Preprint of a chapter in *Advances in Nonlinear Geosciences – A Springer publication*
The final document is available at: https://doi.org/10.1007/978-3-319-58895-7_14

From fractals to stochastics: Seeking theoretical consistency in analysis of geophysical data

Demetris Koutsoyiannis*,†, Panayiotis Dimitriadis*, Federico Lombardo‡,
Spencer Stevens§

Abstract
Fractal-based techniques have opened new avenues in the analysis of geophysical data. On the other hand, there is often a lack of appreciation of both the statistical uncertainty in the results, and the theoretical properties of the stochastic concepts associated with these techniques. Several examples are presented which illustrate suspect results of fractal techniques. It is proposed that concepts used in fractal analyses are stochastic concepts and the fractal techniques can readily be incorporated into the theory of stochastic processes. This would be beneficial in studying biases and uncertainties of results in a theoretically consistent framework, and in avoiding unfounded conclusions. In this respect, a general methodology for theoretically justified stochastic processes, which evolve in continuous time and stem from maximum entropy production considerations, is proposed. Some important modelling issues are discussed with focus on model identification and fitting, often made using inappropriate methods. The theoretical framework is applied to several processes, including turbulent velocities measured every several microseconds, and wind and temperature measurements. The applications shows that several peculiar behaviours observed in these processes are easily explained and reproduced by stochastic techniques.

> *I regard intuition and imagination as immensely important: we need them to invent a theory. But intuition, just because it may persuade and convince us of the truth of what we have intuited, may badly mislead us: it is an invaluable helper, but also a dangerous helper, for it tends to make us uncritical. We must always meet it with respect, with gratitude, and with an effort to be severely critical of it*. (Karl Popper, preface to "The Open Universe: An Argument for Indeterminism", 1982).

1. Introduction

Over the past 30 years or more, considerable literature highlighted the fractal (self-similar, self-affine, multifractal) characteristics of many of complex patterns that characterize geophysical processes. Fractal literature provides a framework in which a

* Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Heroon Polytechneiou 5, 157 80 Zographou, Greece.

† Corresponding author: dk@ntua.gr

‡ Dipartimento di Ingegneria Civile, Edile e Ambientale, Sapienza Università di Roma, Via Eudossiana, 18, 00184, Rome, Italy.

§ Independent researcher, London, UK.

simple process, involving a basic operation repeated many times, can represent natural patterns that can be of extraordinary complexity (Falconer, 2014; Scholz and Mandelbrot, 1989). In a variety of applications, geophysical systems are viewed as fractals that follow certain scaling rules over a broad (even unlimited) range of scales, implying that the degree of their irregularity and/or fragmentation is identical at all those scales. Mathematically, these rules are power laws with exponents being related to a fractal dimension. Roughly speaking, the fractal dimension is a measure of the prominence of complexity of a pattern when viewed at very small scales. Therefore, the fractal dimension is originally a local property, notwithstanding the fact that in fractal literature the local properties are reflected in the global ones (Mandelbrot, 1982).

Finding that a complex system is characterized by fractal (or multifractal) behaviour with particular scaling exponents represents a desideratum for many practicing geophysicists and engineers (von Karman, 1940), because this finding will help in describing the system dynamics with very simple formulae and few parameters, in order to obtain predictions on the future behaviour of the system. Such dynamics is usually denoted as fractal or multifractal, depending on whether it is characterized by one scaling exponent or by a multitude of scaling exponents.

However, if we agree that scientific theories are mental constructs rather than the physical reality per se, then we should also agree that there are no true fractals in nature. Although there are natural phenomena that have been explained in terms of fractal mathematics, “natural fractals” (such as coastlines, turbulence in fluids, cloud boundaries, etc.) can usefully be regarded as such only over an appropriate range of scales, with the fractal description inevitably ceasing to be valid if they are viewed out of this range of scales.

Since asymptotic properties of geophysical processes are crucial for the quantification of future uncertainty, as well as for planning and design purposes, many applications of fractal theory tend to be descriptive rather than predictive (Falconer, 2014; Kantelhardt, 2009). In the foundational treatise on fractals, Mandelbrot (1982) made such a distinction clear, but it has become somewhat blurred in recent literature.

We maintain and show in the following that careful use of stochastics (which includes probability theory, statistics and stochastic processes) can deal with all problems about complex geophysical processes in a more rigorous manner and more effectively than fractals can do.

2. Why not to prefer fractals over stochastics

In spite of the difficulty even mathematicians have in formally defining fractals (Falconer, 2014; Mandelbrot, 1982), their wide popularity stems from the concept of symmetry—in particular, expanding symmetry. From the birth of science and philosophy, symmetry has been closely related to harmony and beauty, and this was to prove decisive for its role in theories of nature. Both ancients and moderns often believed indeed that there is a close association in mathematics between beauty and truth.

A common research theme in the study of complex systems is the pursuit of universal properties that transcend specific system details. In this way, fractal-based techniques have opened new avenues in the analysis of geophysical data. According to Scholz and Mandelbrot (1989):

> *One possible broad explanation of the role of fractals in geophysics may be found in probabilistic limit theorems, and in the existence of classical "universality classes" related to them. The reason is illustrated by the following fact. Wiener's scalar Brownian motion process W(t) is the limit of the linearly rescaled random processes that belong to its very wide domain of attraction. Therefore, it is itself the fixed point of the rescaling process. That is, its graph is a self-affine fractal set, a curve. The argument suggests that geometric shapes relative to probabilistic limit theorems can be expected to be fractal sets.*

On the other hand, the concept of fractals has been closely associated from the outset with mathematical constructions involving infinite operations on simple, deterministically defined, objects. Simple nonlinear dynamical systems were also enrolled in illustrating the emergence of fractal structures. This association with determinism and simplicity has been prominent and shaped the evolution of the fractal literature.

Even when studying more complex systems, such as the evolution of geophysical processes, the intuitive zeal was to make them comply with the simplicity of the archetypal fractal mathematical objects. Thus, several studies attempted to demonstrate that irregular fluctuations observed in natural processes are *au fond* manifestations of underlying deterministic dynamics with low dimensionality, hence rendering probabilistic descriptions unnecessary. If we assume, for example, that the evolutions of all temporal and spatial patterns of geophysics result from deterministic chaos, then we may derive the underlying deterministic rules on the basis of their strange attractors, which have a fractal structure (Grassberger and Procaccia, 1983). However, such an approach is questionable in geophysics (Koutsoyiannis, 2006).

The opposite reading of the same finding would be more sensible, in our view. Specifically, if simple underlying dynamics can produce irregular fluctuations and eventually, unpredictable trajectories, then, *a fortiori*, more complex systems are even more unpredictable. In this line of thought, Koutsoyiannis (2010b) used a caricature geophysical system, which is low dimensional deterministic by construction, and showed that we cannot get rid of uncertainty. Hence, probability theory and its extension, stochastics, become absolutely necessary even for the simplest systems. This argument may also be used in order to criticize the determinist point of view that probability considerations enter into science only if our knowledge is insufficient to enable us to make predictions with certainty (Popper, 1982).

Stochastics has its own rules of calculations and estimations, which go far beyond classical calculus in order to deal with uncertain quantities represented as random variables and stochastic processes. Fractal studies often fail to appreciate this and apply algorithms referring to uncertain quantities with standard mathematical calculations.

They do so even when using stochastic concepts, such as statistical moments, (auto)correlations and power spectra. Thus, they produce results which not only fail to recognize the statistical uncertainty but may be fundamentally flawed, i.e. inconsistent with theory. In the subsections below, we summarize some of the problems often characterizing fractal studies which make us advocate the dedication to proper theoretical concepts, offered by the theory of stochastics.

Ambiguity

Even the very terms *fractal* and *multifractal* remain without an agreed mathematical definition. This is a severe drawback, as without proper definitions we cannot build a scientific theory. The importance of definitions in science has been emphasized in the following philosophical note by the great Russian mathematician Nikolai Luzin:

> *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* (from Graham and Kantor, 2009).

This is not the case with fractals. Instead, fractals are usually identified intuitively; for example, Falconer (2014) refers to a set *F* as a fractal, when:

> *(i) F has a fine structure, i.e. detail on arbitrary small scales;*
> *(ii) F is too irregular to be described in traditional geometrical language, both locally and globally;*
> *(iii) F has some form of self-similarity, perhaps approximate or statistical;*
> *(iv) usually, the "fractal dimension" of F (defined in some way) is greater than its topological dimension;*
> *(v) in most cases of interest, F is defined in a very simple way, perhaps recursively.*

Mandelbrot, who coined the term *fractal* in 1975, tried to theorize about the absence of a definition, arguing just opposite of Luzin:

> *Let me argue that this situation ought not create concern and steal time from useful work. Entire fields of mathematics thrive for centuries with a clear but evolving self-image, and nothing resembling a definition* (Mandelbrot, 1999, p. 14).

One may indeed recall cases where mathematical concepts did not have proper definitions for centuries; *probability* is a characteristic example. However, the expression "*nothing resembling a definition*" may be a gross exaggeration: In the example of probability there never was lack of definitions; the problem was that the definitions were problematic (e.g., suffering from circular logic, like in the previous sentence). Once Kolmogorov (1933) gave a proper definition to probability, he opened new avenues. Certainly, absence of a definition entails domination of intuition over logic, dark over light, or uncritical acting over critical thinking (cf. the excerpt by Luzin above and that by Popper in the opening motto of the paper).

Nevertheless, Mandelbrot's aversion from defining concepts, which he does not regard as "*useful work*" to do, has influenced the entire field of fractals. Even in cases where clear definitions exist, Mandelbrot encourages neglecting them and preferring intuitive notions. The following excerpt provides an example for the well-defined concept of stationarity, which is central in stochastics (see Koutsoyiannis and Montanari, 2016):

> *[Mandelbrot, 1982] observes that "Ordinary words used in scientific discourse combine (a) diverse intuitive meanings, dependent on the user, and (b) formal definitions, each of which singles out one special meaning and enshrines it mathematically. The terms stationary and ergodic are fortunate in that mathematicians agree on them. However, experience indicates that many engineers, physicists, and practical statisticians pay lip service to the mathematical definition, but hold narrower views." That is, many mathematically stationary processes are not intuitively stationary. By and large, those processes exemplify wild randomness, a circumstance that provides genuine justification for distinguishing a narrower and a wider view of stationarity* (Mandelbrot, 1999, p.7).

Even when Mandelbrot attempts to provide a definition for the central concept of a *multifractal*, he bases that definition on the intuitive concept of a "*multibox cartoon*":

> *Definition. The term multifractal denotes the most general category of multibox cartoons. It allows the generator to combine axial boxes and diagonal boxes with non-identical values of* H_i *from* $H_{\min} > 0$ *to* $H_{\max} < \infty$ (Mandelbrot, 1999, p. 45; see section 3 below about the meaning of H).

The ambiguity does not concern merely definitions. "*Peaceful coexistence*" of different numerical values for the same mathematical concept has also been advocated:

> *We are done now with explaining the peaceful coexistence of two values of D: the dimension D = 1/H = 2 applies to that three-dimensional curve, as well as to the trail obtained by projecting on the plane (X, Y). However, the projections of the three dimensional curve on the planes (t, X) and (t, Y) are of dimension D = 2 – H = 1.5* (Mandelbrot, 1999, p. 45).

In fact, when dealing with geophysical processes, one can easily get rid of ambiguity through stochastics. Careful use of stochastics can deal with all problems involving fractals of non-deterministic type in a more rigorous manner and more effectively.

Confusion between local and global properties of processes

Indeed, attempts to remove ambiguity based on stochastics are not rare, as indicated by the following excerpt:

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

Other researchers who seek for clarity also agree on this; for example Veneziano and Langousis (2010, p. 4) state that the most general and mathematically satisfactory definition of fractal dimension is the Hausdorff dimension. Here it is important to note that the Hausdorff dimension expresses a local property, an asymptotic measure as a radius δ for covering the set A tends to zero. This is more evident in the so-called *box-counting dimension*, which is an upper bound for D_{Haus} (Beran et al., 2013, p. 181-182) and is defined as $\dim_{\text{Box}} A = \lim_{\delta \to 0} \log N_\delta / \log \delta$ where N_δ is the minimal number of sets U_i needed for a δ-cover of A.

However, as in the fractal literature it is intuitively believed that the local properties repeat themselves at bigger and bigger scales, and given the general frame of ambiguity, the local properties have been confused with global ones, such as the long-range dependence. Indeed:

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

Even Mandelbrot (1999, p.3) referred to the difference of locality and globality, but in a rather obscure way:

> *The importance of the contrast between mildness and wildness is in part due to its links with a contrast between locality and globality.*

However, this was not enough to hinder the fractal literature from confusing fractal behaviour with long-range dependence.

Gneiting and Schlather (2004) were perhaps the first to clarify the issue and highlight the fact that fractal properties and long-range dependence are independent of each other. They used a process with Cauchy-type autocovariance function, which was first proposed by Yaglom (1987, p. 365) and also referred to by Wackernagel (1995, p. 219; 1998, p. 246), while a similar one was used by Koutsoyiannis (2000) in discrete time. Using this process, they demonstrated that the fractal and Hurst properties (long-range dependence) are two different things, independent to each other: The fractal parameter determines the local properties (the roughness) of the process (as time $t \to 0$) while the Hurst parameter determines the global properties of the process (as $t \to \infty$).

Use of the abstract mathematical objects as if they are natural objects

In mathematical processes the local and global properties can be the same. The obvious example is the Hurst-Kolmogorov (HK) process (see below), also known as fractional Gaussian noise (Mandelbrot and Van Ness, 1968), which is described by a single scaling exponent applicable to all scales. Scale independence or absence of characteristic scales in a process or a phenomenon is mathematically and intuitively attractive. Indeed, it would imply that simple physical dynamics could produce complex phenomena that exhibit startling similarities over all scales. However, in Nature complex phenomena are influenced by different mechanisms and agents, each one acting at a different

characteristic scale, and therefore absence of characteristic scales is only a dream. Besides, the assumption of absence of characteristic time scales would have consequences that would be absurd. Some examples follow:

- The speculation that rivers are fractals with fractal dimension > 1 (e.g. 1.2) has been very popular. However, if that were the case, it would mean that the number of sets of its δ-cover would be a power law of δ with exponent > 1 for arbitrary low δ. As a direct consequence, the geometrical length of the river would be infinite (a curve with dimension > 1 has infinite length; Falconer, 2014) and any particle of water would take infinite time to reach the sea.
- If a Hurst-Kolmogorov process (whose variance is a power law of time scale; equation (16) below) were applicable for arbitrary short time scales, it would entail infinite variance of the instantaneous, continuous-time process which would imply infinite energy.
- If an antipersistent Hurst-Kolmogorov process (with Hurst exponent $H < 0.5$; see below) were applicable for arbitrary short time scales, it would entail negative autocovariance (anti-correlation) for arbitrary small lags which is absurd. For in a natural process, the autocorrelation should tend to 1 as lag tends to 0.

