Climate change impacts on hydrological science: A comment on the relationship of the climacogram with Allan variance and variogram

Demetris Koutsoyiannis
Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Heroon Polytechneiou 5, GR 157 80 Zographou, Greece
(dk@itia.ntua.gr - http://www.itia.ntua.gr/dk)

Version 1 – 11/11/2018

Abstract Inspired by reactions on a talk about climate change impacts on hydrological science, I am presenting detailed comparisons of second-order stochastic tools with particular emphasis on the relationship of the climacogram with the Allan variance and with the variogram.

1 Introduction

In 26 October 2018 I gave a presentation in Moscow, entitled "*Climate change impacts on hydrological science: How the climate change agenda has lowered the scientific level of hydrology*" (Koutsoyiannis, 2018b). This was part of the *School for Young Scientists "Modelling and forecasting of river flows and managing hydrological risks: Towards a new generation of methods"* organized by the Russian Academy of Sciences and the Lomonosov Moscow State University. I received interesting feedback, in particular by Professor A.[1] Here I am replying in detail to a scientific comment related to the stochastic background of my presentation,[2] namely Professor's A statement about the climacogram, which I briefly discussed in my talk as the main stochastic tool necessary to follow my presentation.

In brief, Professor A stated that what I call *climacogram* is well known, is named Allan variance and is well studied since the 1960s. I perceived this statement as implying that my term *climacogram* is superfluous and that the best I did was to copy results known for decades. After the talk he also sent me a couple of papers and other material to see it. I replied orally that the climacogram is just a variance (applied to a stochastic process) and, apparently, I did not claim to have invented the concept of the variance.[3] I could add that the concept of the variance of the time averaged process is not of course my invention as it is contained in stochastics books (e.g. Papoulis 1991; Beran, 1994). What I have done is that I demonstrated, after thorough studies, that this concept has very

[1] I am publishing this comment by uploading it on ResearchGate without naming the professor (who was another invited speaker attending my talk). As ResearchGate allows comments from anyone interested, he can respond to this comment. If he responds that he wishes to be acknowledged by name, I will upload a corrected version with his name. In any case, my plan is to upload a next version once errors are spotted and corrected.

[2] This was the most polite of Professor's A reactions. I am reluctant to further discuss his other, non-scientific reactions, but if he repeats his comments in ResearchGate or in other forums, I will probably provide written replies.

[3] The variance, as a concept and a term, is in common use at least since 1918, after the paper by Fisher (1918) while its square root, the standard deviation, is an even older term, due to Pearson (1894), who attributes the concept to Gauss and Airy.

useful properties and advantages over other tools, and understood that it deserves a name—and I proposed the name *climacogram*. I also said that I do not like a term like Allan variance but if someone suggested a better term to replace *climacogram*, I would adopt it.

However, a more detailed study presented below shows that the Allan variance is **not** the climacogram. Yet it is a second order characteristic of a stochastic process and as any other second-order characteristic, it is related to all others. Namely, other customary second-order characteristics are the autocovariance, the power spectrum, the variogram and some others. The relationships of the Allan variance with some other second-order characteristics are also given below along with the relationships of the climacogram with the variogram.[4]

2 Basic concepts and definitions

This comment is based on the theory and terminology of stochastics, whose possible ignorance may cause difficulties in reading it. When meeting terms not carefully founded on stochastics (as typically is the case in publications referring to the Allan variance, including internet sources, such as Wikipedia), I will try to translate them into the stochastics language. If my translation is found inaccurate or false, I will be happy to change it, and I am also willing to provide any clarification requested.

I am sorry if the text that follows is found too didactic. However, I encourage anyone who has time, not to skip this section. In stochastics, overconfidence about our knowledge does not help, and reading elementary stuff a second or third time may be useful. Besides, those who know stochastics very well, may find errors in my exposition and correct me.

2.1 Stochastic process, stationarity and ergodicity

We recall that a stochastic process $\underline{x}(t)$ is a collection of (usually infinitely many) random variables $\underline{x}$ indexed by t, typically representing time. Time is a continuous variable (real number) but it is customary to discretize it using a time step D. In turn, a random variable, $\underline{x}$, is an abstract mathematical entity, associated with a probability distribution function,

$$F(x) ≔ P\{\underline{x} \le x\} \tag{1}$$

where x is any numerical value (a regular variable), P denotes probability and the symbol "≔"means "is defined as". A random variable $\underline{x}$ becomes identical to a regular variable x_0 only if $F(x) = \mathrm{H}(x - x_0)$, where H is the unit step function, or if its variance is zero (see reminder of the definition of the variance in subsection 2.2).

The stochastic process $\underline{x}(t)$ represents the evolution of a system over time, along with its uncertainty expressed in the language of stochastics. A trajectory or sample function $x(t)$ is a (single) *realization* of $\underline{x}(t)$. If this realization is known at certain points t_i, it is a time series.

[4] During the talk there was another comment by Professor B, about the variogram, which prompted me to try to clarify also the relationship of the climacogram with the variogram.

It is important to stress the fundamental difference of a random variable from a regular variable and of a stochastic process from a time series. To avoid the common practice of confusing the two items in each of the two pairs of concepts, I use a careful notation adopting the so-called Dutch convention (Hemelrijk 1966), i.e., underlining random variables and stochastic processes. Regular variables such as the time t or realizations of $\underline{x}$ are denoted by non-underlined symbols.

Central to the notion of a stochastic process are the concepts of *stationarity* and *nonstationarity*, two widely misunderstood and misused concepts, whose definitions are only possible for (and applies only to) stochastic processes (thus, for example, a time series cannot be stationary, nor nonstationary). A process is called (strict-sense) stationary if its statistical properties are invariant to a shift of time origin, i.e. the processes $\underline{x}(t)$ and $\underline{x}(s)$ have the same statistics for any t and s (see further details, as well as definition of wide-sense stationarity, in Papoulis, 1991; see also further explanations in Koutsoyiannis, 2006, 2011 and Koutsoyiannis and Montanari, 2015). Conversely, a process is nonstationary if some of its statistics are changing in time and their change is described as a deterministic function of time.

Stationarity is also related to *ergodicity*, which in turn is a prerequisite to make inference from data, that is, induction. Without ergodicity inference from data would not be possible. While ergodicity is originally defined in dynamical systems (e.g. Mackey, 1992, p. 48), the ergodic theorem (e.g. Mackey, 1992, p. 54) allows redefining ergodicity within the stochastic processes domain (Papoulis 1991 p. 427; Koutsoyiannis 2010) as will be detailed in the subsection 2.2. From a practical point of view, ergodicity can always be assumed when there is stationarity, while this assumption if fully justified by the theory if the system dynamics is deterministic. Conversely, if nonstationarity is assumed, then ergodicity cannot hold, which forbids inference from data. This contradicts the basic premise in geosciences, where data are the only reliable information in building models and making inference and prediction.

