

Knowable moments for high-order stochastic characterization and modelling of hydrological processes

Demetris Koutsoyiannis

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

Abstract. Classical moments, raw or central, express important theoretical properties of probability distributions but can hardly be estimated from typical hydrological samples for orders beyond two. L-moments are better estimated but they all are of first order in terms of the process of interest; while they are effective in inferring the marginal distribution of stochastic processes, they cannot characterize even second order dependence of processes (autocovariance, climacogram, power spectrum) and thus they cannot help in stochastic modelling. Picking from both categories, we introduce knowable (K-) moments, which combine advantages of both classical and L-moments, and enable reliable estimation from samples and effective description of high-order statistics, useful for marginal and joint distributions of stochastic processes. Further, we extend recent stochastic tools by introducing the K-climacogram and the K-climacospectrum, which enable characterization, in terms of univariate functions, of high-order properties of stochastic processes, as well as preservation thereof in simulations.

Introduction

In a hydrological decade that is featured by the Heraclitean aphorism «Πάντα ῥεῖ» (Panta rhei – everything flows; Montanari et al. 2013), the scientific field of stochastics offers a great deal of opportunities to study change (Koutsoyiannis 2013). In particular the theory of stationary stochastic processes, if understood, enables sophisticated yet simple characterization and modelling of hydrological processes in a framework of unceasing change (Montanari and Koutsoyiannis 2014; Koutsoyiannis and Montanari 2015).

Statistical moments, studied at a multitude of time scales using observed data, constitute the basic tool of stochastic characterization of change and variability. They offer the option to describe probability distributions summarily in order to achieve reasonable simplicity (Feller 1968). However, classical moments, raw or central, cannot be reliably estimated from typical samples for orders beyond two. Lombardo *et al.* (2014) expressed this fact in the title of their paper: “*Just two moments!*”. On the other hand, for non-normal distributions, which are the case in most hydrological and geophysical processes, two summary statistics are not enough to characterize the distributions. Also in stochastic synthesis of those processes, moments of order higher than two are necessary (Dimitriadis and Koutsoyiannis 2018; Koutsoyiannis *et al.* 2018).

L-moments (Hosking 1990) are better estimated but they all are of first order in terms of the process of interest. While they are effective in characterizing independent series or in inferring the marginal distribution of stochastic processes, they cannot characterize and model even second order dependence of processes (and hence change). For example, Hosking and Wallis (2005), while developing an approach for regional frequency analysis based on L-moments, resort to the classical second-order moments (variances and covariances) when they describe correlation.

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

$$\mu'_p := E[\underline{x}^p], \quad \mu_p := E[(\underline{x} - \mu)^p] \quad (1)$$

respectively, where $\mu := \mu'_1 = E[\underline{x}]$ is the mean of the random variable \underline{x} and $E[\]$ denotes expectation; notice the notation of random variables (and hence stochastic processes) with underlined symbols, according to the Dutch convention (Hemelrijk 1966). The standard moment estimators from a sample $\underline{x}_i, i = 1, \dots, n$, are:

$$\hat{\mu}'_p = \frac{1}{n} \sum_{i=1}^n \underline{x}_i^p, \quad \hat{\mu}_p = \frac{b(n,p)}{n} \sum_{i=1}^n (\underline{x}_i - \hat{\mu})^p \quad (2)$$

where $b(n, p)$ is a bias correction factor (e.g. for the variance $\mu_2 =: \sigma^2$, $b(n, 2) = n/(n-1)$ if \underline{x}_i are uncorrelated). The estimators of the noncentral moments $\hat{\mu}'_p$ (or even the central ones if μ is known a priori) are in theory unbiased, but it is impractical to use them in estimation if $p > 2$ (cf. Lombardo *et al.* 2014).

It is well known that for large p and positive x_i (more generally, for x_i satisfying the condition $\max_{1 \leq i \leq n}(x_i) > |\min_{1 \leq i \leq n}(x_i)|$), the following approximate relationship holds:

$$\left(\sum_{i=1}^n x_i^p \right)^{1/p} \approx \max_{1 \leq i \leq n}(x_i) \quad (3)$$

(This is related to the well-known mathematical fact that the maximum norm is the limit of the p -norm as $p \rightarrow \infty$.) A numerical illustration of how fast the convergence of the left-hand side to the right-hand side of equation (3) is provided in Table 1.

Table 1 Illustration of the fact that raising to a power and adding converges fast to the maximum value.

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

Therefore, for relatively large p the estimate of μ'_p will be:

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

(Note that for large p the term $(1/n)$ in the right-hand side is sometimes omitted with a negligible error; see explanation in equation (15) below). Thus, for an unbounded variable \underline{x} and for large p , we can conclude that $\hat{\mu}'_p$, while theoretically is an unbiased estimator of μ'_p , in practice it is more an estimator of an extreme quantity, i.e., the n th order statistic raised to power p , than an estimator of μ'_p . (It is reminded that, by definition, the i th order statistic $\underline{x}_{(i)}$ is equal to the random variable having the i th-smallest value in the sample of size n .) This happens because the convergence of $\hat{\mu}'_p$ to μ'_p is very slow, while the convergence to the maximum value is fast. This is further illustrated in Figure 1 for the eighth moment of a process specified in the figure caption. While even for n as large as 64 000 the sample moment estimate continues to be smaller, by several moments of magnitude, than the theoretical value, the proximity of the moment estimate to the maximum value is evident even for n as small as 10. The jagged shapes of the curves are a clear indication of the dominance of maxima in the moment estimation: the steps occur when a new higher maximum value enters the sample, while the gradual decreases before those are due to the increase of the sample size without a higher maximum value. The ensemble simulation results in the right panel show that the 99% prediction limits from 1000 simulations are not able even to envelop the true value.

As a result, unless p is very small, μ'_p is not a *knowable* quantity: we cannot infer its value from a sample. This is the case even if n is very large as in Figure 1 (see also Figure 2 explained below). Also, the various $\hat{\mu}'_p$ are not independent to each other as they only differ on the power to which the maximum value is raised.

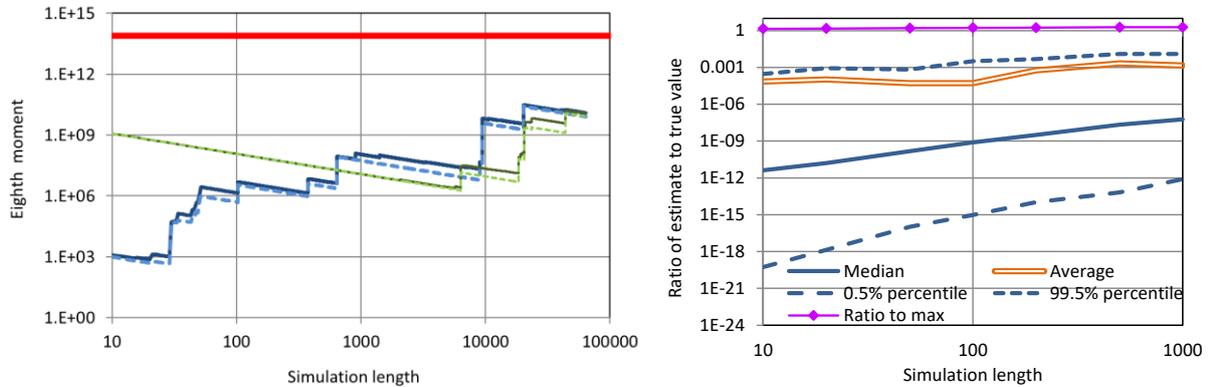


Figure 1 Illustration of the slow convergence of the sample estimate of the eighth noncentral moment to its true value, which is depicted as a thick horizontal line and corresponds to a lognormal distribution $\text{LN}(0,1)$ where the process is an exponentiated Hurst-Kolmogorov process (Koutsoyiannis 2016) with Hurst parameter $H = 0.9$. (Left) The sample moments are estimated from a single simulation of that process with length 64 000, where parts of this time series with sample size n from 10 to 64 000 are used for the estimation. Subsetting of the time series to sample size n was done either from the beginning to the end (thicker lines) or from the end to the beginning (finer lines). Continuous lines in the two cases represent the eighth moment estimates, $\sum_{i=1}^n x_i^p / n$, and dashed lines represent maximum values, $(\max_{1 \leq i \leq n} (x_i))^p / n$. (Right) Sampling distribution of the eighth moment estimator $\sum_{i=1}^n x_i^8 / n$ estimated from 1000 simulated series of length 1000 each and visualized by the 99% prediction limits (percentiles), the median and the average, plotted as ratios to the true value. Theoretically, the ratio should be 1, but it is smaller by many orders of magnitude, and the convergence to 1 is very slow. The ratio to $(\max_{1 \leq i \leq n} (x_i))^8 / n$, also plotted, is close to 1.