All these paradoxes are easily resolved if we abandon the idea of absence of characteristic scales and admit that below (or above) a certain characteristic scale the respective power laws cease to hold.

Hasty use of stochastic concepts

Stochastic concepts such as statistical moments (marginal or joint, e.g. covariances), and spectral densities have been widely used in the fractal literature, usually by making calculations using data and at the same time ignoring the theoretical properties of those concepts. A typical example is the power spectrum, $s(w)$, where w denotes frequency (inverse time scale), and its log-log slope $s^{\#}(w)$. The latter represents the log-log derivative, which for any function $f(x)$ is defined as:

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

The HK process is used as a benchmark as it has a power spectrum with constant slope, i.e. $s(w) \propto w^{\beta}$, where the constant slope $\beta = s^{\#}(w)$ is related to the Hurst parameter H (equation (16) below) by $\beta = 1 - 2H$. The special case $H = 0.5$, which signifies the white noise, corresponds to $\beta = 0$, thus complying with the fact that the white noise spectrum is flat ($s(w)$ = constant). As $H \to 1$, which is the highest possible value, $\beta \to -1$, which is the lowest possible value for a stationary and ergodic process.

However, a huge number of studies exploring several data sets have reported steeper constant slopes, i.e. $\beta < -1$, also suggesting $H > 1$, which is absurd. Other studies assume that slopes $\beta < -1$ are theoretically consistent, also claiming that the particular value $\beta = -2$ corresponds to the power spectrum of the Brownian motion (the integral over time of white noise), which is a nonstationary process. This line of thought is extended

further, in the characterization of processes. Specifically, the power spectrum has been often regarded as a tool to identify whether a process is stationary or nonstationary: values $\beta > -1$ are thought to suggest a stationary process while values $\beta < -1$ are thought to confirm the nonstationarity of the process. The fact is, however, that the entire line of thought is theoretically inconsistent and such numerical results, usually reported, are artefacts due to insufficient data or inadequate estimation algorithms.

Before we describe the details for recovering from the incorrect application of the power spectrum, it would be informative to trace how incorrect results can appear. In the example of Figure 1, 1024 data points have been generated from a stationary stochastic process and the empirical power spectrum, calculated from these data, has been plotted. To apply some smoothing (as per Bartlett's (1948) method), the empirical power spectrum was constructed by averaging from 8 segments, in which the data were separated (since without smoothing the power spectrum would be exceptionally rough). The stochastic process has the theoretical power spectrum with the indicated varying slope (specifically, it is an HHK process, defined in equation (17) below, with parameters $M = 0.5$, $H = 0.8$, $\alpha = \lambda = 1$; see also Koutsoyiannis, 2014). On its right tail the power spectrum has an asymptotic slope of –2, which is not inconsistent nor does it indicate nonstationarity (actually, a right-tail slope of –2 is precisely the slope of a stationary Markov model; see below). In contrast, on its left tail the power spectrum has an asymptotic slope of –0.6, which is strictly > –1 (were it not, it would be inconsistent with theory, as will be detailed below).

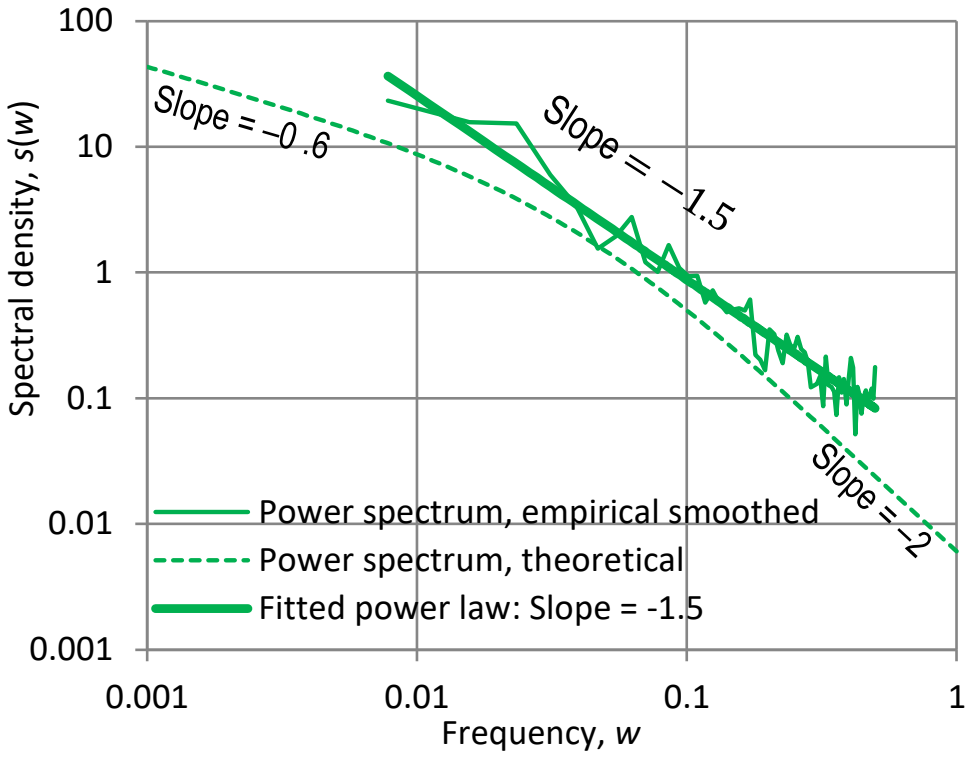

Figure 1 Illustration of inconsistent results derived by hasty use of the power spectrum.

From the shape of the theoretical power spectrum it can be imagined that if the time step and length of the data set were such that we could "see" only at frequencies > 0.1, then we would conclude that we have a constant slope of –2 and, if we followed the standard fractal line of thought, we would claim that the process is (nonstationary) Brownian motion. Of course, all these would be incorrect as the model is purely stationary and not at all related to Brownian motion.

Even with the given data set, which allows us to "see" frequencies much lower than 0.1 (by an order of magnitude or more), the empirical power spectrum may again mislead us. For, even after the aforementioned smoothing, the empirical power spectrum is too rough to recover the underlying model and its parameters. Furthermore, it involves high bias and it suggests a misleading constant slope of –1.5. Just knowing the theoretical properties, as well as the uncertainty and bias of the power spectrum as a stochastic tool, we would avoid making erroneous claims, even though it is doubtful if this would help us to identify the correct model (see Dimitriadis and Koutsoyiannis, 2015). Nonetheless, identifying the model from data and recovering the theoretically consistent asymptotic slopes (–0.6 and –2) are possible but need other methods (CS—see below).

The theoretical properties of the power spectrum which we need to know to avoid false claims include the following:

- Once we make the power spectrum of a process as a function of frequency, we have tacitly assumed a stationary process. In a nonstationary process, both the autocovariance and the spectral density, i.e. the Fourier transform of the autocovariance, are functions of two variables, one being related to "absolute" time (see e.g. Dechant and Lutz, 2015). Thus, there is no meaning in using a stationary representation (setting the power spectrum as a function of frequency only) and, at the same time, claiming nonstationarity. Even though this tactic has been very common, it is inconsistent. Furthermore, we should be aware that the customary Wiener-Khinchin theorem relating autocovariance and power spectrum pertains to stationary processes. This theoretical knowledge will prevent us from making claims of nonstationarity while using formulations and tools pertaining to stationary stochastic processes. In addition, we should be aware that claiming nonstationarity based solely on inductive reasoning is absurd (Koutsoyiannis and Montanari, 2015).
- Once we use the power spectrum of a process for inference, as we always do, we should be aware that inference from data is only possible when the process is ergodic. As shown in the Appendix A, in an ergodic process, the asymptotic slope on the left tail of the power spectrum cannot be steeper than –1. Thus, there is no meaning in reporting slopes in empirical power spectra $s^{\#} < -1$ (e.g. $s^{\#} = -1.5$, as in the example of Figure 1) and at the same time making any claim about the process properties (e.g. of nonstationarity) based on the power spectrum. Actually, such a steep slope, when emerging from processing of data, does not suggest that a process is non-ergodic, it rather identifies inconsistent estimation.

- We should be aware of the close relationship of ergodicity and stationarity (Koutsoyiannis and Montanari, 2015). In particular, a nonstationary process is nonergodic and thus any estimates from data (including those of the power spectrum) are meaningless when we claim nonstationarity.
- As a result of the above listed theoretical points, constant slopes $\beta < -1$ of the power spectrum are invalid and indicate either inadequate length of data or inconsistent estimation algorithm. Likewise, non-constant slopes of power spectrum steeper than –1 ($s^{\#}(w) < -1$) for small frequencies ($w \to 0$) are equally invalid. We note that steep slopes ($s^{\#}(w) < -1$) are mathematically and physically possible for medium and large w—actually they are quite frequent in geophysical processes (see also Koutsoyiannis, 2013a,b; Koutsoyiannis et al., 2013; Dimitriadis and Koutsoyiannis, 2015).

Misspecification / misinterpretation of scaling laws

The applicability of fractal analyses to complex phenomena of the real world essentially relies on the empirical detection of power-law relationships in observational data. Therefore, such analyses heavily rely on available data series and their statistical processing; and since they ask statistical questions, they must rely on probability theory (Stumpf and Porter, 2012).

However, as the inference from data obeys statistical laws and is affected by statistical uncertainty and bias, we should respect these laws in making inference. Some examples can demonstrate that such respect is often not paid in fractal studies. The interested reader could perform a Google search with related terms (e.g. *universal multifractal rainfall*—see also Koutsoyiannis, 2010a) and several studies will be listed that identify multifractal behaviour of rainfall. This is usually done in terms of scaling relationships between raw moments of the averaged process $\underline{x}^{(k)}$ at time scale k, i.e. $\mathrm{E}[(\underline{x}^{(k)})^q]$ (or inverse time scale $\lambda := 1/k$), for several orders of moments q. Such scaling relationships are graphically identified on log-log plots and then the relationship of the scaling exponent (slope) K as a function q (the function $K(q)$) is empirically constructed (even though, according to universal multifractals, there exists a theoretical model for $K(q)$ that one can fit to empirical data; cf. eqn. (2.12) in Tessier et al., 1993).

A graphical example is provided in Figure 2 to illustrate that the entire procedure is problematic from the outset. A time series with length $N = 2^{13} = 8192$ was generated from the HK process with Hurst parameter $H = 0.8$ and Gaussian distribution $N(1,1)$. Some scaling laws seem to appear at a range of time scales. One could be led to assume a multifractal behaviour and specify a $K(q)$ function. All these, however, are spurious. The truth is that there is no multifractal behaviour here. As shown theoretically by Lombardo et al. (2014) for $q = 2$, there is no constant slope K but, as $\lambda \to 0$ (or $k \to \infty$), the slope decreases to $K(q) = 0$. Also the slope empirically estimated for small k (large λ) is too low compared to its theoretical value.

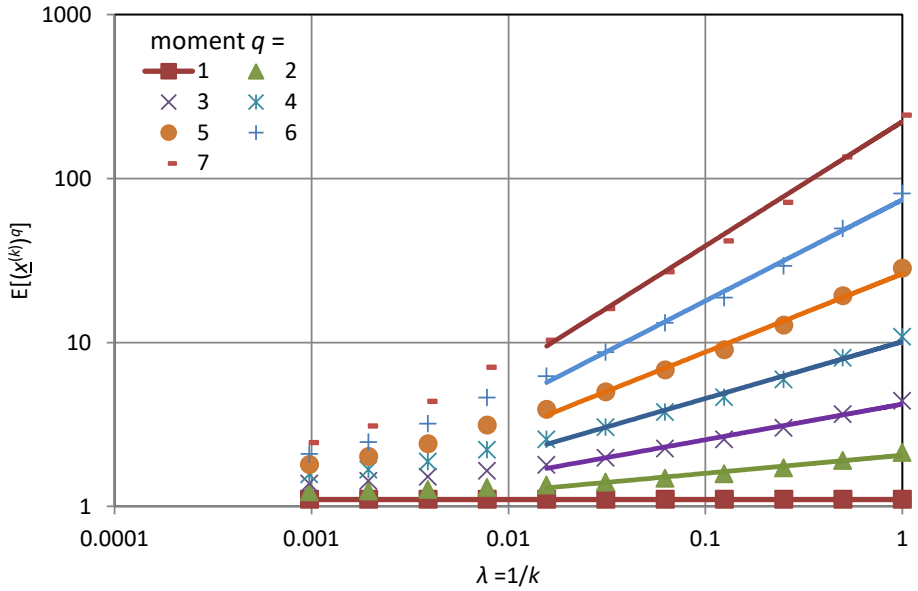

Figure 2 Illustration of spurious scaling laws between raw moments and inverse time scale.

Neglect of statistical bias and variation

The above example illustrates a symptom of a more general tendency in the fractal literature to treat observations (time series) deterministically, confusing random variables with their realizations and ignoring statistical bias and variation. In the example of Figure 2, high-order moments up to $q = 7$ have been used, as actually happens is several multifractal studies (this can be verified in studies that could be located with the Google search mentioned above).

However, high-order moments, which have been popular in multifractal studies, are well known in statistics to have minimal information content and therefore are avoided. This is further illustrated in Figure 3, constructed after Monte Carlo simulation of the fifth moment of a Pareto distribution with shape parameter 0.15 and for sample size $n = 100$ (Papalexiou et al. 2010; see also Lombardo et al., 2014).

Here the theory guarantees that there is no estimation bias, but the distribution function is enormously skewed. The mode is nearly two orders of magnitude less than the mean and the probability that a calculation, based on data, will reach the mean is two orders of magnitude lower than the probability of obtaining the mode. Therefore, there is no meaning in using such uncertain quantity, with so skewed distribution, in any type of inference.

Confusion between different scaling behaviours

Scaling relationships, expressed as power laws between involved quantities, have been central in fractal studies. Yet their meaning has been obscure, while quite different scaling laws with different meanings are confused and regarded to be of the same nature. This is like regarding the different physical laws that involve the product of two quantities (e.g. $F = m\,a$, $W = F\,s$, $m = \rho\,V$, where F, m, a, W, s, ρ and V denote force, mass, acceleration, work, displacement, density and volume) as a manifestation of the same magical law of multiplicative quantities.