2.2 Expectation and its estimation

Functions of random variables, e.g. $\underline{z} = g(\underline{x})$ are random variables. Expected values of random variables are regular variables; for example $\mathrm{E}[\underline{x}]$ and $\mathrm{E}[g(\underline{x})]$ are constants—neither functions of x nor of $\underline{x}$. That justifies the notation $\mathrm{E}[\underline{x}]$ instead of $\mathrm{E}(\underline{x})$ or $\mathrm{E}(x)$ which would imply functions of $\underline{x}$ or x. Specifically, the expectation of a function $g(\,)$ of a continuous random variable $\underline{x}$ is defined as

$$\mathrm{E}\big[g(\underline{x})\big] \coloneqq \int_{-\infty}^{\infty} g(x) f(x) \mathrm{d}x \tag{2}$$

where $f(x)$ is the probability density function, i.e.,

$$f(x) \coloneqq \frac{dF(x)}{dx} \tag{3}$$

For $g(\underline{x}) = x$ the expectation is the mean

$$\mu_x \equiv \mathrm{E}\big[\underline{x}\big] \coloneqq \int_{-\infty}^{\infty} x f(x) \mathrm{d}x \tag{4}$$

while for $g(\underline{x}) = (\underline{x} - \mu_x)^2$ the expectation is the variance

$$\sigma_x^2 \equiv \mathrm{var}[\underline{x}] \coloneqq \mathrm{E}\left[(\underline{x} - \mu_x)^2\right] = \int_{-\infty}^{\infty} (\underline{x} - \mu_x)^2 f(x)\mathrm{d}x \tag{5}$$

Likewise, for two (or more) random variables we can define their joint distribution function, e.g., $F(x, y) \coloneqq P\{\underline{x} \leq x, \underline{y} \leq y\}$, joint density $f(x, y) \coloneqq \partial^2 F(x, y)/\partial x \partial y$, and joint expectations. The simplest case of joint expectation is the covariance:

$$\sigma_{xy} \equiv \mathrm{cov}\left[\underline{x}, \underline{y}\right] \coloneqq \mathrm{E}\left[(\underline{x} - \mu_x)\left(\underline{y} - \mu_y\right)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\underline{x} - \mu_y)\left(\underline{y} - \mu_y\right) f(x, y)\mathrm{d}x \tag{6}$$

In a similar manner, we define expectations for a stochastic process. Assuming that the process $\underline{x}(t)$ is stationary with mean $\mu = \mathrm{E}[\underline{x}(t)]$, the variance of the instantaneous process is

$$\gamma_0 \coloneqq \mathrm{var}[\underline{x}(t)] = \mathrm{E}\left[(\underline{x}(t) - \mu)^2\right] \tag{7}$$

and the autocovariance for time lag h is

$$c(h) \coloneqq \mathrm{cov}[\underline{x}(t), \underline{x}(t + h)] \coloneqq \mathrm{E}[(\underline{x}(t) - \mu)(\underline{x}(t + h) - \mu)] \tag{8}$$

It should be stressed that these expectations are not time averages. Sometimes to make it clearer they are called true or ensemble means, variances, covariances etc. For an ergodic process, they are related to time averages through the following relationship which can serve as a definition of an ergodic process:

$$\underline{\hat{G}}^{(\infty)} \coloneqq \lim_{T\to\infty} \frac{1}{T}\int_0^T g\left(\underline{x}(t)\right) dt = \mathrm{E}[g(\underline{x}(t))] \eqqcolon G \tag{9}$$

We notice that the left-hand side $\underline{\hat{G}}^{(\infty)}$ is a random variable while the right-hand side G is a regular variable; their equality implies that the variance of $\underline{\hat{G}}^{(\infty)}$ is zero.

When dealing with data from a process $\underline{x}(t)$ with a joint distribution function that is unknown, neither the left- nor the right-hand side of (9) can be known a priori. Assuming that we have a time series, at a time step D, with observations $x_\tau \coloneqq x(D\tau)$, τ = 1, …, n, we can approximate the left-hand side by

$$\hat{G} \coloneqq \frac{1}{n}\sum_{\tau=1}^{n} g(x_\tau) \tag{10}$$

The regular variable $\hat{G}$ is called an estimate of the true expectation G. Replacing in equation (10) the values x_τ with the random variables $\underline{x}_\tau \coloneqq \underline{x}(D\tau)$ we define

$$\underline{\hat{G}} \coloneqq \frac{1}{n}\sum_{\tau=1}^{n} g(\underline{x}_\tau) \tag{11}$$

The random variable $\underline{\hat{G}}$ is called an estimator of the true expectation G. This is typically biased (with few exceptions, the most notable being the estimator of the mean), meaning that

$$\mathrm{E}[\underline{\hat{G}}\,] \neq G \tag{12}$$

Therefore, sometimes bias correction factors are used to deal with bias. These factors depend on the stochastic model assumed for the process and are not of general use. For example, the estimator of variance,

$$\underline{\hat{\gamma}}_0 \equiv \frac{1}{n}\sum_{\tau=1}^{n}\left(\underline{x}_\tau - \underline{\hat{\mu}}\right)^2 \tag{13}$$

is well known to be biased. Some think that if we replace n with $n - 1$ in the denominator of the right-hand side, it becomes unbiased. While this is true for uncorrelated samples, this is hardly the case for stochastic processes describing natural phenomena, where this slight change does not make the estimator unbiased and a more sophisticated procedure is required to deal with bias (see Koutsoyiannis 2003, 2016 about a correct assessment of the bias).

Summarizing, there are four different concepts, with slightly different names but very different meaning and content. Unfortunately, these are often confused in the literature and the same symbol and name are used for all, which creates confusion and may result in wrong conclusions. Table 1 clarifies the four different concepts using the variance as an example. Notice in the table that the data can be used only with one of the variance variants, namely the variance estimate, while a theoretical model is necessary to determine any of them.

Table 1 Different variants of the variance.

Name	Symbol	Type of variable	Type of determination
Variance (true)	γ_0	Regular variable	Theoretical calculation from model (by integration)
Variance estimate	$\hat{\gamma}_0$	Regular variable	Estimation from data—but model is also necessary (e.g. to calculate the estimation bias and uncertainty)
Variance estimator	$\underline{\hat{\gamma}}_0$	Random variable	Theoretical calculation from model
Variance estimator limit	$\underline{\hat{\gamma}}_0^{(\infty)}$	Random variable, which for ergodic processes has zero variance and becomes a regular variable	Theoretical calculation from model

3 The climacogram

Let $\underline{x}(t)$ be a stationary and ergodic stochastic process in continuous time t, with variance γ_0 and let

$$\underline{X}(t) \coloneqq \int_0^t \underline{x}(u)du \tag{14}$$

be the cumulative process, which is obviously nonstationary (with stationary increments). The two processes $\underline{x}(t)$ and $\underline{X}(t)$ are illustrated through some realizations thereof in Figure 1, where $\underline{X}(t)$ represents the area under the curve $\underline{x}(t)$. The figure also illustrates the discretization of the process for a time step (or time scale) D, which results in the discrete-time process

$$\underline{x}_\tau^{(D)} := \frac{1}{D} \int_{(\tau-1)D}^{\tau D} \underline{x}(u) du = \frac{\underline{X}(\tau D) - \underline{X}\big((\tau-1)D\big)}{D} \tag{15}$$

which is the time average of $\underline{x}(t)$ over the time interval $[(\tau-1)D, \tau D]$.

