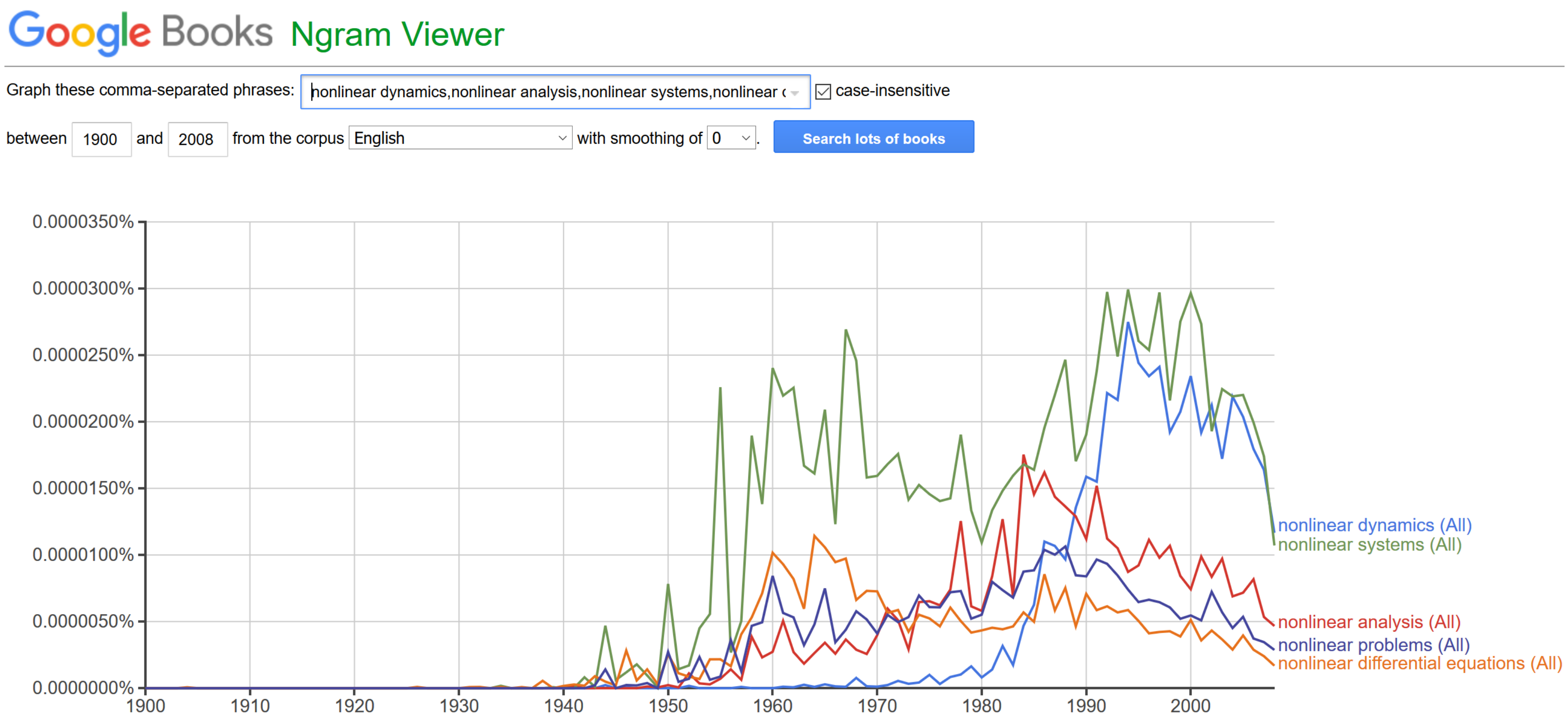
Part 1: On nonlinear dynamics and fractals

Τὸ ἀντίξουν συμφέρον καὶ ἐκ τῶν διαφερόντων καλλίστην ἁρμονίαν καὶ πάντα κατ' ἔριν γίνεσθαι

Opposition unites, the finest harmony springs from difference, and all comes about by strife

Heraclitus; ca. 540-480 BC

30 Years of Nonlinear Dynamics?



http://books.google.com/ngrams/

30 Years of Nonlinear Dynamics?

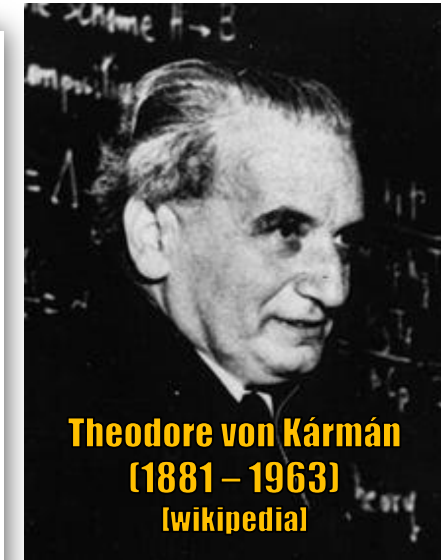
THE ENGINEER GRAPPLES WITH NONLINEAR PROBLEMS[1]

THEODORE VON KÁRMÁN

I do not believe that one could connect justly the name of Gibbs with practical applications of applied mathematics, for his main interest was certainly centered on basic conceptions of mathematical physics. Nevertheless, for example, his beautiful work on graphical methods in thermodynamics is a brilliant example of the presentation of theoretical relations in a form which appeals to the engineer.

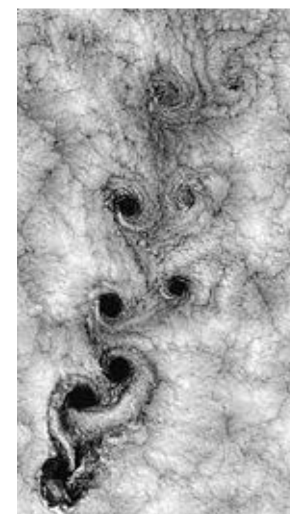
This lecture is intended as an effort to improve the convergence between the viewpoints of mathematics and engineering. Thus, I feel it is not inappropriate to dedicate it to the memory of Josiah Willard Gibbs.

[1] The fifteenth Josiah Willard Gibbs Lecture, delivered at Columbus, Ohio, December 27, 1939, under the auspices of the American Mathematical Society with the cooperation of the American Association for the Advancement of Science.



(von Kármán, 1940)

Kármán Vortex Street caused by wind flowing around the Juan Fernández Islands off the Chilean coast (Wikipedia)



Why do we not prefer fractals (over stochastics)?

- A lot of ambiguity
- Confusion between local and global properties of processes
- Use of the abstract mathematical processes as if they could apply to natural processes
- Hasty use of stochastic concepts
- Misspecification / misinterpretation of scaling laws
- Neglect of statistical bias and variation
- Confusion between different scaling behaviours

Careful use of stochastics can deal with all problems involving fractals of non-deterministic type in a more rigorous manner and more effectively.

Ambiguity

*The terms **fractal** and **multifractal** remain **without an agreed mathematical definition**. Let me argue that this situation ought not create concern and steal time from useful work. Entire fields of mathematics thrive for centuries with a clear but evolving self-image, and nothing resembling a definition.*

(Mandelbrot, 1999, p. 14; Note: Mandelbrot coined the term *fractal* in 1975).

[Mandelbrot, 1982] *observes that "Ordinary words used in scientific discourse combine (a) diverse intuitive meanings, dependent on the user, and (b) formal definitions, each of which singles out one special meaning and enshrines it mathematically. The terms stationary and ergodic are fortunate in that mathematicians agree on them. However, experience indicates that many engineers, physicists, and practical statisticians pay lip service to the mathematical definition, but hold narrower views." That is, **many mathematically stationary processes are not intuitively stationary**. By and large, those processes exemplify wild randomness, a circumstance that provides genuine justification for distinguishing a narrower and a wider view of stationarity.*

(Mandelbrot, 1999, p.7)

Ambiguity (contd.)

*We are done now with explaining the **peaceful coexistence of two values of D**: the dimension $D = 1/H = 2$ applies to that three-dimensional curve, as well as to the trail obtained by projecting on the plane (X, Y). However, the projections of the three dimensional curve on the planes (t,X) and (t,Y) are of dimension $D = 2 - H = 1.5$.*

(Mandelbrot, 1999, p. 45)

***Definition**. The term **multifractal** denotes the most general category of multibox **cartoons**. It allows the generator to combine axial boxes and diagonal boxes with non-identical values of H_i from $H_{min} > 0$ to $H_{max} < \infty$.*

(Mandelbrot, 1999, p. 45)

Contrast

***Each definition is a piece of secret ripped from Nature by the human spirit**. I insist on this: any complicated thing, being illumined by definitions, being laid out in them, being broken up into pieces, will be separated into pieces completely transparent even to a child, **excluding foggy and dark parts that our intuition whispers to us while acting**, separating into logical pieces, **then only can we move further,** towards new successes due to definitions . . .*

(Nikolai Luzin; from Graham and Kantor, 2009)

Attempts to remove ambiguity

There is no "official" consensus on the definition of a fractal. However, what is generally agreed on is that the Hausdorff measure and Hausdorff dimension play a key role. One ***possible definition of a fractal*** *is then for example that it* ***is a set $A \subseteq R^k$ whose Hausdorff dimension $\dim_{\text{Haus}} A$ is not an integer****.*
(Beran et al., 2013, p. 178)

There are many definitions of fractal dimension. The most general and mathematically satisfactory one is the Hausdorff dimension D_{Haus}.
(Veneziano and Langousis, 2010, p. 4)

However:
In the context of time series analysis, fractal behaviour is often mentioned as synonym for long-range dependence. *Though there are strong connections between the two notions, they are also in some sense completely different.*
(Beran et al., 2013, p. 178)

Note: The Hausdorff dimension expresses a local property, as a radius δ for covering the set A tends to zero. This is more evident in the so-called *box-counting dimension*, which is an upper bound for D_{Haus} (Beran et al., 2013, p. 181-182) and is defined as $\dim_{\text{Box}} A = \lim_{\delta \to 0} \log N_\delta / \log \delta$ where N_δ is the minimal number of sets U_i needed for a δ-cover of A.

Confusion between local and global properties of processes

- Most fractal literature confuses fractal behaviour with long-range dependence.
- However, Mandelbrot (1999, p.3) referred to the difference of locality and globality but in a rather obscure way:
  *The importance of the contrast between mildness and wildness is in part due to its links with a contrast between **locality** and **globality**.*
- Gneiting and Schlather (2004), using a Cauchy type autocovariance function* made it clear that the fractal and Hurst properties (long-range dependence) are two different things:
  - The **fractal parameter** determines the **local properties** of the process (as $t \to 0$)
  - The **Hurst parameter** determines the **global properties** of the process (as $t \to \infty$)
- These issues were also elaborated in Beran et al. (2013).
- Koutsoyiannis (2013a,b, 2016) introduced a Cauchy type climacogram which has better properties than the Cauchy type autocovariance function, allowing for negative autocorrelations (antipersistence) at large time lags while ensuring positive autocorrelations at small lags, as demanded for physical consistency.

---

* It was first proposed by Yaglom (1987, p. 365) and also referred to by Wackernagel (1995, p. 219; 1998, p. 246), while a similar autocorrelation was used by Koutsoyiannis (2000) for discrete time processes.

Use of the abstract mathematical objects as if they could apply to natural processes

- In mathematical processes the local and global properties can be the same (e.g. the Hurst-Kolmogorov process, described by a single scaling exponent applicable at all scales).
- Scale independence or absence of characteristic scales in a process or a phenomenon is mathematically attractive.
- However, in Nature these cannot be the case; for example:
  - If the set of points along a river's bed had fractal dimension > 1 (meaning that number of sets of its δ-cover would be a power law of δ with exponent > 1 for arbitrary low δ) then the geometrical length of the river would be infinite and any particle of water would take infinite time to reach the sea.
  - If a Hurst-Kolmogorov process were applicable for arbitrary low time scales, it would entail infinite variance of the instantaneous, continuous-time process which would imply infinite energy (infrared catastrophe).
  - If an antipersistent Hurst-Kolmogorov process (with Hurst exponent H < 0.5) were applicable for arbitrary low time scales, it would entail negative autocovariance (anti-correlation) for arbitrary small lags which is absurd (in a natural process the autocorrelation should tend to 1 as lag tends to 0).

