European Geosciences Union General Assembly 2013
Vienna, Austria, 7-12 April 2013
Session HS7.5/NP8.4 : Hydroclimatic Stochastics

Clustering of extreme events in typical stochastic models

A. Mystegniotis, V. K. Vasilaki, I. Pappa, S. Curceac, D. Saltouridou, N. Efthimiou, G. Papatsoutsos, S. M. Papalexiou and D. Koutsoyiannis

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

Available online: itia.ntua.gr/1332/

1. Abstract

We study the clustering properties of extreme events as produced by typical stochastic models and compare the results with the ones of observed data. Specifically the stochastic models that we use are the AR(1), AR(2), ARMA(1,1), as well as the Hurst-Kolmogorov model. In terms of data, we use instrumental and proxy hydroclimatic time series. To quantify clustering we study the multi scale properties of each process and in particular the variation of standard deviation with time scale as well of the frequencies of similar events (e.g. those exceeding a certain threshold with time scale). To calculate these properties we use either analytical methods when possible, or Monte Carlo simulation.

Acknowledgement: *This research is conducted within the frame of the undergraduate course "Stochastic Methods in Water Resources" of the National Technical University of Athens (NTUA). The School of Civil Engineering of NTUA provided moral support for the participation of the students in the Assembly.*

2. Methodology

The analysis of extreme events plays an important role in engineering practice for water resources design and management. Clustering of extreme events has been reported throughout history, even from the antiquity. A classic example offering long-term information on clustering is the Nilometer time series, which records the annual minimum water levels of the Nile River for a period of 663 years.

In this paper, typical stochastic models are fitted to the statistical characteristics (mean, standard deviation, autocorrelation coefficient of order 1 and 2 , Hurst coefficient) of the Nilometer time-series and are used to generate synthetic series and then assess the clustering behaviour of extremes.

Specifically, we generate 2000 time-series with a sample size of 200 values each, using the following models (collectively referred to as simulation A):

- Autoregressive of order 1 (AR(1)) and 2 (AR(2))
- First order Autoregressive-First order Moving Average (ARMA(1,1))
- Hurst-Kolmogorov(HK, aka FGN)

In addition, to assess the effect of the Hurst coefficient on the clustering behaviour for the HK model, this model is also tested with different Hurst coefficients, spanning from 0.5 to 0.95 (collectively referred to as simulation B).

The analysis is made on several aggregate time scales, from 1 to 20 years, where the upper and lower 5% of the generated values are regarded as extremes.

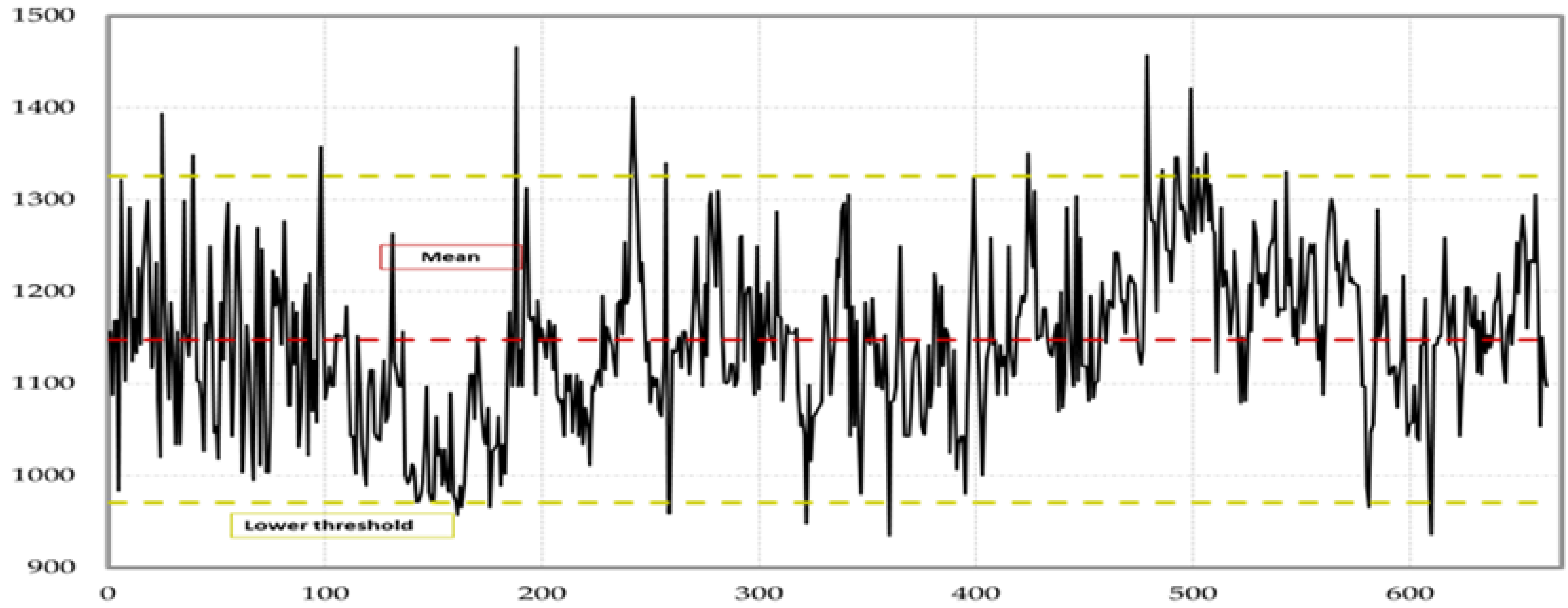
Alternatively, extremes are defined to be those values that are above or below a certain threshold; the thresholds are specified as $m \pm 2s$, where m and s are the mean and standard deviation.

3. Methodology (2)

We also examine the mean value, standard deviation and skewness of the frequency of extreme events, as well as the 99th percentile of that frequency.

The importance of the current year value in order to predict maxima over threshold in a 10-year and 50-year period is also examined, along with the distribution of the temporal distance between two consecutive maxima over threshold.

