This poster participates in OSP

Outstanding Student Poster Contest

European Geosciences Union General Assembly 2016
Vienna, Austria, 17-22 April 2016
Session HS7.4: Change in climate, hydrology and society,
Vol. 18, EGU2016-14577.

Markov vs. Hurst-Kolmogorov behaviour identification in hydroclimatic processes

Panayiotis Dimitriadis, Naya Gournari and Demetris Koutsoyiannis

The poster can be downloaded at: http://www.itia.ntua.gr/

Department of Water Resources
and Environmental Engineering
National Technical University of Athens

1. Introduction

Hydroclimatic processes are usually modeled either by exponential decay of the autocovariance function, i.e. Markov behaviour, or power type decay, i.e. long-term persistence (or else Hurst-Kolmogorov behaviour; Koutsoyiannis, 2015). Hurst was one of the most influential hydrologists of the last century due to his remarkable scientific work. He discovered that hydrological and other geophysical time series exhibited statistical behavior which was named as Hurst-Kolmogorov phenomenon (long-term persistence or long-range dependence). For the Markov process the future state depends entirely on the present state whereas for the HK process entirely on the present as well as the past state.

These two processes include only one parameter, i.e., the lag-1 autocorrelation coefficient and the Hurst coefficient, respectively. However, as simple as may seem to be, it is often quite challenging to determine which one best describes the observed stochastic structure. In hydroclimatic processes, where we usually have limited number of measurements, the above identification becomes even harder and sometimes it is statistically impossible to choose between one another. For the identification and quantification of such behaviours several graphical stochastic tools can be used such as the climacogram, autocovariance, variogram, power spectrum etc. Comparing these tools the climacogram is more accurate with a lower total mean-square error, thus smaller statistical uncertainty (Dimitriadis and Koutsoyiannis, 2015; Dimitriadis et al., 2015). The climacogram comes from the Greek word climax which means scale and is defined as the (plot of) variance (or standard deviation) of the averaged process (assuming stationary) versus averaging time scale (Koutsoyiannis, 2015).

Most methodologies including the above tools are based on the unbiased estimator of the expected value of the standard deviation or variance through least-squares techniques (e.g., Tyralis and Koutsoyiannis, 2011), or based on maximum-likelihood estimators (e.g., Kendziorski, 1999). In this analysis, we explore a methodology that combines both the practical use of a graphical representation of the internal structure of the process as well as the statistical robustness of the maximum probability estimator. For validation and illustration purposes, we apply this methodology to fundamental stochastic processes such as Markov processes with lag-1 autocorrelations ranging from 0.1 to 0.9, and Hurst-Kolmogorov processes, for Hurst coefficients ranging from 0.5 (i.e., white noise) to 0.9.

2. Definitions and notations

For the identification between Markov and HK processes, we adopt the climacogram. Besides the fact that it exhibits smaller uncertainty in comparison with other tools like the autocovariance and power spectrum, it has the advantageous property of developing true identical log-log derivative/slope (abbreviated LLS) at large scales equal to -1, which corresponds to both Markov and white noise processes. Therefore, the climacogram allows for a direct comparison between the HK and Markov mathematical processes to decide which of the two best describes the natural process. The true climacogram, classical estimator and expected value are given by (Koutsoyiannis, 2013; Dimitriadis and Koutsoyiannis, 2015):

$$\gamma(m) := \frac{\operatorname{var}\left[\int_0^m \underline{x}(t)\mathrm{d}t\right]}{m^2}$$

$$\underline{\hat{\gamma}}(k\Delta) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{1}{k}\left(\sum_{l=k(i-1)+1}^{ki}\underline{x}_l^{(\Delta)}\right) - \frac{\sum_{l=1}^{n}\underline{x}_l^{(\Delta)}}{n}\right)^2$$

$$\mathrm{E}\left[\underline{\hat{\gamma}}(k\Delta)\right] = \frac{1-\gamma(n\Delta)/\gamma(k\Delta)}{1-k/n}\gamma(k\Delta)$$

, where γ is the climacogram (^ denotes estimation and underscore is used for random variables), $\underline{x}$ is the random process in continuous time, t denotes time, m is the scale in continuous time, k is the scale in discrete time, n is the length of the sample, $\underline{x}^{(\Delta)}$ is the random process in discrete time and Δ is the time interval.