Let

$$\Gamma(t) := \mathrm{var}[\underline{X}(t)] \tag{16}$$

be the variance of the cumulative process, which is a function of the time t, because the process $\underline{X}(t)$ is nonstationary. Obviously, the variance of the averaged process at any time scale k (e.g., at $k = D$, as shown in Figure 1), is

$$\gamma(k) := \mathrm{var}\left[\frac{\underline{X}(t+k) - \underline{X}(t)}{k}\right] = \frac{\Gamma(k)}{k^2} \tag{17}$$

Notice that, while time t appears in the above term in square brackets, $\gamma(k)$ is not a function of t because of stationarity. The variance of the discrete-time process shown in Figure 1 is

$$\mathrm{var}\left[\underline{x}_\tau^{(D)}\right] = \gamma(D) \tag{18}$$

and does not depend on time τ because the discrete-time process $\underline{x}_\tau^{(D)}$ is also stationary.

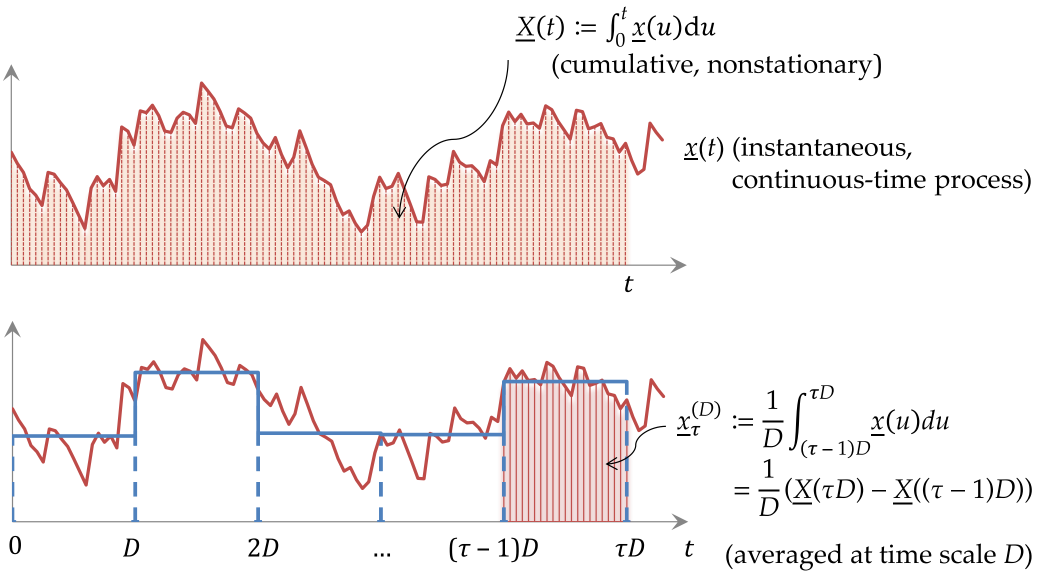

Figure 1 Explanatory sketch for a stochastic process in continuous time and in discrete time. Note that the graphs display a realization of the process (it is impossible to display the process as such) while the notation is for the process per se (from Koutsoyiannis 2017).

I call $\Gamma(t)$ and $\gamma(k)$, as functions of time t and time scale k, the *cumulative climacogram* and the *climacogram* of the process, respectively. As already said, these concepts are not new: for example Papoulis (1991) describes the latter as "*variance* [symbol] *of the time average* [symbol]" which is an accurate phrase but not a name. Also Beran (1994) uses the former concept (e.g. in his Theorem 2.2) using symbolic representation and not a name. Since the 1990s in some publications the term "aggregated variance" has been used, but it is a misnomer because the variance is not aggregated at all—just the time scale varies. As I thought each of the two concepts deserves a proper name but does not have one, in Koutsoyiannis (2010) I coined the Greek[5] term "climacogram" emphasizing the link of the concept to time scale.[6] I still use it as I found no better term.[7]

The climacogram is the second central moment of the process, as a function of time scale, and thus it is a second-order characteristic of the process. It is related by simple one-to-one transformations to any other second-order characteristic of the process. For example, the transformations relating it to the autocovariance function, defined in (8), are (Koutsoyiannis, 2017):

$$\gamma(k) = 2\int_0^1 (1-\chi)c(\chi k)\mathrm{d}\chi\,, \qquad c(h) = \frac{1}{2}\frac{\mathrm{d}^2\big(h^2\gamma(h)\big)}{\mathrm{d}h^2} \tag{19}$$

and those relating it to the power spectrum are

$$\gamma(k) = \int_0^\infty s(w)\ \mathrm{sinc}^2(\pi w k)\,\mathrm{d}w\,, \qquad s(w) = 2\int_0^\infty \frac{\mathrm{d}^2\big(h^2\gamma(h)\big)}{\mathrm{d}h^2}\cos(2\pi w h)\,\mathrm{d}h \tag{20}$$

where the power spectrum is defined as

$$s(w) \coloneqq 4\int_0^\infty c(h)\cos\,(2\pi w h)\,dh \tag{21}$$

Other transformations relating customary second-order characteristic to each other, as well as those relating continuous-time with discrete-time characteristics thereof, can be found in Koutsoyiannis (2017).

The climacogram, like the autocovariance function, is a positive definite function (Koutsoyiannis, 2017) but of the time scale k, rather than the time lag h. It is not as popular as the other tools but it has several good properties due to its simplicity, close relationship to entropy, 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).

[5] Prior to it, I also tried English terms and in particular the *scale(o)gram*, thinking that the international community would be less reluctant to use it, but I found that it is reserved for another concept (https://en.wikipedia.org/wiki/Scaleogram), so my next though is to compose a Greek term.

[6] Climacogram < Greek Κλιμακόγραμμα < [climax (κλίμαξ) = scale] + [gramma (γράμμα) = written, drawn].

[7] Here I reiterate my willingness to discuss and adopt a better term, if someone suggests one. I have already rejected an initial thought to call it gammogram (from the letter γ used to denote it) mostly because of possible weird connotations in the Greek language.