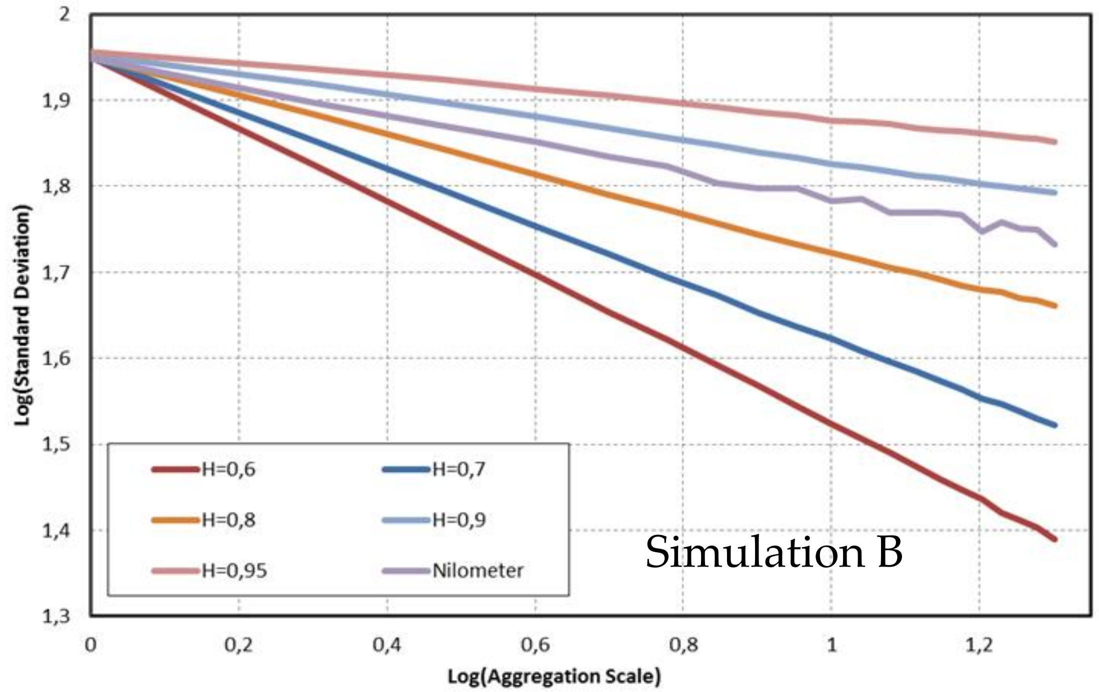
Due to the fact that the normal distribution was used to produce the time series, minima under threshold are not usually mentioned since they do not produce any new results due to symmetry.



Nilometer time series (annual minimum water levels of the Nile in cm) (data from Beran,1994)

4.Standard deviation vs. aggregate time scale

For increasing Hurst exponent, the decrease in standard deviation, with aggregation scale becomes milder. The mild slope is one of the main features that indicates long term persistence and clustering. The Nilometer time-series is well approximated by FGN with H=0.85.

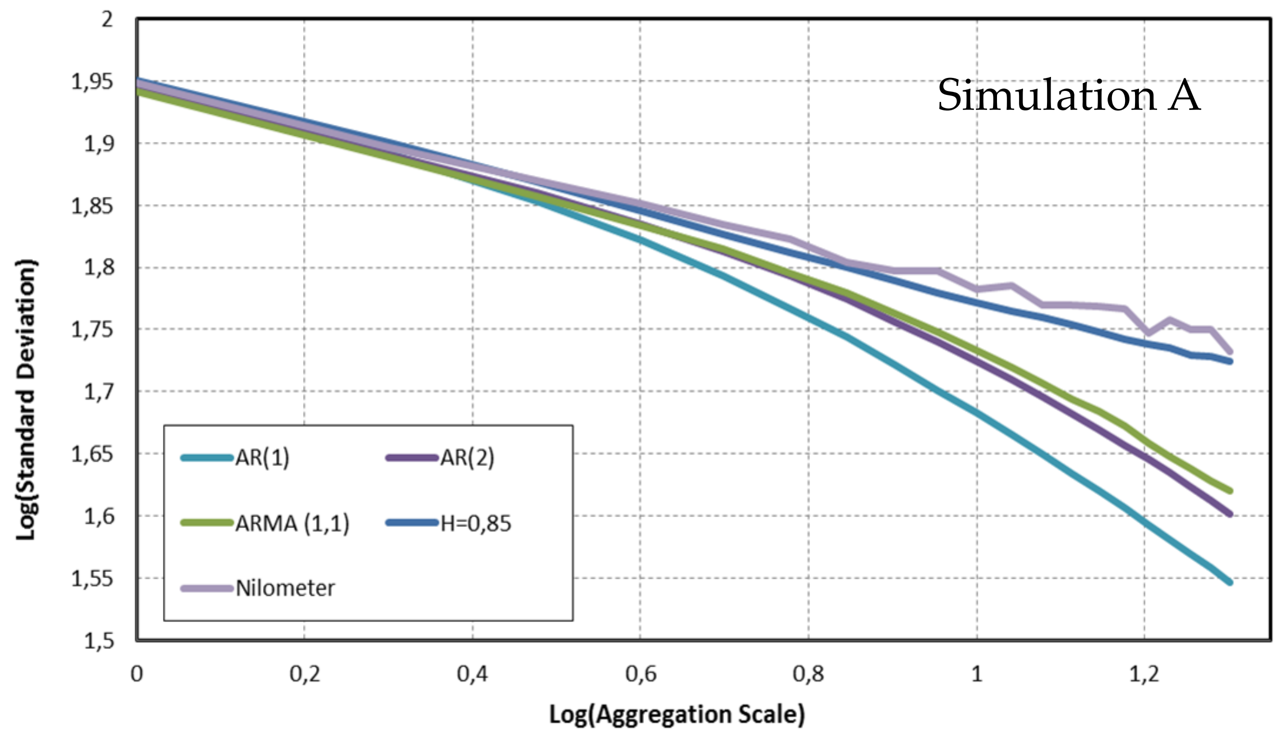


Standard Deviation (unbiased) in aggregation scales (up to 20)

For the FGN case, we have calculated the unbiased standard deviation using the estimator :

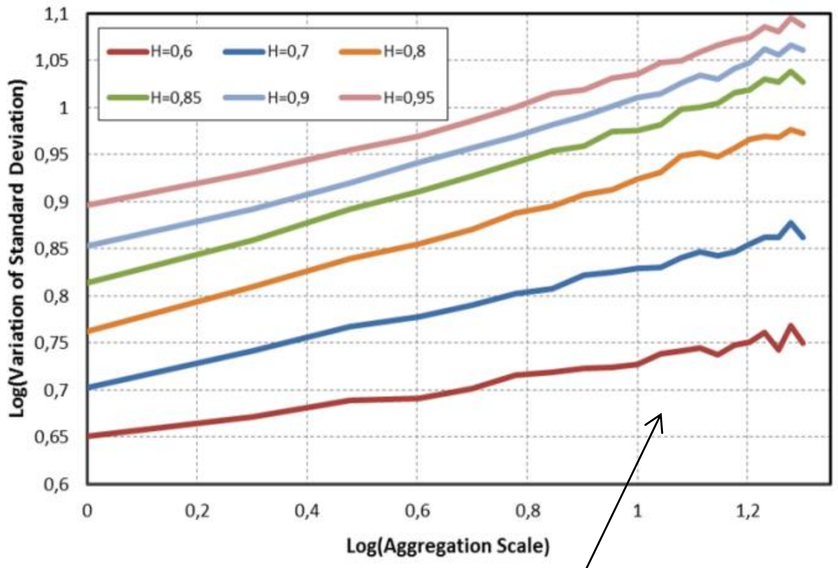
$$\tilde{\tilde{S}} := \sqrt{\frac{n - 1/2}{n - n^{2H-1}}}\, S$$