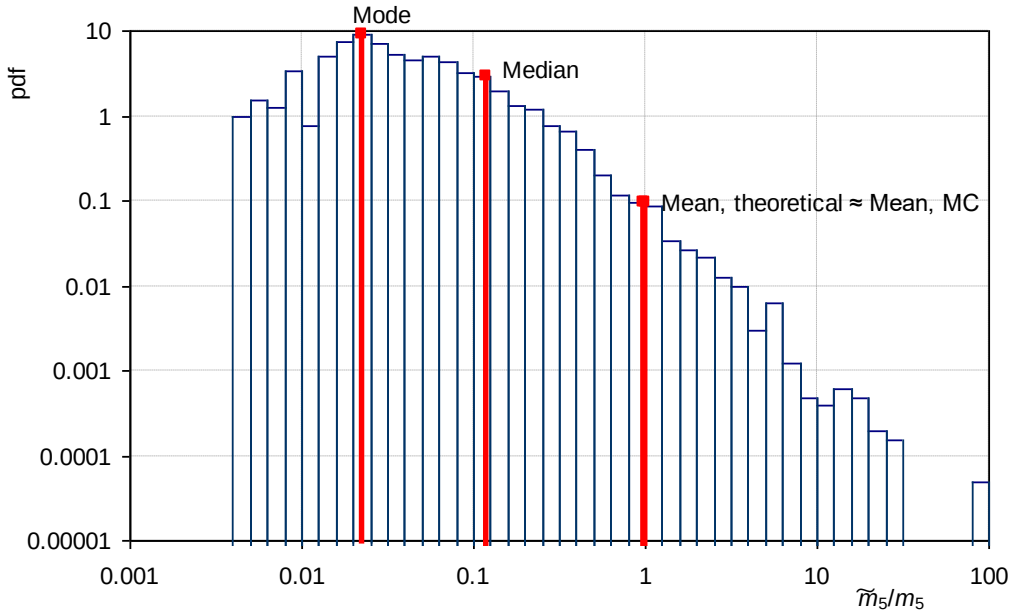

Figure 3 Illustration of the statistical distribution of the estimate $\widetilde{\boldsymbol{m}}_{\mathbf{5}}$ of the fifth moment m_5 of the Pareto distribution (pdf stands for probability density function).

It is thus important to differentiate the unlike types of scaling met in geophysical processes and clarify their meaning. We can distinguish the following types of scaling (where the formal definitions of the various terms are given in section 3):

- *Temporal scaling* indicates dependence in time and is expressed as a power law of some second order property (marginal or joint second central moment) of a process with respect to a quantity related to time. We can further subdivide temporal scaling into:
  - Hurst behaviour, which is expressed as a power function of autocorrelation vs. time lag or climacogram vs. time scale;
  - fractal (local) behaviour, which is expressed as a power function of structure function vs. time lag or climacogram-based structure function (see below) vs. time scale.
- *Spatial scaling* is similar to temporal scaling but indicating dependence in space.
- *State scaling* is totally irrelevant to temporal and spatial scaling; it is related to the marginal distribution of the process and indicates a heavy-tailed distribution (a power law of probability of exceedence vs. state).
- *Scaling of* (high-order) *moments* with time scale; while in theory this cannot be excluded, in most empirical studies it perhaps is an artefact related to other types of scaling and, as explained above, it is usually spurious because high-order moments are not reliably estimated from data.

As already mentioned, in real world systems scaling laws never extend to the entire range of scales. Usually they are asymptotic laws, with different exponents at each edge. Asymptotic scaling laws abound because, in our view, they are a mathematical necessity

(Koutsoyiannis, 2014). The asymptotic behaviour of stochastic properties of processes (such as survival function, autocovariance, structure function, climacogram, etc.) should necessarily tend to zero at one edge (e.g. at infinity) and the decay to zero can be exponential (fast decay) or of power-type (slow decay). In the latter case, the emergence of an asymptotic power law is obvious, whether it holds in the form of scaling in state (heavy-tailed distributions) or in time (long-term persistence). Both cases have been verified in geophysical time series (e.g. O'Connell *et al.*, 2016; Markonis and Koutsoyiannis, 2016, 2013; Dimitriadis and Koutsoyiannis 2017). According to this view, scaling behaviours are just manifestations of enhanced uncertainty and are consistent with the principle of maximum entropy (Koutsoyiannis, 2011; see also below). The connection of scaling with maximum entropy constitutes also a connection of stochastic representations of natural processes with statistical physics.

3. Fundamentals of stochastics for geophysics

In this section, we give a very brief presentation of the most fundamental concepts of stochastics. Later, in section 5 we will show that these concepts suffice to model complex phenomena without making any use of the fractal nomenclature, even though some of these phenomena are thought to belong to the preferential domain of the fractal literature (e.g. turbulence).

The meaning of randomness and stochastics

A deterministic world view is founded on a concept of sharp exactness. A deterministic mathematical description of a system uses regular variables (e.g. x) which are represented as numbers. The change of the system state is represented as a *trajectory* $x(t)$, which is the sequence of a system's states x as time t changes.

In an indeterministic world view there is uncertainty or randomness, where the latter term does not mean anything more than unpredictability or intrinsic uncertainty. A system's description is done in terms of random variables. A random variable $\underline{x}$ is an abstract mathematical entity whose realizations x belong to a set of possible numerical values. A random variable $\underline{x}$ is associated with a probability density (or mass) function $f(x)$. Notice the different notation of random variables (underlined, according to the Dutch notation; Hemelrijk, 1966) from regular ones. The evolution of a system over time is no longer sufficient to be represented as a trajectory but as a stochastic process $\underline{x}(t)$, which is a collection of (usually infinitely many) random variables $\underline{x}$ indexed by t (typically representing time). A realization (sample) $x(t)$ of $\underline{x}(t)$ is a trajectory; if it is known at certain points t_i, $i = 1, 2, \ldots$, it is a time series.

The mathematics of random variables and stochastic processes is termed stochastics, and is composed of probability theory, statistics and stochastic processes. Most natural processes evolve in continuous time but they are observed in discrete time, instantaneously or by averaging. Accordingly, the stochastic processes devised to represent the natural processes should evolve in continuous time and be converted into discrete time, as illustrated in Figure 4.

While a stochastic process denotes, by conception, change (process = change), there should be some properties that are unchanged in time. This implies the concept of stationarity (Koutsoyiannis and Montanari 2015), which is central in stochastics. For the remaining part of this article, the processes are assumed to be stationary, noting that nonstationary processes should be converted to stationary before their study (for example, the cumulative process $\underline{X}(t)$ in Figure 4 is nonstationary, but by differentiating it in time we obtain the stationary process $\underline{x}(t)$). The most customary properties of a stationary stochastic process are its second order properties:

- *Autocovariance function*, $c(h) := \mathrm{Cov}[\underline{x}(t), \underline{x}(t+h)]$.
- *Power spectrum* (also known as *spectral density*), $s(w)$; it is defined as the Fourier transform of the autocovariance function, i.e., by equation (2).
- *Structure function* (also known as *semivariogram* or *variogram*), $v(h) := (1/2)\mathrm{Var}[\underline{x}(t) - \underline{x}(t+h)]$.
- *Climacogram*, $\gamma(k) := \mathrm{Var}[\underline{x}_i^{(k)}]$, where $\underline{x}_i^{(k)}$ is the averaged process over time scale k (see Figure 4 and substitute a varying time scale k for the constant time interval D).

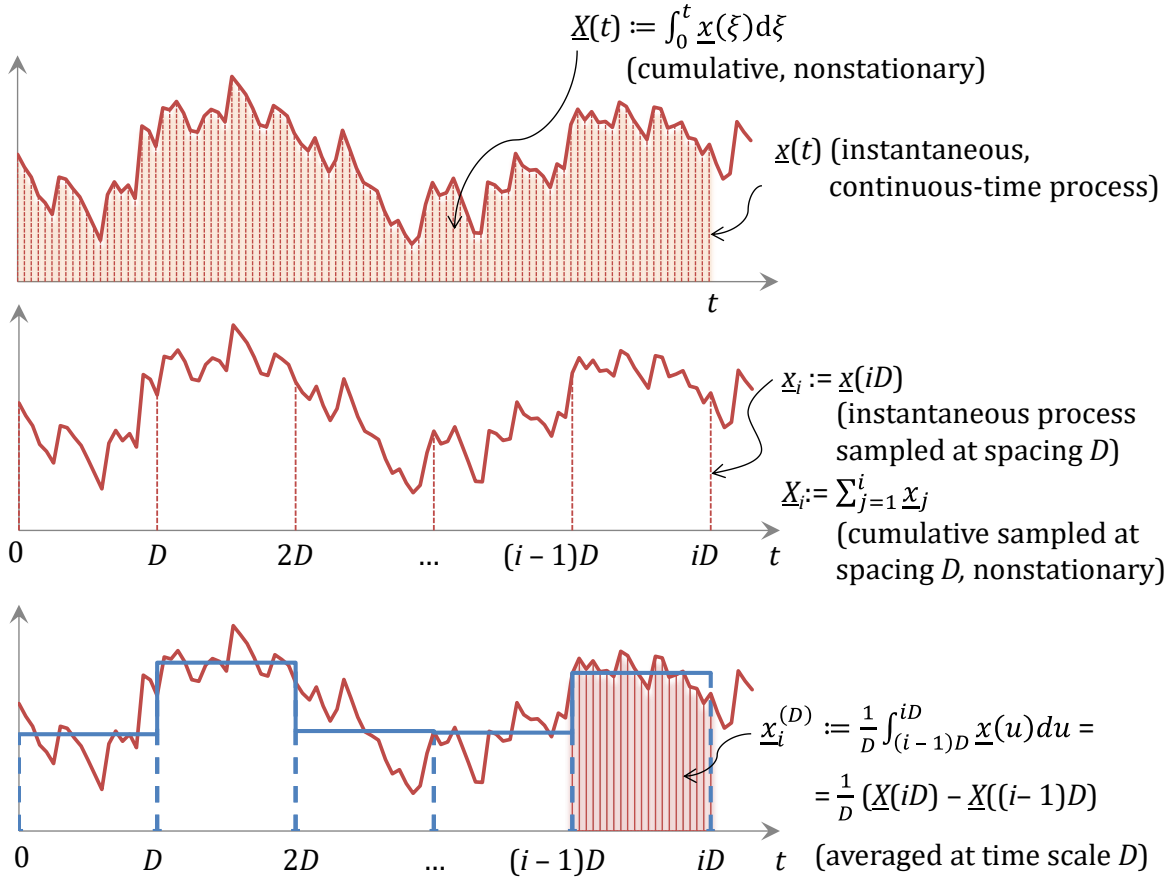

Figure 4 Explanatory sketch for a stochastic process in continuous time and two different representations in discrete time. Note that the graphs display a realization of the process (it is impossible to display the process as such) while the notation is for the process per se.

For time-related quantities, in the above notation and in the next part of this article, we use the following symbols, where Latin letters denote dimensional quantities and Greek letters dimensionless ones, where the latter are convenient when using the discrete-time variants of a process:

- *Time unit* (time step in case of sampling or time scale in case of aggregating or averaging)*, D.*
- *Time*, $t = \tau\ D$ (alternatively for strictly integer i = 1, 2, …, $t = i\ D$ where t is continuous time and i discrete time).
- *Time lag*, $h = \eta\ D$.
- *Time scale* $k = \kappa\ D$.
- *Frequency,* $w = \omega/\ D$, related to time scale by $w = 1/k$, $\omega = 1/\kappa$.

All these properties are transformations of one another, i.e.:

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

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

$$\gamma(k) = 2\int_0^1 (1-v)c(vk)\mathrm{d}v, \qquad c(h) = \frac{1}{2}\frac{\mathrm{d}^2\left(h^2\gamma(h)\right)}{\mathrm{d}h^2} \tag{4}$$

where equation (3) is valid when the variance of the instantaneous process is finite ($\gamma_0 := \gamma(0) \equiv c(0) \neq \infty$).

The climacogram is not as popular as the other tools but it has several good properties due to its simplicity, close relationship to entropy (see below), and more stable behaviour, which is an advantage in model identification and fitting from data. In particular, when estimated from data, the climacogram behaves better than all other tools, which involve high bias and statistical variation (Dimitriadis and Koutsoyiannis, 2015; Koutsoyiannis, 2016). The climacogram involves bias too, but this can be determined analytically and included in the estimation. Furthermore, it enables the definition of additional useful tools as shown in Table 1.

Table 1 Climacogram based metrics of stochastic processes.

Metric / Usefulness	Definition	Comments
Climacogram Useful for the global asymptotic behaviour ($k\rightarrow\infty$)	$\gamma(k) \coloneqq \mathrm{Var}[\underline{x}_i^{(k)}]$	For an ergodic process for $k\rightarrow\infty$, $\gamma(k)\rightarrow 0$ necessarily
Climacogram-based structure function (CSF) Useful for the local asymptotic behaviour ($k\rightarrow 0$)	$\xi(k) \coloneqq \gamma_0 - \gamma(k)$	The definition presupposes that the variance γ_0 is finite
Climacogram-based spectrum (CS) Useful for both the global and local asymptotic behaviour	$\psi(w) \coloneqq \frac{2}{w\gamma_0}\gamma(1/w)\xi(1/w)$ $= \frac{2\,\gamma(1/w)}{w}\left(1-\frac{\gamma(1/w)}{\gamma_0}\right)$	It combines the climacogram and the CSF; valid even for infinite variance

The CSF, $\xi(k)$, behaves like the structure function $v(h)$ and is related to the latter by the same way as the climacogram $\gamma(k)$ is related to the autocovariance function $c(h)$:

$$c(h) = \frac{1}{2}\frac{\mathrm{d}^2\left(h^2\gamma(h)\right)}{\mathrm{d}h^2}, \quad v(h) = \frac{1}{2}\frac{\mathrm{d}^2\left(h^2\xi(h)\right)}{\mathrm{d}h^2} \tag{5}$$

The CS, $\psi(w)$, behaves like the power spectrum; it has same dimensions, and in most cases has precisely the same asymptotic behaviour as the power spectrum, but it is smoother and more convenient in model identification and fitting (see section 5).