4 The Allan variance

The Allan variance is named after Allan (1966), who defined the concept (but did not use that name) as seen in the extract shown in Figure 2, while Figure 3 gives an example of more recent use of the concept from Witt (2001). The general setting and terminology used in both these publications (and other ones investigated) belong to the time series literature. Here I am trying to translate the definition into the language of stochastics.

The average angular frequency $\Omega_\tau(t)$ of an oscillator (to distinguish it from spectral frequency ω) over a time interval τ can be written

$$\Omega_\tau(t) = \frac{1}{\tau}\left[\phi(t+\tau) - \phi(t)\right], \qquad (1)$$

where ϕ is the phase angle in radians. Now the variance of the frequency deviations is the square of the standard deviation σ. Define the time average of a function as

$$\langle f(t)\rangle = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} f(t)\,dt. \qquad (2)$$

One may, therefore, write the square of the standard deviation as follows:

$$\sigma^2 = \langle \Omega_\tau(t)^2\rangle - \langle \Omega_\tau(t)\rangle^2. \qquad (3)$$

One may assume with no loss of generality that the second term in (3) can be set equal to zero by a proper translation, since it is the square of the time average frequency. Therefore, $\Omega_\tau(t)$ is now the frequency deviation from the average value, and $\phi(t)$ the integrally related phase deviation.

Substituting (1) into (3) gives

$$\sigma^2 = \frac{1}{\tau^2}\left[\langle \phi(t+\tau)^2\rangle - 2\langle \phi(t+\tau)\cdot\phi(t)\rangle + \langle \phi^2(t)\rangle\right]. \qquad (4)$$

The time average of $\phi(t+\tau)\cdot\phi(t)$ is the autocovariance function of the phase—denoted $R_\phi(\tau)$. One is justified in assuming that a time translation has no effect on the autocovariance function [1], therefore [4],

$$\sigma^2 = \frac{2}{\tau^2}\left[R_\phi(0) - R_\phi(\tau)\right]. \qquad (5)$$

Figure 2 Extract from Allan (1966) introducing what was later called the Allan variance.

If an infinite time series of the quantity $y(t)$ is divided into adjacent segments of duration τ and if $\bar{y}_k$ is the average value of y in the kth interval, then the Allan variance is defined as

$$\sigma_y^2(\tau) = \frac{\langle(\bar{y}_{k+1} - \bar{y}_k)^2\rangle}{2} \tag{2}$$

where the angular brackets denote an infinite time average. In practice, a finite number M of measurements is carried out with a minimum interval τ_0. Adjacent intervals with durations τ_0, $2\tau_0, 3\tau_0, \ldots$ can be constructed and the Allan variance can be estimated from [6, p. 75]

$$\hat{\sigma}_y^2(\tau) = \frac{\sum_{k=1}^{P_0}[\bar{y}_{k+1}(\tau) - \bar{y}_k(\tau)]^2}{2P_0} \tag{3}$$

where P_0 is the number of pairs, $\mathrm{Int}(M/n) - 1$, of $\bar{y}$ and $\tau = n\tau_0$. Equation (3) is the practical working formula for application of the Allan variance. In many cases illustrated below, it is sufficient to calculate the Allan variance for times such that $\tau = 2^q\tau_0$, where q is a nonnegative integer.

Figure 3 Extract of the definition of Allan variance from Witt (2001).

First it is important to inspect whether or not the definition refers to a stationary stochastic process or another type of process, such as the cumulative process defined above, which is a nonstationary process with stationary intervals. Neither of the two publications shown in Figure 2 and Figure 3 clarifies this, but as will be further illustrated in section 5, the information about stationarity is crucial to proceed. From equation (1) in Allan (1966) one may think that $\phi(t)$ corresponds to the cumulative process, so that $\Omega_\tau(t)$, which is named *average* (angular frequency) correspond to a stationary process. If this was the case, by comparing equation (48) below (corresponding to equation (1) of Allan (1966)) with equations (15) and (17), one would conclude that the climacogram and the Allan variance are identical.

However, the subsequent content suggests otherwise, i.e. that $\phi(t)$ corresponds to a stationary process as detailed in Appendix A. In Witt's (2001) definition, shown in Figure 3, this is clearer as he speaks about "adjacent segments of duration τ" and defines $\bar{y}_k$ as "the average value of in the kth interval".

Having clarified the stationarity of the related processes in Allan and Witt, I use my notation on Witt's variant and replace Witt's τ with my k, k with my τ, and $\bar{y}_k$ with my $\underline{x}_\tau^{(k)}$. Thus, Witt's variant of the Allan variance, which I will denote $\gamma_\mathrm{A}(k)$, is

$$\gamma_\mathrm{A}(k) := \frac{1}{2}\mathrm{var}\left[\underline{x}_{\tau+1}^{(k)} - \underline{x}_\tau^{(k)}\right] = \frac{1}{2}\mathrm{E}\left[\left(\underline{x}_{\tau+1}^{(k)} - \underline{x}_\tau^{(k)}\right)^2\right] \tag{22}$$

This can be determined in terms of the climacogram $\gamma(k)$ as follows:

$$\gamma_{\mathrm{A}}(k) = \frac{1}{2}\mathrm{var}\left[\underline{x}_{\tau+1}^{(k)} - \underline{x}_{\tau}^{(k)}\right] = \frac{1}{2}\mathrm{var}\left[\underline{x}_{\tau+1}^{(k)}\right] + \frac{1}{2}\mathrm{var}\left[\underline{x}_{\tau}^{(k)}\right] - \mathrm{cov}\left[\underline{x}_{\tau+1}^{(k)}, \underline{x}_{\tau}^{(k)}\right] \tag{23}$$

On the other hand we have

$$2\gamma(2k) = \frac{1}{2}\mathrm{var}\left[\underline{x}_{\tau+1}^{(k)} + \underline{x}_{\tau}^{(k)}\right] = \frac{1}{2}\mathrm{var}\left[\underline{x}_{\tau+1}^{(k)}\right] + \frac{1}{2}\mathrm{var}\left[\underline{x}_{\tau}^{(k)}\right] + \mathrm{cov}\left[\underline{x}_{\tau+1}^{(k)}, \underline{x}_{\tau}^{(k)}\right] \tag{24}$$

Using the stationarity assumption and adding the two last equations by parts, we find

$$\gamma_{\mathrm{A}}(k) + 2\gamma(2k) = 2\gamma(k) \tag{25}$$

and finally

$$\gamma_{\mathrm{A}}(k) = 2\big(\gamma(k) - \gamma(2k)\big) \tag{26}$$

Hence the Allan variance is not the climacogram, but is twice the difference of the climacogram at two time scales, k and $2k$. Coincidentally, I have shown in Koutsoyiannis (2017) that the very right-hand side of (26) is related to the *conditional entropy* of a Markovian process for the condition that the past is known. I tried to generalize (approximately) this result for non-Markovian processes and also introduced the quantity

$$\zeta(k) \coloneqq \frac{k\big(\gamma(k) - \gamma(2k)\big)}{\ln 2} \tag{27}$$

as a transformation of the climacogram $\gamma(k)$, and derived the inverse transformation, giving $\gamma(k)$ if $\zeta(k)$ is known. Interestingly, $\zeta(k)$ resembles the power spectrum (namely, its graph plotted versus the frequency $w \coloneqq 1/k$ has an area precisely equal to the variance γ_0 of the instantaneous process and its asymptotic slopes on a log-log plot are equal to those of the power spectrum), and thus I have termed $\zeta(k)$ the climacospectrum.