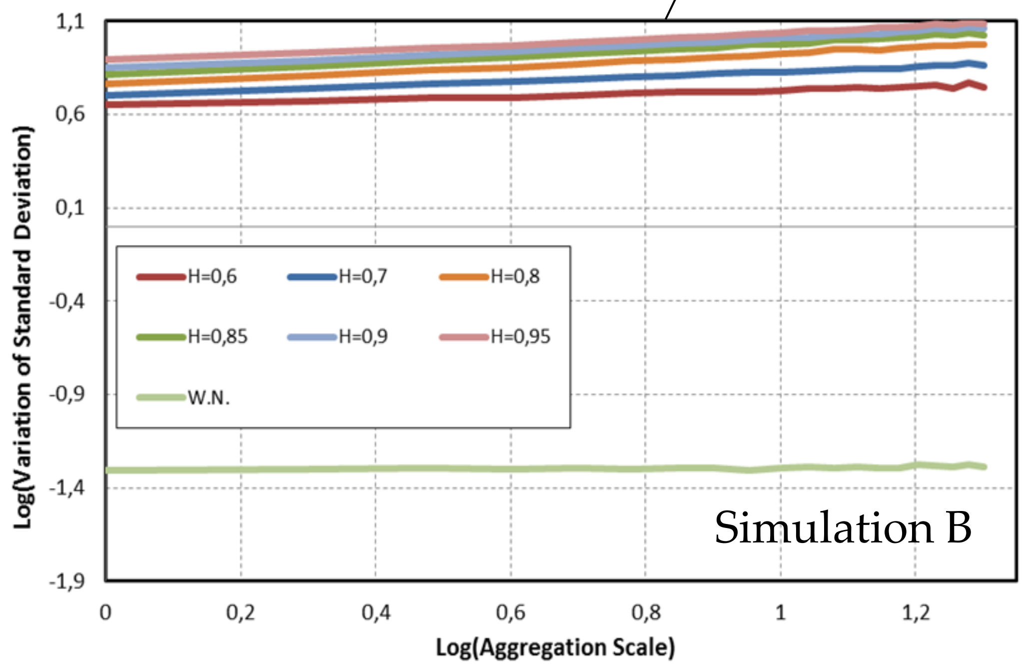
where S is the classical sample estimator of standard deviation and n is the sample size (Koutsoyiannis, 2003).

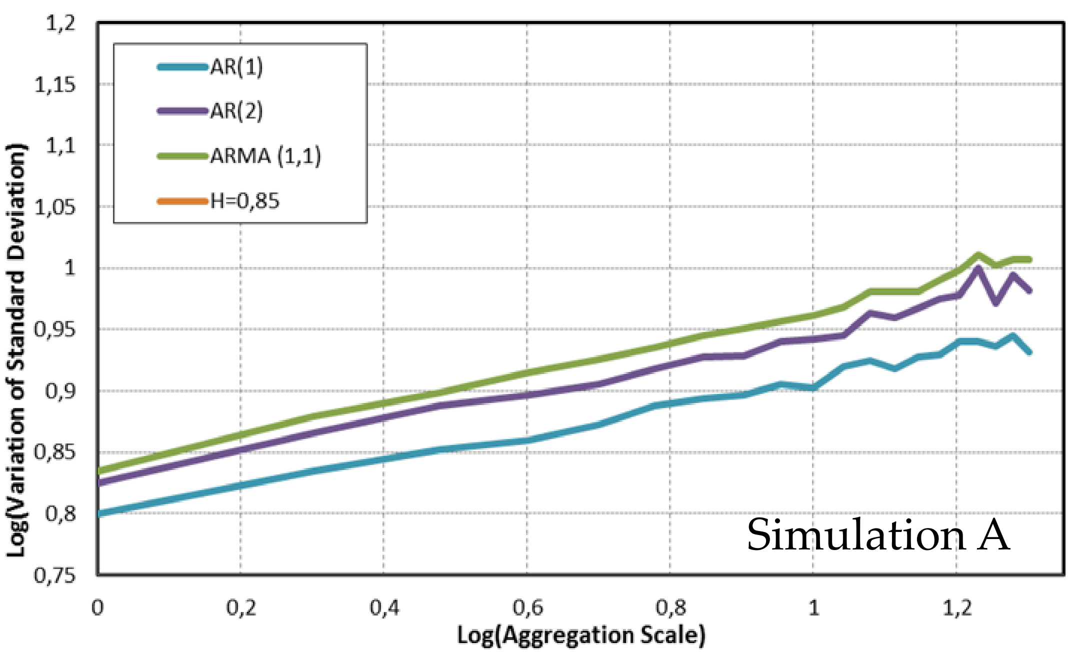


5.Variation of standard deviation with temporal aggregation

For ascending aggregation scale, in all models the standard deviation exhibits higher variability (this can be explained theoretically; cf. Koutsoyiannis 2003). Also these values increase with increased Hurst coefficient (for $H = 0.5$ the standard deviation of sample standard deviation is almost constant for all scales).

