Orlob International Symposium 2016
University California Davis
20-21 June 2016

From time series to stochastics:

A theoretical framework with applications on time scales spanning from microseconds to megayears

Demetris Koutsoyiannis & Panayiotis Dimitriadis
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/en/docinfo/1618/

Part 1: On names and definitions

Schools on names and definitions

The poetic school

What's in a name? That which we call a rose, By any other name would smell as sweet.
—William Shakespeare, "Romeo and Juliet" (Act 2, scene 2)

The philosophico-epistemic school

Ἀρχή σοφίας ὀνομάτων ἐπίσκεψις (The beginning of wisdom is the visiting (inspection) of names)
—Attributed to Antisthenes of Athens, founder of Cynic philosophy

Ἀρχὴ παιδεύσεως ἡ τῶν ὀνομάτων ἐπίσκεψις" (*The beginning of education is the inspection of names*)
—Attributed to Socrates by Epictetus, Discourses, I.17,12,

The beginning of wisdom is to call things by their proper name.
—Chinese proverb paraphrasing Confucius's quote "If names be not correct, language is not in accordance with the truth of things."

On names and definitions (contd.)

The philosophico-epistemic school (contd.)

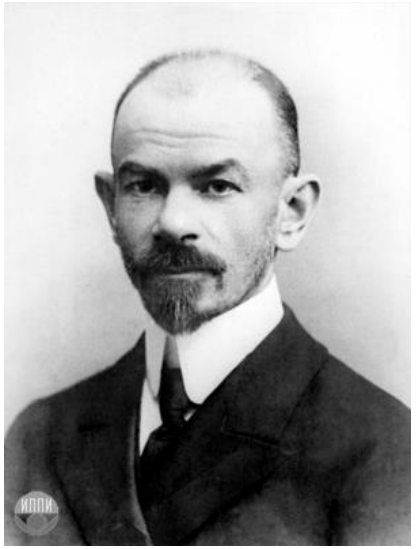
When I name an object with a word, I thereby assert its existence."
—Andrei Bely, symbolist poet and former mathematics student of Dmitri Egorov, in his essay "The Magic of Words"

"Nommer, c'est avoir individu" (to name is to have individuality).
—Nikolai Luzin, leader of the Moscow School of Mathematics (also student of Dmitri Egorov and teacher of Aleksandr Khinchin and Andrey Kolmogorov)

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

Note: The last three quotes are found in a must-read book by Graham and Kantor (2009):
"Naming infinity: A true story of religious mysticism and mathematical creativity"

Giants of the Moscow School of Mathematics



Dmitri Egorov
(1869 – 1931)
[wikipedia]

Nikolai Luzin
(1883 – 1950)
[www.math.nsc.ru/LBRT/g2/english/ssk/case_e.html]

Aleksandr Khinchin
(1894 – 1959)
[wikipedia]

Andrey Kolmogorov
(1903 –1987)
[wikipedia]

Definition of a time series (Wikipedia)

Time series

From Wikipedia, the free encyclopedia

(Redirected from Time series analysis)

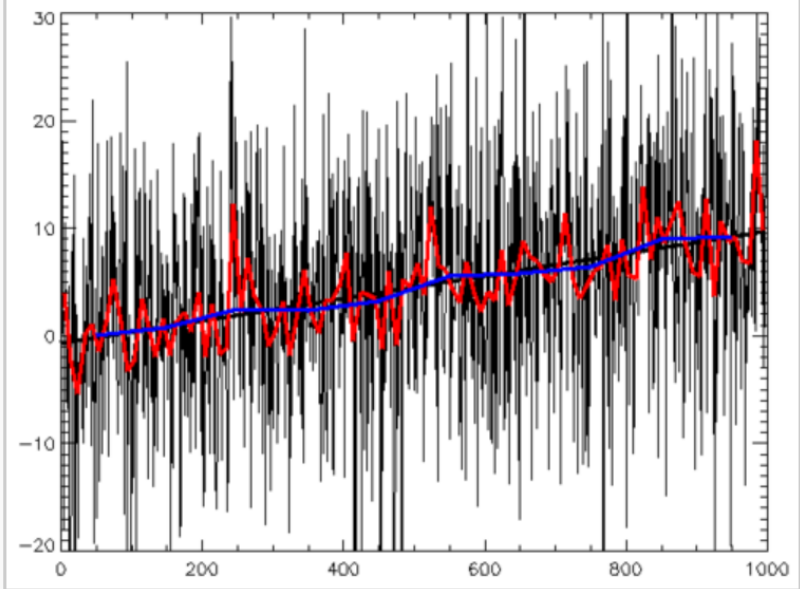
A **time series** is a sequence of data points made:

1) over a continuous time interval

2) out of successive measurements across that interval

3) using equal spacing between every two consecutive measurements

4) with each time unit within the time interval having at most one data point



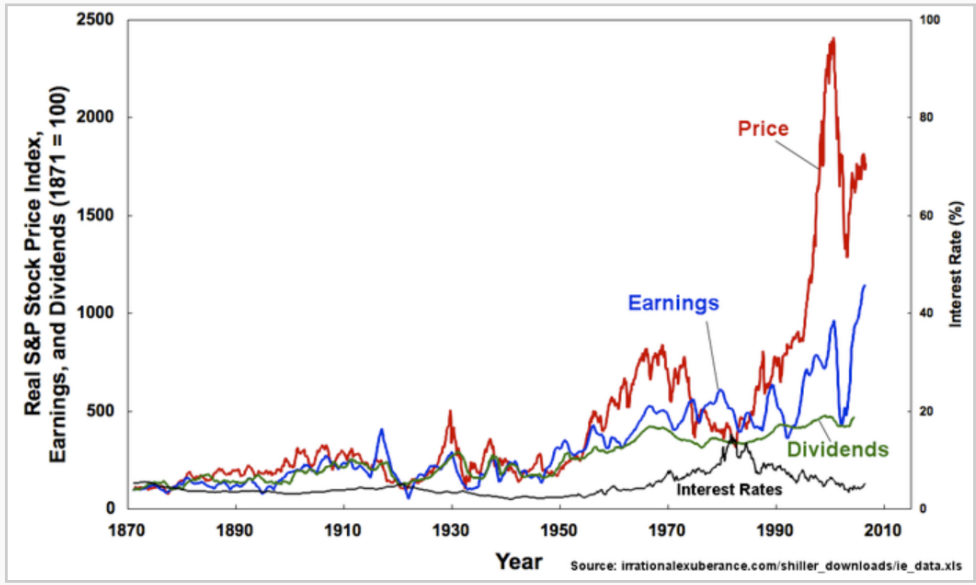
Time series: random data plus trend, with best-fit line and different applied filters

??????

Definition of a stochastic process (Wikipedia)

In probability theory, a **stochastic** (/stoʊˈkæstɪk/) **process**, or often **random process**, is a collection of random variables representing the evolution of some system of random values over time. This is the probabilistic counterpart to a deterministic process (or deterministic system). Instead of describing a process which can only evolve in one way (as in the case, for example, of solutions of an ordinary differential equation), in a stochastic, or random process, there is some indeterminacy: even if the initial condition (or starting point) is known, there are several (often infinitely many) directions in which the process may evolve.

??????



Stock market fluctuations have been modeled by stochastic processes.

Definition of a stationary process (Wikipedia)

Stationary process

From Wikipedia, the free encyclopedia

In mathematics and statistics, a **stationary process** (or **strict(ly) stationary process** or **strong(ly) stationary process**) is a stochastic process whose joint probability distribution does not change when shifted in time. Consequently, parameters such as the mean and variance, if they are present, also do not change over time and do not follow any trends.

