Climacogram vs. autocovariance and power spectrum in stochastic modelling for Markovian and Hurst-Kolmogorov processes (supplementary material)

Panayiotis Dimitriadis* and Demetris Koutsoyiannis

Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Heroon Polytechneiou 5, 158 80 Zographou, Greece

*corresponding author, email: pandim@itia.ntua.gr, tel.: +30-210-772-28-38, fax: +30-210-772-28-32

Keywords: Hurst-Kolmogorov process, classical estimator, Markovian process, stochastic generation

1. Introduction

This is the supplementary material of the main paper 'Climacogram vs. autocovariance and power spectrum in stochastic modelling for Markovian and Hurst-Kolmogorov processes'. Here, we provide investigations (graphical, numerical and analytical) for the Hurst-Kolmogorov (HK) process (sect. 2.1) as well as additional results of the analysis in sect. 3.3 of the main paper, for the investigation of the dimensionless classical estimators of the climacogram (described in sect. 3.2 of the main paper), autocovariance and power spectrum (sect. 2.2). Moreover, in sect. 3, we describe a stochastic generation model (based on the combination of a finite number of Markovian processes) used in the Monte-Carlo sensitivity analysis of sect. 3.3 and 4 of the main paper. Finally, in sect. 4, we describe a methodology for a discrete time series generation of a continuous time Markovian process, for any Δ and *D*, using an ARMA(1,1) model and we show in Fig. 5 the possible errors produced in case that an AR(1) model is used instead. We use the same notation as in the main paper.

2. Additional investigations of the climacogram, autocovariance and power spectrum

The sections below provide additional analyses of sect. 3.2 and 3.3 of the main paper.

2.1 Graphical investigation of the classical estimators of climacogram, autocovariance and power spectrum for an HK process

Here, we apply the same graphical investigation in sect. 3.2 of the main paper for an HK process. In Table 1, we define the climacogram, autocovariance and power spectrum of an HK process based on the framework of sect. 2 of the main paper and in Fig. 1, we present its graphical investigation.

Table 1: Climacogram, autocovariance and power spectrum mathematical expressions of a positively correlated HK process, with $0 < b < 1$, in continuous and discrete time.

Type	HK process	
Autocovariance (continuous)	$c(\tau) = \lambda\lvert\tau/q\rvert^{-b}$ with $b = 2 - 2H$	(1)
Autocovariance (discrete)	$c_d^{(\Delta)}(j) = \lambda \dfrac{\lvert j-1\rvert^{2-b} + \lvert j+1\rvert^{2-b} - 2\lvert j\rvert^{2-b}}{(\Delta/q)^b(1-b)(2-b)}$ for $j \geq 0$, with $c_d^{(\Delta)}(0) = \gamma(\Delta)$	(2)

Climacogram (for continuous and discrete)	$\gamma(m) = \dfrac{2\lambda(m/q)^{-b}}{(1-b)(2-b)}$ with $\gamma(0) \to \infty$	(3)
Power spectrum (continuous)	$s(w) = \dfrac{4\lambda q\ \Gamma(1-b)\mathrm{Sin}\left(\frac{\pi b}{2}\right)}{(2\pi q\lvert w\rvert)^{1-b}}$	(4)
Power spectrum (discrete)	not a closed expression*	-