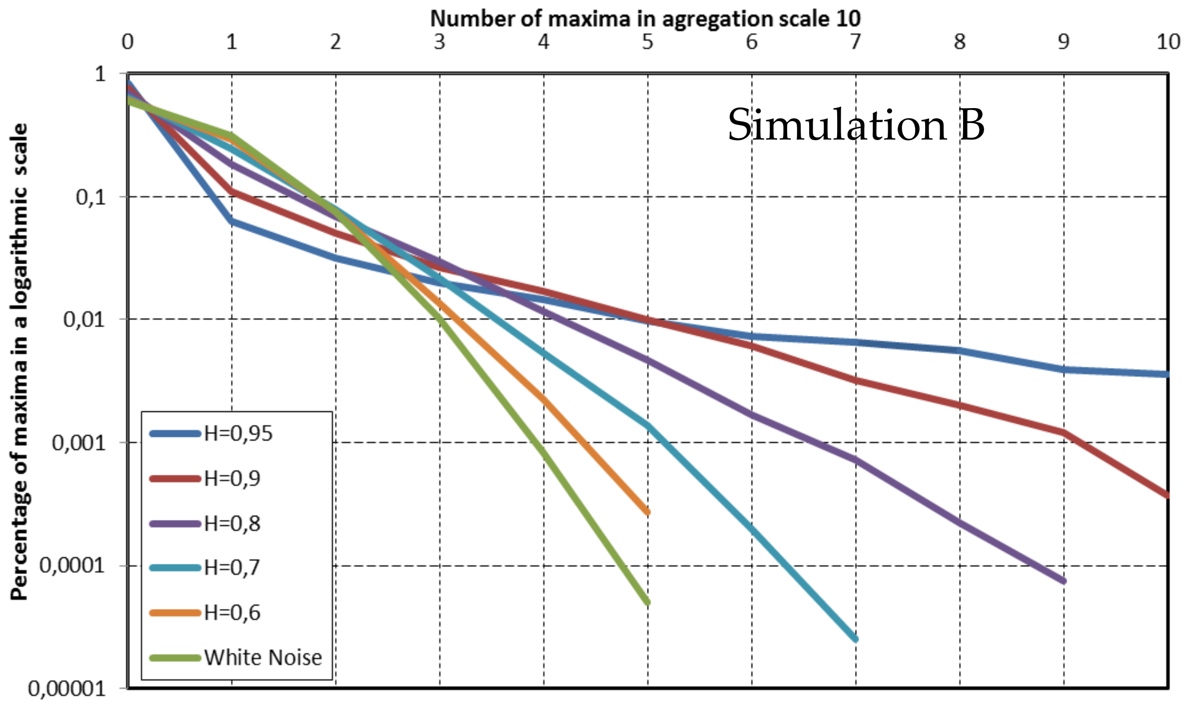


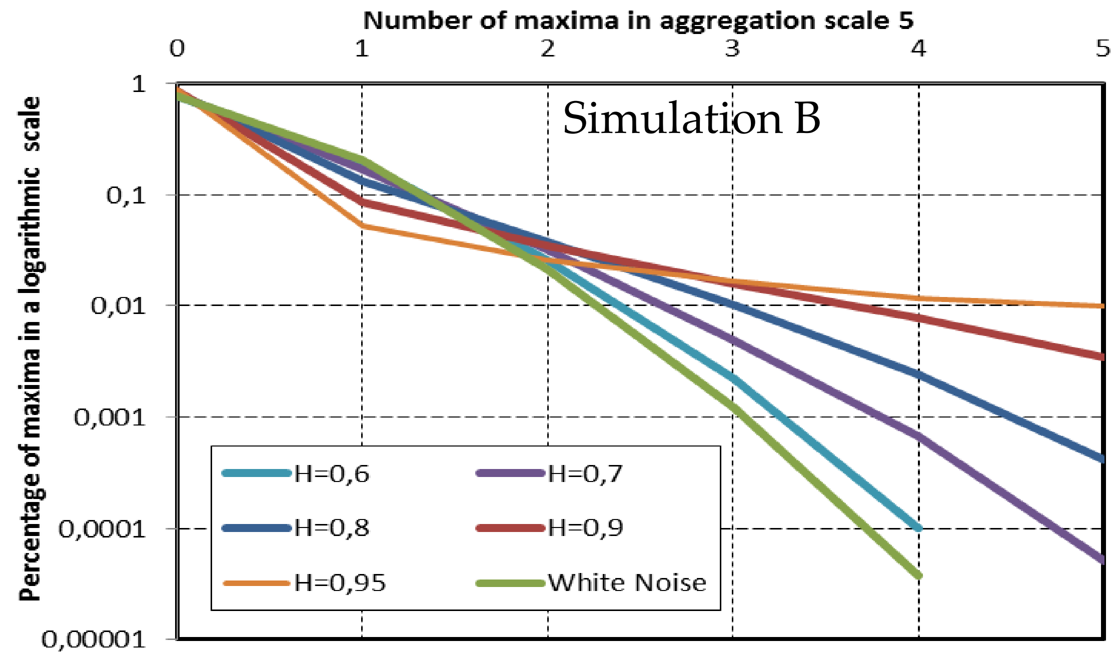


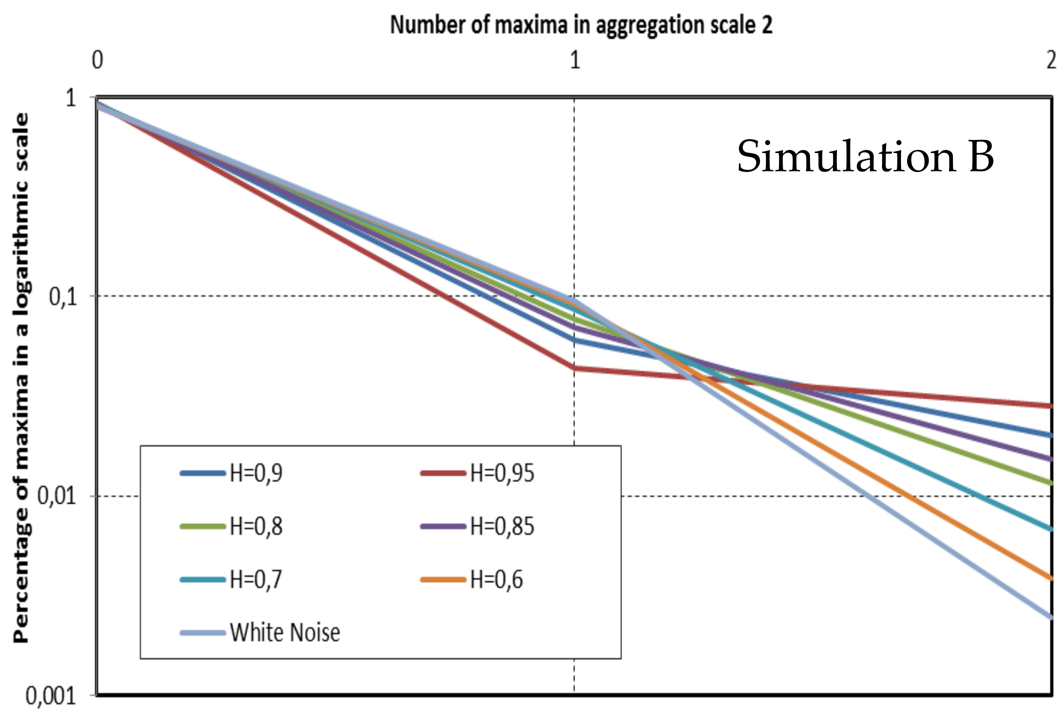


These graphs show the variation of standard deviation with aggregation scales (up to 20). W.N. abbreviates white noise.