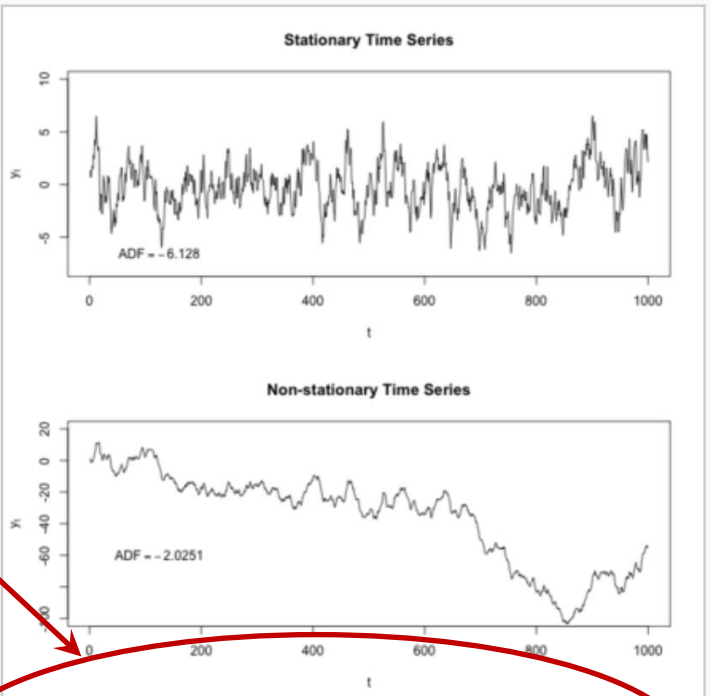
....

Let Y be any scalar random variable, and define a time-series $\{X_t\}$, by

$$X_t = Y \qquad \text{for all } t.$$

Then $\{X_t\}$ is a stationary time series, for which realisations consist of a series of constant values, with a different constant value for each realisation. A law of large numbers

?!?!?!



Two simulated time series processes, one stationary the other non-stationary. The Augmented Dickey–Fuller test is reported for each process and non-stationarity cannot be rejected for the second process.

It seems that, by poetic licence, the terms stochastic process and time series are used interchangeably.

Related definitions in the celebrated book by Kendall and Stuart (1966)

p. 342: Observations on a phenomenon which is moving through time generate an ordered set known as a *time series*. The values assumed by a variable at time t may or may not embody an element of random variation, but in the majority of cases with which we shall be concerned some such element is present, if only as an error of observation.

p. 346: … consider an infinity of values $x(t)$. It is customary and convenient (though not, perhaps, very exact) to speak of continuous time series, when we mean that t is continuous…

p. 404: In the *theory of stochastic processes*, of which stationary *time-series* are a particular case, …

- A definition of a stochastic process is missing.
- A time series is recognized as a series of observations, i.e. numbers which could be a series of values not necessarily associated with a stochastic process.
- However, occasionally the concept of a time series looks to be treated as identical to (or subcase of) that of a stochastic process.

Consistent definitions by the Moscow School of Mathematics

- Kolmogorov (1931) introduced the term *stochastic process* although he cited Bachelier (1900) as having already used stochastic processes (without using that name).
- Kolmogorov (1931) used the term *stationary* to describe a probability density function that is unchanged in time.
- Kolmogorov (1933) introduced the definition of *probability* (based on the measure theory) in an axiomatic manner based on three fundamental concepts (a triplet called *probability space*) and four axioms (non-negativity: normalization, additivity and continuity at zero).
- Khinchin (1934) gave a more formal definition of a *stochastic process* and *stationarity*.
- Kolmogorov (1938) gave a concise presentation of the concepts:

  […] a stationary **stochastic process** in the sense of Khinchin […] **is a set of random variables** x_t depending on the parameter t, $-\infty < t < +\infty$, such that the distributions of the systems

  $$(x_{t_1}, x_{t_2}, \ldots, x_{t_n}) \text{ and } (x_{t_1+\tau}, x_{t_2+\tau}, \ldots, x_{t_n+\tau}) \qquad (1)$$

  coincide for any n, t_1, t_2, … ,t_n, and τ.

Time series models

- Most of the popular knowledge in stochastics originates from so-called time-series books.
- These have given focus on stylized families of models like AR(p), ARMA(p,q), ARIMA(p,d,q), and so on,
  - introduced by Whittle (1951),
  - popularized in the book by Box and Jenkins (1971)
  - extended by Hosking (1981; ARFIMA(p,d,q))
- With the exception of AR(1) and ARMA(1,1) they have several problems:
  - They are too artificial because, being complicated discrete-time models, they do not correspond to a continuous time process, while natural processes typically evolve in continuous time.
  - Their stochastic structure is tightly associated with the number of parameters and usually they become over-parameterized and thus not parsimonious.
  - Their identification, typically based on the estimation of the autocorrelation function from data, usually neglects estimation bias and uncertainly, which in stochastic processes (as opposed to purely random processes) are high.
  - They are unnecessary because synthetic series from a process with any arbitrary autocorrelation structure can be easily generated otherwise.

Concluding remarks of part 1

- Most books could be classified in the "poetic school".
- Nevertheless, there are books on stochastic processes characterized by perfect clarity (following the Khinchin-Kolmogorov conventions), of which *Papoulis* (1991; first edition 1965) is worth mentioning.
- The term *time series* is ambiguous, sometimes denoting a realization of a stochastic process and other times denoting the stochastic process *per se*.
- We can use *Stochastics* as a collective name for *probability theory*, *statistics* and *stochastic processes*.
- Stochastics is much more than numerical calculations. Popular computer programs have made calculations easy and fast, but numerical results may mean nothing, because biases and uncertainties are often tremendous (Lombardo et al., 2014).
- We should be aware that

  real world processes ≠ *models*.
- In real world processes we should avoid false dichotomies such as

  deterministic vs. random

  and unjustified distinctions such as

  signal vs. noise.

