Generic and parsimonious stochastic modelling for hydrology and beyond

Demetris Koutsoyiannis

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

Abstract The old principle of parsimonious modelling of natural processes has regained its importance in the last few years. The inevitability of uncertainty and risk, and the value of stochastic modelling in dealing with them, are also again appreciated, after a period of growing hopes for radical reduction of uncertainty. Yet in stochastic modelling of natural processes several families of models are used which are often non-parsimonious, unnatural or artificial, theoretically unjustified and, eventually, unnecessary. Here we develop a general methodology for more theoretically justified stochastic processes, which evolve in continuous time and stem from maximum entropy production considerations. The discrete-time properties thereof are theoretically derived from the continuous-time ones and a general simulation methodology in discrete time is built, which explicitly handles the effects of discretization and truncation. Some additional modelling issues are discussed with focus on model identification and fitting, which are often made using inappropriate methods.

1   Introduction

The principle of parsimony in explaining and modelling natural phenomena is very old[*] but in the last three decades its importance may have been neglected. The power of computers encouraged many to seek inflationary detailed modelling approaches, hoping that these would provide exact descriptions of real-world systems by representing any detail thereof and would radically reduce uncertainty. After a period of growing hopes for dramatic reduction of uncertainty in modelling of natural, and particularly hydrological, processes, the inevitability of uncertainty and risk is again recognized. At the same time, the value of stochastic modelling in dealing with uncertainty and risk is also again appreciated (Koutsoyiannis *et al.* 2009; Koutsoyiannis 2010; Montanari and Koutsoyiannis 2012; Montanari *et al.* 2013). Stochastic

---

[*] It was first expressed by Aristotle (384–322 BC) in the following manner:

> Ἔστω γὰρ αὕτη ἡ ἀπόδειξις βελτίων τῶν ἄλλων τῶν αὐτῶν ὑπαρχόντων, ἡ ἐξ ἐλαττόνων αἰτημάτων ἢ ὑποθέσεων ἢ προτάσεων.
> *We may assume the superiority, other things being equal, of the demonstration which derives from fewer postulates or hypotheses or propositions.* [Posterior Analytics, I, 25]

> Φανερὸν ὅτι μακρῷ βέλτιον πεπερασμένας ποιεῖν τὰς ἀρχάς, καὶ ταύτας ὡς ἐλαχίστας πάντων γε τῶν αὐτῶν μελλόντων δείκνυσθαι, καθάπερ ἀξιοῦσι καὶ οἱ ἐν τοῖς μαθήμασιν.
> *Obviously, it is much better to assume a finite number of principles, as few as possible yet sufficient to prove what has to be proved, like in what mathematicians demand.* [On the Heavens, III, 4]

The principle was revived by medieval philosophers including Robert Grosseteste (c. 1168-1253), Thomas Aquinas (c. 1225-1274) and William of Ockham (c. 1285-1347), whose statement

> *Plurality is not to be posited without necessity*.

has given the principle the alternative name *Ockham's razor* (see also Gauch 2003).

modelling aims to make probabilistic predictions of future events, conditionally on past observations or not. As analytical calculation of probabilities becomes cumbersome or even infeasible in many cases, stochastic modelling is typically combined with Monte Carlo simulation, which relies on generation of synthetic time series for the processes of interest.

The earlier departure from the principle of parsimony may have also affected stochastic modelling of natural systems. The literature offers a variety of stochastic models which are widely used in simulation and prediction of hydrological and geophysical time series. The dominant approach in stochastic modelling is to choose and fit a model from a repertoire offered in books on time-series analysis. This includes stylized families of models mostly known by their acronyms, such as AR (for autoregressive), MA (for moving average), ARMA (for autoregressive moving average), ARIMA (for autoregressive integrated moving average), ARFIMA (for autoregressive fractionally integrated moving average), and so on. Despite being so common, most of the models are too artificial (in the sense that they were constructed by taking an initially simple model and expanding it mathematically using additional terms), while the typical modelling approach is characterized by the following problems:

- These models refer, by definition (or by construction), to discrete time while natural processes evolve in continuous time. A model defined in discrete time may not necessarily correspond to a continuous-time process.
- With the exception of a few cases, such as AR(1) or ARMA(1, 1), the models are over-parameterized and thus not parsimonious. Moreover, the stochastic structure of the processes is tightly associated with the number of parameters (e.g. to make an autoregressive process with different structure from that of the AR(1) process, we need to use a number of additional autoregression parameters, thus constructing an AR(q) process with $q > 1$). This proves to be a problem in many cases (Koutsoyiannis 2000), as ideally, in parsimonious modelling, the alteration of the stochastic structure of a process should not require additional parameters. In cases where the stochastic model is coupled with a deterministic process-based model (cf. Montanari and Koutsoyiannis 2012) the parameters of the stochastic process are additional to those of the deterministic model and thus lack of parsimony in the stochastic model also affects the entire modelling scheme.
- The identification of the models is typically based on the estimation of the autocorrelation function from data. However, this is characterized by negative bias and uncertainty. While in model identification much importance is typically given to the autocorrelation coefficients for a number of lags, the bias is effortlessly neglected. However, the bias may be very high, particularly when the process is characterized by long-term persistence (LTP), so high that it may even hide the presence of LTP (Koutsoyiannis 2003; Koutsoyiannis and Montanari 2007; Papalexiou *et al.* 2011; see also section 6.1 below).

It is not too difficult to get rid of all these model families and modelling practices, which are in fact unnecessary, and to construct a general method to easily generate synthetic series from processes with other autocorrelation structures. Koutsoyiannis (2000) proposed a methodology for the generation of time series from a process with any arbitrary autocorrelation function, provided that it is mathematically feasible (that is, it is positive definite, or, in other words, corresponds to a non-negative power spectrum). A specific application of this generic

methodology for the Hurst-Kolmogorov process (also known as fractional Gaussian noise; see below) has been given in Koutsoyiannis (2002).

This paper reports some essential developments on several issues that have been made in the last decade and remain unpublished, hoping that the rather simple, parsimonious and general framework the paper provides suffices for most applications of stationary processes, with emphasis on those with LTP, in hydrology and beyond, single-variate or multivariate. The case of cyclostationary processes is also briefly discussed. In brief, the advancements reported next are the following:

- The entire framework is based on the concept of a continuous-time stochastic process (section 2). Inevitably, the generation of time series (similarly to the observation of a natural process) requires discretization of time and this can be done either by instantaneously sampling from the continuous-time process or by aggregating or averaging it at specified time intervals. However, instead of formulating stochastic properties for discrete time, it is possible (and much better; see section 3) to derive the properties of the latter from the continuous-time process.
- The specific processes examined and presented as examples (section 3) are parameter parsimonious as, in addition to the marginal moments of the process and a time-scale parameter, contain no more than two (shape) parameters. While the simulation methodology discussed can be applied with as many parameters as one wishes, over-parameterization should be avoided at any rate, particularly in stochastics (e.g. Lombardo *et al.*, 2014). As far as the dependence structure of the process is concerned, only the parameters that suffice to determine the type of this structure (e.g., short- or long-range dependence or both) should (and usually can) actually be determined.
- The framework has a tight connection with entropic considerations, which provide some guidance about the expected type of dependence and make the models more physically plausible.
- The framework puts special emphasis on the climacogram, which is defined to be the variance (or the standard deviation) of the time-averaged process as a function of time scale of averaging (Koutsoyiannis, 2010). This tool has some convenient theoretical properties but, above all, it is best suited for model identification and parameter estimation. While, like other stochastic tools, it also involves estimation bias, this can be determined analytically and included in the model identification and fitting.
- The framework is complete and stand-alone as it incorporates a general procedure to simulate from any stationary stochastic model.

The framework is illustrated with several examples (section 5.2). These are not real-world case studies but rather designed to underline the simplicity and effectiveness of the procedures in simulation.

2   Definitions and basic concepts

Let $\underline{x}(t)$ be a stationary stochastic process quantifying the instantaneous quantity of a certain hydrological or other physical process at (continuous) time t. The quantity $\underline{x}$ is assumed to be a (continuous) random variable and this is reflected in the notation by underlining it, adopting the so-called Dutch convention (Hemelrijk 1966). Regular variables such as the time t or realizations of $\underline{x}$ are denoted by non-underlined symbols. Since the process is stationary, the marginal

distribution function of $\underline{x}(t)$, $F(x) := P\{\underline{x}(t) \le x\}$, where P denotes probability, as well as the marginal probability density function $f(x) := \mathrm{d}F(x)/\mathrm{d}x$ are real-valued functions of the real (regular) variable x, not functions of time. (Throughout the text, definitions are denoted by the symbol ":=" meaning "is defined as").

While natural processes evolve in continuous time, observation and generation of synthetic time series are only made in discrete time and therefore we always need to consider derived discrete-time processes and to take into account the effect of discretization in the properties thereof. As shown in Figure 1, from the continuous time process, we may define two interesting discrete-time processes (with the discrete time denoted as i), for which we usually observe or generate time series. By sampling at spacing Δ we obtain the sampled instantaneous process

$$\underline{x}_i := \underline{x}(i\Delta) \tag{1}$$

By averaging at time scale Δ we obtain the averaged process

$$\underline{x}_i^{(\Delta)} := (\underline{X}(i\Delta) - \underline{X}((i-1)\Delta))/\Delta \tag{2}$$

where

$$\underline{X}(t) := \int_0^t \underline{x}(\xi)\mathrm{d}\xi \tag{3}$$

is the cumulative process (see Figure 1, upper panel), which obviously is nonstationary. Both the sampled instantaneous process $\underline{x}_i$ and the averaged process $\underline{x}_i^{(\Delta)}$ are stationary. From them we can construct the nonstationary cumulative processes at discrete time i: namely, that of $\underline{x}_i^{(\Delta)}$ is precisely $\underline{X}(i\Delta)/\Delta$, while that of $\underline{x}_i$ is different, i.e.,

$$\underline{X}_i := \sum_{j=1}^i \underline{x}_j \tag{4}$$

Without loss of generality we assume that the mean of the process $\underline{x}(t)$ is zero ($\mathrm{E}[\underline{x}(t)] = 0$) and we denote

$$\gamma_0 := \mathrm{Var}[\underline{x}(t)] \tag{5}$$

its variance, which is also the variance of the sampled process, i.e., $\mathrm{Var}[\underline{x}_i] = \gamma_0$. Let

$$\Gamma(t) := \mathrm{Var}[\underline{X}(t)], \quad \gamma(\Delta) := \mathrm{Var}[\underline{X}(\Delta)/\Delta] = \Gamma(\Delta)/\Delta^2 \tag{6}$$