The continuous-time Markov and HK processes as well as the definition for the Hurst coefficient are given by:

$$\gamma(m) = \frac{2\lambda}{(m/q)^2}\left(m/q + e^{-m/q} - 1\right)$$

$$\gamma(m) = \frac{\lambda}{m^{2-2H}} \quad \text{and} \quad H := 1 + \frac{1}{2}\lim_{m\to\infty}\gamma^{\#}(m)$$

, where λ is the true variance of the process, q is the Markov parameter with $e^{-1/q}$ corresponding to the lag-1 autocorrelation coefficient, H is the Hurst parameter and # denotes the LLS.

3. Methodology

In this work, we explore three common scenarios for the analysis of geophysical processes and for each scenario, we provide appropriate tests to enable the identification between the two processes. For each scenario, we produce 10^4 synthetic timeseries of n = 200 for each one of the Markov processes (i.e., q = 0.5, 1, 2, 5 and 10), for the white noise (i.e., H = 0.5) and for each one of the examined HK processes (i.e., H = 0.6, 0.7, 0.8 and 0.9). For the synthesis of the latter processes we use the 3×AR(1) technique of Koutsoyiannis (2002):

$\underline{x}_i = \underline{A}_i + \underline{B}_i + \underline{C}_i$, where A, B and C, are Markov processes with autocorrelation coefficients and variances:

$$\rho_A = 1.52(H-0.5)^{1.32},\ \rho_B = 0.953 - 7.69(1-H)^{3.85} \text{ and } \rho_C = \begin{cases} 0.932 + 0.087, & H \le 0.76 \\ 0.993 + 0.007H, & H > 0.76 \end{cases}$$

$\lambda_A = (1-c_1-c_2)\gamma_x$, $\lambda_B = c_1\gamma_x$ and $\lambda_C = c_2\gamma_x$, with c_1, c_2 fitting parameters between empirical and model γ.

Firstly, we explore the scenario where we have multiple timeseries of identical length, for example in case of repeatable experiments over the same initial conditions. For this case, the analysis is based on the expected value of the process (Dimitriadis and Koutsoyiannis, 2015) by applying two tests, one to highlight the difference between the expected values of the two process and one for their confidence intervals. At the former test, we plot the fitting error of the expected value for the examined and observed processes for various ranges of scales (from smaller to larger scales and larger to smaller) vs. scale. At the latter test, we estimate which process best fits the confidence intervals (specifically, the q5% and q95%) of the empirical climacogram for various scales. In the second scenario, we have multiple timeseries of different lengths, for example in case of various hydrometeorological stations around the globe with different observational periods. For this case, we apply the same tests as before, but with adjusting a correction factor for the difference in bias to each empirical climacogram (specifically, we multiply it with a ratio of the expected value for the length of the original timeseries and the desired length). Finally in the third scenario, we have only one timeseries and we apply a test based on the most probable value of the climacogram.