6. Frequency of extremes in different temporal aggregation scales for the HK model

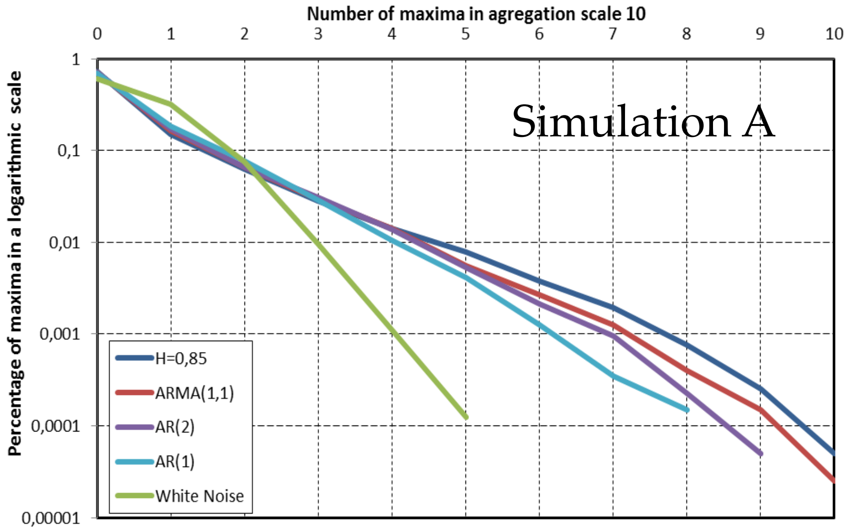


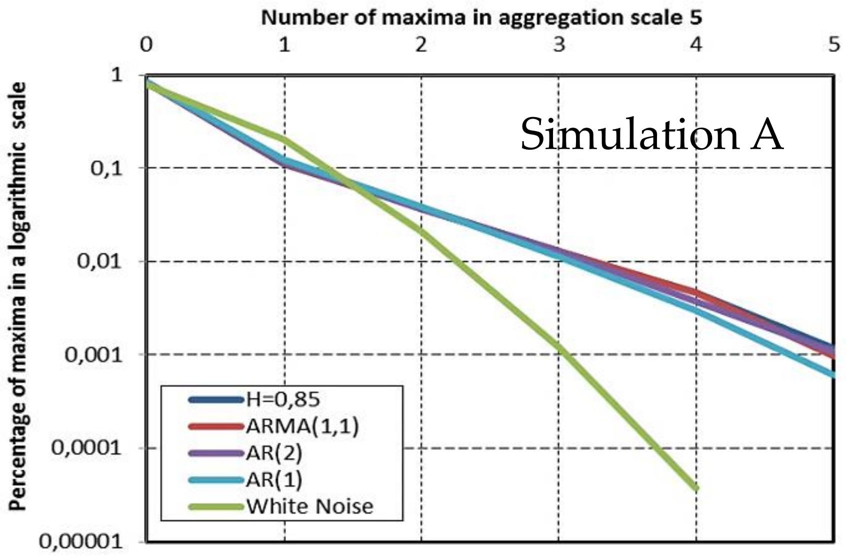


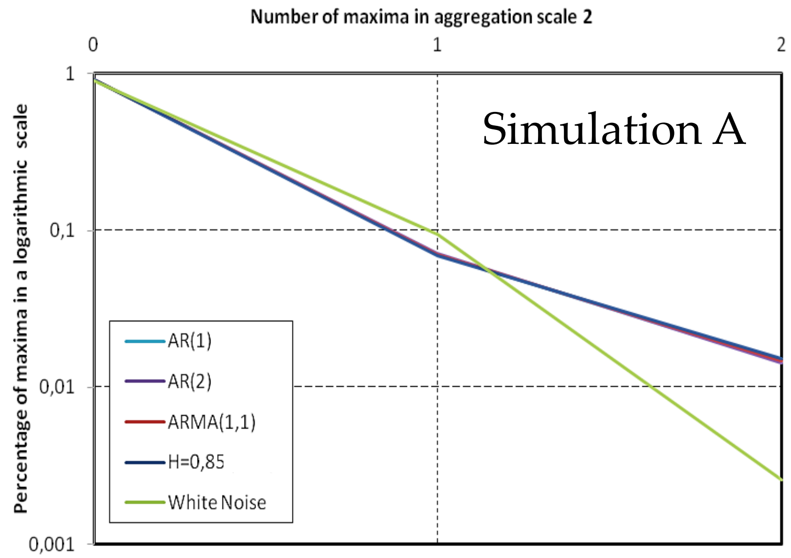


As expected, for small numbers of maxima in aggregate scales, the higher Hurst exponents come along with lower frequencies. This picture is reversed for higher numbers of extremes. In addition, the frequency of high numbers of extremes increases dramatically (by several orders of magnitude) with the increase of the Hurst coefficient. That indicates, also, the stronger clustering that goes with long term persistence.

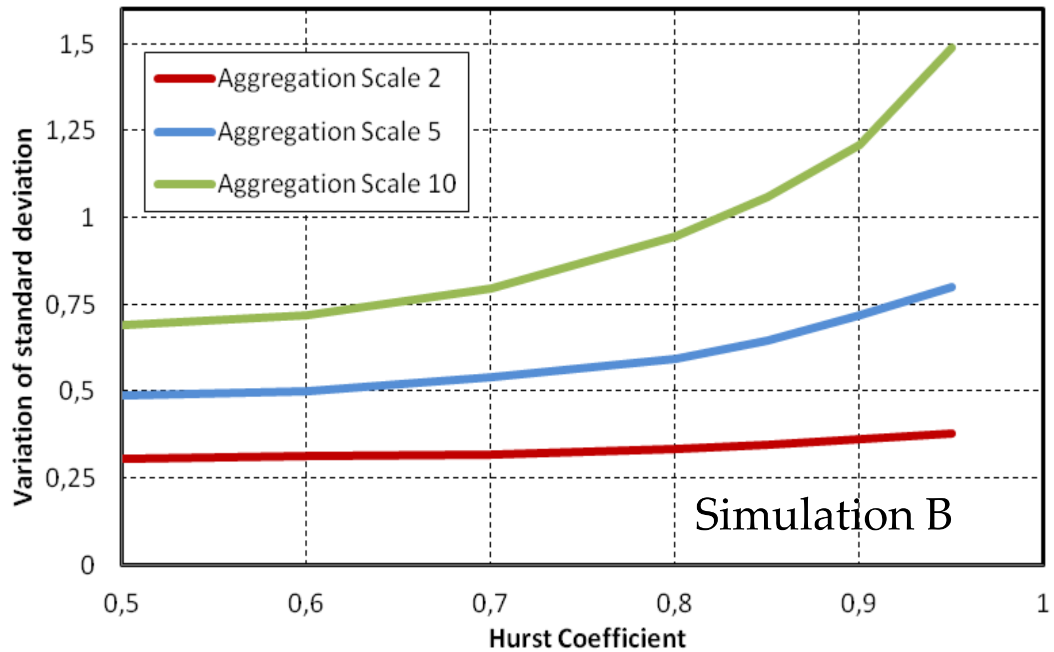
7. Frequency of extremes in different temporal aggregation scales for different stochastic models







For each model we used the percentage of 5% to define what is considered as extreme.
It is obvious that the FGN models result in higher frequencies of large numbers of extremes (for instance >4 in a decade).