* eq. 12 of the main paper couldn't be further analysed

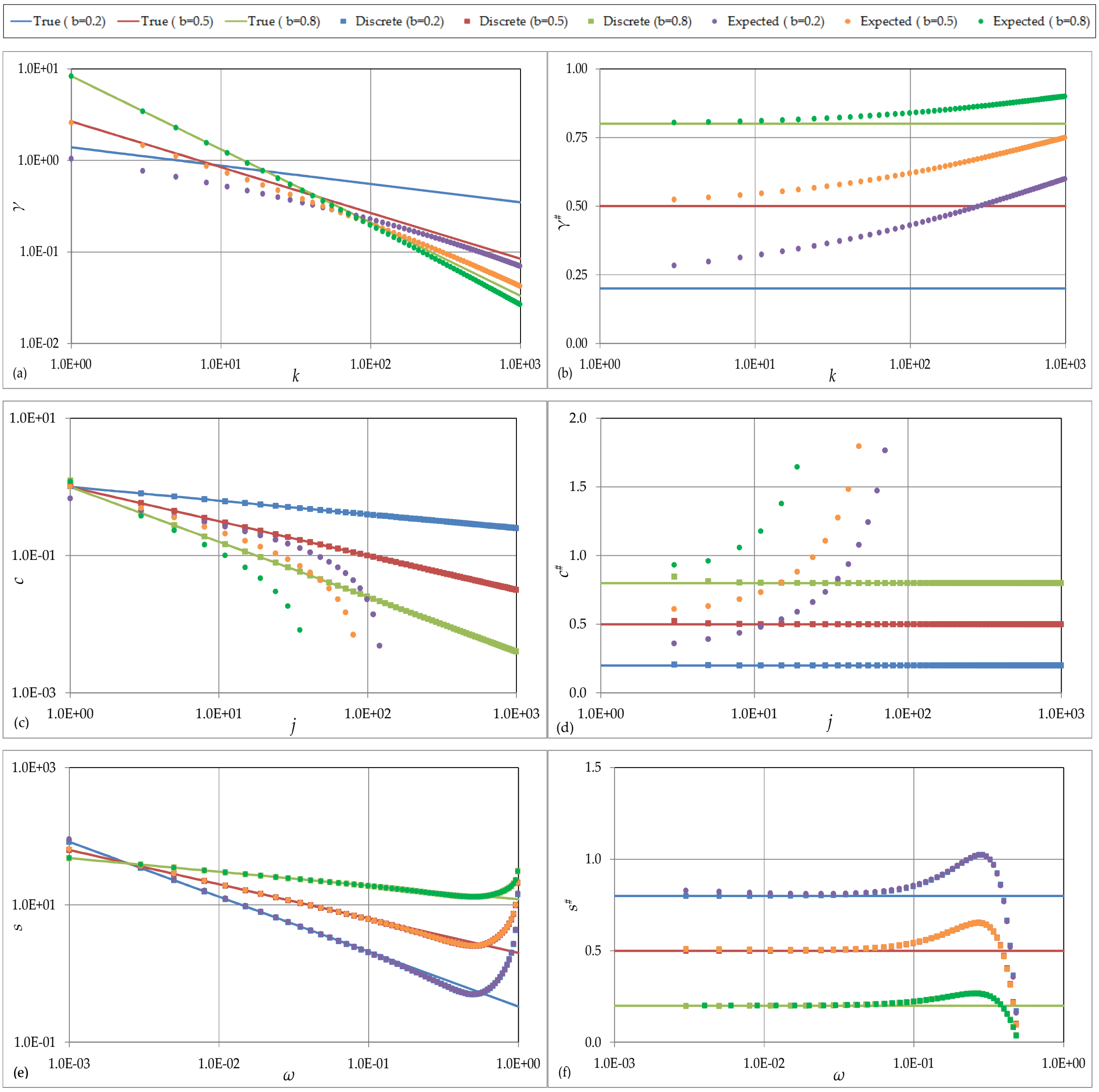

Figure 1: True values in continuous and discrete time and expected values of the climacograms (a), autocovariances (c) and power spectra (e) as well as their corresponding NLDs (b, d and f, respectively) of HK processes with b = 0.2, 0.5 and 0.8, λ = q = 1 and n = 10^3. Note that the continuous and discrete values of the climacogram are identical.

2.2 Additional investigations of the estimators of power spectrum, autocovariance and climacogram

Here, we apply the same analysis as in sect. 3.3 of the main paper for an HK process for $n = 10^3$ (Fig. 2) and for Markovian, HK and gHK processes for $n = 10^2$ (Fig. 3) and $n = 10^4$ (Fig. 4).