be the variances of the cumulative and the averaged process, respectively (note that $\Gamma(0) = 0$ and $\gamma(0) = \gamma_0$). We call $\gamma(\Delta)$ and $\Gamma(t)$, as functions of time scale Δ or time t, the *climacogram* and the *cumulative climacogram* of the process, respectively (see Koutsoyiannis 2010 for the justification of the term and note that sometimes it has been used to describe the standard deviation instead of the variance). We observe that, while $\Gamma(t)$ is the variance of the nonstationary process $\underline{X}(t)$, $\gamma(\Delta)$ is the variance of the stationary process $\underline{x}_i^{(\Delta)}$, i.e., $\gamma(\Delta) = \mathrm{Var}[\underline{x}_i^{(\Delta)}]$ and thus it is not a function of time t.
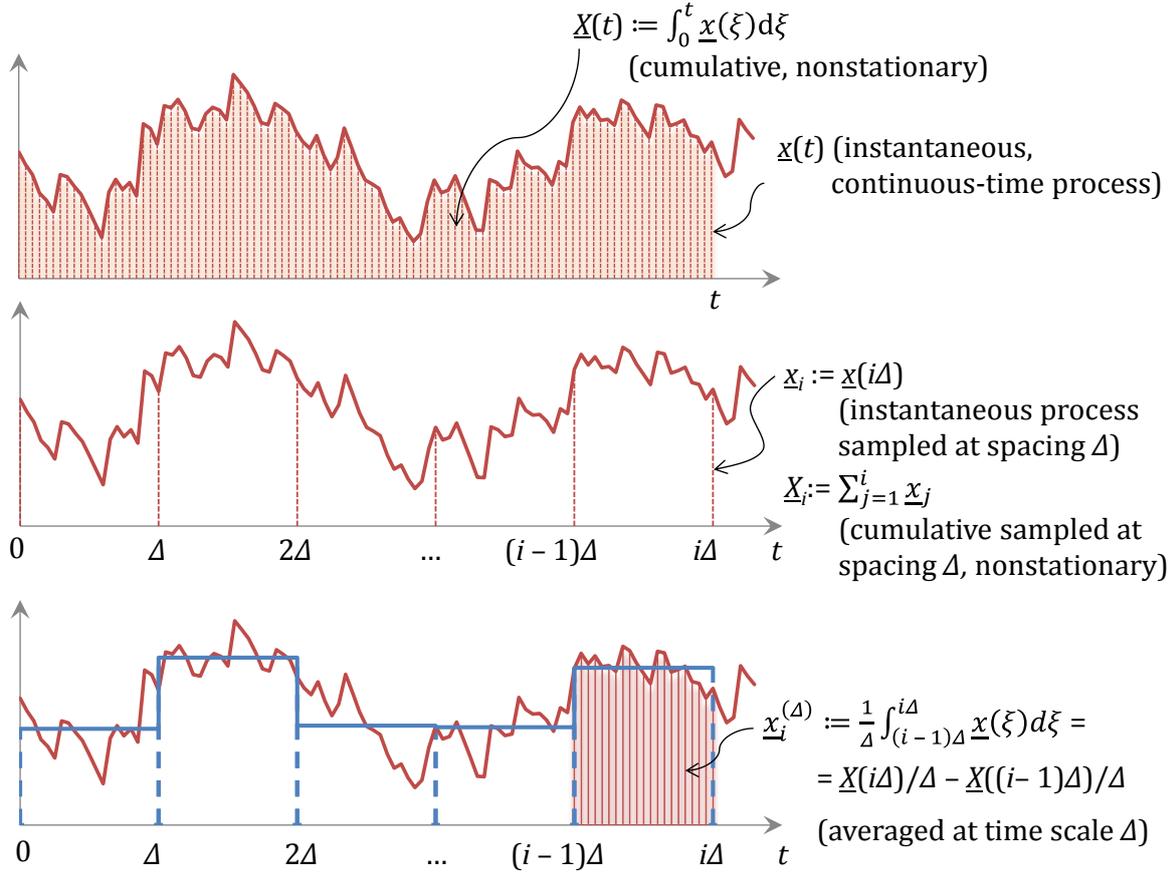
Figure 1 Explanation sketch of the definitions.

Once the climacogram (or the cumulative climacogram) is known, every second-order property of the process, such as the autocovariance, autocorrelation, power spectrum, variogram or structure function, can be obtained from this (Koutsoyiannis 2013b). For example, the autocovariance

$$c(\tau) := \mathrm{Cov}[\underline{x}(t), \underline{x}(t+\tau)] \tag{7}$$

where τ is the lag time, is related to the climacogram by

$$c(\tau) = \frac{1}{2} \frac{\mathrm{d}^2 \Gamma(\tau)}{\mathrm{d}\tau^2} = \frac{1}{2} \frac{\mathrm{d}^2 (\tau^2 \gamma(\tau))}{\mathrm{d}\tau^2} \tag{8}$$

while the inverse transformation is

$$\gamma(\varDelta) = \frac{\Gamma(\varDelta)}{\varDelta^2} = \frac{2}{\varDelta^2} \int_0^{\varDelta} (\varDelta - \tau) c(\tau) \mathrm{d}\tau = 2 \int_0^1 (1 - \xi) c(\xi \varDelta) \mathrm{d}\xi \tag{9}$$

Also, the power spectrum $s(w)$, where w is frequency, is, by definition, the Fourier transform of the autocovariance, i.e.,

$$s(w) := 4 \int_0^{\infty} c(\tau) \cos(2\pi w\tau) \, \mathrm{d}\tau \tag{10}$$

(note that $c(\tau)$ is an even function) while the inverse transformation is

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

From (11), we observe that the entire area under the spectrum for frequencies from 0 to ∞ equals the variance of the process ($c(0) = \gamma_0$); for that reason the multiplier 4 was used in the definition of the power spectrum in (10).

The structure function (also known as semivariogram or variogram) of the process is defined as

$$h(\tau) \coloneqq (1/2) \, \mathrm{Var}\big[\underline{x}(t) - \underline{x}(t + \tau)\big] = \gamma_0 - c(\tau) \tag{12}$$

i.e., it is directly connected to the autocovariance and hence can also be estimated from the climacogram. However, by analogy we can define a climacogram-based structure function (CBSF) as

$$g(\Delta) \coloneqq \gamma_0 - \gamma(\Delta) \tag{13}$$

which has the same asymptotic properties as the standard structure function $h(\tau)$. Combining the climacogram and the CBSF we can also derive a climacogram-based spectrum (CBS), defined as

$$\psi(w) \coloneqq \frac{2}{w\gamma_0} \gamma(1/w) g(1/w) = \frac{2\,\gamma(1/w)}{w}\left(1 - \frac{\gamma(1/w)}{\gamma_0}\right) \tag{14}$$

where $w \equiv 1/\Delta$ is frequency (as in the standard power spectrum). In most of the cases, the asymptotic behaviour of the CBS is the same with that of the standard power spectrum, and thus the CBS constitutes an easily calculated (i.e. without performing a Fourier transform) substitute of the standard power spectrum, particularly useful when handling data series of small length.

The variance (hence the climacogram) is more naturally than any other of the above tools associated with the uncertainty, and hence the entropy, of the process. We recall that the (Boltzmann-Gibbs-Shannon) entropy of the cumulative process $\underline{X}(t)$ with probability density function $f(X; t)$ is a dimensionless quantity defined as:

$$\Phi\big[\underline{X}(t)\big] \coloneqq \mathrm{E}\left[-\ln \frac{f(X;t)}{h(X)}\right] = -\int_{-\infty}^{\infty} \ln \frac{f(X; t)}{h(X)} f(X; \, t) \, dX \tag{15}$$

where $h(X)$ is the density of a background measure. 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.

Assuming that the background measure is the Lebesgue measure, with constant density $h(X) = h = 1[X]^{-1}$, and using simple constraints related to the preservation of the mean and variance, the principle of maximum entropy yields that the probability distribution of $\underline{X}$ will be Gaussian. For a Gaussian process, the entropy depends on its variance $\Gamma(t)$ only and is given as (Papoulis 1991):

$$\Phi[\underline{X}(t)] = (1/2) \ln(2\pi \mathrm{e} \, \Gamma(t)/h^2) \tag{16}$$

Koutsoyiannis (2011) defined entropy production as the time derivative of entropy,

$$\Phi'[\underline{X}(t)] := \mathrm{d}\Phi[\underline{X}(t)]/\mathrm{d}t \qquad (17)$$

and the entropy production in logarithmic time (EPLT) as the derivative in logarithmic time:

$$\varphi(t) \equiv \varphi[\underline{X}(t)] := \Phi'[\underline{X}(t)]\,t \equiv \mathrm{d}\Phi[\underline{X}(t)] \,/\, \mathrm{d}(\ln t) \qquad (18)$$

For a Gaussian process, by virtue of (16), the EPLT is

$$\varphi(t) = \Gamma'(t)\,t \,/\, 2\Gamma(t) \qquad (19)$$

This expresses in dimensionless terms the rate of uncertainty production at time $t > 0$ when nothing is known about the past ($t < 0$) and the present ($t = 0$). 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 past and present. For a Markovian process (see section 3), in which only the present (the latest known value) matters in conditioning, it is relatively easy to calculate the conditional variance. Using detailed results from Koutsoyiannis (2011) for a Markovian process, we can verify that the following relationship holds:

$$\Gamma_{\mathrm{C}}(t) = 2\Gamma(t) - \Gamma(2t)/2 \qquad (20)$$

while the derivative is

$$\Gamma'_{\mathrm{C}}(t) = 2\Gamma'(t) - \Gamma'(2t) \qquad (21)$$

so that the conditional EPLT is

$$\varphi_{\mathrm{C}}(t) = \frac{\Gamma'_{\mathrm{C}}(t)t}{2\Gamma_{\mathrm{C}}(t)} = \frac{(2\Gamma'(t) - \Gamma'(2t))t}{4\Gamma(t) - \Gamma(2t)} \qquad (22)$$

Equations (20)-(22) are strictly valid for a Markovian Gaussian process. However, any persistent (positively correlated) process (under some rather general conditions) can be decomposed into a sum of Markovian process (Koutsoyiannis 2002, 2011), and therefore (20) and (21), by virtue of their linear form (subtraction of two functions), can provide approximations for any persistent Gaussian process. In fact Equation (20) provides a lower bound of the conditional variance for a non-Markovian process: it would be exact only if, by knowing the entire past of $\underline{x}(t)$, we would be able to decompose $\underline{x}(0)$ into the components corresponding to the constituents of the sum of Markovian process. This is not actually the case as the actual conditioning information is lower than assumed to derive (20). An alternative numerical method to evaluate an upper bound of $\Gamma_{\mathrm{C}}(t)$ and consequently $\Gamma'_{\mathrm{C}}(t)$ can be based on the methodology presented in Koutsoyiannis (2005b) and Koutsoyiannis *et al.* (2007) for discretized processes. Nonetheless, if EPLT is to be calculated, Equation (22) may provide a sufficient approximation because, by taking the ratio of the two quantities, the errors in the numerator and the denominator will tend to cancel each other.

3   Effect of discretization

As explained in the Introduction, observation and generation of synthetic time series are made in discrete time and therefore we always need to take into account the effect of discretization in the properties of the discrete-time processes. Once the continuous-time climacogram is known,

the discrete-time autocovariance function of the averaged process can be determined from the following relationship, analogous to (8), which is exact and generic (Koutsoyiannis 2013b):

$$c_j^{(\Delta)} = \frac{1}{2}\frac{\delta_\Delta^2 \Gamma(j\Delta)}{\Delta^2} = \frac{1}{\Delta^2}\left(\frac{\Gamma\big((j+1)\Delta\big) + \Gamma\big((j-1\Delta\big)}{2} - \Gamma(j\Delta)\right) \qquad (23)$$

where δ_Δ^2 is the second order central difference operator for spacing Δ and δ_Δ^2/Δ^2 is the second finite derivative. This is valid for $j > 0$, but it can be written in the following manner so that it is valid for any integer j:

$$c_j^{(\Delta)} = \frac{1}{\Delta^2}\left(\frac{\Gamma(|j+1|\Delta) + \Gamma\big((|j-1|\Delta)}{2} - \Gamma(|j|\Delta)\right) \qquad (24)$$

It is easily seen that for $j = 0$ this results in $c_0^{(\Delta)} = \gamma(\Delta)$, while $c_{-j}^{(\Delta)} = c_j^{(\Delta)}$, as it should.

For the sampled process, assuming that the discrete-time cumulative climacogram

$$\breve{\Gamma}^{(k)} := \mathrm{Var}[\underline{X}_k] \qquad (25)$$

is known for any positive integer k, the following relationship, similar to (23), gives its autocovariance function for any positive integer lag j:

$$\breve{c}_j = \frac{1}{2}\,\delta_1^2\breve{\Gamma}^{(j)} = \frac{\breve{\Gamma}^{(j+1)} + \breve{\Gamma}^{(j-1)}}{2} - \breve{\Gamma}^{(j)} \qquad (26)$$

This holds also for any type of discrete time process, even if it is not linked to a continuous-time one.