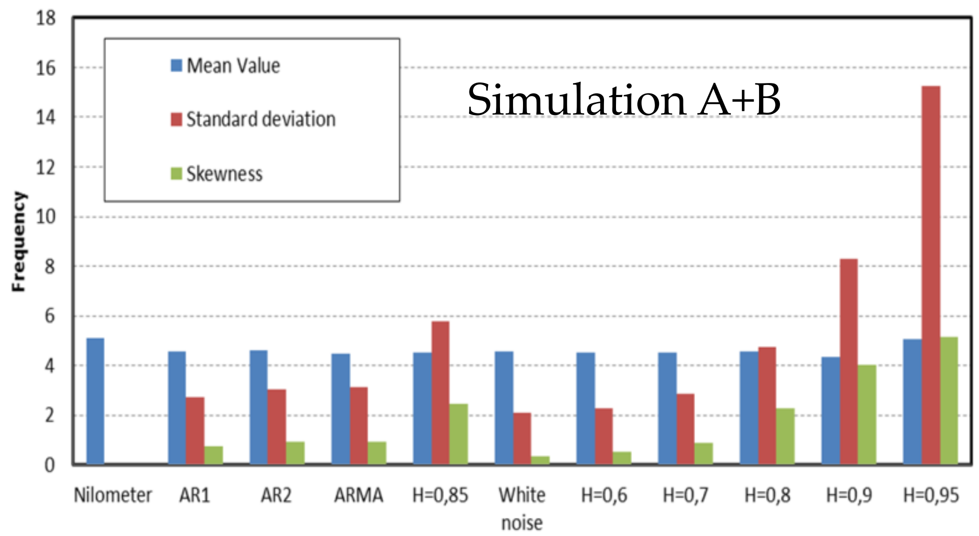


Variation of standard deviation for different aggregation scales vs. Hurst coefficient.

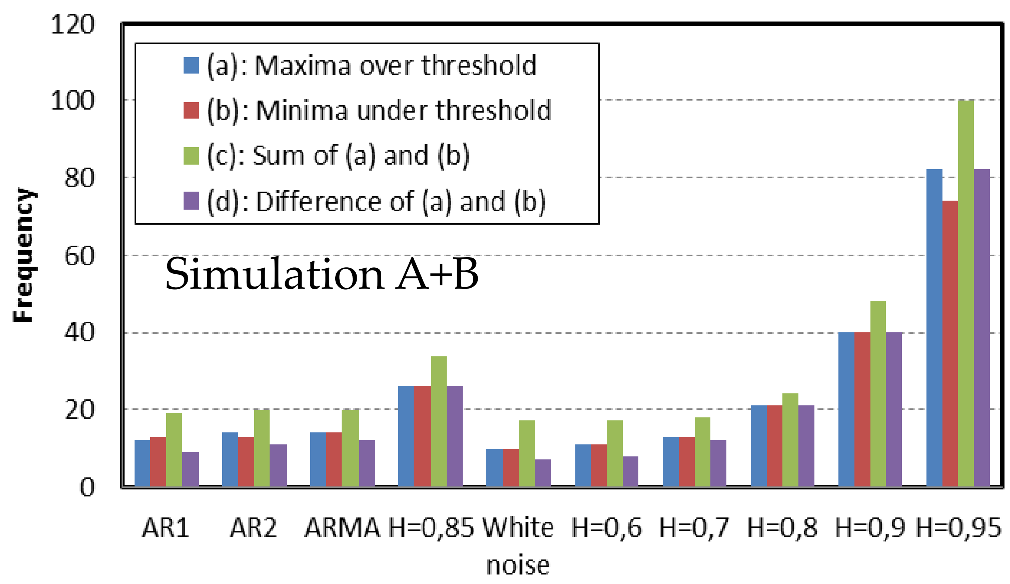
8. Statistical indices and frequency of maxima over threshold

From the chart on the right we can observe the following:

- The mean value remains unchanged regardless of the model used.
- Standard deviation and skewness vary greatly depending on the model used.
- Those two indices are much higher in the HK model and are strongly affected by the Hurst exponent.



Mean value, standard deviation and skewness of the frequency of maxima over threshold.