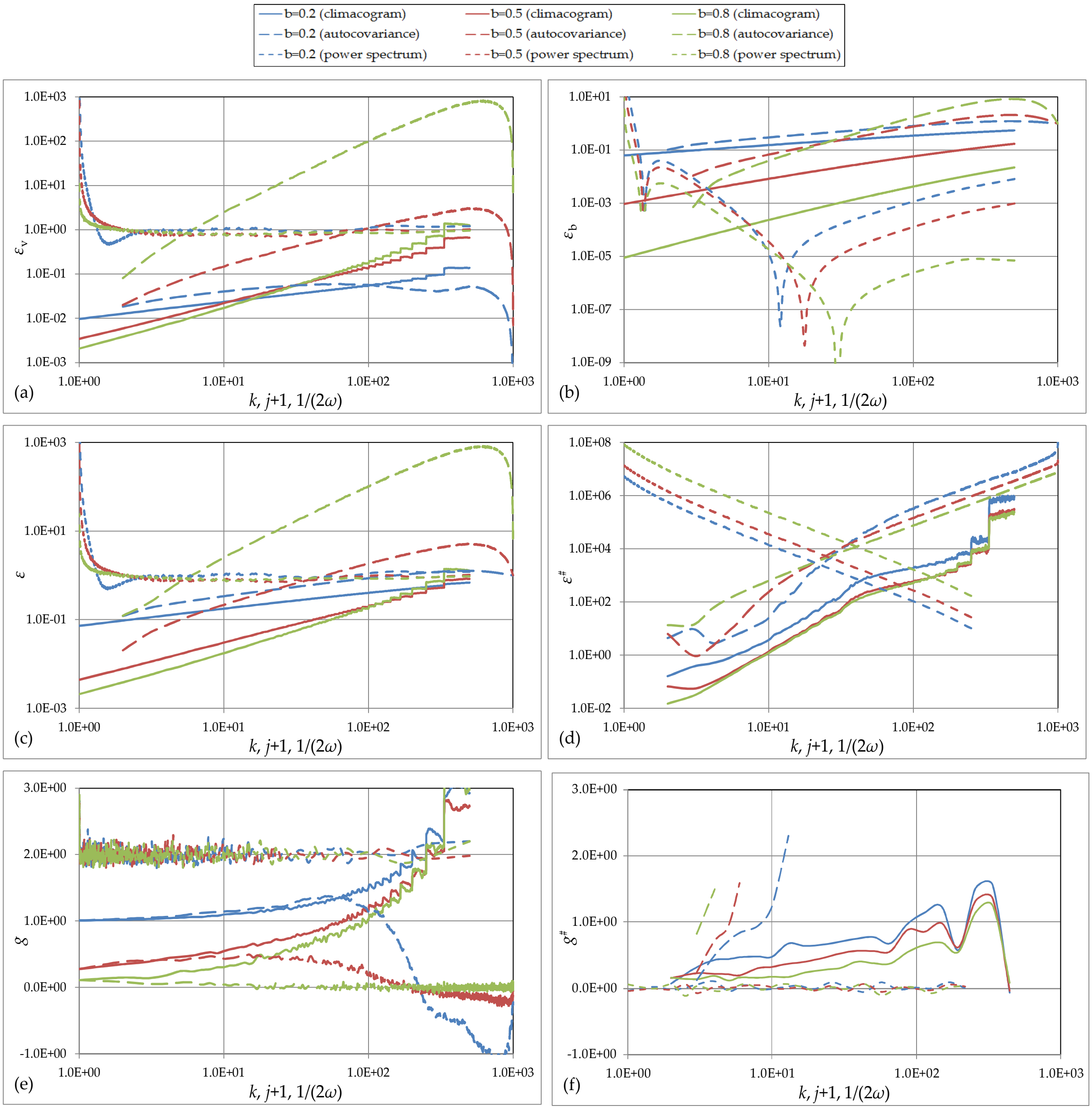

Figure 2: Dimensionless errors of the climacogram estimator (continuous line), autocovariance (dashed line) and power spectrum (dotted line), calculated from 10^4 HK synthetic series with $n = 10^3$ (for $b = 0.2$, q = 1, 10 and 100 and $\lambda = q^{-b}$): (a) ε_{v} (dimensionless MSE of variance); (b) ε_{b} (dimensionless MSE of bias); (c) ε (total dimensionless MSE); and (d) $\varepsilon^{\#}$ (total dimensionless MSE of NLD); as well as the sample skewness of their regular values and NLDs in (e) and (f).

