Lecture Notes on Stochastics

Version 9 – Athens 2017

Encolpion of stochastics

Fundamentals of stochastic 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/)

Preface

Most things are uncertain. Stochastics is the language of uncertainty. I believe the gospel of stochastics is the book by Papoulis (1991). However, as Papoulis was an electrical engineer, his approach may need some additions or adaptations in order to be applied to geophysical processes. The peculiarities of the latter are that:

(a) their modelling relies very much on observational data because geophysical systems are too complex to be studied by theoretical reasoning and deduction, and theories are often inadequate;
(b) the distinction "signal vs. noise" is meaningless;
(c) the samples are small;
(d) they are often characterized by long term persistence, which makes classical statistics inappropriate.

Having studied several hydroclimatic processes, I derived in handwritten notes some equations useful for such processes. Having repeated such derivations several times, because I forgot that I had produced them before, or lost the notes, I decided to produce this document. Some of the equations and remarks contained here can be found in other texts, particularly in Papoulis (but with many differences in notational and other conventions), but some other cannot. I believe they can be useful to other people, researchers and students.

Stochastics is much more than calculations. Popular computer programs have made calculations easy and fast. But numerical results may mean nothing! It is better not to use programs (even better, not to make calculations) if we are unaware of the stochastic properties of the object (see some examples in Papalexiou et al. 2010). Even more so, if we have not assimilated the fundamental concepts of stochastics, including probability theory and statistics. These are difficult to assimilate. I believe those who did not have difficulties in understanding the concepts have inadequate knowledge of stochastics.

Introduction

Most of the popular knowledge in stochastics originates from so-called time-series books. These have given focus on stylized families of models like AR(p), ARMA(p,q), ARIMA(p,d,q), ARFIMA(p,d,q) etc. Most of these are too artificial and unnecessary. They are too artificial because they are over-parameterized and thus not parsimonious, and because, being complicated discrete-time models, they do not correspond to a continuous time process; but natural processes typically evolve in continuous time. They are unnecessary because there exist general methods to generate synthetic series from a process with any arbitrary autocorrelation structure (Koutsoyiannis, 2000, 2016).

Other, more fashionable, texts have given emphasis to mathematical tools like fractals, which are often applied inattentively, treating random variables as if they were deterministic and cursorily using uncontrollable quantities (e.g. high order moments) whose estimates are characterized by extraordinarily high bias and uncertainty (Lombardo et al., 2014).

Even the second order joint moment of a process, the autocorrelation function, is characterized by high bias and uncertainty (Koutsoyiannis and Montanari, 2007; Papalexiou et al., 2010; Lombardo et al., 2014) which are transferred to transformations thereof, e.g. the power spectrum. For this reason, I believe the best concept to use in stochastic modelling, particularly in parameter estimation, is the climacogram (Koutsoyiannis, 2010, 2016; Dimitriadis and Koutsoyiannis, 2015; see also p. 28 below). This is defined to be the variance (or the standard deviation) of the time-averaged process as a function of time scale of averaging. It involves bias too, but its bias can be determined analytically and included in the estimation (see below, p. 29). The climacogram, the autocorrelation function and the power spectrum are transformations one another. The notes below give all details, from definition to theoretical derivation and estimation from data. For the latter the details given are for the climacogram only; once a particular model is fitted based on its climacogram, its autocovariance function and power spectrum can be derived analytically.

Fundamental remarks on random variables

- A *random variable* is not a regular variable and need not necessarily describe something random; it just describes something uncertain, unpredictable, unknown. While a regular variable takes on one value at a time, a random variable is a more abstract mathematical object that takes on all its possible values at once, but not necessarily in a uniform manner; therefore a distribution function should always be associated with a random variable.
- A random variable needs a special notation to distinguish it from a regular variable x; the best notation devised is the so-called Dutch convention (Hemelrijk, 1966), according to which random variables are underlined, i.e. $\underline{x}$. The *distribution function* $F(x) \coloneqq P\{\underline{x} \leq x\}$, where P denotes *probability*, and the *probability density function* $f(x) := \mathrm{d}F(x)/\mathrm{d}x$ are real-valued functions of the real (regular) variable x, not of the random variable $\underline{x}$. Realizations of $\underline{x}$ are typically denoted by the non-underlined symbol x.
- 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.
- Distinguishing random variables from regular ones is essential. For example, if we assume that $\underline{x}$ and $\underline{y}$ are independent random variables uniformly distributed in [0, 1], then it is easy to verify that $P\{\underline{x} \leq \underline{y}\} = 0.5$ (a constant) while $P\{\underline{x} \leq y\} = \max(0, \min(y, 1))$ (a function of y). For $\underline{x}$ and $\underline{y}$ not independent, the conditional expectations $\mathrm{E}[\underline{x}|y]$ and $\mathrm{E}[\underline{x}|\underline{y}]$ are different; the former is a regular variable and a function of y, but the second is a random variable, a function of $\underline{y}$, whose expected value is $\mathrm{E}[\underline{x}]$ (i.e., $\mathrm{E}[\mathrm{E}[\underline{x}|\underline{y}]] = \mathrm{E}[\underline{x}] \neq \mathrm{E}[\underline{x}|y]$). As another example, both conditional entropies $\Phi[\underline{x}|y]$ and $\Phi[\underline{x}|\underline{y}]$ are numbers (not random variables) but while the latter obeys $\Phi[\underline{x}|\underline{y}] \leq \Phi[\underline{x}]$, the former can be either greater or less than the unconditional $\Phi[\underline{x}]$ (Papoulis, 1991, pp. 172, 564); this has led to serious misconceptions in cases of inattentive notation.

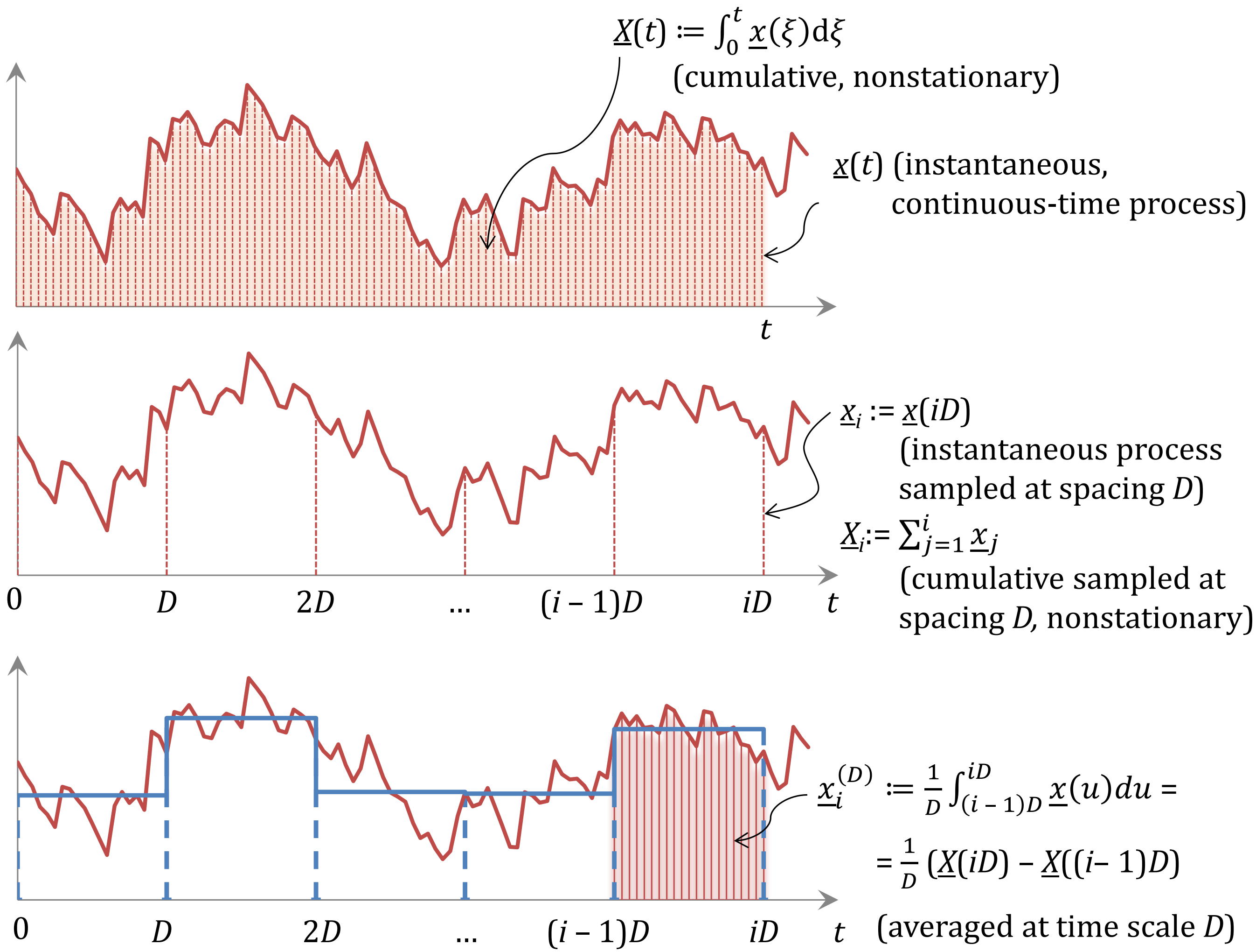
Fundamental remarks on stochastic processes

- A *stochastic process* is a family of infinitely many random variables indexed by a (regular) variable. The index typically represents time and is either a real number, t, in a continuous-time stochastic process $\underline{x}(t)$, or an integer, i, in a discrete-time stochastic process $\underline{x}_i$.
- Realizations, x_i, of a stochastic process, $\underline{x}_i$ or $\underline{x}(t)$, at a finite set of discrete time instances i (or t_i) are called *time series*. Some texts confuse time series and stochastic processes, as they also do with random and regular variables. To avoid such confusion, which may have dramatic effects, we must have always in mind that a stochastic process is a family of random variables, infinitely many for discrete time processes and uncountably infinitely many for continuous time processes. On the other hand, a time series is a finite sequence of numbers.
- Central to the notion of a stochastic process are the concepts of *stationarity* and *nonstationarity*, two widely misunderstood and misused concepts (cf. the popular assertion "stationarity is dead"; see Koutsoyiannis and Montanari, 2014; Montanari and Koutsoyiannis, 2015), 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, 2011b and Koutsoyiannis and Montanari, 2015). Conversely, a process is nonstationary if some of its statistics are changing through time and their change is described as a deterministic function of time.
- A nonstationary process can be handled theoretically without insurmountable difficulties—for example the cumulative process $\underline{X}(t)$ defined below is nonstationary ($\underline{X}(t) \coloneqq \int_0^t \underline{x}(u)\mathrm{d}u$, where $\underline{x}(t)$ is a stationary process); it needs some attention though, as its statistical properties depend on time. However, handling (or even detecting) nonstationarity merely from data may be difficult, if not impossible. Ironically, in many of the cases where nonstationarity has been claimed based on data analyses, the stochastic tools used are meaningful only for stationary processes.

Fundamental remarks on stochastic processes (contd.)

- 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. Ironically, several studies use time series data to estimate statistical properties, as if the process were ergodic, while at the same time what they (cursorily) estimate may falsify the ergodicity hypothesis (see example on p. 22).
- 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) in the following manner: A stochastic process $\underline{x}(t)$ is ergodic if the time average of any (integrable) function $g(\underline{x}(t))$, as time tends to infinity, equals the true (ensemble) expectation $\mathrm{E}[g(\underline{x}(t))]$.
- If the system that is modelled in a stochastic framework has deterministic dynamics (meaning that a system input will give a single system response, as happens for example in most hydrological models) then a theorem applies (Mackey 1992, p. 52), according to which a dynamical system has a stationary probability density *if and only if* it is ergodic. Therefore, a stationary system is also ergodic and vice versa, and a nonstationary system is also non-ergodic and vice versa.
- If the system dynamics is stochastic (a single input could result in multiple outputs), then ergodicity and stationarity do not necessarily coincide. However, recalling that a stochastic process is a model and not part of the real world, we can always conveniently device a stochastic process that is ergodic, provided that we have excluded nonstationarity (see example in Koutsoyiannis and Montanari, 2015).
- In conclusion, 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.

A schematic for definitions and notation



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.