The inverse relationships, which give the (sampled) discrete-time climacograms from the autocovariance function and also hold for any type of discrete-time process, are:

$$\breve{\Gamma}^{(k)} = k\breve{\gamma}_0 + 2\sum_{j=1}^{k-1}(k-j)\breve{c}_j \qquad (27)$$

$$\breve{\gamma}^{(k)} = \frac{\breve{\Gamma}^{(k)}}{k^2} = \frac{1}{k}\left(\breve{\gamma}_0 + 2\sum_{j=1}^{k-1}\left(1-\frac{j}{k}\right)\breve{c}_j\right) \qquad (28)$$

where $\breve{\gamma}_0 \equiv \breve{c}_0$ is the variance of the discrete time process. The following recursive relationship facilitates calculation of the climacogram:

$$\breve{\Gamma}^{(k)} = 2\breve{\Gamma}^{(k-1)} - \breve{\Gamma}^{(k-2)} + 2\breve{c}_{k-1}, \quad \breve{\Gamma}^{(0)} = 0, \quad \breve{\Gamma}^{(1)} = \breve{\gamma}_0 \qquad (29)$$

Specifically, it is easily verified that recursive application of (29) results in (27).

The direct calculation of the discrete-time power spectra from that in continuous time is feasible but somewhat involved. Specifically, discretization makes the power spectrum a periodic function with period Δ and since it is, by definition, an even function it suffices to determine it for frequencies w in the interval $(0, 1/2\Delta)$ only, where the upper bound of this interval $(1/2\Delta)$ is known as the Nyquist frequency. Usually, in discrete time we use a dimensionless frequency ω and we also standardize the power spectrum so that its area remains equal to the variance, i.e.,

$$\omega := w\Delta, \quad \check{s}_{\mathrm{d}}(\omega) := \frac{\check{s}(w)}{\Delta}, \quad s_{\mathrm{d}}^{(\Delta)}(\omega) = \frac{s^{(\Delta)}(w)}{\Delta} \tag{30}$$

Thus, for the sampled process the following relationship connects $\check{s}_{\mathrm{d}}(\omega)$ with that in continuous time, $s(w)$ (cf. Papoulis, 1991, p. 336):

$$\check{s}_{\mathrm{d}}(\omega) = \frac{\check{s}(w)}{\Delta} = \frac{1}{\Delta} \sum_{j=-\infty}^{\infty} s\left(\frac{\omega + j}{\Delta}\right) = \frac{1}{\Delta} \sum_{j=-\infty}^{\infty} s\left(w + \frac{j}{\Delta}\right) \tag{31}$$

To find the relationship of $s_{\mathrm{d}}^{(\Delta)}(\omega)$ of the averaged process with $s(w)$, we utilize (31) 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 window Δ, i.e., $\underline{y}(t) := \frac{1}{\Delta} \int_{t-\Delta}^{t} \underline{x}(\xi)\mathrm{d}\xi$ (an operation known as a low-pass filtering; cf. Papoulis 1991, p. 325, also noticing the different notational and other conventions), is:

$$s_y(w) = s(w) \ \mathrm{sinc}^2(\pi w \Delta) \tag{32}$$

where $\mathrm{sinc}(x) := \sin(x)/x$. By sampling $\underline{y}(t)$ at intervals Δ, we find:

$$
\begin{aligned}
s_{\mathrm{d}}^{(\Delta)}(\omega) &= \frac{s^{(\Delta)}(w)}{\Delta} = \frac{1}{\Delta} \sum_{j=-\infty}^{\infty} s\left(\frac{\omega + j}{\Delta}\right) \mathrm{sinc}^2\big(\pi(\omega + j)\big) \\
&= \frac{1}{\Delta} \sum_{j=-\infty}^{\infty} s\left(w + \frac{j}{\Delta}\right) \mathrm{sinc}^2\big(\pi(w\Delta + j)\big)
\end{aligned} \tag{33}
$$

Alternatively the power spectrum of a discretized process can be calculated directly from the autocovariance function of the discretized process. Thus,

$$\check{s}_{\mathrm{d}}(\omega) = 2\check{c}_0 + 4 \sum_{j=1}^{\infty} \check{c}_j \cos(2\pi\omega j), \quad s_{\mathrm{d}}^{(\Delta)}(\omega) = 2c_0^{(\Delta)} + 4 \sum_{j=1}^{\infty} c_j^{(\Delta)} \cos(2\pi\omega j) \tag{34}$$

for the sampled and the averaged process, respectively. The former expression can also be applied for a discrete-time process that is not linked to a continuous-time process. Note that the first term in each of the right-hand parts of (34) equals the respective variance, i.e., $\check{c}_0 \equiv \check{\gamma}_0 (\equiv \gamma_0$ in case that the discrete-time process is sampled from a continuous-time process) and $c_0^{(\Delta)} \equiv \gamma(\Delta)$. The inverse transformation is

$$\check{c}_j = \int_0^{1/2} \check{s}_{\mathrm{d}}(\omega) \cos(2\pi\omega j) \, \mathrm{d}\omega, \quad c_j^{(\Delta)} = \int_0^{1/2} s_{\mathrm{d}}^{(\Delta)}(\omega) \cos(2\pi\omega j) \, \mathrm{d}\omega \tag{35}$$

Equations (34) and (35) are the discrete-time analogues of the continuous-time equations (10) and (11). Notice that, even in the discrete-time case, the inverse transformation is an integral, not a sum.

4 Example processes

Three example processes have been chosen to analyse here on the basis that they have some properties which make them physically plausible and thus candidate to be used in hydrological

and other applications. Specifically, they correspond to extremization of entropy production—in particular extremization of EPLT, which as shown in Koutsoyiannis (2011) is equivalent to extremization of entropy production. The first two, the Markov process and the Hurst-Kolmogorov (HK) process, are well-known processes which Koutsoyiannis (2011) found to correspond to EPLT extremization for asymptotic times, $t \to 0$ and $t \to \infty$; the third process, the hybrid Hurst-Kolmogorov process (HHK), is introduced here as a physically plausible modification and generalization of the HK process.
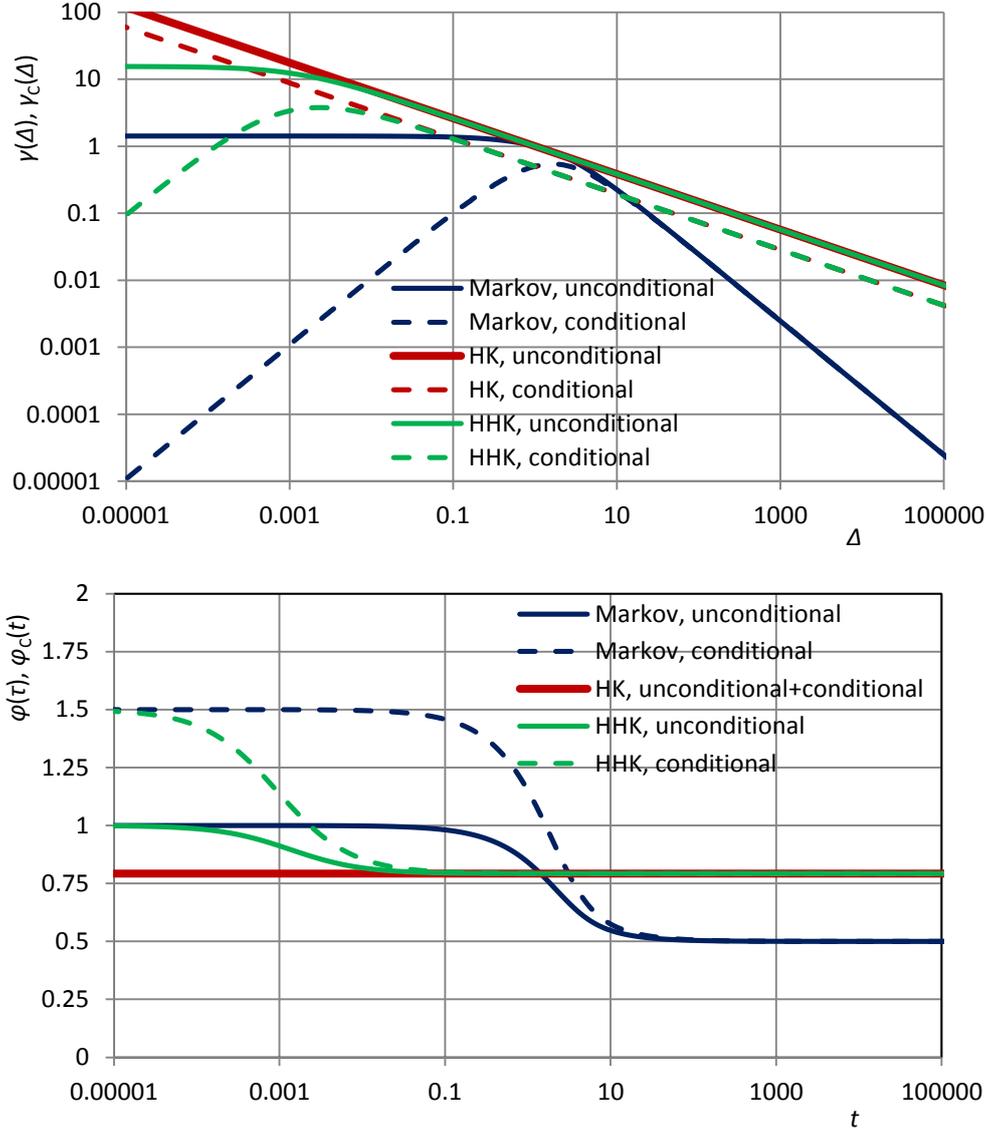
Figure 2 (upper) Climacograms and (lower) EPLTs of the three example processes (see text and Tables 1-2 for the definitions of processes). At time scale $\Delta = 1$ all three processes have the same variance $\gamma(1) = 1$ and the same autocovariance for lag 1, $c_1^{(1)} = 0.5$. Their parameters are (see text for their definitions): for the Markov process $a = 0.8686$, $\lambda = 1.4176$; for the HK process $a = 0.0013539$, $\lambda = 15.5032$, $H = 0.7925$; for the HHK process $a = 0.0013539$, $\lambda = 15.5093$, $\kappa = 0.5$, $H = 0.7925$ (note that for the HK process the parameter set $a = \lambda = 1$ is equivalent to the above, but the former set was preferred in order to be comparable to the HHK).

While the three processes will be described in the following subsections, it is instructive to overview at this point their entropic properties. These are illustrated in Figure 2 which depicts

their climacograms and their EPLTs *vs.* time. At time scale $\Delta = 1$ all three processes have the same variance $\gamma(1) = 1$ and the same autocovariance for lag 1, $c_1^{(1)} = 0.5$. These two values are used as constraints in the extremization of EPLT, as detailed in Koutsoyiannis (2011). As seen in Figure 2, the Markov process has the highest EPLT for $t \to 0$ ($\varphi(0) = 1$, $\varphi_C(0) = 1.5$) and the lowest for $t \to \infty$ ($\varphi(\infty) = \varphi_C(\infty) = 0.5$). Conversely, the HK process has constant EPLT, $\varphi(t) = \varphi_C(t) = H = 0.7925$, where H is known as the Hurst coefficient and its value 0.7925 is fully determined from the above constraints (see section 4.2). This EPLT of the HK process is lowest for $t \to 0$ and highest for $t \to \infty$. The HHK process has the highest EPLT both for $t \to 0$ ($\varphi(0) = 1$, $\varphi_C(0) = 1.5$ as in the Markov case) and for $t \to \infty$ ($\varphi(\infty) = \varphi_C(\infty) = H = 0.7925$, as in the HK case).