Hasty use of stochastic concepts

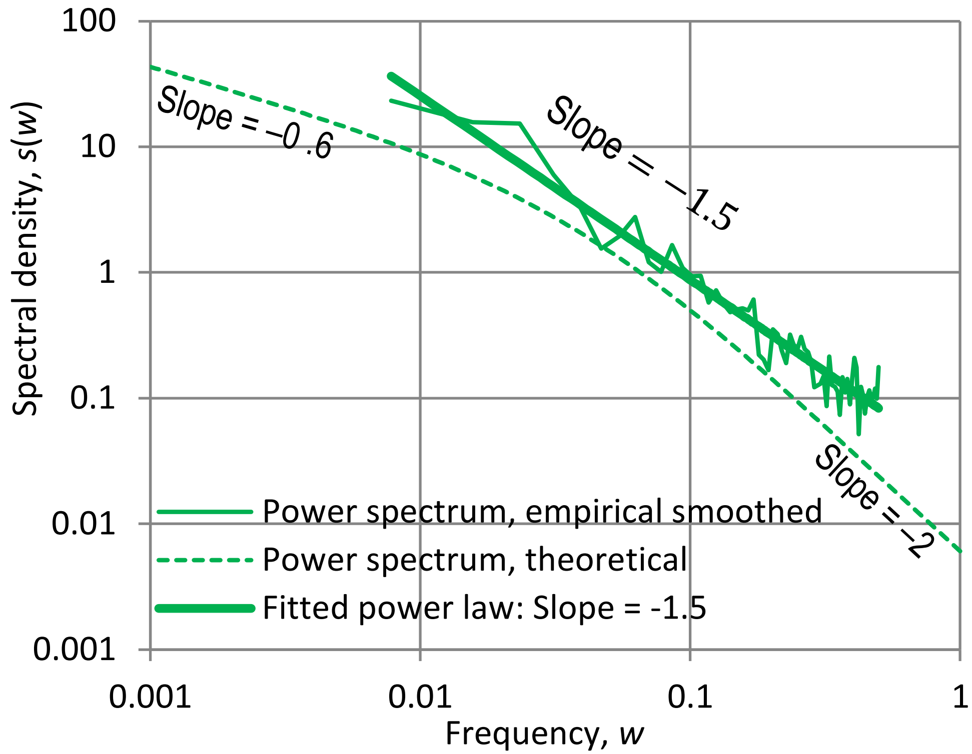
- Use of formulations and tools of stationary stochastic processes while making claims of nonstationarity (e.g. in a nonstationary process, the autocovariance and the spectral density are functions of two variables, one being related to “absolute” time—see Dechant and Lutz—2015, but they have typically been presented as functions of one variable).
- Use of statistical estimations from data while claiming nonstationarity; however, a nonstationary process is nonergodic and thus the estimates are meaningless.
- A classical example: reporting logarithmic slopes in empirical power spectra $s^{\#} < -1$ (e.g. $s^{\#} = -1.5$, etc.) for small frequencies w (tending to zero). Note that the superscript # denotes logarithmic derivative, i.e.

$$f^{\#}(x) := \frac{\mathrm{d}(\ln f(x))}{\mathrm{d}\,(\ln x)} = \frac{x f'(x)}{f(x)} \tag{1}$$

- A slope $s^{\#}(w) < -1$ is mathematically and physically possible for large w but infeasible for $w \to 0$ (**see proof in Appendix**). Reported values $s^{\#} < -1$ for small w are invalid and are due to inconsistent estimation algorithms (stemming from the fact that the periodogram constructed from empirical autocovariances is too rough and the estimation of slopes from this is too uncertain; cf. Koutsoyiannis, 2013a,b; Koutsoyiannis et al., 2013; Dimitriadis and Koutsoyiannis, 2015).

Hasty use of stochastic concepts (contd.)

- It is not difficult to use inappropriate estimators and get inconsistent results, as exemplified in the graph of the power spectrum.
- In the example, 1024 data points have been generated from a stochastic process which has the theoretical power spectrum with the indicated varying slope (specifically, an HHK process—see below—with parameters $\kappa = 0.5$, $H = 0.8$, $\alpha = \lambda = 1$—see Koutsoyiannis, 2014).
- The standard empirical power spectrum is too rough to recover the underlying model and its parameters, and even by smoothing (here by averaging from 8 segments) relies in high bias and a misleading constant slope of −1.5.
- The theoretically consistent asymptotic slopes (−0.6 and −2) can be recovered by other methods (CBS—see below).



Misspecification / misinterpretation of scaling laws

- Experiment: A Google search with terms *universal multifractal rainfall* was performed (similar to the experiment in Koutsoyiannis, 2010).
- The first (highest PageRank) paper was chosen and its first figure is reproduced here.

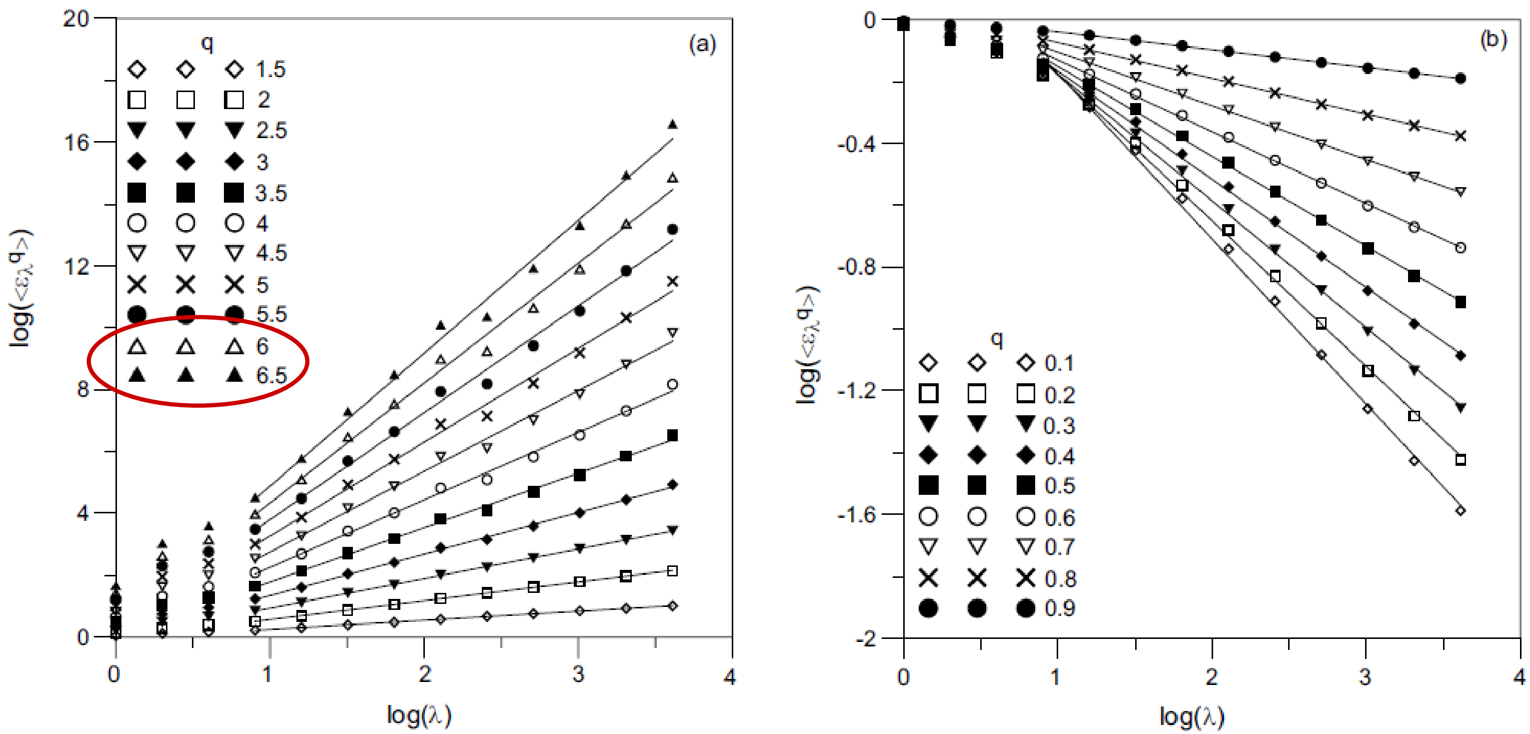
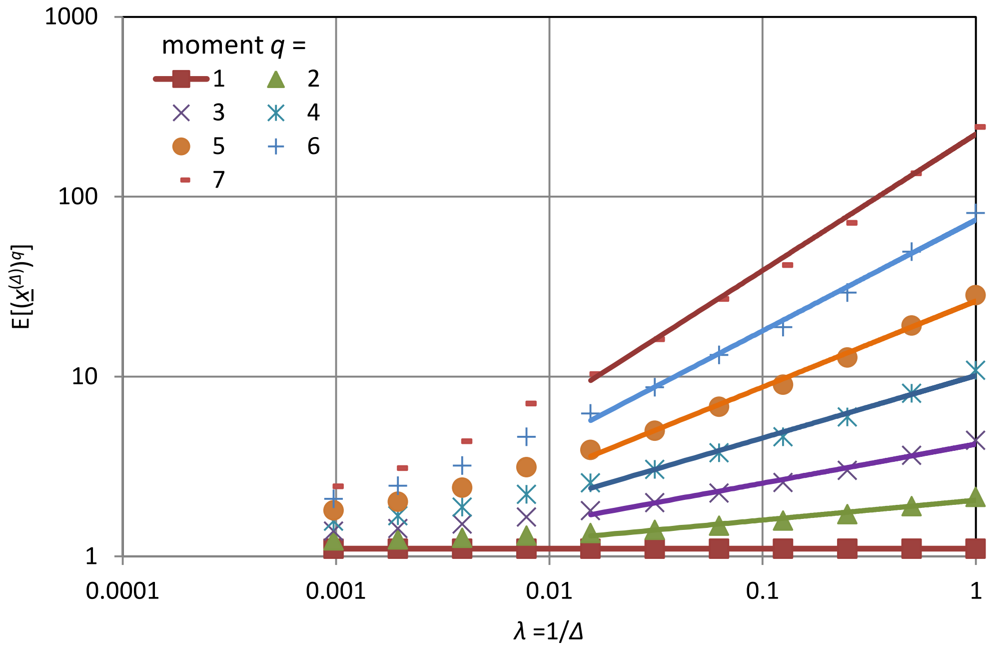


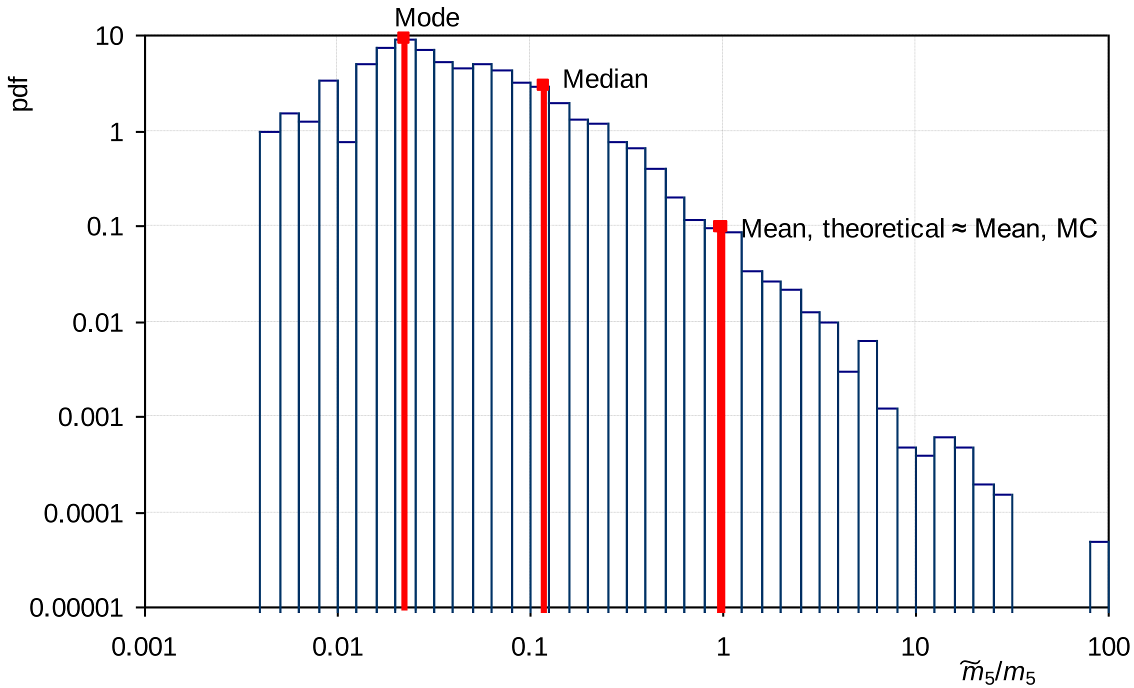
Fig. 1. Log-log plot of the qth moments of the rainfall intensity on the time scales from 1 hour to almost 6 months versus the scale ratio λ. (a) For moments larger than 1; (b) for moments smaller than 1.

Misinterpretation of scaling laws (contd.)

- In the example illustrated in the graph, a time series with length $N = 2^{13} = 8192$ was generated from the Hurst-Kolmogorov process (see below) with Hurst coefficient H = 0.8 and Gaussian distribution N(1,1).
- Some scaling laws seam to appear at a range of time scales, which are spurious. One could be misled to assume a multifractal behaviour and specify a $K(q)$ function, where $K(q)$ is the log-log slope of the raw moment $\mathrm{E}[(\underline{x}^{(\Delta)})^q]$ vs. the inverse time scale $\lambda = 1/\Delta$ for specified q.
- The truth is that there is no multifractal behaviour here. As shown theoretically by Lombardo et al. (2014), (a) there is no constant slope (e.g., as $\lambda \to 0$, or $\Delta \to \infty$, $K(q) = 0$); also the slope empirically estimated for small Δ (large λ) is too low compared to its theoretical value.
- This is a symptom of a more general tendency in the fractal literature to treat observations (time series) deterministically, confusing random variables with their realizations and ignoring statistical bias and variation.