Definitions and notation

Type	Continuous time	Discrete time by aggregating or averaging at time interval D	Discrete time by sampling at spacing D (and aggregating at scale κD, where κ is integer)
Stochastic processes	$\underline{x}(t)$: instantaneous, stationary $\underline{X}(t) := \int_0^t \underline{x}(\xi)\mathrm{d}\xi$: cumulative, nonstationary	$\underline{X}_i^{(D)} := \underline{X}(iD) - \underline{X}((i-1)D)$: aggregated, stationary intervals of $\underline{X}(t)$ $\underline{x}_i^{(D)} := \underline{X}_i^{(D)} / D$: averaged	$\underline{x}_i := \underline{x}(iD)$: sampled instantaneous $\underline{X}_i := \sum_{j=1}^{i} \underline{x}_j$: cumulative $\underline{\check{X}}_i^{(\kappa)} := \underline{X}_{i\kappa} - \underline{X}_{(i-1)\kappa}$: aggregated $\underline{\check{x}}_i^{(\kappa)} := \underline{\check{X}}_i^{(\kappa)}/\kappa$: averaged
Characteristic variances	$\gamma_0 := \mathrm{Var}[\underline{x}(t)]$ $\Gamma(t) := \mathrm{Var}[\underline{X}(t)]$ $\gamma(t) := \mathrm{Var}[\underline{X}(t)/t] = \Gamma(t)/t^2$ Note: $\Gamma(0) = 0;\ \gamma(0) = \gamma_0$	$\mathrm{Var}[\underline{X}_i^{(D)}] = \Gamma(D)$ $\mathrm{Var}[\underline{x}_i^{(D)}] = \gamma(D)$	$\gamma_0\ (= \mathrm{Var}[\underline{x}_i])$ $\check{\Gamma}^{(\kappa)} := \mathrm{Var}[\underline{\check{X}}_i^{(\kappa)}]$ $\check{\gamma}^{(\kappa)} := \mathrm{Var}[\underline{\check{x}}_i^{(\kappa)}] = \check{\Gamma}^{(\kappa)}/\kappa^2$ Note: $\check{\Gamma}^{(1)} = \check{\gamma}^{(1)} = \gamma_0$
Autocovariance function	$c(h) := \mathrm{Cov}[\underline{x}(t), \underline{x}(t+h)]$ Note: $c(0) \equiv \gamma_0 = \gamma(0)$	$c_j^{(D)} := \mathrm{Cov}[\underline{x}_i^{(D)}, \underline{x}_{i+j}^{(D)}]$ Note: $c_0^{(D)} \equiv \gamma(D)$	$\check{c}_j := \mathrm{Cov}[\underline{x}_i, \underline{x}_{i+j}] = c(jD)$ Note: $\check{c}_0 \equiv \gamma_0$
Power spectrum (spectral density)	$s(w) :=$ $2\int_{-\infty}^{\infty} c(h)\cos(2\pi w h)\,\mathrm{d}h$	$s_\mathrm{d}^{(D)}(\omega) :=$ $2\sum_{j=-\infty}^{\infty} c_j^{(D)} \cos(2\pi\omega j)$ $s^{(D)}(w) = D s_\mathrm{d}^{(D)}(wD)$	$\check{s}_\mathrm{d}(\omega) := 2\sum_{j=-\infty}^{\infty} \check{c}_j \cos(2\pi\omega j)$ $\check{s}(w) = D s_\mathrm{d}(wD)$

Note: all equations for the averaged process are valid also if the time interval D is replaced by any time scale k.

Notes on notation

1. Notation of time-related quantities (Latin letters denote dimensional quantities, while Greek letters denote dimensionless ones

   - *Discretization time unit* (*time step* in case of sampling or *time scale* in case of aggregating or averaging), D.
   - *Time*, $t = \tau D$ (alternatively for strictly integer $i = 0, 1, 2, \ldots$, $t = i D$ where i is discrete time).
   - *Time lag*, $h = \eta D$ (alternatively for strictly integer $j = 0, 1, 2, \ldots$, $h = j D$, where j is discrete time lag).
   - *Time scale* $k = \kappa D$.
   - *Frequency* (inverse time scale), $w = \omega / D$, related to time scale by $w = 1/k$, $\omega = 1/\kappa$.

2. Both w and ω are real numbers ranging in $(-\infty, \infty)$ for a continuous-time process, while for a discrete-time process w ranges in $[-1/2D, 1/2D]$ and ω in $[-½, ½]$. As both the autocovariance function and the power spectrum are even functions, we make all calculations for a continuous-time process in $(0, \infty)$, and for a discrete-time process in $[0, 1/2D]$ for w and in $[0, ½]$ for ω (see below).

3. The standard deviations of the instantaneous and averaged processes are denoted, respectively, as $\sigma_0 \coloneqq \sqrt{\gamma_0}$, $\sigma(D) \coloneqq \sqrt{\gamma(D)}$.

4. The power spectrum of the continuous-time (instantaneous) process, denoted as $s(w)$, is twice the cosine Fourier transform of the autocovariance function of the process, while those of the discrete-time processes, denoted as $s_{\mathrm{d}}^{(D)}(\omega)$ and $\check{s}_{\mathrm{d}}(\omega)$, are twice the inverse finite cosine Fourier transforms of the respective autocovariance functions. The convention of the multiplying factor 2 in the Fourier transforms was adopted so that the integral of the spectrum on positive frequencies only equals the variance of the process. The spectra versions$s^{(D)}(w)$ and $\check{s}(w)$ (with $w = \omega/D$) are comparable have the same dimensions same as the continuous-time spectrum $s(w)$.

Computation of variances and covariances

The climacogram and the autocovariance function of a process either continuous- or discrete-time, are transformations one another. The climacogram of the continuous time process, as well as those of the averaged or aggregated process, can be calculated from the autocovariance function of the continuous time process as follows (cf. Papoulis, 1991, p. 299):

$$\Gamma(k) = 2\int_0^k (k-h)c(h)\mathrm{d}h \tag{1}$$

$$\gamma(k) = \frac{\Gamma(k)}{k^2} = \frac{2}{k^2}\int_0^k (k-h)c(h)\mathrm{d}h = 2\int_0^1 (1-\xi)c(\xi k)\mathrm{d}\xi \tag{2}$$

The inverse formula, by which we can find the autocovariance if the climacogram is known, is easily derived by taking the second derivative in (1) also using Leibniz's integral rule:

$$c(h) = \frac{1}{2}\,\frac{\mathrm{d}^2\Gamma(h)}{\mathrm{d}h^2} = \frac{1}{2}\,\frac{\mathrm{d}^2\big(h^2\gamma(h)\big)}{\mathrm{d}h^2} \tag{3}$$

The climacogram of the sampled aggregated and the sampled averaged process cannot be found in this way, but needs to be calculated in a purely discrete-time framework (cf. Papoulis, 1991, p. 432; Koutsoyiannis, 2010), as shown in next page.

Using l'Hôpital's rule on (2), the following useful asymptotic properties are derived:

$$c(0) = \gamma(0), \quad c'(0) = 3\gamma'(0) \tag{4}$$

Computation of variances and covariances (contd.)

The following relationships give the climacogram from the autocovariance function for the discrete-time sampled process.

$$\check{\Gamma}^{(\kappa)} = \kappa\gamma_0 + 2\sum_{j=1}^{\kappa-1}(\kappa - j)\check{c}_j \tag{5}$$

$$\check{\gamma}^{(\kappa)} = \frac{\check{\Gamma}^{(\kappa)}}{\kappa^2} = \frac{1}{\kappa}\left(\gamma_0 + 2\sum_{j=1}^{\kappa-1}\left(1 - \frac{j}{\kappa}\right)c_j^*\right) \tag{6}$$

The following recursive relationship facilitates calculation of the climacogram:

$$\check{\Gamma}^{(\kappa)} = 2\check{\Gamma}^{(\kappa-1)} - \check{\Gamma}^{(\kappa-2)} + 2\check{c}_{\kappa-1}, \quad \check{\Gamma}^{(0)} = 0, \quad \check{\Gamma}^{(1)} = \check{\gamma}_0 \tag{7}$$

The following inverse relationships give the autocovariance functions from the climacograms for the discrete-time processes of interest.

$$c_j^{(D)} = \frac{1}{2}\frac{\delta_D^2 \Gamma(jD)}{D^2} = \frac{1}{D^2}\left(\frac{\Gamma\big((j+1)D\big)+\Gamma\big((j-1)D\big)}{2} - \Gamma(jD)\right) \tag{8}$$

$$\check{c}_j = \frac{1}{2}\,\delta_1^2\check{\Gamma}^{(j)} = \frac{\check{\Gamma}^{(j+1)} + \check{\Gamma}^{(j-1)}}{2} - \check{\Gamma}^{(j)} \tag{9}$$

Here δ_D^2 is the second order central difference operator for spacing D and δ_D^2/D^2 is the second finite derivative. Both formulae are exact (not approximations). The latter holds also for any type of discrete time process, even if it is not linked to a continuous-time one. Both are valid for $j > 0$, but can be written in the following manner so as to be valid for any integer j:

$$c_j^{(D)} = \frac{1}{D^2}\left(\frac{\Gamma(|j+1|D)+\Gamma\big((|j-1|D\big)}{2} - \Gamma(|j|D)\right), \; \check{c}_j = \frac{\check{\Gamma}^{(|j+1|)}+\check{\Gamma}^{(|j-1|)}}{2} - \check{\Gamma}^{(|j|)} \tag{10}$$

Power spectrum

The power spectrum of the continuous-time process is calculated from the autocovariance function as

$$s(w) = 4\int_0^\infty c(h)\cos(2\pi w h)\,\mathrm{d}h \tag{11}$$

The inverse transformation is

$$c(h) = \int_0^\infty s(w)\cos(2\pi w h)\,\mathrm{d}w \tag{12}$$

The power spectrum of a discrete-time process is calculated from the autocovariance function as

$$s_\mathrm{d}^{(D)}(\omega) = 2c_0^{(D)} + 4\sum_{j=1}^\infty c_j^{(D)}\cos(2\pi\omega j), \qquad \check{s}_\mathrm{d}(\omega) = 2\check{c}_0 + 4\sum_{j=1}^\infty \check{c}_j\cos(2\pi\omega j) \tag{13}$$

while $s^{(D)}(w)$ and $\check{s}(w)$ are readily derived from their definitions.

The inverse transformation is

$$c_j^{(D)} = \int_0^{1/2} s_\mathrm{d}^{(D)}(\omega)\cos(2\pi\omega j)\,\mathrm{d}\omega\,, \qquad \check{c}_j = \int_0^{1/2} \check{s}_\mathrm{d}(\omega)\cos(2\pi\omega j)\,\mathrm{d}\omega \tag{14}$$

Notice that, even in the discrete case, the inverse transformation is an integral, not a sum.

Relationships between power spectra of discrete and continuous time

The direct calculation of the discrete-time power spectra from that of the continuous time is feasible but involved.

Specifically, the following relationship connects $\check{s}_{\mathrm{d}}(\omega)$ with $s(w)$ (cf. Papoulis, 1991, p. 336):

$$\check{s}_{\mathrm{d}}(\omega) = \frac{\check{s}(w)}{D} = \frac{1}{D} \sum_{j=-\infty}^{\infty} s\left(\frac{\omega + j}{D}\right) = \frac{1}{D} \sum_{j=-\infty}^{\infty} s\left(w + \frac{j}{D}\right), w = \omega \,/\, D \qquad (15)$$

To find the relationship of $s_{\mathrm{d}}^{(D)}(\omega)$ with $s(w)$, we utilize (15) after we have applied a moving average operation to the process. The power spectrum of the process $\underline{y}(t)$ obtained as the moving average of $\underline{x}(t)$ at time scale D, i.e., $\underline{y}(t) \coloneqq \frac{1}{D}\int_{t-D}^{t} \underline{x}(\xi)\mathrm{d}\xi$ is (cf. Papoulis, 1991, p. 325, also noticing the different notational and other conventions):

$$s_y(w) = s(w)\ \operatorname{sinc}^2(\pi w D) \qquad (16)$$

where $\operatorname{sinc}(x) \coloneqq \sin(x)/x$. By sampling $\underline{y}(t)$ at intervals D, we find (for $w = \omega \,/\, D$):

$$s_{\mathrm{d}}^{(D)}(\omega) = \frac{s^{(D)}(w)}{D} = \frac{1}{D}\sum_{j=-\infty}^{\infty} s\left(\frac{\omega+j}{D}\right)\ \operatorname{sinc}^2\big(\pi(\omega + j)\big) = \frac{1}{D}\sum_{j=-\infty}^{\infty} s\left(w + \frac{j}{D}\right)\ \operatorname{sinc}^2\big(\pi(wD + j)\big) \qquad (17)$$

Relationship between climacogram and power spectrum

As both the climacogram and the power spectrum are transformations of the autocovariance function, the two are also related to each other by simple transformations. Thus, by virtue of (3), expressing $c(h)$ in terms of $\gamma(h)$ in (11) and using known properties of the Fourier transform, we find

$$s(w) = 2\,w^2 \frac{\mathrm{d}^2}{\mathrm{d}w^2} \int_0^\infty \gamma(h) \cos(2\pi w h)\,\mathrm{d}h \tag{18}$$

which after algebraic manipulations becomes

$$s(w) = -2 \int_0^\infty (2\pi w h)^2 \gamma(h) \cos(2\pi w h)\,\mathrm{d}h \tag{19}$$

On the other hand, combining (2) and (12) we find

$$\gamma(k) = 2 \int_0^1 (1-\xi) \int_0^\infty s(w) \cos(2\pi w \xi k)\,\mathrm{d}w\,\mathrm{d}\xi = 2 \int_0^\infty s(w) \int_0^1 (1-\xi) \cos(2\pi w \xi k)\,\mathrm{d}\xi\,\mathrm{d}w \tag{20}$$

and after algebraic manipulations, we find the following equation giving directly the climacogram from the power spectrum (which could be also derived by combining (12) and (16) for $h = 0$):

$$\gamma(k) = \int_0^\infty s(w)\ \mathrm{sinc}^2(\pi w k)\,\mathrm{d}w \tag{21}$$

The autocorrelation function

The *autocorrelation function* is the autocovariance function standardized by variance, i.e.,

$$\rho(h) \coloneqq \frac{c(h)}{\gamma_0}, \qquad \rho_j^{(D)} \coloneqq \frac{c_j^{(D)}}{\gamma(D)}, \qquad \check{\rho}_j \coloneqq \frac{\check{c}_j}{\gamma_0} \tag{22}$$

for the continuous-time, the averaged and the sampled process, respectively. All these autocorrelation functions are bounded from below and above, ranging in [−1, 1].

