CONSTANȚA,
JUNE 21-23 2018

Hydrofractals '18

Ovidius University of Constanta,
Constanta, Romania

Unknowable and knowable moments: Are they relevant to hydrofractals?

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

Presentation available online: http://www.itia.ntua.gr/1846/

The general framework: Seeking theoretical consistency in analysis of geophysical data (Using *stochastics*)

R^G Home Questions Jobs Search

Project

Seeking Theoretical Consistency in Analysis of Geophysical Data (Using Stochastics)

Demetris Koutsoyiannis · Panayiotis Dimitriadis · ... · Show all 4 collaborators

Goal: Analysis of geophysical data is (explicitly or implicitly) based on stochastics, i.e. the mathematics of random variables and stochastic processes. These are abstract mathematical objects, whose properties distinguish them from typical variables that take on numerical values. It is important to understand these properties before making calculations with data, otherwise the results may be meaningless (not even wrong).

Lab: Laboratory of Hydrology and Water Resources Development

Why stochastics in geophysics and hydrofractals?

- **Geophysics** is the branch of physics that **relies most decisively on data**.
- **Geophysical data are numbers** but to treat them we need to use **stochastics, not arithmetic.**
- **Stochastics is the mathematics of random variables and stochastic processes.**
- These are abstract mathematical objects, whose properties distinguish them from typical variables that take on numerical values.
- It is important to understand these properties before making calculations with data, otherwise the results may be meaningless (not even wrong).
- The numerical data allow us to **estimate** (not to determine precisely) **expectations**.
- Expectations are defined as **integrals** of products of functions. For a continuous random variable $\underline{x}$ with probability density function $f(x)$, the expectation of an arbitrary function g of $\underline{x}$ (where $g(\underline{x})$ is a random variable per se), the **expectation** of $g(\underline{x})$ is defined as $\theta \coloneqq \mathrm{E}\left[g\left(\underline{x}\right)\right] \coloneqq \int_{-\infty}^{\infty} g(x) f(x) \mathrm{d}x$.
- Central among expectations are the **moments**, in which $g(\underline{x})$ is a power of $\underline{x}$ (or a linear expression of $\underline{x}$).
- To estimate true parameters θ from data we need estimators; the **estimator** $\underline{\hat{\theta}}$ of θ is a random variable depending on the stochastic process of interest $\underline{x}(t)$ and is a model per se, not a number.
- The **estimate** $\hat{\theta}$ is a number, calculated by using the observations and the estimator.
- Characteristic statistics of the estimator $\underline{\hat{\theta}}$ are its **bias**, $\mathrm{E}\left[\underline{\hat{\theta}}\right] - \theta$, and its **variance** $\mathrm{var}\left[\underline{\hat{\theta}}\right]$. When $\mathrm{E}\left[\underline{\hat{\theta}}\right] = \theta$ the estimator is called unbiased.
- Estimation is made possible thanks to two concepts of stochastics: **stationarity** and **ergodicity**.

Memorable moments in the history of stochastics

Ludwig Boltzmann

(1844 –1906, Universities of Graz and Vienna, Austria, and Munich, Germany)

1877 Explanation of the concept of **entropy** in probability theoretic context.

1884/85 Introduction of the notion of **ergodic*** systems which however he called "isodic"

* The term is etymologized from Greek words but which ones exactly is uncertain (options: (a) έργον + οδός; (b) έργον + είδος; (c) εργώδης; see Mathieu, 1988).

George D. Birkhoff

(1884 – 1944; Princeton, Harvard, USA)

1931 Discovery of the **ergodic (Birkhoff–Khinchin) theorem**

Aleksandr Khinchin

(1894 – 1959; Moscow State University, Russia)

1933 Purely measure-theoretic proof of the **ergodic (Birkhoff–Khinchin) theorem**

1934 Definition of **stationary stochastic processes** and probabilistic setting of the **Wiener-Khinchin theorem** relating autocovariance and power spectrum

Andrey N. Kolmogorov

(1903 – 1987; Moscow State University, Russia)

1931 Introduction of the term **stationary** to describe a probability density function that is unchanged in time

1933 Definition of the concepts of **probability** & **random variable**

1937-1938 Probabilistic exposition of the **ergodic (Birkhoff–Khinchin) theorem** and **stationarity**

1947 Definition of **wide sense stationarity**

Stationarity and nonstationarity

Central to the notion of a stochastic process are the concepts of **stationarity** and **nonstationarity**, two widely misunderstood and broadly misused concepts (Koutsoyiannis and Montanari, 2015); their definitions apply only to stochastic processes (e.g., time series cannot be stationary, nor nonstationary).

Reminder of definitions

Following Kolmogorov (1931, 1938) and Khinchin (1934), a process is **(strict-sense) stationary** if its **statistical properties are invariant to a shift of time origin**, i.e. the processes $\underline{x}(t)$ and $\underline{x}(t')$ have the same statistics for any t and t' (see further details in Papoulis, 1991; see also further explanations in Koutsoyiannis, 2006, 2011b and Koutsoyiannis and Montanari, 2015).

Following Kolmogorov (1947), a stochastic process is **wide-sense stationary** if its mean is constant and its autocovariance depends on time difference only, i.e.:

$$\mathrm{E}[\underline{x}(t)] = \mu = \text{constant}, \quad \mathrm{E}[(\underline{x}(t) - \mu)\,(\underline{x}(t+\tau) - \mu)] = c(\tau) \qquad (1)$$

Conversely, a process is **nonstationary** if some of its statistics are changing through time and **their change is described as a deterministic function of time**.

Ergodicity

Stationarity is also related to **ergodicity**, which in turn is a prerequisite to make inference from data, that is, induction.

By definition (e.g., Mackey, 1992, p. 48; Lasota and Mackey, 1994, p. 59), a **transformation of a dynamical system is ergodic if all its invariant sets are trivial (have zero probability)**. 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).

The **ergodic theorem** (Birkhoff, 1931; Khinchin, 1933; see also Mackey, 1992, p. 54) allows **redefining ergodicity within the stochastic processes domain** (Papoulis 1991 p. 427; Koutsoyiannis 2010) in the following manner: A stochastic process $\underline{x}(t)$ is ergodic if the **time average** of any (integrable) function $g(\underline{x}(t))$, **as time tends to infinity, equals the true** (ensemble) **expectation** $\mathrm{E}[g(\underline{x}(t))]$, i.e., $\lim_{T\to\infty} \frac{1}{T}\int_0^T g\left(\underline{x}(t)\right) dt = \mathrm{E}[g(\underline{x}(t))]$.

If the system that is modelled in a stochastic framework has **deterministic dynamics** (meaning that a system input will give a single system response, as happens for example in most hydrological models) then a theorem applies (Mackey 1992, p. 52), according to which a dynamical system has a stationary probability density *if and only if* it is ergodic. Therefore, **a stationary system is also ergodic and vice versa**, and **a nonstationary system is also non-ergodic and vice versa**.

If the system **dynamics is stochastic** (a single input could result in multiple outputs), then **ergodicity and stationarity do not necessarily coincide**. However, recalling that a stochastic process is a model and not part of the real world, **we can always conveniently device a stochastic process that is ergodic** (see example in Koutsoyiannis and Montanari, 2015).

Both stationarity and ergodicity are immortal!*

From a practical point of view ergodicity can always be assumed when there is stationarity. Without stationarity and ergodicity inference from data would not be possible. Ironically, several studies use time series data to estimate statistical properties, as if the process were ergodic, while at the same time what they (cursorily) estimate may falsify the ergodicity hypothesis.

Misuse example 1: By analysing the time series x_τ (where τ denotes time), I concluded that it is nonstationary and I identified an increasing trend with slope b.

Corrected example 1: I analysed the time series x_i based on the hypothesis that the stochastic process $\underline{x}_\tau - b\tau$ is stationary and ergodic, which enabled the estimation of the slope b.

Misuse example 2: From the time series x_τ, I calculated the power spectrum and found that its slope for low frequencies is steeper than –1, which means that the process is nonstationary.

Possible correction (a) of example 2: I cursorily interpreted a slope steeper than –1 in the power spectrum as if indicated nonstationary, while a simple explanation would be that the frequencies on which my data enable calculation of the power spectrum values are too high.

Possible correction (b) of example 2: I cursorily applied the concept of the power spectrum of a stationary stochastic process, forgetting that the empirical power spectrum of a stationary stochastic process is a (nonstationary) stochastic process per se. The high variability of the latter (or the inconsistent numerical algorithm I used) resulted in a slope for low frequencies steeper than –1, which is absurd. Such a slope would suggest a non-ergodic process while my calculations were based on the hypothesis of a stationary and ergodic process.

Possible correction (c) of example 2: I cursorily applied the concept of the power spectrum of a stationary stochastic process using a time series which is realization of a nonstationary stochastic process and I found an inconsistent result; therefore, I will repeat the calculations recognizing that the power spectrum of a nonstationary stochastic process is a function of two variables, frequency and “absolute” time.

*Montanari and Koutsoyiannis (2014)

Moments of what order?

The classical definitions of raw and central moments of order p are:

$$\mu'_p \coloneqq \mathrm{E}\left[\underline{x}^p\right], \qquad \mu_p \coloneqq \mathrm{E}\left[\left(\underline{x} - \mu\right)^p\right] \tag{2}$$

respectively, where $\mu \coloneqq \mu'_1 = \mathrm{E}\left[\underline{x}\right]$ is the mean of the random variable $\underline{x}$. Their standard estimators from a sample $\underline{x}_i$, $i = 1, \ldots, n$, are

$$\underline{\hat{\mu}}'_p = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i^p\,, \; \underline{\hat{\mu}}_p = \frac{b(n,p)}{n}\sum_{i=1}^{n}\left(\underline{x}_i - \hat{\mu}\right)^p \tag{3}$$

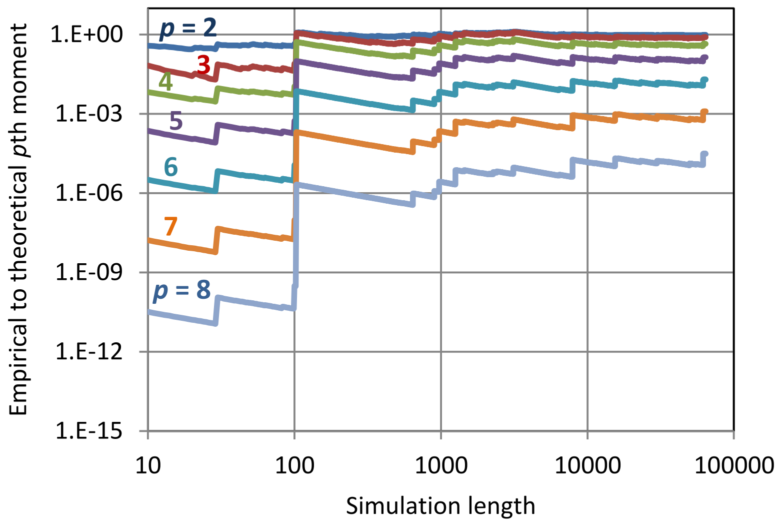
where $a(n, p)$ is a bias correction factor (e.g. for the variance $\mu_2 =: \sigma^2$, $b(n, 2) = n/(n - 1)$).