4. Is climacogram unbiased?

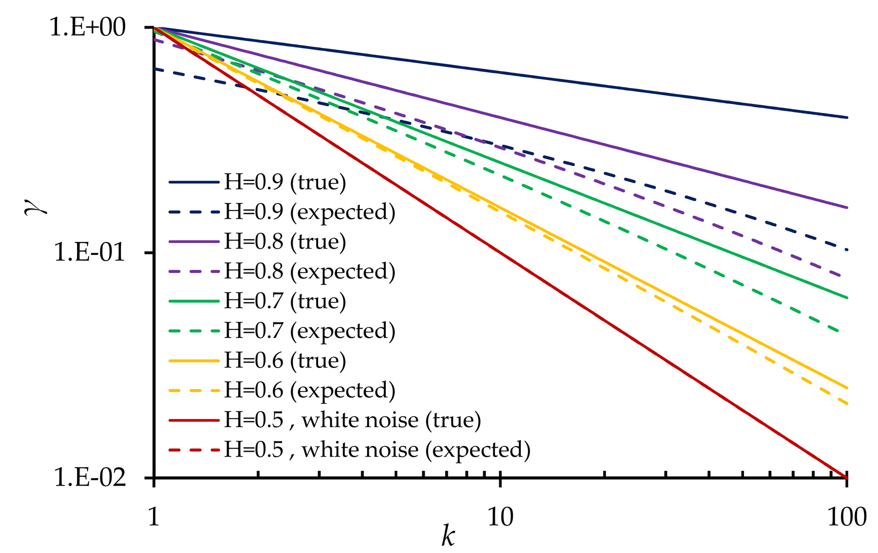

Figure 1. True vs. expected values between various HK processes.

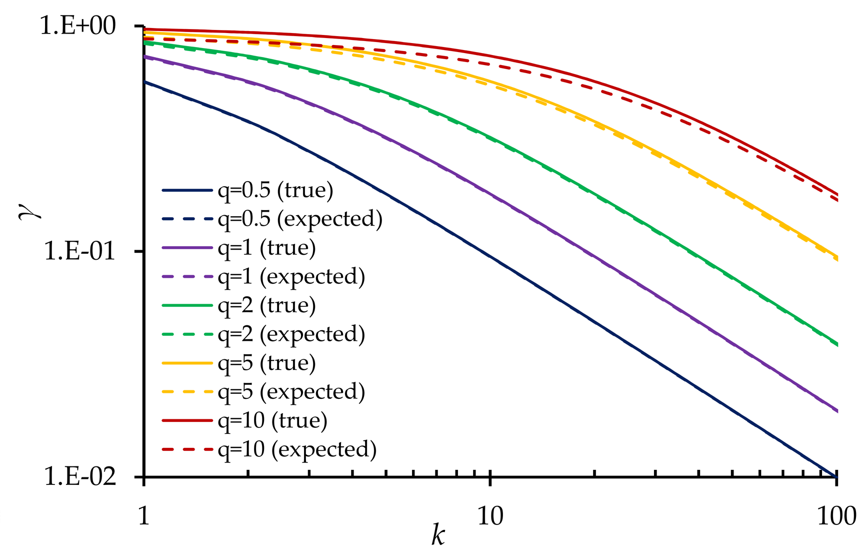

Figure 2. True vs. expected values between various Markov processes.

In the Figures above we observe that the expected value of the climacogram may be different from its true value, especially for large Hurst coefficients. Only in case of White Noise (*H=0.5)* the climacogram is unbiased (equivalently, zero autocorrelation). Moreover, the bias in case of a Markov process is negligible for small lag-1 autocorrelation coefficients, but it can be significant for large ones. In stochastic modelling we apply the expected value a process without considering that it may be different than its true value since it is impossible to have a timeseries of infinite length or equivalently infinite timeseries of finite length.

5. What if we have multiple timeseries of identical length?

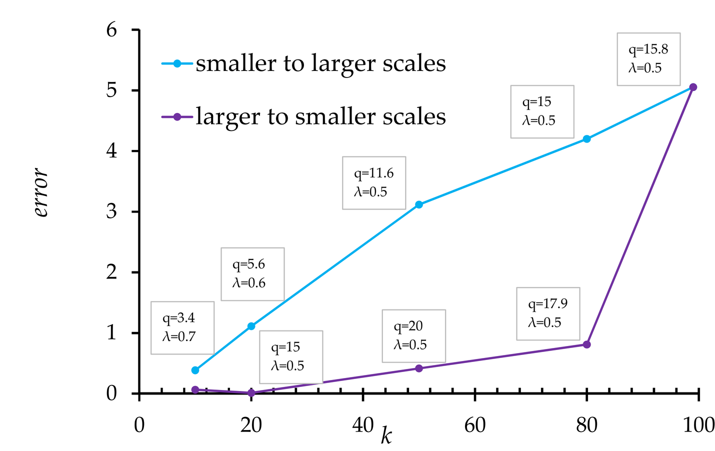

Figure 3. Fitting error of expected HK process (H=0.9).