Second order properties at discrete time

Once the continuous-time properties are known, the discrete-time ones can be readily calculated. For example, and assuming a time interval D for discretization, as in Figure 4, the autocovariance of the averaged process is:

$$c_\eta^{(D)} = \mathrm{Cov}\left[\underline{x}_\tau^{(D)}, \underline{x}_{\tau+\eta}^{(D)}\right] = \frac{1}{D^2}\left(\frac{\Gamma(|\eta+1|D) + \Gamma\left((|\eta-1|D\right)}{2} - \Gamma(|\eta|D)\right) \tag{6}$$

where $\Gamma(D) \coloneqq \mathrm{Var}[\underline{X}(D)] = D^2\gamma(D)$. Also, the power spectrum of the averaged process can be calculated from:

$$s_\mathrm{d}^{(D)}(\omega) = 2c_0^{(D)} + 4\sum_{\eta=1}^{\infty} c_\eta^{(D)}\cos(2\pi\eta\omega) \tag{7}$$

where $s_\mathrm{d}^{(D)}(\omega) \coloneqq s^{(D)}(w)/D$ (nondimensionalized spectral density), whereas the discrete-time power spectrum $s^{(D)}(w)$ is related to the continuous-time one by (Koutsoyiannis, 2016)

$$s^{(D)}(\omega) = \sum_{j=-\infty}^{\infty} s\left(w + \frac{j}{D}\right)\mathrm{sinc}^2(\pi(wD+j)) \tag{8}$$

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

Cautionary notes for model fitting

Model identification and fitting is much more important than commonly thought. Even the statistical literature has paid little attention to the fact that 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 to estimate statistics.

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.

It is commonly thought that the standard estimator of the variance from a sample of size n is unbiased if we divide the sum of squared deviations from mean by $n - 1$ instead of n (equation (9)). This is correct only if the assumed model is the white noise. Otherwise, the estimation is biased and, if the process has long-range dependence, the bias can be substantial. The climacogram, which is none other than the variance, needs

to consider this bias. Actually, it is easy to analytically estimate the bias and the effect of discretization, once a model has been assumed in continuous time.

Let us consider a process with climacogram $\gamma(k)$, from which we have a time series for an observation period T (multiple of the time step D), each one giving the averaged process $\underline{x}_i^{(D)}$ at a time step D. We form time series for scales that are multiples of D, i.e., $k = \kappa D$, $\kappa = 1, 2, \ldots,$ and we wish to estimate the variance at any such scale (including that at scale D, for $\kappa = 1$). The standard estimator $\hat{\underline{\gamma}}(k)$ of the variance $\gamma(k)$ is

$$\hat{\underline{\gamma}}(k) \coloneqq \frac{1}{n-1}\sum_{i=1}^{n}\left(\underline{x}_i^{(k)} - \underline{x}_1^{(T)}\right)^2 = \frac{1}{T/k-1}\sum_{i=1}^{T/k}\left(\underline{x}_i^{(k)} - \underline{x}_1^{(T)}\right)^2 \tag{9}$$

where by inspection it is seen that $\underline{x}_1^{(T)}$ is the sample mean, while it was assumed that T is a multiple of k so that the sample size is $n = T/D$ (if not, we should replace T with $\lfloor T/k \rfloor k$, where $\lfloor . \rfloor$ denotes the floor of a real number). It can be then shown (Koutsoyiannis, 2011, 2016) that the bias can be calculated from

$$\mathrm{E}\left[\hat{\underline{\gamma}}(k)\right] = \chi(k,T)\gamma(k), \quad \chi(k,T) = \frac{1-\gamma(T)/\gamma(k)}{1-k/T} = \frac{1-(k/T)^2\Gamma(T)/\Gamma(k)}{1-k/T} \tag{10}$$

Entropy and entropy production

As already mentioned, the emergence of scaling from maximum entropy considerations may provide the theoretical background in modelling complex natural processes by scaling laws.

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:

$$\Phi\left[\underline{X}(t)\right] \coloneqq \mathrm{E}\left[-\ln\frac{f(\underline{X};t)}{m(\underline{X})}\right] = -\int_{-\infty}^{\infty}\ln\frac{f(X;\,t)}{m(X)}f(X;\,t)dX \tag{11}$$

where $m(X)$ 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):

$$\varphi(t) \equiv \varphi[\underline{X}(t)] \coloneqq \Phi'[\underline{X}(t)]\, t \equiv \mathrm{d}\Phi[\underline{X}(t)] \,/\, \mathrm{d}(\ln t) \tag{12}$$

For a Gaussian process with constant density of background measure, $m(X) \equiv m$, the entropy depends on its variance $\Gamma(t)$ only and is:

$$\Phi[\underline{X}(t)] = (1/2)\ln(2\pi \mathrm{e}\,\Gamma(t)/m^2),\ \varphi(t) = \Gamma'(t)\,t \,/\, 2\Gamma(t) \tag{13}$$

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:

$$\Gamma_\mathrm{C}(t) \approx 2\Gamma(t) - \frac{\Gamma(2t)}{2}, \quad \varphi_\mathrm{C}(t) = \frac{\Gamma_\mathrm{C}'(t)t}{2\Gamma_\mathrm{C}(t)} \approx \frac{\left(2\Gamma'^{(t)} - \Gamma'^{(2t)}\right)t}{4\Gamma(t) - \Gamma(2t)} \tag{14}$$

Resulting processes from maximizing entropy production

Koutsoyiannis (2011) assumed that the behaviour seen in natural processes is consistent with extremization of entropy production and provided a framework to derive processes maximizing entropy production. Using simple constraints in maximization, such as known variance at the scale $k = D = 1$, and lag one autocovariance for the same time scale, the following processes extremizing the EPLT $\varphi(t)$ and $\varphi_C(t)$ can be derived, which are also depicted in Figure 5 in terms of their EPLT and climacograms.

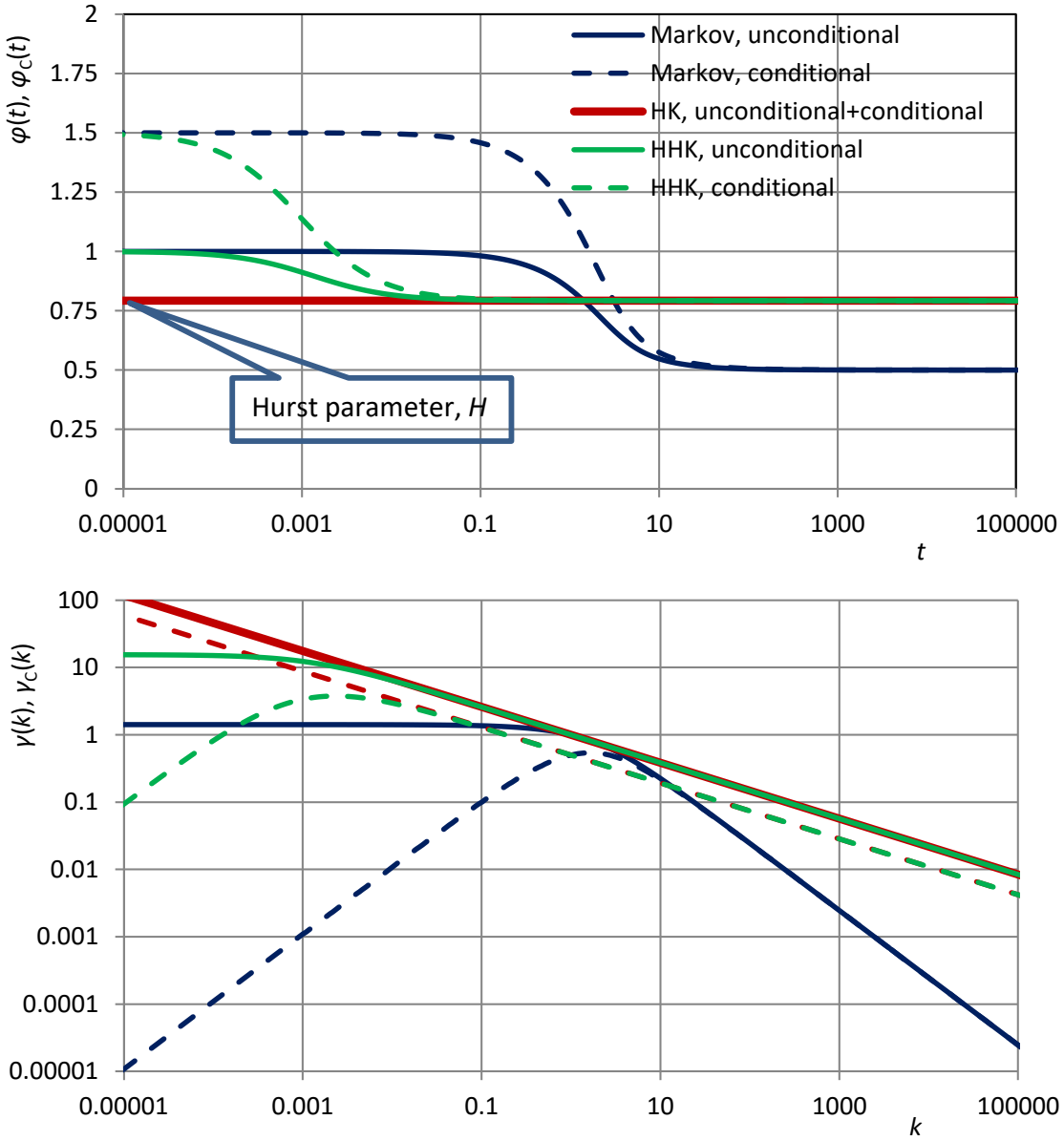

Figure 5 EPLTs (upper) and climacograms (lower) of the three processes extremizing entropy production. At time scale $k = 1$ all three processes have the same variance, $\gamma(1) = 1$, and the same autocovariance for lag 1, $\boldsymbol{c}_1^{(1)} = 0.5$. Their parameters are (see text for their definitions): for the Markov process $\alpha = 0.8686$, $\lambda = 1.4176$; for the HK process $a = 0.0013539$, $\lambda = 15.5032$, $H = 0.7925$; for the HHK process $a = 0.0013539$, $\lambda = 15.5093$, $M = 0.5$, $H = 0.7925$ (adapted from Koutsoyiannis 2016).

- A *Markov process:*

$$c(h) = \lambda \mathrm{e}^{-h/\alpha}, \gamma(k) = \frac{2\lambda}{k/\alpha}\left(1 - \frac{1-\mathrm{e}^{-k/\alpha}}{k/\alpha}\right) \quad (15)$$

  maximizes entropy production for small times but minimizes it for large times.
- A *Hurst-Kolmogorov* (HK) process:

$$\gamma(k) = \lambda(\alpha/k)^{2-2H} \quad (16)$$

  maximizes entropy production for large times but minimizes it for small times.
- A *Hybrid Hurst-Kolmogorov* (HHK) process

$$\gamma(k) = \lambda(1 + (k/\alpha)^{2M})^{\frac{H-1}{M}} \quad (17)$$

  maximizes entropy production both at small and large time scales.

In these definitions α and λ are scale parameters with dimensions of $[t]$ and $[x^2]$, respectively. The parameter H (in honour of Hurst) is the Hurst parameter which determines the global properties of the process (as $k \to \infty$). The parameter M (in honour of Mandelbrot) is the fractal parameter which determines the local properties (as $k \to 0$). Both H and M are dimensionless numbers in the interval $(0, 1)$. In the HHK process, locality and globality are clearly independent of each other, each one characterized by an asymptotic power law. Hence, it allows explicit control of both asymptotic logarithmic slopes of the CS $\psi^{\#}(k)$ and the power spectrum $s^{\#}(w)$. In the special case where $H = M = 0.5$, HHK is practically indistinguishable from a Markov process, even though not precisely identical. Furthermore, as $\alpha \to 0$, the process tends to a pure HK process with the same Hurst parameter H. Also, for any specific parameter set, HHK exhibits Markov behaviour for small time scales (if $M = 0.5$, or similar to Markov if $M \neq 0.5$) and Hurst behaviour for large time scales, as seen in Figure 5.

The HHK process is consistent with natural behaviours and remedies known inconsistencies of the HK process (discussed in subsection "*Use of the abstract mathematical objects as if they are natural objects*"), while retaining the persistence or antipersistence properties. Specifically, the variance of the instantaneous process is always finite ($\gamma_0 = \gamma(0) = \lambda$), while even for $0 < H < 0.5$ the initial part of the autocovariance function for small lags is positive for all variants of the process (continuous time, discrete time, either sampled or averaged, for a small time interval D).

4. Simulation of stochastic processes respecting their fractal properties

Monte Carlo (stochastic) simulation is an important numerical method for resolving problems that have no analytical solution. Obviously, simulation is performed in discrete time, at a convenient discretization step. The following method based on the so-called symmetric moving average (SMA) scheme (Koutsoyiannis, 2000, 2016) can be used to exactly simulate any Gaussian process, with any arbitrary autocovariance function (provided that it is mathematically feasible). It can also approximate, with controlled accuracy, any non-Gaussian process with any arbitrary autocovariance function and any marginal distribution function.

The symmetric moving average scheme

The SMA scheme can directly generate time series x_i (where for simplicity we have omitted the time interval D in the notation) from any process $\underline{x}_i$ with any type of dependence by:

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

where a_l are coefficients calculated from the autocovariance function and $\underline{v}_i$ is white noise averaged in discrete-time. Assuming that the power spectrum $s_{\mathrm{d}}^{(D)}(\omega)$ of the averaged discrete-time process is known (from the equations listed above), 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

$$s_{\mathrm{d}}^{a}(\omega) = \sqrt{2s_{\mathrm{d}}^{(D)}(\omega)} \tag{19}$$

Thus, to calculate a_l we first determine $s_{\mathrm{d}}^{a}(\omega)$ from the power spectrum of the process and then we invert 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 (18) to

$$\underline{x}_i = \sum_{l=-q}^{q} a_{|l|}\underline{v}_{i+l} \tag{20}$$

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

$$a_l = a'_l + a'' \tag{21}$$

where the coefficients a'_l are calculated from inverting 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:

$$\gamma(D) = \sum_{l=-q}^{q} a_{|l|}^2 = \sum_{l=-q}^{q} \left(a'_{|l|} + a''\right)^2 \tag{22}$$

Solving for a'', this yields:

$$a'' = \sqrt{\frac{\gamma(D) - \Sigma a'^2}{2q+1} + \left(\frac{\Sigma a'}{2q+1}\right)^2} - \frac{\Sigma a'}{2q+1} \tag{23}$$

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

Handling of moments higher than second order

In addition to being general for any second order properties (autocovariance function), the SMA method can explicitly preserve higher order marginal moments. Here is should be made clear that, while, as already mentioned, high-order moments cannot be estimated reliably from data, non-Gaussianity is very commonly verified empirically and also derived by theoretical reasoning (Koutsoyiannis 2005, 2014). An easy manner to simulate non-Gaussian (e.g., skewed) distributions is to calculate theoretically (not from the data) their moments and then explicitly preserve these moments in simulation. Preservation of three or four central moments usually provides good approximations to the theoretical distributions. Apparently, by preserving four moments, a non-Gaussian distribution is not precisely preserved. What can be assumed to be preserved is a Maximum Entropy (ME) approximation of the distribution constrained by the known moments. For four known moments of the variable $\underline{x}$ this approximation should be an exponentiated fourth-order polynomial of x (Jaynes, 1957; Papoulis, 1991), which can be written as

$$f(x) \coloneqq \frac{1}{\lambda_0} \mathrm{e}^{-\left(\frac{x}{\lambda_1}+\mathrm{sign}(\lambda_2)\left(\frac{x}{\lambda_2}\right)^2+\left(\frac{x}{\lambda_3}\right)^3+\left(\frac{x}{\lambda_4}\right)^4\right)} \tag{24}$$

where λ_i are parameters, all with dimensions $[x]$ (with $\lambda_4 \geq 0$).