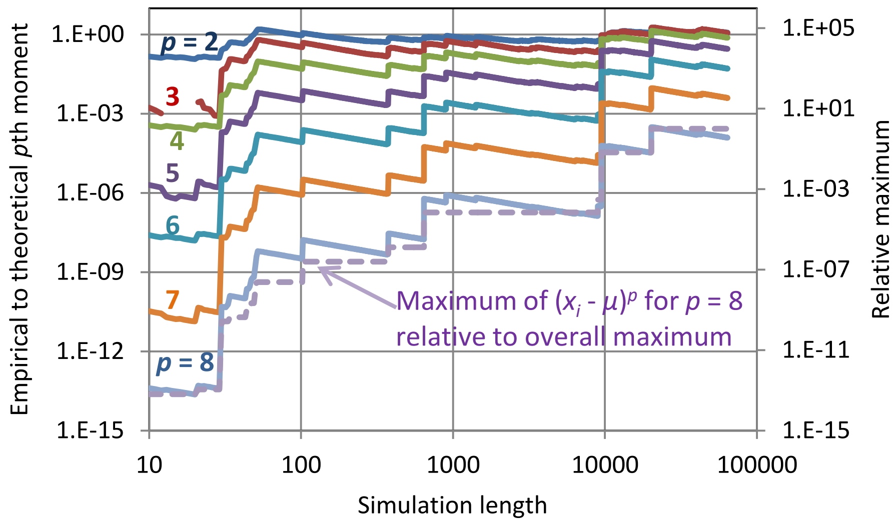
The estimators of the raw moments $\underline{\hat{\mu}}'_p$ (as well as central moments μ_p if μ is a priori known) are in theory unbiased, but it is practically impossible to use them in estimation for $p > 2$.

cf. Lombardo et al. (2014), "**Just two moments**".

The ergodic theorem enables, in theory, estimation of moments from data as $n \to \infty$, but what happens for finite n?

Moments of what order?: Illustration of slow convergence





Empirical central moments of order p from a single simulation of non-Gaussian white noise with lognormal distribution LN(0,1), standardized by the corresponding theoretical moment, versus the simulation length.

The limit as the simulation length → ∞ is 1, but for relatively high p it deviates from the limit by many orders of magnitude and the convergence is slow.

Same as in the left panel but from a single simulation of an exponentiated Hurst-Kolmogorov process with Hurst parameter H = 0.9; the marginal distribution is again with lognormal LN(0,1).

The deviation of the empirical curves from the limit (i.e. 1) is even greater than in the white noise.

The shape of the curves of empirical moments is similar to that of the maximum value as a function of simulation length.

The Hurst-Kolmogorov process and the Hurst parameter will be explained soon.

Moments of what order?: The reason of slow convergence

What is the result of **raising to a power and adding**, i.e. $\sum_{i=1}^{n} x_i^p$ – like in estimating moments?

Linear, p = 1	Pythagorean, p = 2	Cubic, p = 3	High order, p = 8
$3 + 4 = 7$	$3^2 + 4^2 = 5^2$	$3^3 + 4^3 = 4.5^3$	$3^8 + 4^8 \approx 4^8$
$3 + 4 + 12 = 19$	$3^2 + 4^2 + 12^2 = 13^2$	$3^3 + 4^3 + 12^3 = 12.2^3$	$3^8 + 4^8 + 12^8 \approx 12^8$

Symbolically, for relatively large p the estimate of μ'_p is*:

$$\hat{\mu}'_p = \frac{1}{n}\sum_{i=1}^{n} x_i^p \approx \frac{1}{n}\left(\max_{1\le i\le n}(x_i)\right)^p \tag{4}$$

Thus, for an unbounded variable $\underline{x}$ and for large p, we can conclude that $\underline{\hat{\mu}}'_p$ **is not an estimator of μ'_p** but one of an extreme quantity, i.e., the nth order statistic raised to power p.

Thus, unless p is very small, **μ'_p is not a *knowable* quantity**: we cannot infer its value from a sample. **This is the case even if n is very large**!

Also, the various $\underline{\hat{\mu}}'_p$ are not independent to each other as they only differ on the power to which the maximum value is raised.

---

* This is precise if x_i are positive; see also p. 27. Note that for large p the term $(1/n)$ in the rightmost part of the equation could be omitted with a negligible error.

Moments of how many random variables?

In stochastic processes we are interested not only about marginal moments but also joint ones.

However, the estimation problems become more complicated in joint moments.

Ideally, moments of a **single variable are preferable**, but how can we describe **dependence** based on one variable?

Autocovariance and its equivalent standardized form, i.e., **autocorrelation**, have been the most customary tools to characterize dependence, but **can they be described using a single variable**?

The **variance of the process averaged at specified time scale** k provides a mathematically equivalent but statistically more advantageous means to this aim.

Consider the second-order dependence of any two random variables $\underline{x}_1$ and $\underline{x}_2$ with means μ_i and standard deviations σ_i, i = 1, 2. The variance of the average of the two variables contains the same information as the covariance thereof. We note, though, that if the variables denote different physical quantities, it is necessary to make them compatible before taking the average, which can be made by standardizing with their standard deviations. In other words, we define:

$$\rho_{12} \coloneqq \mathrm{var}\left[\frac{1}{2}\left(\frac{\underline{x}_1}{\sigma_1} + \frac{\underline{x}_2}{\sigma_2}\right)\right] \tag{5}$$

which is related to the classical (Pearson) correlation coefficient by:

$$\rho_{12} = \frac{1}{2} + \frac{1}{2}\mathrm{cov}\left[\frac{\underline{x}_1}{\sigma_1}, \frac{\underline{x}_2}{\sigma_2}\right] = \frac{1}{2} + \frac{1}{2}r_{12}, \qquad r_{12} \coloneqq \frac{\mathrm{Cov}\left[\underline{x}_1, \underline{x}_2\right]}{\sigma_1\sigma_2} = \mathrm{cov}\left[\frac{\underline{x}_1}{\sigma_1}, \frac{\underline{x}_2}{\sigma_2}\right] \tag{6}$$

where $-1 \le r_{12} \le 1$, while $0 \le \rho_{12} \le 1$.

The notion of the climacogram

Unlike r_{12}, the notion of ρ_{12} could be readily expanded to many variables. Assuming that the variables $\underline{x}_1, \dots, \underline{x}_\kappa$ are identically distributed (and thus have common variance σ^2, so that standardization is no longer needed before taking the variance), we define the so-called climacogram:

$$\gamma_\kappa \coloneqq \mathrm{var}\left[\underline{X}_\kappa/\kappa\right], \qquad \underline{X}_\kappa \coloneqq \underline{x}_1 + \cdots + \underline{x}_\kappa \tag{7}$$

so that $\underline{X}_\kappa/\kappa$ is the average, satisfying $0 \le \gamma_\kappa \le \sigma^2$.

The climacogram is readily adapted to a continuous-time stochastic process $\underline{x}(t)$:

$$\gamma(k) \; \coloneqq \; \mathrm{var}[\underline{X}(k)/k], \qquad \underline{X}(k) \coloneqq \int_0^k \underline{x}(t)\mathrm{d}t \tag{8}$$

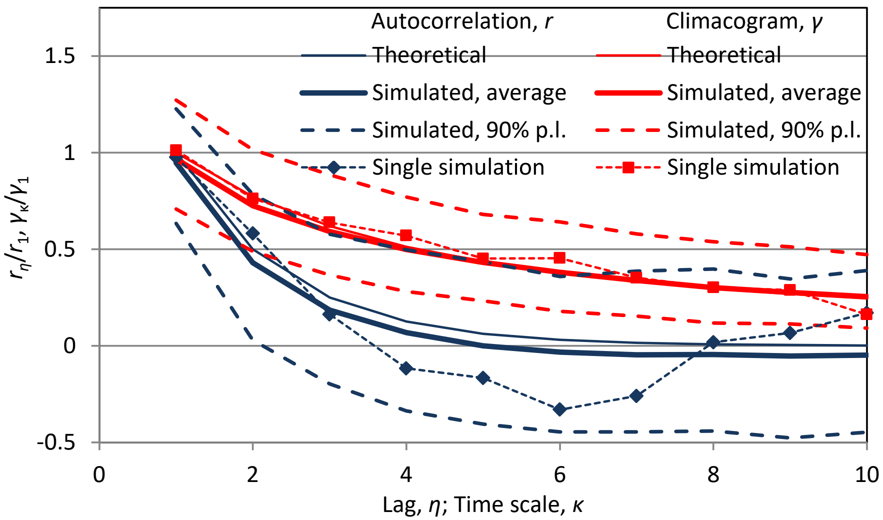
It can also be expanded to describe the dependence of different processes, replacing the concept of cross-covariance of two stationary processes $\underline{x}(t)$ and $\underline{y}(t)$ by the *cross-climacogram*:

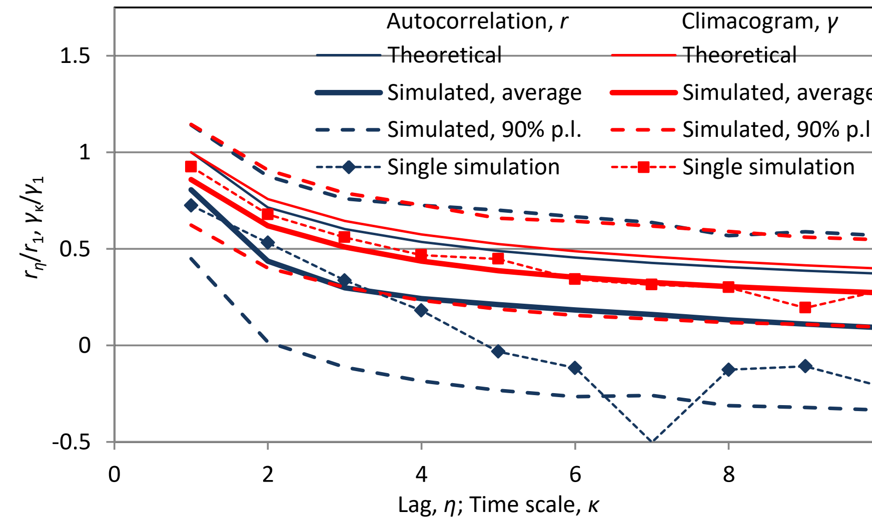
$$\gamma_{xy}^{\eta}(k) \coloneqq \sigma_x \sigma_y \, \mathrm{var}\left[\frac{\underline{X}(k)}{k\sigma_x} + \frac{\underline{Y}\big((\eta+1)k\big) - \underline{Y}(\eta k)}{k\sigma_y}\right], \qquad \underline{X}(k) \coloneqq \int_0^k \underline{x}(t)\mathrm{d}t\,, \underline{Y}(k) \coloneqq \int_0^k \underline{y}(t)\mathrm{d}t \tag{9}$$

where η is time lag.

Note: $\underline{x}(k)$ and $\underline{x}_\kappa$ are assumed to be stationary stochastic processes, which entails that the cumulative processes $\underline{X}(k)$ and $\underline{X}_\kappa$ are nonstationary.

Single-variable vs. multivariate characterization of dependence





Simulation results for autocovariance and climacogram (standardized by their value at lag 1 and scale 1, respectively) for sample size n = 100 for a Markov process with lag-one correlation r = 0.5

As in left panel but for a Hurst-Kolmogorov process with Hurst parameter H = 0.8

The sample estimates of autocovariance/autocorrelation have higher bias and higher variability than the climacogram estimates. Furthermore, the climacogram bias and uncertainty are easy to control as they can be calculated analytically (and be known a priori; see Koutsoyiannis, 2016).

In single simulations the sample estimates of autocovariance/autocorrelation has a rougher and more scattered shape than that of the climacogram estimates.

As will be seen soon, additional advantages of the climacogram are (a) its close relationship with entropy production and (b) its expandability to high-order moments.

Mathematical equivalence of climacogram, autocovariance and power spectrum

Related characteristics	Direct Relationship	Inverse relationship	Ref.
Climacogram $\gamma(k)$ ↔ Autocovariance $c(h)$	$\gamma(k) = 2\int_0^1 (1-\chi)c(\chi k)\mathrm{d}\chi$	$c(h) = \frac{1}{2}\frac{\mathrm{d}^2\left(h^2\gamma(h)\right)}{\mathrm{d}h^2}$	(10)
Power spectrum $s(w)$ ↔ Autocovariance $c(h)$	$s(w) \coloneqq 4\int_0^\infty c(h)\cos(2\pi w h)\,\mathrm{d}h$	$c(h) = \int_0^\infty s(w)\cos(2\pi w h)\,\mathrm{d}w$	(11)
Climacogram $\gamma(k)$ ↔ Power spectrum $s(w)$	$\gamma(k) = \int_0^\infty s(w)\ \mathrm{sinc}^2(\pi w k)\,\mathrm{d}w$	$s(w) \coloneqq 2\int_0^\infty \frac{\mathrm{d}^2\left(h^2\gamma(h)\right)}{\mathrm{d}h^2}\cos(2\pi w h)\,\mathrm{d}h$	(12)

The **autocovariance function is the second derivative of the climacogram**. The fact that the **estimation of the second derivative from data is too uncertain** explains why the shape of the empirical autocorrelation function is very rough.