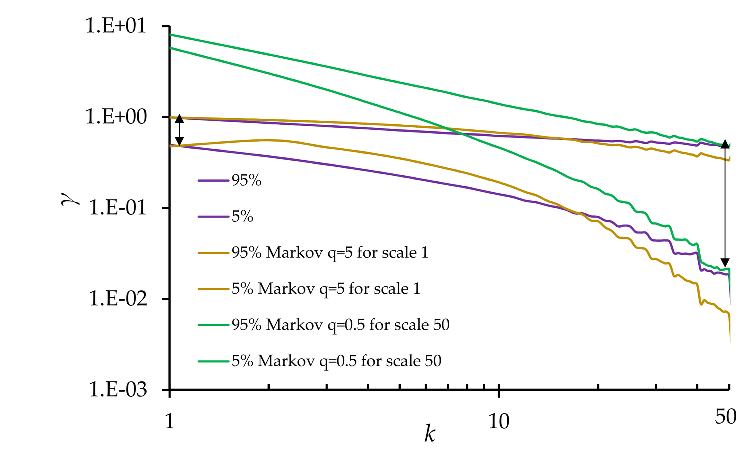

Figure 4. Range of confidence intervals of HK process (H=0.9).

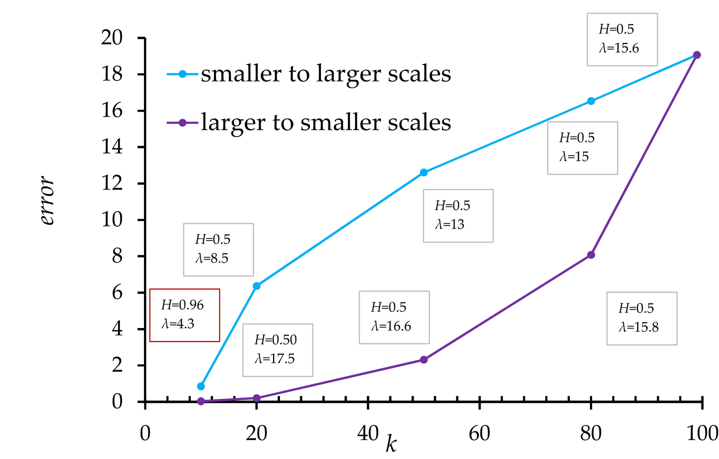

Figure 5. Fitting error of expected Markov process (q=10).

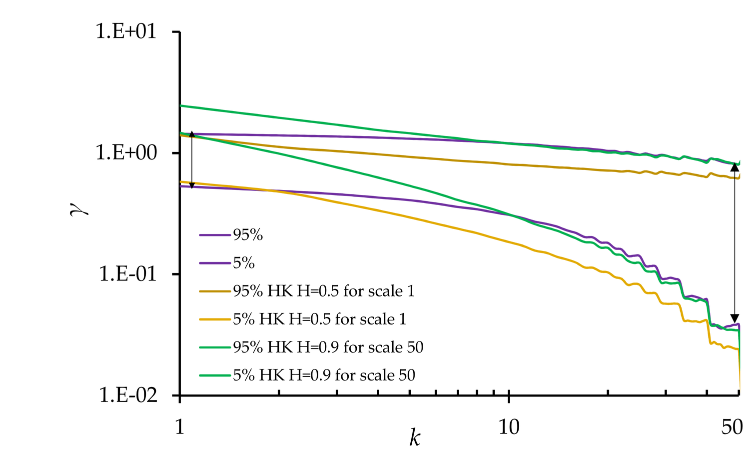

Figure 6. Range of confidence intervals of Markov process (q=10).