Picking from both concepts of classical and L-moments, below we introduce a general category of *knowable* or K-moments, which combine advantages of (and virtually contain) both categories, classical and L-moments. They enable reliable estimation from samples (in some cases even more reliable than L-moments) and effective description of statistics of high order, useful for marginal and joint distributions of stochastic processes. They also allow multiscale characterization of high-order properties of stochastic processes using univariate functions, thus avoiding the common multivariate functions expressing joint high-order moments. This is made possible by extending the notion of climacogram (variance of the averaged process vs. time scale of averaging; Koutsoyiannis 2010, 2016) and climacospectrum (transformation of the climacogram; Koutsoyiannis 2017). Specifically, using K-moments we introduce the K-climacogram and the K-climacospectrum, which in addition to characterization, in terms of univariate functions, of high-order properties of stochastic processes, enable preservation thereof in simulations. Overall, the manuscript discusses all these issues, from the definition and rationale of K-moments, K-climacogram and K-climacospectrum, to their use in characterizing hydrological time series (with some case studies) as well as in stochastic simulation of processes.

Definition and rationale of K-moments

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

$$K_{pq} := (p - q + 1)E\left[(2F(\underline{x}) - 1)^{p-q}(\underline{x} - \mu)^q\right], \quad p \geq q \quad (5)$$

The usefulness of the factor $(p - q + 1)$ will be explained in a while. Likewise, we define noncentral K-moments as:

$$K'_{pq} := (p - q + 1)E\left[\left(F(\underline{x})\right)^{p-q} \underline{x}^q\right], \quad p \geq q \quad (6)$$

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

$$\hat{F}(x_{(i)}) = \frac{i - 1}{n - 1} \quad (7)$$

thus taking values from 0 to 1 precisely and irrespective of the values $x_{(i)}$; likewise, $2F(x_{(i)}) - 1$ is estimated as

$$2\hat{F}(x_{(i)}) - 1 = \frac{2i - n - 1}{n - 1} \quad (8)$$

taking values from -1 to 1 precisely and irrespective of the values $x_{(i)}$. Hence, the estimators are:

$$\hat{K}'_{pq} = \frac{p-q+1}{n} \sum_{i=1}^n \binom{i-1}{n-1}^{p-q} \underline{x}_{(i)}^q, \quad \hat{K}_{pq} = \frac{p-q+1}{n} \sum_{i=1}^n \binom{2i-n-1}{n-1}^{p-q} (\underline{x}_{(i)} - \hat{\mu})^q \quad (9)$$

We note that to form these estimators our main provision was simplicity while unbiasedness was not given priority. Hence these estimators can be biased, particularly for high p . The study of the estimation bias and the possible adjustment of the estimators to become unbiased are interesting problems, but they fall out the scope of this paper (see also the Conclusions section).

To illustrate the rationale of the definition of K-moments, we assume that the distribution mean is close to the median, so that $F(\mu) \approx 1/2$. This equality is precise for a symmetric distribution. The quantity whose expectation is taken in (5) is:

$$A(\underline{x}) := (2F(\underline{x}) - 1)^{p-q} (\underline{x} - \mu)^q = B(\underline{x})(\underline{x} - \mu)^q, \quad B(\underline{x}) := (2F(\underline{x}) - 1)^{p-q} \quad (10)$$

Now the m th derivative of $B(\underline{x})$ for $m < p - q$ will be a sum of terms, each one of which will contain $(2F(\underline{x}) - 1)$ to some positive power. Thus, if we evaluate this derivative at the point $\underline{x} = \mu$, by virtue of the assumption $F(\mu) \approx 1/2$, we will get zero. For $m \geq p - q$ this is no longer the case, and therefore the Taylor expansion of $B(\underline{x})$ will have nonzero terms for powers $\geq p - q$. Hence the Taylor expansion of $A(\underline{x})$ will have nonzero terms for powers $\geq p$ only. More specifically, under the assumption $F(\mu) = 1/2$, it can be verified that the Taylor expansion is:

$$A(\underline{x}) = (2f(\mu))^{p-q} (\underline{x} - \mu)^p + (p-q)(2f(\mu))^{p-q-1} f'(\mu) (\underline{x} - \mu)^{p+1} + O((\underline{x} - \mu)^{p+2}) \quad (11)$$

where $f(x)$ is the probability density function of \underline{x} . Clearly then, K_{pq} depends on μ_p as well as classical central moments of \underline{x} of order higher than p . The independence of K_{pq} from classical moments of order $< p$ means that the information contained in each K_{pq} is free of information related to classical moments μ_i of order $i < p$, and this makes K_{pq} a good *knowable* surrogate of the unknowable μ_p . We note that the assumption $F(\mu) \approx 1/2$ is only used for the explanation of the rationale and is not necessary for the precise definition.

To further explain the rationale we discuss the inclusion of the multiplicative term $(p - q + 1)$ in definitions (5) and (6). A first reason for the inclusion is that, as a rule for positive x , it makes K-moments increasing functions of p , which is intuitive and consistent to the behaviour of classical moments. Another, more important, reason is that, as p becomes large, by virtue of this multiplicative term, K'_{pq} shares similar asymptotic properties with $\hat{\mu}_p^{q/p}$ (the estimate, not the true $\mu_p^{q/p}$). To illustrate this for $q = 1$ and for independent variables \underline{x}_i , we consider the variable $\underline{z}_p := \max_{1 \leq i \leq p} \underline{x}_i$ and denote $f(\cdot)$ and $h(\cdot)$ the probability densities of \underline{x}_i and \underline{z}_p , respectively. Then (Papoulis 1990, p. 209):

$$h(z) = pf(z)(F(z))^{p-1} \quad (12)$$

This is none other than the derivative of $(F(z))^p$, while for independent variables this latter, i.e., the product of p instances of $F(z)$, is the probability distribution of \underline{z}_p . Thus, according to (6),

$$E[\underline{z}_p] = pE\left[\left(F(\underline{x})\right)^{p-1} \underline{x}\right] = K'_{p1} \quad (13)$$

On the other hand, because of equation (4), for positive \underline{x} and large $p \rightarrow n$,

$$\left(\mathbb{E} \left[\hat{\mu}'_p \right] \right)^{1/p} = \left(\mathbb{E} \left[\left(\frac{1}{n} \sum_{i=1}^n x_i^p \right) \right] \right)^{1/p} \approx \left(\mathbb{E} \left[\frac{1}{n} \max_{1 \leq i \leq n} (x_i^p) \right] \right)^{1/p} \approx n^{-1/p} \mathbb{E} \left[\max_{1 \leq i \leq n} x_i \right] \approx \mathbb{E} [Z_n] \quad (14)$$

because for large n and $p \rightarrow n$, $n^{-1/p} \rightarrow 1$. In other words, as $p \rightarrow n$,

$$\left(\mathbb{E} \left[\hat{\mu}'_p \right] \right)^{1/p} \approx K'_{p1} \quad (15)$$

This property reconciles theoretical and sample moments, as it reflects the fact that asymptotically $\hat{\mu}'_p$ is indeed an estimator of a theoretical quantity that surprisingly can be closer to K'_{p1} than to μ'_p .

Illustration of the above remarks is provided in Figure 2, where theoretical moments for the normal distribution $N(0,1)$ and lognormal distribution $LN(0,1)$ are plotted and compared with sample estimates from synthetic series generated for these distributions with sample sizes ten times higher than the maximum p shown in graphs, i.e., 1000. It is evident in the figure that for the lognormal distribution and for $p > 3$ the sample estimates are irrelevant to (are smaller by orders of magnitudes than) the theoretical quantities they are supposed to estimate. Evidently, classical moments and their estimates do not describe the same thing. For the normal distribution the estimates of classical moments are impressively close to the theoretical K-moments and far from the theoretical classical moments. For the lognormal distribution, the proximity is less impressive yet the asymptotic tendency described by equation (15) is evident as is the divergence of $\hat{\mu}'_p$ from μ'_p . All in all, Figure 2 provides a graphical justification of the notion of unknowable vs. knowable (this will also be verified in Figure 7 with real-world data).

Summarizing, as p becomes large (approaching n), estimates of both classical and K-moments, central or noncentral, become estimates of expressions involving extremes (like $(\max_{1 \leq i \leq p} (x_i))^q$, $(\max_{1 \leq i \leq p} (x_i - \mu))^q$, etc., where the exponents q should be replaced with p for classical moments). (For negatively skewed distributions these quantities may involve minimum, instead of maximum quantities; this can be understood considering that if \underline{x} is negatively skewed, then $-\underline{x}$ is positively skewed, while $\min_{1 \leq i \leq p} (-x_i) = -\max_{1 \leq i \leq p} x_i$). For the K-moments this behaviour is consistent with their theoretical definition. For the classical moments and cumulants this is an inconsistency.