Neglect of statistical bias and variation



- The example illustrates that estimates of high-order moments, which have been popular in multifractal studies, 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; see also Lombardo et al., 2014).
- Here the theory guarantees 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.

Confusion between different scaling behaviours

- Temporal scaling
  - It indicates dependence in time and is expressed as a power law of either:
    - ✓ autocorrelation vs. time lag and/or climacogram vs. time scale (Hurst);
    - ✓ structure function vs. time lag and/or differential climacogram vs. time scale (fractal/local behaviour).
- Spatial scaling
  - It is similar to temporal scaling but indicating dependence in space.
- State scaling
  - Totally irrelevant to temporal/spatial scaling; it is related to the marginal distribution of the process and indicates heavy-tailed distributions (power laws of probability of exceedence vs. state).
- Scaling of high-order moments with time scale
  - Perhaps an artefact related to other types of scaling; usually spurious because high-order moments are not reliably estimated from data.

- As already mentioned, scaling laws never extend to the entire range of scales.
- Usually they are asymptotic laws, with different exponents at each edge.
- Scaling laws, as asymptotic ones, abound because they are a mathematical necessity (Koutsoyiannis, 2014).

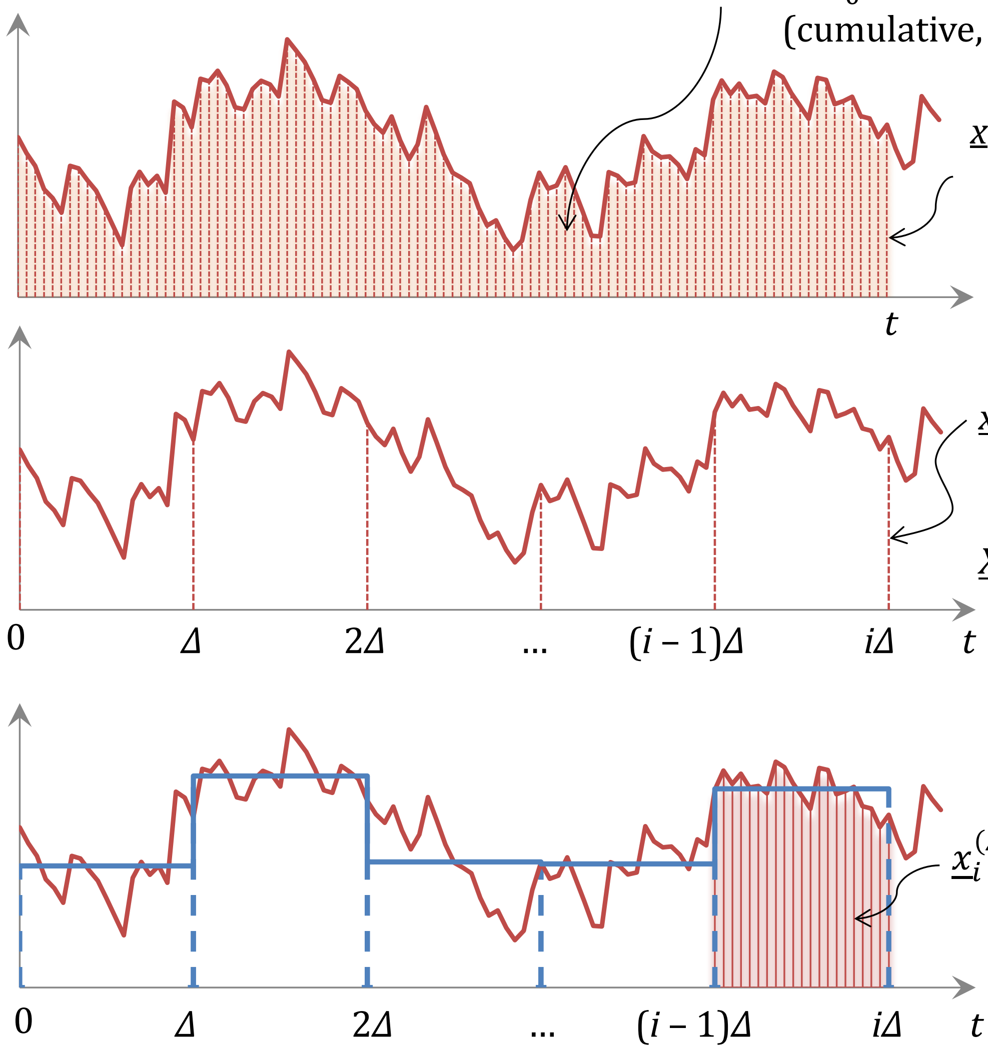
Part 2: Stochastics

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 $f(x)$. Notice the different notation of random variables (underlined, Hemelrijk, 1966) from regular ones.
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

From continuous time to discrete time processes

$\underline{X}(t) \coloneqq \int_0^t \underline{x}(\xi)\mathrm{d}\xi$
(cumulative, nonstationary)



$\underline{x}(t)$ (instantaneous, continuous-time process)

$\underline{x}_i \coloneqq \underline{x}(i\Delta)$
(instantaneous process sampled at spacing Δ)

$\underline{X}_i \coloneqq \sum_{j=1}^{i} \underline{x}_j$
(cumulative sampled at spacing Δ, nonstationary)

$\underline{x}_i^{(\Delta)} \coloneqq \frac{1}{\Delta}\int_{(i-1)\Delta}^{i\Delta} \underline{x}(\xi)d\xi =$
$= \underline{X}(i\Delta)/\Delta - \underline{X}((i-1)\Delta)/\Delta$
(averaged at time scale Δ)

Most natural processes evolve in *continuous time* but they are observed in *discrete time*, instantaneously or by averaging

Important note: The graphs display a realization of the process while the notation is for the process per se.

Second order properties of a stationary stochastic process

- **Autocovariance function**, $c(\tau) := \mathrm{Cov}[\underline{x}(t), \underline{x}(t+\tau)]$, where τ is time lag.
- **Power spectrum** (*spectral density*), $s(w)$, where w is frequency (inverse time).
- **Structure function** (*semivariogram* or *variogram*), $h(\tau) \coloneqq \frac{1}{2}\mathrm{Var}\left[\underline{x}(t) - \underline{x}(t+\tau)\right]$.
- **Climacogram**, $\gamma(\Delta)$, where Δ denotes time scale, so that $\gamma(\Delta) := \mathrm{Var}[\underline{x}_i^{(\Delta)}]$.
- All these properties are transformations of one another, i.e.:

$$s(w) = 4\int_0^\infty c(\tau)\cos(2\pi w\tau)\,\mathrm{d}\tau, \quad c(\tau) = \int_0^\infty s(w)\cos(2\pi w\tau)\,\mathrm{d}w \tag{2}$$

$$h(\tau) = c(0) - c(\tau), \quad c(\tau) = c(0) - h(\tau) \tag{3}$$

$$\gamma(\Delta) = 2\int_0^1 (1-\xi)c(\xi\Delta)\mathrm{d}\xi, \quad c(\tau) = \frac{1}{2}\frac{\mathrm{d}^2\left(\tau^2\gamma(\tau)\right)}{\mathrm{d}\tau^2} \tag{4}$$

- In estimation from data, the climacogram behaves better than all other tools, which involve high bias and statistical variation (Dimitriadis and Koutsoyiannis, 2015 Koutsoyiannis, 2016). The climacogram involves bias too, but this can be determined analytically and included in the estimation.

Second order properties at discrete time

- Once the continuous-time properties are determined, the discrete-time ones can be calculated.
- For example, the autocovariance of the averaged process is:

$$c_j^{(\Delta)} = \mathrm{Cov}\left[\underline{x}_i^{(\Delta)}, \underline{x}_{i+j}^{(\Delta)}\right] = \frac{1}{\Delta^2}\left(\frac{\Gamma(|j+1|\Delta)+\Gamma((|j-1|\Delta)}{2} - \Gamma(|j|\Delta)\right) \quad (5)$$

where $\Gamma(\Delta) \coloneqq \mathrm{Var}[\underline{X}(\Delta)] = \Delta^2\gamma(\Delta)$.

- Also, the power spectrum of the averaged process can be calculated from:

$$s_\mathrm{d}^{(\Delta)}(\omega) = 2c_0^{(\Delta)} + 4\sum_{j=1}^{\infty} c_j^{(\Delta)}\cos(2\pi\omega j) \quad (6)$$

where $\omega \coloneqq w\Delta$, $s_\mathrm{d}^{(\Delta)}(\omega) = s^{(\Delta)}(w)/\Delta$ (nondimensionalized frequency and spectral density, respectively).

- More details and additional cases can be found in Koutsoyiannis (2013b, 2016).

Climacogram based metrics of stochastic processes

Metric / Usefulness	**Definition**	**Comments**
Climacogram For the global asymptotic behaviour ($\Delta \to \infty$)	$\gamma(\Delta) \coloneqq \mathrm{Var}[\underline{x}_i(\Delta)]$	For an ergodic process for $\Delta \to \infty$ $\gamma(\Delta) \to 0$ necessarily
Differential climacogram (DC) (or climacogram-based structure function, CBSF) For the local asymptotic behaviour ($\Delta \to 0$)	$g(\Delta) \coloneqq \gamma_0 - \gamma(\Delta)$ 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
Climacogram-based spectrum (CBS) For both the global and local asymptotic behaviour	$\psi(w) \coloneqq \frac{2}{w\gamma_0}\gamma(1/w)g(1/w)$ $= \frac{2\,\gamma(1/w)}{w}\left(1 - \frac{\gamma(1/w)}{\gamma_0}\right)$ where $w \equiv 1/\Delta$ 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 DC is related to the structure function $h(\tau)$ by the same way as the climacogram is related to the autocovariance function $c(\tau)$:

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

Cautionary notes for model fitting

- Direct estimation of **any statistic** of a process (except perhaps for the mean) is not possible merely from the data; **we always need to assume a model**.
- Any statistical estimator $\underline{\hat{s}}$ of a true parameter s is biased either strictly (meaning: $\mathrm{E}\left[\underline{\hat{s}}\right] \neq s$) or loosely (meaning: $\mathrm{mode}\left[\underline{\hat{s}}\right] \neq s$).
- Model fitting is necessarily based on discrete-time data and needs to consider the effects of (a) discretization and (b) bias.
- The climacogram provides easy means to analytically estimate from its true expression (that in continuous time) both effects.
- As an example, we consider a process with climacogram $\gamma(\Delta)$, from which we have a time series for an observation period T (multiple of Δ), each one giving the averaged process $\underline{x}_i^{(\Delta)}$ at a time step Δ, so that the sample size is $n = T/\Delta$.
- The standard estimator $\underline{\hat{\gamma}}(\Delta)$ of the variance $\gamma(\Delta)$ of the averaged process is

$$\underline{\hat{\gamma}}(\Delta) := \frac{1}{n-1} \sum_{i=1}^{n} \left(\underline{x}_i^{(\Delta)} - \underline{x}_1^{(T)}\right)^2 = \frac{1}{T/\Delta - 1} \sum_{i=1}^{T/\Delta} \left(\underline{x}_i^{(\Delta)} - \underline{x}_1^{(T)}\right)^2 \quad (7)$$

- As shown in Koutsoyiannis (2011, 2016) the bias can be calculated from

$$\mathrm{E}\left[\underline{\hat{\gamma}}(\Delta)\right] = \eta(\Delta, T)\gamma(\Delta), \quad \eta(\Delta, T) = \frac{1-\gamma(T)/\gamma(\Delta)}{1-\Delta/T} = \frac{1-(\Delta/T)^2\Gamma(T)/\Gamma(\Delta)}{1-\Delta/T} \quad (8)$$

Entropy and entropy production

- The Boltzmann-Gibbs-Shannon entropy of a cumulative process $\underline{X}(t)$ with probability density function $f(X;t)$ is a dimensionless quantity defined as:

$$\Phi[\underline{X}(t)] := \mathrm{E}\left[-\ln\frac{f(X;t)}{h(X)}\right] = -\int_{-\infty}^{\infty} \ln\frac{f(X;t)}{h(X)} f(X;\ t)\, dX \tag{9}$$

  where $h(X)$ is the density of a background measure (typically Lebesgue).

- The entropy production in logarithmic time (EPLT) is a dimensionless quantity, the derivative of entropy in logarithmic time (Koutsoyiannis, 2011):

$$\varphi(t) \equiv \varphi[\underline{X}(t)] := \Phi'[\underline{X}(t)]\, t \equiv \mathrm{d}\Phi[\underline{X}(t)] \,/\, \mathrm{d}(\ln t) \tag{10}$$

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

$$\Phi[\underline{X}(t)] = (1/2)\ln(2\pi \mathrm{e}\, \Gamma(t)/h^2),\ \ \varphi(t) = \Gamma'(t)\, t \,/\, 2\Gamma(t) \tag{11}$$

- When the past ($t < 0$) and the present ($t = 0$) are observed, instead of the unconditional variance $\Gamma(t)$ we should use a variance $\Gamma_{\mathrm{C}}(t)$ conditional on the past and present:

$$\Gamma_{\mathrm{C}}(t) \approx 2\Gamma(t) - \Gamma(2t)/2, \quad \varphi_{\mathrm{C}}(t) = \frac{\Gamma_{\mathrm{C}}'(t)t}{2\Gamma_{\mathrm{C}}(t)} \approx \frac{(2\Gamma'(t) - \Gamma'(2t))t}{4\Gamma(t) - \Gamma(2t)} \tag{12}$$

Resulting processes from maximizing entropy production

- A Markov process:

$$c(\tau) = \lambda \mathrm{e}^{-\tau/\alpha},$$

$$\gamma(\Delta) = \frac{2\lambda}{\Delta/\alpha}\left(1 - \frac{1-\mathrm{e}^{-\Delta/\alpha}}{\Delta/\alpha}\right) \qquad (13)$$

maximizes entropy production for small times but minimizes it for large times.