6. What if we have multiple timeseries of different length?

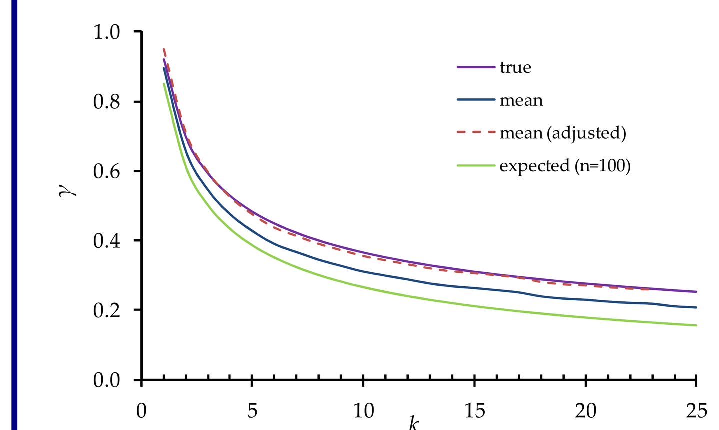

Figure 7. Comparison between mean (no adjustment for bias) and adjusted mean HK process (H=0.8).

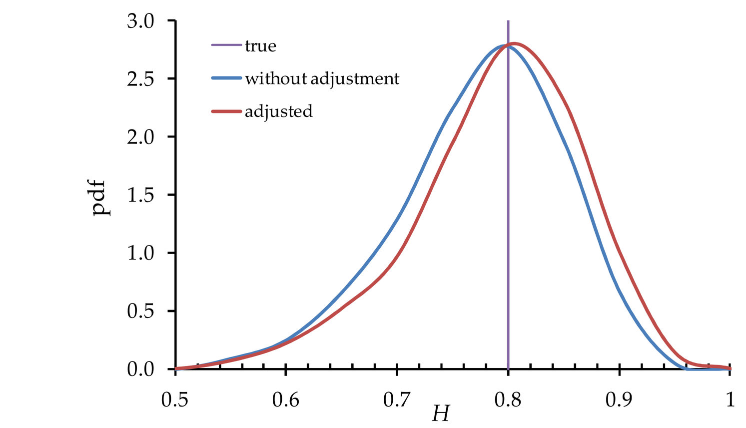

Figure 8. Hurst coefficient distribution functions for HK process (H=0.8) with and without adjustement for bias.

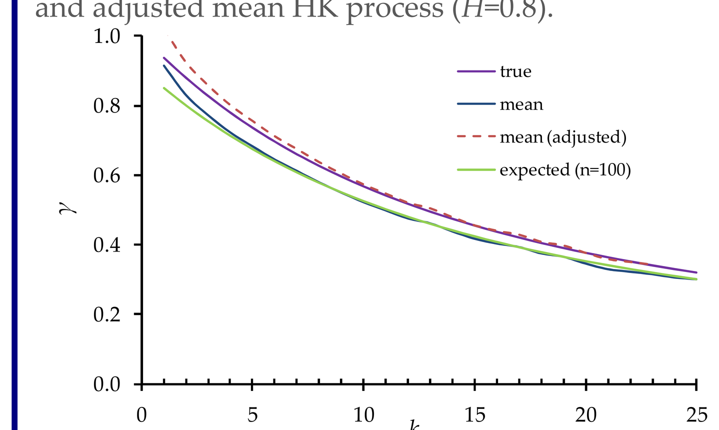

Figure 9. Comparison between mean (no adjustment for bias) and adjusted mean Markov process (q=5).

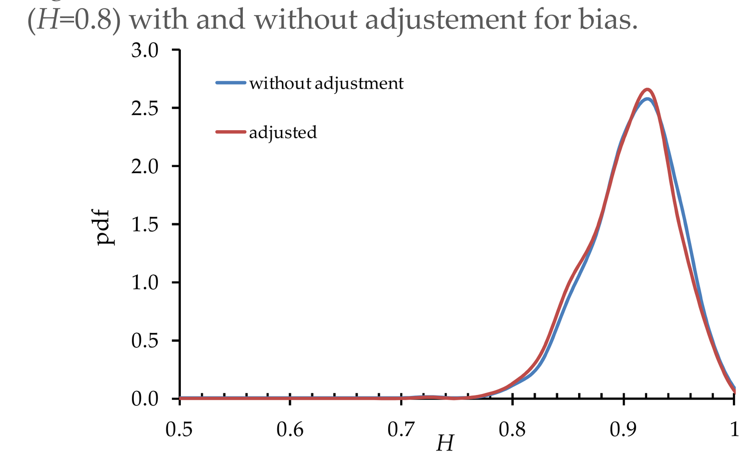

Figure 10. Hurst coefficient distribution functions for Markov process (q=5) with and without adjustment for bias.