In some processes (including the white noise and the Hurst-Kolmogorov process) the instantaneous variance $\gamma_0 = \infty$, while $c(h) < \infty$ for $h \neq 0$, so that $\rho(h) = \check{\rho}_j = 0$ for any lag h or j (except $h = j = 0$, in which $\rho(0) = \check{\rho}_0 = 1$). However, if the processes are ergodic (and normally they should be as explained in p. 5), in the averaged process the variance is finite, so that generally $\rho_j^{(D)} \neq 0$.

The structure function and the climacogram-based structure function

The *structure function*, also known as *semivariogram* or *variogram*, is defined to be

$$v(t,u) \coloneqq \frac{1}{2}\mathrm{Var}\big[\underline{x}(t) - \underline{x}(u)\big] \equiv \frac{1}{2}\Big(\mathrm{Var}\big[\underline{x}(t)\big] + \mathrm{Var}\big[\underline{x}(u)\big]\Big) - \mathrm{Cov}\big[\underline{x}(t), \underline{x}(u)\big] \tag{23}$$

In a stationary process, for $h \coloneqq u - t$,

$$v(h) \coloneqq \frac{1}{2}\mathrm{Var}\big[\underline{x}(t) - \underline{x}(t+h)\big] \equiv \mathrm{Var}\big[\underline{x}(t)\big] - \mathrm{Cov}\big[\underline{x}(t), \underline{x}(t+h)\big] \tag{24}$$

which in the various processes examined here takes the form

$$v(h) \coloneqq \gamma_0 - c(h),\ h_j^{(D)} \coloneqq \gamma(D) - c_j^{(D)},\ \check{h}_j \coloneqq \gamma_0 - \check{c}_j \tag{25}$$

By analogy, we define the climacogram-based structure function CSF:

$$\xi(k) \coloneqq \gamma_0 - \gamma(k),\ \ \xi_\kappa^{(D)} \coloneqq \gamma(D) - \gamma(\kappa D),\ \ \check{\xi}^{(\kappa)} \coloneqq \gamma_0 - \check{\gamma}^{(\kappa)} \tag{26}$$

The CSF, ξ, is related to the structure function, v, by the same way as the climacogram, γ, is related to the autocovariance function, c:

$$c(h) = \frac{1}{2}\,\frac{\mathrm{d}^2\big(h^2\gamma(h)\big)}{\mathrm{d}h^2},\ v(h) = \frac{1}{2}\,\frac{\mathrm{d}^2\big(h^2\xi(h)\big)}{\mathrm{d}h^2} \tag{27}$$

By nondimensionalizing $v(h)$ or $\xi(k)$ (by dividing by γ_0) we get the *nondimensionalized structure function*, which is the complement of ρ from 1, and the *nondimensionalized climacogram-based structure function* (NCSF), i.e.,

$$1 - \rho(h),\ \ 1 - \rho_j^{(D)},\ \ 1 - \check{\rho}_j,\ \ 1 - \gamma(k)/\gamma_0,\ \ 1 - \gamma(\kappa D)/\gamma(D),\ \ 1 - \check{\gamma}^{(\kappa)}/\gamma_0$$

Autocorrelation function and structure function in processes with infinite variance

In cases where $\gamma_0 = \infty$, the structure function is $v(h) = \infty$, while $\rho(h) = 0$ and $1 - \rho(h) = 1$ for any lag $h \neq 0$. This is an improper behaviour.

However, assuming that the process is ergodic, such problems do not appear in the averaged process, in which we will have:

$$0 < \gamma(k) < \infty, \quad 0 \leq 1 - \rho_j^{(D)} \leq 2$$

The geostatistics literature includes the so-called *intrinsic* models in which $\gamma_0 = \infty$ and $c(h) = \infty$, while $v(h) = \gamma_0 - c(h) < \infty$ (e.g. $v(h) = a\, h^b$), so that $1 - \rho(h) = 0$ for any $h \neq \infty$. In such models the infinities are transferred even in the averaged process:

$$\gamma(D) = c_j^{(D)} = \infty, \; 1 - \rho_j^{(D)} = 1 - \check{\rho}_j = 0$$

In particular, the property $\gamma(D) = \infty$ means that the process is non-ergodic (in ergodic processes $\gamma(D)$ should tend to zero as D tends to infinity; see p. 17). For these reasons, as explained in p. 4, such processes are not recommended in modelling. There are much better alternatives (see examples below).

The climacogram-based spectrum (CS)

While the power spectrum is a magnificent tool for stochastic processes, its estimation from data is problematic. An empirical spectrum (periodogram) constructed from empirical autocovariances is too rough. Techniques to smooth the power spectrum need too many data to be applied, and they may also involve uncontrolled biases.

A substitute of the power spectrum which has similarities in its properties, is the climacogram-based spectrum (CS) defined as

$$\psi(w) \ := \frac{2}{w\gamma_0}\gamma(1/w)\xi(1/w) = \frac{2\,\gamma(1/w)}{w}\left(1 - \frac{\gamma(1/w)}{\gamma(0)}\right) \tag{28}$$

In contrast to the empirical periodogram, the empirical $\psi(w)$ is pretty smooth. Its value at $w = 0$ equals that of the power spectrum; indeed from (1), (2), (11) we obtain $\psi(0) = s(0) = k\,\gamma(k)|_{k\to\infty}$ $= 4\int_0^\infty c(h)\mathrm{d}h$.

At frequencies a where the power spectrum has peaks, signalling periodicities, the CS has smooth troughs (negative peaks; see example in p. 43), while it has discontinuities at frequencies $(m + ½)a$, where m is any integer.

The CS combines the climacogram and the CSF; it is valid for both finite and infinite variance. In processes with infinite variance ($\gamma(0) = c(0) = \infty$) the CS simplifies to

$$\psi(w) = \frac{2\,\gamma(1/w)}{w} \tag{29}$$

Some important stochastic properties

The climacogram $\gamma(k)$ denotes variance and therefore should be nonnegative for any time scale k. It is be positive and finite or even infinite for $k = 0$. For (mean) ergodic processes it should necessary tend to 0 for $k \to \infty$ (Papoulis, 1991, p. 429). Thus,

$$\gamma(k) > 0, \qquad \gamma(\infty) = 0 \tag{30}$$

As the autocovariance $c(h)$ equals the variance for $h = 0$, it follows that $c(0) > 0$. For $h \neq 0$, $c(h)$ can take on negative values as well. However, $c(h)$ must be a positive definite function (see e.g. Stewart, 1976), a property which, among other things, makes it bounded from below and above by $\pm\, c(0)$. Ergodicity also imposes a constraint about its asymptotic behaviour (Papoulis, 1991, p. 430); in conclusion:

$$c(0) > 0, \ \ |c(h)| \leq c(0), \qquad \frac{1}{k}\int_0^k c(h)\,\mathrm{d}h \xrightarrow[k\to\infty]{} 0 \tag{31}$$

In order for the function $c(h)$ to be positive definite, its Fourier transform, i.e. the power spectrum $s(w)$ should be nonnegative. Thus,

$$s(w) \geq 0 \tag{32}$$

Additional properties of $s(w)$ are discussed in next section.

The autocovariance $c(h)$ is often a nonnegative and nonincreasing function. In this case $\gamma(k)$ is nonincreasing too. To see this, we take the derivative with respect to k from (2) and we find

$$\gamma'(k) = \tfrac{4}{k^3}\int_0^k (h - k/2)c(h)\mathrm{d}h \tag{33}$$

The term $(h - k/2)$ within the integral is symmetric with respect to $k/2$ (negative for $h < k/2$ and positive for $h > k/2$). As $c(h)$ is nonincreasing, its values for $h < h/2$ are greater than those for $h > k/2$. Clearly then the negative product prevails over the positive product and thus $\gamma'(k) < 0$.

Asymptotic power laws and the log-log derivative

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

Power laws are functions of the form

$$f(x) \propto x^b \tag{34}$$

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

$$f^{\#}(x) \coloneqq \frac{\mathrm{d}(\ln f(x))}{\mathrm{d}\,(\ln x)} = \frac{x f'(x)}{f(x)} \tag{35}$$

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

Asymptotic properties of the power spectrum

The asymptotic behaviour of the log-log derivative of the power spectrum $s(w)$ is quite important to characterize the process. Generally, this log-log derivative is

$$s^{\#}(w) \coloneqq \frac{\mathrm{d}(\ln s(w))}{\mathrm{d}\,(\ln w)} = \frac{w\, s'(w)}{s\,(w)} = \frac{a(w)}{s(w)} \tag{36}$$

where $a(w) \coloneqq ws'(w)$. We will find its asymptotic behaviour for $w \to 0$, i.e. $s^{\#}(0)$. Note that continuous time is assumed for the process as well as the spectrum. From (11), the derivative is

$$s'(w) = -4 \int_0^\infty 2\pi h\, c(h) \sin(2\pi w h)\, \mathrm{d}h \tag{37}$$

We define:

$$A(w) \coloneqq a(w) + s(w) = 4 \int_0^\infty c(h)\,(\cos(2\pi w h) - 2\pi w h \sin(2\pi w h))\,\mathrm{d}h \tag{38}$$

and we observe that

$$A(0) = 4 \int_0^\infty c(h)\,\mathrm{d}h = s(0) \tag{39}$$

In other words

$$a(0) + s(0) = s(0) \tag{40}$$

If $0 < s(0) < \infty$ then (40) simplifies to

$$a(0) = 0 \tag{41}$$

and hence

$$0 < s(0) < \infty \Rightarrow s^{\#}(0) = 0 \tag{42}$$

Asymptotic properties of the power spectrum (contd.)

If $s(0) = 0$ then (41) is still valid, but the log-log derivative $s^{\#}(0)$ becomes an indeterminate quantity (0/0). This should necessarily be positive, so that $s^{\#}(w)$>0 for w >0. Thus,

$$s\ (0) = 0 \Rightarrow s^{\#}(0) > 0 \tag{43}$$

If $s(0) = \infty$ (hence $s'(0) = -\infty$, $a(0) < 0$, while $s(w)$ is continuous), then (because $s(w)$ is nonnegative) (40) results in

$$\lim_{w\to 0}\bigl(a(w) + s(w)\bigr) > 0 \tag{44}$$

and

$$\lim_{w\to 0}\left(\frac{a(w)}{s(w)} + 1\right) > 0 \tag{45}$$

Consequently,

$$s(0) = \infty \Rightarrow s^{\#}(0) > -1 \tag{46}$$

In conclusion, the asymptotic log-log derivative of the power spectrum for $w \to 0$ can never be lower (steeper) than –1 and more specifically it ranges as follows:

$$\begin{array}{ll} s^{\#}(0) > 0, & s(0) = 0 \\ s^{\#}(0) = 0, & 0 < s(0) < \infty \\ -1 < s^{\#}(0) < 0, & s(0) = \infty \end{array} \tag{47}$$

The asymptotic log-log derivative for $w \to \infty$ should necessarily be non-positive, without other restrictions, i.e.:

$$s^{\#}(\infty) \leq 0 \tag{48}$$

A comment on the asymptotic slope of the power spectrum

We often see publications reporting logarithmic slopes in empirical power spectra $s^\# < -1$ (e.g. $s^\# = -1.5$, etc.). Does this mean that (46) is wrong or that the reality does not comply with this theoretical property?