The Allan's (1966) variant of Allan variance is slightly different from Witt's as Allan does not use the factor ½ in his definition. On the other hand, denoting the autocovariance at time scale k and discrete time lag $\eta \coloneqq h/k$ as $c_{\eta}^{(k)} \coloneqq \mathrm{cov}\left[\underline{x}_{\tau}^{(k)}, \underline{x}_{\tau+\eta}^{(k)}\right]$ and observing that $\mathrm{var}\left[\underline{x}_{\tau}^{(k)}\right] = c_0^{(k)}$ and $\mathrm{cov}\left[\underline{x}_{\tau+1}^{(k)}, \underline{x}_{\tau}^{(k)}\right] = c_1^{(k)}$ we can write (23) in terms of covariances as

$$\gamma_{\mathrm{A}}(k) = c_0^{(k)} - c_1^{(k)} \tag{28}$$

which is the equivalent with Allan's equation (5). Here I note that, in contrast to the climacogram, which has the convenient property that its values are identical for the continuous- and discrete-time representation, the autocovariance values in discrete time, $c_{\eta}^{(k)}$, are different from those in continuous time, $c(k\eta)$. Namely, the former are determined from the climacogram and not the continuous-time autocovariance (Koutsoyiannis, 2017), i.e. from:

$$c_\eta^{(k)} = \frac{1}{k^2}\left(\frac{\Gamma(|\eta+1|k) + \Gamma((|\eta-1|k)}{2} - \Gamma(|\eta|k)\right)$$
$$= \frac{(\eta+1)^2\gamma(|\eta+1|k) + (\eta-1)^2\gamma((|\eta-1|k)}{2} - \gamma(|\eta|k) \tag{29}$$

Hence for η = 0 and 1,

$$c_0^{(k)} = \frac{1}{k^2}\Gamma(k) = \gamma(k), \qquad c_1^{(k)} = \frac{1}{k^2}\left(\frac{\Gamma(2k)}{2} - \Gamma(k)\right) = 2\gamma(2k) - \gamma(k) \tag{30}$$

and combining equations (28) and (30) we recover (and thus verify) equation (26).

5 The variogram

For a stochastic process $\underline{x}(t)$, stationary or nonstationary, the variogram (also known as *semivariogram* or *structure function*) is defined to be

$$v(t,u) \coloneqq \frac{1}{2}\mathrm{Var}[\underline{x}(t) - \underline{x}(u)] \equiv \frac{1}{2}(\mathrm{Var}[\underline{x}(t)] + \mathrm{Var}[\underline{x}(u)]) - \mathrm{Cov}[\underline{x}(t), \underline{x}(u)] \tag{31}$$

In this definition, it is clear that the variogram is a function of two time variables, t and u. However, it is commonly used in applications as a function of a single variable, the time lag $h = u - t$, i.e.,

$$v(h) \coloneqq \frac{1}{2}\mathrm{Var}[\underline{x}(t) - \underline{x}(t+h)]$$
$$\equiv \frac{1}{2}(\mathrm{Var}[\underline{x}(t)] + \mathrm{Var}[\underline{x}(t+h)]) - \mathrm{Cov}[\underline{x}(t), \underline{x}(t+h)] \tag{32}$$

Apparently, the latter form is legitimate for a stationary process but it is also used indiscriminately for nonstationary ones. As will be shown below this is legitimate for some types of nonstationary stochastic processes but not for any case of a nonstationary stochastic process. The fact that it can be used in the same form, as a function of a single variable, for stationary and nonstationary processes, has been regarded as an advantage of the tool. However, I contend the opposite, i.e. that it is a disadvantage for these reasons:

- Inference from data is possible only for ergodic processes and ergodic processes are necessarily stationary. Therefore, to make inference we must first transform the nonstationary to a stationary process. Even if we missed that, there should exist such a transformation behind the scene and it is much better, in terms of transparency and validity of mathematical derivations and calculations, if we are conscious about which the underlying stationarity is.
- A theoretically consistent model cannot be developed without clarifying whether the process is stationary or not, and if not, without defining the type of nonstationarity and proposing a model for it.

- In particular for the variogram, if we clarify the underlying process, it turns out that the relationship of the variogram to other second-order characteristics of the process is different for stationary and nonstationary process (see next subsections), and it is important to be aware of the differences.
- The indiscriminate and unaware use of certain hypotheses, which in fact are valid for certain conditions, entails a risk of using them while the conditions do not hold and draw false conclusions.

For these reasons I study below some cases where it is or it is not legitimate to use the variogram as a function of lag, i.e., $v(h)$ as in equation (32), noting that it is always legitimate (albeit inconvenient and ineffective) to use the variogram as a function of two variables, i.e., $v(t, u)$ as in equation (31). I also derive, in each of the cases, the variogram's relation with other stochastic tools and, in particular, the climacogram, which I propose as the benchmark tool for many reasons, among which is its unique property not to be affected by discretization, while all other second-order tools are affected.

5.1 Stationary processes

Apparently, in a stationary process (even a wide-sense stationary one), time translations do not affect second-order properties and thus (31) for lag $h \coloneqq u - t$ entails (32), which can be written as

$$v(h) \coloneqq \frac{1}{2}\mathrm{Var}\big[\underline{x}(t) - \underline{x}(t+h)\big] \equiv \mathrm{Var}\big[\underline{x}(t)\big] - \mathrm{Cov}\big[\underline{x}(t), \underline{x}(t+h)\big] \tag{33}$$

and, considering (8) and (19), it takes the form

$$v(h) \coloneqq \gamma_0 - c(h) = \gamma(0) - \frac{1}{2}\frac{\mathrm{d}^2\big(h^2\gamma(h)\big)}{\mathrm{d}h^2} \tag{34}$$

This is the continuous-time representation. In discrete type, by replacing $\underline{x}(t)$ with $\underline{x}_\tau^{(D)}$ and using (29) we find

$$\begin{aligned} v_\eta^{(D)} &\coloneqq \gamma(D) - c_\eta^{(D)} \\ &= \gamma(D) - \frac{(\eta-1)^2}{2}\gamma\big((|\eta - 1|D\big) + \eta^2\gamma(|\eta|D) - \frac{(\eta+1)^2}{2}\gamma(|\eta+1|D) \end{aligned} \tag{35}$$

Notable is the difference of the latter expression from the continuous-time one. As a particular case, for $\eta = 1$ we easily find that

$$v_1^{(D)} \coloneqq \gamma(D) - c_1^{(D)} = 2\gamma(D) - 2\gamma(2D) \tag{36}$$

and comparison with (26) shows that the discrete-time variogram for lag 1 is identical to the Allan variance.