The subsections below provide all details of the three processes, while additional interesting continuous-time processes are given in Gneiting *et al.* (2012) and Koutsoyiannis (2013b).

4.1   The Markov process

By definition, a Markov process is one in which the future does not depend on the past when the present is known (Papoulis, 1991). As a consequence of this, the autocovariance function of the process is a pure exponential function (Equation (T1.4) in Table 1, in which λ (dimension $[x]^2$) is a state-scale parameter and α (dimension $[t]$) is time-scale parameter, while all other second-order properties of the process are determined using the framework of Sections 2 and 3 and are given in Table 1.

The process in continuous time is also known as the Ornstein–Uhlenbeck process. By inspecting Table 1 and comparing with known equations in the time-series literature, it is seen that in discrete time the process is identical to an AR(1) process if it is sampled at equidistant times, or a special case of an ARMA(1,1) process if it is averaged at time windows with constant length.

Major advantages of this process are its simplicity (all its second-order properties are given by simple equations in Table 1) and parsimony (it contains only one parameter, α, in addition to those describing its marginal distribution). However, the Markovian property, which is an ideal case not verified in nature, does not make it quite plausible for natural phenomena.

4.2   The Hurst-Kolmogorov process

The HK process can be defined based on its climacogram, which is a power function of time scale (Equation (T2.2) in Table 2). The autocovariance function and hence the power spectrum are then easily determined (see Table 2).

The process in continuous time is also known as fractional Gaussian noise (FGN) due to Mandelbrot and van Ness (1968), although these authors used a more complicated approach to define it. In essence, though, the mathematical process had been earlier proposed by Kolmogorov (1940), while Hurst (1951) pioneered the detection in geophysical time series of the behaviour described by this process; hence the name we use for this process. Because this process has infinite instantaneous variance, the sampled process in discrete time is not meaningful (many properties take infinite values). However, the averaged process is well behaving with all of its properties (including its variance) finite, which makes it quite useful in applications.

Table 1 Properties of the Markov process.

Property	Formula	Ref.
Variance		
Continuous-time and sampled process (Instantaneous)	$\gamma_0 = \gamma(0) = c(0) = \check{\gamma}_0 = \lambda$	(T1.1)
Averaged process, scale Δ (Climacogram)	$\gamma(\Delta) = \frac{2\lambda}{\Delta/\alpha}\left(1 - \frac{1-e^{-\Delta/\alpha}}{\Delta/\alpha}\right)$	(T1.2)
Sampled process, scale $k\Delta$ (Climacogram)	$\check{\gamma}^{(k)} = \frac{\lambda}{k\,(1-\rho)^2}\left(1 - \rho^2 - \frac{2\rho(1-\rho^k)}{k}\right), k = 1, 2, \dots$ where $\rho = e^{-\Delta/\alpha}$	(T1.3)
Autocovariance function		
Continuous-time process, lag τ	$c(\tau) = \lambda e^{-\tau/\alpha}$	(T1.4)
Averaged process, lag $\tau = j\Delta$	$c_j^{(\Delta)} = \frac{\lambda(1-e^{-\Delta/\alpha})^2}{(\Delta/\alpha)^2}\, e^{-(j-1)\Delta/\alpha}, j = 1,2, \dots$	(T1.5)
Sampled process, lag $\tau = j\Delta$	$\check{c}_j = \lambda e^{-j\Delta/\alpha} = \lambda\rho^j$ $(j = 0,1,2,\dots)$	(T1.6)
Power spectrum		
Continuous-time process, frequency w	$s(w) = \frac{4\alpha\lambda}{1+(2\pi\alpha w)^2}$	(T1.7)
Averaged process, frequency $w = \omega/\Delta$	$s^{(\Delta)}(w) = 4\alpha\lambda\left(1 - \frac{1}{\Delta/\alpha}\frac{(1-\cos(2\pi\Delta w))\sinh(\Delta/\alpha)}{\cosh(\Delta/\alpha)-\cos(2\pi\Delta w)}\right)$	(T1.8)
Sampled process, frequency $w = \omega/\Delta$	$\check{s}^{(\Delta)}(w) = \frac{2\lambda\Delta\sinh(\Delta/\alpha)}{\cosh(\Delta/\alpha)-\cos(2\pi\Delta w)}$	(T1.9)

The HK process is almost equally simple and parsimonious with the Markov process; again it contains only one parameter, H, in addition to those describing its marginal distribution. Notice that the process variance is controlled by the product $\lambda\,\alpha^{2-2H}$, so that λ and α, are not in fact separate parameters. Despite that, we prefer the formulation shown in Table 2 with three nominal parameters for dimensional consistency: α and λ are scale parameters with dimensions $[t]$ and $[x]^2$, respectively, while H, the Hurst coefficient, is dimensionless in the interval (0, 1).

For $H = 0.5$ the process reduces to pure white noise. For $0.5 < H < 1$ the process exhibits LTP (also called long-range dependence). For $0 < H < 0.5$ it is called an antipersistent process. Most of the expressions shown in Table 2 are valid in the all cases. However, the autocovariance $c(\tau)$ has different expressions in the three cases, as shown in Table 2. Specifically, for $H < 0.5$, the autocovariance $c(\tau)$ is negative for any lag $\tau > 0$, tending to $-\infty$ as $\tau \to \infty$. However, at $\tau = 0$, $c(0) = +\infty$, because this is the variance of process which cannot be negative; thus, there is a discontinuity at $\tau = 0$. Consequently, the averaged process has positive variance and all covariances negative. Such a process is not physically realistic because real-world events at near times are always positively correlated, which means that for small τ, $c(\tau)$ should be positive. Also, the infinite variance cannot appear in nature (it would be associated with infinite energy); even the white noise, whose autocovariance is a Dirac delta function (corresponding to infinite

variance, see (T2.3) for $H = 1/2$), cannot describe a natural process. Thus, the HK process can describe natural phenomena only for $0.5 < H < 1$ and for time scales not too small. Furthermore, values $H > 1$ that sometimes are being reported in the literature are mathematically invalid (Koutsoyiannis 2013b) and are results of inconsistent algorithms.

Table 2 Properties of the HK process.

Property	Formula	Ref.
Variance		
Continuous-time and sampled process (Instantaneous)	$\gamma_0 = \gamma(0) = c(0) = \breve{\gamma}_0 = +\infty$	(T2.1)
Averaged process, scale Δ (Climacogram)	$\gamma(\Delta) = \lambda(\alpha/\Delta)^{2-2H}$	(T2.2)
Autocovariance function		
Continuous-time process, lag τ	$c(\tau) = \lambda H(2H-1)(\alpha/\tau)^{2-2H}$ $\quad(H > 1/2)$ $c(\tau) = \lambda\,\delta(\tau/\alpha)$ $\quad\quad\quad\quad\quad(H = 1/2)$ $c(\tau) = \lambda H(2H-1)\big((\alpha/\tau)^{2-2H}$ $\quad\quad + \delta''(\tau/\alpha)\big)$ $\quad\quad\quad(H < 1/2)$	(T2.3)
Averaged process, lag $\tau = j\Delta$	$c_j^{(\Delta)} = \lambda(\alpha/\Delta)^{2-2H}\left(\frac{|j-1|^{2H}+|j+1|^{2H}}{2} - |j|^{2H}\right)$	(T2.4)
Power spectrum		
Continuous-time process, frequency w	$s(w) = \dfrac{2\alpha\lambda\,\Gamma(2H+1)\sin(\pi H)}{(2\pi\alpha w)^{2H-1}}$	(T2.5)

Notes: (a) $\Gamma(\,)$ is the gamma function, $\delta(\,)$ is the Dirac delta function and $\delta''(\,)$ is its second derivative. (b) The sampled process is not meaningful as its variance (instantaneous, $\breve{\gamma}_0$, or at any time scale, $\breve{\gamma}^{(k)}$) is infinite and its spectrum, $\breve{s}(w)$, is also infinite. (c) The power spectrum of the averaged process exists (it is finite for $w > 0$) but it does not have a closed expression.

Most of the functions in Table 2 describe simple scaling laws, i.e., are depicted as straight lines if plotted in double-logarithmic plots. More formally, these slopes are represented by log-log derivatives, such as

$$\gamma^\#(\Delta) := \frac{\mathrm{d}(\ln\gamma(\Delta))}{\mathrm{d}(\ln\Delta)} = \frac{\Delta\,\gamma'(\Delta)}{\gamma(\Delta)} \tag{36}$$

and likewise for $c^\#(\tau)$, $s^\#(w)$, etc. It can be seen that $\gamma^\#(\Delta) = 2H - 2$ for any $\Delta \geq 0$, $c^\#(\tau) = 2H - 2$ for any $\tau > 0$ —and also for $\tau = 0$ in the persistent process. The slope of the power spectrum is also constant, i.e., $s^\#(w) = \psi^\#(w) = 1 - 2H$, which is negative for the persistent process, positive for the antipersistent process and zero for the purely random process. As footnoted in Table 2, the power spectrum of the sampled process diverges while that of the averaged process exists but does not have a closed expression. Numerically the latter can be evaluated from (34) and generally approaches the continuous-time power spectrum whose expression is given in (T2.5). For small frequencies the approximation is perfect, but as w

approaches $1/2\Delta$ (the Nyquist frequency), the approximation worsens until the slope of $s^{(\Delta)}(w)$ becomes exactly zero at $w = 1/2\Delta$. Interestingly, in the antipersistent process there is a range of intermediate scales (close to $\sim 1/5\Delta$) where the log-log slope becomes much steeper than $1 - 2H$, close to +2. However, by choosing a higher Δ, this range moves to higher frequencies, so that as $\Delta \to 0$, this high-slope area disappears.

Notably, the HK process is the only one in which the log-log derivatives of all second order properties are constant for any Δ, τ and w, thereby defining simple scaling laws. In all other processes these log-log derivatives vary, yet scaling laws are defined asymptotically, i.e., for Δ, τ or w tending to 0 or tending to ∞. In any process other than the HK the respective asymptotic values of the log-log derivatives are different in the two asymptotic cases, 0 and ∞. Interestingly, combining (36) with (9) and (19) we conclude that for Gaussian processes the EPLT is related to the log-log derivative of the climacogram by

$$\varphi(t) = 1 + \tfrac{1}{2}\, \gamma^{\#}(t) \tag{37}$$

This makes EPLT maximization equivalent to maximization of $\gamma^{\#}(t)$. The above results confirm that in the HK process the EPLT is constant, $\varphi(t) = H$, and also enable generalization of the Hurst coefficient for any process by

$$H = \varphi(\infty) = 1 + \tfrac{1}{2}\, \gamma^{\#}(\infty) \tag{38}$$

4.3   The Hybrid Hurst-Kolmogorov process

This process in introduced here in terms of its climacogram, described by the Cauchy-type relationship

$$\gamma(\Delta) = \lambda(1 + (\Delta/\alpha)^{2\kappa})^{\frac{H-1}{\kappa}} \tag{39}$$

Here α and λ are scale parameters with dimensions $[t]$ and $[x]^2$, respectively, H is the Hurst coefficient as in the HK process, a dimensionless scaling parameter in the interval (0, 1), and κ is a second scaling parameter, again dimensionless and in the interval (0, 1), which will be called the fractal parameter. As we will see, H determines the global properties of the process (as $t \to \infty$) and κ determines the local properties (as $t \to 0$).

This process remedies both inconsistencies of the HK process discussed above, while retaining the persistence or antipersistence properties. Specifically, the variance of the instantaneous process is always finite ($\gamma_0 = \gamma(0) = \lambda$), while even for $0 < H < 0.5$ the initial part of the autocovariance function for small lags is positive for all variants of the process (continuous time, discrete time, either sampled or averaged, for a small time interval Δ).