The **power spectrum is the Fourier transform of the autocovariance and entails an even rougher shape** and more uncertain estimation than in the autocovariance (Dimitriadis and Koutsoyiannis, 2015).

See more advantages of the climacogram over autocovariance and power spectrum in Dimitriadis and Koutsoyiannis (2015) and Koutsoyiannis (2016).

Asymptotic power laws and the log-log derivative

It is quite common that functions $f(x)$ defined in $[0, \infty)$, whose limits at 0 and ∞ exist, are associated with asymptotic power laws as $x \to 0$ and ∞ (Koutsoyiannis, 2014b).

Power laws are functions of the form

$$f(x) \propto x^b \tag{13}$$

A power law is visualized in a graph of $f(x)$ plotted in logarithmic axis vs. the logarithm of x, so that the plot forms a straight line with slope b. Formally, the slope b is expressed by the **log-log derivative** (LLD):

$$f^{\#}(x) \coloneqq \frac{\mathrm{d}(\ln f(x))}{\mathrm{d}\,(\ln x)} = \frac{x f'(x)}{f(x)} \tag{14}$$

If the power law holds for the entire domain, then $f^{\#}(x) = b =$ constant. Most often, however, $f^{\#}(x)$ is not constant. Of particular interest are the **asymptotic values** for $x \to 0$ and ∞, symbolically $f^{\#}(0)$ and $f^{\#}(\infty)$, which **define two asymptotic power laws**.

Definition and importance of entropy

Historically entropy was introduced in thermodynamics but later it was given a rigorous definition within probability theory (owing to Boltzmann, Gibbs and Shannon). Thermodynamic and probabilistic entropy are essentially the same thing (Koutsoyiannis, 2013, 2014a; but others have different opinion).

Entropy is a dimensionless measure of uncertainty defined as follows:

For a **discrete random variable** $\underline{z}$ with probability mass function $P_j := P\{\underline{z} = z_j\}$:

$$\Phi[\underline{z}] := \mathrm{E}[-\ln P(\underline{z})] = -\sum_{j=1}^{w} P_j \ln P_j \quad (15)$$

For a **continuous random variable** $\underline{z}$ with probability density function $f(z)$:

$$\Phi[\underline{z}] := \mathrm{E}\left[-\ln \frac{f(\underline{z})}{m(\underline{z})}\right] = -\int_{-\infty}^{\infty} \ln \frac{f(z)}{m(z)} f(z)\mathrm{d}z \quad (16)$$

where $m(z)$ is the density of a background measure (usually $m(z) = m = 1[z^{-1}]$).

Entropy acquires its importance from the **principle of maximum entropy** (Jaynes, 1957), which postulates that the entropy of a random variable should be at maximum, under some **conditions, formulated as constraints, which incorporate all knowledge that is deduced** (deterministically) about this variable.

Its physical counterpart, the tendency of **entropy to become maximal (2nd Law of thermodynamics) is the driving force of natural change**.

Entropy production in stochastic processes

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

$$\Phi'[\underline{X}(t)] := \mathrm{d}\Phi[\underline{X}(t)]/\mathrm{d}t \qquad (17)$$

A more convenient (and dimensionless) measure is the **entropy production in logarithmic time** (EPLT):

$$\varphi(t) \equiv \varphi[\underline{X}(t)] := \Phi'[\underline{X}(t)]\, t \equiv \mathrm{d}\Phi[\underline{X}(t)] \,/\, \mathrm{d}(\ln t) \qquad (18)$$

For a Gaussian random variable, the entropy depends on its variance only (Papoulis, 1991). Hence, for a Gaussian process the entropy will depend on the climacogram $\gamma(t)$ only:

$$\Phi[\underline{X}(t)] = (1/2)\,\ln(2\pi \mathrm{e}\, t^2\, \gamma(t)/m^2) \qquad (19)$$

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

$$\varphi(t) = 1 + \gamma'(t)\, t \,/\, 2\gamma(t) = 1 + ½\, \gamma^{\#}(t) \qquad (20)$$

That is, **EPLT** is visualized and estimated by the **slope of a log-log plot of the climacogram.**

When the past and the present are observed, instead of the unconditional variance $\gamma(t)$, we should use a variance $\gamma_{\mathrm{C}}(t)$ conditional on the known past and present. This turns out to equal a **differenced climacogram** (Koutsoyiannis, 2017):

$$\gamma_{\mathrm{C}}(k) = \varepsilon\big(\gamma(k) - \gamma(2k)\big), \qquad \varepsilon = \frac{1}{1 - 2^{\gamma^{\#}(\infty)}} \qquad (21)$$

The **conditional entropy production in logarithmic time (CEPLT)** becomes:

$$\varphi_C(t) = 1 + ½\gamma_C^{\#}(t) \qquad (22)$$

The climacospectrum

By slightly modifying the differenced climacogram (in order to make it integrable in $(0, \infty)$), i.e., by multiplying it with k, we can obtain an additional tool, which resembles the power spectrum and thus is referred to as the **climacospectrum** (Koutsoyiannis, 2017):

$$\zeta(k) \coloneqq \frac{k\big(\gamma(k)-\gamma(2k)\big)}{\ln 2} \tag{23}$$

The climacospectrum is also written in an alternative manner in terms of frequency $w = 1/k$:

$$\tilde{\zeta}(w) \coloneqq \zeta(1/w) = \frac{\gamma(1/w)-\gamma(2/w)}{(\ln 2)w} \tag{24}$$

The inverse transformation, i.e., that giving the climacogram $\gamma(k)$ once the climacospectrum $\zeta(k)$ is known, is:

$$\gamma(k) = \ln 2 \sum_{i=0}^{\infty} \frac{\zeta(2^i k)}{2^i k} = \gamma(0) - \ln 2 \sum_{i=1}^{\infty} \frac{\zeta(2^{-i} k)}{2^{-i} k} \tag{25}$$

As also happens with the power spectrum, the entire area under the curve $\tilde{\zeta}(w)$ is precisely equal to the variance $\gamma(0)$ of the instantaneous process. The climacospectrum has also the same asymptotic behaviour with the power spectrum, i.e.,

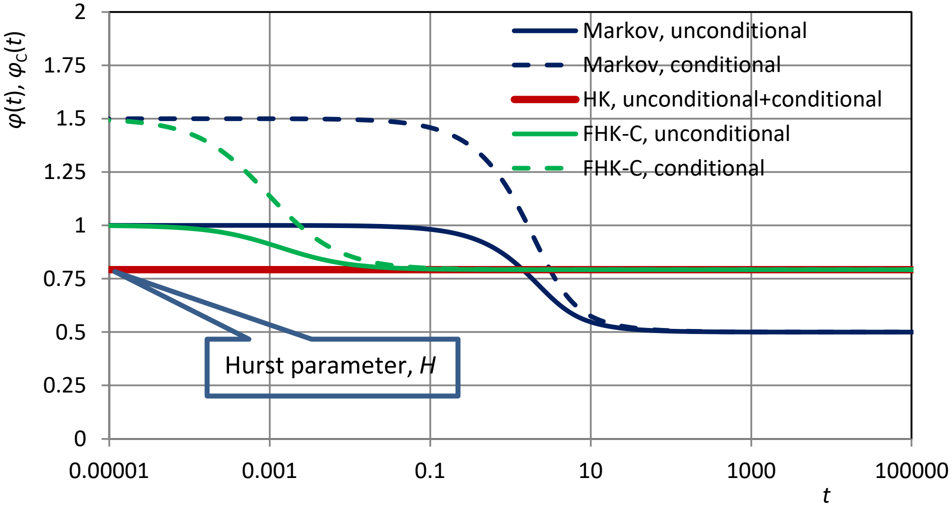
$$\tilde{\zeta}^{\#}(0) = -\zeta^{\#}(\infty) = s^{\#}(0), \qquad \tilde{\zeta}^{\#}(\infty) = -\zeta^{\#}(0) = s^{\#}(\infty) \tag{26}$$

This property holds almost always, with the exception of the cases where $\zeta^{\#}(0)$ is a specific integer ($\zeta^{\#}(\infty) = -1$ or $\zeta^{\#}(0) = 3$).

The climacospectrum is also connected with the CEPLT trough:

$$\varphi_C(k) = ½\left(1 + \zeta^{\#}(k)\right) = ½\left(1 - \tilde{\zeta}^{\#}(1/k)\right) \tag{27}$$

Examples of stochastic processes and their entropy production



All three processes have same:

variance $\gamma(1) = 1$;

autocovariance for lag 1, $c_1^{(1)} = 0.5$;

fractal parameter $M = 0.5$.

The HK and FHK processes have Hurst parameter $H = 0.7925$.

1. A **Markov process** maximizes entropy production for small times ($t \to 0$) but minimizes it for large times ($t \to \infty$):

$$c(h) = \lambda \mathrm{e}^{-h/\alpha}, \ \gamma(k) = \frac{2\lambda}{k/\alpha}\left(1 - \frac{1-\mathrm{e}^{-k/\alpha}}{k/\alpha}\right) \tag{28}$$

2. A **Hurst-Kolmogorov (HK) process** maximizes entropy production for large times ($t \to \infty$) but minimizes it for small times ($t \to 0$):

$$\gamma(k) = \lambda(\alpha/k)^{2-2H} \tag{29}$$

3. A **Filtered Hurst-Kolmogorov (FHK) process** with a generalized Cauchy-type climacogram (FHK-C) maximizes entropy production both for large ($t \to \infty$) and small times ($t \to 0$):

$$\gamma(k) = \lambda(1 + (k/\alpha)^{2M})^{\frac{H-1}{M}} \tag{30}$$

The parameters a and λ are scale parameters. The parameter H is the Hurst parameter and determines the global properties of the process with the notable property $H = \varphi(\infty) = \varphi_C(\infty)$. The parameter M (for Mandelbrot) is the fractal parameter. Both M and H are dimensionless parameters varying in the interval (0, 1] with $M < ½$ or $> ½$ indicating a rough or a smooth process, respectively, and with $H < ½$ or $> ½$ indicating an antipersistent or a persistent process, respectively (see also the graph in p. 20).

Asymptotic scaling of second order properties

EPLT and the CEPLT are related to LLDs (slopes of log-log plots) of second order tools such as **climacogram, climacospectrum, power spectrum**, etc. With a few exceptions, these slopes are nonzero asymptotically, hence entailing **asymptotic scaling** or **asymptotic power laws** with the **LLDs being the scaling exponents**. It is intuitive to expect that an emerging asymptotic scaling law would provide a good approximation of the true law for a range of scales.

If the scaling law was appropriate for the entire range of scales, then we would have a simple scaling law. Such simple scaling sounds attractive from a mathematical point of view, but it turns out to be **impossible in physical processes** (Koutsoyiannis, 2017; see also the graph in p. 20).