- A Hurst-Kolmogorov (HK) process:

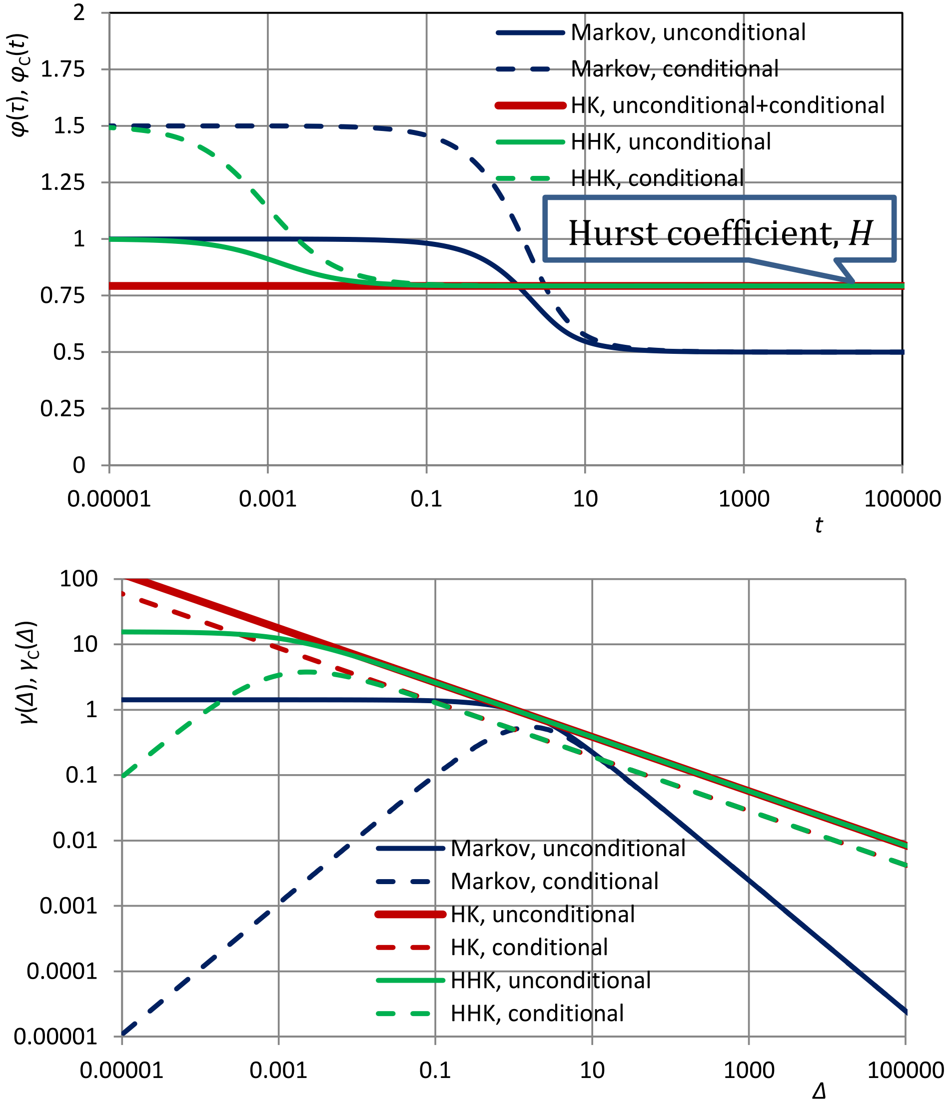
$$\gamma(\Delta) = \lambda(\alpha/\Delta)^{2-2H} \qquad (14)$$

maximizes entropy production for large times but minimizes it for small times

- A Hybrid Hurst Kolmogorov process

$$\gamma(\Delta) = \lambda\left(1 + (\Delta/\alpha)^{2\kappa}\right)^{\frac{H-1}{\kappa}} \qquad (15)$$

maximizes entropy production both at small and large time scales.



Part 3: Simulation of stochastic processes (at discrete time)

The symmetric moving average scheme

- The so-called symmetric moving average (SMA) method (Koutsoyiannis 2000) can directly generate time series with any arbitrary autocorrelation function provided that it is mathematically feasible:

$$\underline{x}_i = \sum_{l=-\infty}^{\infty} a_{|l|} \underline{v}_{i+l} \tag{16}$$

where a_j are coefficients calculated from the autocovariance function and $\underline{v}_i$ is white noise averaged in discrete-time.

- Assuming that we work for the averaged discrete-time process with power spectrum $s_{\mathrm{d}}^{(\Delta)}(\omega)$, it has been shown (Koutsoyiannis 2000) that the Fourier transform $s_{\mathrm{d}}^{a}(\omega)$ of the a_l series of coefficients is related to the power spectrum of the discrete time process as

$$s_{\mathrm{d}}^{a}(\omega) = \sqrt{2 s_{\mathrm{d}}^{(\Delta)}(\omega)} \tag{17}$$

- Thus, to calculate a_l we first determine $s_{\mathrm{d}}^{a}(\omega)$ from the power spectrum of the process and then we inverse the Fourier transform to estimate all a_l.

Handling of truncation error

- It is expected that the coefficients a_l will decrease with increasing l and will be negligible beyond some q ($l > q$), so that we can truncate (16) to read

$$\underline{x}_i = \sum_{l=-q}^{q} a_{|l|} \underline{v}_{i+l} \tag{18}$$

- This would introduce some truncation error in the resulting autocovariance function. To adjust for this on the variance, we the calculate the a_l from

$$a_l = a'_l + a'' \tag{19}$$

where the coefficients a'_l are calculated from inversing the Fourier transform of either $s_{\mathrm{d}}^{a}(\omega)$ or $s_{\mathrm{d}}^{a}(\omega)(1 - \mathrm{sinc}(2\pi\omega q))$(two options; Koutsoyiannis, 2016).

- The constant a'' is determined so that the variance is exactly preserved:

$$\gamma(\Delta) = \sum_{l=-q}^{q} a_{|l|}^2 = \sum_{l=-q}^{q} \left(a'_{|l|} + a''\right)^2 \tag{20}$$

- Solving for a'', this yields:

$$a'' = \sqrt{\frac{\gamma(\Delta) - \Sigma\alpha\prime^2}{2q+1} + \left(\frac{\Sigma\alpha\prime}{2q+1}\right)^2} - \frac{\Sigma\alpha\prime}{2q+1} \tag{21}$$

where $\Sigma\alpha' \coloneqq \sum_{l=-q}^{q} a'_{|l|}$ and $\Sigma\alpha'^2 \coloneqq \sum_{l=-q}^{q} a'^2_{|l|}$.

Handling of moments higher than second

- In addition to being general for any second order properties (autocovariance function), the SMA method can explicitly preserve higher marginal moments.
- Specifically, to produce a discrete-time process $\underline{x}_i$ with coefficient of skewness $C_{\mathrm{s},x}$ we need to use a white-noise process $\underline{v}_i$ with coefficient of skewness:

$$C_{\mathrm{s},v} = C_{\mathrm{s},x} \frac{\left(\sum_{l=-q}^{q} a_{|l|}^2\right)^{3/2}}{\sum_{l=-q}^{q} a_{|l|}^3} \quad (22)$$

- Likewise, to produce a process $\underline{x}_i$ with coefficient of kurtosis $C_{\mathrm{k},x}$ the process $\underline{v}_i$ should have coefficient of kurtosis:

$$C_{\mathrm{k},v} = \frac{C_{\mathrm{k},x}\left(\sum_{l=-q}^{q} a_{|l|}^2\right)^2 - 6\sum_{l=-q}^{q}\sum_{k=-q}^{q} a_{|l|}^2 a_{|k|}^2}{\sum_{l=-q}^{q} a_{|l|}^4} \quad (23)$$

- See details in Dimitriadis and Koutsoyiannis (2016).
- Note that the method can also be used in multivariate processes, represented by vectors (Koutsoyiannis, 2000).

Simple marginal distributions for generation of non-Gaussian white noise

- Four-parameter distributions are needed to preserve skewness and kurtosis.
- For light-tailed distributions of $\underline{v}$ we can use an extended and standardized version of the Kumaraswamy distribution (ESK) with distribution function:

$$F(v) = 1 - \left(1 - \left(\frac{v-c}{d}\right)^{a}\right)^{b} \tag{24}$$

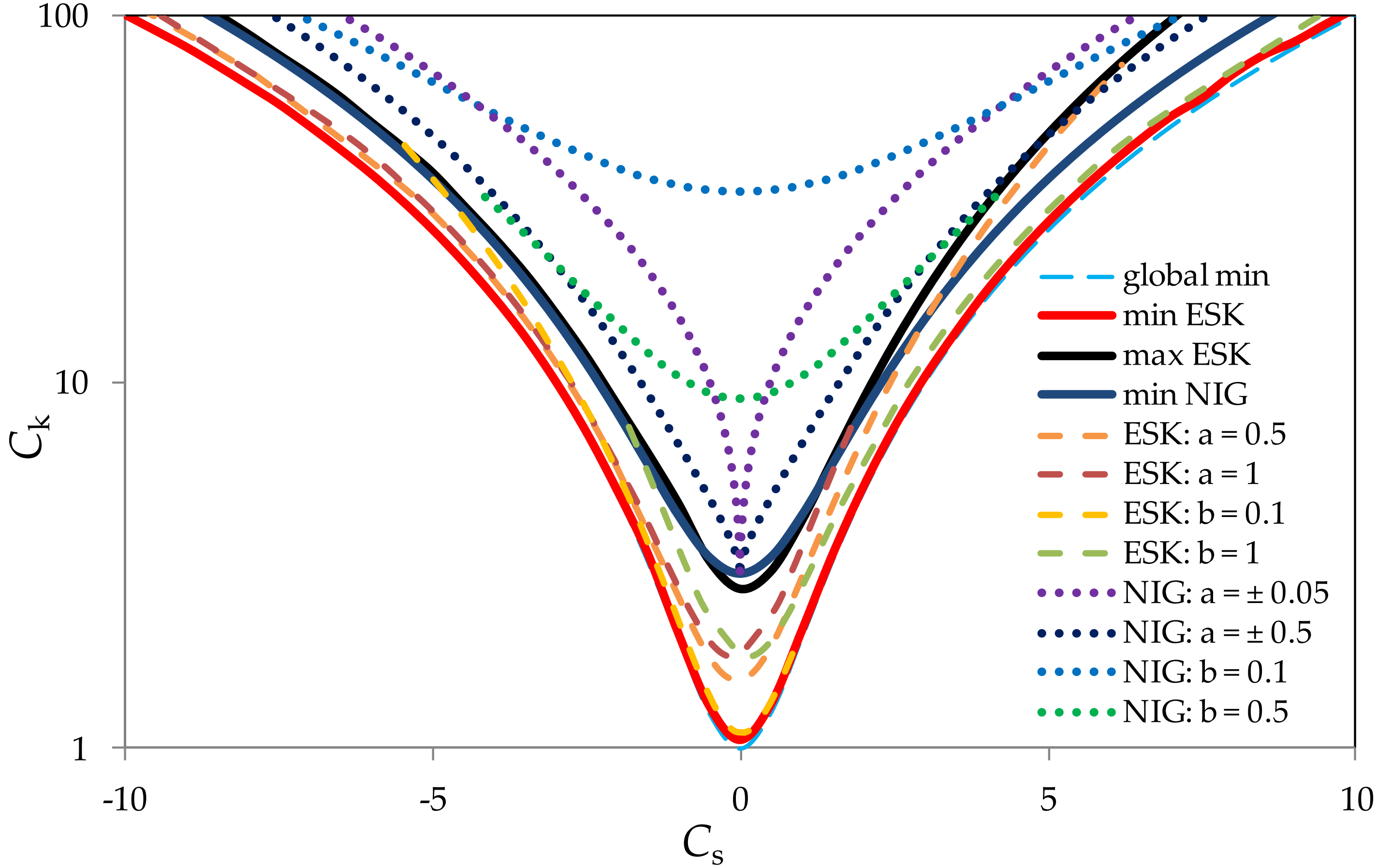
- For heavy-tailed distributions we can use the Normal-Inverse Gaussian (NIG) with probability density:

$$f(v) = \frac{\sqrt{a^2+b^2}\mathrm{e}^{b+a(v-c)/d}}{\pi d\sqrt{1+((v-c)/d)^2}} \mathrm{K}_1\left(\sqrt{a^2+b^2}\sqrt{1+((v-c)/d)^2}\right) \tag{25}$$

  with K_1 denoting a modified Bessel function of the third kind. Even though its mathematical form is involved, its moments are calculated analytically and the generation from the distribution is easy.