First we should point out that a slope $s^\#(w) < -1$ is mathematically and physically possible for large w. However it is infeasible for $w \to 0$. Therefore, reported values $s^\# < -1$ for small w are spurious and are due to inadequate sample size or inconsistent estimation algorithms (stemming from the fact that the periodogram constructed from empirical autocovariances is too rough and the estimation of slopes from this is too uncertain; cf. Koutsoyiannis, 2013a; Lombardo et al., 2013; Dimitriadis and Koutsoyiannis, 2015). Such results do not put into question the validity of (46) but are just invalid results.

For additional support of this argument, let us assume the contrary, i.e., that for frequency range $0 \le w \le \varepsilon$ (with ε however small) the log-log derivative is $s^\#(w) = \beta$, or else $s(w) = \alpha\, w^\beta$ where α and β are constants, with $\beta < -1$. We notice in (21) that the fraction within the integral takes significant values only for $w < 1/k$ (cf. Papoulis, 1991, p. 433). Hence, assuming a scale $k \gg 1/\varepsilon$, and with reference to (21) we may write:

$$\gamma(k) = \int_0^\infty s(w)\,\mathrm{sinc}^2(\pi w k)\,\mathrm{d}w \approx \int_0^\varepsilon \alpha w^\beta\ \mathrm{sinc}^2(\pi w k)\,\mathrm{d}w \tag{49}$$

On the other hand, it is easy to verify that, for $0 < w < 1/D$,

$$\mathrm{sinc}(\pi w k) \ge 1 - wk \ge 0 \tag{50}$$

and since $\varepsilon \gg 1/k$, while the function in the integral (49) is nonnegative,

$$\gamma(k) \approx \int_0^\varepsilon \alpha w^\beta\ \mathrm{sinc}^2(\pi w k)\,\mathrm{d}w \ge \int_0^{1/k} \alpha w^\beta\ \mathrm{sinc}^2(\pi w k)\,\mathrm{d}w \ge \int_0^{1/k} \alpha w^\beta (1 - wk)^2\,\mathrm{d}w \tag{51}$$

Comment on the asymptotic slope of the power spectrum (contd.)

Substituting $\omega = wk$ in (51), we find:

$$\gamma(k) \geq ak^{-\beta-1} \int_0^1 \omega^\beta (1-\omega)^2 \, \mathrm{d}\omega \tag{52}$$

To evaluate the integral in (52) we take the limit for $q \to \infty$ of the integral:

$$B(q) \coloneqq \int_{1/q}^1 \omega^\beta (1-\omega)^2 \, \mathrm{d}\omega \; = \frac{1-q^{-1-\beta}}{1+\beta} - 2\,\frac{1-q^{-2-\beta}}{2+\beta} + \frac{1-q^{-3-\beta}}{3+\beta} \tag{53}$$

Clearly, the limit of $B(q)$ as $q \to \infty$ depends on that of the term with the highest exponent, i.e. $q^{-1-\beta}$. For $\beta <$ –1 this term diverges and thus, $B(0) = +\infty$. Thus, by virtue of the inequality (52), $\gamma(D) = \infty$. Therefore, the process is non-ergodic (see p. 17).* This analysis generalizes a result by Papoulis (1991, p. 434) who shows that an impulse at $w = 0$ corresponds to a non-ergodic process.

In a non-ergodic process there is no possibility to infer statistical properties from the samples, so the statistical analyses are in vain and hence the reported results not meaningful (see also p. 5).

Sometimes reported slopes $s^\# < -1$ are interpreted as indications of nonstationarity. Such interpretations are equally invalid because even the definition of the power spectrum as a function of frequency only (as well as those of autocorrelation and climacogram as functions of lag and scale, respectively) assumes stationarity.

---

* It is interesting to note that, if $|\beta| < 1$, the integral in (49) can be evaluated to give:

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

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

Asymptotic properties of the climacogram, CSF and CS

The asymptotic behaviour of the climacogram is the same with that of the autocovariance function (under some general conditions and with the exception where $\gamma^{\#}(\infty) = -1$ or -2^{*}).

The asymptotic behaviour of the CSF is the same with that of the structure function (under similar conditions and with the exception where $\xi^{\#}(0) = 1$ or 2; the proof is similar to that below, by virtue of (27)).

The asymptotic behaviour of the CS is the same with that of the power spectrum (under some general conditions and with some exceptions not fully investigated yet; cf. Stein, 1999). In most processes the log-log derivatives $\psi^{\#}(w)$ of CS at frequencies (or resolutions) $w \to 0$ and $w \to \infty$ are precisely equal to those of the power spectrum $s^{\#}(w)$, as illustrated in the examples below (pp. 33-44).

The logarithmic slope $\psi^{\#}(w)$ of CS is connected with that of the climacogram and the CSF by:

$$\psi^{\#}(w) = -1 - \left(2 - \frac{1}{1-\gamma(1/w)/\gamma(0)}\right)\gamma^{\#}(1/w) \tag{54}$$

$$\psi^{\#}(w) = -1 - \left(2 - \frac{1}{1-\xi(1/w)/\gamma(0)}\right)\xi^{\#}(1/w) \tag{55}$$

From (54), if $\gamma(0) = \infty$, then $\psi^{\#}(w) = -1 - \gamma^{\#}(1/w)$. Also, for $w = 0$, $\psi^{\#}(0) = -1 - \gamma^{\#}(\infty)$. Since $\gamma'(\infty) \leq 0$ (cf. p. 17) and hence $\gamma^{\#}(\infty) \leq 0$, it follows that $\psi^{\#}(0) \geq -1$ (same behaviour as in $s^{\#}(0)$).

When $0 < \gamma(0) < \infty$, $\gamma^{\#}(0) = 0$, so that in $\psi^{\#}(\infty)$ an expression $\infty \cdot 0$ appears in the right-hand side of (54). In that case, as $\xi(0) = 0$, (55) gives $\psi^{\#}(\infty) = -1 - \xi^{\#}(0)$.

———————

*__Proof__: We assume that $\gamma(k)$ has first and second derivative which $\to 0$ as $k \to \infty$. We use l'Hopital's rule to find:

$$\lim_{h\to\infty} h^b c(h) = \lim_{h\to\infty} \frac{c(h)}{h^{-b}} = \lim_{h\to\infty} \frac{1}{2}\frac{\mathrm{d}^2\left(h^2\gamma(h)\right)/\mathrm{d}h^2}{h^{-b}} = \lim_{h\to\infty}\left(\frac{\gamma(h)}{h^{-b}} + 2\frac{\gamma'(h)}{h^{-b-1}} + \frac{1}{2}\frac{\gamma''(h)}{h^{-b-2}}\right) = \frac{1}{2}(b-1)(b-2)\lim_{h\to\infty}\left(\frac{\gamma(h)}{h^{-b}}\right) = \frac{1}{2}(b-1)(b-2)\lim_{h\to\infty} h^b\gamma(h)$$

Unless $b = 1$ or $b = 2$, the limit $\lim_{h\to\infty} h^b c(h)$ is 0, finite or ∞, if and only if $\lim_{h\to\infty} h^b \gamma(h)$ is 0, finite or ∞, respectively.

Fractal parameter and Hurst parameter

A fractal parameter is a local property of a stochastic process characterizing its behaviour at small time scales. Specifically (Gneiting and Schlather, 2004), assuming that for small lags h or small time scales k the structure function or, respectively, the CSF is given as

$$v(h) \coloneqq \gamma_0 - c(h) \;\propto\; h^a \text{ or } \xi(k) \coloneqq \gamma_0 - \gamma(k) \;\propto\; k^a \tag{56}$$

The fractal parameter (M in honour of Mandelbrot) and the fractal dimension D_{F} are

$$M = \frac{\alpha}{2} = \frac{v^\#(0)}{2} = \frac{\xi^\#(0)}{2}, \quad D_{\mathrm{F}} = 2 - \frac{a}{2} = 2 - \frac{v^\#(0)}{2} = 2 - \frac{\xi^\#(0)}{2} \tag{57}$$

Conversely, the Hurst parameter is a global property of a stochastic process characterizing its behaviour at large time scales. Specifically (cf. Gneiting and Schlather, 2004), assuming that for large time scales k the climacogram is given as

$$\gamma(k) \;\propto\; k^{-b} \tag{58}$$

the Hurst parameter (H in honour of Hurst) is

$$H = 1 - \frac{b}{2} = 1 + \frac{\gamma^\#(\infty)}{2} \tag{59}$$

Typically, these slopes are reflected in the power spectrum, i.e. $s^\#(\infty) = -a - 1$, $s^\#(0) = b - 1$, so that

$$M = \frac{\xi^\#(0)}{2} = -\frac{1}{2} - \frac{s^\#(\infty)}{2} \tag{60}$$

$$H = 1 + \frac{\gamma^\#(\infty)}{2} = \frac{1}{2} - \frac{s^\#(0)}{2} \tag{61}$$

while in most cases the slopes of the spectrum $\psi^\#$ could be used instead of those of the spectrum $s^\#$.

Definition and importance of entropy

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

Entropy is a dimensionless measure of uncertainty defined as follows:

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

$$\Phi[\underline{z}] := \mathrm{E}[-\ln P(\underline{z})] = -\sum_{j=1}^{w} P_j \ln P_j \tag{62}$$

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

$$\Phi[\underline{z}] := \mathrm{E}\left[-\ln \frac{f(\underline{z})}{m(\underline{z})}\right] = -\int_{-\infty}^{\infty} \ln \frac{f(z)}{m(z)} f(z)\,\mathrm{d}z \tag{63}$$

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

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

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

Entropy production in stochastic processes

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

$$\Phi'[\underline{X}(t)] := \mathrm{d}\Phi[\underline{X}(t)]/\mathrm{d}t \tag{64}$$

A more convenient (and dimensionless) measure is the entropy production (i.e. the derivative) in logarithmic time (EPLT):

$$\varphi(t) \equiv \varphi[\underline{X}(t)] := \Phi'[\underline{X}(t)]\, t \equiv \mathrm{d}\Phi[\underline{X}(t)] \,/\, \mathrm{d}(\ln t) \tag{65}$$

For a Gaussian process, the entropy depends on its variance $\Gamma(t)$ only and is given as (cf. Papoulis, 1991):

$$\Phi[\underline{X}(t)] = (1/2)\ln(2\pi \mathrm{e}\, \Gamma(t)/m^2) \tag{66}$$

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

$$\varphi(t) = \Gamma'(t)\, t \,/\, 2\Gamma(t) \tag{67}$$

When the past and the present are observed, instead of the unconditional variance $\Gamma(t)$ we should use a variance $\Gamma_{\mathrm{C}}(t)$ conditional on the known past and present. This turns out to be:

$$\Gamma_{\mathrm{C}}(t) \approx 2\Gamma(t) - \Gamma(2t)/2 \tag{68}$$

The climacogram and the climacogram-based metrics compared to more standard metrics

- In stochastic processes, almost all classical statistical estimators are biased and uncertain; in processes with LTP bias and uncertainty are very high.
- In the climacogram (variance), bias and uncertainty are easy to control as they can be calculated analytically (and a priori known; see below).
- The autocovariance function is the second derivative of the climacogram.
    - Estimation of the second derivative from data is too uncertain and makes a very rough graph.
    - Estimation of autocovariance is too biased in processes with LTP.
- The power spectrum is the Fourier transform of the autocovariance and entails an even rougher shape and more uncertain estimation than in the autocovariance (see also Dimitriadis and Koutsoyiannis, 2015).
- An additional advantage of the climacogram is its close relationship with EPLT. Specifically, combining equations (67), (2) and (35) we conclude that for Gaussian processes the EPLT is

$$\varphi(t) = 1 + \tfrac{1}{2}\, \gamma^{\#}(t)$$

- This entails that the Hurst parameter H equals the global EPLT, $\varphi(\infty)$.