The third and fourth moments are more conveniently expressed in terms of the coefficients of skewness and kurtosis, respectively. 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 (Koutsoyiannis, 2000):

$$C_{\mathrm{s},v} = C_{\mathrm{s},x} \frac{\left(\sum_{l=-q}^{q} a_{|l|}^2\right)^{3/2}}{\sum_{l=-q}^{q} a_{|l|}^3} \tag{25}$$

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 (Dimitriadis and Koutsoyiannis, 2017):

$$C_{\mathrm{k},v} = \frac{C_{\mathrm{k},x}\left(\sum_{l=-q}^{q} a_{|l|}^2\right)^2 - 6\sum_{l=-q}^{q-1}\sum_{k=l+1}^{q} a_{|l|}^2\, a_{|k|}^2}{\sum_{l=-q}^{q} a_{|l|}^4} \tag{26}$$

Four-parameter distributions are needed to preserve skewness and kurtosis; details are provided by Dimitriadis and Koutsoyiannis (2017). Illustration of the very good performance of the method in the generation of non-Gaussian white noise is provided in Figure 6 for popular distribution functions such as Weibull, gamma, lognormal and Pareto.

It is finally noted that the method can also be used in multivariate processes, represented by vectors of random variables (Koutsoyiannis, 2000).

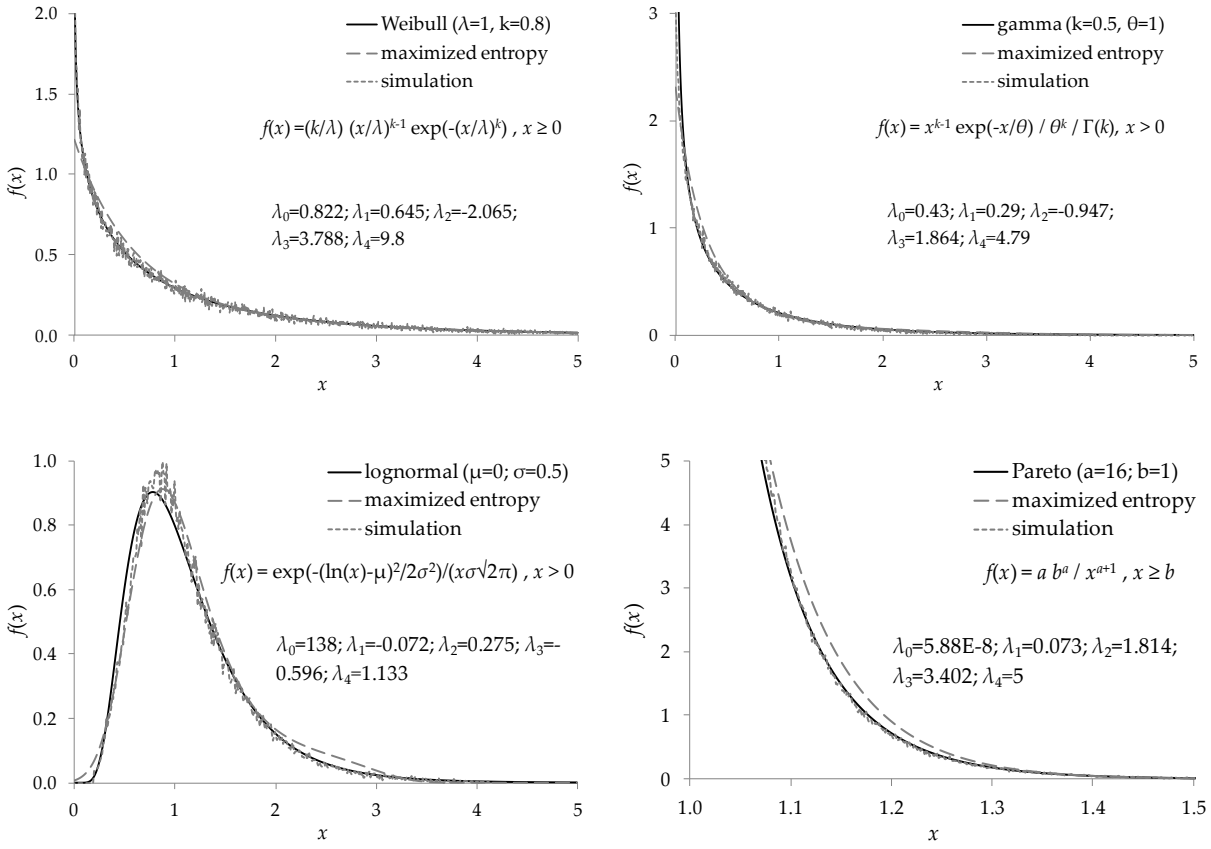

Figure 6 Various two-parameter probability density functions along with their fitted ME approximations and the empirical probability density from a single synthetic time series with $n = 10^5$ (from Dimitriadis and Koutsoyiannis, 2017).

5. Applications

Application 1: Turbulence

Estimation of high-order moments involves large uncertainty and cannot be reliable in the typically short time series of geophysical processes. However, in laboratory experiments at sampling intervals of μs, very large samples can be formed which can support the reliable estimation of high-order moments. Here we use grid-turbulence data made available on the Internet by the Johns Hopkins University (http://www.me.jhu.edu/meneveau/datasets/datamap.html). This dataset consists of 40 time series with $n = 36\times10^6$ data points of longitudinal wind velocity along the flow direction, all measured at a sampling time interval of 25 μs by X-wire probes placed downstream of the grid (Kang et al., 2003).

By standardizing all series (see Dimitriadis et al., 2016; Dimitriadis and Koutsoyiannis, 2017) we formed a sample of $40 \times 36 \times10^6 = 1.44 \times10^9$ values to estimate the marginal distribution, and an ensemble of 40 series, each with 36×10^6 values to estimate the dependence structure through the climacogram. Based on this dataset we built a stochastic model of turbulence, which to verify we performed stochastic

simulation using the SMA framework with n =10^6 values and compared the synthetic data with the measurements using several tools.

In terms of the marginal distribution, the time series are nearly-Gaussian but not exactly Gaussian. There are slight deviations from normality toward positive skewness, as indicated by the coefficient of skewness, which is 0.2 instead of 0, and that of kurtosis, which is 3.1 instead of 3, as well as from the plot of the probability function shown in Figure 7. This divergence of fully developed turbulent processes from normality has been also justified theoretically (Wilczek et al., 2011). Interestingly, these slight differences from normality result in highly non-normal distribution of the white noise $\underline{v}_i$ of the SMA model (skewness $C_{s,v}$ = 3.26; kurtosis $C_{k,v}$ = 12.30!); this should have substantial effects in some aspects of turbulence.

For the stochastic dependence of the turbulent velocity process, after some exploratory analysis, we assumed a model consisting of the sum of two equally weighted processes, an HHK and a Markovian:

$$\gamma(k) = \frac{\lambda}{2}(1 + (k/\alpha)^{2M})^{\frac{H-1}{M}} + \frac{\lambda}{k/\alpha}\left(1 - \frac{1 - e^{-k/\alpha}}{k/\alpha}\right) \tag{27}$$

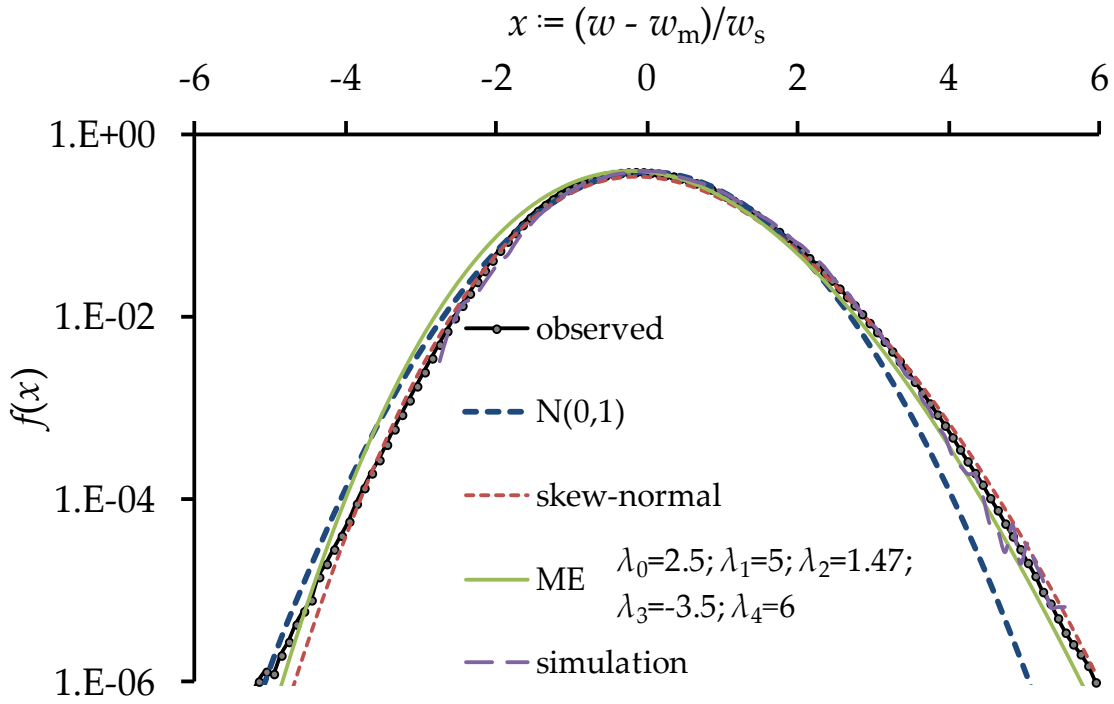

Figure 7 Probability density function of the measured turbulent velocity w standardized, in each time series, by the mean w_m and standard deviation w_s, compared to that of a single simulation using the SMA scheme preserving the first four moments; the standard normal distribution $N(0,1)$ and the skew normal (both not used in simulation) are also shown. The ME approximation, also shown in the figure, is the one used in simulations.

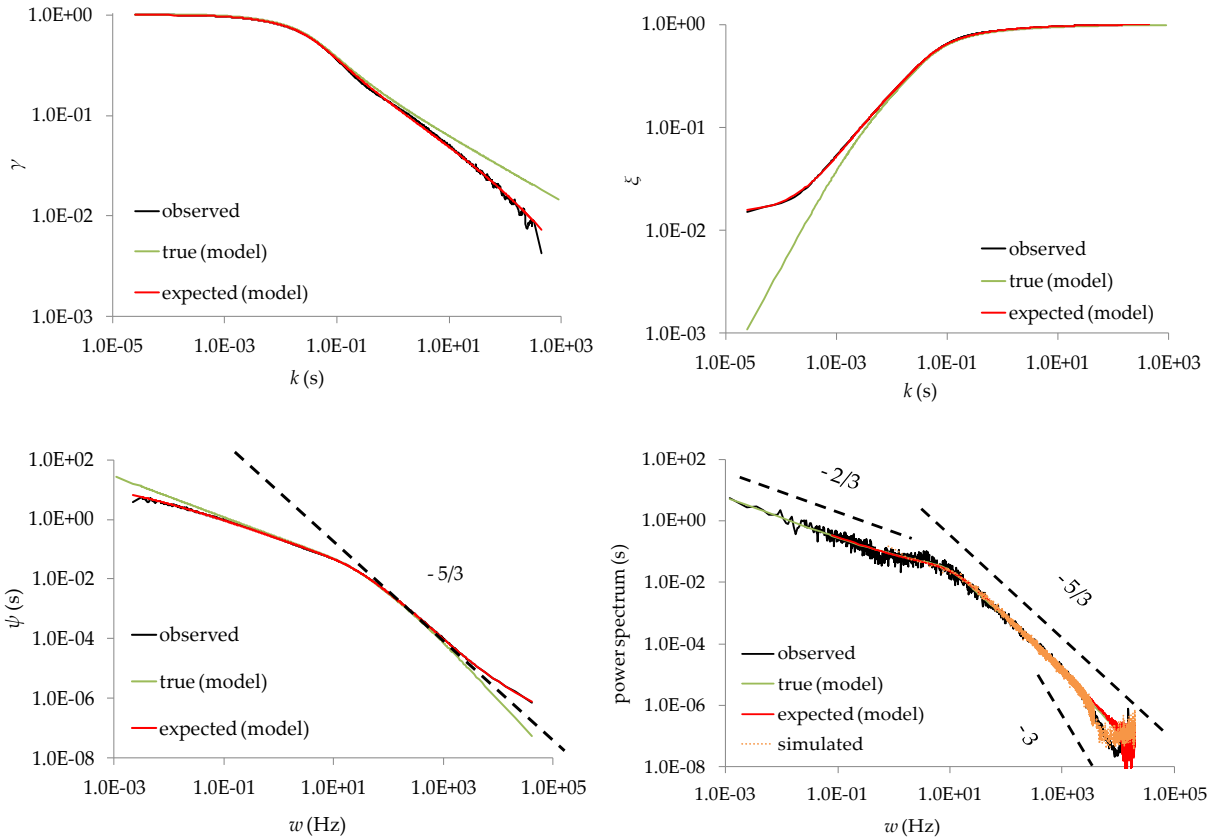

Figure 8 Empirical, true and expected values of the climacogram (upper left), CSF (upper right), CS (lower left) and power spectrum (lower right). The "observed" is the average from the 40 time series.

We fitted the model to the climacogram, the structure function, the CS and the power spectrum, calculated as the average of the 40 series. The fitting is shown in Figure 8; the four parameters of the model are estimated as: $\lambda = 1$, $\alpha = 14$ ms, $M = 1/3$, $H = 5/6$. As seen in Figure 8, the model is indistinguishable from the data, measured or synthesized, when the climacogram or its derivatives CSF and CS are used. Note that the comparison of the empirical quantities is not made with the true ones but with the expected, in order to take account of the bias.

The power spectrum is much rougher than the other tools, yet a good model fit can be clearly seen. Kolmogorov's "5/3" law of turbulence (K41 self-similar model; Kolmogorov, 1941) is also evident in the power spectrum for $w > 10$ Hz. Steepening of the power spectrum slope for even larger frequencies ($w > 1000$ Hz), which has also reported in several studies, is also apparent in Figure 8. This, however seems to be a numerical effect (resulting from discretization and bias), as the same behaviour appears also in the simulated data from a model whose structure (equation (27)) does not include anything that would justify steepening of the slope.