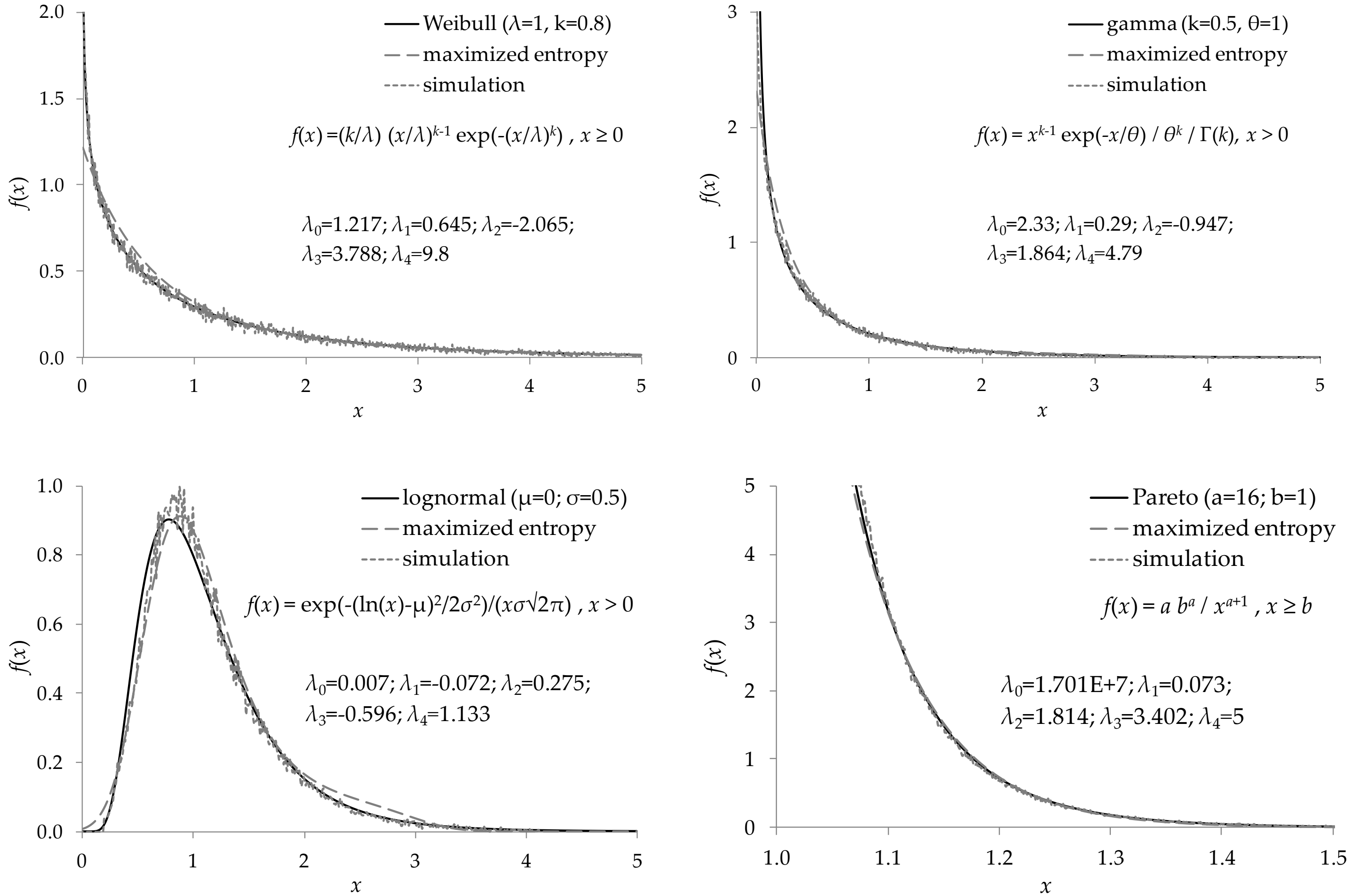
- In both cases *v* is the value of the random variable, *a* and *b* are dimensionless shape parameters, *c* is location parameter and *d* scale parameter; *c* and *d* have same dimensions as *v* (see details in Dimitriadis and Koutsoyiannis, 2016).

Range of skewness and kurtosis covered by the two distributions



Isopleths of parameters *a* or *b* of the ESK and the NIG distribution for the indicated skewness and kurtosis.

Performance in the generation of non-Gaussian white noise



Four two-parameter probability density functions, their approximations by maximum entropy distributions using four moments, i.e., $f(x) = \lambda_0 \exp(-\sum_{i=1}^{4}(x/\lambda_i)^i)$, and by the empirical density from a single synthetic time series with n = 10^5.

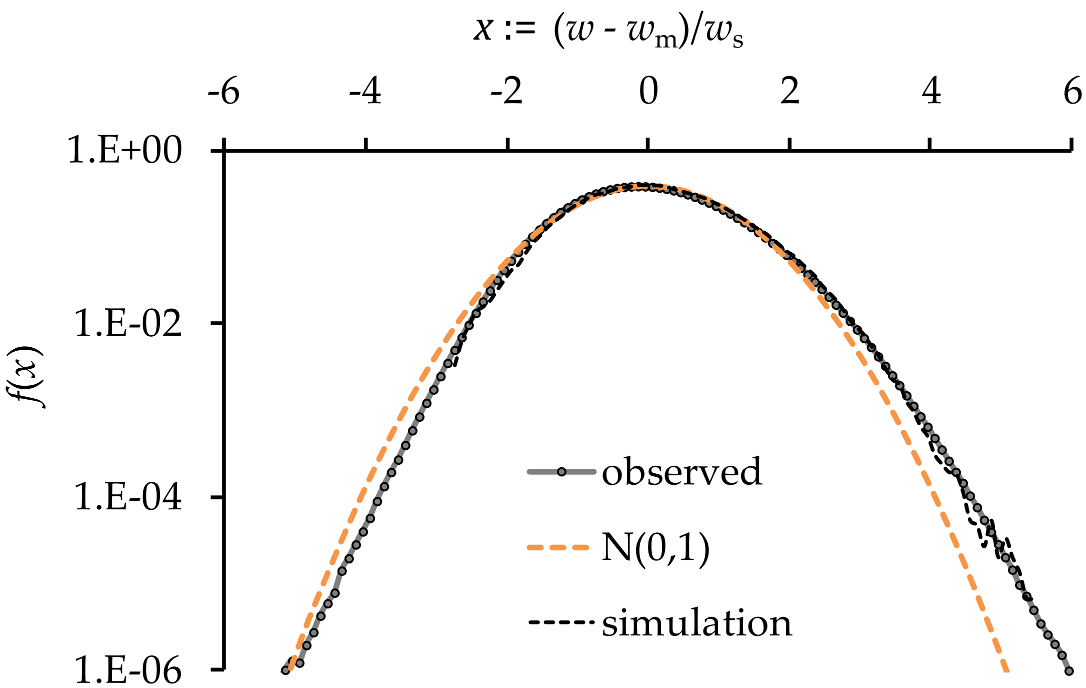
Part 4: Applications

Application 1: Microscale (turbulence)

- Estimation of high moments involves large uncertainty and cannot be reliable in the typically short time series of geophysical processes.
- On the contrary, high moments can be reliably estimated from large samples recorded in laboratory experiments at sampling intervals of μs.
- Here we use grid-turbulence data provided by the Johns Hopkins University (http://www.me.jhu.edu/meneveau/datasets/datamap.html).
- This dataset consists of 40 time series with $n = 36\times10^6$ data points of longitudinal wind velocity along the flow direction, all measured at a sampling time interval of 25 μs by X-wire probes placed downstream of the grid (Kang et al., 2003).
- By standardizing all series we formed a sample of $40 \times 36 \times10^6 = 1.44 \times10^9$ values to estimate the marginal distribution, and an ensemble of 40 series, each with 36×10^6 values to estimate the dependence structure through the climacogram.
- We also performed simulation using the SMA framework with $n = 10^6$ values.

Marginal distribution

- The time series are nearly-Gaussian but not exactly Gaussian (skewness = **0.23**; kurtosis = **3.08**). This divergence of fully developed turbulent processes from normality has been also derived theoretically (Wilczek et al., 2011).
- Interestingly, these small differences from normality result in highly non-normal distribution of the white noise of the SMA model (skewness = **3.26**; kurtosis = **12.30**!)

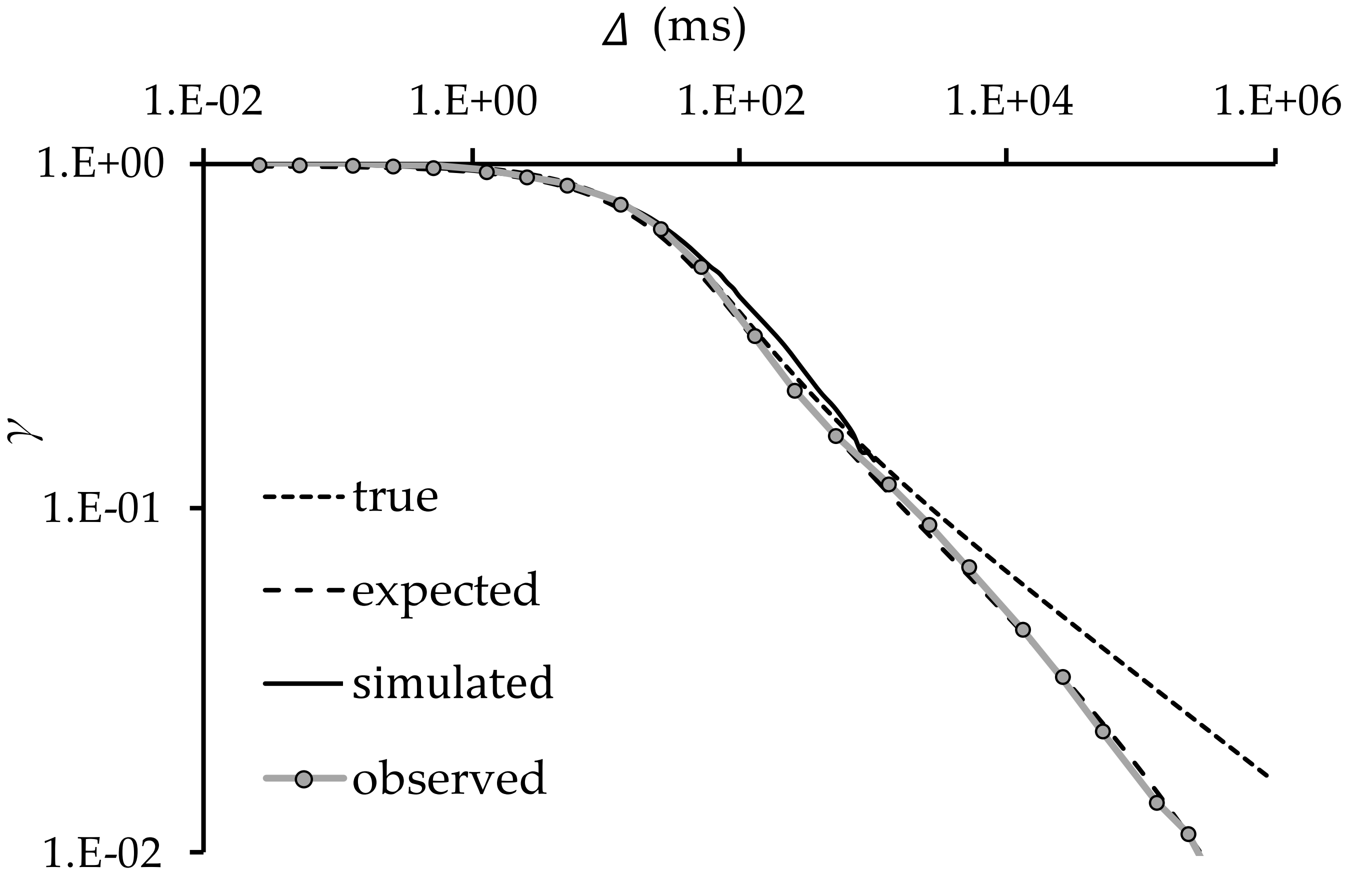


Probability density function of the mean standardized time series of turbulent velocity compared to that of a single simulation using the SMA scheme preserving the first four moments; the standard normal distribution N(0,1) is also shown.