Part 2: Important issues in stochastics

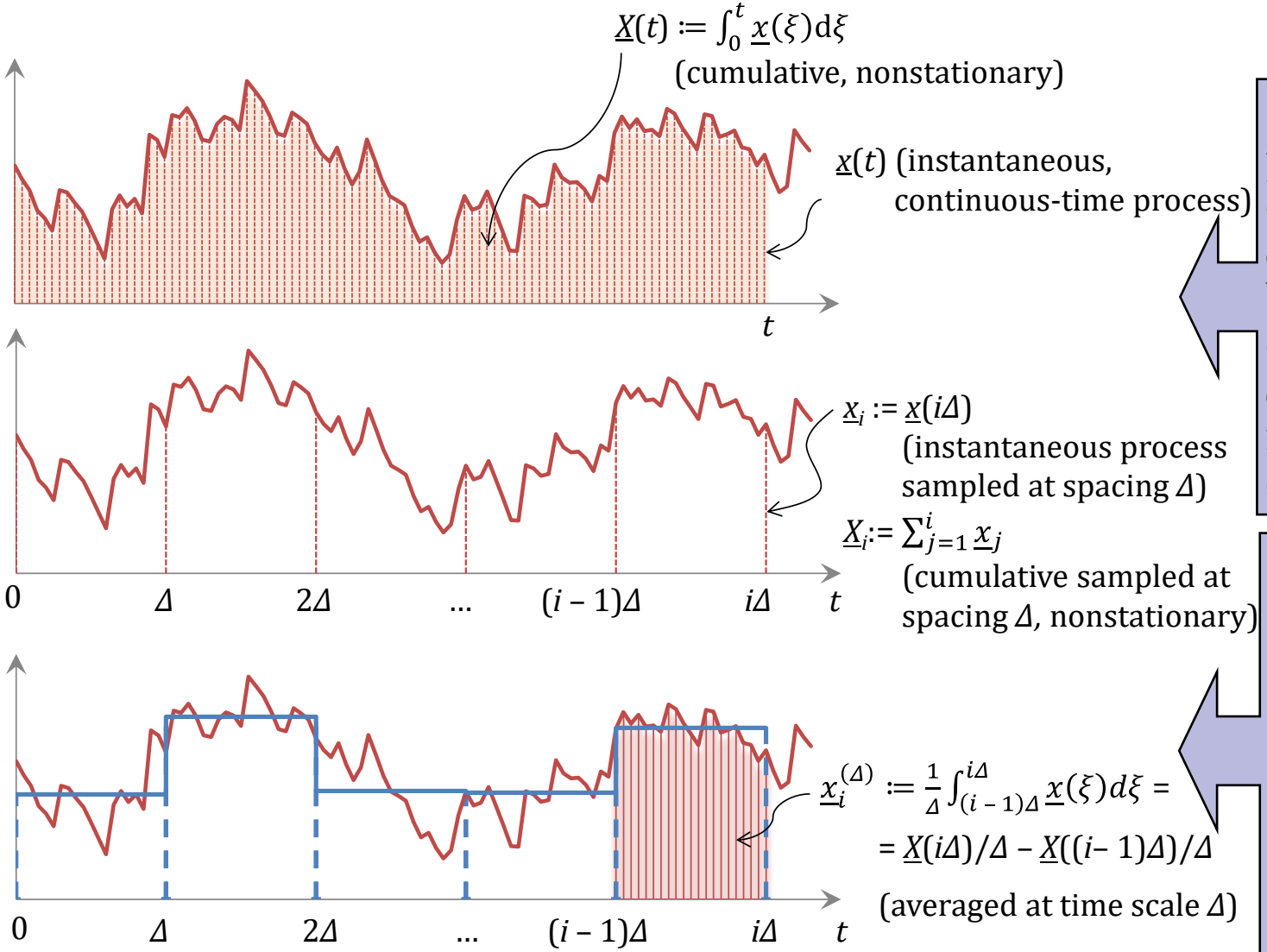
Fundamental concepts of stochastic processes

- Fundamental to stochastics is the concept of a random variable which should be distinguished from its realizations.
- A *random variable* is not a regular variable, while "random" means uncertain, unpredictable, unknown.
- While a regular variable takes on one value at a time, a random variable is a more abstract mathematical object that takes on all its possible values at once, but not necessarily in a uniform manner; therefore a distribution function $F(x)$ should always be associated with a random variable.
- A random variable needs a special notation to distinguish it from a regular variable x; the best notation devised is the so-called Dutch convention (Hemelrijk, 1966), according to which random variables are underlined, i.e. $\underline{x}$.
- A *stochastic process* is a family of infinitely many random variables indexed by a (regular) variable. The index typically represents time and is either a real number, t, in a continuous-time stochastic process $\underline{x}(t)$, or an integer, i, in a discrete-time stochastic process $\underline{x}_i$.
- Realizations, x_i, of a stochastic process, $\underline{x}_i$ or $\underline{x}(t)$, at a finite set of discrete time instances i (or t_i) are called *time series*.

Stationarity and ergodicity in stochastic processes

- Central to the notion of a stochastic process are the concepts of *stationarity* and *nonstationarity*, two widely misunderstood and misused concepts, whose definitions are only possible for (and applies only to) stochastic processes (thus, for example, a time series cannot be stationary, nor nonstationary).
- A process is called (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 also Koutsoyiannis and Montanari, 2015).
- 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.
- A nonstationary process should be handled theoretically (on the basis of deduction) rather than empirically.
- Another also misused concept is that of *ergodicity* (see definition in Papoulis, 1991). If a process is non-ergodic, then its statistics cannot be estimated from time series.
- For most applications, stationarity and ergodicity entail one another.
- Ironically, numerous studies claiming nonstationarity based on data analyses, use stochastic tools that are meaningful only for stationary and ergodic processes.
- Claiming, handling, or detecting nonstationarity needs to be based on deduction; doing those merely from data may be difficult, if not impossible.

From continuous time to discrete time processes



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*), $h(\tau) := \frac{1}{2}\mathrm{Var}[\underline{x}(t) - \underline{x}(t+\tau)]$.
- **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 uncertainty (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 of 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 (2013, 2016).

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

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

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

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

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:

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

  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):

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

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

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

- 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:

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

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:

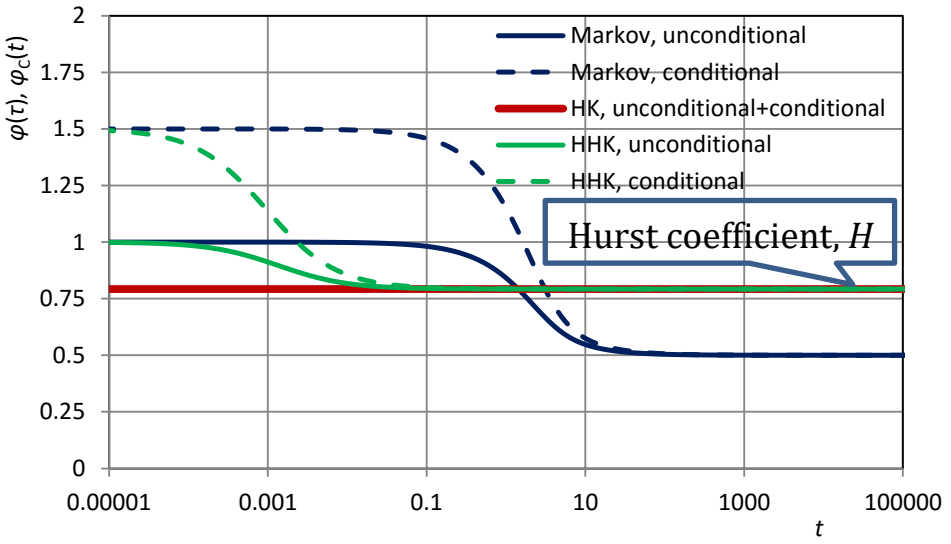
$$\gamma(\Delta) = \lambda(\alpha/\Delta)^{2-2H} \qquad (14)$$

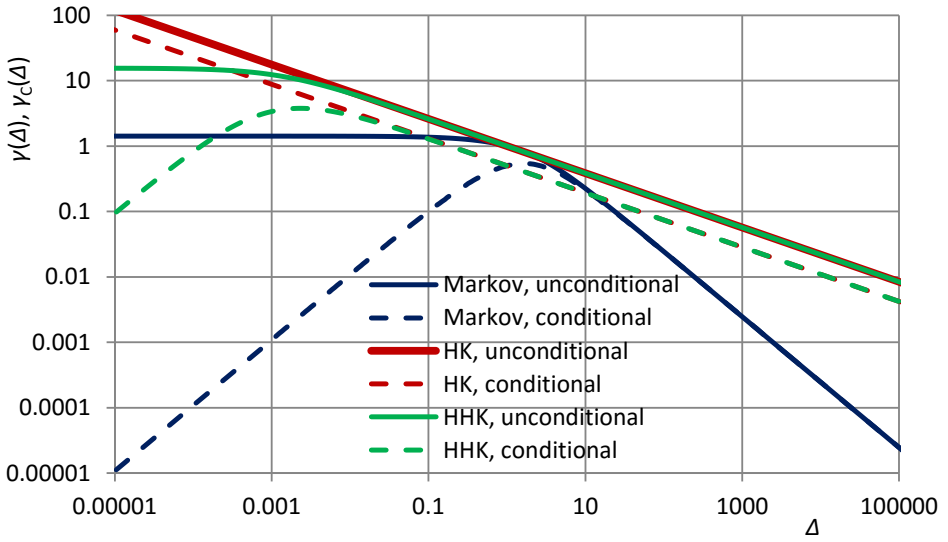
maximizes entropy production for large times but minimizes it for small times