It is thus physically more realistic to expect **two different types of asymptotic scaling** laws, one in each of the ends of the continuum of scales. The respective scaling exponents are the following:

1. Local scaling or **smoothness** or **fractal behaviour**, for small time scales ($k \to 0$) or high frequencies ($w \to \infty$):

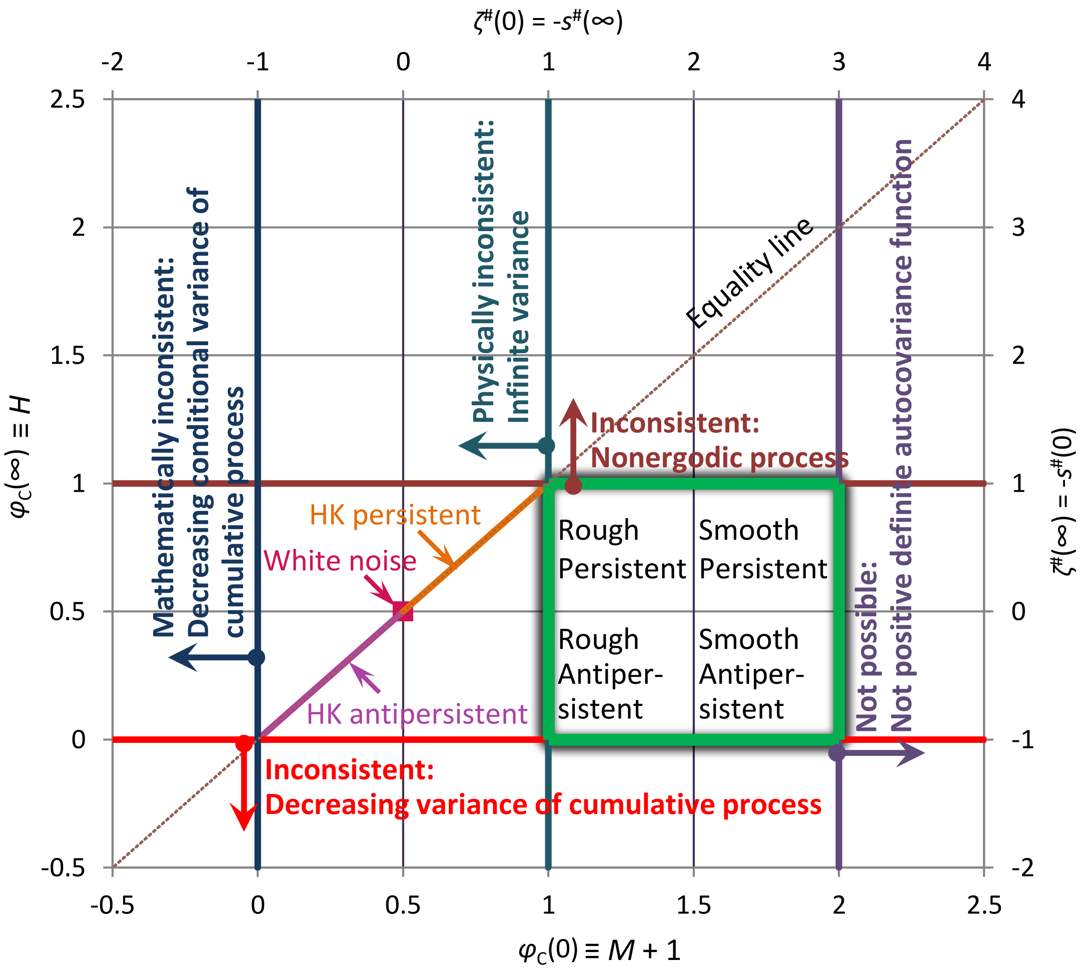
$$\gamma_{\mathrm{C}}^{\#}(0) = \xi^{\#}(0) = \ v^{\#}(0) = \zeta^{\#}(0) - 1 = \ 2\varphi_C(0) - 2 = -s^{\#}(\infty) - 1 \ = \ 2M \tag{31}$$

2. Global scaling or **persistence** or **Hurst- Kolmogorov behaviour**, for $k \to \infty$ or $w \to 0$**:**

$$\gamma_{\mathrm{C}}^{\#}(\infty) = \gamma^{\#}(\infty) = c^{\#}(\infty) = \zeta^{\#}(\infty) - 1 = \ 2\varphi_C(\infty) - 2 = -s^{\#}(0) - 1 = 2H - 2 \tag{32}$$

Here, the emergence of scaling has been related to maximum entropy considerations, and this may provide the theoretical background in modelling complex natural processes by such scaling laws. Generally, **scaling laws are a mathematical necessity** and could be constructed for virtually any continuous function defined in $(0, \infty)$. In other words, there is no magic in power laws, except that they are, logically and mathematically, a necessity.

Bounds of scaling



Bounds of asymptotic values of CEPLT, $\varphi_C(0)$ and $\varphi_C(\infty)$, and corresponding bounds of the log-log slopes of power spectrum and climacospectrum (from Koutsoyiannis, 2017)

The "**green square**" represents the **physically realistic** region.

On its left lies another equal square denoting a **mathematically feasible but physically inconsistent** region.

The reasons why a process out of the square would be physically or mathematically inconsistent are also marked (note that $s^\#$ can, by exception, take on values out of the square when $\varphi_C(0) = 2$ or $\varphi_C(\infty) = 0$).

The lines $\varphi_C(0) = 3/2$ and $\varphi_C(\infty) = 1/2$ define "neutrality" (which is represented by a Markov process) and support the classification of stochastic processes into the indicated four categories (smaller squares within the "green square").

From unknowable to knowable moments: Definition of K-moments

To derive **knowable moments** for high orders p, in the expectation defining the pth moment we raise $(\underline{x} - \mu)$ to a **lower power** $q < p$ and for the remaining $(p - q)$ terms we **replace** $(\underline{x} - \mu)$ with $(2F(\underline{x}) - 1)$, where $F(x)$ is the distribution function. This leads to the following definition of the central *K-moment* of order (p, q):

$$K_{pq} \coloneqq (p - q + 1)\mathrm{E}\left[\left(2F\left(\underline{x}\right) - 1\right)^{p-q}(\underline{x} - \mu)^q\right] \tag{33}$$

Likewise, we define the non-central K-moment of order (p, q) as:

$$K'_{pq} \coloneqq (p - q + 1)\mathrm{E}\left[\left(F\left(\underline{x}\right)\right)^{p-q} \underline{x}^q\right] \tag{34}$$

The quantity $\left(2F\left(\underline{x}\right) - 1\right)^{p-q}$ is estimated from a sample without using powers of $\underline{x}$. Specifically, for the ith element of a sample $x_{(i)}$ of size n, sorted in ascending order, $F(x_{(i)})$ and $(2F(x_{(i)}) - 1)$ are estimated as,

$$\hat{F}\left(x_{(i)}\right) = \frac{i-1}{n-1},\ 2\hat{F}\left(x_{(i)}\right) - 1 = \frac{2i-n+1}{n-1} \tag{35}$$

taking values in (0, 1) and (–1, 1), respectively, irrespective of the values $x_{(i)}$. Hence, the estimators are:

$$\underline{\hat{K}}'_{pq} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{i-1}{n-1}\right)^{p-q} \underline{x}_{(i)}^q, \quad \underline{\hat{K}}_{pq} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{2i-n+1}{n-1}\right)^{p-q}\left(\underline{x}_{(i)} - \hat{\mu}\right)^q \tag{36}$$

Rationale of the definition

1. Assuming that the distribution mean is close to the median, so that $F(\mu) \approx 1/2$ (this is precisely true for a symmetric distribution), the quantity whose expectation is taken in (33) is $A(x) \coloneqq \left(2F\left(\underline{x}\right) - 1\right)^{p-q} (\underline{x} - \mu)^q$ and its Taylor expansion is

$$A(x) = \left(2f(\mu)\right)^{p-q} (\underline{x} - \mu)^p + (p-q)\left(2f(\mu)\right)^{p-q-1} f'(\mu)(\underline{x} - \mu)^{p+1} + O((\underline{x} - \mu)^{p+2}) \quad (37)$$

where $f(x)$ is the probability density function of $\underline{x}$. Clearly then, K_{pq} depends on μ_p as well as classical moments of $\underline{x}$ of order higher than p. The **independence of K_{pq} from classical moments of order < p makes it a good knowable surrogate of the unknowable μ_p**.

2. As p becomes large, by virtue of the multiplicative term $(p - q + 1)$ in definition (33), K_{pq} shares similar asymptotic properties with $\hat{\mu}_p^{q/p}$ (the estimate, not the true $\mu_p^{q/p}$). To illustrate this for $q = 1$, we consider the variable $\underline{z} \coloneqq \max_{1 \le i \le p} \underline{x}_i$ and denote $f(\ )$ and $h(\ )$ the probability densities of $\underline{x}_i$ and $\underline{z}$, respectively. Then (Papoulis, 1990, p. 209):

$$h(z) = p f(z)\left(F(z)\right)^{p-1} \quad (38)$$

and thus, by virtue of (34),

$$\mathrm{E}\left[\underline{z}\right] = p\mathrm{E}\left[\left(F\left(\underline{x}\right)\right)^{p-1} \underline{x}\right] = K'_{p1} \quad (39)$$

On the other hand, for positive $\underline{x}$ and large $p \to n$,

$$E[\underline{\hat{\mu}}'^{\,1/p}_p] = E\left[\left(\frac{1}{n}\sum_{i=1}^{n} \underline{x}_i^p\right)^{1/p}\right] \approx E\left[\max_{1 \le i \le n} x_i\right] = \mathrm{E}\left[\underline{z}\right] = K'_{p1} \quad (40)$$

Note also that the multiplicative term $(p - q + 1)$ in definition (33) and (34) makes K-moments generally increasing functions of p.

Asymptotic properties of moment estimates

Generally, as p becomes large (approaching n), estimates of both classical and K moments, central or non-central, become estimates of expressions involving extremes such as $\left(\max_{1\le i\le p} x_i\right)^q$ or $\max_{1\le i\le p}(x_i-\mu)^q$. For negatively skewed distributions these quantities can also involve minimum, instead of maximum quantities.

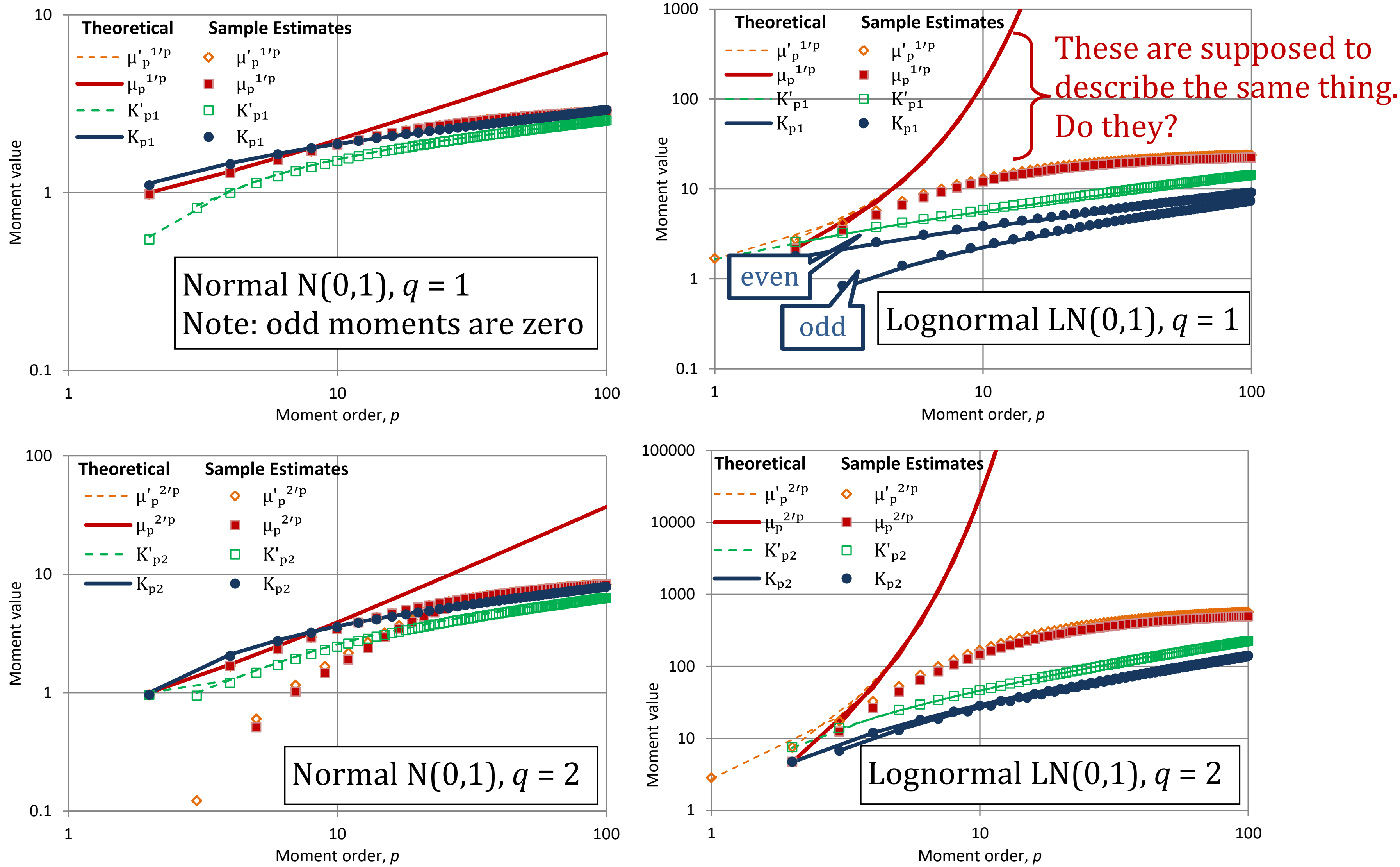
For the K-moments this is consistent with their theoretical definition. For the classical moments this is an inconsistency.