Stochastic dependence of the turbulent velocity process

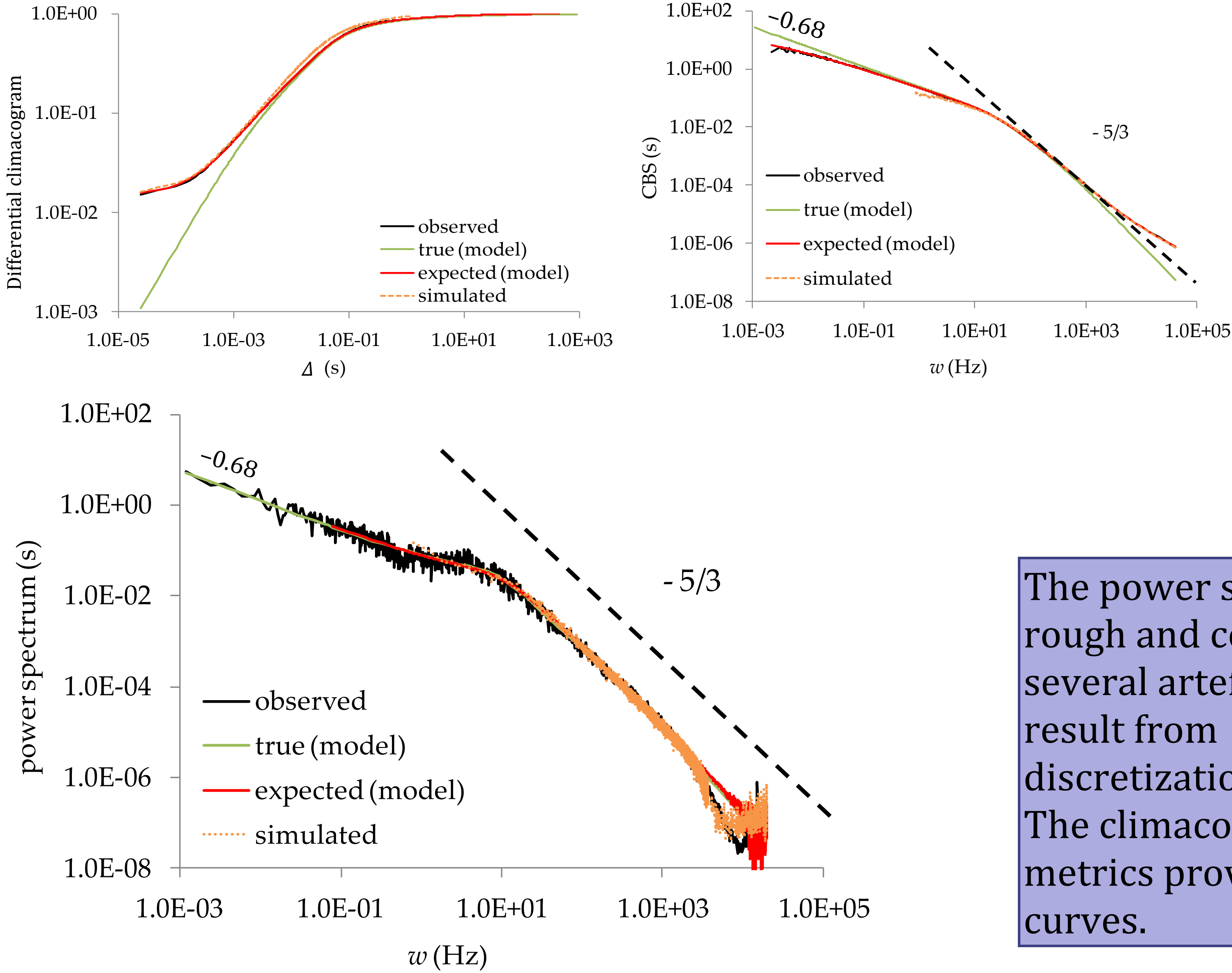
Sum of two equally weighted processes, an HHK and a Markov:

$$\gamma(\Delta) = \frac{\lambda}{2}\left(1 + (\Delta/\alpha)^{2\kappa}\right)^{\frac{H-1}{\kappa}} + \frac{\lambda}{\Delta/\alpha}\left(1 - \frac{1 - \mathrm{e}^{-\Delta/\alpha}}{\Delta/\alpha}\right) \quad (26)$$



Climacogram of the turbulent velocity process (observed is the average from the 40 time series); the five parameters of the model are estimated as: $\lambda = 1.016$, $\alpha = 14$ ms, $\kappa = 0.375$, $H = 0.84$.

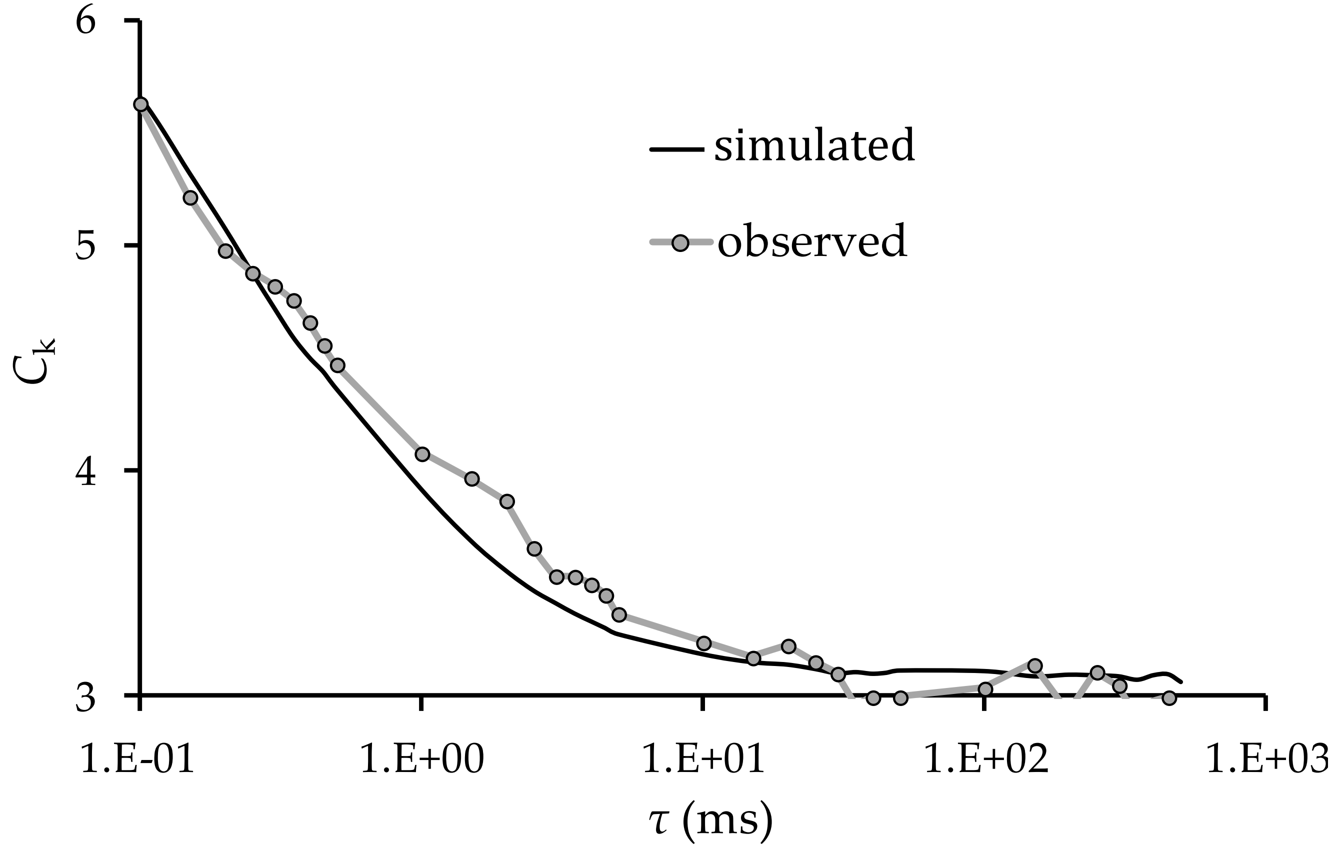
Other second-order properties of the model compared to data



The power spectrum is rough and contains several artefacts that result from discretization and bias. The climacogram based metrics provide better curves.

Kurtosis of velocity increments

The change of kurtosis of the velocity increments (differences) with increased time distance, τ (lag), is related to the intermittent behaviour of turbulence (Batchelor and Townsend, 1949). Therefore it is important to preserve this variation.

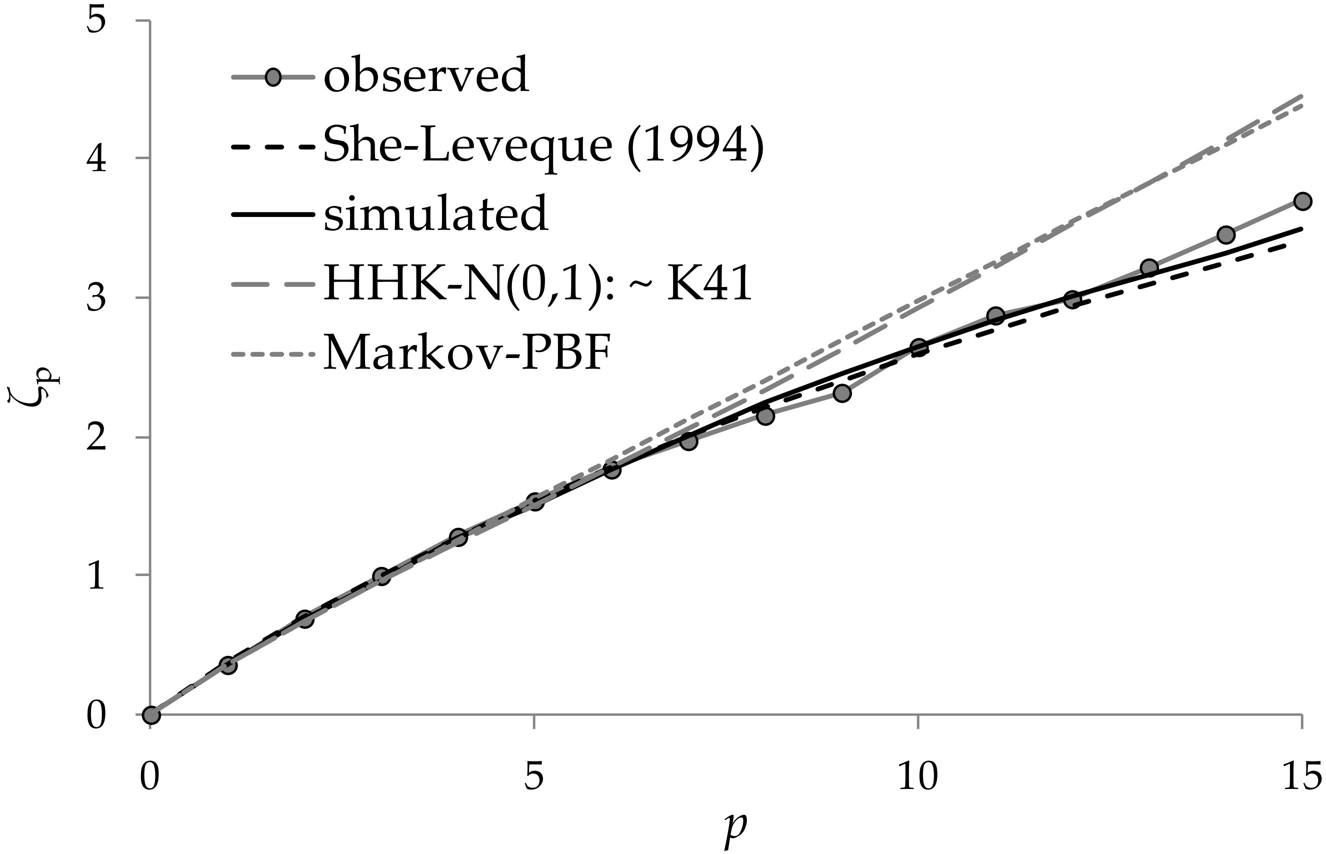


Empirical and simulated kurtosis vs. lag.

Not a mystery to have **large kurtosis** (**> 5** here) in velocity increments, even though the velocity is almost normal (**kurtosis = 3.08**)

Even higher moments

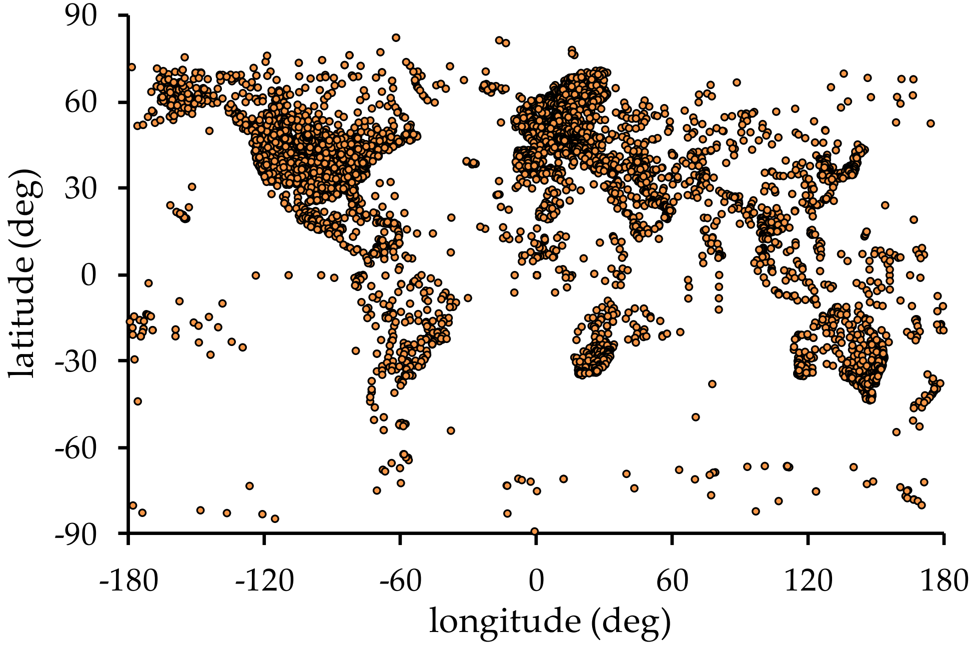
$S_p := \mathrm{E}[|\underline{x}_i - \underline{x}_{i+r}|^p] \approx r^{\zeta_p}$



Not a mystery that empirical values depart from the straight line (regarded as manifestation of intermittency). No provision is needed to reproduce it; a good parsimonious model (with Hurst behaviour and slightly non-Gaussian distribution) suffices.