- A Hybrid Hurst Kolmogorov process

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

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

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

- 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:

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

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

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:

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

- 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:

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

- 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:

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

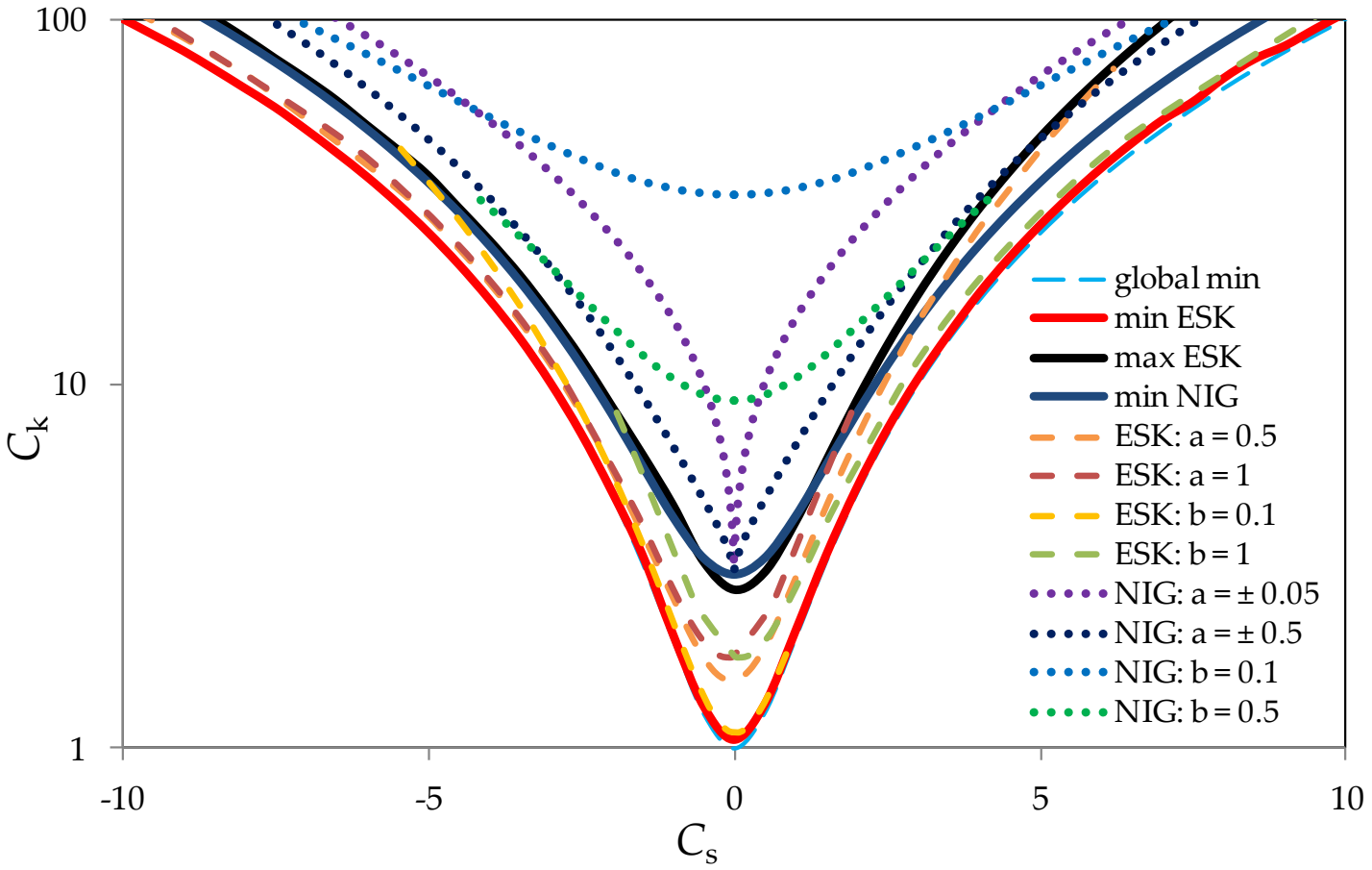
- For heavy-tailed distributions we can use the Normal-Inverse Gaussian (NIG) with probability density:

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

  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.

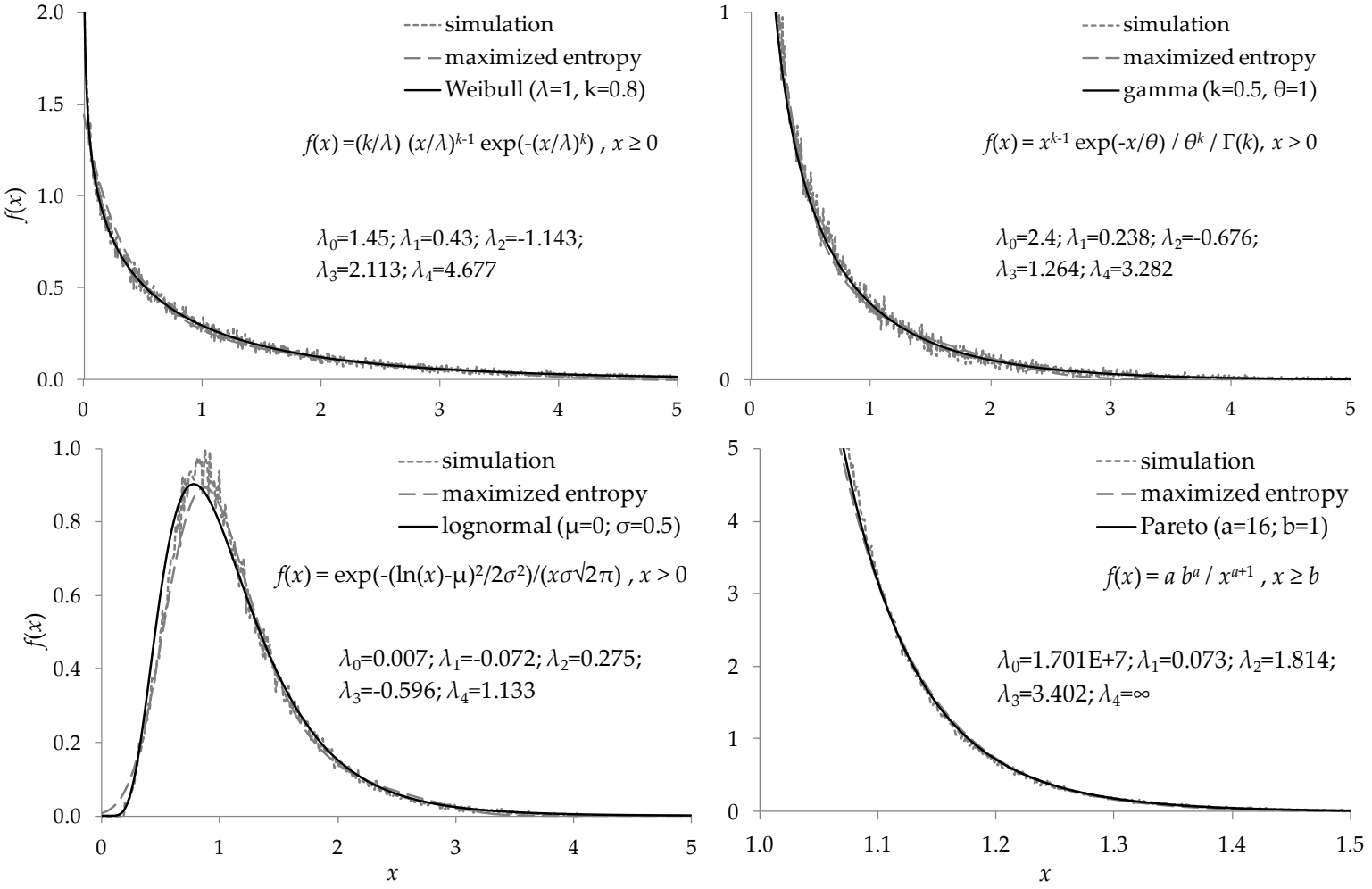
- In both cases v is the value of the random variable, a and b are dimensionless shape parameters, c is location parameter and d scale parameter; c and d have same dimensions as v (see details in Dimitriadis and Koutsoyiannis, 2016).