It is extremely insightful to investigate the high-order properties of the velocity increments, i.e., differences of velocities at adjacent times with a certain time distance (lag) h. In particular, the variation of high-order moments of the velocity increments with increasing h (i.e., the moments $v_p := \mathrm{E}[|\underline{x}(t) - \underline{x}(t+h)|^p]$ for $p > 2$) has been

associated with the intermittent behaviour of turbulence and has been mentioned as the intermittent effect (Frisch, 2006, sect. 8.3), first discovered in turbulence by Batchelor and Townsend (1949). Therefore it is important to preserve this variation. The model in equation (27) does not make any effort for such preservation. However, as seen in Figure 9, these are preserved well and effortless. Therefore, it is no longer puzzling to have large kurtosis (even > 5) in velocity increments, even though the velocity is almost normal. No additional assumption, model component, or even model parameter is necessary. Similar good preservation appears also for the skewness of velocity increments (Figure 9).

The huge data size in this application allows evaluation of even higher moments and construction of a plot (Figure 10) of the exponent ζ_p vs. moment order p of an assumed scaling relationship

$$v_p \coloneqq \mathrm{E}\left[\left|\underline{x}(t) - \underline{x}(t+h)\right|^p\right] \approx h^{\zeta_p} \tag{28}$$

which has been very common in the literature. Again the agreement between the simulated and measured data is impressive, particularly if we bear in mind the fact that no provision has been made to this aim. Some more simulations have been used to investigate this further and a number of additional curves have been plotted in Figure 10. It is thus seen that the HHK model alone fails to preserve this actual behaviour if a Gaussian distribution is assumed; it rather approached the K41 self-similar model (Kolmogorov, 1941) as reproduced by Frisch (2006, Fig. 8.8). Similar results are obtained if a Markov dependence structure is assumed along with the modelled marginal distribution based on the empirical moments (Figure 7). Interestingly, if we combine the modelled distribution (Figure 7) and the modelled climacogram (equation (27)), then we adequately preserve the intermittent effect without the need for any other mono-fractal (such as the β-model) nor multi-fractal models (cf., Frisch, 2006, sect. 8.5) and not even the She-Leveque model (1994), which is also plotted in Figure 10 (Frisch, 2006, sect. 8.6.4, 8.6.5) and behaves also well against the empirical data.

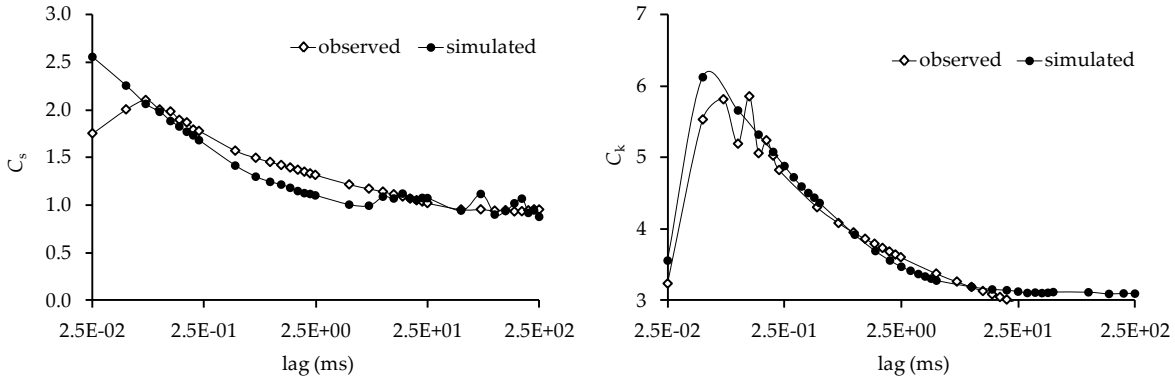

Figure 9 Empirical and simulated coefficients of skewness (left) and kurtosis (right) of the velocity increments vs. lag.

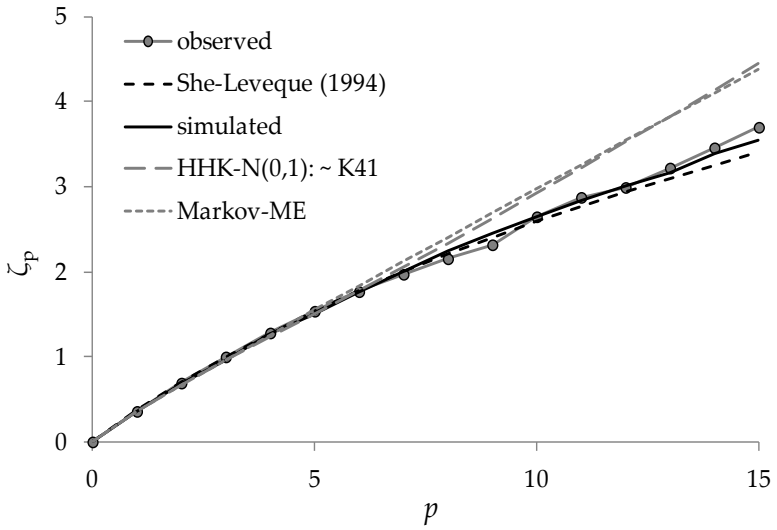

Figure 10 Empirical values of the scaling exponent ζ_p vs. moment order p of the scaling relationship (28).

In conclusion, this application shows that all important properties of turbulence, including its short- and long-term characteristics, as well as intermittency, can be very well modelled without any mystery but using a parsimonious stochastic model, theoretically justified on the basis of the maximization of entropy production (Koutsoyiannis, 2011), with both Hurst and fractal behaviours and slightly non-Gaussian distribution (with skewness of 0.2 and kurtosis of just 3.1).

Application 2: Wind

Understanding atmospheric motion in the form of wind is essential to many fields in geophysics. Wind is considered one of the most important processes in hydrometeorology since, along with temperature, it drives climate dynamics. Currently, the interest for modelling and forecasting of wind has increased due to the importance of wind power production in the frame of renewable energy resources development.
For the investigation of the large scale of atmospheric wind speed, we use over 15000 meteorological stations around the globe (Figure 11, upper) recorded mostly by anemometers and with hourly resolution (Integrated Surface Database—ISD; http://www.ncdc.noaa.gov/isd). In total, we analyse almost 4000 stations from different sites and climatic regimes by selecting time series that are still operational, with at least one year length of data, at least one non-zero measurement per three hours on average and at least 80% of non-zero values for the whole time series (Figure 11, middle). This data set is referred to below as "global".

By standardizing all series we formed a sample of $\sim 0.5\times 10^9$ values to estimate the marginal distribution, and an ensemble of 3886 series, each with $\sim 10^5$ values on average, to estimate the dependence structure through the climacogram. A known problem of field measurements of wind (particularly those originating from over 70 years ago), is that the technology of measuring devices has been rapidly changed (Manwell et al., 2010, sect. 2.8.3).

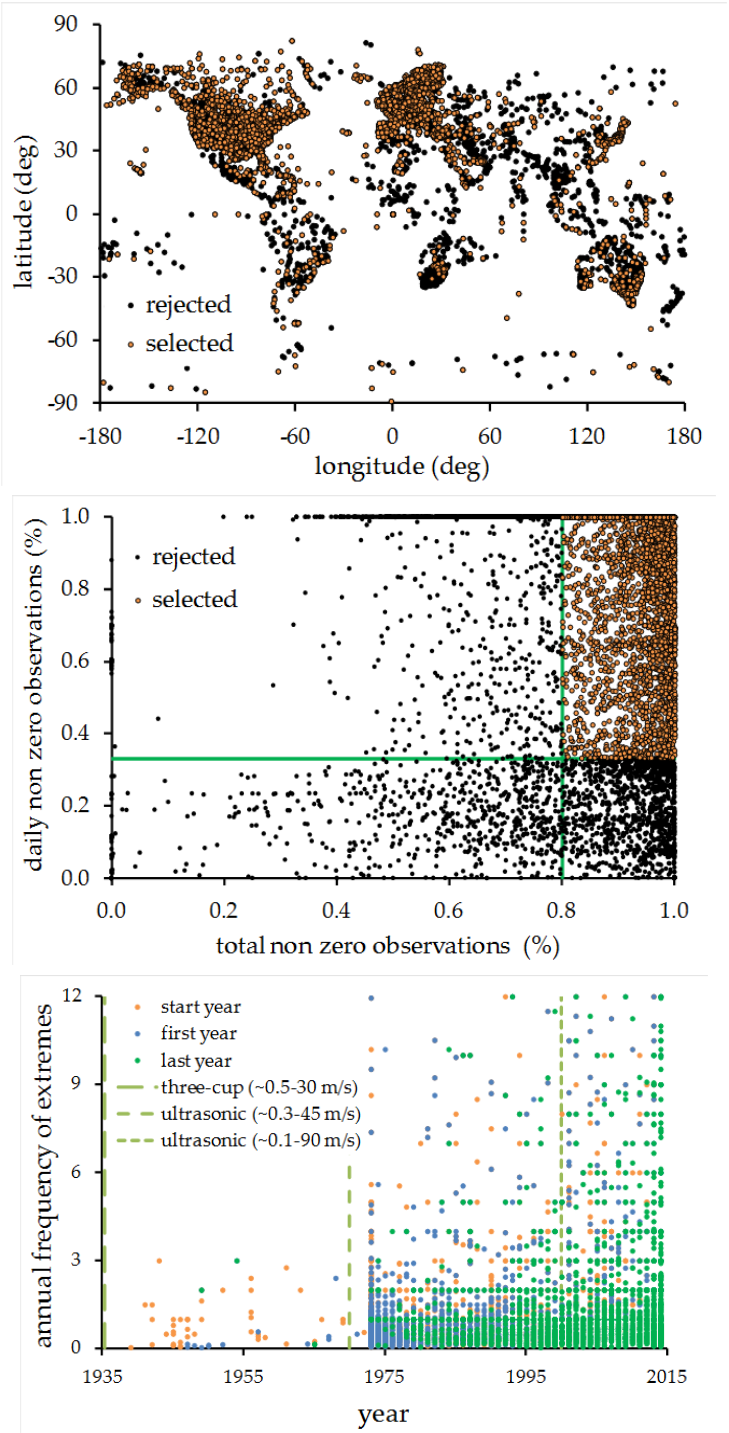

Figure 11 (upper) Distribution of the wind speed stations over the globe; (middle) sketch about the selection of the stations in the analysis; (lower) evolution of the frequency of measured extremes in the stations (where the 'start' year denotes the first operational year of the station and the 'first' and 'last' year denote the first and last year that an extreme value was recorded, respectively).

For example, in Figure 11 (lower) we illustrate a rather virtual increase of extreme wind events after the 1970s which is mainly due to the inability of older devices to properly measure wind speeds over 30 m/s (i.e., category I of Saffir–Simpson hurricane wind scale). Furthermore, in common anemometer instrumentation there is a lower

threshold of speed that could be measured, usually within the range 0.1 – 0.5 m/s (e.g., www.pce-instruments.com). It should be noted that, as the recorded wind speed decreases, so does the instrumental accuracy and it may be a good practice to always set the minimum threshold to 0.5 m/s to avoid measuring the errors of the instrument (e.g., zero or extremely low values) in place of the actual wind speed that can never reach an exact zero value.

In an attempt to incorporate smaller scales, starting from the microscale of turbulence, we include again the dataset of the previous application of turbulence, using it as an indicator of the similar statistical properties of small scale wind (Castaing et al., 1990). In addition to the 40 time series of the longitudinal turbulent velocity, here we also use another 40 time series of transverse velocity, measured at the same points with the longitudinal one; again each time series has $n = 36\times10^6$ data points with a sampling interval of 25 μs. The coefficients of skewness and kurtosis are estimated as 0.1 and 3.1 for the transverse velocity, respectively. Stochastic similarities between small scale atmospheric wind and turbulent processes abound in the literature as for example in terms of the marginal distribution (Monahan, 2013 and references therein), of the distribution of fluctuations (Böttcher et al., 2007 and references therein), of the second-order dependence structure (Dimitriadis et al., 2016 and references therein) and of higher-order behaviour such as intermittency (e.g., Mahrt, 1989).

Finally, to link the large and small scale of atmospheric wind we analyse an additional time series, referred to as “medium”, provided by NCAR/EOL of one-month length and with a 10 Hz resolution. This time series has been recorded by a sonic anemometer on a meteorological tower located at Beaumont KS and it includes over 25×10^6 longitudinal and transverse wind speed measurements (http://data.eol.ucar.edu/; Doran, 2011).

The statistical characteristics based on moments up to fourth order are shown in Figure 12; interestingly, there appears to be a rather well defined relationship between mean and standard deviation. The plot of coefficient of kurtosis vs. coefficient of skewness indicates that Weibull distribution falls close to the lower bound of the scatter of empirical points.

Numerous works have been conducted for the distribution of the surface wind speed (see Appendix B for a sample of recent studies). The Weibull distribution has proven very useful in describing the wind magnitude distribution for over three decades (Monahan, 2006 and references therein). However, various studies illustrate empirical as well as physically-based deviations from the Weibull distribution (Drobinski and Coulais, 2012 and references therein). Due to the discussed limitations of properly measuring wind speed most studies have focused on a local or small scale. In such cases where there is limited empirical evidence, we could search for a physical justification for the left and right tail of the probability function.