A common property of both classical and K moments is that symmetrical distributions have all their odd moments equal to zero.

For unbounded variables both classical and K moments are non-decreasing functions of p, separately for odd and even p.

In geophysical processes we can justifiably assume that the variance $\mu_2 \equiv \sigma^2 \equiv K_{22}$ **is finite** (an infinite variance would presuppose infinite energy to materialize, which is absurd). Hence, high order K-moments K_{p2} will be finite too, even if classical moments μ_p diverge to infinity beyond a certain p (i.e., in heavy tailed distributions).

Justification of the notion of *unknowable* vs. *knowable*



Note: Sample sizes are ten times higher than the maximum p shown in graphs, i.e., 1000.

Relationship among different moment types

The **classical moments** can be recovered as a **special case of K-moments**: $M_p \equiv K_{pp}$. In particular, in **uniform distribution, classical and K-moments are proportional to each other**:

$$K'_{pq} \coloneqq (p-q+1)\mu'_p, \;\; K_{pq} \coloneqq (p-q+1)\mu_p \tag{41}$$

The **probability weighted moments** (PWM), defined as $\beta_p \coloneqq \mathrm{E}\left[\underline{x}\left(F\left(\underline{x}\right)\right)^p\right]$, are a **special case of K- moments** corresponding to q = 1:

$$K'_{p1} = p\beta_{p-1} \tag{42}$$

The **L-moments** are defined as $\lambda_p \coloneqq \frac{1}{p}\sum_{k=0}^{p-1}(-1)^k\binom{p-1}{k}\mathrm{E}\left[\underline{x}_{(p-k):p}\right]$, where $\underline{x}_{k:p}$ is the kth order statistic in a sample of size p. **L-moments are also related to PWM and through them to K-moments.** The relationships for the different types of moments for the first four orders are:

$$\begin{aligned}
&K'_{11} = \mu = \beta_0, && K_{11} = 0 \\
&K'_{21} = 2\beta_1, && K_{21} = 2(K'_{21} - \mu) = 4\beta_1 - 2\beta_0 = 2\lambda_2 \\
&K'_{31} = 3\beta_2, && K_{31} = 4(K'_{31} - \mu) - 6(K'_{21} - \mu) = 12\beta_2 - 12\beta_1 + 2\beta_0 = 2\lambda_3 \\
&K'_{41} = 4\beta_3, && K_{41} = 8(K'_{41} - \mu) - 16(K'_{31} - \mu) + 12(K'_{21} - \mu) \\
& && \quad = 32\beta_3 - 48\beta_2 + 24\beta_1 - 4\beta_0 = \frac{8}{5}\lambda_4 + \frac{12}{5}\lambda_2
\end{aligned} \tag{43}$$

Both **PWM and L-moments** are better estimated from samples than classical moments but they are all of first order in terms of the random variable of interest. **PWM and L-moments** are **good to characterize independent series or to infer the marginal distribution** of stochastic processes, but they **cannot characterize** even **second order dependence** of processes; **K-moments can**.

Characteristics a marginal distribution using K-moments

Within the framework of K-moments, while respecting the rule of thumb "Just two moments" in terms of the power of $\underline{x}$, i.e. q = 1 or 2, we can obtain knowable statistical characteristics for much higher orders p.

In this manner, for $p > 1$ we have two alternative options to define statistical characteristics related to moments of the distribution, as in the table below. (Which of the two is preferable depends on the statistical behaviour, and in particular, the mean, mode and variance, of the estimator.)

Characteristic	**Order p**	**Option 1**	**Option 2**	**Option 3***
Location	1	$K'_{11} = \mu$ (the classical mean)		
Variability	2	$K_{21} = 2(K'_{21} - \mu) = 2\lambda_2$	$K_{22} = \mu_2 = \sigma^2$ (the classical variance)	
Skewness (dimensionless)	3	$\frac{K_{31}}{K_{21}} = \frac{\lambda_3}{\lambda_2}$	$\frac{K_{32}}{K_{22}}$	$\frac{K_{32}}{K_{22}^{3/2}} = \frac{\mu_3}{\sigma^3}$
Kurtosis (dimensionless)	4	$\frac{K_{41}}{K_{21}} = \frac{4}{5}\frac{\lambda_4}{\lambda_2} + \frac{6}{5}$	$\frac{K_{42}}{K_{22}}$	$\frac{K_{42}}{K_{22}^2} = \frac{\mu_3}{\sigma^4}$

* Option 3 is based on the classical moments and is not recommended for distribution fitting.

Statistical behaviour of variability, skewness and kurtosis indices

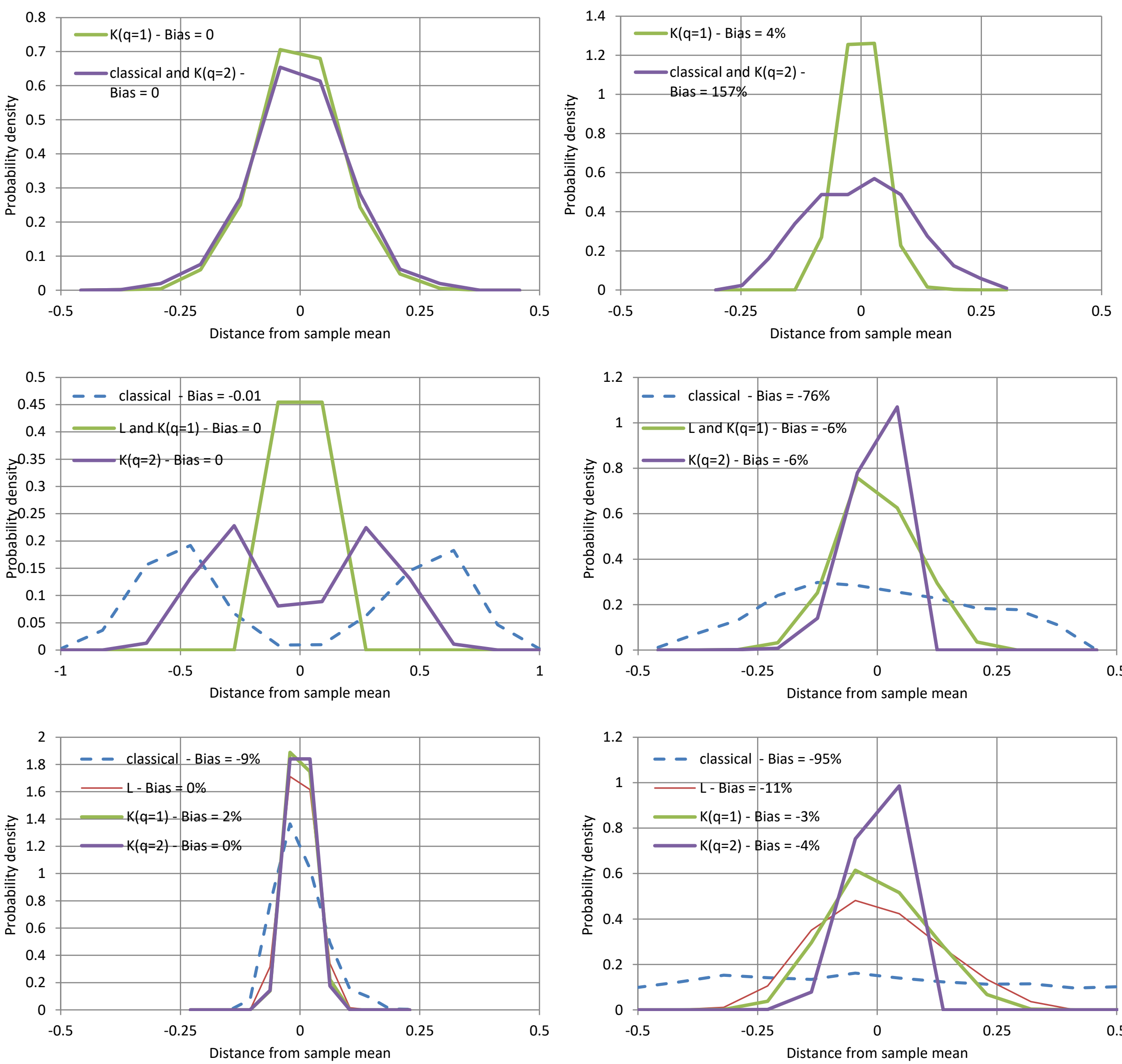


Illustration of the probability density function of:

(upper) variability index ($K_{11}/K_{21}, \mu/\sigma \;\equiv K_{11}/\sqrt{K_{22}}$; note that the latter is inverse of the common coefficient of variation);

(middle) skewness index ($\mu_3^{1/3}/\sigma, K_{31}/K_{21}$, $\mathrm{sign}(K_{32})\sqrt{|K_{32}|/K_{22}}$);

(lower) kurtosis index ($\mu_4^{1/4}/\sigma, \lambda_4/\lambda_2, K_{41}/K_{21}$, $\sqrt{K_{42}/K_{22}}$).

The panels of the left column correspond to the normal distribution N(0,1) and those of the right column to the lognormal distribution LN(0, 2).

High order moments for stochastic processes: the K-climacogram and the K-climacospectrum

The full description of the third-order, fourth-order, etc., properties of a stochastic process requires functions of 2, 3, …, variables.

For example, the **third order properties are expressed in terms of the two-variable function**:

$$c_3(h_1, h_2) := \mathrm{E}[(\underline{x}(t) - \mu)\,(\underline{x}(t + h_1) - \mu)\,(\underline{x}(t + h_2) - \mu)] \quad (44)$$

Such a description is **not parsimonious and its accuracy holds only in theory**, because sample estimates are not reliable.

This problem is remedied if we **introduce single-variable descriptions for any order** p, expanding the idea of the climacogram and climacospectrum based on K-moments.