7. What if we have one timeseries?

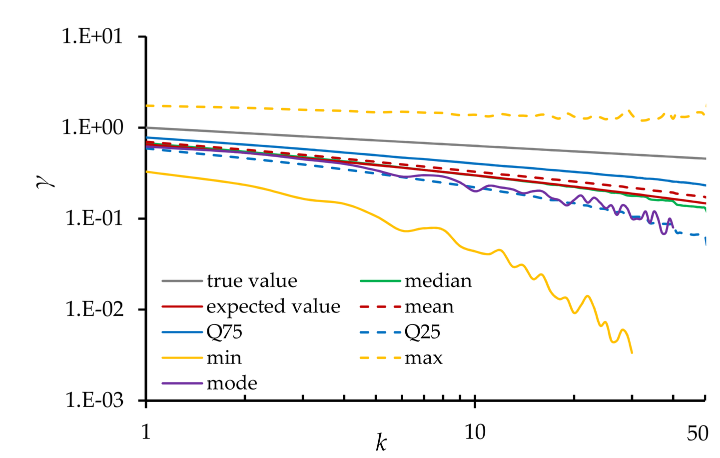

Figure 11. Climacogram characteristics of HK process (H=0.9).

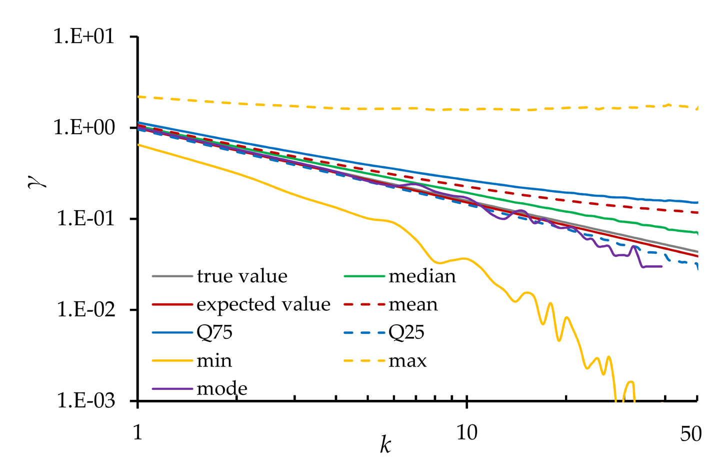

Figure 12. Climacogram characteristics of HK process (H=0.6).

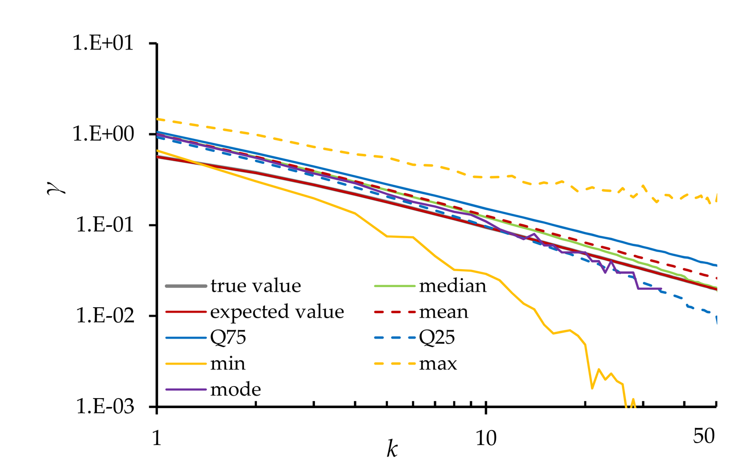

Figure 13. Climacogram characteristics of Markov process (q=0.5).