Statistical estimation—Averaged process

The following analysis is based on Koutsoyiannis (2011a, 2016). We assume that we have $n = \lfloor T_o/k \rfloor = T/k$ observations of the averaged process at scale $k = \kappa D$, $\underline{x}_i^{(k)}$, where T_o is the observation period, $\lfloor\ \rfloor$ denotes the floor of a real number, and $T \coloneqq \lfloor T_o/k \rfloor k$ is the observation period rounded off to an integer multiple of k. The (unbiased) estimator of the common mean μ of the instantaneous process $\underline{x}(t)$ as well as of the discrete process $\underline{x}_i^{(k)}$ is

$$\underline{\bar{x}}^{(k)} \coloneqq \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i^{(D)} = \frac{\underline{X}(T)}{nD} = \frac{\underline{X}(T)}{T} = \underline{x}_1^{(T)} \tag{69}$$

The standard (but biased except for white noise) estimator $\underline{\hat{\gamma}}(k)$ of the variance $\gamma(k)$ of the averaged process $\underline{x}_i^{(k)}$ is

$$\underline{\hat{\gamma}}(k) \coloneqq \frac{1}{n-1}\sum_{i=1}^{n}\left(\underline{x}_i^{(k)} - \underline{\bar{x}}^{(k)}\right)^2 = \frac{1}{T/D-1}\sum_{i=1}^{T/D}\left(\underline{x}_i^{(k)} - \underline{\bar{x}}^{(k)}\right)^2 \tag{70}$$

The following general equation can be used to estimate the bias of $\underline{\hat{\gamma}}(D)$

$$\mathrm{E}\left[\underline{\hat{\gamma}}(k)\right] = \frac{1}{1-1/n}\left(\mathrm{Var}\left[\underline{x}_i^{(k)}\right] - \mathrm{Var}\left[\underline{\bar{x}}^{(k)}\right]\right) \tag{71}$$

Statistical estimation—Averaged process (contd.)

The general equation (71) in this case becomes

$$\mathrm{E}\left[\hat{\underline{\gamma}}(k)\right] = \frac{1}{1 - k/T}\,(\gamma(k) - \gamma(T)) = \frac{1}{1 - k/T}\left(\frac{\Gamma(k)}{k^2} - \frac{\Gamma(T)}{T^2}\right) \tag{72}$$

or

$$\mathrm{E}\left[\hat{\underline{\gamma}}(k)\right] = \eta(k, T)\gamma(k) \tag{73}$$

where the bias correction coefficient η is

$$\eta(k, T) = \frac{1 - \gamma(T)/\gamma(k)}{1 - k/T} = \frac{1 - (k/T)^2\Gamma(T)/\Gamma(k)}{1 - k/T} \tag{74}$$

From (72) and (73) it can be verified that both $\hat{\underline{\gamma}}_1(k)$ and $\hat{\underline{\gamma}}_2(k)$ defined in (75) below are unbiased estimators of $\gamma(k)$ (see also Dimitriadis and Koutsoyiannis, 2017):

$$\hat{\underline{\gamma}}_1(k) \coloneqq \hat{\underline{\gamma}}(k)/\eta(k, T), \qquad \hat{\underline{\gamma}}_2(k) \coloneqq (1 - k/T)\hat{\underline{\gamma}}(k) + \gamma(T) \tag{75}$$

Important note: It becomes clear from the above equations that direct estimation of the variance $\gamma(k)$ for any time scale k, let alone that of the instantaneous process γ_0, is not possible merely from the data. We need to know the ratio $\gamma(T)$ / $\gamma(k)$ and thus we should assume a stochastic model which evidently influences the estimation of $\gamma(k)$. Once the model is assumed and its parameters estimated based on the data, we can expand our calculations to estimate the variance for any time scale, including that of the instantaneous scale γ_0.

Statistical estimation—Sampled process

If the observed process is the sampled ($\underline{x}_i$) at time steps D, rather than the averaged one ($\underline{x}_i^{(D)}$), then (69) becomes

$$\underline{\bar{x}} := \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i = \underline{\breve{x}}_1^{(n)} \tag{76}$$

The standard (and again biased) estimator $\underline{\hat{\breve{\gamma}}}^{(k)}$ of the variance $\breve{\gamma}^{(k)}$ of the sampled process is

$$\underline{\hat{\breve{\gamma}}}^{(\kappa)} := \frac{1}{n-1}\sum_{i=1}^{n}\left(\underline{\breve{x}}_i^{(\kappa)} - \underline{\bar{x}}\right)^2 \tag{77}$$

In a similar manner it is shown that (73) and (74) become respectively

$$\mathrm{E}\left[\,\underline{\hat{\breve{\gamma}}}^{(\kappa)}\right] = \eta'(D, n, \kappa)\breve{\gamma}^{(\kappa)} \tag{78}$$

$$\eta'(D, n, \kappa) = \frac{1 - \breve{\gamma}^{(n)}/\breve{\gamma}^{(\kappa)}}{1 - 1/n} \approx \frac{1 - \gamma(T)/\gamma(\kappa D)}{1 - \kappa D\ /\ T} \tag{79}$$

Important note: Even if our target is to estimate the variance of the instantaneous process, $\breve{\gamma}^{(1)} = \gamma(0) = \gamma_0$, again this is not possible merely from the data. We need to know the ratio $\breve{\gamma}^{(n)}/\gamma_0$ and thus we should again, as above, assume a stochastic model. Once the model is assumed and its parameters estimated, we can expand our calculations to estimate $\breve{\gamma}^{(\kappa)}$ for any multiple κ of the time step D.

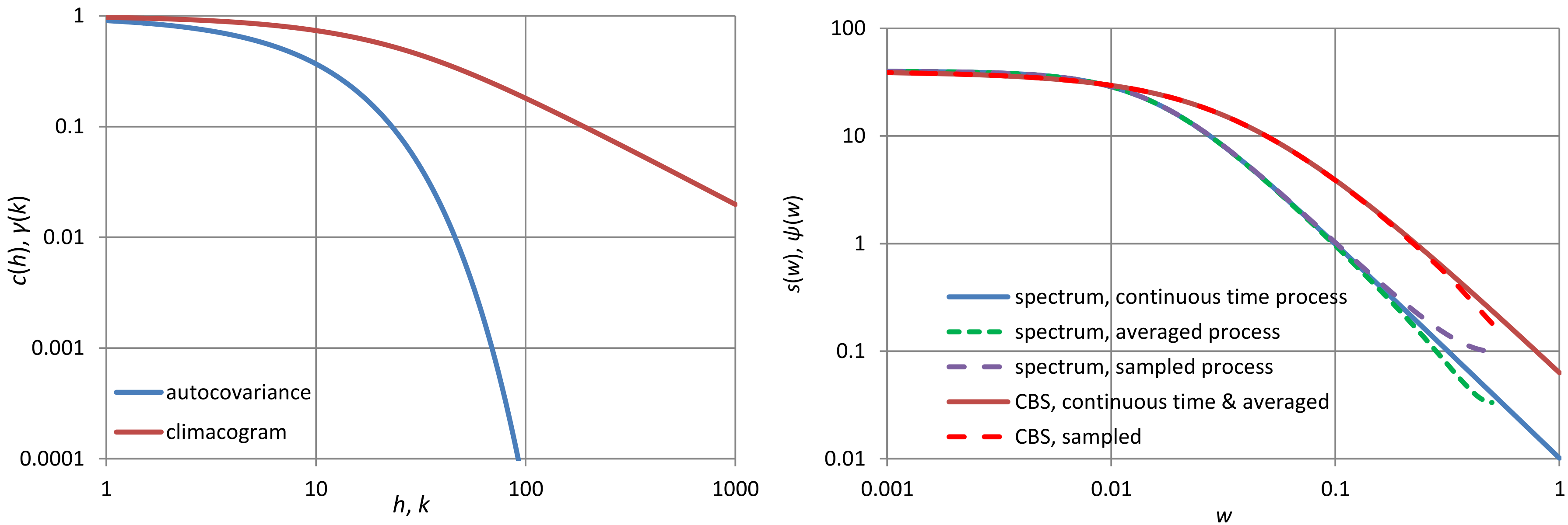
Example 1: The Markov process

Type	Continuous time	Discrete time, averaged	Discrete time, sampled
Variance of instantaneous process	$\gamma_0 = \gamma(0) = c(0) = \lambda$	Not applicable	$\check{\gamma}_0 = \gamma_0 = \gamma(0) = \lambda$
Climacogram (Variance at scale $k = \kappa D$)	$\gamma(k) = \frac{2\lambda}{k/a}\left(1 - \frac{1-e^{-k/a}}{k/a}\right)$	As in continuous time	$\check{\gamma}^{(\kappa)} = \frac{\lambda}{\kappa\,(1-\rho)^2}\left(1-\rho^2 - \frac{2\rho(1-\rho^\kappa)}{\kappa}\right), \kappa = 1, 2, \dots$ where $\rho = e^{-D/a}$
Autocovariance function for lag $h = jD$	$c(h) = \lambda \mathrm{e}^{-h/a}$	$c_j^{(D)} = \frac{\lambda\left(1-e^{-D/a}\right)^2}{(D/a)^2}\ \mathrm{e}^{-(j-1)D/a}$, $j = 1, 2, \dots$	$\check{c}_j = \lambda \mathrm{e}^{-jD/a} = \lambda\rho^j$, $j = 1, 2, \dots$
Power spectrum for frequency $w = \omega/D$	$s(w) = \frac{4a\lambda}{1+(2\pi a w)^2}$	$s^{(D)}(w) = 4a\lambda\left(1 - \frac{1}{D/a}\frac{(1-\cos(2\pi D w))\sinh(D/a)}{\cosh(D/a)-\cos(2\pi D w)}\right)$	$\check{c}_j(w) = \frac{2\lambda D \sinh(D/a)}{\cosh(D/a)-\cos(2\pi D w)}$
Asymptotic slopes, global	$\gamma^{\#}(\infty) = 2\sigma^{\#}(\infty) = -1$ ($c^{\#}(\infty) = -\infty$)* $\psi^{\#}(0) = s^{\#}(0) = 0$		
Asymptotic slopes, local	$\gamma^{\#}(0) = \sigma^{\#}(0) = 0$ $g^{\#}(0) = h^{\#}(0) = 1$ $\psi^{\#}(\infty) = s^{\#}(\infty) = -2$		

* With respect to the observations on p. 24, it is seen that the Markov process belongs to the exceptions because $\gamma^{\#}(\infty) = -1$; this explains why $c^{\#}(\infty) = -\infty$.

Note: It is readily verifiable that the sampled discrete-time process is equivalent to the AR(1) process (notice that $c_j^* = \lambda\rho^j$) while the averaged discrete-time process is not an AR(1) but an ARMA(1,1) process.

Example 1: The Markov process (contd.)



Notes

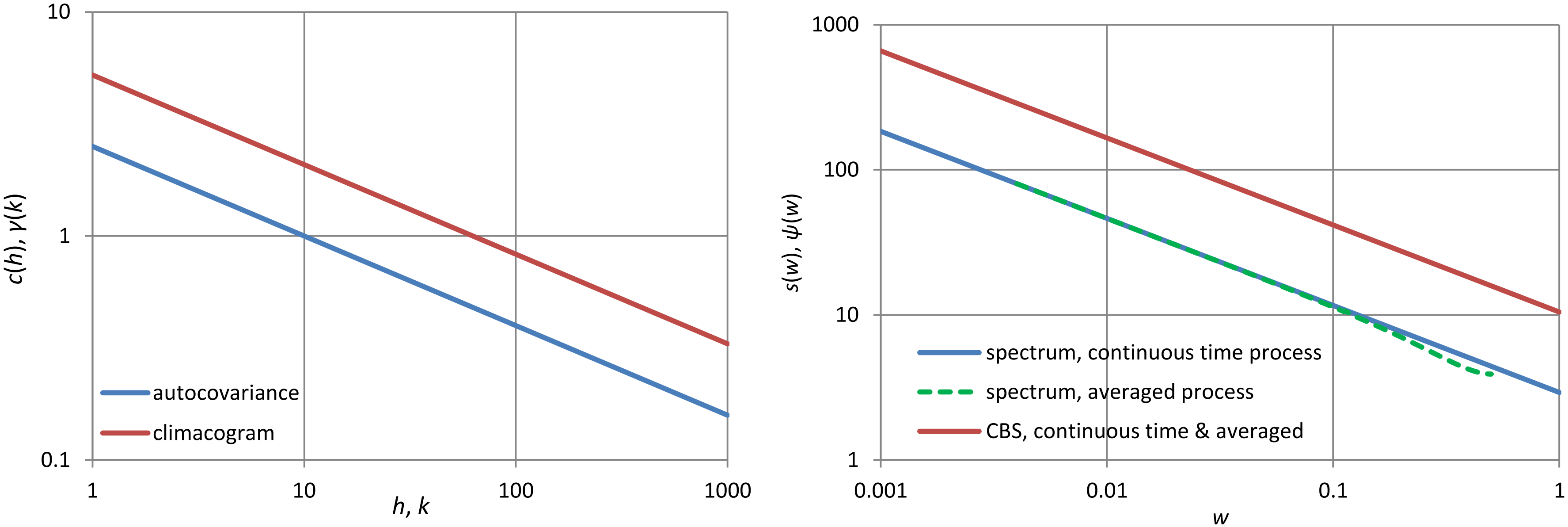
1. The parameters are $\lambda = 1$, $a = 10$ and the spacing is $D = 1$, resulting in $\rho = 0.905$.
2. The plotted autocovariance function and climacogram are for the continuous-time process, whereas those of the discretized processes are visually indistinguishable.
3. The power spectra of the discretized processes fail to capture the slopes for $w > 1 / 10D$, while for $w = 1 / 2D$ they give a slope which is precisely zero (this is a consequence of the fact that the discrete-time spectrum is symmetric around $w = 1 / 2D$).
4. The CS performs better in identifying the asymptotic slopes (even in the worst case of the sampled process).