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

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

It is worthwhile to note that the problem of unknowability appears also in cumulants, which, as will be seen below, are quite useful in stochastic synthesis, but cannot be estimated from data reliably. In particular, the cumulant generating function is:

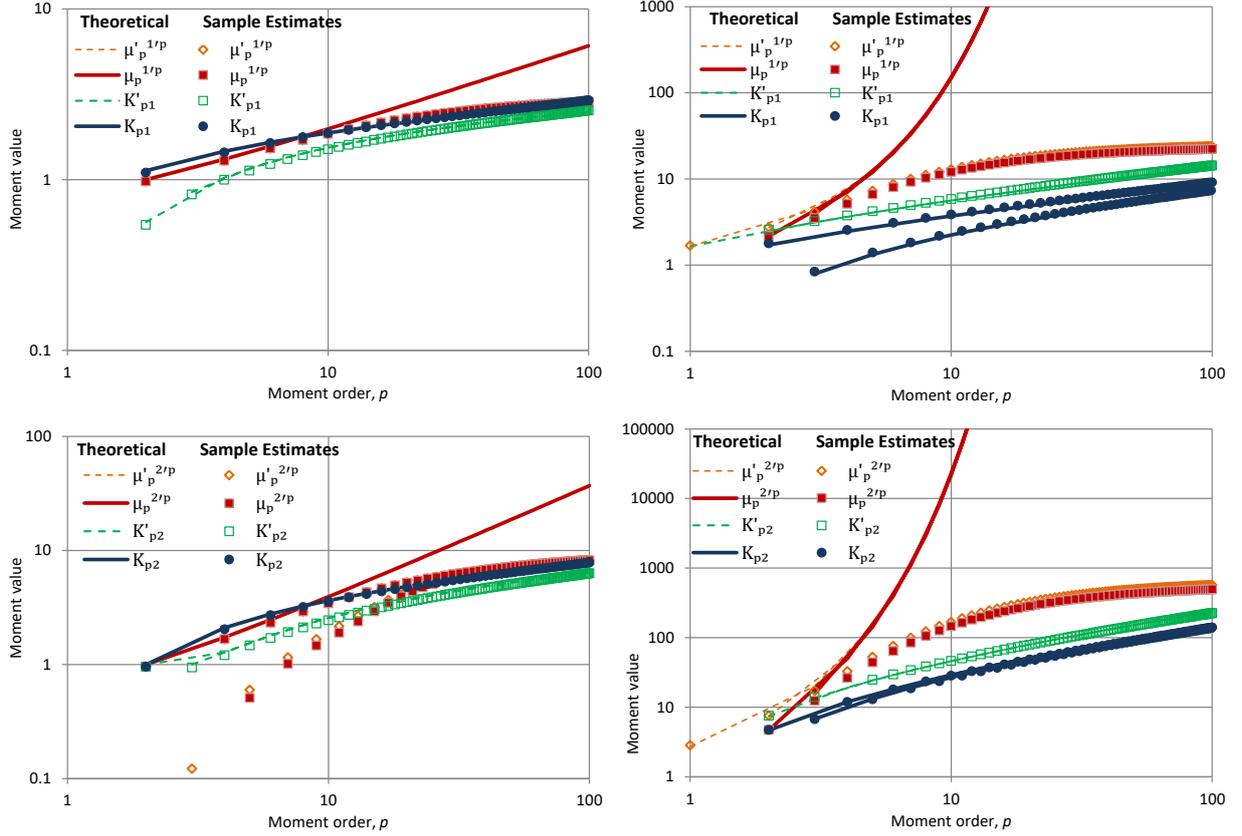


Figure 2 Illustration of agreement and disagreement of moment estimates with theoretical moments for: normal distribution $N(0,1)$ (left) and lognormal distribution $LN(0,1)$ (right), where the K -moments correspond to $q = 1$ (upper row) and $q = 2$ (lower row). Note that in the normal distribution theoretical moments of odd order are zero (and thus not shown in the logarithmic plots) and that points corresponding to μ'_p are (almost) indistinguishable from those corresponding to μ_p . In the lognormal distribution the moments of odd order are depicted by the lower blue lines and series of points (circles), while the upper ones depict even moments.

$$K(t) := \ln E[e^{t\underline{x}}] \quad (16)$$

and its standard estimator is (Ghosh and Beran 2006):

$$\hat{K}(t) = \ln \frac{\sum_{i=1}^n e^{t\underline{x}_i}}{n} \quad (17)$$

Considering that one form of the generalized (Kolmogorov) mean of variables $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ is precisely the estimator $\hat{K}(t)$, as well as the fact that (Lange *et al.* 2014):

$$\lim_{t \rightarrow \infty} \frac{1}{t} \ln \frac{\sum_{i=1}^n e^{t\underline{x}_i}}{n} = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \left(\sum_{i=1}^n e^{t\underline{x}_i} \right) - \lim_{t \rightarrow \infty} \frac{\ln n}{t} \approx \max_{1 \leq i \leq n} \underline{x}_i \quad (18)$$

we conclude that, for large t and independent variables \underline{x}_i ,

$$\frac{1}{t} E[\hat{K}(t)] \approx E \left[\max_{1 \leq i \leq n} \underline{x}_i \right] = E[Z_n] = K'_{n1} \quad (19)$$

In other words, again the estimate can be closer to $K'_{n1}t$ than to $K(t)$. The cumulants κ_p , are obtained from a power series expansion of the cumulant generating function, i.e.,

$$K(t) = \sum_{p=1}^{\infty} \kappa_p \frac{t^p}{p!} \quad (20)$$

and are related to moments of similar order (see equation (27) below). Since for large t the estimator $\hat{K}(t)$ is in fact estimator not of $K(t)$ but of an extreme quantity, again cumulants, similar to standard moments, are unknowable for large p .

Relationships among different moment types

The classical moments can be recovered as a special case of K-moments:

$$K'_{pp} \equiv \mu'_p, \quad K_{pp} \equiv \mu_p \quad (21)$$

while other interesting special cases are

$$K'_{p0} = 1, \quad K_{p0} = \frac{(-1)^p + 1}{2} = \begin{cases} 1 & p \text{ even} \\ 0 & p \text{ odd} \end{cases}, \quad K'_{11} \equiv \mu, \quad K_{22} \equiv \sigma^2 \quad (22)$$

A particular case where classical and K-moments almost coincide (more precisely, are proportional to each other) is the uniform distribution, in which:

$$K'_{pq} := (p - q + 1)\mu'_p, \quad K_{pq} := (p - q + 1)\mu_p \quad (23)$$

The probability weighted moments (PWM; Greenwood *et al.* 1979) are also closely related to the (noncentral) K- moments. In particular, the most common PWM form $\beta_p := E \left[\underline{x} \left(F(\underline{x}) \right)^p \right]$ is the proportional to the noncentral K- moment corresponding to $q = 1$:

$$K'_{p1} = p\beta_{p-1} \quad (24)$$

The L-moments are defined by (Hosking 1990):

$$\lambda_p := \frac{1}{p} \sum_{k=0}^{p-1} (-1)^k \binom{p-1}{k} E[\underline{x}_{(p-k):p}] \quad (25)$$

where $\underline{x}_{k:p}$ denotes the k th order statistic in an independent sample of size p . L-moments are also related to PWM and through them to K-moments. In particular, the relationships for the different types of moments for the first four orders are:

$$\begin{aligned} K'_{11} &= \mu = \beta_0, & K_{11} &= 0 \\ K'_{21} &= 2\beta_1, & K_{21} &= 2(K'_{21} - \mu) = 4\beta_1 - 2\beta_0 = 2\lambda_2 \\ K'_{31} &= 3\beta_2, & K_{31} &= 4(K'_{31} - \mu) - 6(K'_{21} - \mu) = 12\beta_2 - 12\beta_1 + 2\beta_0 = 2\lambda_3 \\ K'_{41} &= 4\beta_3, & K_{41} &= 8(K'_{41} - \mu) - 16(K'_{31} - \mu) + 12(K'_{21} - \mu) \\ & & &= 32\beta_3 - 48\beta_2 + 24\beta_1 - 4\beta_0 = \frac{8}{5}\lambda_4 + \frac{12}{5}\lambda_2 \end{aligned} \quad (26)$$

Finally, the cumulants are related to the classical noncentral moments by the following recursive relationships (Smith 1995):

$$\mu'_p = \sum_{i=0}^{p-1} \binom{p-1}{i} \kappa_{p-i} \mu'_i, \quad \kappa_p = \mu'_p - \sum_{i=1}^{p-1} \binom{p-1}{i} \kappa_{p-i} \mu'_i \quad (27)$$

In the above moment categories, the initial values (necessary for recursive relationships) are: $\kappa_0 = \mu_1 = K_{11} = K_{2p+1,0} = 0$, $\mu_0 = \mu'_0 = K'_{p0} = K_{2p,0} = 1$, $\kappa_1 = \mu'_1 = K'_{11} = \beta_0 = \mu$. Relationships between central and noncentral moments are given in the Appendix.

Basic characteristics of marginal distribution

Within the framework of K-moments and according to the rule of thumb “Just two moments” we may assume that the power of \underline{x} , i.e. q , should be taken $q = 1$ or 2 and obtain knowable statistical characteristics for much higher order p . In this manner, for $p > 1$ we have two alternative options to define statistical characteristics related to moments of the distribution, the most customary of which are shown in Table 2. Which of the two options is preferable depends on the statistical behaviour and in particular, the mean, mode and variance of the estimator. For completeness, Table 2 contains also as Option 3 the classical case but this is not recommended for the fitting phase of a distribution.

For illustration of the numerical values of the statistical characteristics resulting from the various options of Table 2, Table 3 provides some comparisons for distribution functions of common use. These values are for theoretical (not sample) moments and have been calculated analytically (by integration) or numerically (by numerical integration). Numerical calculation of theoretical moments, when analytical integration is infeasible, is an easy matter and involves no difficulty; thus the existence of an analytical solution of theoretical moments of a certain distribution should not be regarded as an important criterion for choosing that distribution. The important issue for model fitting is whether the moments are knowable or not, in the sense of their estimation from a sample; their theoretical values are always knowable once the distribution parameters have been specified.