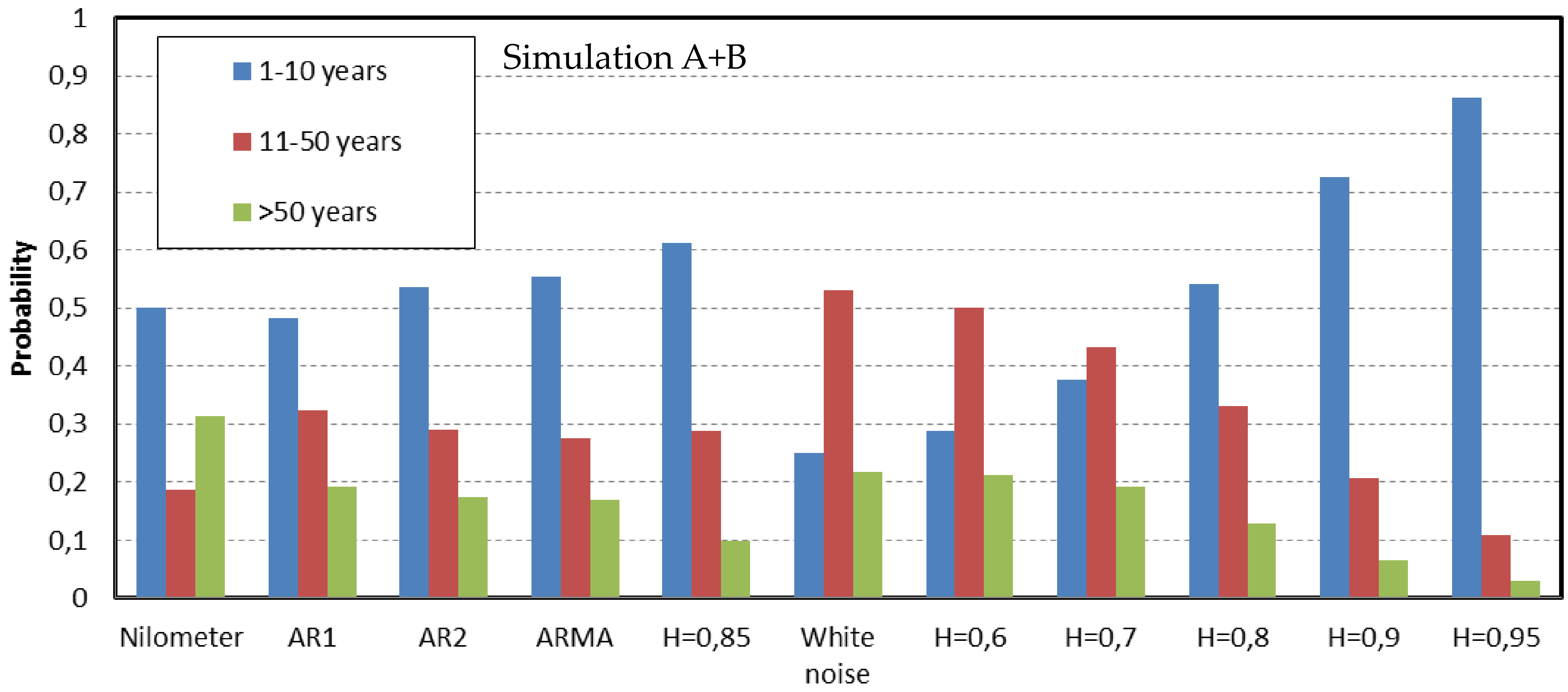


Frequency of the 99th percentiles of (a) maxima over threshold, (b) minima under threshold, (c) sum of (a) and (b), and (d) difference of (a) and (b).

The Hurst- Kolmogorov model results in much higher frequency of extreme events than all the other models. Unsurprisingly, the rate of increase, with ascending Hurst exponent, of the 99th percentile of the numbers of extremes over or under threshold, matches the rate of increase of the standard deviation. The fact that the difference between the extremes over and under threshold is as high as the total frequency of the maxima over threshold if the Hurst exponent is high enough, means that in those cases (at the 99th percentile) only maxima over threshold are observed.

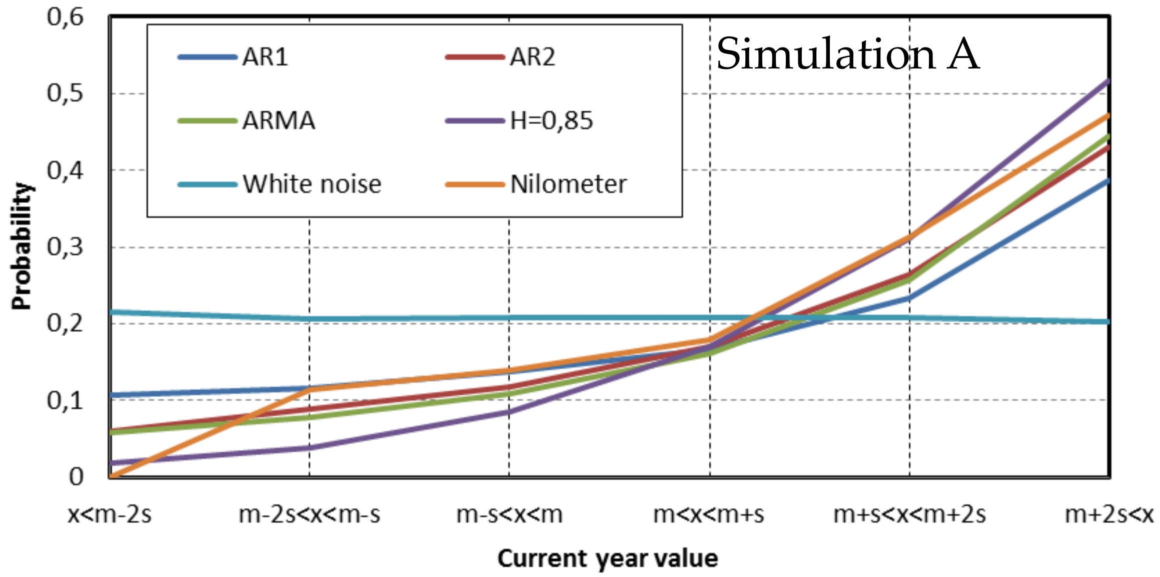
9. Distribution of temporal distances between maxima over threshold

Probability that the temporal distance between maxima belongs to one of the three specified classes



We observe that the Hurst- Kolmogorov model is the one that favours clustering the most, as it yields the highest probability that the time distance between two maxima over threshold be 10 years or less, and the lowest probability to be more than 50 years. An increase of the Hurst exponent amplifies this behaviour.

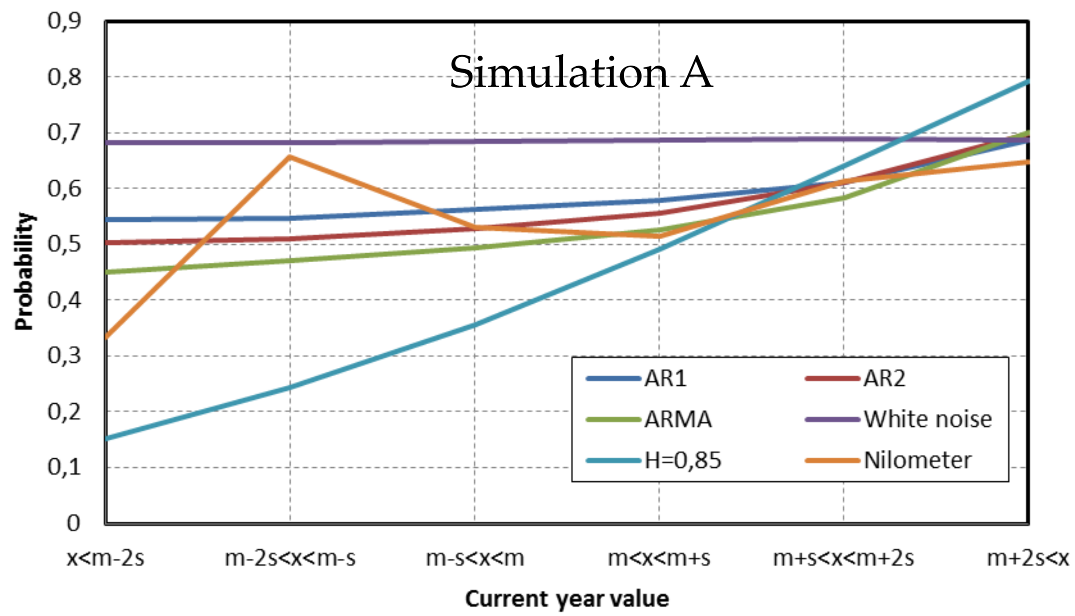
10. Conditional probability of the next maximum over threshold