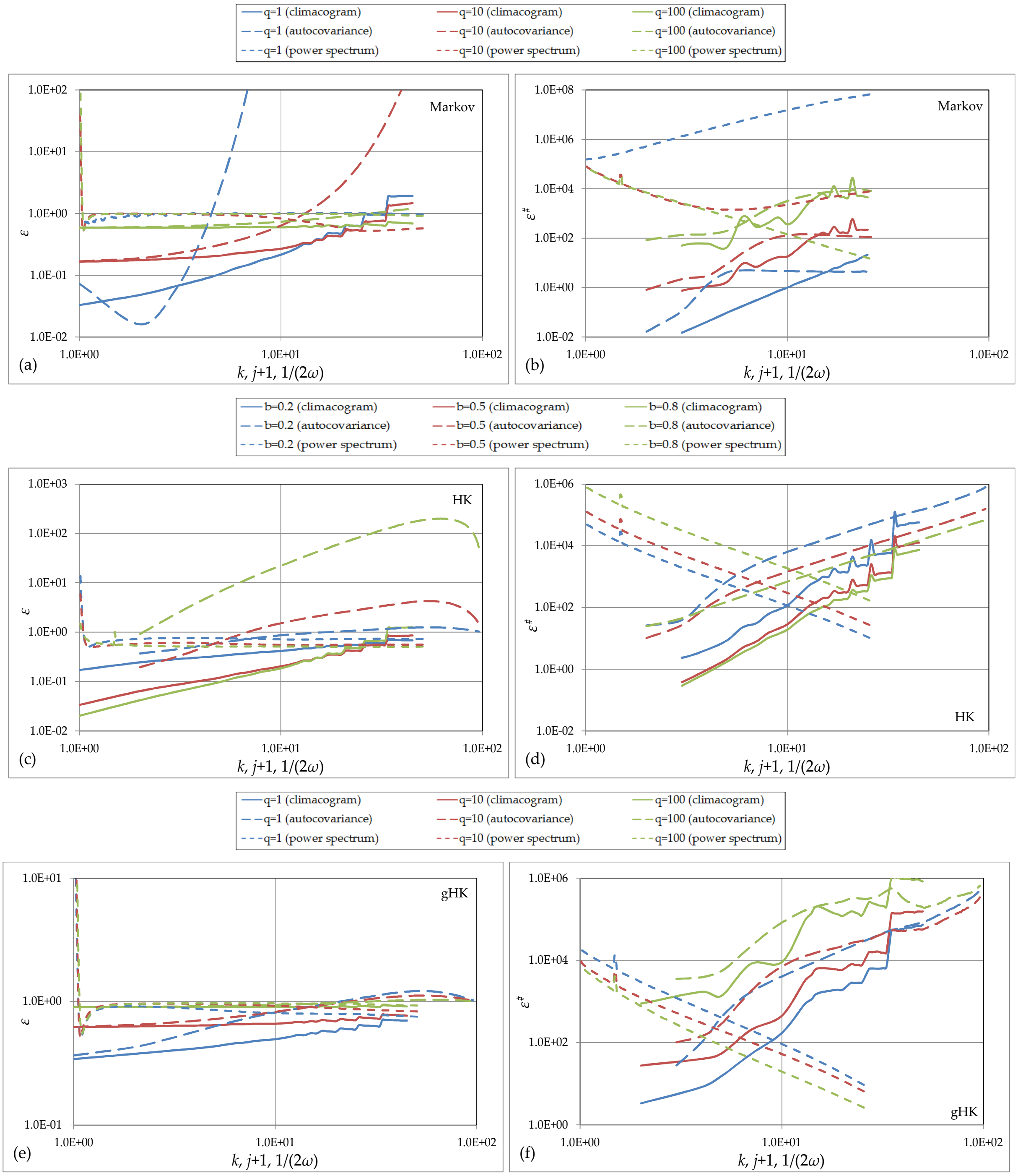

Figure 3: Dimensionless errors ε and $\varepsilon^{\#}$ of the climacogram estimator (continuous line), autocovariance (dashed line) and power spectrum (dotted line), calculated from 10^4 Markovian (for q = 1, 10 and 100, in a and b), HK (for b = 0.2, 0.5 and 0.8, in c and d) and gHK (for b = 0.2, q = 1, 10 and 100 and $\lambda = q^{-b}$, in e and f) synthetic series of $n = 10^2$.