5.2 Cumulative process

The cumulative process $\underline{X}(t)$ is nonstationary and thus the analysis of subsection 5.1 does not hold. However, its definition from equation (14) suggests that its derivative

$\underline{x}(t) = \mathrm{d}\underline{X}(t)/\mathrm{d}t$ exits[8] and is a stationary process. Thus, in this case the transformation that makes the nonstationary process $\underline{X}(t)$ stationary is differentiation. So, we can base our further analyses on the stationarity (and ergodicity) of the derivative.

For the cumulative process we start our analysis from the general definition (31) replacing $\underline{x}(t)$ with $\underline{X}(t)$:

$$v(t,u) = \frac{1}{2}\mathrm{Var}\left[\underline{X}(t) - \underline{X}(u)\right] \tag{37}$$

As $\underline{X}(t)$ has stationary intervals, time translation does not affect the second order structure of the difference $\underline{X}(t) - \underline{X}(u)$, and we can write

$$v(t,u) = \frac{1}{2}\mathrm{Var}\left[\underline{X}(t-t) - \underline{X}(u-t)\right] = \frac{1}{2}\mathrm{Var}\left[\underline{X}(u-t)\right] \tag{38}$$

(because by definition $\underline{X}(0) \equiv 0$). Noticing that the rightmost part is indeed a function of the time lag $h \coloneqq u - t$, and using (16) and (17) we get

$$v(h) = \frac{\Gamma(h)}{2} = \frac{h^2 \gamma(h)}{2} \tag{39}$$

The difference of (39) from (34) or (35) is spectacular.

5.3 Nonstationary processes

Here I illustrate the non-legitimate use of the univariate form of the variogram in the general case of a nonstationary process. As a simple example, I assume a nonstationary process $\underline{x}(t)$ with time varying mean $\mu(t)$ and variance $\sigma^2(t)$, where $\mu(t)$ and $\sigma(t)$ are deterministic functions of time t.

In this case the transformation

$$\underline{y}(t) \coloneqq \frac{\underline{x}(t) - \mu(t)}{\sigma(t)} \tag{40}$$

makes a process $\underline{y}(t)$ that can be assumed (wide-sense) stationary, with $E\left[\underline{y}(t)\right] = 0$, $\mathrm{var}\left[\underline{y}(t)\right] = 1$, and covariance $c_y(h)$; this latter is any arbitrary function of the lag h only, provided that it is positive definite. The question is, can the covariance or the variogram of $\underline{x}(t)$ also be univariate functions of the time lag?

To study the first part of the question, from (40) we easily find that

$$\begin{aligned} \mathrm{cov}\left[\underline{x}(t), \underline{x}(t+h)\right] &= \mathrm{cov}[\sigma(t)\underline{y}(t), \sigma(t+h)\underline{y}(t+h)] \\ &= \sigma(t)\sigma(t+h)\,\mathrm{E}\left[\underline{y}(t), \underline{y}(t+h)\right] = \sigma(t)\sigma(t+h)c_y(h) \end{aligned} \tag{41}$$

If the function on the left-hand side was a function of h, $c(h)$, then

$$c(h) = \sigma(t)\sigma(t+h)c_y(h) \tag{42}$$

For $h = 0$, noticing that $c_y(0) = \mathrm{var}\left[\underline{y}\right] = 1$, we find

---

[8] Note that not all stochastic processes are differentiable or integrable; see conditions of existence in Papoulis (1991).

$$c(0) = \sigma(t)^2 \tag{43}$$

which entails that $\sigma(t) =$ constant, a result that contradicts the assumption of a time varying variance.

Now we will make a similar analysis for the variogram, assuming that it can be expressed as a univariate function $v(h)$ as in equation (32). This means that

$$v(h) = \frac{1}{2}\left(\mathrm{Var}[\underline{x}(t)] + \mathrm{Var}[\underline{x}(t+h)]\right) - \mathrm{Cov}[\underline{x}(t), \underline{x}(t+h)] \tag{44}$$

which results in

$$v(h) = \frac{\sigma^2(t) + \sigma^2(t+h)}{2} - \sigma(t)\sigma(t+h)c_y(h) \tag{45}$$

Clearly, this has the solution $\sigma(t) = \sigma =$ constant; in this case

$$v(h) = \sigma^2(1 - c_y(h)) \tag{46}$$

but this solution contradicts the assumption that the variance is time varying. Except for the trivial solution $\sigma(t) = \sigma$, (45) cannot hold for an arbitrary function $\sigma(t)$, as demonstrated in Appendix B.

Thus, neither the autocovariance, nor the variogram can be functions of a single variable in the general case of a nonstationary process. However, if nonstationarity holds only for the mean, then both the autocovariance and the variogram become functions of a single variable because none of the above results depends on the function $\mu(t)$. In this respect, even if we regarded as an advantage the applicability of the same tool to stationary and nonstationary processes indistinguishably (which I do not) the variogram would not have an advantage over the autocovariance (or the climacogram).

6 Discussion and conclusion

When dealing with stochastics, overconfidence and certainty in assertions are not useful[9] because stochastic concepts and tools are delicate. For this reason, I am posting this document for discussion, hoping that possible errors or wrong assertions would be spotted and corrected.

The above discourse indicates some of the virtues of the climacogram but it has more. To refer to just one additional, it is readily expandable beyond a second-order representation of a stochastic process, still providing characterization of high-order properties of processes in terms of univariate functions of time scale (Dimitriadis and Koutsoyiannis, 2018; Koutsoyiannis, 2018a; Koutsoyiannis et al., 2018).

Because of its virtues, I am using the climacogram as benchmark for comparison with other stochastic tools. Here comparisons showed that the climacogram is not identical to Allan variance and to the variogram, but, evidently, it is related to them and the relationships have thoroughly been studied above. In particular it has been shown that

[9] I believe they did not even benefit Professor A in this particular case.

the Allan variance is twice the difference of the climacogram at scales k and $2k$; also, it equals the variogram for lag one in a discrete time scale representation.

The fact that stochastics deals with uncertainty should not mislead us to think that stochastic concepts are, or need to be, uncertain per se.[10] In this respect, we owe a lot to the Moscow school of mathematics, for its substantial contribution for the solid and rigorous foundation of probability (e.g., Kolmogorov, 1933) and stochastic processes (e.g., Kolmogorov, 1931, 1937, 1938, 1947; Khintchine, 1933, 1934). However, I must admit that clarity is not appreciated by everyone in the scientific community. In my other talk in the same event (Koutsoyiannis, 2018c) I referred to this fact using the following quotations, suggesting the existence of two opposite schools of thought.