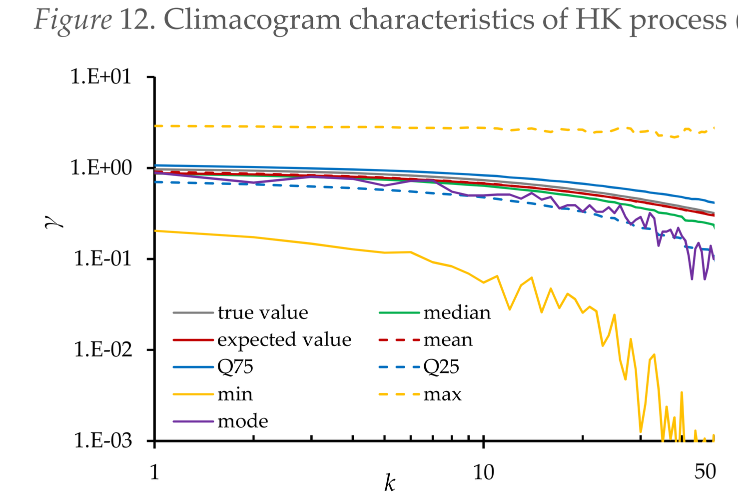

Figure 14. Climacogram characteristics of Markov process (q=10).

8. What if we have one timeseries? (cont.)

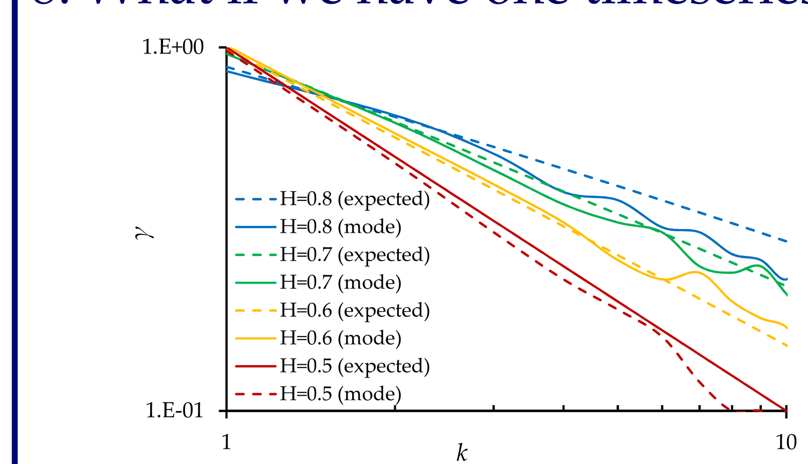

Figure 15. Expected value and mode for all examined HK processes.

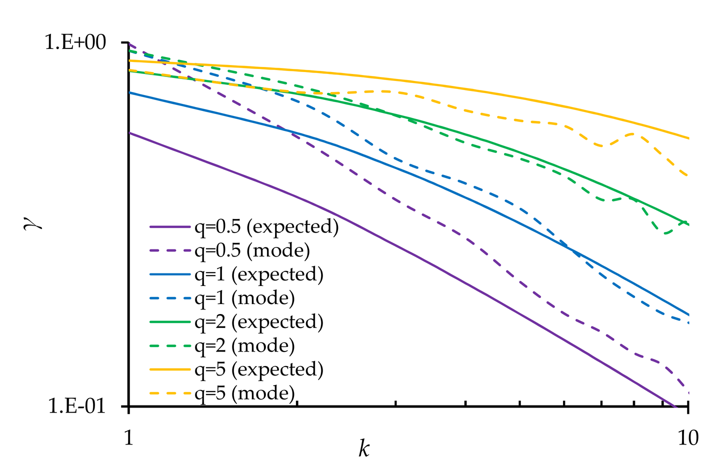

Figure 16. Expected value and mode for all examined Markov processes.