A similar process was proposed by Gneiting and Schlather (2004), which has a Cauchy-type autocorrelation function rather than climacogram. However, the climacogram-based definition in (39) is superior, because it allows for antipersistent processes for $0 < H < 0.5$, which have positive autocovariance for small lags and negative for large lags. (In the process with Cauchy-type autocorrelation function autocovariances are either all positive or all negative). The name "hybrid" for the process defined in (39) reflects the fact that it incorporates both the Markov and the HK processes. Specifically, in the special case with $H = \kappa = 0.5$, HHK is practically indistinguishable from a Markov process (even though not precisely identical, as can be seen by comparing (39) with (T1.2)). Furthermore, as $\alpha \to 0$, the process tends to a pure HK process with the same Hurst coefficient H. Also, for any specified parameter set, HHK exhibits Markov

behaviour for small time scales (if $\kappa = 0.5$, or similar to Markov if $\kappa \neq 0.5$) and Hurst behaviour for large time scales, as seen in Figure 2.

The physical realism of the process is its strongest advantage. This is expressed by the fact that it maximizes entropy production both for small ($t \to 0$) and large ($t \to \infty$) times, as well as by the properties discussed above in comparison with the HK process. Empirical testing of the consistency of the process with nature is not quite easy, as the availability of observed time series of natural processes that have both short observation scale and large sample size is quite limited. Note that, if the available time series had short observation scale and small sample size, the process would look as Markovian, while if it had large sample size but at large discretization scale, HHK will look indistinguishable from the pure HK. There is one exception offered by laboratory measurements of turbulence, which can have high temporal resolution as well as very large sample size. And indeed the behaviour observed from time series of turbulence is consistent with that of an HHK model (see Koutsoyiannis 2013a, who fitted a special case of the HHK model to turbulence data).

The expressions of most of the HHK second-order characteristics are too complex, but the framework provided in Sections 2 and 3 suffices to evaluate numerically all of its properties based on (39). As an exception, the autocovariance in continuous time has a rather simple expression:

$$c(\tau) = \gamma(\tau) \frac{1 + (3H + 2\kappa H - 2\kappa - 1)(\tau/a)^{2\kappa} + H(2H-1)(\tau/a)^{4\kappa}}{1 + 2(\tau/a)^{2\kappa} + (\tau/a)^{4\kappa}} \tag{40}$$

A further important feature of this process is that it allows explicit control of the asymptotic behaviour of all properties ($\gamma(\Delta)$, $c(\tau)$, s(w), $\psi(w)$) at both ends (for Δ, τ, $w \to 0$ or ∞), which are different at each end, opposite to the HK process, which implies simple scaling laws as described above. Despite the expressions of these properties being involved, the asymptotic properties are easy to calculate, i.e., $\gamma^{\#}(\infty) = c(\infty) = 2H - 2$, $\gamma^{\#}(0) = c^{\#}(0) = 0$, $s^{\#}(0) = \psi^{\#}(0) = 1 - 2H$, $s^{\#}(\infty) = \psi^{\#}(\infty) = -1 - 2\kappa$. From the last two values we conclude that the fractal properties of the process, which are local properties expressing the asymptotic behaviour for small scales or large frequencies, depend on the parameter κ only (with the fractal dimension being equal to 2 – κ; cf. Gneiting and Schlather 2004). In contrast, the behaviour for large scales or small frequencies depends on the Hurst coefficient H only.

5 Methodology for stochastic simulation

5.1 The general algorithm

The so-called symmetric moving average (SMA) method (Koutsoyiannis 2000) can directly generate time series with any arbitrary autocorrelation function provided that it is mathematically feasible. The SMA method consists of the following generation equation:

$$\underline{x}_i = \sum_{l=-\infty}^{\infty} a_{|l|} \underline{v}_{i+l} \tag{41}$$

where a_j are coefficients calculated from the autocovariance function and $\underline{v}_i$ is white noise averaged in discrete-time (and not necessarily Gaussian; see section 6.4). Assuming that we work for the sampled discrete-time process with autocovariance $\check{c}_j$ and power spectrum $\check{s}_\mathrm{d}$ (the

procedure is the same for the averaged process), it has been shown (Koutsoyiannis 2000) that the Fourier transform $s_d^a(\omega)$ of the a_l series of coefficients is related to the power spectrum of the discrete time process as

$$s_d^a(\omega) = \sqrt{2\check{s}_d(\omega)} \tag{42}$$

Thus, to calculate a_l we first determine $s_d^a(\omega)$ from the power spectrum of the process and then, using an equation analogous to (35) we estimate all coefficients a_l. In other words, the generation of time series from any possible autocorrelation function is a matter of applying a couple of times the Fourier transform to estimate some coefficients. Note that the coefficients are internal constants of the model, not model parameters.

It is expected that the coefficients a_l will decrease with increasing l and will be negligible beyond some q ($l > q$), so that we can truncate (41) to read

$$\underline{x}_i = \sum_{l=-q}^{q} a_{|l|}\underline{v}_{i+l} \tag{43}$$

This would introduce some truncation error, as the resulting autocovariance function

$$\tilde{c}_j = \sum_{l=-q}^{q-j} a_{|l|}a_{|j+l|}, \qquad j = 0,1,2,\dots \tag{44}$$

would not be identical to $\check{c}_j$. In processes with short-range dependence the error is small even for a relatively small truncation limit q (cf. section 5.2.3). However, if the dependence is of long-range type with high H, then there may be a problem. While in hydrology the H values are rather moderate, recent studies have estimated values $H > 0.9$ for some geophysical series, e.g., Earth's temperature (Koutsoyiannis and Montanari, 2007; Koutsoyiannis 2013a, Markonis and Koutsoyiannis 2013). In such cases q should be high to achieve good approximation of the autocovariance function $\check{c}_j$ and even of the variance $\check{c}_0$. For example, Koutsoyiannis (2002) suggested $q = 250\ 000$ if $H = 0.9$. This entails high computational burden.

However, a small q is possible, if we restrict our interest to the preservation of a moderate number of autocovariance terms $\check{c}_j$, say, two to three orders of magnitude. Our desideratum in this case is to restrict q in the same order of magnitude as that of the required number of autocovariance terms. Koutsoyiannis (2000) provided an iterative numerical algorithm for that problem, which does not make use of the Fourier transform, but uses an optimization procedure. While that algorithm is general enough, an optimization for, say, 1000 terms a_l, which become the control variables in the optimization problem, may be cumbersome.

Here we study a different procedure, which is direct and explicitly considers the truncation in the Fourier analysis. Specifically, instead of deriving a_l from $s_d^a(\omega)$, we consider a modified spectrum $\tilde{s}_d^a(\omega)$ of a_l which should give $\tilde{c}_j = 0$ for $j > q$, while retaining $\tilde{c}_j$ close to the actual c_j for small lags. We thus estimate $\tilde{s}_d^a(\omega)$ as the sum of two terms:

$$\tilde{s}_d^a(\omega) = s_d^a(\omega)(1 - \text{sinc}(2\pi\omega q)) + \varepsilon\,\text{sinc}(2\pi\omega q) \tag{45}$$

where ε is a constant. The rationale of the first term is that it represents a high-pass filtered process (Papoulis 1991, p. 325; cf. with (32) noticing that this is a low pass filter and that s_d^a is proportional to the square root of $\check{s}_d$), which suppresses low frequencies or high time scales ($> q$). The second term, whose inverse Fourier transform is the rectangular function (so that, with

respect to equation (34), it results in coefficients equal to $\varepsilon/(4q+2)$ for lags $\leq q$ and 0 for lags $> q$), replaces the truncated part with a constant component for the lags of interest. In conclusion, both terms result in virtually zero coefficients for lags $> q$.

Let a_l' be the coefficients a_l corresponding to the first term of (45), while those corresponding to the second term are equal to each other for all l, $a'' = \varepsilon/(4q+2)$. Their sum

$$a_l = a_l' + a'' \tag{46}$$

will give the final coefficients a_l to be applied for the generation. The coefficients a_l' can be calculated from (35) by using $s_d^a(\omega)(1 - \text{sinc}(2\pi\omega q))$ in place of $\check{s}_d(\omega)$. The constant a'' requires the weight ε to be determined. Applying (44) for $j = 0$, so as to estimate (and preserve exactly) the variance, $\check{\gamma}_0 \equiv \check{c}_0$ we obtain

$$\check{\gamma}_0 = \sum_{l=-q}^{q} a_{|l|}^2 = \sum_{j=-q}^{q} \left(a_{|l|}' + a''\right)^2 \tag{47}$$

Solving for a'', after algebraic manipulations we find:

$$a'' = \sqrt{\frac{\check{\gamma}_0 - \Sigma\alpha'^2}{2q+1} + \left(\frac{\Sigma\alpha'}{2q+1}\right)^2} - \frac{\Sigma\alpha'}{2q+1} \tag{48}$$

where $\Sigma\alpha' := a_0' + 2(a_1' + \cdots + a_q')$ and $\Sigma\alpha'^2 := a_0'^2 + 2(a_1'^2 + \cdots + a_q'^2)$.

This method of truncation with adjustment will be referred to as method 1. A further approximation which constitutes a simpler calculation denoted as method 2, is to use $s_d^a(\omega)$ directly (rather than $s_d^a(\omega)(1 - \text{sinc}(2\pi\omega q))$) to estimate a_l' (by taking the inverse transform of $s_d^a(\omega)$ as in (35) but using $s_d^a(\omega)$ in place of $\check{s}_d(\omega)$) only for $j = 0$ to q, set all terms beyond q equal to 0, and then estimate a'' again from (48).

In both methods, it is expected that $\check{\gamma}_0 - \Sigma\alpha'^2 > 0$. If due to numerical imperfections it happens that $\check{\gamma}_0 - \Sigma\alpha'^2 < 0$, then it advisable not to use equations (45)-(48), but rather modify a_l' by proportional adjustment, i.e.,

$$a_l = a_l' \sqrt{\check{\gamma}_0/\Sigma\alpha'^2} \tag{49}$$

This adjustment will be referred to as method 3.

5.2 Application

The preservation of the autocovariance function of any process by the SMA algorithm is theoretically guaranteed and therefore there is no need to test it through applications (although some applications can be found in Koutsoyiannis 2000). What is more meaningful to examine is the influence of truncation and the ability of the proposed methods of adjustment to deal with the truncation error. This is done in the following subsections, where we use the three models described above with parameter values as listed in the caption of Figure 2 (in some cases additional parameter sets are tested). It is recalled that in the case illustrated, the Hurst coefficient for the HK and the HHK process is relatively high, around 0.8.