Example 2: The Hurst-Kolmogorov (HK) process

Type	Continuous time	Discrete time, averaged	Discrete time, sampled
Variance of instantaneous process	$\gamma_0 = \gamma(0) = c(0) = +\infty$	Not applicable	$\check{\gamma}_0 = \gamma_0 = \gamma(0) = \infty$
Climacogram (Variance at scale $k = \kappa D$)	$\gamma(k) = \lambda(a/k)^{2-2H}$	As in continuous time	$\check{\gamma}^{(\kappa)} = +\infty$
Autocovariance function for lag $h = jD$	$H > ½$: $c(h) = \lambda H(2H-1)(a/h)^{2-2H}$ $H = ½$: $c(h) = \lambda\, \delta(h/a)$ $H < ½$: $c(h) = \lambda H(2H-1)\left((a/h)^{2-2H} + \delta''(h/a)\right)$	$c_j^{(D)} = \lambda(a/D)^{2-2H} \times \left(\frac{\lvert j-1\rvert^{2H}+\lvert j+1\rvert^{2H}}{2} - \lvert j\rvert^{2H}\right)$	$\check{c}_j = c(jD), j = 1, 2, \ldots$
Power spectrum for frequency $w = \omega/D$	$s(w) = \frac{4a\lambda\, H(2H-1)\Gamma(2H-1)\sin(\pi H)}{(2\pi a w)^{2H-1}}$	$s^{(D)}(w)$: not a closed expression	$\check{s}(w) = +\infty$
Slopes (constant for any w and D)	$\gamma^{\#}(D) = 2\sigma^{\#}(D) = 2H - 2$ $\psi^{\#}(w) = s^{\#}(w) = 1 - 2H$		

Notes: (a) $\Gamma(\,)$ is the gamma function, $\delta(\,)$ is the Dirac delta function and $\delta''(\,)$ is its second derivative.
(b) While the power spectra of the continuous-time and the averaged process converge except for $w = 0$, that of the sampled discrete-time process diverges, because its variance is ∞ for any w. To see that it diverges, we apply the definition of the discrete-time power spectrum (like in eq. (13)) for nondimensionalized frequency $\omega = 1/4$. The first term is ∞ and in order for $\check{s}(w)$ to converge the second term (the sum) must be $-\infty$. This sum can be written as a quantity proportional to $4^{2H-2} + 8^{2H-2} + \cdots - (2^{2H-2} + 6^{2H-2} + \cdots)$, which equals $2^{2H}(2^{2H-1} - 1)\,\zeta(2-2H)$, where $\zeta()$ is the zeta function. This however is a finite quantity and thus $\check{s}(w) = \infty$.

Example 2: The Hurst-Kolmogorov (HK) process (contd.)



Notes

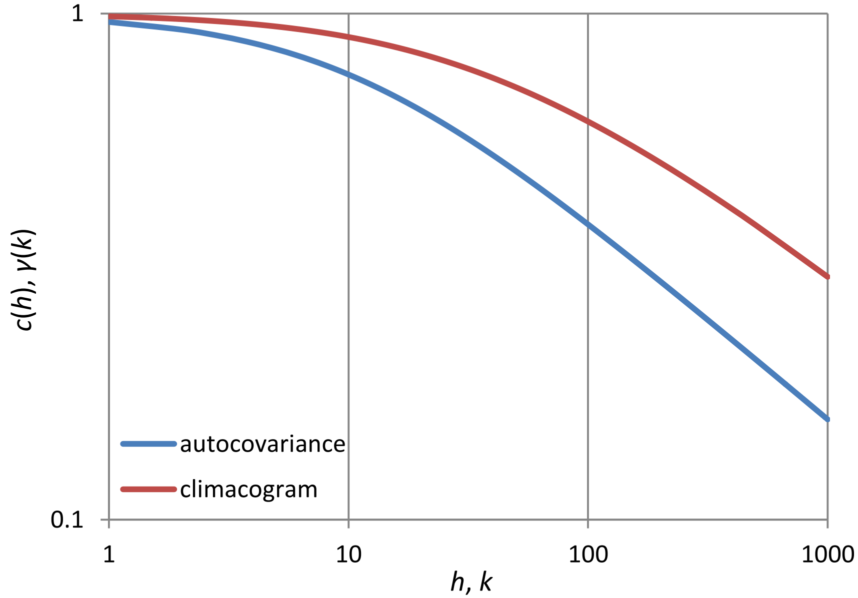
1. The parameters are $\lambda = 25/12$, $a = 10$, $H = 0.8$ (so that $\lambda H(2H - 1) = 1$)and the spacing is $D = 1$. Notes 2-4 of Example 1 are also valid in this example.
2. Notice that the model parameters are in essence two, i.e. H and ($\lambda\, a^{2-2H}$). Despite that, we prefer the formulation shown with three nominal parameters for dimensional consistency: the units of a and λ are [t] and [x^2], respectively, while H is dimensionless.
3. The process is most often used with $½ \le H < 1$, being called in this case a persistent process, while for $0 < H < ½$ it is called an antipersistent process. In the latter case, the autocovariance $c(h)$ is negative for any lag $h > 0$ and tends to $-\infty$ as $h \to 0$, having a discontinuity at 0 so that $c(0) = +\infty$. Consequently, the averaged process has positive variance and all covariances negative. Such a process is not physically realistic as for small h, $c(h)$ should be positive. More realistic antipersistent process is discussed in other examples below.
4. Clearly, the case $H > 1$ is invalid for reasons explained in pp. 19-22.

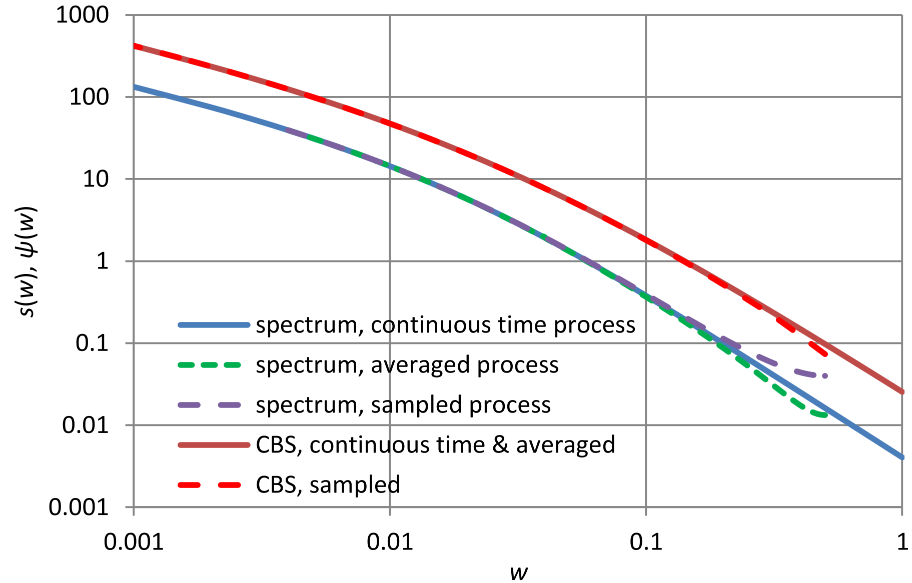
Example 3: A modified finite-variance HK process

Type	Continuous time	Discrete time, averaged	Discrete time, sampled
Variance of instantaneous process	$\gamma_0 = \gamma(0) = c(0) = \lambda$	Not applicable	$\check{\gamma}_0 = \gamma_0 = \gamma(0) = \lambda$
Climacogram (Variance at scale $k = \kappa D$)	$\gamma(k) = \lambda(a / k) \times \left(\frac{a}{k}\left(1 + \frac{a}{k}\right)^{2H} - \frac{a}{k} - 2H\right)$	As in continuous time	$\check{\gamma}^{(\kappa)} = \frac{\check{\Gamma}^{(\kappa)}}{\kappa^2}$, where $\check{\Gamma}^{(\kappa)}$ is calculated from (7)
Autocovariance function for lag $h = jD$	$c(h) = \frac{\lambda H(2H-1)}{(1+h/a)^{2-2H}}$	$c_j^{(D)}$: From (8)	$\check{c}_j = \frac{\lambda H(2H-1)}{(1+jD/a)^{2-2H}}, j = 1, 2, \ldots$
Power spectrum for frequency $w = \omega/D$	$s(w)$: closed expression too complex	$s^{(D)}(w)$: not a closed expression	$\check{s}(w)$: not a closed expression
Asymptotic slopes, global	$\gamma^{\#}(\infty) = 2\sigma^{\#}(\infty) = 2H - 2$ $\psi^{\#}(0) = s^{\#}(0) = 1 - 2H$		
Asymptotic slopes, local	$\gamma^{\#}(0) = \sigma^{\#}(0) = 0$ $\psi^{\#}(\infty) = s^{\#}(\infty) = -2$		

Note: In this example the asymptotic slopes of the power spectrum are both nonzero and different in the cases $w \to 0$ and $w \to \infty$. The slopes of the CS are identical with those of the spectrum.

Example 3: A modified finite-variance HK process (contd.)





Note: The parameters are $\lambda = 25/12$, $a = 10$, $H = 0.8$ (so that $\lambda H(2H - 1) = 1$) and the spacing is $D = 1$. Notes 2-4 of Example 1 are also valid in this example.

A practical conclusion: When slope estimations are based on data, it is safer to use frequencies or resolutions $w < 1\ /\ 10D$, or five times smaller than the absolute maximum that the observation period allows (in theory, this is $1\ /\ 2D = n\ /\ 2T$, where T is the observation period and n is the number of observations). Likewise, when estimating the climacogram and the CS, although in theory it is possible to calculate the variance for time scales up to $T/2 = (n/2)D$, it is safer to use time scales 5 times smaller, i.e. up to $T/10 = (n/10)D$.

Example 4: The Hybrid Hurst-Kolmogorov process (HHK)

An interesting process is defined by a Cauchy-type climacogram (Koutsoyiannis, 2016):

$$\gamma(k) = \lambda\left(1 + (k/a)^{2M}\right)^{\frac{H-1}{M}} \tag{80}$$

The expressions of its characteristics are too complex, but the framework provided above suffices to evaluate numerically all of its properties based on (80). As an exception, the autocovariance in continuous time has a rather simple expression:

$$c(h) = \gamma(h)\frac{1 + (3H + 2MH - 2M - 1)(h/a)^{2M} + H(2H - 1)(h/a)^{4M}}{1 + 2(h/a)^{2M} + (h/a)^{4M}} \tag{81}$$

Furthermore, it has simple asymptotic properties. Namely, its global properties are

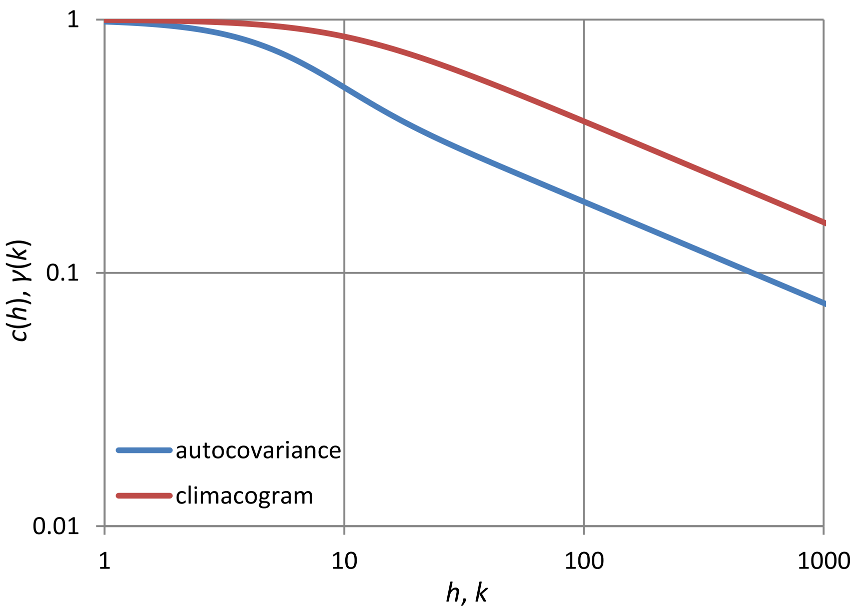
$$\gamma^{\#}(\infty) = c^{\#}(\infty) = 2\sigma^{\#}(\infty) = 2H - 2,\ \psi^{\#}(0) = s^{\#}(0) = 1 - 2H$$