K-climacogram: $$\gamma_{pq}(k) = (p - q + 1)\mathrm{E}\left[\left(2F\left(\underline{X}(k)/k\right) - 1\right)^{p-q}\left(\underline{X}(k)/k - \mu\right)^q\right] \quad (45)$$

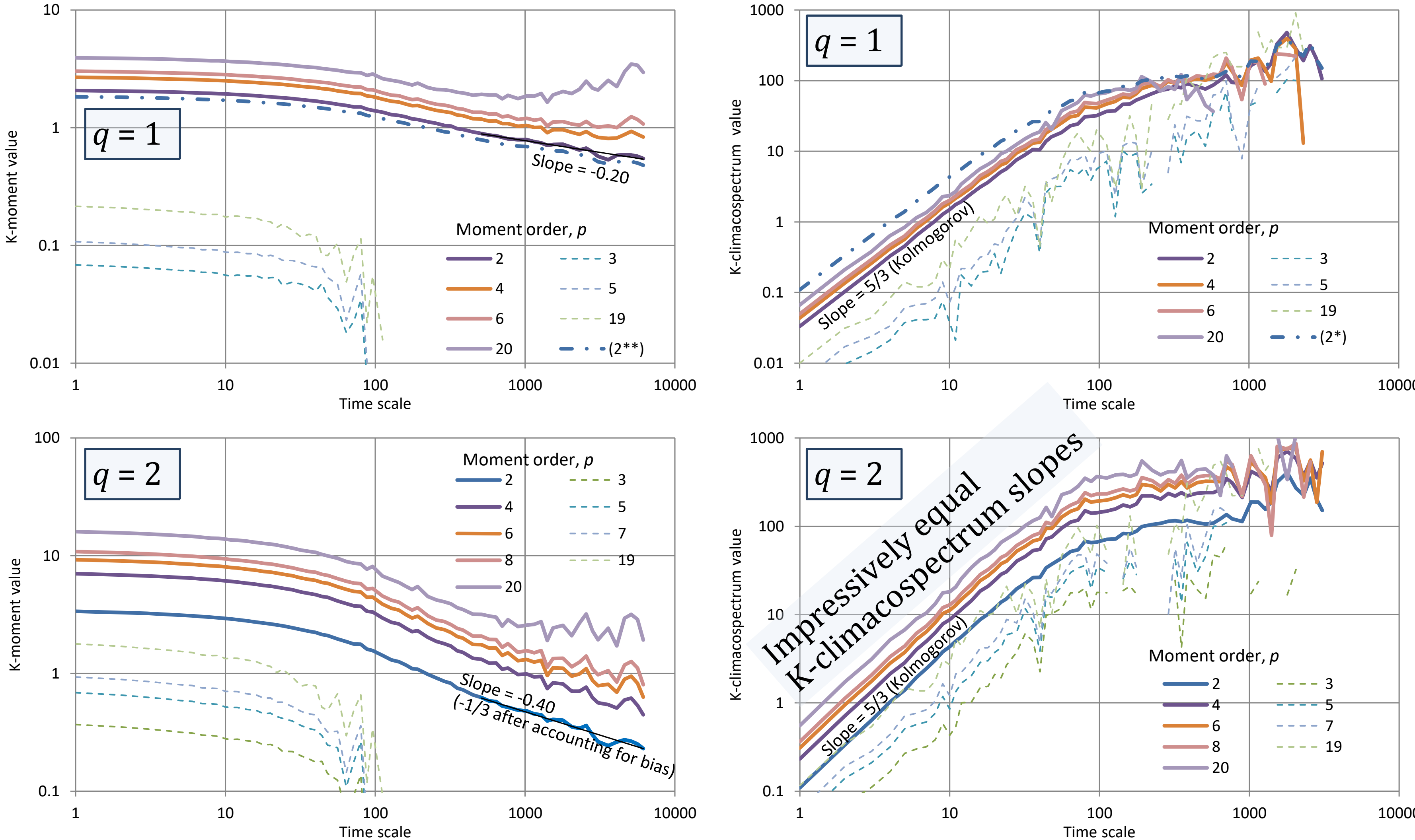
K-climacospectrum: $$\zeta_{pq}(k) = \frac{k\left(\gamma_{pq}(k) - \gamma_{pq}(2k)\right)}{\ln 2} \quad (46)$$

where $\gamma_{22}(k) \equiv \gamma(k)$ and $\zeta_{22}(k) \equiv \zeta(k)$.

While the standard climacogram $\gamma_{22}(k) \equiv \gamma(k)$ provides a description precisely equivalent to the classical, this is not the case for $q > 2$. In this case, the single-variable K-climacogram description is obviously not equivalent to the multivariate high-order one. However, it suffices to define the marginal distribution at any scale k.

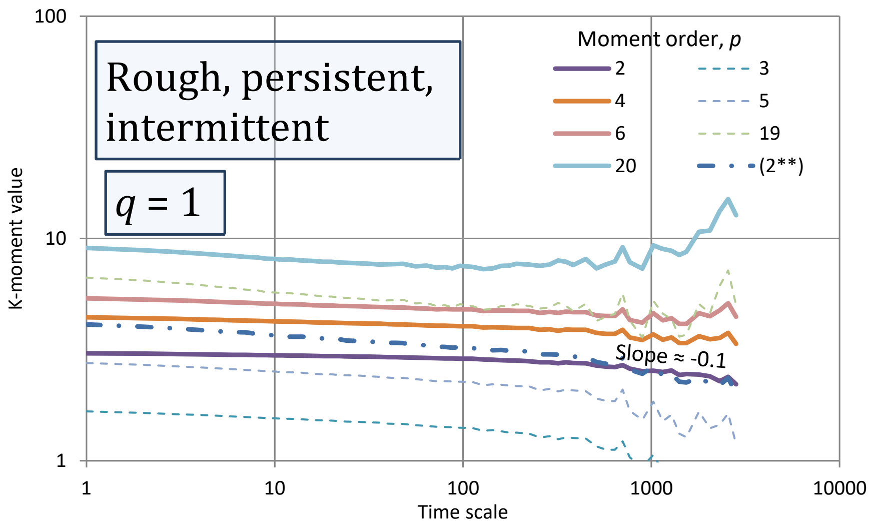
Example 1: Turbulent velocity

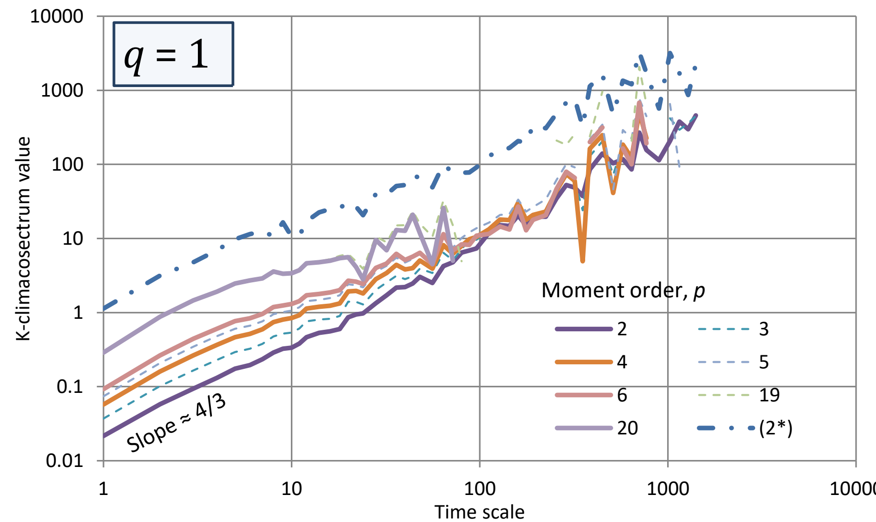
Slightly rough, persistent, slightly intermittent

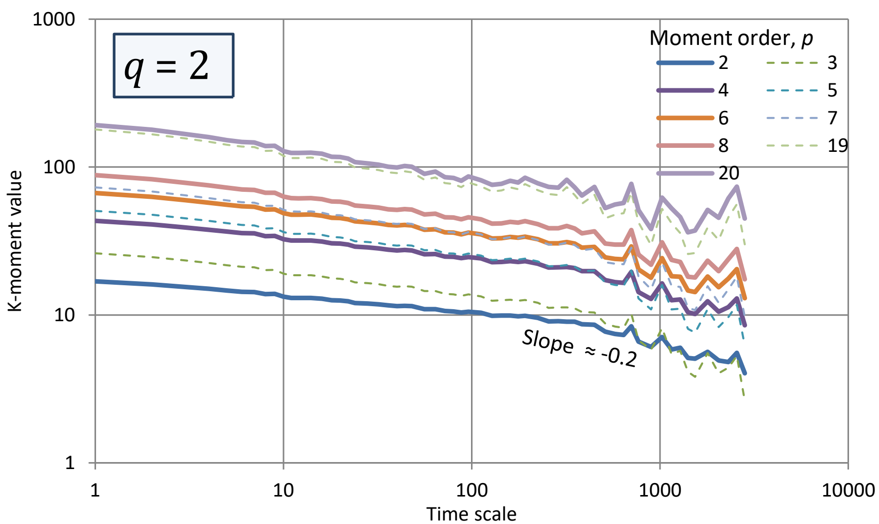


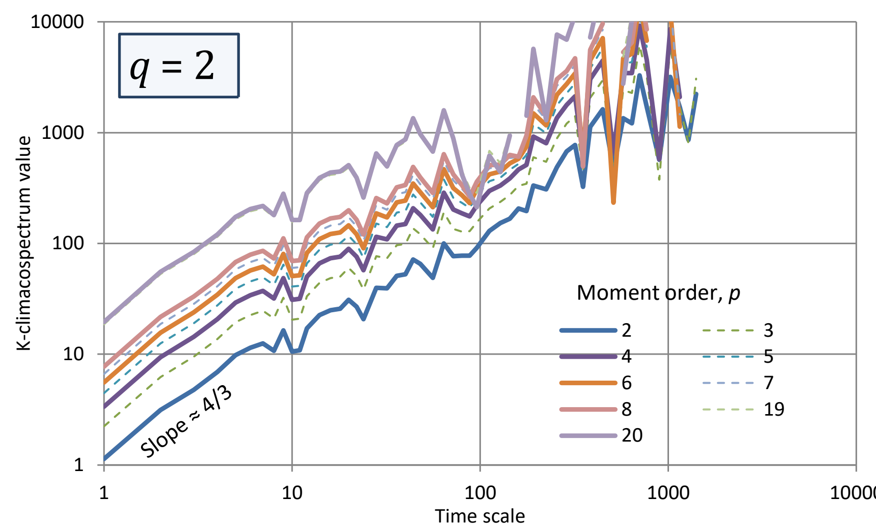
Data: 60 000 values of turbulent velocity along the flow direction (Kang, 2003; Koutsoyiannis 2017, Dimitriadis and Koutsoyiannis, 2018); the original series was averaged so that time scale 1 corresponds to 0.5 s.
Note: Plot (2*) is constructed from the variance and (2**) corresponds to standard deviation.

Example 2: Rainfall rate at Iowa measured every 10 s





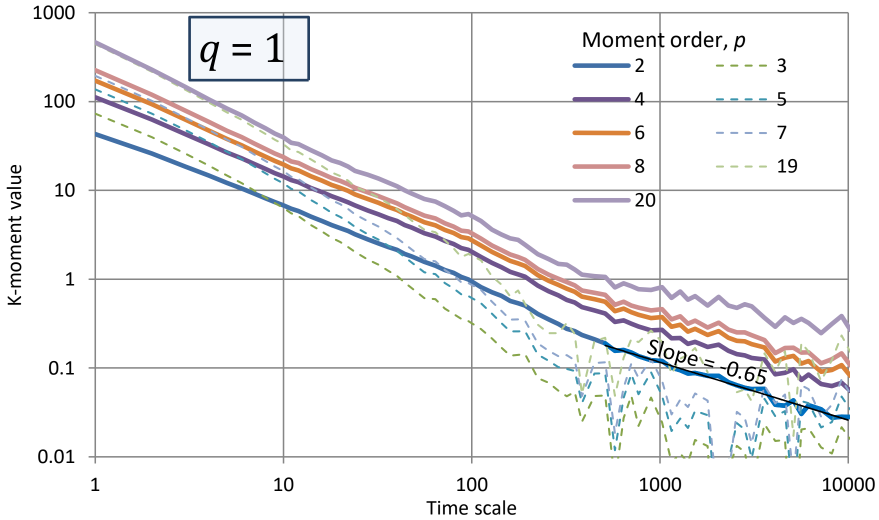


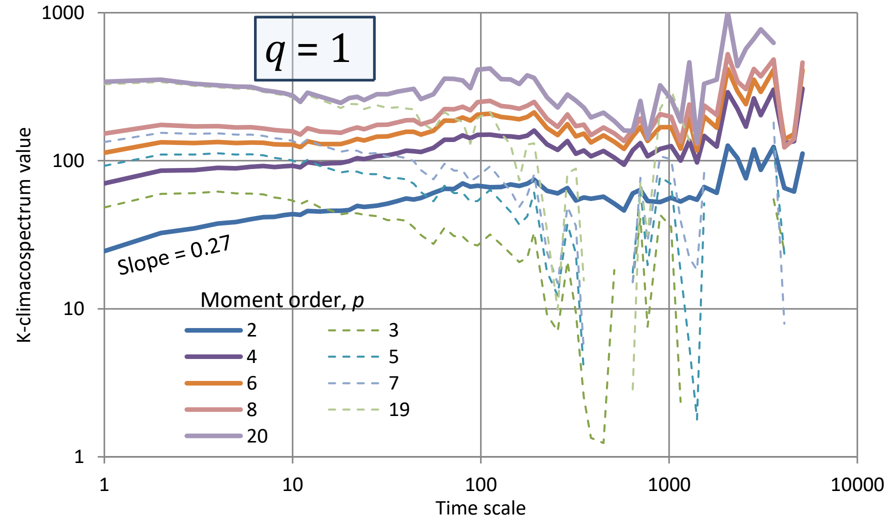


Data: 29542 values of rainfall at Iowa measured at temporal resolution of 10 s (merger of seven events from Georgakakos et al. 1994; see also Lombardo et al. 2012). Plot (2*) is constructed from the variance and (2**) corresponds to standard deviation.