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 $n = 10^5$.

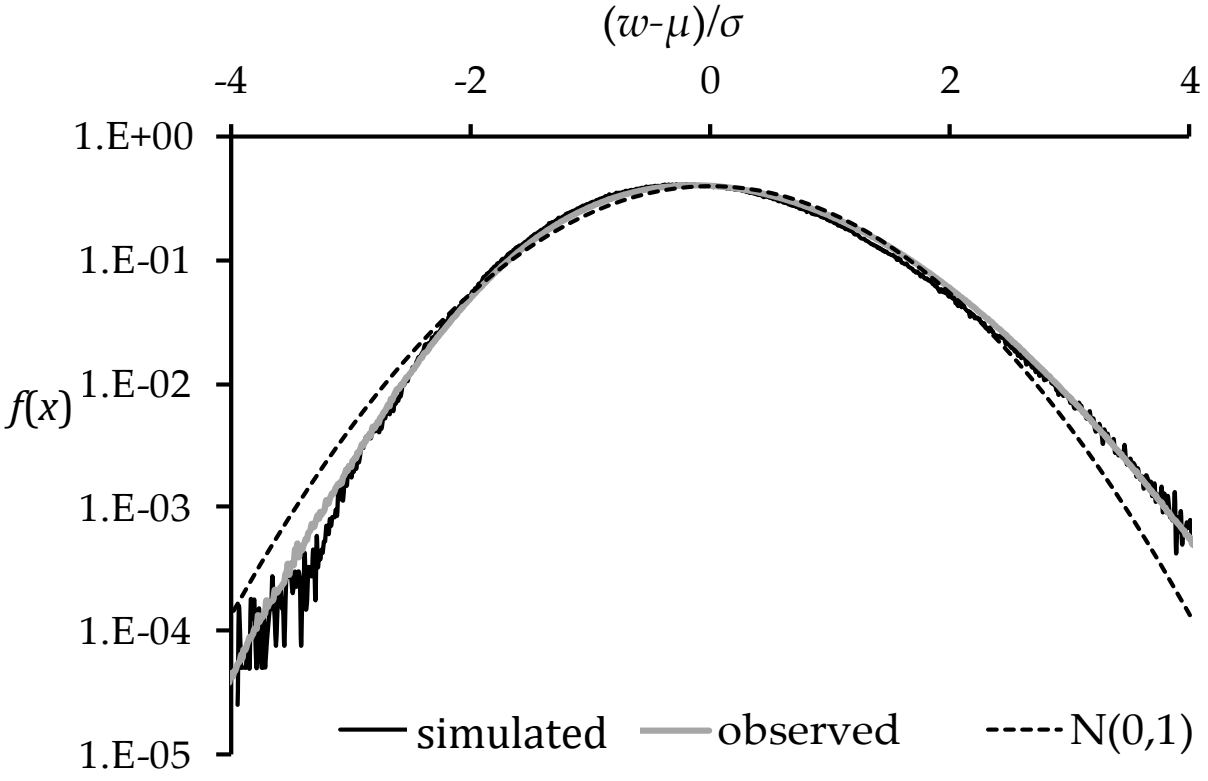
Part 4: Applications

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 $n = 36 \times 10^6$ 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 $40 \times 36 \times 10^6 = 1.44 \times 10^9$ 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**!)

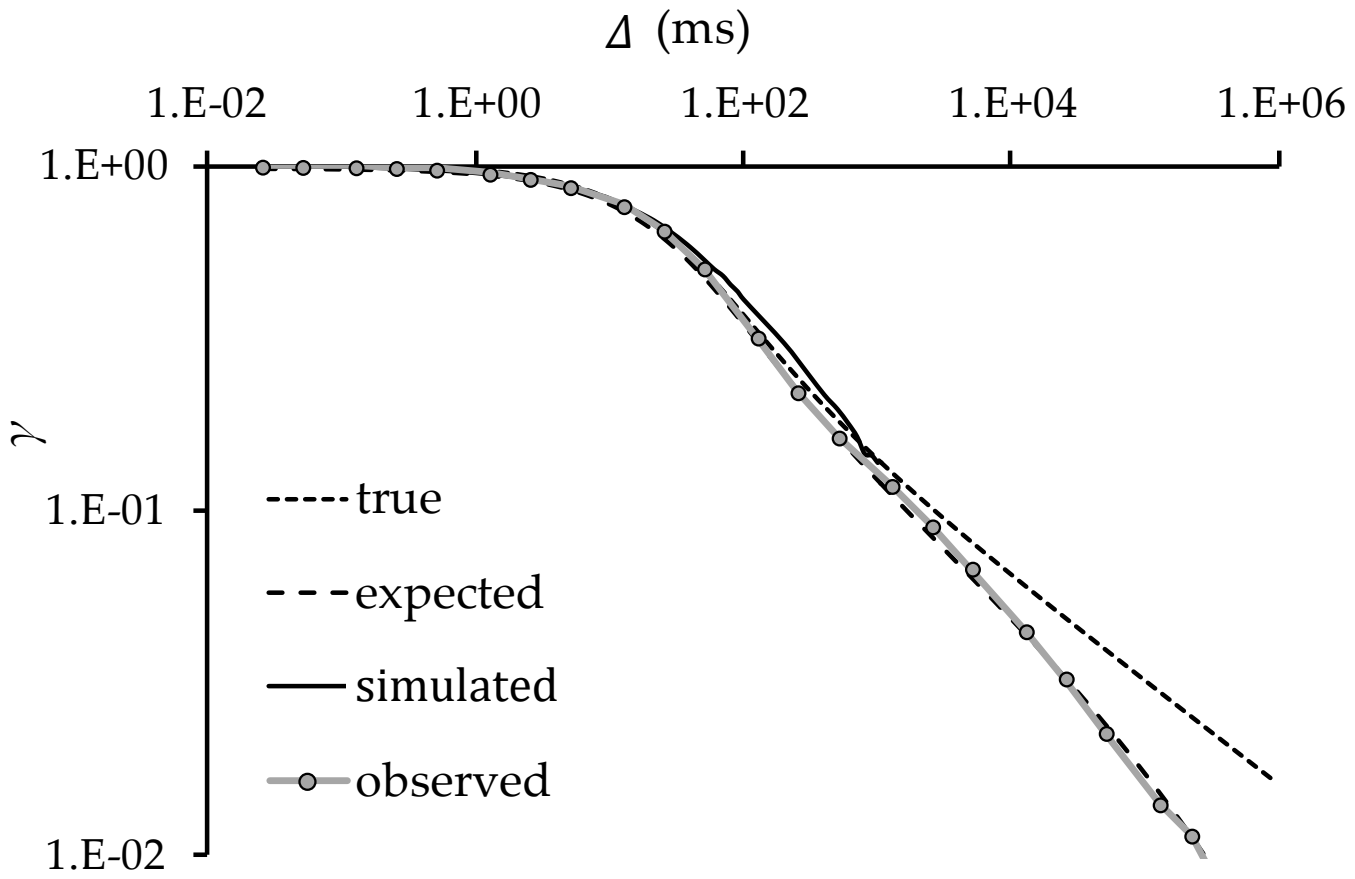


Probability density function of the mean standardized time series of turbulent velocity compared to that of a single simulation using the SMA scheme preserving the first four moments; the standard normal distribution N(0,1) is also shown.