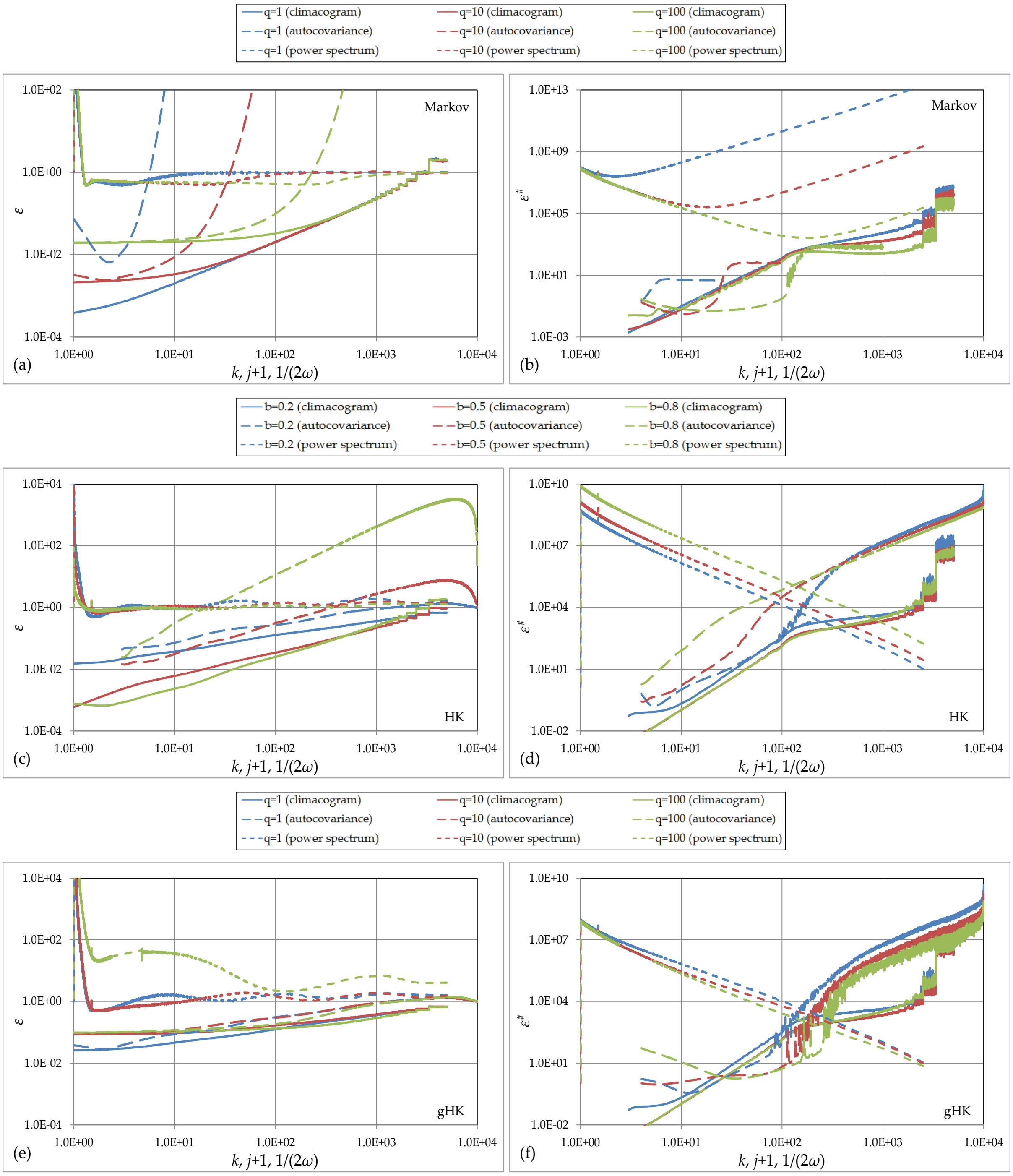

Figure 4: Dimensionless errors ε and $\varepsilon^{\#}$ of the climacogram estimator (continuous line), autocovariance (dashed line) and power spectrum (dotted line), calculated from 10^4 Markovian (for q = 1, 10 and 100, in a and b), HK (for b = 0.2, 0.5 and 0.8, in c and d) and gHK (for b = 0.2, q = 1, 10 and 100 and $\lambda = q^{-b}$, in e and f) synthetic series of $n = 10^4$.

3. Methodology for synthesis of a Gaussian stochastic process as a sum of Markovian processes

Here, we describe a methodology to produce synthetic Gaussian distributed time series of a target process $\underline{x}$ based on an aggregation of Markovian processes, denoted as $\underline{y}$, used in sect. 3.2, 3.3 and 4 of the main paper. We seek the Markovian climacograms whose sum fits the climacogram of our target true continuous time process, represented by a function $f(k\Delta)$, with k the discrete time scale and $D = \Delta > 0$ the time step. We could use the autocovariance or power spectrum but the climacogram has the advantage of reduced computational cost, as it has identical expressions for continuous and discrete time (for $D = \Delta > 0$). We denote $g(k\Delta, q, \lambda)$ the true climacogram of a Markovian process (see in Table 4 of the main paper):

$$g(k\Delta, q, \lambda) := \frac{2\lambda}{(k\Delta/q)^2}\left(k\Delta/q + e^{-k\Delta/q} - 1\right) \tag{5}$$

where λ and q are parameters corresponding to the variance and a characteristic time scale of the process, respectively.

Our target is to approximate $f(k\Delta)$ with the sum of a finite number N of functions $g(k\Delta, q_l, \lambda_l)$ for $l = 1$ to N, i.e. for all integral scales from $k = 1$ to n, where n is the number of data produced in the synthetic time series. We seek $q_l > 0$ and $\lambda_l > 0$ such as for all scales $k \geq 1$:

$$f(k\Delta) \approx \sum_{l=1}^{N} g(k\Delta, q_l, \lambda_l) \tag{6}$$

The basic assumption of this methodology is that the Markovian parameters q_l are connected to each other in a pre-defined way. Here, we choose a simple relationship based on only two parameters p_1 and p_2:

$$q_l = p_1 {p_2}^{l-1} \tag{7}$$

If we know p_1 and p_2, we can calculate analytically parameters λ_l (expressed by the matrix $\boldsymbol{\Lambda} \geq \boldsymbol{0}$) from the equation below, since the ratio $g(k\Delta, q_l, \lambda_l)/\lambda_l$ is independent of λ_l for Markovian processes:

$$\boldsymbol{A\Lambda} = \boldsymbol{I} \rightarrow \boldsymbol{\Lambda} = \boldsymbol{A}^{-1}\boldsymbol{I} \tag{8}$$

where $\boldsymbol{A} = \begin{bmatrix} \frac{g(\Delta, q_1, \lambda_1)/\lambda_1}{f(\Delta)} & \cdots & \frac{g(\Delta, q_N, \lambda_N)/\lambda_N}{f(\Delta)} \\ \vdots & \ddots & \vdots \\ \frac{g(n\Delta, q_1, \lambda_1)/\lambda_1}{f(n\Delta)} & \cdots & \frac{g(n\Delta, q_N, \lambda_N)/\lambda_N}{f(n\Delta)} \end{bmatrix}$, $\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{bmatrix}$, $\boldsymbol{I} = \left.\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}\right\} N$ and

$\boldsymbol{A}^{-1} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$, the left inverse of $\boldsymbol{A}$ (for $n > N$).