5.2.1 HHK process

Figure 3(a) presents the autocovariance function of the HHK process with and without truncation at q = 32 in comparison with the theoretical one. The 32 terms used are too few for a process with LTP but serve well illustration. If we make truncation without any adaptation, then even the first 32 autocovariance terms are not well approximated and there is a 3.3% underprediction of the variance (not shown in figure). However, truncation methods 1 and 2 remedy this problem, with method 1 performing somewhat better than method 2. Figure 3(c) shows that to correct the autocovariance function the sequence of a_l needs to be raised in its right tail. Figure 3(b, d) shows that this raise adds some periodicity to the power spectrum as well as to the Fourier transform of the sequence of a_l. Generally, the results are quite satisfactory and obviously they would be more satisfactory if more a_l terms were used (q > 32).
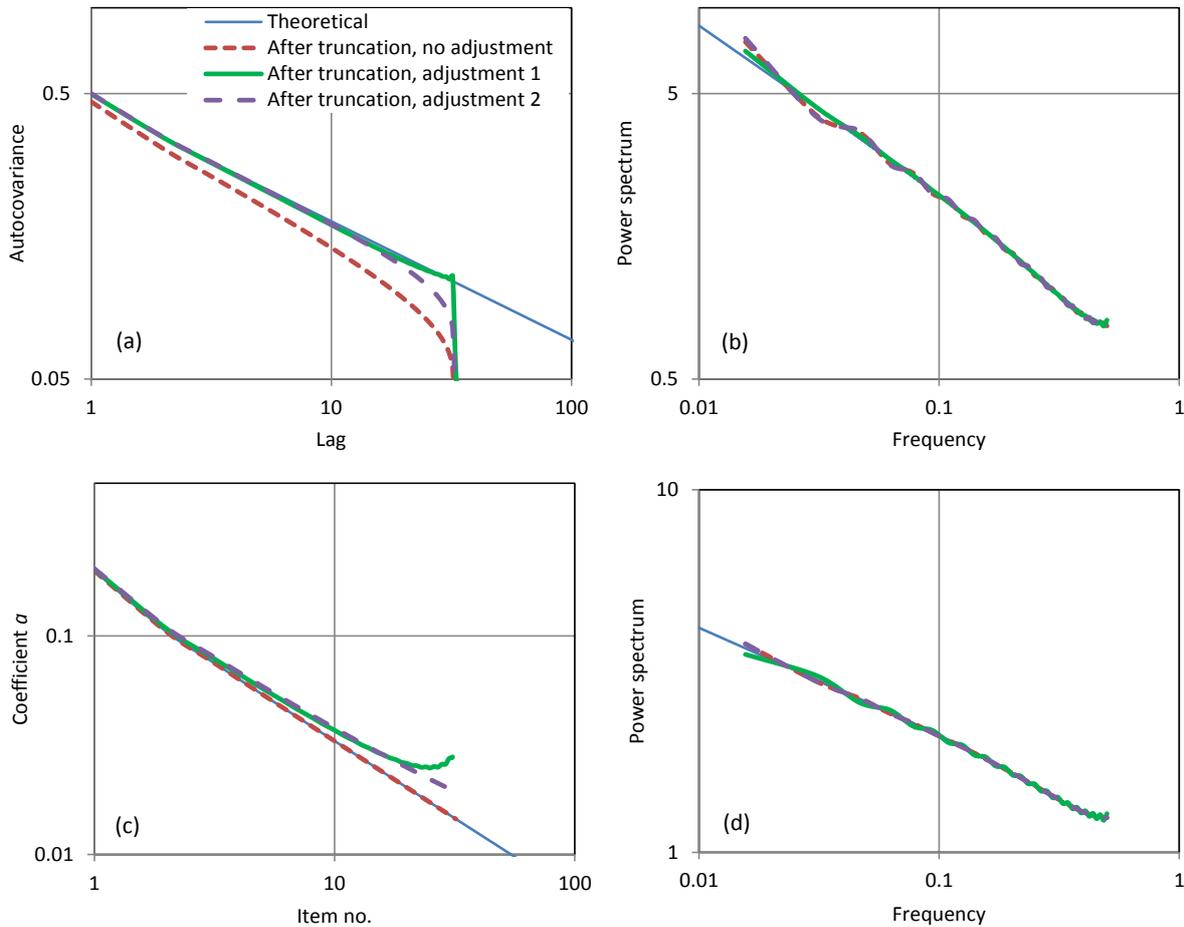
Figure 3 Comparison of the theoretical properties of the HHK process (discrete time, averaged) with parameter values as shown in Figure 2 with the properties after truncation at q = 32: (a-b) autocovariance and power spectrum of the process; (c-d) calculated coefficients a_l and their Fourier transform.

5.2.2 HK process

The application of the same method to the HK process gave quite similar results, so similar that they would be indistinguishable from those in Figure 3 and thus not necessary to reproduce here. The reason that they look similar is that for the time scales and frequencies that can be modelled with q = 32 terms, the HK process is indistinguishable from the HHK.

However, the simplicity of the HK process allows presenting a more systematic comparison. Moreover, this simplicity allows an analytical determination of the sequence of the

untruncated sequence of a_l. Specifically, as shown in Koutsoyiannis (2002) the coefficients a_l can be approximated as

$$a_l = \frac{\sqrt{(2-2H)\gamma(\Delta)}}{3-2H}(|l+1|^{H+0.5} + |l-1|^{H+0.5} - 2|l|^{H+0.5}) \tag{50}$$

where the multiplying term in the right hand side was found after a numerical investigation, so that the sum of all a_j^2 equals $\gamma(\Delta)$ as it should. Here, by refining the approximation using the same method, the following improvement was obtained:

$$a_l = \sqrt{\frac{2(1-H)\gamma(\Delta)}{(3/2-H)^2 + 0.2(1/2-H)^2}}\left(\frac{|l+1|^{H+0.5} + |l-1|^{H+0.5}}{2} - |l|^{H+0.5}\right) \tag{51}$$
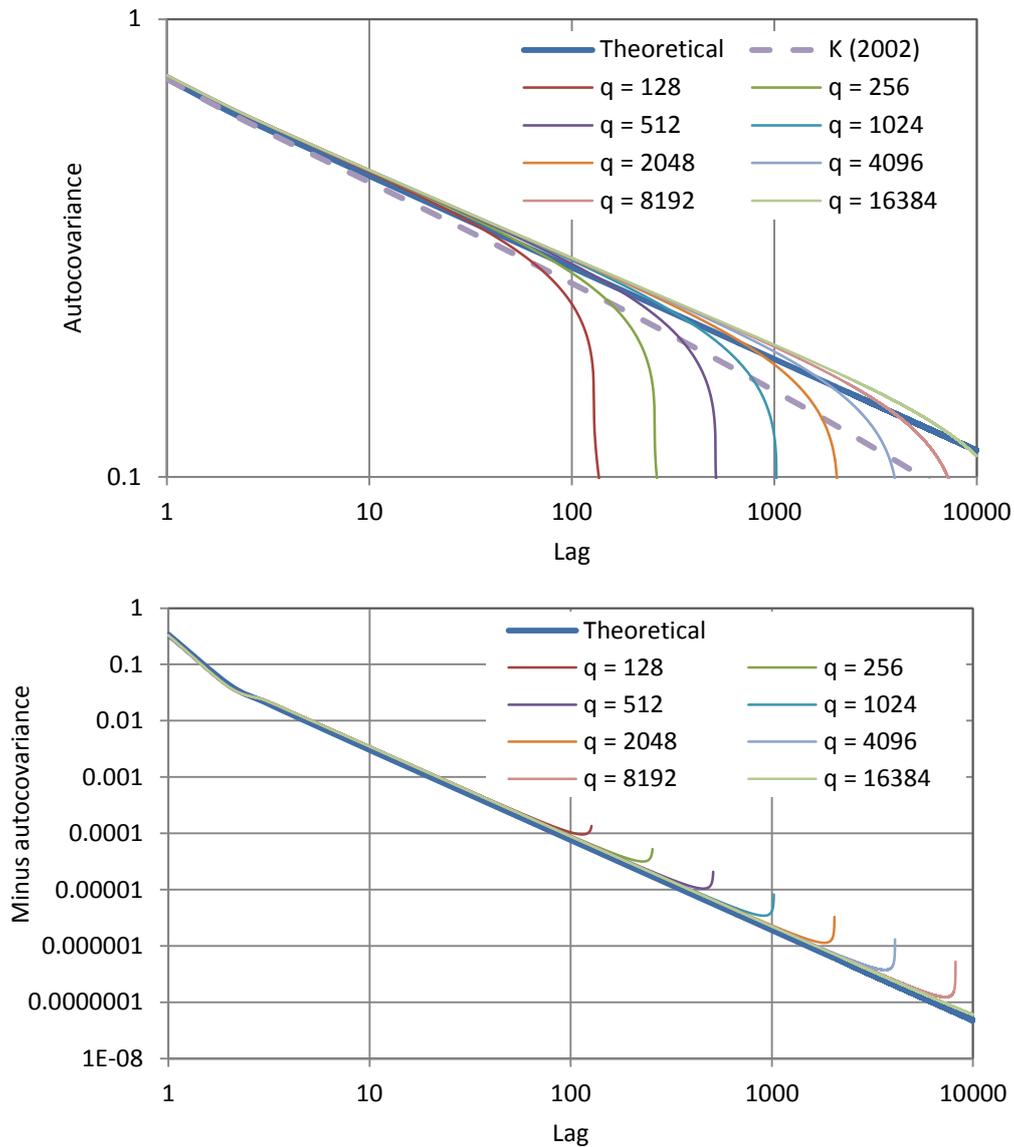
Figure 4 Comparison of the theoretical autocovariance functions of the HK proses with those obtained after truncation with the indicated number q of a_l terms: (upper) persistent case with $H = 0.9$ and adjustments with method 2; K (2002) refers to the results in Koutsoyiannis (2002) without adjustment using q = 250 000 terms; (lower) antipersistent case with $H = 0.2$ and adjustments with method 3. The variance $\gamma(\Delta)$ is 1 in both cases.

For $H > 0.5$, the proximity of the power spectrum of the averaged process with that of the continuous-time process (equation (T2.5) in Table 2) allows the theoretical derivation of a more consistent equation, i.e.,

$$a_l = \sqrt{\frac{2\Gamma(2H+1)\sin(\pi H)\,\gamma(\Delta)}{\Gamma^2(H+3/2)\,(1+\sin(\pi H))}} \left(\frac{|l+1|^{H+0.5} + |l-1|^{H+0.5}}{2} - |l|^{H+0.5} \right) \tag{52}$$

For $H < 0.5$ the proximity is not good and thus equation (52) does not perform well. However, equation (51) is a very good approximation and is also valid for $H < 0.5$.

Equation (51) is readily applied with adjustment method 2 or 3, without any need to calculate a Fourier transform. Results of the application are shown in Figure 4 for a persistent ($H = 0.9$) and an antipersistent ($H = 0.2$) case for a varying number of terms q. As shown in the figure, the method performs quite well giving good approximation of the autocorrelation function for lags approaching q and outperforms the case where no adjustments are used, even if in the latter case a very high number of terms were used.

5.2.3 Markov process

Coming again to the application designated in Figure 2 but now with the Markov model, some results are depicted in Figure 5. The results support the conclusion that, due to the exponential decay of autocovariance, in this model only a few terms a_l are required (in this case, even $q = 8$ is too much, given that autocovariance values of the order of 10^{-3} can be neglected) and that no adjustment due to the truncation is actually needed. While the results in Figure 5 are for the averaged process, the application was made also for the sampled process with quite similar results which are aimless to reproduce here.
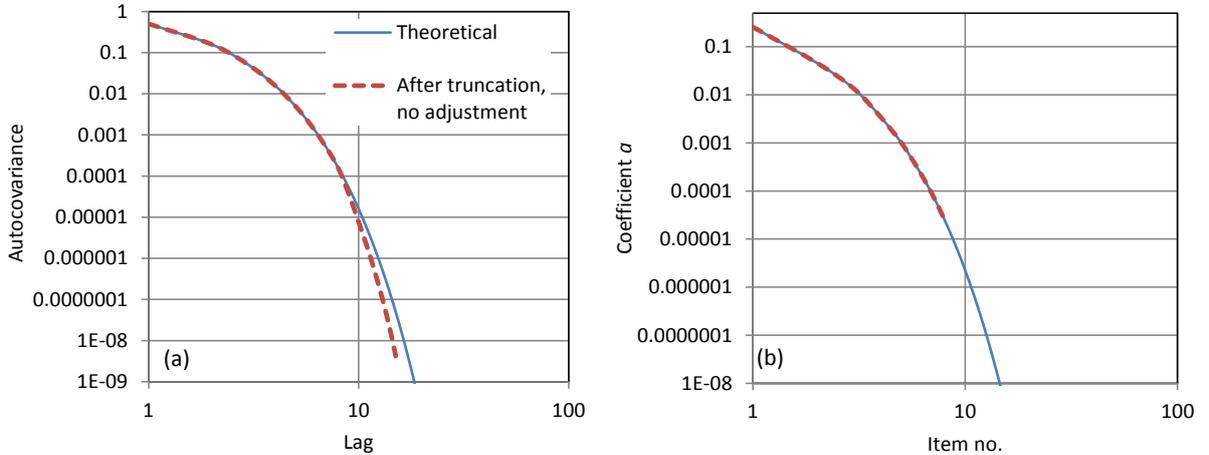

Figure 5 (a) Comparison of the theoretical autocovariance function of the Markov process (discrete time, averaged) with parameter values as shown in Figure 2 with that after truncation at $q = 8$ without adjustment; (b) corresponding coefficients a_l.