Stochastic dependence of the turbulent velocity process

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

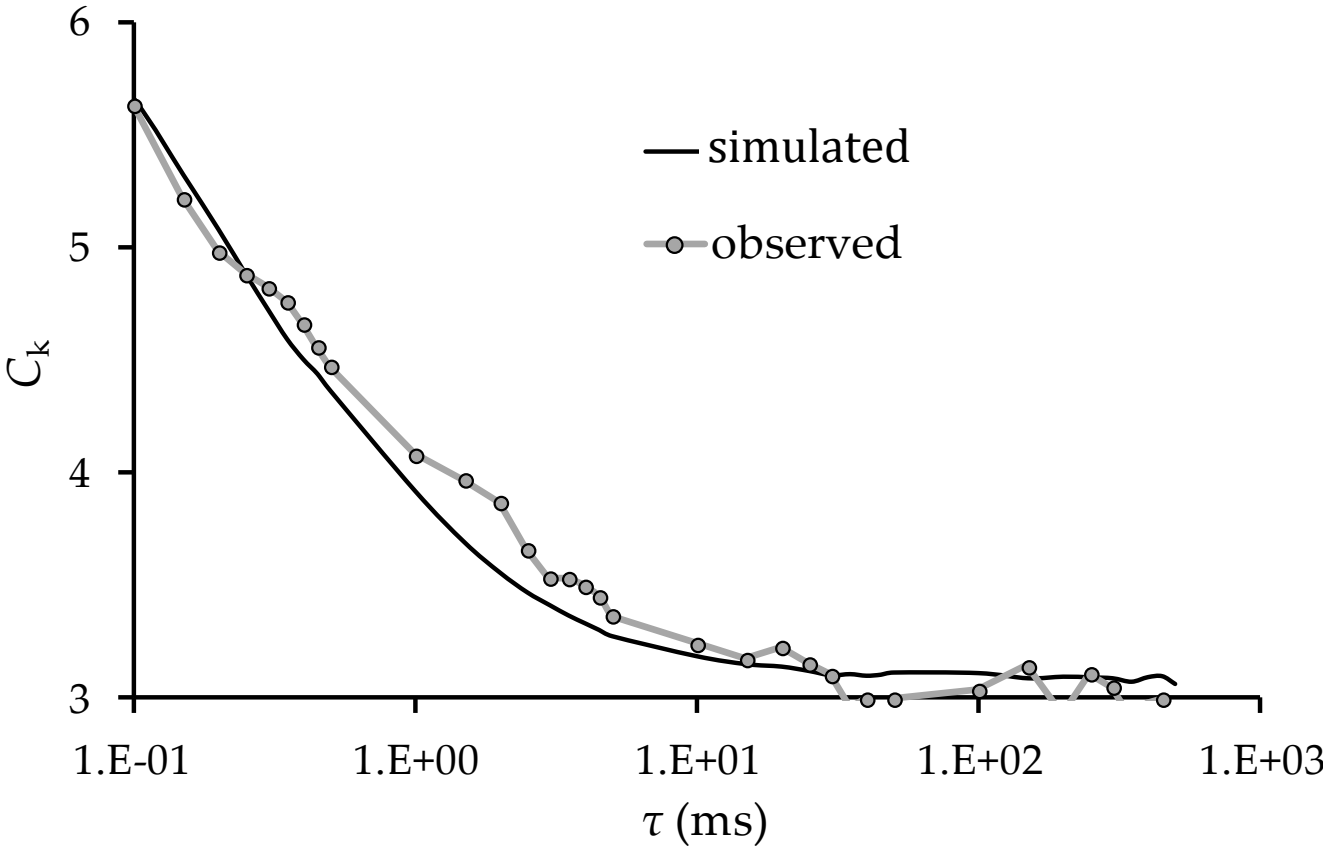
$$\gamma(\Delta) = \frac{\lambda}{2}\left(1 + (\Delta/\alpha_1)^{2\kappa}\right)^{\frac{H-1}{\kappa}} + \frac{\lambda}{\Delta/\alpha_2}\left(1 - \frac{1 - \mathrm{e}^{-\Delta/\alpha_2}}{\Delta/\alpha_2}\right) \qquad (26)$$



Climacogram of the turbulent velocity process (observed is the average from the 40 time series); the five parameters of the model are estimated as: $\lambda = 1.017$, $\alpha_1 = 10$ ms, $\alpha_2 = 15$ ms, $\kappa = 0.4$, $H = 0.85$.

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.

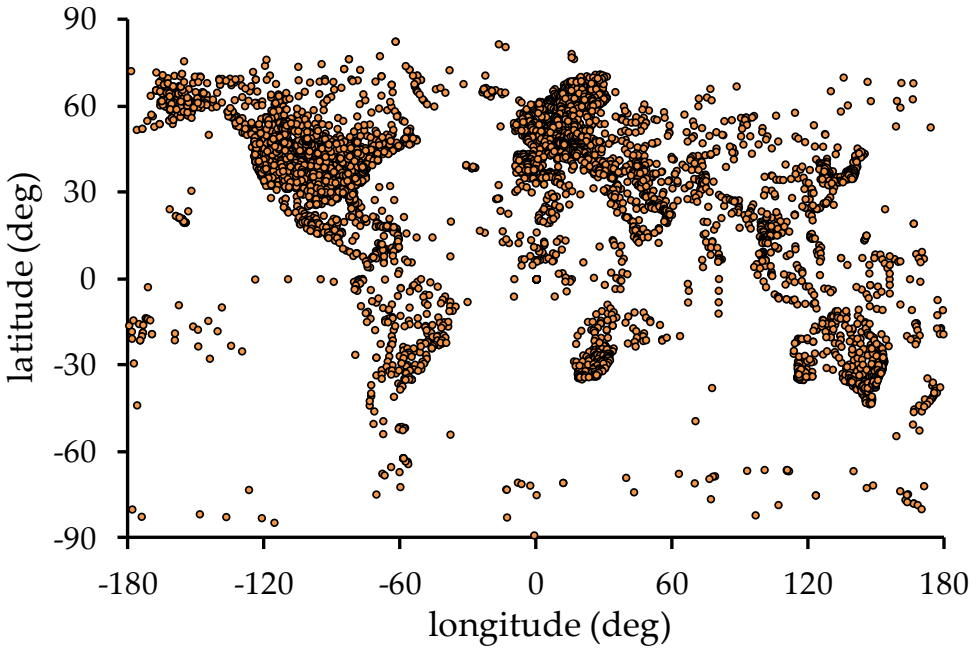


Empirical and simulated kurtosis vs. lag.