Table 2 Typical marginal statistical characteristics of distributions using different moment categories.

Characteristic	Order p	Option 1	Option 2	Option 3*
Location	1	$K'_{11} = \mu$ (the classical mean)		
Variability	2	$K_{21} = 2(K'_{21} - \mu) = 2\lambda_2$	$K_{22} = \mu_2 = \sigma^2$ (the classical variance)	
Skewness (dimensionless)	3	$\frac{K_{31}}{K_{21}} = \frac{\lambda_3}{\lambda_2}$	$\frac{K_{32}}{K_{22}}$	$\frac{K_{33}}{K_{22}^{3/2}} = \frac{\mu_3}{\sigma^3}$
Kurtosis (dimensionless)	4	$\frac{K_{41}}{K_{21}} = \frac{4\lambda_4}{5\lambda_2} + \frac{6}{5}$	$\frac{K_{42}}{K_{22}}$	$\frac{K_{44}}{K_{22}^2} = \frac{\mu_4}{\sigma^4}$

* Not recommended for distribution fitting.

Table 3 Numerical illustration of the variation of typical marginal statistical characteristics for some customary distributions.

Characteristic Distribution function*	Variability		Skewness			Kurtosis		
	K_{21} (1) [†]	K_{22} (2)	$\frac{K_{31}}{K_{21}}$ (1)	$\frac{K_{32}}{K_{22}}$ (2)	$\frac{K_{33}}{K_{22}^{3/2}}$ (3)	$\frac{K_{41}}{K_{21}}$ (1)	$\frac{K_{42}}{K_{22}}$ (2)	$\frac{K_{44}}{K_{22}^2}$ (3)
U(0,1)	0.3333	0.08333	0	0	0	1.2	1.8	1.8
N(0,1)	1.128	1	0	0	0	1.465	2.103	3
LN(0, 1/2)	0.6262	0.3647	0.2409	0.8146	1.750	1.335	2.239	8.898
LN(0, 1)	1.716	4.671	0.4625	1.455	6.185	1.434	2.548	113.9
LN(0, 2)	12.45	2926.4	0.7909	1.965	414.4	1.712	2.961	9.2×10^6
GP(1,1/6)	0.2181	0.06	0.4118	1.273	3.810	1.386	2.406	38.67
GP(1,1/4)	0.3810	0.2222	0.4545	1.429	7.071	1.418	2.522	∞
GP(1,1/3)	0.6	0.75	0.5	1.6	∞	1.455	2.657	∞

* U: Uniform; N: Normal; LN: Lognormal; GP: Generalized Pareto with lower bound zero; the numbers in parentheses are the parameter values of the distributions.

[†] The numbers in parentheses refer to the Options defined in Table 2.

It is seen in Table 3 that the values of the classical statistical measures of skewness and particularly kurtosis (Option 3) can be extraordinarily high and impractical, even though they are by definition nondimensionalized (a fortiori, the respective moments are even higher by orders of magnitude). On the contrary, both Options 1 and 2 provide reasonable and intuitive values while keeping the consistency in terms of their variation (increasing/decreasing behaviour with respect to change of distribution parameters and order of moments). Finally, as already explained, the results from Options 1 and 2 are finite (provided that mean and variance are finite) even when those of Option 3 are infinite.

It is useful to explore the variability of the sample estimates of these characteristics. Figure 3 illustrates this for two of the cases of Table 3, namely the normal distribution N(0,1) and the lognormal distribution LN(0, 2). The estimates from the three options are compared in terms of the resulting probability density functions of the empirical statistics corresponding to a sample with size 100, where the density functions are estimated by Monte Carlo simulations with 1000 repetitions. For a fair comparison of the statistics of different options, all of them were first transformed into the same “units” with the mean μ and then standardized by the respective parameter of variability, again with same units; the latter equals K_{21} for Option 1 ($q = 1$) and $\sigma \equiv \sqrt{K_{22}}$ for Options 2 ($q = 2$) and 3 (classical statistics). The exact definitions in each case are shown in the caption of Figure 3.

First of all, the plot indicates that the classical statistics have always worse performance, sometimes extraordinarily worse, than the K-statistics, as the spread of their density is wider and its peak smaller. Second, for the skewed distribution case (LN; right column) the statistics of Option 2 ($q = 2$) have the best performance. There is one exception though, the upper right panel, where Option 1 clearly outperforms Option 2, both in terms of the spread and bias. The

high bias of this statistic results from that in the estimation of the variance. For the case of the symmetric distribution (N ; left column) K-statistics are better than classical and L ones, and Option 1 outperforms Option 2 in terms of skewness estimation. These results are quite relevant for distribution fitting but certainly additional and more systematic analyses are needed in order to shape an optimal fitting procedure.

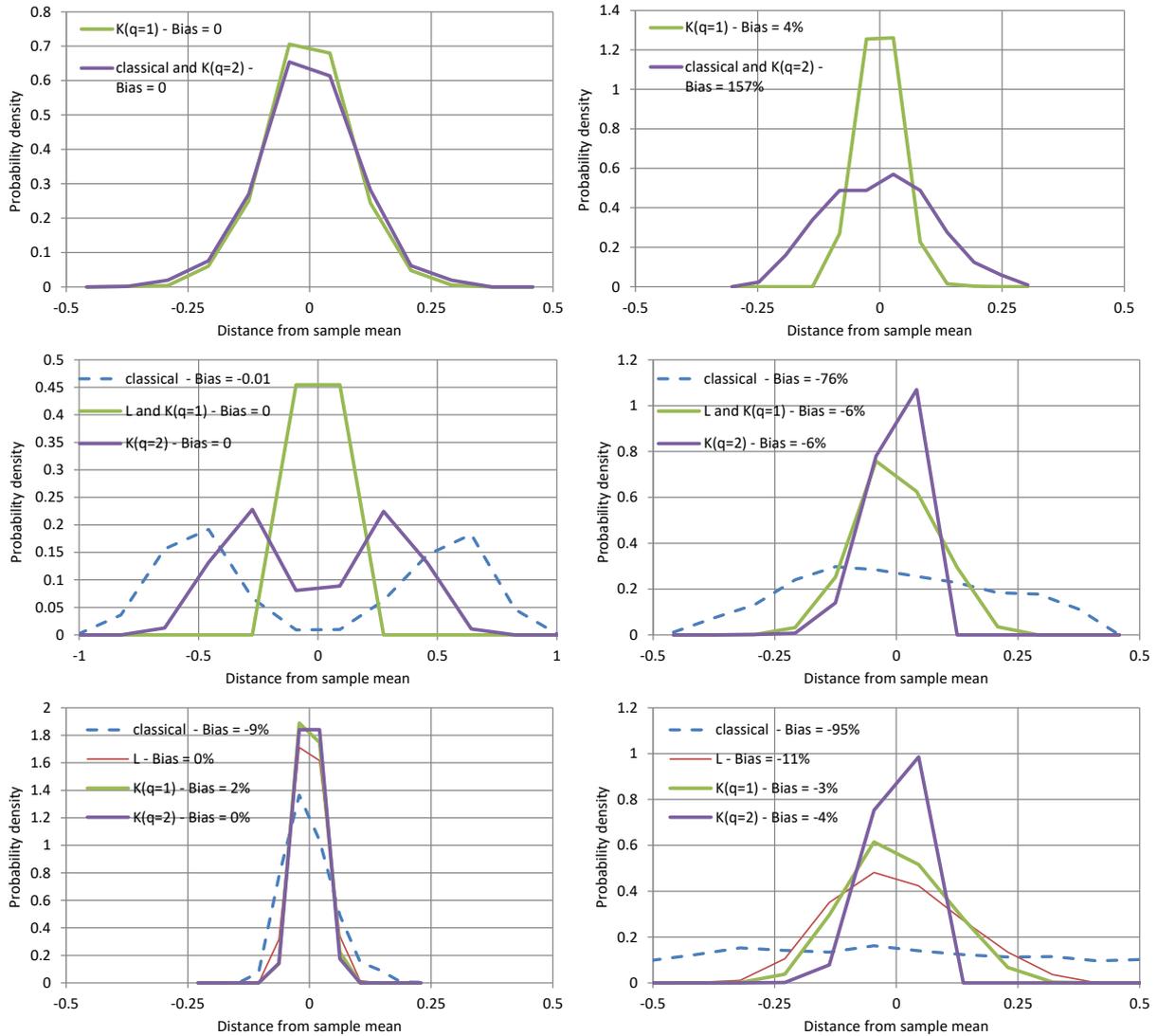


Figure 3 Illustration of the probability density function of: (upper) variability index (K_{11}/K_{21} , $\mu/\sigma \equiv K_{11}/\sqrt{K_{22}}$; note that the latter is inverse of the common coefficient of variation); (middle), skewness index ($\mu_3^{1/3}/\sigma$, K_{31}/K_{21} , $\text{sign}(K_{32})\sqrt{|K_{32}|/K_{22}}$); (lower) kurtosis index ($\mu_4^{1/4}/\sigma$, λ_4/λ_2 , K_{41}/K_{21} , $\sqrt{K_{42}/K_{22}}$). The panels of the left column correspond to the normal distribution $N(0,1)$ and those of the right column to the lognormal distribution $LN(0, 2)$. The L-statistics, also plotted, differ from the K-statistics only for $p = 4$ (see Table 2). The densities are depicted after shifting to zero mean, while the bias (difference of the simulated mean from the theoretical value of Table 3, as a percentage of the latter) is also given in the legend of each panel.