> *Each definition is a piece of secret ripped from Nature by the human spirit. I insist on this: any complicated thing, being illumined by definitions, being laid out in them, being broken up into pieces, will be separated Into pieces completely transparent even to a child, excluding foggy and dark parts that our intuition whispers to us while acting, separating into logical pieces, then only can we move further, towards new successes due to definitions . . .*
>
> Nikolai Luzin (from Graham and Kantor, 2009)

> *Let me argue that this situation* [absence of a definition] *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.*
>
> Benoit Mandelbrot (1999, p. 14)

My clear preference is for Luzin's school.

Appendix A

To demonstrate that $\phi(t)$ in Allan (1966) corresponds to a stationary process, we assume the opposite, i.e. we consider a stochastic process $\underline{\phi}(t)$ that is nonstationary with stationary intervals, so that the intervals $\underline{\phi}(t+\tau) - \underline{\phi}(t)$ form a stationary process for a fixed τ and $t = \tau, 2\tau, 3\tau, \ldots$. Obviously, in this case,

$$\mathrm{E}\left[\underline{\phi}(t+\tau) - \underline{\phi}(t)\right] = \mu\tau \tag{47}$$

where μ denotes average (e.g., $\mu = \mathrm{E}[\underline{\phi}(2\tau) - \underline{\phi}(\tau)]$, so that if the process were stationary, then $\mu = 0$).

In accord with Allan (1966) we define

$$\underline{\Omega}_\tau(t) \coloneqq \frac{1}{\tau}\left(\underline{\phi}(t+\tau) - \underline{\phi}(t)\right) \tag{48}$$

for which

$$\mathrm{E}\left[\underline{\Omega}_\tau(t)^2\right] = \frac{1}{\tau^2}\left(\mathrm{E}\left[\underline{\phi}(t+\tau)^2\right] - 2\mathrm{E}\left[\underline{\phi}(t+\tau)\underline{\phi}(t)\right] + \mathrm{E}\left[\underline{\phi}(t)^2\right]\right) \tag{49}$$

---

[10] I would not write this document if I believed otherwise.

and thus

$$\begin{aligned}\operatorname{var}\left[\underline{\Omega}_\tau(t)\right] &= \mathrm{E}\left[\underline{\Omega}_\tau(t)^2\right] - \mathrm{E}\left[\underline{\Omega}_\tau(t)\right]^2 \\ &= \frac{1}{\tau^2}\left(\mathrm{E}\left[\underline{\phi}(t+\tau)^2\right] - 2\mathrm{E}\left[\underline{\phi}(t+\tau)\underline{\phi}(t)\right] + \mathrm{E}\left[\underline{\phi}(t)^2\right]\right) - \mu^2\tau^2\end{aligned} \tag{50}$$

Furthermore, from (48) we get

$$\underline{\phi}(t+\tau) = \underline{\phi}(t) + \tau\underline{\Omega}_\tau(t) \tag{51}$$

and hence

$$\mathrm{E}\left[\underline{\phi}(t+\tau)^2\right] = \mathrm{E}\left[\underline{\phi}(t)^2\right] + 2\tau\mathrm{E}\left[\underline{\phi}(t)\underline{\Omega}_\tau(t)\right] + \tau^2\mathrm{E}\left[\underline{\Omega}_\tau(t)^2\right] \neq \mathrm{E}\left[\underline{\phi}(t)^2\right] \tag{52}$$

as it cannot be supported that the nonstationary process $\underline{\phi}(t)$ is such that $2\mathrm{E}\left[\underline{\phi}(t)\underline{\Omega}_\tau(t)\right] + \tau\mathrm{E}\left[\underline{\Omega}_\tau(t)^2\right] \equiv 0$ for any t and τ (and such an assumption, which practically would not be different from the simpler assumption of wide-sense stationarity, would be stated in the paper if it was the case).

Now, noting that the equations derived here in terms of expectations should correspond to the equations of Allan (1966), which are given in terms of time average limits, we observe that:

(a) Equation (50) differs from Allan's equation (4) in the term $-\mu^2\tau^2$ (which would be zero if the process was stationary).

(b) Allan's note that "*a time translation has no effect on the autocovariance function*" contradicts inequality (52), because $\mathrm{E}\left[\underline{\phi}(t)^2\right]$ equals the autocovariance according to Allan's definition for $\tau = 0$.

Furthermore, Allan's reference [4], i.e. Searle et al. (1964), clearly speaks about "*a stationary time function*."

This analysis proves that Allan's process $\underline{\phi}(t)$ is stationary as the opposite would be contradictory.

Appendix B

Writing (45) after replacing t with $t + h$ and h with $-h$ we get

$$v(-h) = \frac{\sigma^2(t+h) + \sigma^2(t)}{2} - \sigma(t+h)\sigma(t)c_y(h) \tag{53}$$

Comparing the last equation with (45) we get $v(-h) = v(h)$, which means that $v(h)$, if it exists, will be an even function. Now, we write again (45) without replacing t but replacing h with $-h$ and get

$$v(-h) = \frac{\sigma^2(t) + \sigma^2(t-h)}{2} - \sigma(t)\sigma(t-h)c_y(h) \tag{54}$$

Subsequently, we subtract (45) and (54) by parts and find

$$0 = \frac{\sigma^2(t+h) - \sigma^2(t-h)}{2} - \sigma(t)c_y(h)\big(\sigma(t+h) - \sigma(t-h)\big) \tag{55}$$

In the trivial case of constant variance we have $\sigma(t+h)-\sigma(t-h)=0$ and equation (55) holds true. Otherwise, i.e., if $\sigma(t+h)-\sigma(t-h)\neq 0$, we divide (55) by $\sigma(t+h)-\sigma(t-h)$ and get

$$0=\frac{\sigma(t+h)+\sigma(t-h)}{2}-\sigma(t)c_y(h) \tag{56}$$

or

$$\sigma(t+h)+\sigma(t-h)=2\sigma(t)c_y(h) \tag{57}$$

As in the right-hand side the function is separable in terms of t and h, for the left-hand side we try the same, i.e., a solution of the form $\sigma(t+h)=A(t)B(h)$. Interchangeability of t and h implies that $A(t)B(h)=A(h)B(t)$, which entails $B(h)=A(h)$, so that (57) is written as

$$A(t)\big(A(h)+A(-h)\big)=2\sigma(t)c_y(h) \tag{58}$$

so that necessarily $A(t)=C\sigma(t)$ for some constant C. Hence

$$C^2\big(\sigma(h)+\sigma(-h)\big)=2c_y(h) \tag{59}$$

and setting h = 0 in the above and noticing that $c_y(0)=1$ we conclude that $C=1/\sqrt{\sigma(0)}$ and $A(t)=\sigma(t)/\sqrt{\sigma(0)}$.