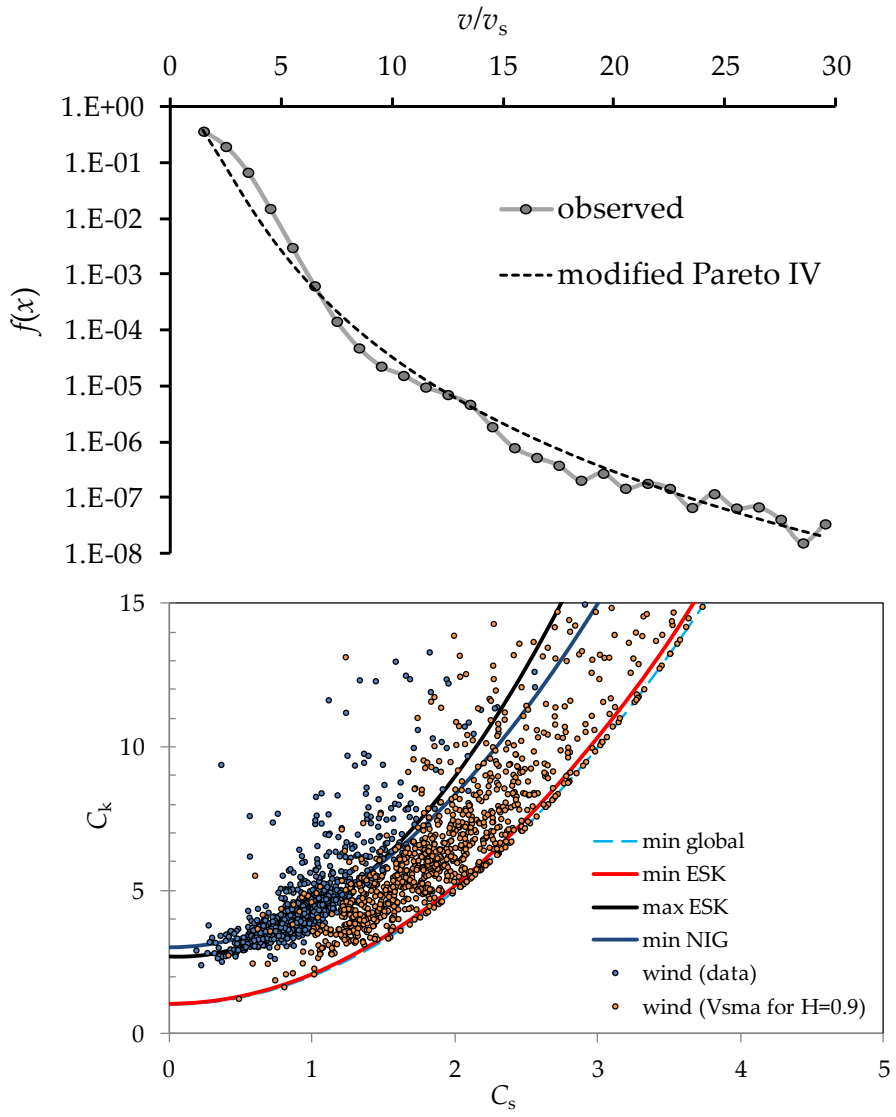
Not a mystery to have **large kurtosis** (**> 5** here) in velocity increments, even though the velocity is almost normal (**kurtosis = 3.08**)

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 $\sim 10^9$ 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 $n = 3 \times 10^5$ values.



Marginal distribution



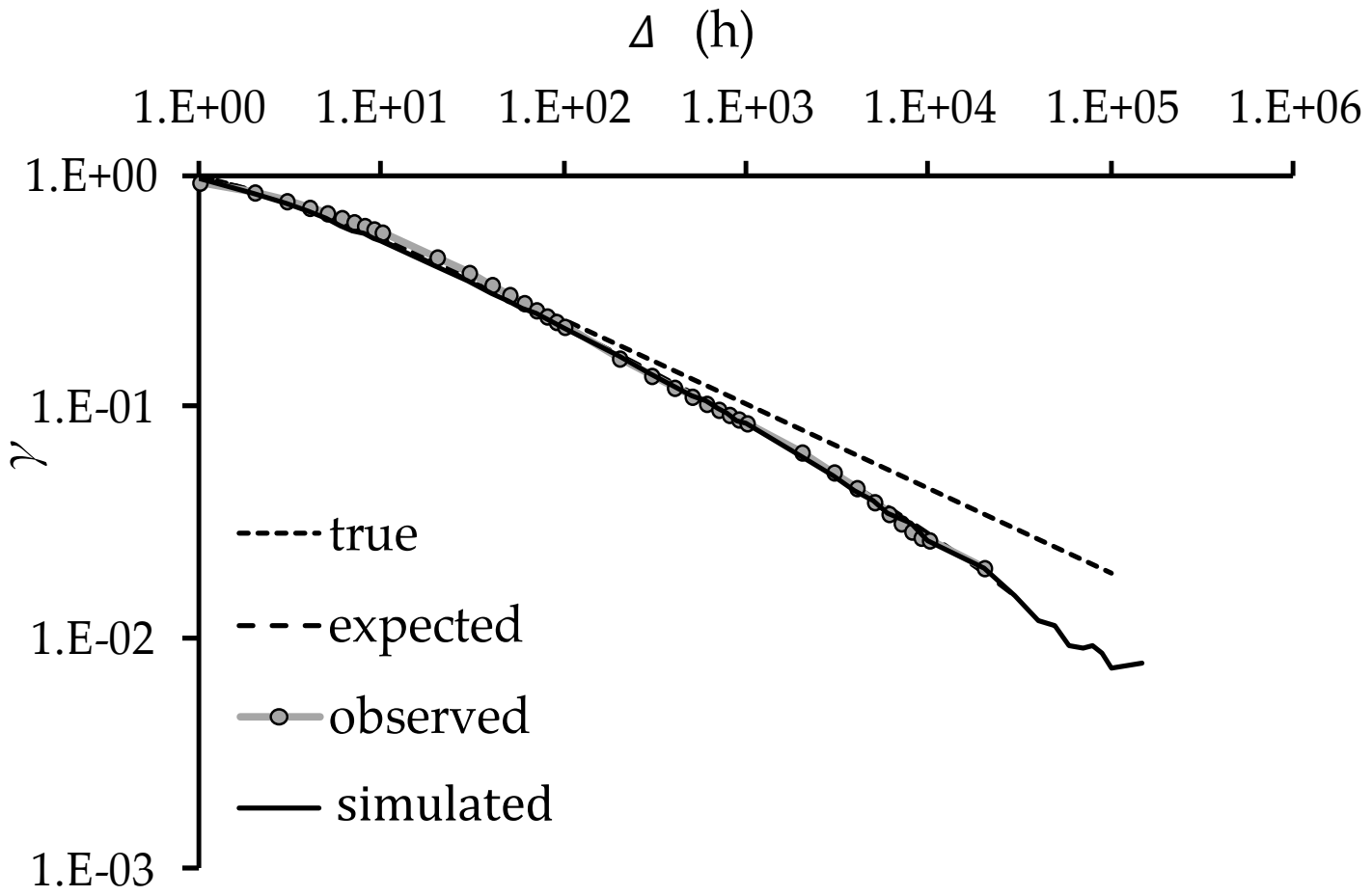
Wind speed distribution (from ~10^9 values):