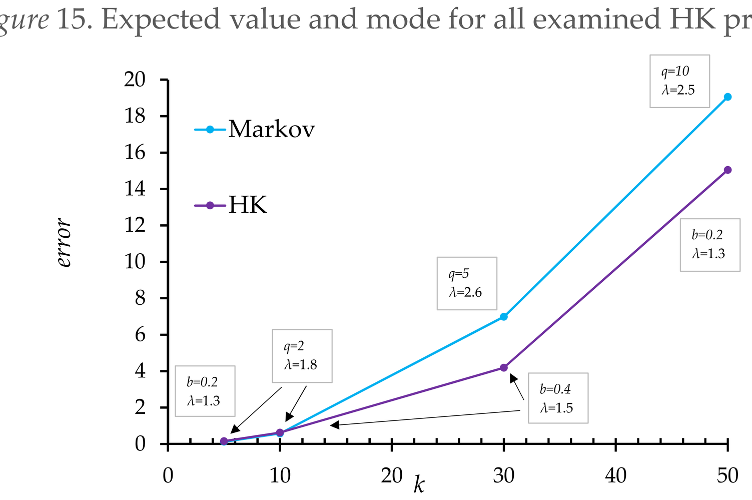

Figure 17. Fitting error for HK process (H=0.9).

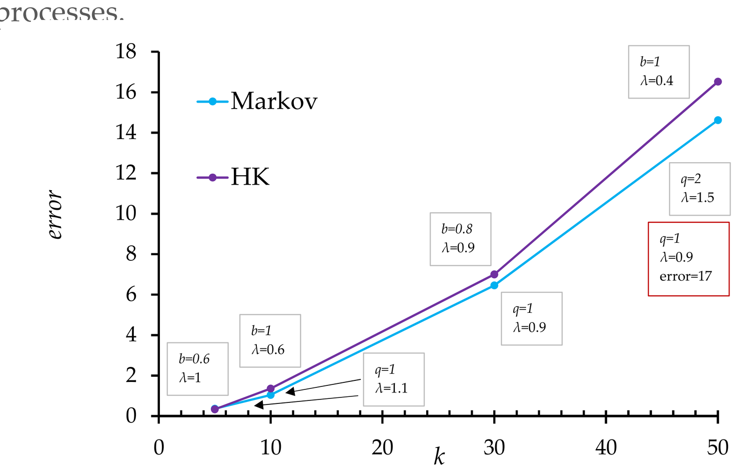

Figure 18. Fitting error for Markov process (q=1).

9. Comments and conclusions

In this work, a common problem in stochastic analysis is tackled, that is the identification from data of an exponential-type decay (i.e., Markov behaviour) or power-type decay (i.e., HK behaviour) of the autocorrelation function of a random variable. We explore methodologies and propose several tests to ease this dilemma. Our analysis is based on the climacogram whose bias can be estimated accurately (sect. 2, 3 and 4). The two tests we propose are based on the expected value and confidence intervals of the examined mathematical process. In case the expected value of the empirical process can be estimated within reasonable accuracy (sect. 5 and 6), we can determine the fitting error for various scales between the expected value of each process and the empirical one. As we move from smaller to larger scales and contrariwise, we can easily determine which model has the lowest error and thus, which one we should prefer (Fig. 3 and 4). In case we only have limited scales and the two processes give equally small fitting errors, we can apply a second test based on the range of the confidence intervals for both processes (Fig. 5 and 6). Finally, in case we have only one timeseries of the random variable then we can only apply the first test but instead of using the expected value we should use the most probable one (sect. 7 and 8).

References

Dimitriadis, P., and D. Koutsoyiannis, Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst–Kolmogorov processes, *Stochastic Environmental Research & Risk Assessment*, 29 (6), 1649–1669, doi:10.1007/s00477-015-1023-7, 2015.

Dimitriadis, P., D. Koutsoyiannis, and P. Papanicolaou, Stochastic similarities between the microscale of turbulence and hydrometeorological processes, *Hydrological Sciences Journal*, doi:10.1080/02626667.2015.1085988, 2015.

Kendziorski, C.M., Evaluation Maximum Likelihood Estimation Methods to Determine the Hurst Coefficient, *Physica A*, Vol. 273, No. 3-4, 1999, pp. 439-451. http://dx.doi.org/10.1016/S0378-4371(99)00268-X.

Koutsoyiannis, D., Generic and parsimonious stochastic modelling for hydrology and beyond, *Hydrological Sciences Journal*, 61 (2), 225–244, doi:10.1080/02626667.2015.1016950, 2016.

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