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/
1. Introduction
Hydroclimatic processes are usually modeled either by exponential decay of the autocovariance function, i.e., Markov behavior, or power type decay, i.e., long-term persistence or long-range dependency. For the Markov process the future state depends entirely on the present state whereas for the HK process entirely on the past as well as the present.

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 another. For the identification and quantification of such behaviors several graphical stochastic processes can be used such as the climacogram, autocovariance, spectrum 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., 2018). The climacogram comes from the Greek word climax which means and is defined as the plot of variance (or standard deviation) of the averaged process (say 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., Tyrilas and Koutsoyiannis, 2011), or based on maximum-likelihood estimators (e.g., Kendziorz, 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 find stochastic parameters such as the HK coefficients (by adjusting from 0.1 to 0.9), and Markov 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 advantage 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, 2015; Dimitriadis and Koutsoyiannis, 2015):

\[
\gamma(n) = \frac{1}{n^2} \sum_{k=1}^{n} \left( \frac{\sum_{i=1}^{n} x(i+k) - \sum_{i=1}^{n} x(i)}{n} \right)^2
\]

\[
\hat{\gamma}(n) = \frac{1}{1 - \lambda}(n^2 \hat{\gamma} p(n)) - \frac{1}{\lambda^2} \hat{\gamma} p(n) \sum_{i=1}^{n} \frac{1}{i(n-i)}
\]

\[
E(\gamma(n)) = \frac{1}{1 - \lambda}(n^2 E(\gamma p(n))) - \frac{1}{\lambda^2} E(\gamma p(n)) \sum_{i=1}^{n} \frac{1}{i(n-i)}
\]

where: \(p\) is the climacogram (~ denotes estimation and underline is used for random variables), \(\gamma\) is the random process in continuous time, \(\gamma\) denotes time, \(\lambda\) is the scale in discrete time, \(n\) is the length of the sample, \(\gamma\) is the random process in discrete time and \(\lambda\) is the time interval.

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

\[
\hat{\gamma} p(n) = \frac{(n^2 \hat{\gamma} p(n))}{\lambda} \text{ and } \hat{\gamma} p(n) = \frac{1}{\lambda^2} \sum_{i=1}^{n} \frac{1}{i(n-i)}
\]

\[
E(\gamma p(n)) = \frac{1}{\lambda^2} \sum_{i=1}^{n} \frac{1}{i(n-i)}
\]

where: \(\hat{\gamma} p(n)\) is the Hurst parameter with \(\hat{\gamma} p(n) = \gamma p(n)\) corresponding to the lag-1 autocorrelation coefficient, \(H\) is the Hurst parameter and \(\gamma\) 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 tools to enable the identification between the two processes. For each scenario, we produce 100 synthetic time series \(n = 200\) for each one of the Markov processes \((\lambda = 0.5, 1, 2, 10)\), for the white noise (i.e., \(H = 0.5\)) and for each one of the examined HK processes (i.e., \(H = 0.1, 0.7, 0.8, 0.9\)). For each scenario, we apply two types of identification: one based on the climacogram (Koutsoyiannis, 2015) and one based on the autocovariance function (Kendziorski, 1999).

In the Inferences 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 equivalent, 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 of 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.

4. Is climacogram unbiased?

In Figures 1 and 2 we observe that the climacogram unbiasedness decreases as Hurst coefficient increases.

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

In Figures 3 and 4 we observe that the expected value of the climacogram is unbiased for all Hurst coefficients.

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

In Figures 5 and 6 we observe that the expected value of the climacogram is biased for all Hurst coefficients.

7. What if we have one timeseries?

In Figures 7 and 8 we observe that the expected value of the climacogram is biased for all Hurst coefficients.

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

In Figures 9 and 10 we observe that the expected value of the climacogram is unbiased for all Hurst coefficients.

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 behavior) or power type decay (i.e. HK behavior) of the autocovariance function of a random variable. We explore methodologies and propose several tests to ease this dilemma. Our analysis is based on the climacogram and the autocovariance function (see sect. 2.3 and 4). The two tools impose basic conditions on the reliability 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 the process and the empirical one. As we move from smaller to larger scales and conversely, we can easily determine which model has the lowest error and bias, which one we should choose (Fig. 3 and 4). In case we have limited timeseries 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. 3 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