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

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

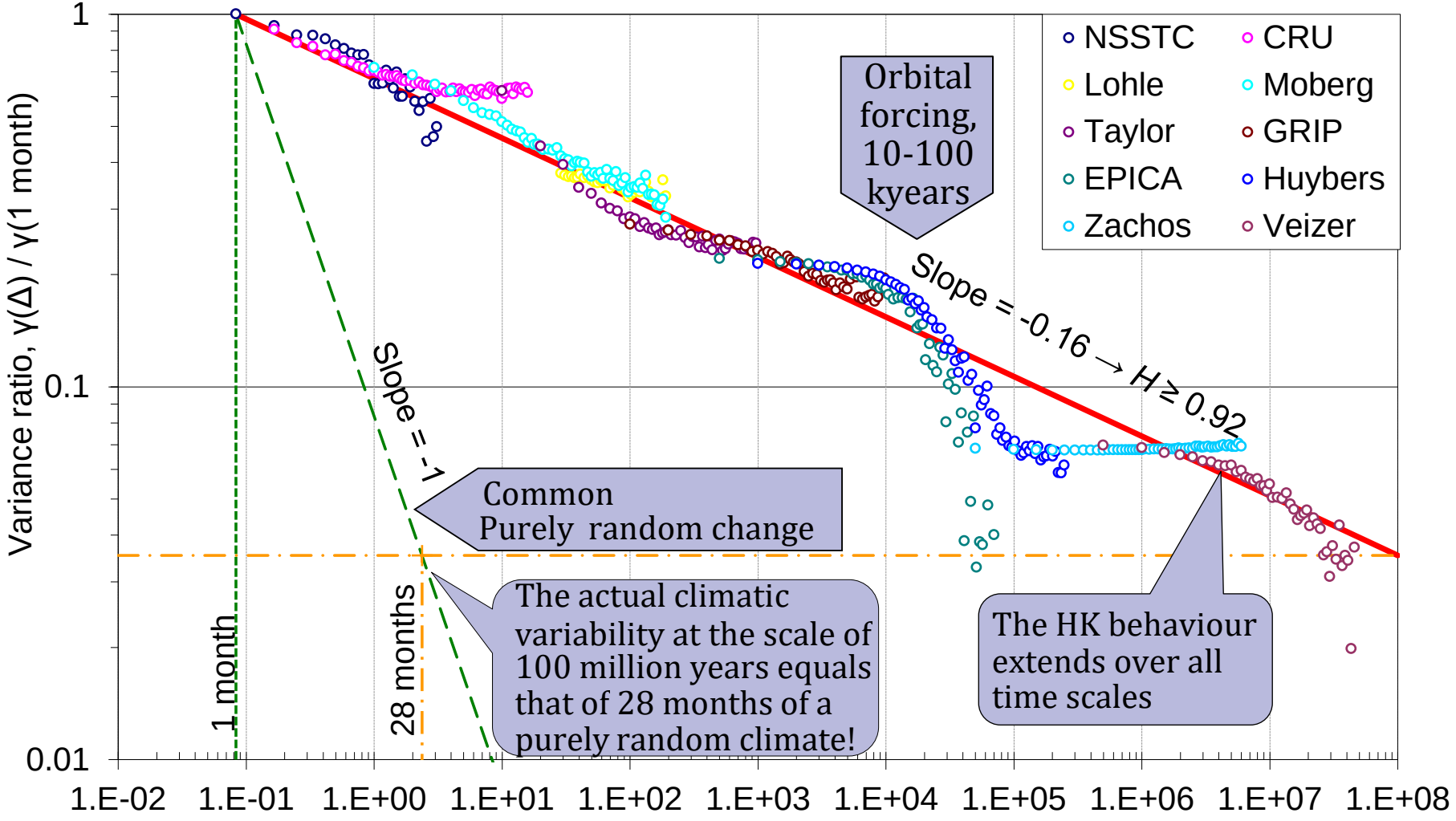
Sample skewness and kurtosis coefficients of 1000 hourly wind stations as well as of the corresponding white noise process of the SMA model.

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.

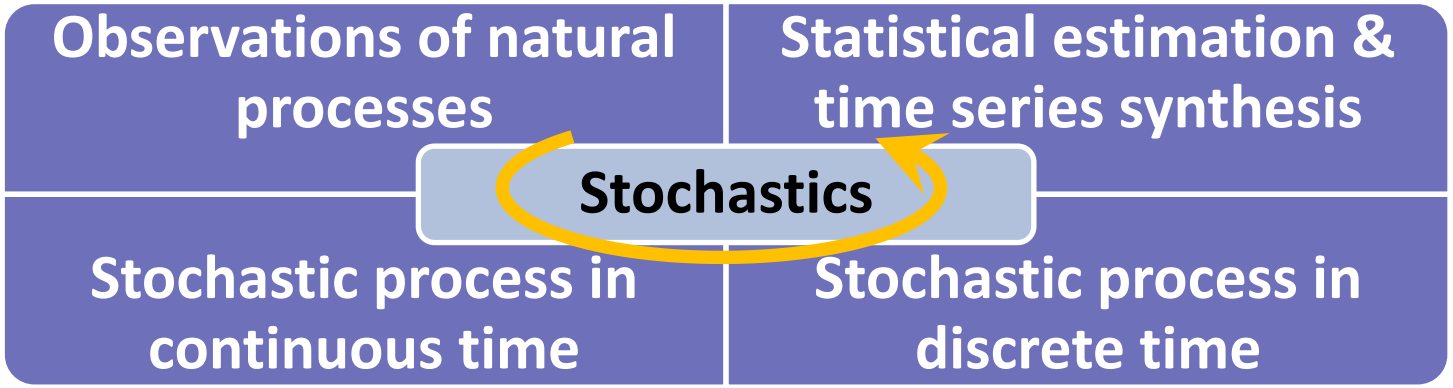
Application 3: Megascale (temperature)



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

Epilogue

- Natural processes evolve in continuous time but can be observed in discrete time.
- Observations cannot be handled unless we construct a model of the process.
- Stochastic processes in continuous time offer a strong basis for modelling and interpretation of natural behaviours.
- Calculating values of sample statistics without considering their statistical properties (bias and uncertainty) can yield misleading results.
- A general methodology for construction of synthetic time series is possible provided that we have a good understanding of stochastics.
- Thanks to Andrey Kolmogorov, we have a well-founded mathematical theory of stochastics.



References

Bachelier, M.L. (1900), Théorie de la speculation, *Annales scientifiques de l'École Normale Supérieure*, 17, 21–86.

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.

Box, G. E., and Jenkins, G. M. (1970), *Time Series Analysis, Forecasting and Control*, Holden-Day, 553 pages, San Francisco, USA.

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. (2016), Stochastic synthesis approximating any process dependence and distribution, in preparation.

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

Hemelrijk, J. (1966), Underlining random variables, *Statistica Neerlandica*, 20 (1), 1–7. doi: 10.1111/j.1467-9574.1966.tb00488.x.

Hosking, J. R. M. (1981), Fractional differencing, *Biometrika*, 68 (1), 165–176. doi: 10.1093/biomet/68.1.165.

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.

Kendall, M. G., and Stuart, A. (1966), *The Advanced Theory of Statistics*, *vol. 3: Design and Analysis, and Time-series*, 552 pp., Griffin, London, UK.

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, 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).

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

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

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., and Montanari, A. (2015), Negligent killing of scientific concepts: the stationarity case, *Hydrological Sciences Journal*, 60 (7-8), 1174–1183, doi:10.1080/02626667.2014.959959.

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.

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.

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

Whittle, P. (1951), Hypothesis testing in times series analysis. PhD thesis, Uppsala: Almqvist & Wiksells Boktryckeri AB.

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.