As minimization objective for the system of eq. 7 and 8, in order to estimate the parameters p_1 and p_2, first we use the dimensionless error $\varepsilon_{\mathrm{sg}}$ between the sum of Markovian climacograms and $f(k\Delta)$, to locate initial values and then, we use the error $\varepsilon_{\mathrm{mg}}$ (maximum absolute dimensionless residual), for fine tuning and distributing the error equally to all scales.

$$\varepsilon_{\mathrm{sg}} = \sum_{k=1}^{n} \left\{\frac{\sum_{l=1}^{N} g(k\Delta, q_l, \lambda_l) - f(k\Delta)}{f(\mathrm{k}\Delta)}\right\}^2 \tag{9}$$

$$\varepsilon_{\mathrm{mg}} = \max_{k=1,\dots,n} \left|\frac{\sum_{l=1}^{N} g(k\Delta, q_l, \lambda_l) - f(k\Delta)}{f(k\Delta)}\right| \tag{10}$$

Thus, we can estimate parameters p_1 and p_2 by minimizing eq. 9 and 10, then q_l can be easily found from eq. 7 and parameters λ_l from eq. 8. Finally, the synthetic discrete time series for the X process can be estimated as:

$$\underline{x}_i^{(\Delta)} = \sum_{l=1}^{N} \underline{y}_i^{(\Delta)}(l) \tag{11}$$

where $\underline{y}_i^{(\Delta)}(l)$ is the discrete time Markovian process corresponding to the climacogram $g(k\Delta, q_l, \lambda_l)$ in eq. 6, with parameters q_l and λ_l, each one produced following the methodology in sect. 4 below.

The above methodology has been tested in simple processes such as HK, gHK and combination of them and with Markovian processes and therefore, for other types of processes (e.g. anti-correlated ones with $1 < b \leq 2$) one should be cautious when applying it. For the purpose of the analysis in sect. 3 and 4 of the main paper, we applied the above methodology for HK and gHK processes for $\lambda = 1$ and for a variety of b, q/Δ and n values. In Tables 2 to 4, we present the results from this analysis. Note that we choose N, for each n and each process, as the minimum value of the sum of Markovian processes giving $\varepsilon_{\mathrm{mg}} \leq 1\%$.

Table 2: Parameters p_1 and p_2 estimated to fit different types of HK and gHK processes (for λ = 1) with a sum of Markovian processes for $n = 10^2$.

Process	b	q/Δ	p_1	p_2	N	ε_{mg} (‰)
HK	0.2	-	0.069	47.358	3	6
HK	0.5	-	0.122	22.196	3	8
HK	0.8	-	0.101	17.045	3	9
gHK	0.2	1	2.888	10.656	3	5
gHK	0.2	10	11.424	27.168	2	1
gHK	0.2	100	611.13	-	1	2
gHK	0.5	1	1.789	7.695	3	9
gHK	0.5	10	9.232	12.514	2	2
gHK	0.5	100	243.46	-	1	4
gHK	0.8	1	1.373	6.559	3	9
gHK	0.8	10	7.676	8.807	2	2
gHK	0.8	100	151.54	-	1	6

Table 3: Parameters p_1 and p_2 estimated to fit different types of HK and gHK processes (for λ = 1) with a sum of Markovian processes for $n = 10^3$.

Process	b	q/Δ	p_1	p_2	N	ε_{cm} (‰)
HK	0.2	-	0.379	10.356	5	2
HK	0.5	-	0.251	9.490	5	5
HK	0.8	-	0.103	8.958	5	4
gHK	0.2	1	2.656	11.873	4	3
gHK	0.2	10	0.852	43.042	3	6
gHK	0.2	100	111.54	27.331	2	1
gHK	0.5	1	1.964	10.505	4	7
gHK	0.5	10	8.744	5.801	4	2
gHK	0.5	100	89.976	12.591	2	2
gHK	0.8	1	1.362	8.240	4	7
gHK	0.8	10	6.900	5.112	4	2
gHK	0.8	100	74.712	8.861	2	3

Table 4: Parameters p_1 and p_2 estimated to fit different types of HK and gHK processes (for $\lambda = 1$) with a sum of Markovian processes for $n = 10^4$.

Process	b	q/Δ	p_1	p_2	N	ε_{cm} (‰)
HK	0.2	-	0.665	18.217	5	7
HK	0.5	-	0.200	11.400	6	6
HK	0.8	-	0.053	17.044	5	8
gHK	0.2	1	2.695	12.006	5	4
gHK	0.2	10	20.809	12.793	4	5
gHK	0.2	100	7.743	44.342	3	7
gHK	0.5	1	2.226	12.176	5	10
gHK	0.5	10	14.831	10.788	4	10
gHK	0.5	100	84.308	5.835	4	2
gHK	0.8	1	1.115	6.220	6	3
gHK	0.8	10	10.132	8.149	4	9
gHK	0.8	100	66.249	5.123	4	2