One may ask whether this framework offers anything for the Markov case in comparison to the ARMA(1, 1) model (for the averaged process) or the AR(1) model (for the sampled process). It should thus be made clear that these two classical models are useful in simulation and need not be replaced by the current framework. The strength of the current framework is its applicability to *any* stochastic process in the same form. In addition, the SMA scheme is simpler and more convenient for any other stochastic process except these two. To see this, we consider two examples, the AR(2) model and a bivariate AR(1) model. In each of these models the

generation expression will give the variable $\underline{x}_i$ as a weighted sum of two past variables (for two and one time step before, respectively) and some white noise variables, while the SMA model in the form (41) or (43) will be a weighted sum of white noise variables only. Now, if we try to reproduce the fourth moment (kurtosis) of the process and thus raise the generation expression to the fourth power, in the AR(2) model and the bivariate AR(1) model some joint fourth-order moments will emerge, which are not possible to handle. In the SMA model no such terms will appear.

6   Final notes

6.1   Model identification and fitting

The typical procedure to choose an appropriate stochastic model, based on an observed time series, is to construct the empirical autocorrelogram of the time series and estimate which stochastic process (e.g., of AR or ARMA type) is suitable and how many autocorrelation terms should be preserved. It is not too difficult to illustrate that this technique can completely distort the underlying process. Figure 6(a) depicts the autocorrelogram of a time series with length 100, which does not seem to have any relationship with the theoretical autocorrelation function of the model from which was constructed. The latter is the HHK one with parameters as in the caption of Figure 2. Clearly, the empirical autocorrelation does not give any hint that the time series stems from a process with LTP. With that autocorrelogram one would conclude that an AR(1) model with a lag-1 autocorrelation of about 0.4 would be appropriate. Note that, considering the annual scale which cancels out seasonality, a length 100 of a time series is unusually high in hydrology and in other geophysical fields. For shorter time series the results would be even more disappointing.

The reasons for the failure of the autocorrelogram to capture the real behaviour of the process are two. First, from (8) and (23) it is seen that the autocorrelation (autocovariance standardized by variance) is by nature the second derivative of the climacogram. Estimation of the second derivative from data is too uncertain and makes a very rough graph. The second reason is that autocorrelation is biased, as explained in Koutsoyiannis (2003).
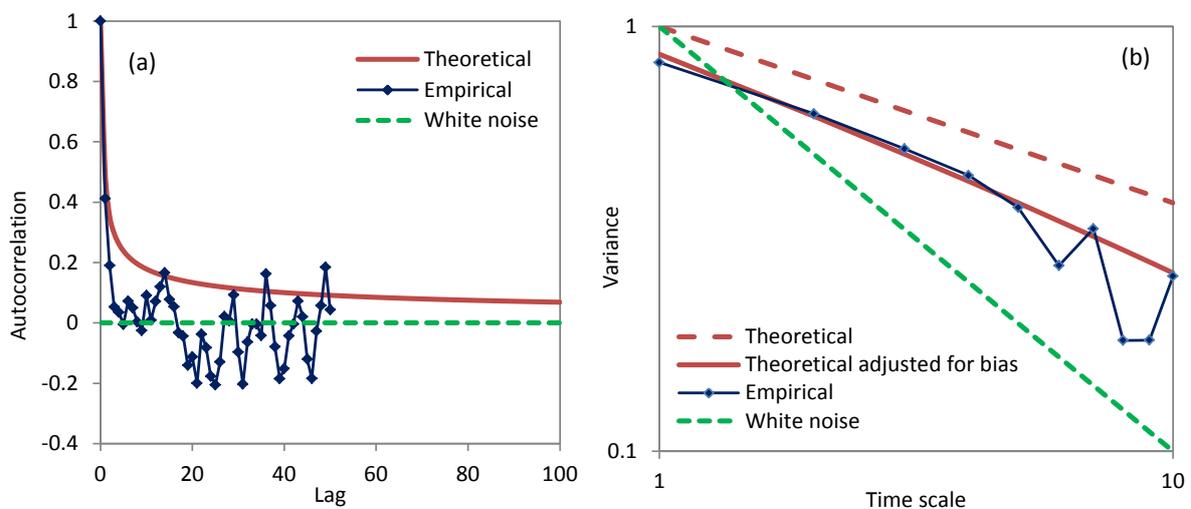

Figure 6 (a) Autocorrelogram and (b) climacogram of a time series of 100 terms generated from the HHK model with parameters as in the caption of Figure 2 and $q = 128$.

It is, thus, more natural to directly use the climacogram instead of the autocorrelogram for model identification. For our example time series, this is illustrated in Figure 6(b), which indicates that the long-term persistence is well captured by the empirical climacogram and the parameter H is correctly estimated ($H = 0.79$, based on the method presented in Koutsoyiannis 2003, and Tyralis and Koutsoyiannis 2011).

Here it should be noted that in stochastic processes, particularly ones with LTP, almost all classical statistical estimators are biased and the climacogram is not excepted from this rule. However, the bias of the climacogram is easy to calculate a priori and actually this has been made in the plot of Figure 6(b), where the empirical climacogram is not compared with the theoretical one but with the expected one considering bias. The big distance between the two climacograms in Figure 6(b), theoretical and adjusted for bias, indicates how big the bias can be and suggests the importance of considering the bias even for model identification, let alone model fitting.

Specifically, the bias in the climacogram can be considered as follows. As shown in Koutsoyiannis (2011), assuming that we have $n = T/\Delta$ observations of the averaged process $\underline{x}_i^{(\Delta)}$, where the observation period T is an integer multiple of Δ, the expected value of the empirical (sample) climacogram $\hat{\gamma}(\Delta)$ is

$$\mathrm{E}\left[\underline{\hat{\gamma}}(\Delta)\right] = \frac{\gamma(\Delta) - \gamma(T)}{1 - \Delta/T} = \frac{1}{1 - \Delta/T}\left(\frac{\Gamma(\Delta)}{\Delta^2} - \frac{\Gamma(T)}{T^2}\right) = \frac{1 - \gamma(T)/\gamma(\Delta)}{1 - \Delta/T}\gamma(\Delta) \tag{53}$$

where $\hat{\gamma}(\Delta)$ is given by the standard statistical estimator, i.e.,

$$\underline{\hat{\gamma}}(\Delta) := \frac{1}{n-1}\sum_{i=1}^{n}\left(\underline{x}_i^{(\Delta)} - \underline{\bar{x}}^{(\Delta)}\right)^2 = \frac{1}{T/\Delta - 1}\sum_{i=1}^{T/\Delta}\left(\underline{x}_i^{(\Delta)} - \underline{\bar{x}}^{(\Delta)}\right)^2 \tag{54}$$

whereas $\bar{x}^{(\Delta)}$ is the standard (unbiased) statistical estimator of the common mean μ of the instantaneous process $\underline{x}(t)$ as well as of the discrete processes $\underline{x}_i$ and $\underline{x}_i^{(\Delta)}$, i.e.,

$$\underline{\bar{x}}^{(\Delta)} := \frac{1}{n}\sum_{i=1}^{n}\underline{x}_i^{(\Delta)} = \frac{\underline{X}(T)}{n\Delta} = \frac{\underline{X}(T)}{T} = \underline{x}_1^{(T)} \tag{55}$$

If the observed process is the sampled ($\underline{x}_i$) at time steps Δ, rather than the averaged one ($\underline{x}_i^{(\Delta)}$), then (53) becomes (Koutsoyiannis 2013b):

$$\mathrm{E}\left[\underline{\hat{\tilde{\gamma}}}^{(k)}\right] = \frac{\check{\gamma}^{(k)} - \check{\gamma}^{(n)}}{1 - 1/n} = \frac{1 - \check{\gamma}^{(n)}/\check{\gamma}^{(k)}}{1 - 1/n}\check{\gamma}^{(k)} \approx \frac{1 - \gamma(T)/\gamma(k\Delta)}{1 - \Delta/T}\check{\gamma}^{(k)} \tag{56}$$

where

$$\underline{\hat{\check{\gamma}}}^{(k)} := \frac{1}{n-1}\sum_{i=1}^{n}\left(\underline{\check{x}}_i^{(k)} - \underline{\bar{x}}\right)^2 \tag{57}$$

$$\underline{\bar{x}} := \frac{1}{n}\sum_{i=1}^{n}\underline{x}_i = \underline{\check{x}}_1^{(n)} \tag{58}$$

$$\underline{\check{x}}_i^{(k)} := (\underline{X}_{ik} - \underline{X}_{(i-1)k})/\Delta \tag{59}$$

Notice that in this case $\breve{\gamma}^{(1)} = \gamma(0)$.

It becomes clear from the above equations that direct estimation of the variance $\gamma(\Delta)$, let alone that of the instantaneous process $\gamma(0)$, is not possible merely from the data. We need to know the ratio $\gamma(T)\,/\,\gamma(\Delta)$ and thus we should assume a stochastic model which evidently influences the estimation of $\gamma(\Delta)$. 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 the instantaneous one $\gamma(0)$. This applies to both cases of discretization, sampled and averaged, and therefore the need to assume a model in order to estimate the process statistics is imperative, irrespective of the type of observations, instantaneous or averaged. One may counter-argue that the standard statistical estimation offered in statistics books is free of model assumptions in estimating the true (population) parameters from data. However, this is not correct: a model is always assumed. In classical statistics the assumption is the independence in time and hence the model is the white noise process. In fact, such independence can hardly be a realistic assumption for natural processes.

In other words, model identification and fitting cannot be separated from each other. The climacogram, thanks to its analytical expression of bias and its smother shape, in comparison to that of the autocorrelogram, offers a basis to perform both tasks simultaneously. Once the model is identified, at a later step its parameters can be fine-tuned by using additional methods, for example the maximum likelihood method. Tyralis and Koutsoyiannis (2011) have streamlined the maximum likelihood method for the HK process, in which the autocovariance matrix introduced in calculations is unavoidably huge. The same method can be adapted for the HHK process without difficulty. In addition, Tyralis and Koutsoyiannis (2011) compared several parameter estimation methods and concluded that the maximum likelihood is generally superior, while other methods they examined performed much worse, except two climacogram-based methods. These latter performed only slightly worse than the maximum likelihood method but have the advantage of combining model identification and parameter estimation.

The maximum likelihood method (in most cases approximated by a least squares objective function) has also been in common use in ARIMA-based stochastic modelling. Obviously, this is a much more reliable practice in comparison to fitting the model through empirical estimates of autocorrelations, based on the Yule–Walker equations. Yet in standard modelling practice, the empirical autocorrelation function remains the primary tool for model identification and, as explained above, this is not a good modelling practice. If the model identified is inappropriate, then seeking the proper values of its parameters is less important and perhaps misleading. In particular, if the actual process is characterized by LTP, then a time series may contain rich patterns even if the underlying process is quite simple and parsimonious (Papalexiou *et al.* 2011). However, if the model identification failed to detect the LTP, then it is quite likely that it would need a complicated model with many parameters to describe these rich patterns.

A more thorough investigation of problems related to model identification and parameter estimation has been made by Dimitriadis and Koutsoyiannis (2015) who compared the climacogram, autocovariance and power spectrum and concluded that the former is more advantageous.