Writing (57) for $t=h$ we get

$$\sigma(h+h)+\sigma(0)=2\sigma(h)c_y(h) \tag{60}$$

or

$$\big(\sigma(h)\big)^2/\sigma(0)+\sigma(0)-2\sigma(h)c_y(h)=0 \tag{61}$$

This is a quadratic equation in terms of $\sigma(h)$ and its discriminant is $4(c_y(h)^2-1)$. As $c_y(h)$, being autocovariance, is a positive definite function with $c_y(0)=1$, the discriminant is negative and thus there is no solution of the quadratic equation for $\sigma(h)$.

For further illustration I provide two specific examples, in which I do not use the above theoretical results. In the first one I consider the case $\sigma(t)=at^b$, in which

$$\begin{aligned} v(h)&=\frac{a^2t^{2b}}{2}\left(1+\left(1+\frac{h}{t}\right)^{2b}\right)-a^2t^{2b}\left(1+\frac{h}{t}\right)^{b}c_y(h)\\ &=\frac{a^2t^{2b}}{2}\left(1+\left(1+\frac{h}{t}\right)^{2b}-2\left(1+\frac{h}{t}\right)^{b}c_y(h)\right)\end{aligned} \tag{62}$$

where the right-hand side is a function of t (not only of h), unless b = 0, which corresponds to constant variance. The linear case, b = 1, and the square-root case, b = 1/2, do not eliminate dependence on t.

In the second example I assume $\sigma(t)=a\mathrm{e}^{bt}$ in which

$$v(h)=\frac{a^2\mathrm{e}^{2bt}(1+\mathrm{e}^{2bh})}{2}-a^2\mathrm{e}^{2bt+h}c_y(h)=\frac{a^2\mathrm{e}^{2bt}}{2}\big(1+\mathrm{e}^{2bh}-2\mathrm{e}^{h}c_y(h)\big) \tag{63}$$

where again the right-hand side is a function of t (not only of h), except in the trivial case of constant variance (b = 0). Interestingly, in this case the dependence of the right-hand side on t is separable from the dependence on h, while in the previous example this happens only when $h/t \to 0$.

References

Allan, D.W. (1966), Statistics of atomic frequency standard, *Proceedings of the IEEE*, 54(2), 221–230.

Beran, J. (1994), *Statistics for Long-Memory Processes*, Chapman & Hall, New York.

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*,* 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.

Fisher, R.A. (1918), The correlation between relatives on the supposition of Mendelian inheritance, *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 52(2), 399-433.

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

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

Khintchine, A. (1933), Zu Birkhoffs Lösung des Ergodenproblems, *Mathematische Annalen*, 107, 485-488.

Khintchine, A., 1934. Korrelationstheorie der stationären stochastischen Prozesse. *Mathematische Annalen*, 109 (1), 604–615.

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

Kolmogorov, A.N. (1933), Grundbegriffe der Wahrscheinlichkeitsrechnung (Ergebnisse der Mathematik und Ihrer Grenzgebiete), Berlin (2nd English Edition: Foundations of the Theory of Probability, 84 pp. Chelsea Publishing Company, New York, 1956).

Kolmogorov, A.N. (1937), Ein vereinfachter Beweis der Birkhoff-Khintchineschen Ergodensatzes. *Mat. Sb.*, 2, 367-368.

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

Kolmogorov, A.N. (1947), Statistical theory of oscillations with continuous spectrum. Collected papers on the 30th anniversary of the Great October Socialist Revolution, Vol. 1, Akad. Nauk SSSR, Moscow-Leningrad, pp. 242-252. (English edition: Kolmogorov, A.N., 1992, *Selected Works of A. N. Kolmogorov - Volume 2, Probability Theory and Mathematical Statistics*, A. N. Shiryayev ed., Kluwer, Dordrecht, The Netherlands, pp. 321-330).

Koutsoyiannis, D. (2003), Climate change, the Hurst phenomenon, and hydrological statistics, *Hydrological Sciences Journal,* 48 (1), 3–24, doi: 10.1623/hysj.48.1.3.43481, 2003.

Koutsoyiannis, D. (2006), Nonstationarity versus scaling in hydrology, *Journal of Hydrology*, 324, 239–254.

Koutsoyiannis, D. (2010), A random walk on water, *Hydrology and Earth System Sciences*, 14, 585–601, doi: 10.5194/hess-14-585-2010.

Koutsoyiannis, D. (2011), Hurst-Kolmogorov dynamics and uncertainty, *Journal of the American Water Resources Association*, 47 (3), 481–495, doi: 10.1111/j.1752-1688.2011.00543.x.

Koutsoyiannis, D. (2016), Generic and parsimonious stochastic modelling for hydrology and beyond, *Hydrological Sciences Journal*, *61*, 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. (2018a), Knowable moments for high-order stochastic characterization and modelling of hydrological processes, *Hydrological Sciences Journal* (in review; preprint available online: doi: 10.13140/RG.2.2.14876.59529).

Koutsoyiannis, D. (2018b), Climate change impacts on hydrological science: How the climate change agenda has lowered the scientific level of hydrology, *School for Young Scientists "Modelling and forecasting of river flows and managing hydrological risks: Towards a new generation of methods" (2018)*, Russian Academy of Sciences, Lomonosov Moscow State University, doi: 10.13140/RG.2.2.11110.06727.

D. Koutsoyiannis, (2018c) Modelling extreme rainfall in the era of climate change concerns: Towards a consistent stochastic methodology, S*chool for Young Scientists "Modelling and forecasting of river flows and managing hydrological risks: Towards a new generation of methods" (2018)*, Russian Academy of Sciences, Lomonosov Moscow State University, Moscow, doi: 10.13140/RG.2.2.22015.25766.

Koutsoyiannis, D., Dimitriadis, P., Lombardo, F, and Stevens, S. (2018), From fractals to stochastics: Seeking theoretical consistency in analysis of geophysical data, *Advances in Nonlinear Geosciences*, edited by A.A. Tsonis, 237–278, Springer, doi: 10.1007/978-3-319-58895-7_14.

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.

Mackey, M.C. (1992), *Time's Arrow: The Origins of Thermodynamic Behavior,* Dover, Mineola, NY, USA, 175 pp.,

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

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

Searle, C.L., Posner, R.D., Badessa, R.S., and Bates, Y.J. (1965), Computer-aided calculation of frequency stability, *1964 Proc. Symp. on the Definition and Measurement of Short-Term Frequency Stability*, NASA SP-80, 273-277, US Govt. Printing Office, Washington DC.

Pearson, K. (1894), On the dissection of asymmetrical frequency curves, *Philosophical Transactions of the Royal Society A*, 185, 71–110, doi: 10.1098/rsta.1894.0003.

Witt, T.J. (2001), Using the Allan variance and power spectral density to characterize DC nanovoltmeters, *IEEE Transactions on Instrumentation and Measurement*, 50 (2) 445-448.