4. Generation of a discrete time series from of a continuous time Markovian process

Here, we present a methodology to synthesize a discrete time representation of continuous time Markovian process, with parameters q and λ (used in sect. 3.2, 3.3 and 4 of the main paper). We assume sample size n and $D = \Delta > 0$ (see in Table 4 of the main paper for its autocovariance, climacogram and power spectrum expressions). First we try to approximate the continuous time Markovian process in discrete time by an AR(1) model with variance $\lambda_{\text{AR}(1)}$, shape parameter $q_{\text{AR}(1)}$ and autocovariance $\lambda_{\text{AR}(1)} e^{-j\Delta/q_{\text{AR}(1)}}$, for $j \geq 0$. We find that the AR(1) model either underestimates all autocovariances of the process for lags $j \geq 1$, when we set the variance correctly (from eq. 17 in Table 4 of the main paper) to $\lambda_{\text{AR}(1)} = \gamma(\Delta) = \frac{2\lambda}{(\Delta/q)^2}\left(\Delta/q + e^{-\Delta/q} - 1\right) \leq \lambda$, or overestimates this variance, when we set it equal to continuous time Markovian variance, i.e. $\lambda'_{\text{AR}(1)} = \gamma(0) = \lambda$ (where in both cases the correct shape parameter $q_{\text{AR}(1)} = q$ is used). Keeping the variance equal to $\lambda_{\text{AR}(1)}$ and setting the ratio of the lag 1 autocovariance (or first-order autocorrelation coefficient) over the discrete variance to $' = \frac{c_d^{(\Delta)}(1)}{\gamma(\Delta)} = \frac{\left(1-e^{-\Delta/q}\right)^2}{\left(\Delta/q+e^{-\Delta/q}-1\right)}$ (instead of its proper value, i.e. $a = e^{-\Delta/q}$), the model correctly estimates the lag 0 and 1 discrete time Markovian autocovariances but leads to high overestimation of all the rest autocovariances, i.e. for lags $j > 1$. Only in case of very small Δ/q (or $\Delta \ll D$), i.e. when $a \approx a' \approx 1$, $c_d^{(\Delta)}(1) \approx a\gamma(\Delta)$ and $\lambda_{\text{AR}(1)} \approx \lambda$, a single AR(1) model can well approximate a discrete time representation of a continuous time Markovian process (for $\Delta = 0$, the model AR(1) can exactly represent it). Specifically, for $\Delta/q \lesssim 2.5\%$, we have $\frac{|a'-a|}{a'} \lesssim 1\%$ and thus, the AR(1) autocovariance small deviates from the discretized continuous time Markovian one, while for large Δ/q, the error produced can be quite large (see analysis displayed in Fig. 5 for $D = \Delta > 0$, while in case of $D \neq \Delta > 0$, the errors produced can be significant).

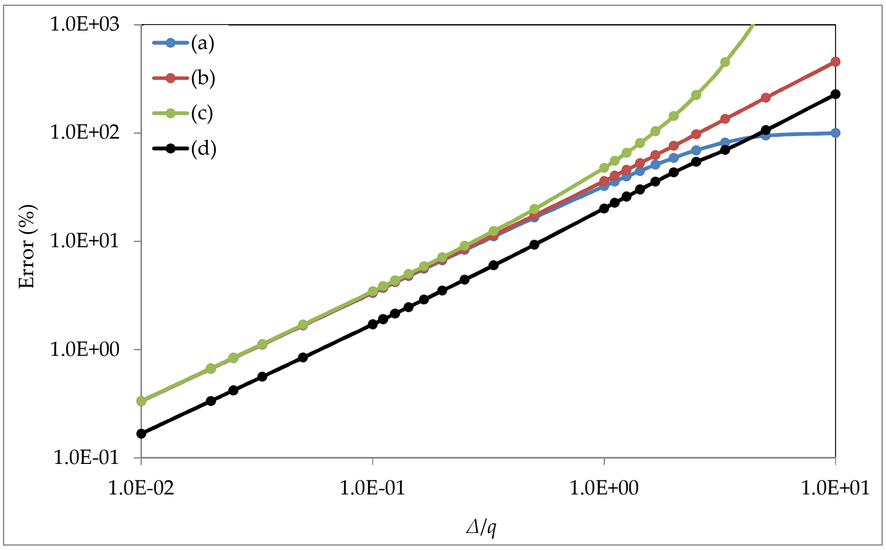

Figure 5: Dimensionless error $\max_{j=0,\dots,n-1}\left|\frac{c'^{(\Delta)}_d(j)-c^{(\Delta)}_d(j)}{c^{(\Delta)}_d(j)}\right|$, between a Markovian process in discrete time (with parameters q, λ and $c_d^{(\Delta)}(j>0)=\frac{\lambda\left(1-e^{-\Delta/q}\right)^2}{(\Delta/q)^2}\,\mathrm{e}^{-(|j|-1)\Delta/q}$, with $c_d^{(\Delta)}(0)=\gamma(\Delta)=\frac{2\lambda}{(\Delta/q)^2}\left(\Delta/q+e^{-\Delta/q}-1\right)$) and an AR(1) model with a discrete time autocovariance $c'^{(\Delta)}_d(j)$, $q_{\mathrm{AR}(1)}=q$ and a scale parameter equal to: (a) the discrete time Markovian variance (blue), i.e. $\lambda_{\mathrm{AR}(1)}=\frac{2\lambda}{(\Delta/q)^2}\left(\Delta/q+e^{-\Delta/q}-1\right)$; (b) the variance of the continuous time Markovian process (red), i.e. $\lambda'_{\mathrm{AR}(1)}=\lambda$; (c) the variance used to correctly estimate all autocovariances except the lag 0 one (green), i.e. $\lambda''_{\mathrm{AR}(1)}=\frac{c_d^{(\Delta)}(1)}{e^{-\Delta/q}}=\frac{\lambda e^{\Delta/q}\left(1-e^{-\Delta/q}\right)^2}{(\Delta/q)^2}$ and (d) a variance in between $\lambda_{\mathrm{AR}(1)}$ and λ (black), i.e. $\lambda'''_{\mathrm{AR}(1)}=\left(\lambda_{\mathrm{AR}(1)}+\lambda\right)/2$.