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

Second order properties of stationary stochastic processes are most typically expressed by the autocovariance function:

$$c(h) := \text{cov}[\underline{x}(t), \underline{x}(t+h)] \quad (28)$$

whereas an equivalent description is its Fourier transform, i.e. the power spectrum:

$$s(w) := 4 \int_0^{\infty} c(h) \cos(2\pi wh) dh \quad (29)$$

While autocovariance and its equivalent standardized form, i.e., autocorrelation, have been the most customary tools to characterize dependence, they are neither the only nor the most effective ones. Instead, the variance of the process averaged at a specified time scale k provides a mathematically equivalent but statistically more advantageous means to this aim. To see this, let us consider the second-order dependence of any two random variables \underline{x}_1 and \underline{x}_2 with means μ_i and standard deviations σ_i , $i = 1, 2$. The variance of the average of the two variables contains the same information as the covariance thereof. We note, though that if the variables denote different physical quantities, it is necessary to make them compatible before taking the average, which can be made by standardizing with their standard deviations. In other words, we define:

$$\rho_{12} := \text{var} \left[\frac{1}{2} \left(\frac{\underline{x}_1}{\sigma_1} + \frac{\underline{x}_2}{\sigma_2} \right) \right] \quad (30)$$

which can be expanded to yield:

$$\rho_{12} = \frac{1}{4} \text{E} \left[\left(\frac{\underline{x}_1 - \mu_1}{\sigma_1} + \frac{\underline{x}_2 - \mu_2}{\sigma_2} \right)^2 \right] = \frac{1}{2} + \frac{1}{2} \text{cov} \left[\frac{\underline{x}_1}{\sigma_1}, \frac{\underline{x}_2}{\sigma_2} \right] = \frac{1}{2} + \frac{1}{2} r_{12} \quad (31)$$

Here r_{12} is the classical (Pearson) correlation coefficient, i.e.,

$$r_{12} := \frac{\text{cov}[\underline{x}_1, \underline{x}_2]}{\sigma_1 \sigma_2} = \text{cov} \left[\frac{\underline{x}_1}{\sigma_1}, \frac{\underline{x}_2}{\sigma_2} \right] \quad (32)$$

satisfying $-1 \leq r_{12} \leq 1$ with the values $-1, 0, 1$ representing fully anti-correlated, uncorrelated and fully correlated variables, respectively. Obviously, the same information as in r_{12} is provided by ρ_{12} , which satisfies $0 \leq \rho_{12} \leq 1$ with the values $0, 1/2, 1$ representing fully anti-correlated, uncorrelated and fully correlated variables, respectively.

Unlike r_{12} , the notion of ρ_{12} could be readily expanded to many variables. Assuming that the variables $\underline{x}_1, \dots, \underline{x}_k$ are identically distributed with common variance σ^2 , so that standardization is no longer needed before taking the variance, we define the so-called climacogram, $\gamma_k := \text{var}[\underline{X}_k/\kappa]$, where $\underline{X}_k := \underline{x}_1 + \dots + \underline{x}_k$, and so \underline{X}_k/κ is the average, satisfying $0 \leq \gamma_k \leq \sigma^2$.

Furthermore, the climacogram is readily adapted to a continuous-time stochastic process $\underline{x}(t)$, namely,

$$\gamma(k) := \text{var}[\underline{X}(k)/k], \quad \underline{X}(k) := \int_0^k \underline{x}(t)dt \quad (33)$$

The climacogram provides a description fully equivalent to that of autocovariance as for a continuous-time stochastic process the two tools are connected by:

$$\gamma(k) = 2 \int_0^1 (1 - \chi)c(\chi k)d\chi, \quad c(h) = \frac{1}{2} \frac{d^2(h^2\gamma(h))}{dh^2} \quad (34)$$

As shown by Dimitriadis and Koutsoyiannis (2015) and Koutsoyiannis (2016) the climacogram offers several advantages over autocovariance and power spectrum. A surrogate of the power spectrum, again with several advantages over it, is the climacospectrum (Koutsoyiannis 2017) defined as

$$\zeta(k) := \frac{k(\gamma(k) - \gamma(2k))}{\ln 2} \quad (35)$$

The climacogram can be further expanded to describe the dependence of different processes, replacing the concept of cross-correlogram of two stationary processes $\underline{x}(t)$ and $\underline{y}(t)$ by the *standardized cross-climacogram* (SCC) for scale k and lag h :

$$\rho_{xy}(k, h) := \text{var} \left[\frac{\underline{X}(k)}{2\sqrt{\Gamma_x(k)}} + \frac{\underline{Y}(k+h) - \underline{Y}(h)}{2\sqrt{\Gamma_y(k)}} \right] = \text{var} \left[\frac{\underline{X}(k)/k}{2\sqrt{\gamma_x(k)}} + \frac{(\underline{Y}(k+h) - \underline{Y}(h))/k}{2\sqrt{\gamma_y(k)}} \right] \quad (36)$$

where $\underline{Y}(k)$ is defined in a similar manner with $\underline{X}(k)$. Likewise, we could replace the cross-covariance by the *cross-climacogram* (CC) and the *cumulative cross-climacogram* (CCC):

$$\gamma_{xy}(k, h) := \rho_{xy}(k, h)\sqrt{\gamma_x(k)\gamma_y(k)}, \quad \Gamma_{xy}(k, h) := \rho_{xy}(k, h)\sqrt{\Gamma_x(k)\Gamma_y(k)} \quad (37)$$

Now, coming to high-order properties of a stationary stochastic process, we observe that full description thereof requires functions of many variables. For example, the third-order properties are expressed in terms of a function of two time lags h_1 and h_2 :

$$c_3(h_1, h_2) := E[(\underline{x}(t) - \mu)(\underline{x}(t+h_1) - \mu)(\underline{x}(t+h_2) - \mu)] \quad (38)$$

Such a description is not parsimonious and its accuracy holds only in theory because, as we have seen, sample estimates of classical high-order moments are not reliable. Therefore we introduce single-variable descriptions for any order p , expanding the idea of the climacogram and climacospectrum based on K-moments. Specifically we define the K-climacogram as:

$$\gamma_{pq}(k) := (p - q + 1)E\left[(2F(\underline{X}(k)/k) - 1)^{p-q}(\underline{X}(k)/k - \mu)^q\right] \quad (39)$$

and the K-climacospectrum as:

$$\zeta_{pq}(k) := \frac{k(\gamma_{pq}(k) - \gamma_{pq}(2k))}{\ln 2} \quad (40)$$

where $\gamma_{22}(k) \equiv \gamma(k)$ and $\zeta_{22}(k) \equiv \zeta(k)$. Even though for $q > 2$ the K-moment description is no longer equivalent to the multivariate high-order one (the former is a function of a scalar variable while the latter is a function of a vector variable), it suffices to fully define the marginal distribution at any scale k .

To illustrate these two tools we use some examples with real world data. In the first example shown in Figure 4 the data originate from turbulence measurements. Specifically 60 000 values of turbulent velocity along the flow direction from a grid-turbulence experiment are used. The original series (described in Kang 2003; see also Koutsoyiannis 2017, and Dimitriadis and Koutsoyiannis 2018) was upscaled (averaged) so that time scale 1 corresponds to 0.5 s. It is impressive to see that all K-climacogram plots for all moment orders are similar and parallel to each other while there is full correspondence of slopes at $q = 1$ and $q = 2$ (in the latter the asymptotic slope is twice that in the former, as theoretically expected because $q = 2$ entails squaring). A deviation from the rule of parallelism appears for very large moment order ($p = 20$), and very large time scale (>1000) but only for $q = 1$. Whether this signifies some real behaviour (e.g. in terms of large-scale extremes) or is a pure statistical effect needs to be explored further. Another interesting behaviour shown is the small skewness of the distribution (see details about skewness in Dimitriadis and Koutsoyiannis 2018), reflected by the slightly positive values of the odd-order moments for small scales. As theoretically expected (due to the central limit theorem) the skewness disappears at large scales (>100).

The K-climacograms highlight the process behaviour on large time scales; in particular the persistence, quantified by the Hurst parameter H (which equals the K-climacogram slope for $q = 1$, plus one; see Koutsoyiannis, 2017, for further explanation of the Hurst parameter). On the other hand, they tend to mask it for small ones, because for finite variance the plot, for theoretical reasons (Koutsoyiannis 2017), should be (and actually is) a horizontal line as time scale $k \rightarrow 0$. Visibility at small time scales is regained by the K-climacospectra. These are shown in the right column of Figure 4 and verify an impressive agreement with Kolmogorov's "5/3" law at small scales, notably the same for all K-moment orders.

In the second example, shown in Figure 5, we use rainfall rate data at Iowa measured every 10 s. A sample of 29 542 values of rainfall at temporal resolution of 10 s was formed by merging measurements of seven events at Iowa by Georgakakos *et al.* (1994), which was also investigated in several other studies (e.g., Lombardo *et al.* 2012).