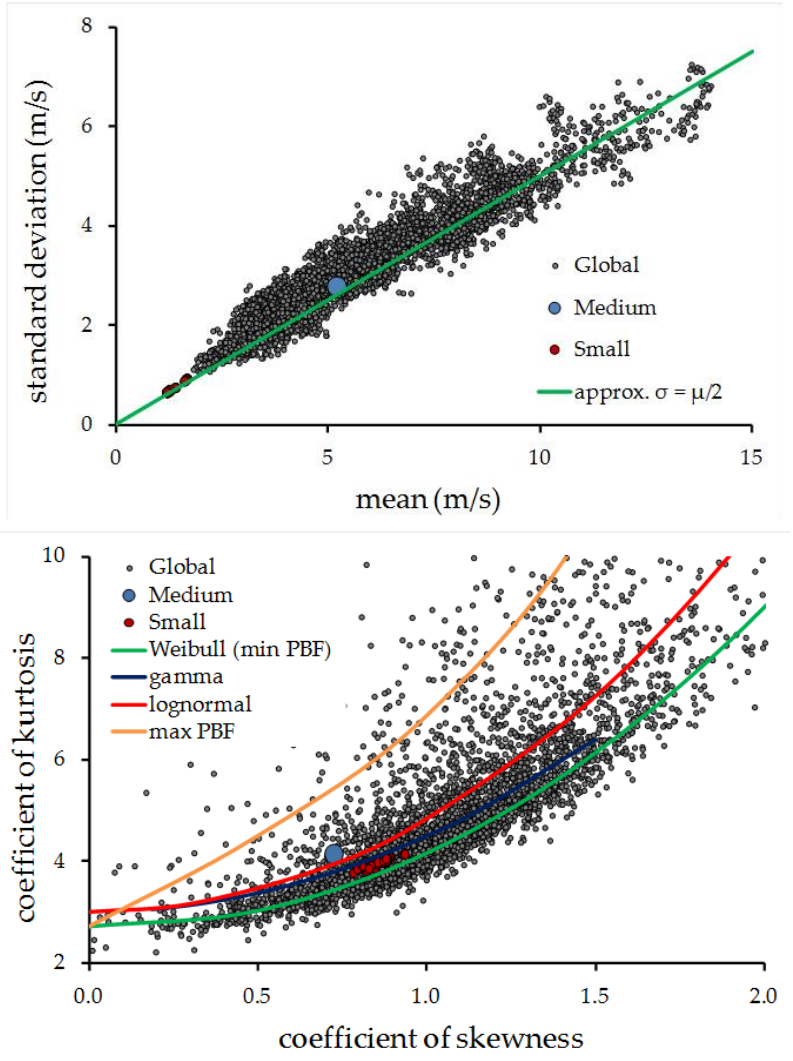

Figure 12 Standard deviation vs mean (upper) and coefficient of kurtosis vs. coefficient of skewness of all time series used in Application 2.

It can easily be proven that the length (norm) of a vector of random variables with uncorrelated Gaussian distributions with zero mean and equal variance follows the Rayleigh distribution. However, there is empirical and theoretical evidence (Application 1) that the small-scale distribution of turbulence is not Gaussian and it is expected that this should also be the case for the components of wind speed. Through Monte-Carlo experiments we illustrate that correlated non-Gaussian components result in a distribution close to Weibull and is in agreement with small and medium scale observations (an example is shown in Figure 13).

The distribution of the "global" time series appear to deviate from Weibull, gamma and lognormal distributions, and is closer to a distribution with a much heavier tail:

$$F(v) = 1 - \left(1 + \left(\frac{v}{\alpha v_s}\right)^b\right)^{-c/b} \tag{29}$$

where $v > 0$ is the wind speed, v_s is the standard deviation of the wind speed process; α is a scale parameter and b and c are the shape parameters of the marginal distribution, all three dimensionless. For this distribution we use the name Pareto- Burr-Feller (PBF)

to give credit to the engineer V. Pareto, who discovered a family of power-type distributions for the investigation of the size distribution of incomes in a society (Singh and Maddala, 1976), to Burr (1942) who identified and analysed (but without giving a justification) a function first proposed as an algebraic form by Bierens de Haan, and to Feller (1971) who linked it to the Beta function and distribution. Other names such as Pareto type IV or Burr type VII are also in use for the same distribution. Interestingly, the PBF distribution has two different asymptotic properties, i.e., the Weibull distribution for low wind speeds and the Pareto distribution for large ones. The derivation of PBF from maximum entropy has been studied in Yari and Borzadaran (2010). The PBF has been used in a variety of independent fields (see Brouers, 2015). Therefore, it seems that there is a strong physical as well as empirical justification for applying the PBF to the analysis of the wind process.

The distribution fitted to all data sets is shown in Figure 14 and the fitted parameters are α = 3.5, b = 1.9, c = 8.5. The mean estimated climacograms from the data (Figure 15) indicate that the model of equation (27) is also applicable for the wind speed at all scales with parameters estimated as $\lambda \approx 1$, $M = 1/3$, $H = 5/6$ and α = 6 h.

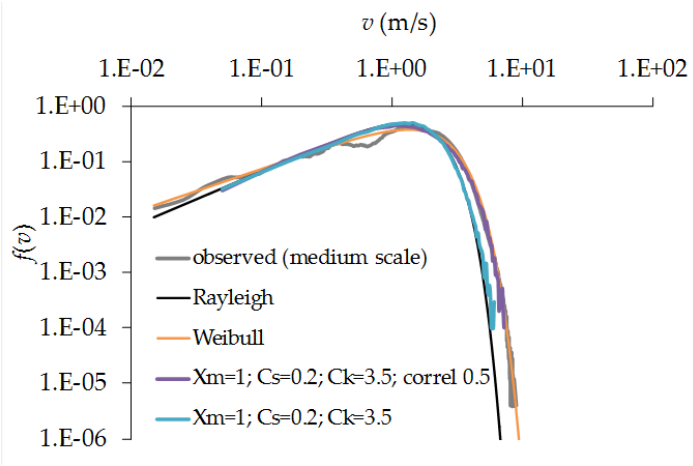

Figure 13 Probability density function of the medium scale time series along with theoretical and Monte Carlo generated distributions.

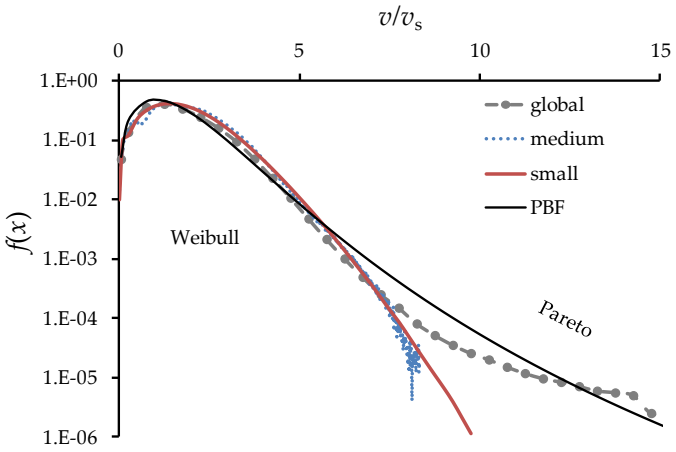

Figure 14 Probability density function of the velocity of grid-turbulent data (small) and of the wind speed of the medium and global scale time series along with fitted theoretical distributions.

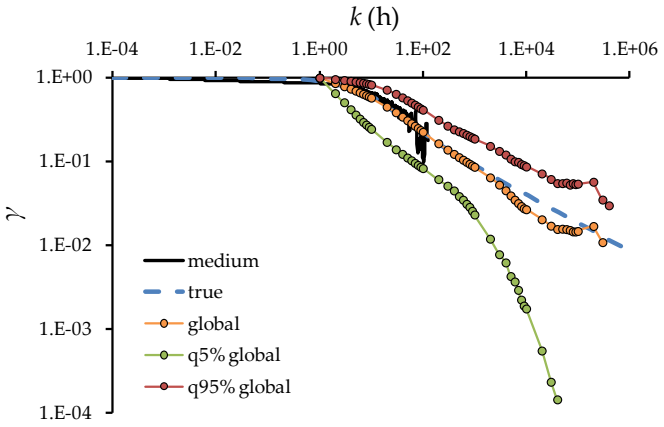

Figure 15 Climacogram of the wind speed process estimated from the medium and global series.

Application 3: Temperature

In this last application we analyse the dependence structure of the air temperature process close to surface. For the microscale structure, we use a 10 Hz resolution time series recorded for a two-month period via a sonic anemometer at Beaumont, USA (https://data.eol.ucar.edu/dataset/45.910). For the macro-scale structure, we use a global database of hourly air temperature (https://www.ncdc.noaa.gov/data-access/land-based-station-data). In total, we analyse over 5000 stations from different sites and climatic regimes by selecting time series with at least 1 year length and at least one measurement per three hours (Figure 16).

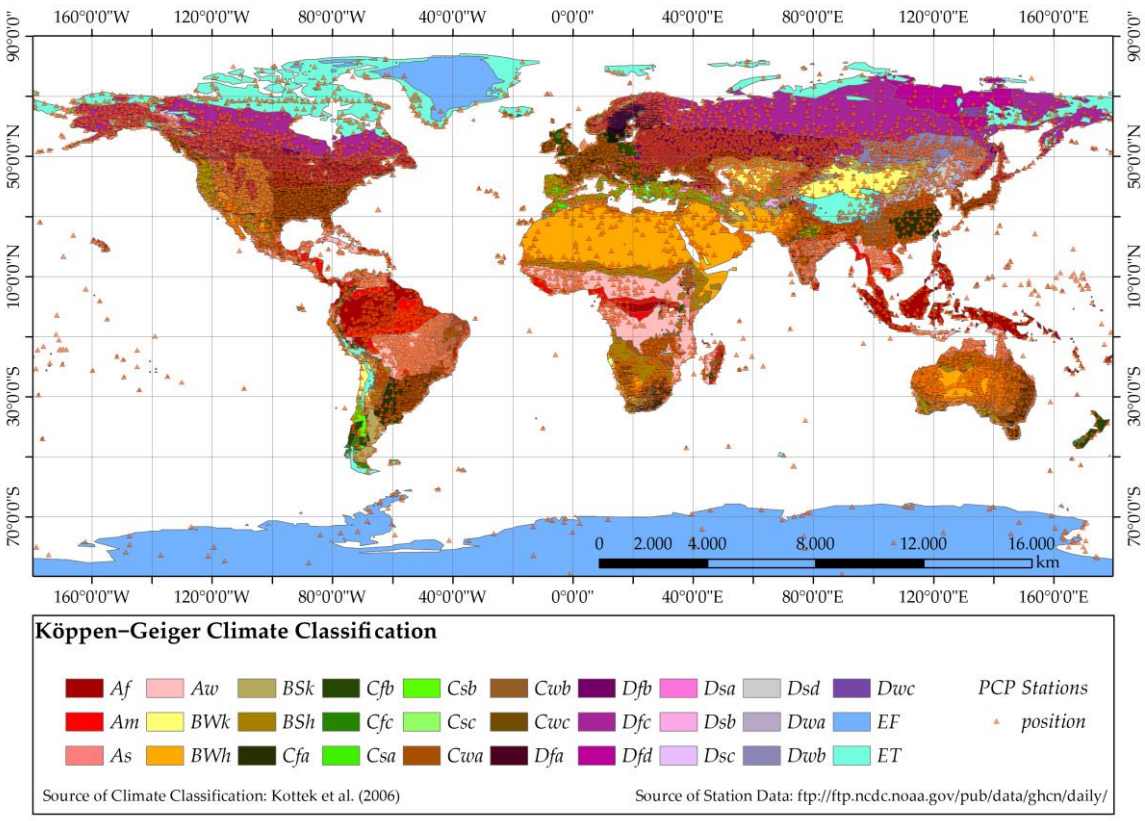

Figure 16 Locations of the selected hourly time series of air temperature.

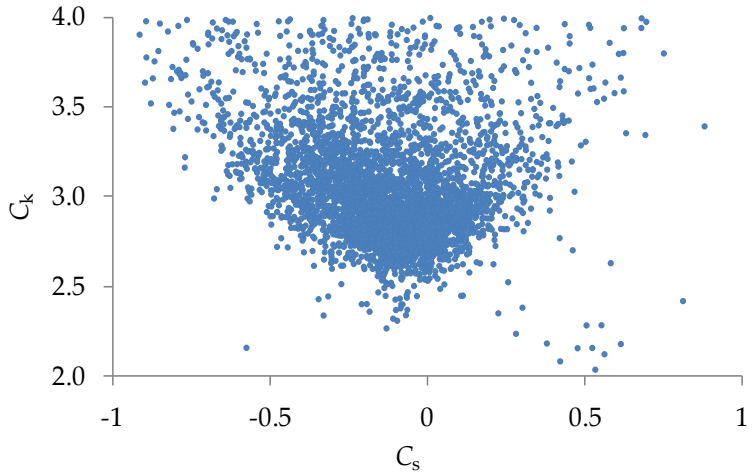

Figure 17 Coefficient of skewness vs. coefficient of kurtosis for ~90% of the macro-scale temperature time series.

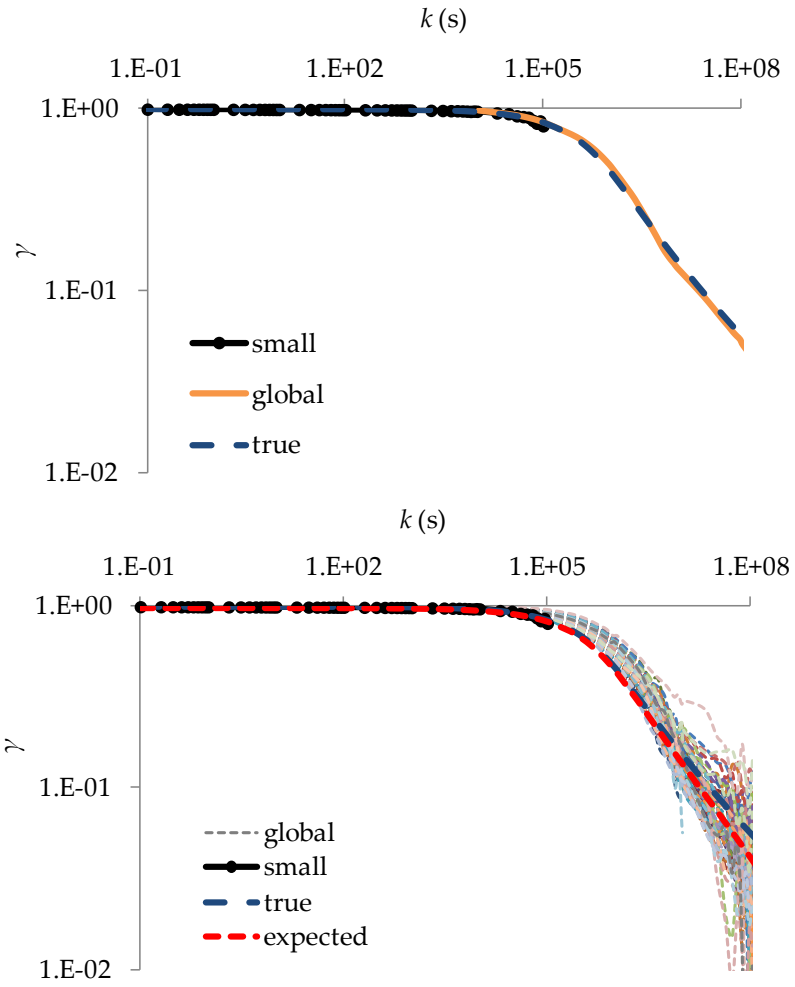

Figure 18 Climacogram of the normalized temperature for the micro-scale time series (small) and the set of hourly air temperature time series (global; upper: average climacogram; lower: climacograms of 100 different time series), compared to the fitted model of equation (27) (true and expected).