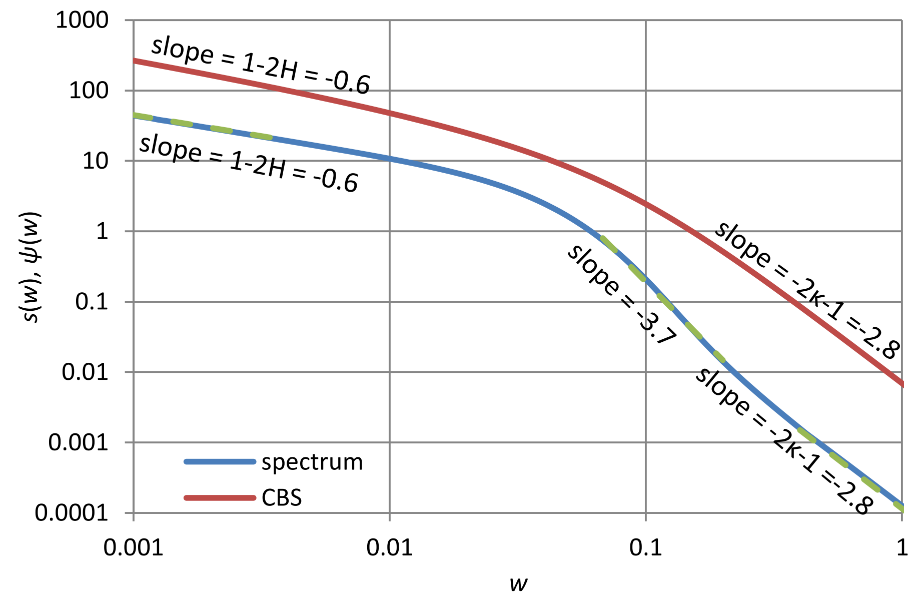
and its local properties are

$$g^{\#}(0) = h^{\#}(0) = 2M,\ \psi^{\#}(\infty) = s^{\#}(\infty) = -2M - 1,\ \gamma^{\#}(0) = \sigma^{\#}(0) = 0$$

An important feature of this process is that it allows explicit control of both asymptotic slopes as the global properties depend on the parameter H only, while the local properties depend on M only. It realistically represents both persistent and antipersistent processes; in the latter case ($0 < H < 0.5$), it has positive autocovariance for small lags and negative for large lags (in contrast to Example 3, where autocovariances are either all positive or all negative). The name “hybrid” for the process reflects the fact that it incorporates both the Markov and the HK processes. Specifically, in the special case with $H = M = 0.5$, HHK is practically indistinguishable from a Markov process (even though not precisely identical). Furthermore, as $a \to 0$, the process tends to a pure HK process with the same Hurst parameter H.

Example 4: The Hybrid Hurst-Kolmogorov process (contd.)





Notes

1. The parameters are λ = 1, a = 10, H = 0.8, M = 0.9 (only the continuous-time properties are depicted).
2. For small time scales HHK exhibits behaviour similar to Markov or, for M = 0.5 the process is practically indistinguishable from Markov. For large time scales it exhibits Hurst behaviour.
3. An intermediate steep slope that appears in the power spectrum is artificial and does not indicate a scaling behaviour.

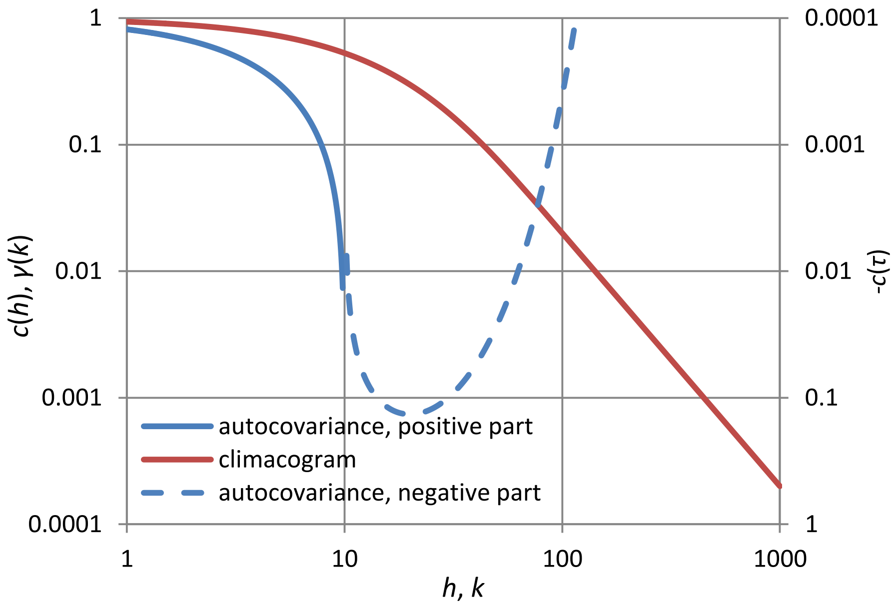
Example 5: A simple antipersistent process

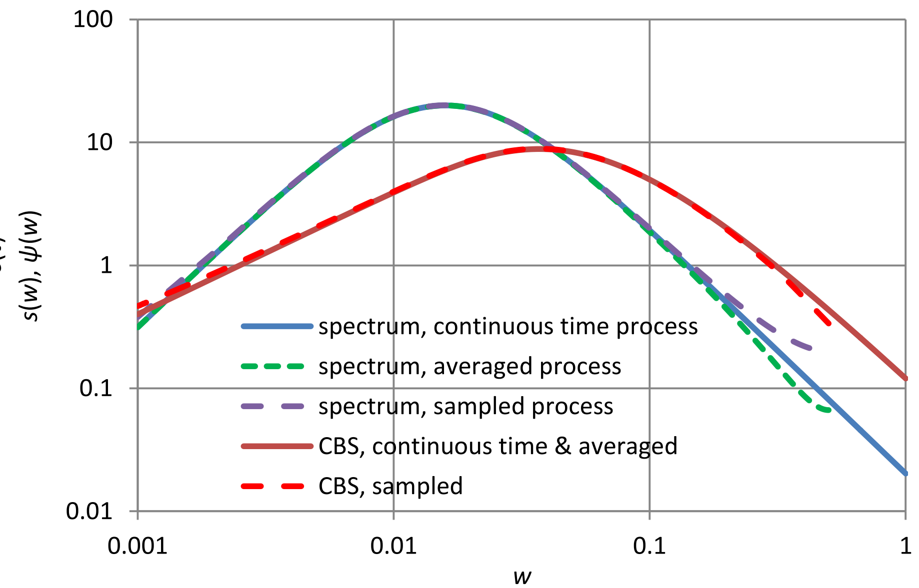
Type	Continuous time	Discrete time, averaged	Discrete time, sampled
Variance of instantaneous process	$\gamma_0 = \gamma(0) = c(0) = \lambda$	Not applicable	$\check{\gamma}_0 = \gamma_0 = \gamma(0) = \lambda$
Climacogram Variance at scale $k = \kappa D$)	$\gamma(k) = \frac{2\lambda}{k/a} \times \left(\frac{1-e^{-k}}{k/a} - e^{-k/a}\right)$	As in continuous time	$\check{\gamma}^{(\kappa)} = \frac{\check{\Gamma}^{(\kappa)}}{\kappa^2}$, where $\check{\Gamma}^{(k)}$ is calculated from (7)
Autocovariance function for lag $h = jD$	$c(h) = \lambda(1 - h/a)\mathrm{e}^{-h/a}$	$c_j^{(D)}$: From (8)	$\check{c}_j = \lambda(1 - jD/a)\mathrm{e}^{-jD/a}$ $(j \geq 0)$
Power spectrum for frequency $w = \omega/D$	$s(w) = 8\lambda a \left(\frac{2\pi a w}{1+(2\pi a w)^2}\right)^2$	$s^{(D)}(w)$: closed expression too complex	$\check{s}(w)$: closed expression too complex
Asymptotic slopes, global	$\gamma^{\#}(\infty) = 2\sigma^{\#}(\infty) = -2$ $\psi^{\#}(0) = 1;\ s^{\#}(0) = 2$		
Asymptotic slopes, local	$\gamma^{\#}(0) = \sigma^{\#}(0) = 0$ $\psi^{\#}(\infty) = s^{\#}(\infty) = -2$		

Notes

1. It is readily verifiable that $4\int_0^{\infty} c(h)\mathrm{d}h = \psi(0) = s(0) = 0$ (while of course $c(0) = \gamma(0) = \lambda > 0$). This is the condition making the process antipersistent.
2. In this process for $h > a$ the autocovariance is consistently negative—but for small h it is positive.
3. A similar antipersistent process but with autocovariance becoming again positive after a certain h is the one with autocovariance $c(h) = \lambda\left(1 - (5/3)(h/a) + (1/3)(h/a)^2\right)\mathrm{e}^{-h/a}$.

Example 5: A simple antipersistent process (contd.)





Notes

1. The parameters are λ = 1, a = 10, and the spacing is D = 1. Notes 2-4 of Example 1 are also valid in this example.
2. As shown in the figures, antipersistence is manifested in the positive slopes in the power spectrum and the CS. Clearly, these slopes are positive only for low frequencies w. For high w they should necessarily be negative. (Processes with increasing slopes for *all* frequencies, or negative autocorrelations for *all* lags, are physically unrealistic, even though they can be mathematically feasible, e.g. the HK process for H < 0.5).
3. Due to the shape of the autocorrelation function, this process has been termed in the geostatistics literature "the hole-effect model".

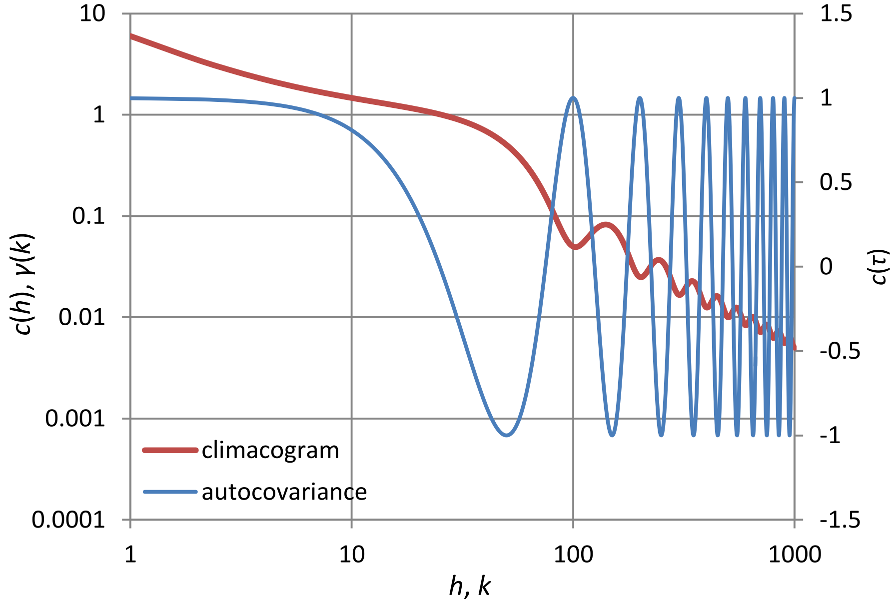
Example 6: A periodic process with white noise