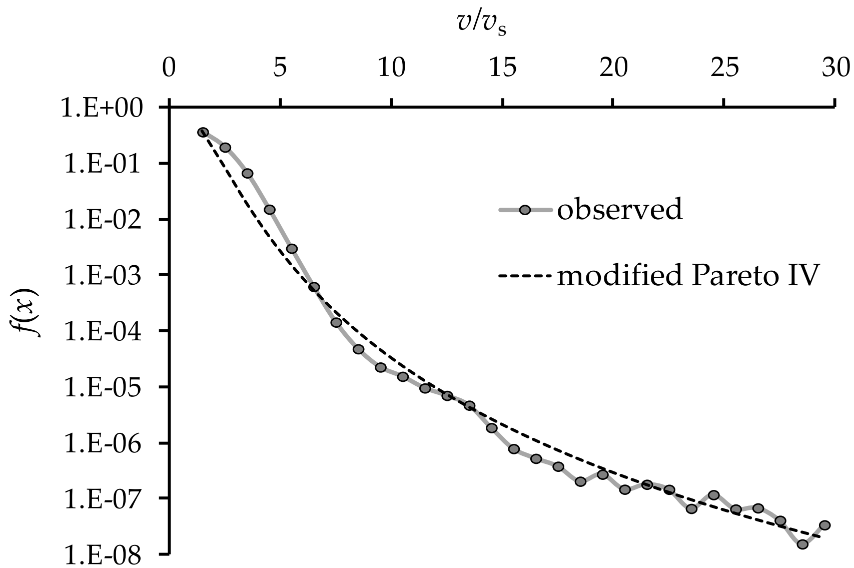
Application 2: Medium scale (wind)

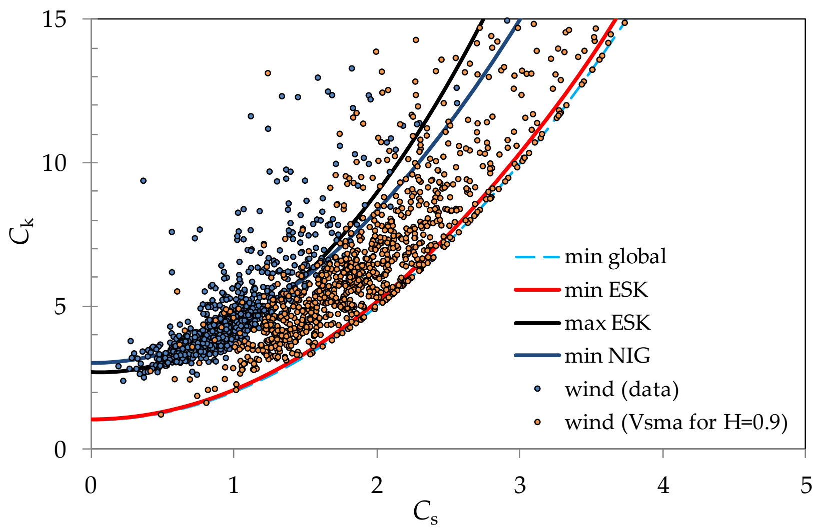
- We can estimate high moments in geophysical processes accurately only after analyzing thousands of short time series.
- Here we use hourly wind speed data by NOAA (www.ncdc.noaa.gov).
- This dataset consists of 15 000 time series around the globe with 10 min average measurements every one hour. After several quality and quantity tests we ended up with approximately 3500 stations.
- By standardizing all series we formed a sample of ~10^9 values to estimate the marginal distribution, and an ensemble of 3500 series, each with 3×10^5 values on the average, to estimate the dependence structure through the climacogram.



- We also performed multiple simulations using the SMA framework with n =3 ×10^5 values.

Marginal distribution





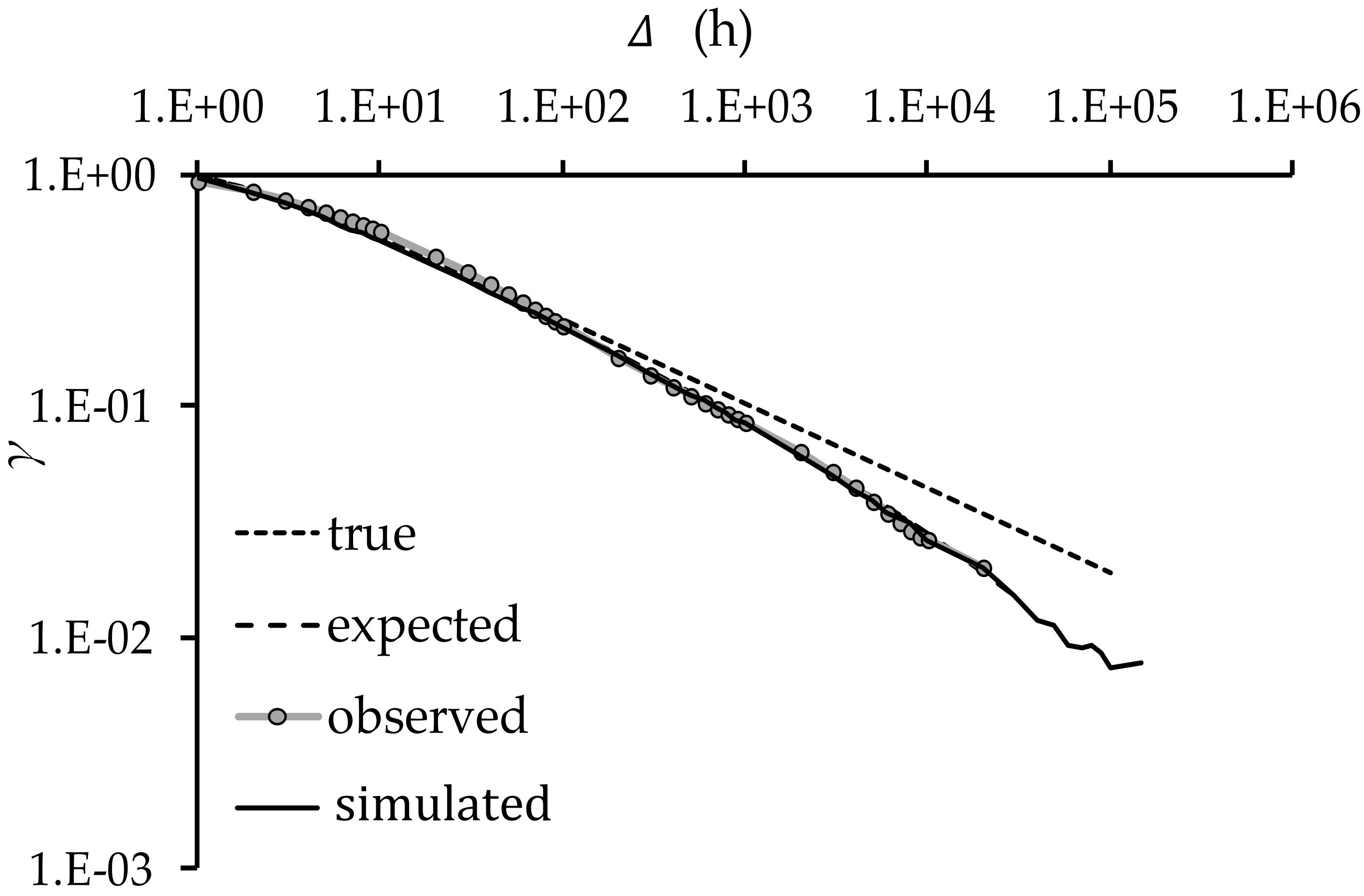
Wind speed distribution (from ~10^9 values):

$$F(v) = 1 - (1 + (v/\alpha v_s)^2)^{-\beta/2} \quad (27)$$

where $\alpha = 2$ and $\beta = 3$.

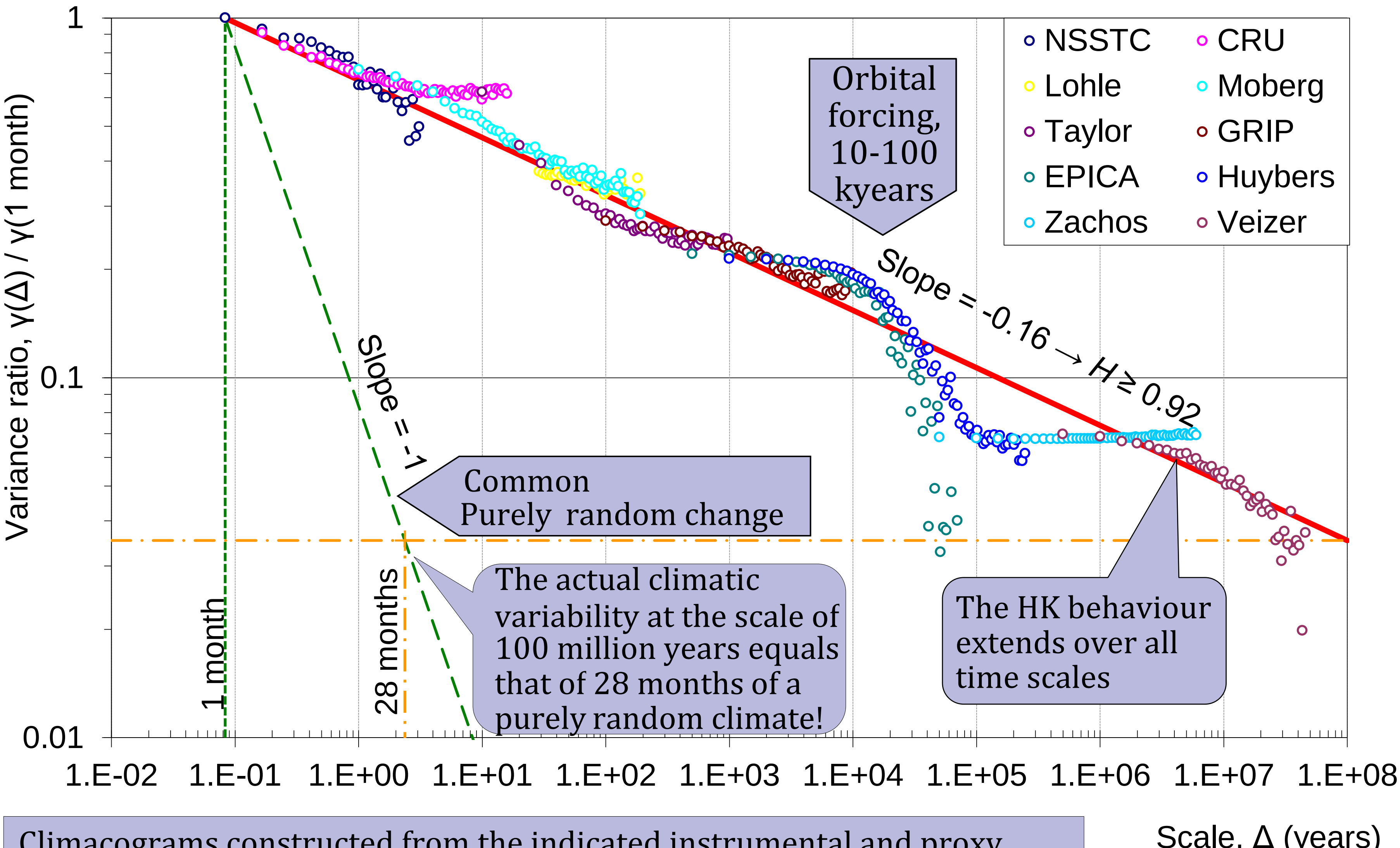
Sample skewness and kurtosis coefficients of 1000 hourly wind stations as well as of the corresponding white noise process of the SMA model.

Stochastic dependence of the wind process



Climacogram of the wind speed process (observed is the average from the 3500 time series); the four parameters of the model are estimated as: α = 1 h, κ = 0.5, λ = 1.3 and H = 0.82.