The K-climacograms show prominent persistence with Hurst parameter $H \approx 1 - 0.1 = 0.9$, while the very large order K-climacogram for $q = 1$ indicates an effect similar to that of the turbulence data discussed above. The skewness is also prominent almost for the entire range of time scales. The K-climacospectra suggest a slope of about 4/3 for small scales, so that according to the classification by Koutsoyiannis (2017) the process is rough-persistent, like turbulence.

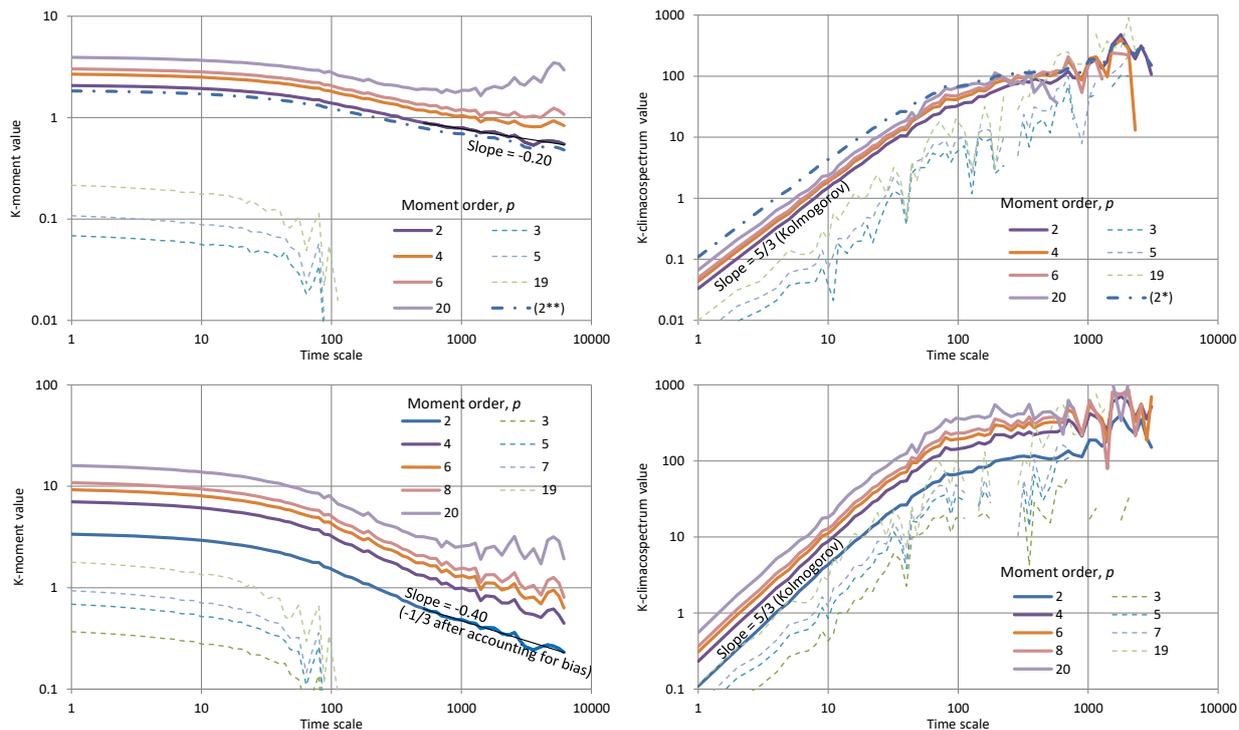


Figure 4 K-climacograms (left) and K-climacospectra (right) of turbulent velocity measured every 0.5 s, where the K-moments correspond to $q = 1$ (upper row) and $q = 2$ (lower row). Plot (2^*) is constructed from the variance and (2^{**}) corresponds to standard deviation.

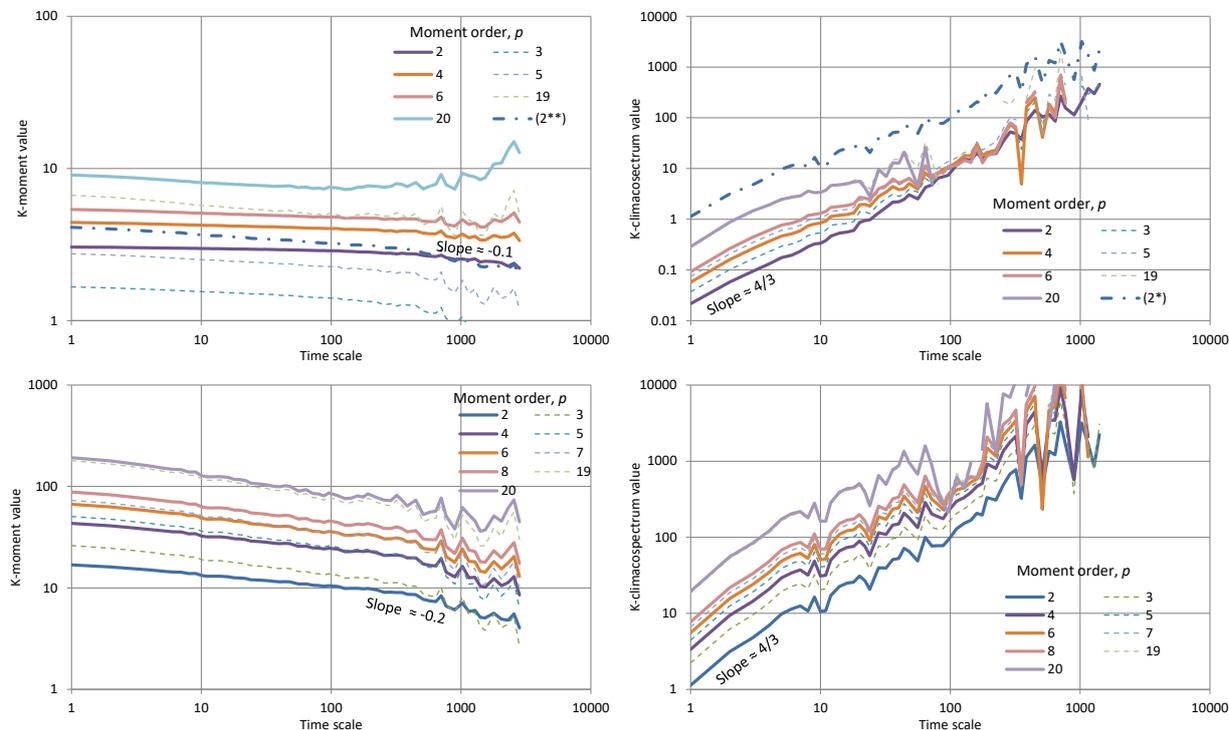


Figure 5 K-climacograms (left) and K-climacospectra (right) of rainfall rate at Iowa measured every 10 s, where the K-moments correspond to $q = 1$ (upper row) and $q = 2$ (lower row). Plot (2^*) is constructed from the variance and (2^{**}) corresponds to standard deviation.

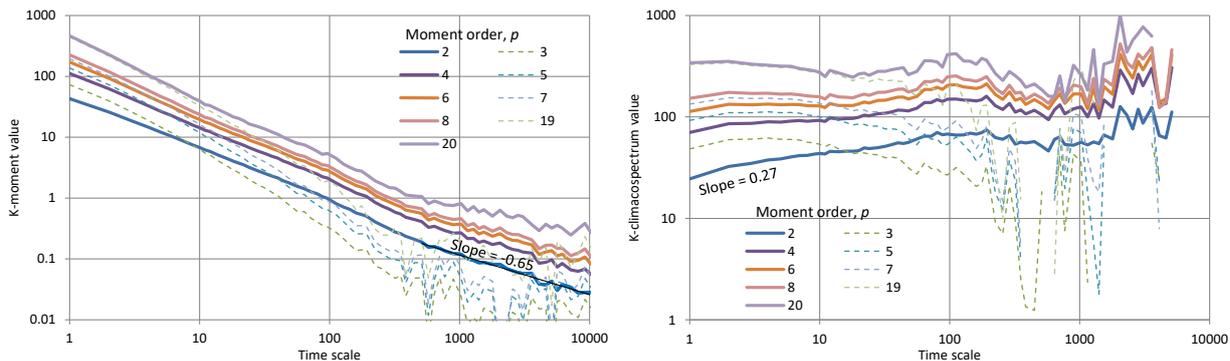


Figure 6 K-climacograms (left) and K-climacospectra (right) of daily rainfall at Padova, where the K-moments correspond to $q = 2$.