However, it is known that the discrete time representation of the Markovian process corresponds to an ARMA(1,1) model (as mentioned in Koutsoyiannis 2002, 2013a), denoted as $\underline{y}$. Here, we show its algorithm for the general case of $D \neq \Delta$, where $c(\tau)=e^{-|\tau|/q}$ now leads to:

$$c_d^{(\Delta,D)}(j)=\frac{1}{\Delta^2}\int_0^{\Delta}\int_{jD}^{jD+\Delta}c(x-y)\mathrm{d}x\mathrm{d}y=\frac{\lambda\left(1-e^{-\Delta/q}\right)^2}{(\Delta/q)^2}e^{-(Dj-\Delta)/q}=c_d^{(\Delta=D)}(j)e^{-j(D-\Delta)/q} \quad (12)$$

Thus, ARMA(1,1) algorithm for the general case of modelling a Markovian process, i.e. $c(\tau)=e^{-|\tau|/q}$, in discrete time, is as follows:

$$\underline{y}_i^{(\Delta,D)}=a_1\underline{y}_{i-1}^{(\Delta,D)}+\underline{v}_i+a_2\underline{v}_{i-1}\text{ , } i=1,\dots,n \quad (13)$$

where

$$a_1=e^{-D/q} \quad (14)$$

a parameter related to the shape of the process, with $0<a_1\leq 1$,

$$\underline{v}_i=N\left(\mu_{\underline{v}},\sigma_{\underline{v}}\right) \quad (15)$$

is the discrete time Gaussian white noise process, with mean value:

$$\mu_{\underline{v}}=\frac{1-a_1}{1+a_2}\mu_{\underline{y}} \quad (16)$$

with $\mu_{\underline{y}}$ the mean of $\underline{y}$. The parameters a_2 and $\sigma_{\underline{v}}$ and can be found from the solution of the equations:

$$c_d^{(\Delta)}(0)=a_1c_d^{(\Delta)}(1)+\left(1+a_1a_2+{a_2}^2\right){\sigma_{\underline{v}}}^2 \quad (17)$$

$$c_d^{(\Delta)}(1)=a_1c_d^{(\Delta)}(0)+a_2{\sigma_{\underline{v}}}^2 \quad (18)$$

where

$$c_d^{(\Delta,D)}(0) = \gamma(\Delta) = \frac{2\lambda}{(\Delta/q)^2}\left(\Delta/q + e^{-\Delta/q} - 1\right) \tag{19}$$

$$c_d^{(\Delta,D)}(1) = \frac{\lambda\left(1-e^{-\Delta/q}\right)^2}{(\Delta/q)^2} e^{-(D-\Delta)/q} \tag{20}$$

the autocovariance of the continuous time Markovian process (represented in discrete time) for lag zero and one, respectively.

Eq. C.6 and C.7 result in a 2nd order polynomial for a_2:

$$a_2{}^2 + a_2 \frac{2a_1 c_d^{(\Delta,D)}(1) - (1 + a_1{}^2)\gamma(\Delta)}{c_d^{(\Delta,D)}(1) - a_1\gamma(\Delta)} + 1 = 0$$

with $c_d^{(\Delta,D)}(1) \geq a_1\gamma(\Delta)$ (the equality holds only for $q \to \infty$) and it has two real positive solutions:

$$a_2 = \frac{-B \pm \sqrt{B^2 - 4}}{2} > 0 \tag{21}$$

$$\text{with } B = \frac{2a_1 c_d^{(\Delta,D)}(1) - (1 + a_1{}^2)\gamma(\Delta)}{c_d^{(\Delta,D)}(1) - a_1\gamma(\Delta)} \leq -2 \tag{22}$$

and also,

$$\sigma_{\underline{v}} = \sqrt{\frac{\gamma(\Delta) - a_1 c_d^{(\Delta,D)}(1)}{1 + a_1 a_2 + a_2{}^2}} \tag{23}$$

Acknowledgement

This supplementary material of the main paper 'Climacogram vs. autocovariance and power spectrum in stochastic modelling for Markovian and Hurst-Kolmogorov processes' was partly funded by the Greek General Secretariat for Research and Technology through the research project "Combined REnewable Systems for Sustainable ENergy DevelOpment" (CRESSENDO; programme ARISTEIA II; grant number 5145).

References

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

Koutsoyiannis, D. (2013a), Encolpion of stochastics: Fundamentals of stochastic processes, 12 pages, *Department of Water Resources and Environmental Engineering – National Technical University of Athens*, Athens.