Example 3: Daily rainfall at Padova

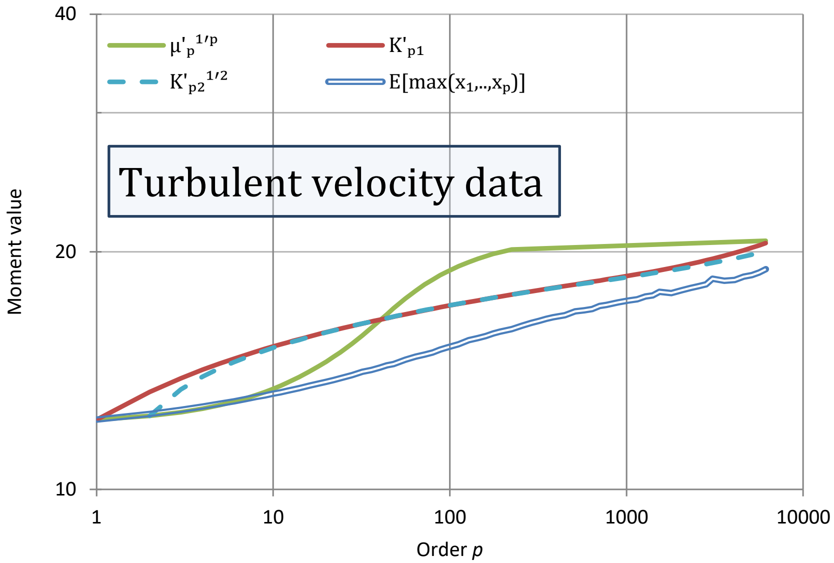
Very rough, slightly persistent, highly intermittent

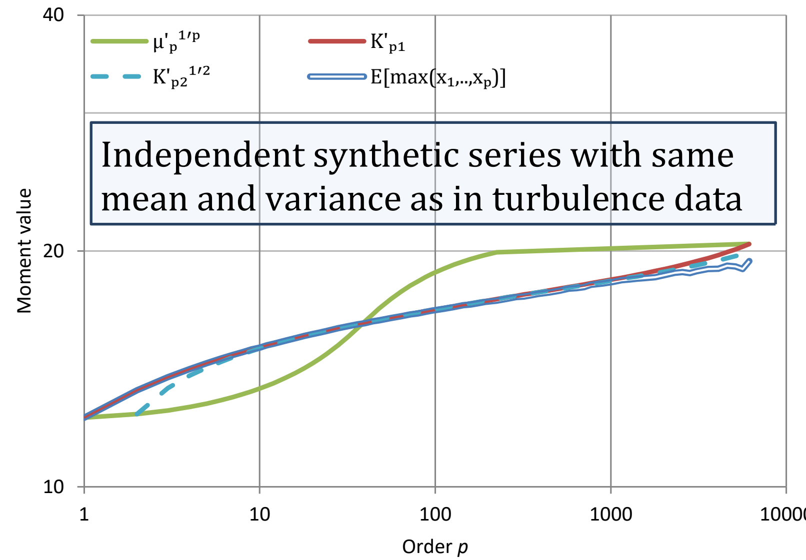


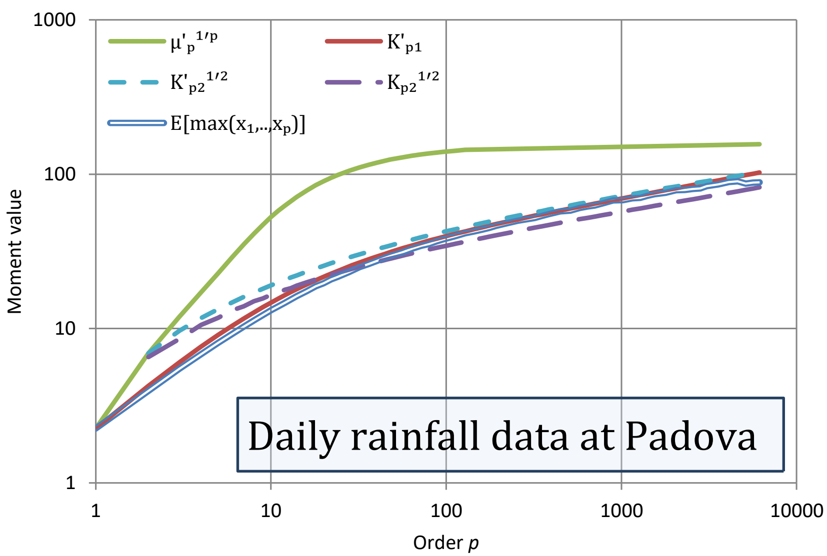


Data: 100 442 values of daily rainfall at Padova (the longest rainfall record existing worldwide; Marani and Zanetti, 2015).

K-moments vs. moment order







Note 1: Each curve is in fact a series of connected points whose shape is smooth by itself (not artificially smoothed).

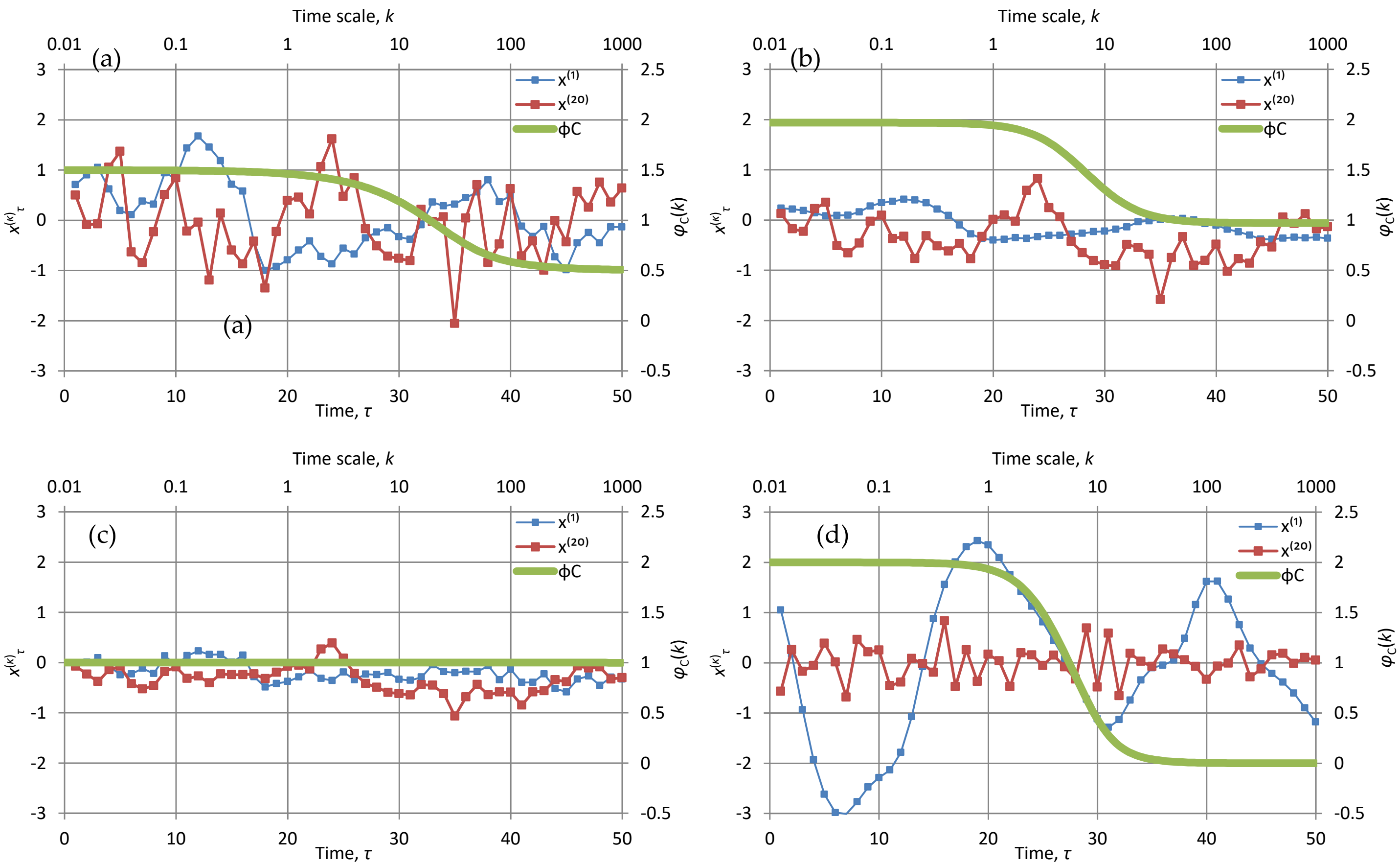
Note 2: Moments of order *p* approximately represent maxima for a time window of length *p*. For independent processes $\mathrm{E}\left[\max\left(\underline{x}_1, \dots, \underline{x}_p\right)\right]$ should be equal to K'_{p1}, but when there is dependence the two quantities differ. The former reflects the joint distribution and the latter the marginal one. The difference is related to the dependence structure.

Generic stochastic simulation methodology

The symmetric moving average (SMA) method (Koutsoyiannis, 2000) can generate time series with any arbitrary autocorrelation function from $\underline{x}_i = \sum_{l=-q}^{q} a_{|l|} \underline{v}_{i+l}$, which transforms white noise $\underline{v}_i$, not necessarily Gaussian, to a process $\underline{x}_i$ with the specified autocorrelation. The modelling steps (Koutsoyiannis, 2018) are summarized as follows:

1. We estimate K-moments for q = 1 and 2 we choose a marginal distribution for it based on K-moments and possibly relevant theoretical considerations (e.g. entropy maximization).
2. We construct the climacogram and climacospectrum, and we choose a suitable model of second-order dependence (see a repertoire of models in Koutsoyiannis 2016, 2017).
3. We estimate the marginal and joint distribution parameters of the model (with appropriate provision for fitting issues, such as bias, e.g., as in Koutsoyiannis 2016).
4. Based on the model parameters we calculate theoretically (and not estimate from data) the classical moments of the process of interest.
5. From theoretical relationships of moments with cumulants we calculate the cumulants of the process of interest.
6. We readily calculate the cumulants of the white noise process and hence its moments.
7. We choose an appropriate distribution for the white noise, calculate its parameters theoretically from its moments and generate a random sample with the required length.
8. Filtering with the SMA scheme we synthesize the simulated series for the process of interest.
9. We construct K-climacograms from the original and synthetic data and compare for relevant moment orders > 2.
10. If a disagreement is found in step 9, then we repeat the process separating the entire range of relevant scales to parts, building different models for each part, and coupling the separate models using a model coupling (disaggregation) scheme such as that in Koutsoyiannis (2001).