6.2 Multivariate modelling

Most applications of hydrological simulation require multivariate setting as it is usually important to simulate cross-correlation coefficients among different processes. While the above discourse is for single-variate setting, the multivariate extension is direct and easier than in other modelling approaches, as already indicated in section 5.2.3. The details required for this extension are given in Koutsoyiannis (2000).

6.3 Prediction

Whenever future prediction, rather than unconditional simulation, is required, while past observations are available, then those observations can be used to condition the distribution of the future realizations of the process $x(t)$. A general method to handle this case, which has been described in Koutsoyiannis (2000), can easily be implemented within a Monte Carlo prediction framework (Efstratiadis *et al.* 2014). Such type of prediction is vital in water management applications (Koutsoyiannis *et al.* 2003). Conditioning on past data within a Monte Carlo framework is particularly important when LTP is present, as clarified by Koutsoyiannis *et al.* (2007) using a classical statistical framework and by Tyralis and Koutsoyiannis (2014) using a Bayesian framework.

6.4 Non-Gaussian distributions

If the marginal distribution of $x(t)$ is Gaussian, then no consideration additional to already exposed is required. However, non-Gaussian distributions (mostly positively skewed and possibly heavy tailed), are very common in hydrology and their emergence has some theoretical justification again based on entropic considerations (Koutsoyiannis 2005a, 2014). For non-Gaussian distributions two strategies are commonly used. The first is to convert the process, using a nonlinear transformation, to a Gaussian process (e.g. Koutsoyiannis *et al.* 2008). This is not quite easy as the transformation cannot be invariant with respect to the time scale (Lombardo *et al.* 2012). The second strategy is to explicitly preserve, at the time scales of interest, higher-order marginal moments, in particular the third moment or, equivalently, the skewness coefficient. The SMA model allows easy handling of high-order moments, as already indicated in section 5.2.3. The details of the latter technique have been studied in Koutsoyiannis (1999, 2000). It should be emphasized, though, that high-order moments cannot be reliably estimated from data (Lombardo *et al.* 2014) and hence a model for the marginal distribution is again required to better specify the values of the moments to be preserved.

6.5 Intermittency

Hydrological processes are often intermittent at small time scales (e.g. rainfall). Intermittency is not easy to handle. Some techniques have been implemented in Koutsoyiannis *et al.* (2003) and Efstratiadis *et al.* (2014). However, they contain some *ad hoc* elements and therefore further research is needed to develop a fully consistent theoretical approach.

6.6 Cyclostationarity

At sub-annual time scales, hydrological and other geophysical processes exhibit periodicity in their statistical properties. This behaviour is also known as seasonality or cyclostationarity. A

popular technique to deal with this problem is to "stationarize" the time series by standardizing by the mean and standard deviation of each separate "season" (e.g., month or day). However, it should be kept in mind that such "stationarization" applies only to the marginal distribution (in particular, only to the first two moments of the marginal distribution if this popular linear transformation procedure is used). It has no effect on the dependence structure, which may also be characterized by cyclostationarity. Two different techniques can deal more effectively with this problem. One (Langousis and Koutsoyiannis 2006) generates times series of each "season" separately, thus accounting for long-term (multiyear) persistence, and then modifies the separate time series to establish proper autocorrelations between consecutive months. The other technique, more widely implemented so far in hydrology, uses a disaggregation approach. This first generates the annual time series, thus accounting for LTP, and then disaggregates them into sub-annual (e.g., monthly or daily) quantities (Koutsoyiannis 2001; Efstratiadis *et al.* 2014).

7   Conclusions

It is more natural and better from many aspects to formulate the stochastic processes that are used in modelling of hydrological (and other geophysical) processes in continuous time rather than using the recipes of discrete-time processes (e.g. AR, ARMA, etc.) of the literature. The theoretical framework provided here allows derivation, with easily applicable equations, of the stochastic properties in discrete time, whether these correspond to sampling of the continuous-time process at equidistant times or averaging it at time intervals with equal length.

Among continuous-time processes, the HHK process introduced and studied here, proves to have such properties that make it a good candidate for several hydrological applications. First of all, it is associated with maximum entropy production both at small and large time scales, thus combining Markovian behaviour at small scales and Hurst-Kolmogorov behaviour at large scales. It also incorporates both these processes as special cases. Second it is parameter parsimonious as, in addition to marginal distributional properties, it uses one time-scale parameter and up to two dimensionless parameters to define the dependence structure. One of the latter (κ) determines its behaviour on small scales (the fractal properties) and the other one (H) the long-term persistence (the HK properties). Third, it is free of some unnatural properties of the HK process *per se*, namely the infinite variance of the instantaneous process and the discontinuity of its autocovariance function in the case of antipersistence ($H < 1/2$), which results in negative autocorrelation even for arbitrarily short lags.

Despite those shortcomings, the HK process which is also thoroughly studied here provides a perfect approximation of the HHK process for intermediate and large time scales, being even more parsimonious than the HHK process and enabling analytical expressions of its properties in continuous time and of most of those in discrete time (for the averaged process).

Whichever of the three above processes is used, or even anyone else, the general simulation methodology (SMA) allows easy application and also correction of possible truncation errors. The results of the case studies examined show that these errors can be almost eliminated by the adjustment methodologies studied.

As regards model identification and fitting, an example is provided that illustrates the inability to specify the appropriate model if the common practice based on the empirical autocorrelogram is used. The framework supports the use of the climacogram for this purpose. While, like other stochastic tools, the climacogram also involves estimation bias, its bias can be

determined analytically and included in the model identification and fitting. Furthermore, it is shown that the latter two tasks are closely related to each other. Even fundamental statistical estimators, such as the sample variance, cannot be applied without specifying a model type.

Overall, the methodology developed is not demanding in terms of computational means. For instance, all examples presented here have been prepared in common spreadsheet software. However, it is planned that all presented advances be implemented in the next version of the stochastic simulation framework *Castalia* (Efstratiadis *et al.* 2014).

References

Dimitriadis, P., and Koutsoyiannis, D., 2015. Climacogram vs. autocovariance and power spectrum in stochastic modelling for Markovian and Hurst-Kolmogorov processes. *Stochastic Environmental Research and Risk Assessment* (in press).

Efstratiadis, A., *et al.*, 2014. A multivariate stochastic model for the generation of synthetic time series at multiple time scales reproducing long-term persistence. *Environmental Modelling and Software*, 62, 139–152, doi:10.1016/j.envsoft.2014.08.017.

Gauch, H. G., Jr., *Scientific Method in Practice*, Cambridge University Press, Cambridge, 2003.

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

Gneiting, T., Sevciková, H., and Percival, D.B., 2012. Estimators of fractal dimension: Assessing the roughness of time series and spatial data. *Statistical Science*, 27(2), 247–277.

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

Hurst, H.E., 1951. Long term storage capacities of reservoirs. *Transactions of the American Society of Civil Engineers*, 116, 776–808.

Jaynes, E.T. 1957. Information theory and statistical mechanics, *Physical Review*, 106 (4), 620-630.

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

Koutsoyiannis, D., 1999. Optimal decomposition of covariance matrices for multivariate stochastic models in hydrology. *Water Resources Research*, 35 (4), 1219–1229.

Koutsoyiannis, D., 2000. A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. *Water Resources Research*, 36 (6), 1519–1533.

Koutsoyiannis, D., 2001. Coupling stochastic models of different time scales. *Water Resources Research*, 37 (2), 379–392.

Koutsoyiannis, D., 2002. The Hurst phenomenon and fractional Gaussian noise made easy. *Hydrological Sciences Journal*, 47 (4), 573–595.

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

Koutsoyiannis, D., 2005a. Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling. *Hydrological Sciences Journal*, 50 (3), 381–404.

Koutsoyiannis, D., 2005b. Uncertainty, entropy, scaling and hydrological stochastics, 2, Time dependence of hydrological processes and time scaling. *Hydrological Sciences Journal*, 50 (3), 405–426.

Koutsoyiannis, D., 2010. HESS Opinions: "A random walk on water". *Hydrology and Earth System Sciences*, 14, 585–601.

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

Koutsoyiannis, D., 2013a. Hydrology and Change. *Hydrological Sciences Journal*, 58 (6), 1177–1197.

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

Koutsoyiannis, D., 2014. Entropy: from thermodynamics to hydrology. *Entropy*, 16 (3), 1287–1314.

Koutsoyiannis, D., Efstratiadis, A., and Georgakakos, K., 2007. Uncertainty assessment of future hydroclimatic predictions: A comparison of probabilistic and scenario-based approaches. *Journal of Hydrometeorology*, 8 (3), 261–281.

Koutsoyiannis, D., *et al.*, 2003. A decision support system for the management of the water resource system of Athens. *Physics and Chemistry of the Earth*, 28 (14-15), 599–609.

Koutsoyiannis, D., *et al.*, 2009. HESS Opinions: "Climate, hydrology, energy, water: recognizing uncertainty and seeking sustainability". *Hydrology and Earth System Sciences*, 13, 247–257.

Koutsoyiannis, D., and Montanari, A., 2007. Statistical analysis of hydroclimatic time series: Uncertainty and insights. *Water Resources Research*, 43 (5), W05429, doi:10.1029/2006WR005592.

Koutsoyiannis, D., Onof, C., and Wheater, H.S., 2003. Multivariate rainfall disaggregation at a fine timescale. *Water Resources Research*, 39 (7), 1173, doi:10.1029/2002WR001600.

Koutsoyiannis, D., Yao, H., and Georgakakos, A., 2008. Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods. *Hydrological Sciences Journal*, 53 (1), 142–164.

Langousis, A., and Koutsoyiannis, D., 2006. A stochastic methodology for generation of seasonal time series reproducing overyear scaling behaviour. *Journal of Hydrology*, 322, 138–154.

Lombardo, F., Volpi, E., and Koutsoyiannis, D., 2012. Rainfall downscaling in time: Theoretical and empirical comparison between multifractal and Hurst-Kolmogorov discrete random cascades, *Hydrological Sciences Journal*, 57 (6), 1052–1066.

Lombardo, F., *et al.*, 2014. Just two moments! A cautionary note against use of high-order moments in multifractal models in hydrology. *Hydrology and Earth System Sciences*, 18, 243–255.

Mandelbrot, B.B. and van Ness, J.W., 1968. Fractional Brownian motions, fractional noises and applications. *SIAM review*, 10 (4), 422–437.

Markonis, Y., and Koutsoyiannis, D. 2013. Climatic variability over time scales spanning nine orders of magnitude: Connecting Milankovitch cycles with Hurst–Kolmogorov dynamics. *Surveys in Geophysics*, 34 (2), 181–207.

Montanari, A., and Koutsoyiannis, D., 2012. A blueprint for process-based modeling of uncertain hydrological systems. *Water Resources Research*, 48, W09555, doi:10.1029/2011WR011412.

Montanari, A., *et al*., 2013. "Panta Rhei – Everything Flows", Change in Hydrology and Society – The IAHS Scientific Decade 2013-2022. *Hydrological Sciences Journal*, 58 (6), 1256–1275.

Papalexiou, S.M., Koutsoyiannis, D., and Montanari, A., 2011. Can a simple stochastic model generate rich patterns of rainfall events?. *Journal of Hydrology*, 411 (3-4), 279–289.

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

Tyralis, H., and Koutsoyiannis, D., 2011. Simultaneous estimation of the parameters of the Hurst-Kolmogorov stochastic process. *Stochastic Environmental Research & Risk Assessment*, 25 (1), 21–33.

Tyralis, H., and Koutsoyiannis, D., 2014. A Bayesian statistical model for deriving the predictive distribution of hydroclimatic variables. *Climate Dynamics*, 42 (11-12), 2867–2883, doi:10.1007/s00382-013-1804-y.