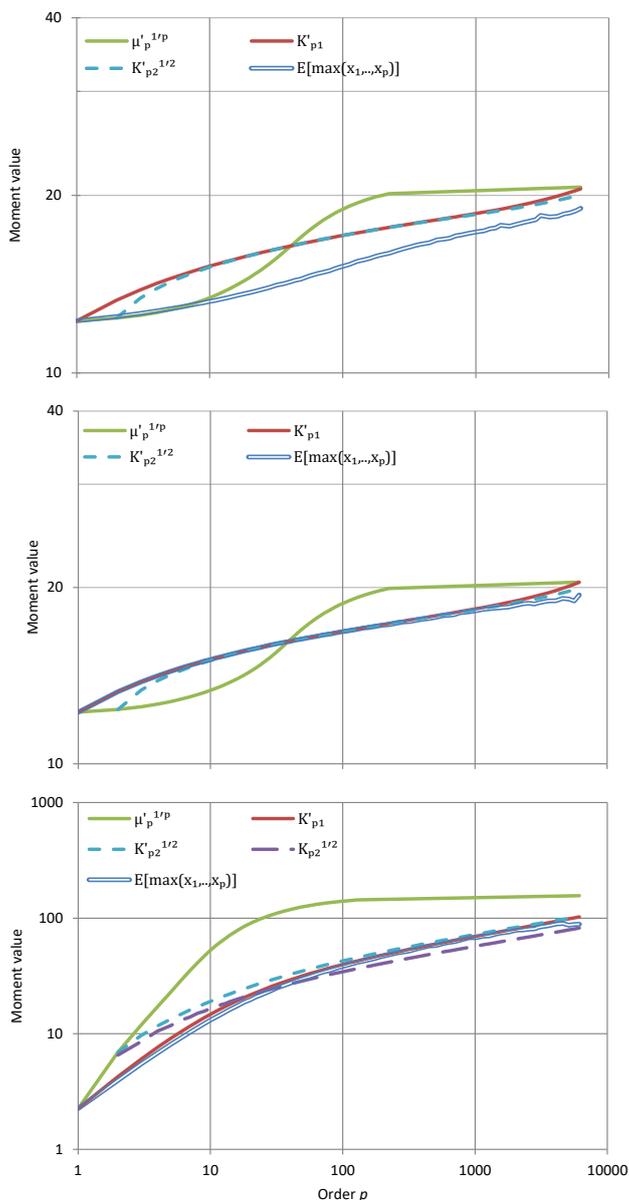


Figure 7 K-moments vs. moment order for: (upper) turbulence velocity data; (middle) synthetic data with same mean and variance as the turbulence data; (lower) daily rainfall data at Padova. Note that each curve is in fact a series of connected points whose shape is smooth by itself (not artificially smoothed).

The last example, shown in Figure 6, is the exploration of daily rainfall data at Padova; this is the longest rainfall record existing worldwide (Marani and Zanetti 2015), starting at 1725 and containing 100 442 values. For large time scales the K-climacograms for $q = 2$ indicate a Hurst behaviour with a slope of $-0.65 > -1$, corresponding to a Hurst parameter $H = 1 - 0.65/2 = 0.68$, consistent with findings in Iliopoulou *et al.* (2018). At the daily and multi-daily scale the process appears to be closer to white noise with climacogram slope -0.83 (against -1 of white noise) and climacospectrum close to constant. However, these are artefacts because, as seen in the previous example, at scales finer than daily the behaviour is far different from white noise. The general shape of the climacogram in Figure 6, with decreasing slope as we go to larger scales, is consistent with a behaviour observed by Markonis and Koutsoyiannis (2016).

Finally, the fact that the noncentral K-moments reflect well the behaviour of maxima in a process is illustrated in Figure 7, where moments are plotted against their order p for the turbulence and Padova rainfall series. In addition to K-moments, the sample estimate of the maximum for a time window of length p , i.e., $E[\max(\underline{x}_1, \dots, \underline{x}_p)]$, is also plotted. The latter plot generally agrees with those of K'_{p1} and $\sqrt{K'_{p2}}$. Some deviations are due to the temporal dependence in the process because $E[\max(\underline{x}_1, \dots, \underline{x}_p)]$ reflects the joint distribution, while K'_{p1} reflects the marginal one. This is further illustrated in the middle panel of the figure, which is constructed from uncorrelated synthetic data with same mean and variance as the turbulence data. It is clearly seen that, because of independence, the plot corresponding to $E[\max(\underline{x}_1, \dots, \underline{x}_p)]$ has been shifted up and now coincides with K'_{p1} . Even the plot of the classical $\mu_p^{1/p}$ is not far from the other ones, because it again asymptotically reflects maxima as explained above. In the last panel of the Padova rainfall series the central moment $\sqrt{K_{p2}}$ has also been plotted and it is seen that, because of the high skewness of the distribution, it does not differ substantially from $\sqrt{K'_{p2}}$. Furthermore, because the temporal dependence in these rainfall data is weak, the plot corresponding to $E[\max(\underline{x}_1, \dots, \underline{x}_p)]$ is very close to that corresponding to K'_{p1} .

Stochastic simulation

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 so-called symmetric moving average (SMA) method (Koutsoyiannis 2000, 2016) can 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 an arbitrary autocovariance function and any marginal distribution function. In particular, the approximation up to fourth order moments has been studied in Dimitriadis and Koutsoyiannis (2018) and Koutsoyiannis *et al.* (2018), while in the former work it was shown that the method can perform for even higher

orders. Here we provide a more general formulation based on cumulants, which can handle explicitly moments of arbitrarily high order.

The SMA scheme can directly generate time series from any process \underline{x}_i with any type of dependence by:

$$\underline{x}_i = \sum_{l=-r}^r a_{|l|} \underline{v}_{i+l} \quad (41)$$

where a_l are coefficients calculated from the autocovariance function and \underline{v}_i is (generally non-Gaussian) white noise averaged in discrete-time. In theory, the limit r should be ∞ but in practice a truncation to a specific finite r is made (see Koutsoyiannis 2016 for methods to handle the truncation error).

It should be stressed that the weights a_l are not model parameters estimated from data but internal coefficients determined by theoretical calculations. Assuming that the power spectrum $s_d(\omega)$ of \underline{x}_i in discrete time is known (from the climacogram or equivalently, from the autocovariance function, i.e., from (29) and (34) whose discrete-time versions can be found in Koutsoyiannis 2016, 2017), the Fourier transform $s_d^a(\omega)$ of the a_l series of coefficients has been shown (Koutsoyiannis 2000) to be:

$$s_d^a(\omega) = \sqrt{2s_d(\omega)} \quad (42)$$

Thus, to calculate a_l we first determine $s_d^a(\omega)$ from the power spectrum of the process and then we invert the Fourier transform to estimate all a_l .

With respect to the preservation of moments with order > 2 , we utilize the properties of cumulants of independent variables, and particularly homogeneity and additivity. For the p th cumulant, κ_p of \underline{x}_i , by virtue of (41), these properties result in

$$\kappa_p = \sum_{l=-r}^r a_{|l|}^p \kappa_p^v \quad (43)$$

where κ_p^v is p th cumulant of \underline{v}_i . Observing that the zeroth cumulant is zero, we can estimate κ_p^v from κ_p by

$$\kappa_p^v = \frac{\kappa_p}{2 \sum_{l=1}^r a_{|l|}^p} \quad (44)$$

Based on the above discourse, we can formulate the following steps of a general simulation strategy, starting from the observed data (noting that alternative modelling strategies can be seen in a series of references provided by Dimitriadis and Koutsoyiannis 2018):

1. We estimate K-moments for $q = 1$ and 2 , and we choose a marginal distribution for the process based on K-moments and possibly relevant theoretical considerations (e.g. entropy maximizing distribution).
2. We construct the climacogram and climacospectrum, and we choose a suitable model of second-order dependence (see a repertoire of models in Koutsoyiannis 2016, 2017).

3. We estimate the marginal and joint distribution parameters of the model (with appropriate provision for fitting issues, such as bias, e.g., as in Koutsoyiannis 2016).
4. Based on the model parameters we calculate theoretically (and not estimate from data) the classical moments of the process of interest.
5. From equation (27) we calculate the cumulants of the process of interest.
6. From equation (44) we calculate the cumulants of the white noise process and from (27) we calculate its moments.
7. We choose an appropriate distribution for the white noise, calculate its parameters theoretically from its moments and generate a random sample with the required length.
8. Filtering with equation (41) we synthesize the simulated series for the process of interest.
9. We construct K-climacograms from the original and synthetic data and compare for relevant moment orders > 2 .
10. If a disagreement is found in step 9, then we repeat the process separating the entire range of relevant scales to parts, building different models for each part, and coupling the separate models using a model coupling (disaggregation) scheme such as that in Koutsoyiannis (2001) (see also Lombardo *et al.* 2012, 2017).

Even though a full presentation of a case study with all these steps is out of the scope of this theoretical paper, a proof of concept for the feasibility and effectiveness of simulation for high-order moments using several data sets has been already made in Dimitriadis and Koutsoyiannis (2018), albeit without using the K-climacograms and indirectly using the cumulants (i.e. by finding direct relationships between moments of the process of interest and white noise). The agreement of simulated and original moments at all scales was impressive, so that step 10 was not necessary.

Conclusions and discussion

The concept of knowable (K-)moments introduced here resolves the well-known (e.g. Lombardo *et al.* 2014) problem of inability to estimate high-order moments from typical (or even large) samples, thus complementing the idea of L-moments in a manner that can also model joint (not only marginal) distributions of stochastic processes and describe the dependence.

The concept of K-climacogram, also introduced here, provides an effective and knowable means to parsimoniously characterize the dependence structure of a process for moments higher than 2 and detect possible scaling behaviours for large scales (i.e. persistence) and the possible change of the scaling laws with moment order. It is noted though that the data sets studied here support a rather single scaling law for all orders.