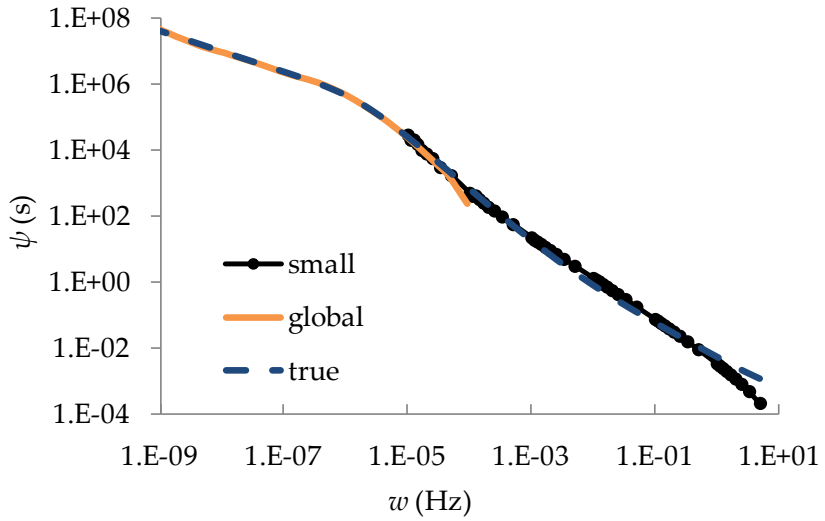

Figure 19 CS of the normalized temperature for the micro-scale time series (small) and the set of hourly air temperature time series (global; average from all time series), compared to the fitted model of equation (27) (true).

It can be assumed that the air temperature process follows a Gaussian distribution (Koutsoyiannis 2005). Indeed, Figure 17 shows that the 90% of the time series have skewness around 0±1 and kurtosis around 3±1. We normalize all time series and we estimate the dependence structure through the climacogram, autocovariance and power spectrum.

The mean estimated climacograms from the data (Figure 18) and the CS (Figure 19) indicate that, interestingly, the model of equation (27) is also applicable here with parameters estimated as $\lambda \approx 1$, $M = 1/3$, $H = 5/6$ and $\alpha = 3.3$ d.

6. Concluding remarks

Stochastics offers a strong basis for modelling and interpretation of natural behaviours and can directly incorporate, in a rigorous manner, useful concepts from the fractal literature, removing the ambiguity characterizing many fractal studies. Stochastics offers all tools for data analysis, inductive inference and prediction with quantified uncertainty, but above all it offers the basis for a logical world view.

We owe the well-founded and rigorous mathematical theory of stochastics to Kolmogorov (1931, 1933, 1938), including the foundation of scaling processes (Kolmogorov, 1940). This theory has often been distorted but there exist textbooks consistent with it (e.g. Papoulis, 1991).

Calculating values of sample statistics without considering their statistical properties (bias and statistical variation) can yield misleading results. Without proper attention to the underlying stochastics, we can even "identify" phenomena that do not exist and take statistical sampling effects as natural behaviours.

A general methodology for data analysis and construction of synthetic time series is possible provided that we have a good understanding of stochastics. In particular, the applications presented here suggest a promising characterization of different

geophysical processes in a unified manner and with a simple and parsimonious stochastic model, appropriate for a range of scales spanning several orders of magnitude.

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

This proof is summarized here from Koutsoyiannis (2013b) and Koutsoyiannis et al. (2013).

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

$$\gamma(k) = \int_0^{\infty} s(w)\ \operatorname{sinc}^2(\pi w k)\, \mathrm{d}w \tag{30}$$

The sinc² function within the integral takes significant values only for $w < 1/k$ (cf. Papoulis, 1991, p. 433). Assuming a scale $k \gg 1/\varepsilon$,

$$\gamma(k) = \int_0^{\infty} s(w) \operatorname{sinc}^2(\pi w k)\, \mathrm{d}w \approx \int_0^{\varepsilon} \alpha w^{\beta}\ \operatorname{sinc}^2(\pi w k)\, \mathrm{d}w \tag{31}$$

On the other hand, it can be easily seen that, for $0 < w < 1/k$, the following inequality holds:

$$\operatorname{sinc}(\pi w k) \ge 1 - wk \ge 0 \tag{32}$$

Since $\varepsilon \gg 1/k$, while the function in the integral (31) is nonnegative,

$$\begin{aligned} &\gamma(k) \approx \textstyle\int_0^{\varepsilon} \alpha w^{\beta}\ \operatorname{sinc}^2(\pi w k)\, \mathrm{d}w \ge \int_0^{1/k} \alpha w^{\beta}\ \operatorname{sinc}^2(\pi w k)\, \mathrm{d}w \ge \\ &\textstyle\int_0^{1/k} \alpha w^{\beta} (1 - wk)^2\, \mathrm{d}w \end{aligned} \tag{33}$$

By substituting $\omega = wk$ into equation (33), we find:

$$\gamma(k) \ge a k^{-\beta-1} \int_0^{1} \omega^{\beta} (1 - \omega)^2\, \mathrm{d}\omega \tag{34}$$

To evaluate the integral in (34) we take the limit for $q \to \infty$ of the integral:

$$B(q) \coloneqq \int_{1/q}^{1} \omega^{\beta} (1 - \omega)^2\, \mathrm{d}v\ = \frac{1 - q^{-1-\beta}}{1 + \beta} - 2\frac{1 - q^{-2-\beta}}{2 + \beta} + \frac{1 - q^{-3-\beta}}{3 + \beta} \tag{35}$$

Clearly, the limit of $B(q)$ as $q \to \infty$ depends on that of the term with the highest exponent, i.e. $q^{-1-\beta}$. For β < –1 this term diverges and thus, $B(0)$ = +∞. Then, by virtue of the inequality (34), $\gamma(k)$= ∞. For a (mean) ergodic processes $\gamma(k)$ should necessary tend to 0 for $k \to \infty$ (Papoulis, 1991, p. 429). Therefore, the process is non-ergodic.

It is interesting to note here that, when $|\beta| < 1$, the integral in (31) can be evaluated to give:

$$\gamma(k) \approx \alpha \int_0^\infty w^\beta \ \mathrm{sinc}^2(\pi w \Delta)\, \mathrm{d}w = \frac{\alpha \Gamma(1+\beta)\, \mathrm{sinc}(\pi\beta/2)}{2(1-\beta)(2\pi)^\beta k^{1+\beta}} \tag{36}$$

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

This analysis for $\beta < -1$ generalizes a result by Papoulis (1991, p. 434) who shows that an impulse at $w = 0$ corresponds to a non-ergodic process.

Appendix B: Literature review on the distribution function of wind speed

A large variety of distributions in the literature (with the most common to be Gaussian, gamma, Weibull, lognormal, Pareto and generalizations thereof as well as mixtures with each other) show equally good agreement with atmospheric wind measurements recorded at different sites around the globe with different climatic conditions.

Table B1 Recent publications on the distribution function of wind speed.

Reference	General characteristics	Proposed distribution	Comments
Aksoy et al. (2004)	1 station; 4 years	Weibull	Markov chain
Monahan (2006)	Global; sea-surface; wind speed	Weibull	Non Rayleigh
Bottcher et al. (2007)	Laboratory; 4 atmospheric stations; wind components	Castaing (1990)	Standard deviation with a lognormal model for intermittency
Kiss and Janosi (2008)	Reanalysis data over Europe	Generalized gamma	Non Rayleigh; non Weibull
Morgan et al. (2011)	178 offshore time series; 10-min wind speed	Kappa	14 distribution tested; non Weibull; non Rayleigh
Lo Brano et al. (2011)	Wind speed over Palermo	Burr	Tested: Weibull, Rayleigh, lognormal, gamma, inverse-Gaussian, Pearson V
Drobinski and Coulais (2012)	3 stations; high altitude; wind components	Rayleigh-Rice	Non Weibull, Elliptical distribution to model skewness
Wu et al. (2013)	Inner Mongolia region	Lognormal	Weibull; logistic
Ouarda et al. (2015)	9 stations in United Arab Emirates	Kappa, general-ized gamma	18 distributions tested with mixture properties

A sample of recent publications are listed in Table B1 along with the proposed distributions. However, some distributions seem to exhibit good agreement with data at the left or right tail mostly due to different lengths of the examined time series, while arguably most distributions do not exhibit good agreement for the whole range.

References

Aksoy, H.Z., Fuad, T., Aytek, A., Erdem N., (2004), Stochastic generation of hourly mean wind speed data, *Renewable Energy*, (29), 2111-2131.

Bartlett, M.S., (1948), Smoothing periodograms from time series with continuous spectra, *Nature*, 161(4096), 686-687, doi: 10.1038/161686a0.

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

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

Bottcher, F., Barth, S., and Peinke, J., (2007), Small and large scale fluctuations in atmospheric wind speeds, *Stochastic Environmental Research and Risk Assessment*, 21, 299–308.

Brouers, F. (2015), The Burr XII distribution family and the maximum entropy principle: power-law phenomena are not necessarily nonextensive, Open J Stat 5:730–741.

Burr, I.W., (1942), Cumulative Frequency Functions, *Annals of Mathematical Statistics*, 13, 215–235.

Castaing, B., Gagne, Y. and Hopfinger, E.J., (1990), Velocity probability density functions of high Reynolds number turbulence, *Physica D*, 46, 177-200.

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

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

Dimitriadis, P., and Koutsoyiannis, D. (2017), Stochastic synthesis approximating any process dependence and distribution, *Stochastic Environmental Research & Risk Assessment* (in review).

Dimitriadis, P., Koutsoyiannis, D., and Papanicolaou, P. (2016), Stochastic similarities between the microscale of turbulence and hydrometeorological processes, *Hydrological Sciences Journal*, 61 (9), 1623–1640.

Doran, C. (2011), Anemometer - Sonic at ABLE Beaumont Site Data. Version 1.0. UCAR/NCAR - Earth Observing Laboratory, http://data.eol.ucar.edu/dataset/45.910. Accessed 07 Jan 2017.

Drobinski, P., and Coulais, C., (2012), Is the Weibull distribution really suited for wind statistic modelling and wind power evaluation, *Journal of Physics*, Conference Series 753: 5 – 8.

Falconer, K. (2014), *Fractal Geometry: Mathematical Foundations and Applications*, 3rd edition, John Wiley & Sons, Chichester, UK.

Feller, W., (1970), *An Introduction to Probability and its Applications*, Vol II, 2nd ed. John Wiley & Sons, New York.

Frisch, U. (2006), Turbulence: The Legacy of A. N. Kolmogorov, *Cambridge University Press*, Cambridge.

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

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

Grassberger, P. and Procaccia, I. (1983), Characterization of strange attractors, *Physical Review Letters*, 50 (5), 346–349.

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

Jaynes, E.T. (1957), Information theory and statistical mechanics, *Phys. Rev.,* 106**,** 620.

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.

Kantelhardt, J. W. (2009), Fractal and Multifractal Time Series, In Meyers, R. A. (ed.) *Encyclopedia of Complexity and Systems Science*, vol. LXXX, 3754–3778, Springer, Berlin, DE.

Kiss, P., Janosi, I.M., Comprehensive empirical analysis of ERA-40 surface wind speed distribution over Europe, *Energy Convers Manage*, 49(8), 2142-51, doi:10.1016/j.enconman.2008.02.003.

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

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

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

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

Kolmogorov, A.N., (1941), Dissipation energy in locally isotropic turbulence, *Doklady Akademii Nauk SSSR*, 32, 16-18,.

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. (2006), On the quest for chaotic attractors in hydrological processes, *Hydrological Sciences Journal*, 51(6), 1065–1091.

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

Koutsoyiannis, D. (2010b), A random walk on water, *Hydrology and Earth System Sciences*, 14, 585–601.

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

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

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

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

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

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

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

Lo Brano, V., Orioli, A., Ciulla, G., Culotta, S., (2011), Quality of wind speed fitting distributions for the urban area of Palermo, Italy. *Renew Energy*, 36, 1026-39.

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.

Mahrt, L., (1989), Intermittency of atmospheric turbulence, *J. Atmos. Sci.*, 46, 79–95.

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

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

Mandelbrot, B. B. and Van Ness, J. W. (1968), Fractional Brownian motions, fractional noises and applications, *SIAM Review*, 10, 422–437.

Manwell, J.F., McGowan, J.G., Rogers A.L., (2010), *Wind Energy Explained*, 2nd ed., Amherst, USA: Wiley.

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.

Markonis, Y. and Koutsoyiannis, D. (2016), Scale-dependence of persistence in precipitation records, *Nature Climate Change,* 6, 399–401.

Monahan, A.H., (2006), The probability distribution of sea surface wind speeds, Part I, Theory and sea winds observations, *J. Clim.*, 19, 497-520.

Monahan, A.H., (2013), The Gaussian statistical predictability of wind speeds, J. Clim., 26, 5563–5577.

Morgan E.C., Lackner M., Vogel R.M. and Baise L.G. (2011),Probability distributions for offshore wind speeds, *Energy Conversion and Management*, 52(1),15-26.

O'Connell, P. E., Koutsoyiannis, D., Lins, H. F., Markonis, Y., Montanari, A. and Cohn, T. A. (2016), The scientific legacy of Harold Edwin Hurst (1880 – 1978), *Hydrological Sciences Journal*, 61 (9), 1571–1590.

Ouarda T.B.M.J., Charron, C., Shin, J.Y., Marpu P.R., Al-Mandoos, A.H., Al-Tamimi, M.H., et al., (2015), Probability distributions of wind speed in the UAE, *Energy Convers. Manag.*, 93,414-34.

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

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

Popper, K. R. (1982), *The Open Universe: An Argument For Indeterminism*, Hutchinson, London, UK.

Scholz, C. H., Mandelbrot, B. B. (1989), *Fractals in Geophysics*, Birkhäuser Verlag, Basel, DE.

She, Z.S., and Leveque, E., (1994), Universal scaling laws in fully developed turbulence, *Phys. Rev. Lett.*, 72, 336.

Singh, S.K. and Maddala, G.S. (1976), A function for size distribution of incomes, *Econometrica*, 44, 963-970.

Stumpf, M. P. H. and Porter, M. A. (2012), Critical truths about power laws, *Science*, 335, 665–666.

Tessier, Y., Lovejoy, S., and Schertzer, D. (1993), Universal multifractals: theory and observations for rain and clouds, *Journal of Applied Meteorology*, 32(2), 223-250.

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

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

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

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

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

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

Wua J., Wanga J., Chib D., (2013), Wind energy potential assessment for the site of Inner Mongolia in China, *Renew Sustain Energy Rev*,21:215–28.

Yari, G.H., and Borzadaran, G.R. M., (2010), Entropy for Pareto-types and its order statistics distributions, *Communications in Information and Systems*, 10(3), 193–201.