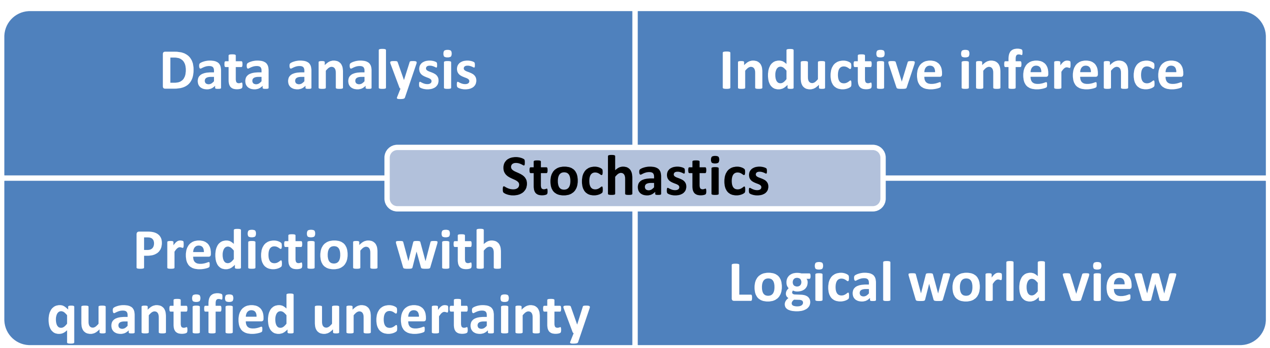
Application 3: Megascale (temperature)



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

Epilogue

- Stochastic processes in continuous time offer a strong basis for modelling and interpretation of natural behaviours.
- We owe the well-founded and rigorous mathematical theory of stochastics to Kolmogorov (1931, 1933, 1938), including the foundation of scaling processes (Kolmogorov, 1940). This theory has often be distorted but there exist textbooks consistent with it (e.g. Papoulis, 1991).
- Calculating values of sample statistics without considering their statistical properties (bias and statistical variation) can yield misleading results.
- Without proper attention to the underlying stochastics, we can even “identify” phenomena that do not exist and take statistical sampling effects as natural behaviours.
- A general methodology for data analysis and construction of synthetic time series is possible provided that we have a good understanding of stochastics.



Appendix: Proof of infeasibility of too steep slopes in power spectrum for low frequencies

- This proof is summarized here from Koutsoyiannis (2013b) and Koutsoyiannis et al. (2013).
- Let us assume the contrary, i.e., that for frequency range $0 \leq w \leq \varepsilon$ (with ε however small) the log-log derivative is $s^{\#}(w) = -\beta$, or else $s(w) = \alpha\, w^{-\beta}$ where α and β are constants, with $\beta > 1$.
- As a result of (2) and (4) the climacogram is related to power spectrum by:

$$\gamma(\Delta) = \int_0^\infty s(w)\ \mathrm{sinc}^2(\pi w \Delta)\,\mathrm{d}w \qquad (28)$$

- The sinc^2 function within the integral takes significant values only for $w < 1/\Delta$ (cf. Papoulis, 1991, p. 433). Hence, assuming a scale $\Delta \gg 1/\varepsilon$, and with reference to (28) we may write:

$$\gamma(\Delta) = \int_0^\infty s(w)\,\mathrm{sinc}^2(\pi w \Delta)\,\mathrm{d}w \approx \int_0^\varepsilon \alpha w^{-\beta}\ \mathrm{sinc}^2(\pi w \Delta)\,\mathrm{d}w \qquad (29)$$

- On the other hand, it is easy to verify that, for $0 < w < 1/\Delta$,

$$\mathrm{sinc}(\pi w \Delta) \geq 1 - w\Delta \geq 0 \qquad (30)$$

- Since $\varepsilon \gg 1/\Delta$, while the function in the integral (29) is nonnegative,

$$\gamma(\Delta) \approx \int_0^\varepsilon \alpha w^{-\beta}\ \mathrm{sinc}^2(\pi w \Delta)\,\mathrm{d}w \geq \int_0^{1/\Delta} \alpha w^{-\beta}\ \mathrm{sinc}^2(\pi w \Delta)\,\mathrm{d}w \geq \int_0^{1/\Delta} \alpha w^{-\beta}(1 - w\Delta)^2\,\mathrm{d}w \qquad (31)$$

- Substituing $\xi = w\Delta$ in (31), we find:

$$\gamma(\Delta) \geq a\Delta^{\beta-1} \int_0^1 \xi^{-\beta}(1-\xi)^2\,\mathrm{d}\xi \qquad (32)$$

Appendix (contd.)

- To evaluate the integral in (32) we take the limit for $q \to 0$ of the integral:

$$B(q) \coloneqq \int_q^1 \xi^{-\beta} (1-\xi)^2 \, \mathrm{d}\xi \; = \frac{q^{1-\beta}-1}{\beta-1} - 2\,\frac{q^{2-\beta}-1}{\beta-2} + \frac{q^{3-\beta}-1}{\beta-3} \tag{33}$$

- Clearly, for $\beta > 1$ the first term of the latter integral diverges for $q \to 0$, i.e., $B(0) = \infty$ and thus, by virtue of the inequality (32), $\gamma(\Delta) = \infty$. For a (mean) ergodic processes $\gamma(\Delta)$ should necessary tend to 0 for $\Delta \to \infty$ (Papoulis, 1991, p. 429). Therefore, the process is non-ergodic.† This analysis generalizes a result by Papoulis (1991, p. 434) who shows that an impulse at $w = 0$ corresponds to a non-ergodic process.
- In a non-ergodic process there is no possibility to infer statistical properties from the samples, so the statistical analyses are in vain and hence the reported results not meaningful.
- Sometimes reported slopes $s^{\#} < -1$ are interpreted as indications of nonstationarity. Such interpretations are equally invalid because even the definition of the power spectrum as a function of frequency only (as well as those of autocorrelation and climacogram as functions of lag and scale, respectively) assumes stationarity.

† It is interesting to note that, if $|\beta| < 1$, the integral in (29) can be evaluated to give:

$$\gamma(\Delta) \approx \alpha \int_0^\infty w^{-\beta} \; \mathrm{sinc}^2(\pi w \Delta) \, \mathrm{d}w = \frac{\sin(\pi\beta/2)}{(\pi\beta/2)} \, \frac{(2\pi)^\beta \alpha \Gamma(1-\beta)}{2(\beta+1)\Delta^{1-\beta}}$$

Clearly, for $\Delta \to \infty$, the last expression gives $\gamma(\Delta) \to 0$ and thus for $|\beta| < 1$ the process is mean ergodic.

References

Batchelor, G. K. and Townsend, A. A. (1949), The nature of turbulent motion at large wave-numbers, *Proc. R. Soc. Lond. A* ,199, 238-255.

Beran, J., Feng, Y., Ghosh, S. and Kulik, R. (2013), *Long-Memory Processes, Long-Memory Processes*, Springer.

Dechant, A. and Lutz, E. (2015), Wiener-Khinchin theorem for nonstationary scale-invariant processes, *Physical Review Letters*, 115(8), p.080603

Dimitriadis, P., and Koutsoyiannis, D. (2015), Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst–Kolmogorov processes, *Stochastic Environm. Research & Risk Assessment*, 29 (6), 1649–1669, doi:10.1007/s00477-015-1023-7.

Dimitriadis, P., and Koutsoyiannis, D. (2016), Stochastic synthesis approximating any process dependence and distribution (in preparation).

Gneiting, T., and Schlather, M. (2004), Stochastic models that separate fractal dimension and the Hurst effect, *Society for Industrial and Applied Mathematics Review*, 46 (2), 269-282.

Graham, L. and Kantor, J.-M. (2009), *Naming Infinity: A True Story of Religious Mysticism and Mathematical Creativity*, Harvard University Press.

Kang, H. S., Chester, S., and Meneveau, C. (2003), Decaying turbulence in an active-grid-generated flow and comparisons with large-eddy simulation, *Journal of Fluid Mechanics*, 480, 129–160.

Hemelrijk, J., Underlining random variables, *Statistica Neerlandica*, 20 (1), 1-7, 1966.

Kolmogorov, A. N. (1931), Uber die analytischen Methoden in der Wahrscheinlichkcitsrechnung, Math. Ann. 104, 415-458. (English translation: On analytical methods in probability theory, In: Kolmogorov, A.N., 1992. Selected Works of A. N. Kolmogorov - Volume 2, Probability Theory and Mathematical Statistics A. N. Shiryayev, ed., Kluwer, Dordrecht, The Netherlands, pp. 62-108).

Kolmogorov, A. N. (1933), *Grundbegrijfe der Wahrscheinlichkeitsrechnung*, Ergebnisseder Math. (2), Berlin. (2nd English Edition: Foundations of the Theory of Probability, 84 pp. Chelsea Publishing Company, New York, 1956).

Kolmogorov, A. N. (1938), A simplified proof of the Birkhoff-Khinchin ergodic theorem, *Uspekhi Matematicheskikh Nauk*, 5, 52–56. (English edition: Kolmogorov, A.N., 1991, Selected Works of A. N. Kolmogorov - Volume 1, Mathematics and Mechanics, Tikhomirov, V. M. ed., Kluwer, Dordrecht, The Netherlands, pp. 271–276).

Kolmogorov, A. N. (1940), Wiener spirals and some other interesting curves in a Hilbert space. *Dokl. Akad. Nauk SSSR*, 26, 115–118. English translation in: V.M. Tikhomirov, ed., 1991, Selected works of A.N. Kolmogorov, Volume I: Mathematics and mechanics, 324–326. Springer, Berlin.

Koutsoyiannis, D. (2000), A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, *Water Resources Research*, 36 (6), 1519–1533.

Koutsoyiannis, D. (2010), Some problems in inference from time series of geophysical processes (solicited), *European Geosciences Union General Assembly 2010, Geophysical Research Abstracts,* Vol. 12, Vienna, EGU2010-14229, European Geosciences Union (http://www.itia.ntua.gr/973/).

Koutsoyiannis, D. (2011), Hurst-Kolmogorov dynamics as a result of extremal entropy production, *Physica A*, 390 (8), 1424–1432.

References (contd.)

Koutsoyiannis, D. (2013a), Climacogram-based pseudospectrum: a simple tool to assess scaling properties, *European Geosciences Union General Assembly 2013, Geophysical Research Abstracts, Vol. 15*, Vienna, EGU2013-4209, European Geosciences Union, (http://itia.ntua.gr/1328).

Koutsoyiannis, D. (2013b), *Encolpion of Stochastics: Fundamentals of Stochastic Processes*, Athens: Department of Water Resources and Environmental Engineering – National Technical University of Athens (http://www.itia.ntua.gr/1317/).

Koutsoyiannis, D. (2014), Random musings on stochastics (Lorenz Lecture), AGU 2014 Fall Meeting, San Francisco, USA, American Geophysical Union, doi:10.13140/RG.2.1.2852.8804 (http://www.itia.ntua.gr/en/docinfo/1500/).

Koutsoyiannis, D. (2016), Generic and parsimonious stochastic modelling for hydrology and beyond, *Hydrological Sciences Journal*, 61 (2), 225–244, doi:10.1080/02626667.2015.1016950.

Koutsoyiannis, D., Lombardo, F., Volpi, E., and Papalexiou S. M. (2013), Is consistency a limitation? — Reply to "Further (monofractal) limitations of climactograms" by Lovejoy et al., Comment in the review of "Just two moments! A cautionary note against use of high-order moments in multifractal models in hydrology" by Lombardo et al., *Hydrol. Earth Syst. Sci. Discuss.*, 10, C5397, (http://www.hydrol-earth-syst-sci-discuss.net/10/C5397/2013/hessd-10-C5397-2013-supplement.pdf).

Lombardo, F., Volpi, E., Koutsoyiannis, D., and Papalexiou, S. M. (2014), Just two moments! A cautionary note against use of high-order moments in multifractal models in hydrology, *Hydrology and Earth System Sciences*, 18, 243–255.

Mandelbrot, B. B. (1982), *The Fractal Geometry of Nature*, W. H. Freeman, New York.

Mandelbrot, B. B. (1999), *Multifractals and 1/f Noise: Wild Self-Affinity in Physics (1963-1976)*, Springer.

Markonis, Y., and Koutsoyiannis, D. (2013), Climatic variability over time scales spanning nine orders of magnitude: Connecting Milankovitch cycles with Hurst–Kolmogorov dynamics, *Surveys in Geophysics*, 34(2), 181–207.

Papalexiou, S. M., Koutsoyiannis, D., and Montanari, A. (2010), Mind the bias!, *STAHY Official Workshop: Advances in statistical hydrology*, Taormina, Italy, International Association of Hydrological Sciences.

Papoulis, A. (1991), *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill.

Yaglom, A. M. (1987), *Correlation Theory of Stationary and Related Random Functions*, Springer.

Veneziano, D. and Langousis, A. (2010), Scaling and fractals in hydrology, *Advances in Data-based Approaches for Hydrologic Modeling and Forecasting*, edited by B. Sivakumar and R. Berndtsson, 145 pages, World Scientific.

von Kármán, T. (1940), The engineer grapples with nonlinear problems, *Bull. Amer. Math. Soc.*, 46, 615-683, doi: 10.1090/S0002-9904-1940-07266-0.

Wackernagel, H. (1995), *Multivariate Geostatistics*, Springer, Berlin.

Wackernagel, H. (1998), *Multivariate Geostatistics*, 2nd completely revised edition, Springer, Berlin.

Wilczek, M., Daitche, A., and Friedrich, R. (2011), On the velocity distribution in homogeneous isotropic turbulence: correlations and deviations from Gaussianity, Journal of Fluid Mechanics, 676, 191-217.