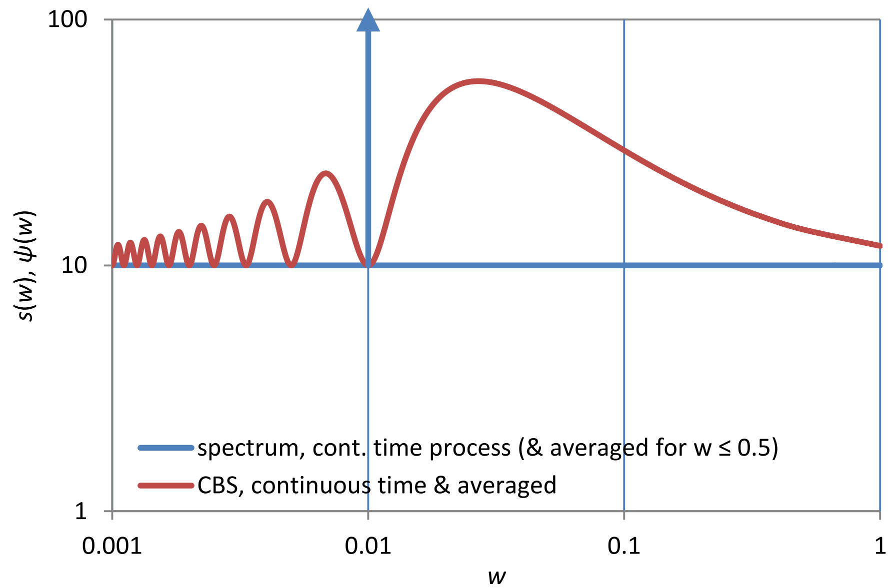
Type	Continuous time	Discrete time, averaged	D.t., sampled
Variance of instantaneous process	$\gamma_0 = \gamma(0) = c(0) = \infty$	Not applicable	$\check{\gamma}_0 = \gamma_0 = \gamma(0) = \infty$
Climacogram (Variance at scale $k = \kappa D$)	$\gamma(k) = \frac{\lambda_1}{k/a} + \lambda_2 \mathrm{sinc}^2\left(\frac{\pi k}{a}\right)$ for $k \neq (m + ½)a$, $m \in \mathrm{N}^0$ (discontinuities at $k=(m+½)a$)	As in continuous time	$\check{\gamma}^{(\kappa)} = \infty$
Autocovariance function for lag $h = jD$	$c(h) = \lambda_1 \delta(h/a) + \lambda_2 \cos(2\pi\, h/a)$	$c_j^{(D)} = \frac{\lambda_2}{(2\pi D/a)^2} \times \left(2\cos\left(\frac{2\pi D j}{a}\right) - \cos\left(\frac{2\pi D(j+1)}{a}\right) - \cos\left(\frac{2\pi D(j-1)}{a}\right)\right)$	$\check{c}_j = \lambda_2 \cos(2\pi\, Dj/a)$ $(j \geq 1)$
Power spectrum for frequency $w = \omega/D$	$s(w) = 2\lambda_1 a + \lambda_2 a\, \delta(aw - 1)$	$s^{(D)}(w) = 2\lambda_1 a + \lambda_2 a\, \delta(aw - 1)$ (as in continuous time)	$\check{s}(w) = \infty$
Asymptotic slopes for $\lambda_1 > 0$ [and for $\lambda_1 = 0$; but not valid for $s^{\#}(\,)$]	$\gamma^{\#}(\infty) = 2\sigma^{\#}(\infty) = -1\ [-2]$ $\psi^{\#}(0) = s^{\#}(0) = 0\ [+1]$ $\gamma^{\#}(0) = \sigma^{\#}(0) = -1\ [0]$ $\psi^{\#}(\infty) = s^{\#}(\infty) = 0\ [-3]$		

Notes

1. Strictly speaking, the periodic component is a deterministic rather than a stochastic process. In this respect, the process should be better modelled as a cyclostationary one.
2. However, the fact that the autocorrelation is a function of the lag h only, allows the process to be treated as a typical stationary stochastic process.
3. The infinite variance of the instantaneous process is a result of the white noise component, which in continuous time has infinite variance.

Example 6: A periodic process with white noise (contd.)





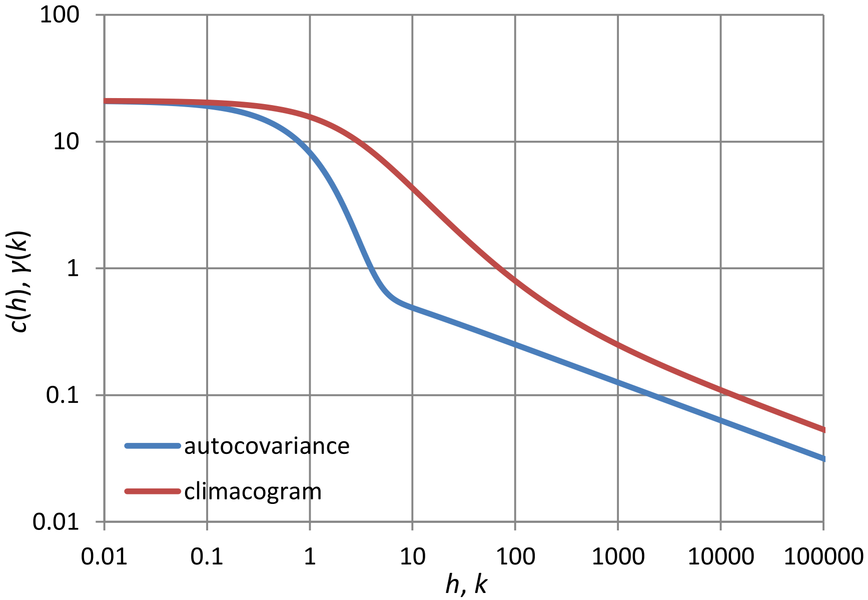
Notes

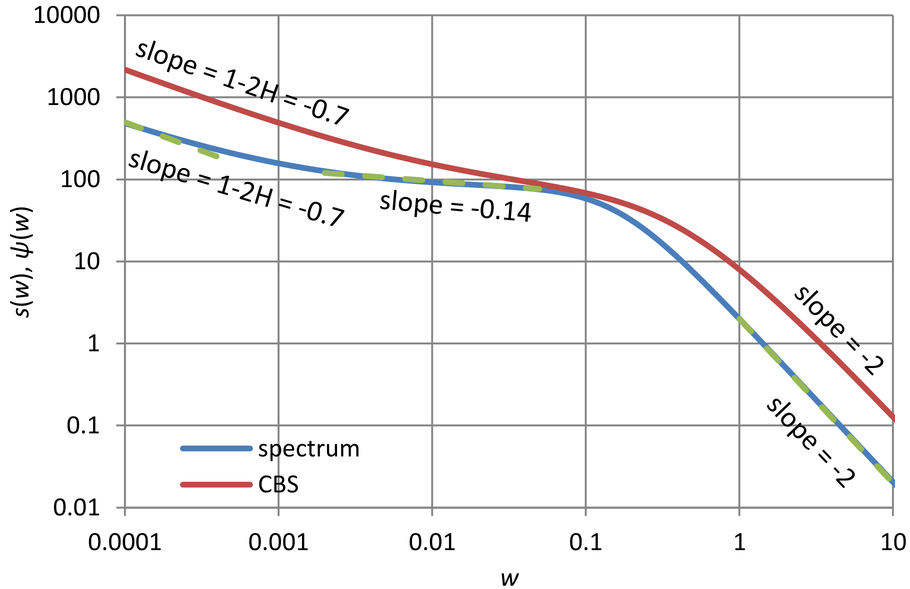
1. The parameters are $\lambda_1 = 0.05$, $\lambda_2 = 1$ (so that the random noise is 5% of the variability implied by periodicity), $a = 100$, and the spacing is $D = 1$.
2. While asymptotically (for $w \to 0$ and ∞) the spectrum and the CS behave similarly, they have marked differences for other w. In particular, the power spectrum has a peak (an infinite impulse) at $h = 1/a$; the CS has a trough (negative peak) at this value as well as at $h = 1/ka$ for integer k (as the variance is diminished at these resolutions). Also CS has discontinuities at frequencies $(m + ½)a$, where m is any integer (not plotted here).
3. Thus, the power spectrum is more robust in identifying periodicities, yet the CS provides a hint about their existence.

Example 7: A composite long-range and short-range dependence

The process whose characteristics are shown in the figures below is composed of a modified HK process (p. 36) and a Markov process (p. 32) and has autocovariance:

$$c(h) = \lambda_1 (1 + h/a)^{2H-2} + \lambda_2 \mathrm{e}^{-h/a} \tag{82}$$





Notes

1. The parameters are $\lambda_1 = 1$, $\lambda_2 = 20$, $a = 10$, $H = 0.85$ (only the continuous time properties are depicted).
2. As in the Example 4, again an intermediate slope (in this case a mild one) appears in the power spectrum. Again this is artificial, here imposed by the Markov process, and does not indicate a scaling behaviour.

References

- Dimitriadis, P., and D. Koutsoyiannis, 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, 2015.
- Dimitriadis, P., and Koutsoyiannis, D. (2017), Stochastic synthesis approximating any process dependence and distribution, *Stochastic Environmental Research & Risk Assessment* (in review).
- Hemelrijk, J., Underlining random variables, *Statistica Neerlandica*, 20 (1), 1-7, 1966.
- Gneiting, T., and M. Schlather, Stochastic models that separate fractal dimension and the Hurst effect, *Society for Industrial and Applied Mathematics Review*, 46 (2), 269-282, 2004.
- Jaynes, E.T. Information theory and statistical mechanics, *Physical Review*, 106 (4), 620-630, 1957.
- Koutsoyiannis, D., A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, *Water Resources Research*, 36 (6), 1519–1533, 2000.
- Koutsoyiannis, D., Nonstationarity versus scaling in hydrology, *Journal of Hydrology*, 324, 239–254, 2006.
- Koutsoyiannis, D., A random walk on water, *Hydrology and Earth System Sciences*, 14, 585–601, 2010.
- Koutsoyiannis, D., Hurst-Kolmogorov dynamics as a result of extremal entropy production, *Physica A*, 390 (8), 1424–1432, 2011a.
- Koutsoyiannis, D., Hurst-Kolmogorov dynamics and uncertainty, *Journal of the American Water Resources Association*, 47 (3), 481–495, 2011b.
- Koutsoyiannis, D., Climacogram-based pseudospectrum: a simple tool to assess scaling properties, *European Geosciences Union General Assembly 2013, Geophysical Research Abstracts, Vol. 15*, Vienna, EGU2013-4209, European Geosciences Union, 2013a (http://itia.ntua.gr/1328).
- Koutsoyiannis, D., Physics of uncertainty, the Gibbs paradox and indistinguishable particles, *Studies in History and Philosophy of Modern Physics*, 44, 480–489, 2013b.
- Koutsoyiannis, D., Entropy: from thermodynamics to hydrology, *Entropy*, 16 (3), 1287–1314, 2014a.
- Koutsoyiannis, D., Random musings on stochastics (Lorenz Lecture), *AGU 2014 Fall Meeting*, San Francisco, USA, American Geophysical Union, 2014b (http://www.itia.ntua.gr/en/docinfo/1500/).
- Koutsoyiannis, D., Generic and parsimonious stochastic modelling for hydrology and beyond, *Hydrological Sciences Journal*, 61 (2), 225–244, doi: 10.1080/02626667.2015.1016950, 2016.
- Koutsoyiannis, D., and A. Montanari, Statistical analysis of hydroclimatic time series: Uncertainty and insights, *Water Resources Research*, 43 (5), W05429, doi:10.1029/2006WR005592, 2007.

References (contd.)

- Koutsoyiannis, D., and A. Montanari, Negligent killing of scientific concepts: the stationarity case, *Hydrological Sciences Journal*, 60 (7-8), 1174–1183, doi: 10.1080/02626667.2014.959959, 2015.
- Lombardo, F., E. Volpi, and D. Koutsoyiannis, Effect of time discretization and finite record length on continuous-time stochastic properties, *IAHS - IAPSO - IASPEI Joint Assembly*, Gothenburg, Sweden, International Association of Hydrological Sciences, International Association for the Physical Sciences of the Oceans, International Association of Seismology and Physics of the Earth's Interior, 2013 (http://itia.ntua.gr/1380).
- Lombardo, F., E. Volpi, D. Koutsoyiannis, and S.M. Papalexiou, 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, 2014.
- Mackey, M.C., *Time's Arrow: The Origins of Thermodynamic Behavior*, Dover, Mineola, NY, USA, 175 pp., 2003.
- Montanari, A., and D. Koutsoyiannis, Modeling and mitigating natural hazards: Stationarity is Immortal!, *Water Resources Research*, 50 (12), 9748–9756, doi: 10.1002/2014WR016092, 2014.
- Papalexiou, S.M., D. Koutsoyiannis, and A. Montanari, Mind the bias!, *STAHY Official Workshop: Advances in statistical hydrology*, Taormina, Italy, International Association of Hydrological Sciences, 2010.
- Papoulis, A., *Probability, Random Variables and Stochastic Processes*, 3rd edn., McGraw-Hill, New York, 1991.
- Stein, M.L., *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, 1999.
- Stewart, J., Positive definite functions and generalizations, an historical survey, *Rocky Mountain J. Math*., 6(3), 1976.

Acknowledgment

I am grateful to Federico Lombardo and Panayiotis Dimitriadis for their remarks and corrections of some errors in earlier versions.