The complementary concept of K-climacospectrum, is another effective tool again to characterize the dependence structure of a process for high-order moments, particularly at the opposite (lower) end of time scales. It can also detect possible scaling behaviours (i.e.

smoothness or fractality), with exponents varying with moment order (multifractality). It is noted though that the data sets studied here support a rather unique scaling law for all orders. It is thus interesting to further study potential multifractality of geophysical processes and in particular isolate from the real behaviour of the processes the effect of using inappropriate estimators, as implied in Lombardo *et al.* (2014) and Koutsoyiannis *et al.* (2018).

The well-known concept of cumulants, if combined with the other tools described above, can streamline the generation of skewed white noise and in turn simulate any non-Gaussian distribution by preserving its high-order moments.

One can imagine a great deal of further work to explore, organize and exploit the tools introduced here. This includes: (a) detecting and characterizing natural behaviours (such as scaling, persistence, roughness/fractality, intermittence); (b) exploiting the use of the framework in studying extremes on multiple scales (such as in constructing ombrian, also called IDF, relationships); (c) compiling a set of analytical or numerical results/tables for customary distribution types; (d) statistically characterizing the bias and variability of the tools; (e) standardizing an optimal model fitting procedure; (f) studying the entire scheme in disaggregation mode; and (g) extending the framework to multivariate processes, represented by vectors of random variables (cf. Koutsoyiannis, 2000) and multidimensional stochastic fields (cf. Koutsoyiannis *et al.* 2011).

Appendix: Relationships of central and noncentral moments

It is reminded that classical central and noncentral moments are related to each other by

$$\mu'_p = \sum_{i=0}^p \binom{p}{i} \mu^{p-i} \mu_i, \quad \mu_p = \sum_{i=0}^p \binom{p}{i} (-\mu)^{p-i} \mu'_i \quad (45)$$

Similar relationships can be obtained between central and noncentral K-moments. In particular, for $q = 1$ and 2 the following relationships hold (their proof is not too difficult):

$$\begin{aligned} K'_{p1} &= \left(\frac{1}{2}\right)^{p-1} \sum_{i=0}^{p-1} \binom{p}{i} K_{p-i,1} + \mu \\ K_{p1} &= \sum_{i=0}^{p-1} \binom{p}{i} (-1)^i 2^{p-1-i} (K'_{p-i,1} - \mu) \\ K'_{p2} &= \left(\frac{1}{2}\right)^{p-2} \sum_{i=0}^{p-2} \binom{p-1}{i} (K_{p-i,2} + 2\mu K_{p-i-1,1}) + \mu^2 \\ K_{p2} &= \sum_{i=0}^{p-2} \binom{p-1}{i} (-1)^i (2)^{p-2-i} (K'_{p-i,2} - 2\mu K'_{p-i,1} + \mu^2) \end{aligned} \quad (46)$$

Acknowledgements. This research was made on the occasion of two invitations: partly for a seminar in Università Roma Tre and partly for a plenary talk at Hydrofractals '18. I sincerely thank Elena Volpi and Alin Carsteanu for each of these two invitations. I also thank Aldo Fiori and Salvatore Grimaldi for their feedback during the seminar in Rome, as well as Panayiotis Dimitriadis and Theano (Any) Iliopoulou for their comments on a first draft of the manuscript. I gratefully acknowledge the constructive and impressively penetrating reviews by Eric Gaume, Federico Lombardo and an anonymous reviewer, which helped me improve the paper substantially.

Funding information. No funding was provided for this research.

Conflicts of interest. No conflict of interest exists.

References

- Dimitriadis, P., and Koutsoyiannis, D., 2015. Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst–Kolmogorov processes. *Stochastic Environmental Research & Risk Assessment*, 29 (6), 1649–1669, doi: 10.1007/s00477-015-1023-7.
- Dimitriadis, P., and Koutsoyiannis, D., 2018. Stochastic synthesis approximating any process dependence and distribution. *Stochastic Environmental Research & Risk Assessment*, doi: 10.1007/s00477-018-1540-2.
- Feller, W., 1968. *An Introduction to Probability Theory and Its Applications*, vol. I, 3rd edition, London-New York-Sydney-Toronto, John Wiley & Sons.
- Georgakakos, K. P., Cârsteanu, A. A., Sturdevant, P. L. and Cramer, J. A., 1994. Observation and analysis of Midwestern rain rates. *Journal of Applied Meteorology*, 33, 1433-1444.
- Ghosh, S., and Beran, J., 2006. On estimating the cumulant generating function of linear processes. *Annals of the Institute of Statistical Mathematics*, 58 (1), 53–71.
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R., 1979. Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. *Water Resour. Res.*, 15 (5), 1049–1054, doi: 10.1029/WR015i005p01049.
- Hemelrijk, J., 1966. Underlining random variables. *Statistica Neerlandica*, 20 (1), 1-7.
- Hosking, J. R. M., 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B*, 52, 105–124.
- Hosking, J. R. M., and Wallis, J. R., 2005. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press.
- Iliopoulou, T., Papalexiou, S. M., Markonis, Y., and Koutsoyiannis, D., 2018. Revisiting long-range dependence in annual precipitation. *Journal of Hydrology*, 556, 891–900, doi: 10.1016/j.jhydrol.2016.04.015.
- 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.
- Koutsoyiannis, D., 2000. A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. *Water Resources Research*, 36 (6), 1519–1533, doi: 10.1029/2000WR900044.
- Koutsoyiannis, D., 2001. Coupling stochastic models of different time scales. *Water Resources Research*, 37 (2), 379–391, doi: 10.1029/2000WR900200.

- Koutsoyiannis, D., 2010. A random walk on water. *Hydrology and Earth System Sciences*, 14, 585–601, doi: 10.5194/hess-14-585-2010.
- Koutsoyiannis, D., 2013. Hydrology and change. *Hydrological Sciences Journal*, 58 (6), 1177–1197, doi: 10.1080/02626667.2013.804626.
- Koutsoyiannis, D., 2016. Generic and parsimonious stochastic modelling for hydrology and beyond. *Hydrological Sciences Journal*, 61 (2), 225–244, doi: 10.1080/02626667.2015.1016950.
- Koutsoyiannis, D., 2017. Entropy production in stochastics. *Entropy*, 19 (11), 581, doi: 10.3390/e19110581.
- Koutsoyiannis, D., Dimitriadis, P., Lombardo, F., and Stevens, S., From fractals to stochastics: Seeking theoretical consistency in analysis of geophysical data, *Advances in Nonlinear Geosciences*, edited by Tsonis, A. A., 237–278, doi: 10.1007/978-3-319-58895-7_14, Springer, 2018.
- Koutsoyiannis, D., and Montanari, A., 2015. Negligent killing of scientific concepts: the stationarity case. *Hydrological Sciences Journal*, 60 (7-8), 1174–1183, doi: 10.1080/02626667.2014.959959.
- Koutsoyiannis, D., Paschalis, A., and Theodoratos, N., 2011. Two-dimensional Hurst-Kolmogorov process and its application to rainfall fields. *Journal of Hydrology*, 398 (1-2), 91–100.
- Lange, M., Zühlke, D., Holz, O., Villmann, T. and Mittweida, S.G., 2014. Applications of lp-norms and their smooth approximations for gradient based learning vector quantization. In *Proc. of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2014)*. 271-276, Louvain-La-Neuve, Belgium.
- Lombardo, F., Volpi, E., and Koutsoyiannis, D. (2012), Rainfall downscaling in time: Theoretical and empirical comparison between multifractal and Hurst-Kolmogorov discrete random cascades, *Hydrological Sciences Journal*, 57 (6), 1052–1066.
- Lombardo, F., Volpi, E., Koutsoyiannis, D., and Papalexiou, S. M., 2014. Just two moments! A cautionary note against use of high-order moments in multifractal models in hydrology. *Hydrology and Earth System Sciences*, 18, 243–255, doi:10.5194/hess-18-243-2014.
- Lombardo, F., Volpi, E., Koutsoyiannis, D., and Serinaldi, F., 2017. A theoretically consistent stochastic cascade for temporal disaggregation of intermittent rainfall. *Water Resources Research*, 53 (6), 4586–4605, doi: 10.1002/2017WR020529.
- Marani, M., and Zanetti, S., 2015. Long-term oscillations in rainfall extremes in a 268 year daily time series. *Water Resources Research*, 51, 639–647.
- Markonis, Y., and Koutsoyiannis, D., 2016. Scale-dependence of persistence in precipitation records. *Nature Climate Change*, 6, 399–401, doi: 10.1038/nclimate2894.
- Montanari, A., and Koutsoyiannis, D., 2014. Modeling and mitigating natural hazards: Stationarity is immortal!. *Water Resources Research*, 50 (12), 9748–9756, doi: 10.1002/2014WR016092.
- Montanari, A. et al. ,2013, “Panta Rhei – Everything Flows”, Change in Hydrology and Society – The IAHS Scientific Decade 2013-2022. *Hydrological Sciences Journal*, 58 (6), 1256–1275, doi: 10.1080/02626667.2013.809088.
- Papoulis, A., 1990. *Probability and Statistics*. Prentice-Hall, New Jersey.
- Smith, P. J., 1995. A recursive formulation of the old problem of obtaining moments from cumulants and vice versa. *The American Statistician*, 49 (2), 217–218.