Some results of simulations for several models



(a) **Markov**;

(b) FHK, with **CEPLT** close to the **absolute maximum** ($H = M = 0.97$);

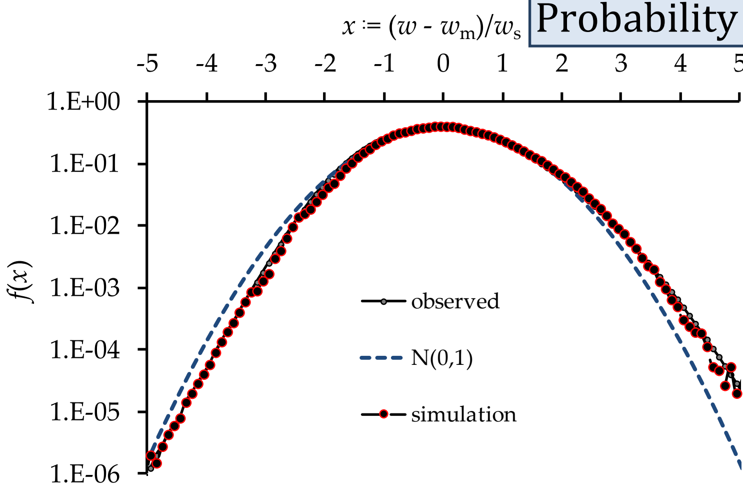
(c) FHK, close to "**red noise**", i.e., with CEPLT close to the absolute maximum for large scales ($H = 0.99$) and close to the absolute minimum for small scales ($M = 0.01$);

(d) process with the **blackbody** spectrum, i.e. with CEPLT equal to the absolute minimum (0) for large scales and to the absolute maximum (2) for small scales.

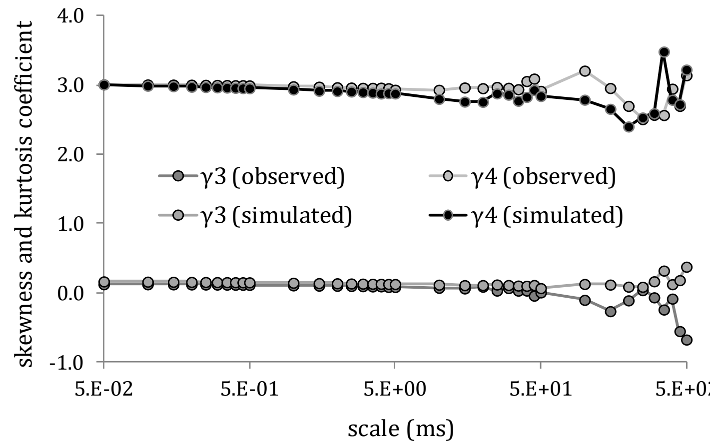
The first fifty terms of times series at time scales $k = 1$ and 20 of time series produced by various models, along with "stamps" of the models (green lines plotted with respect to the secondary axes) represented by the CEPLT, $\varphi_C(k)$. In all cases the discretization time scale is $D = 1$, the characteristic time scale $a = 10$, and the characteristic variance scale λ is chosen so that for time scale D, $\gamma(D) = 1$. The mean is 0 in all cases and the marginal distribution is normal (see details in Koutsoyiannis, 2017).

Impressive results of homogenous turbulence simulation (preserving the 2-climacogram and four marginal moments)

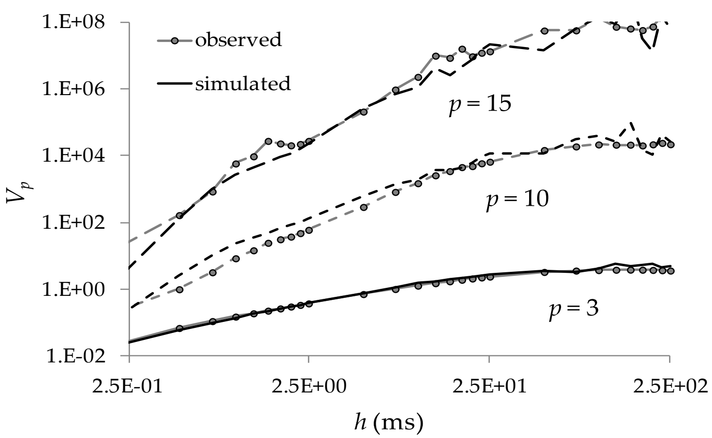
Probability density



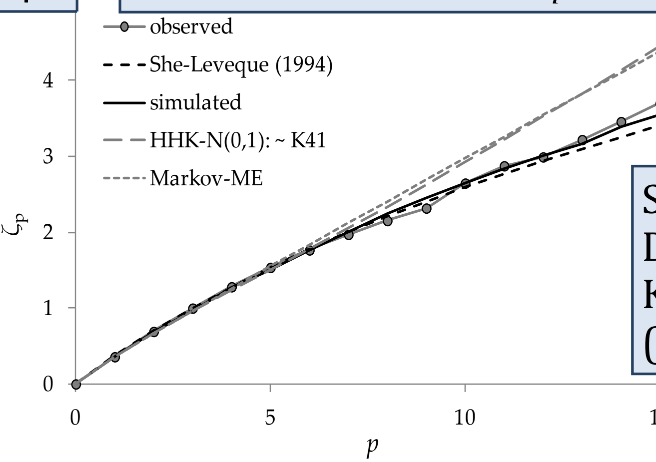
Coefficients of skewness & kurtosis



Generalized structure function $V_p(h) := \mathrm{E}[|\underline{x}_i - \underline{x}_{i+h}|^p$



Scaling exponent ζ_p in $V_p(h) \approx h^{\zeta_p}$



See details in Dimitriadis & Koutsoyiannis (2018)

Concluding remarks

Awareness of stochastics is important in analysing and modelling geophysical processes.

Pursuit of simplicity and parsimony is a guide for good modelling.

Moments of single variables are simpler than joint moments and can replace the latter offering some advantages and practically at no cost.

In this regard, (second-order) climacograms contain precisely the same information as autocovariance and power spectrum, displaying better statistical properties.

High-order classical moments are in fact unknowable for typical geophysical samples; however, high-order properties of distributions can be inferred from knowable moments (the K-moments).

The K-climacogram and the K-climacospectrum enable viewing the statistical properties and possible scaling behaviours of a process at a wide range of scales and without applying laborious transformations.

The turbulence case study indicates that simplicity can achieve impressive results: by preserving just the second-order climacogram, along with the marginal third and fourth moments at a single scale, we can reproduce the observable behaviour of homogeneous turbulence.

The question if knowable moments and K-climacograms are relevant to hydrofractals may have a positive reply but requires further research.

References

- Birkhoff, G. D., 1931. Proof of the ergodic theorem. *Proc. Nat. Acad. Sci.*, 17, 656–660.
- Boltzmann, L., 1877. Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht. *Wiener Ber.*, 76, 373–435 (in German).
- Boltzmann, L., 1884/85. Über die Eigenschaften monocyklischer und anderer damit verwandter Systeme. *Crelles J.*, 98, 68–94.
- Dimitriadis, P., and Koutsoyiannis, D., 2015. Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst–Kolmogorov processes. *Stochastic Environmental Research & Risk Assessment*, 29 (6), 1649–1669, doi:10.1007/s00477-015-1023-7.
- Dimitriadis, P., and Koutsoyiannis, D., 2018. Stochastic synthesis approximating any process dependence and distribution. *Stochastic Environmental Research & Risk Assessment*, doi:10.1007/s00477-018-1540-2.
- Jaynes, E.T., 1957. Information theory and statistical mechanics. *Physical Review*, 106 (4), 620-630.
- Georgakakos, K. P., Cârsteanu, A. A., Sturdevant, P. L., and Cramer, J. A., 1994. Observation and snalysis of Midwestern rain rates. *Journal of Applied Meteorology*, 33, 1433-1444.
- 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.
- Khintchine, A., 1933. Zu Birkhoffs Lösung des Ergodenproblems. *Math. Ann.*, 107, 485-488.
- Khintchine, A., 1934. Korrelationstheorie der stationären stochastischen Prozesse. *Mathematische Annalen*, 109 (1), 604–615.
- Kolmogorov, A. N., 1931. Uber die analytischen Methoden in der Wahrscheinlichkcitsrechnung, *Mathematische Annalen*, 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. *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Ergebnisse der Mathematik und Ihrer Grenzgebiete), Berlin (2nd English Edition: Foundations of the Theory of Probability, 84 pp. Chelsea Publishing Company, New York, 1956).
- Kolmogorov, A. N., 1937. Ein vereinfachter Beweis der Birkhoff-Khintchineschen Ergodensatzes. *Mat. Sb.*, 2, 367-368.
- Kolmogorov, A.N., 1938. A simplified proof of the Birkhoff-Khinchin ergodic theorem. *Uspekhi Mat. 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., 1947. Statistical theory of oscillations with continuous spectrum. *Collected papers on the 30th anniversary of the Great October Socialist Revolution*, Vol. 1, Akad. Nauk SSSR, Moscow-Leningrad, pp. 242-252. (English edition: 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. 321-330).
- Koutsoyiannis, D., 2000. A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. *Water Resources Research*, 36 (6), 1519–1533,.

References (2)

- Koutsoyiannis, D., 2001. Coupling stochastic models of different time scales, *Water Resources Research*, 37 (2), 379–391, doi: 10.1029/2000WR900200.
- Koutsoyiannis, D., 2006. Nonstationarity versus scaling in hydrology. *Journal of Hydrology*, 324, 239–254.
- Koutsoyiannis, D., 2010. A random walk on water, *Hydrology and Earth System Sciences*, 14, 585–601, doi: 10.5194/hess-14-585-2010.
- Koutsoyiannis, D., 2011a. Hurst-Kolmogorov dynamics as a result of extremal entropy production. *Physica A*, 390 (8), 1424–1432.
- Koutsoyiannis, D., 2011b. Hurst-Kolmogorov dynamics and uncertainty. *Journal of the American Water Resources Association*, 47 (3), 481–495, doi: 10.1111/j.1752-1688.2011.00543.x.
- Koutsoyiannis, D., 2013. Physics of uncertainty, the Gibbs paradox and indistinguishable particles. *Studies in History and Philosophy of Modern Physics*, 44, 480–489.
- Koutsoyiannis, D., 2014a. Entropy: from thermodynamics to hydrology. *Entropy*, 16 (3), 1287–1314.
- Koutsoyiannis, D., 2014b. Random musings on stochastics (Lorenz Lecture). *AGU 2014 Fall Meeting*, San Francisco, USA, American Geophysical Union, doi: 10.13140/RG.2.1.2852.8804.
- 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., 2017. Entropy production in stochastics. *Entropy*, 19 (11), 581, doi: 10.3390/e19110581.
- Koutsoyiannis, D., 2018. Knowable moments for high-order characterization and modelling of hydrological processes. Preprint, doi: 10.13140/RG.2.2.14876.59529.
- Koutsoyiannis, D., and Montanari, A., 2015. Negligent killing of scientific concepts: the stationarity case. *Hydrological Sciences Journal*, 60 (7-8), 1174–1183.
- Lasota, A., and Mackey, M. C. , 1994. *Chaos, Fractals and Noise*, Springer-Verlag.
- Lombardo, F., Volpi, E., and Koutsoyiannis, D., 2012. Rainfall downscaling in time: Theoretical and empirical comparison between multifractal and Hurst-Kolmogorov discrete random cascades. *Hydrological Sciences Journal*, 57 (6), 1052–1066.
- 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, doi: 10.5194/hess-18-243-2014.
- Mackey, M.C., 2003. *Time's Arrow: The Origins of Thermodynamic Behavior*. Dover, Mineola, NY, USA, 175 pp.,
- Marani, M., and Zanetti, S., 2015. Long-term oscillations in rainfall extremes in a 268 year daily time series. *Water Resources Research*, 51, 639–647.
- Mathieu. M., 1988. On the origin of the notion 'Ergodic Theory'. *Expositiones Mathematicae*, 6, 373-377.
- A. Montanari, and D. Koutsoyiannis, Modeling and mitigating natural hazards: Stationarity is immortal!, *Water Resources Research*, 50 (12), 9748–9756, doi:10.1002/2014WR016092, 2014.
- Papoulis, A., 1990. *Probability and Statistics*, Prentice-Hall.
- Papoulis, A., 1991. *Probability, Random Variables and Stochastic Processes*, 3rd edn., McGraw-Hill, New York.