Probability of encountering a maximum over threshold in a 10 year period, conditional on the current year value being in each of the indicated classes

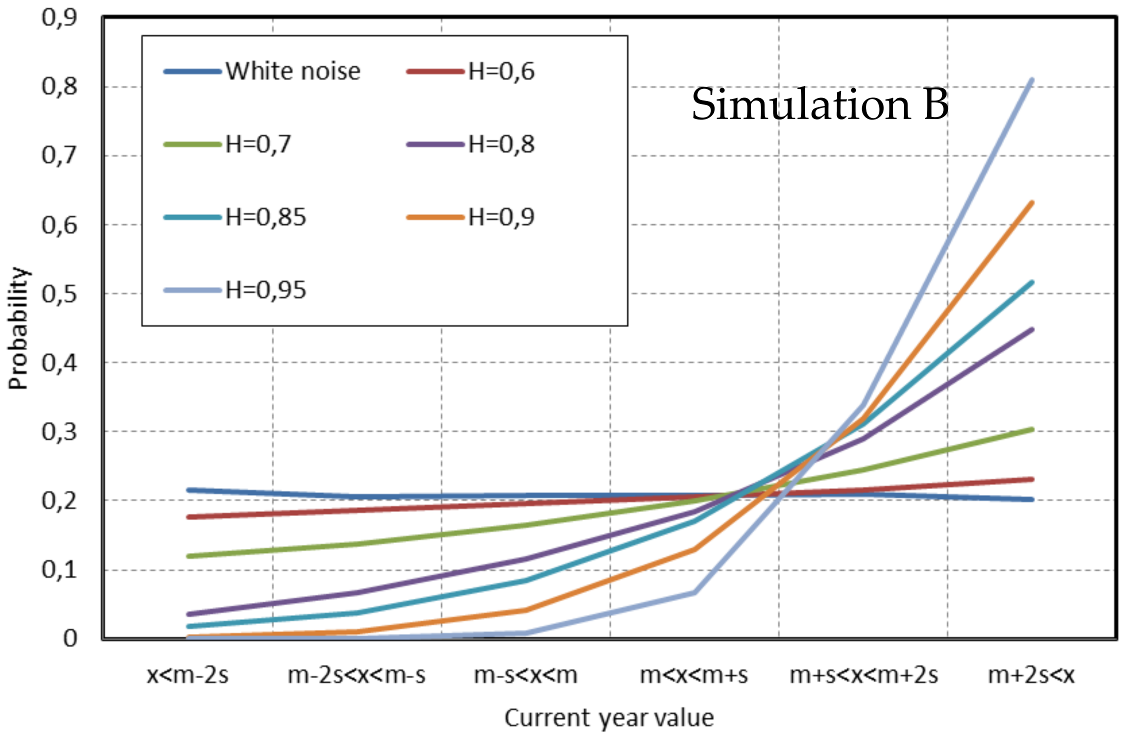
Although the curve of the Hurst- Kolmogorov model is steeper, all models produce comparable results in this time frame, and the observed probability is highly dependent on the current year value (except for white noise).

Probability of encountering a maximum over threshold in a 50 year period, conditional on the current year value being in each of the indicated classes

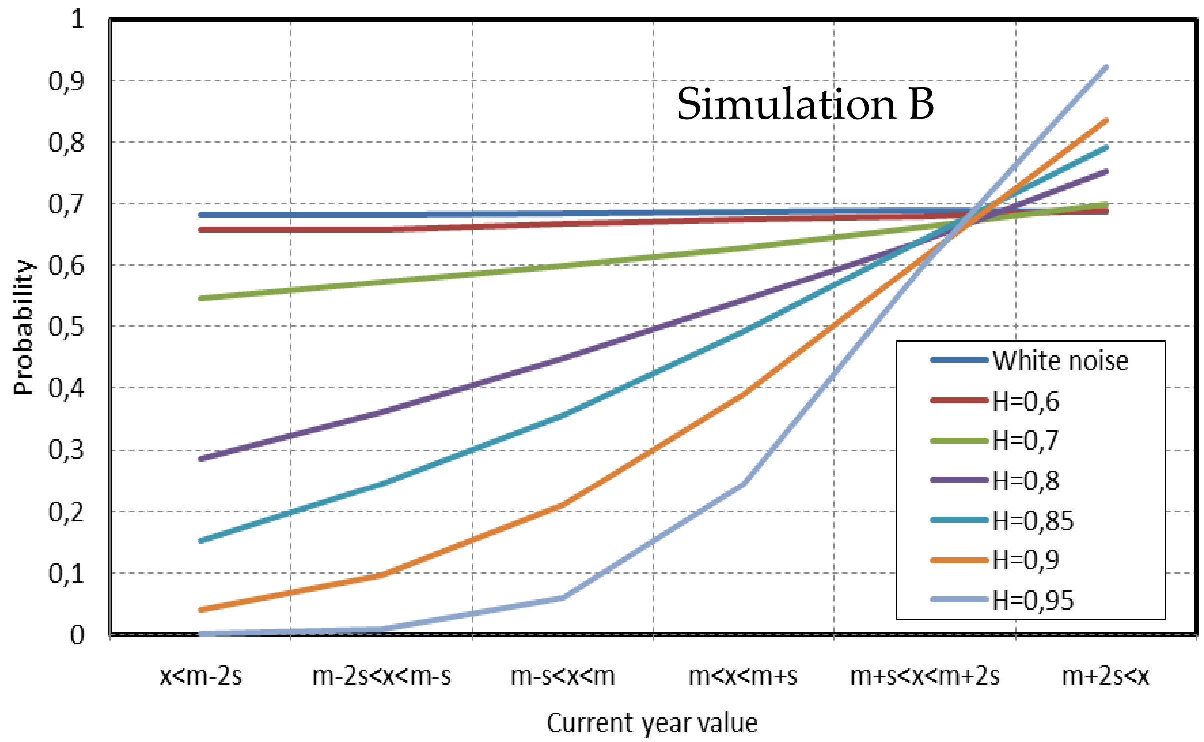


- For the AR1, AR2 and ARMA(1,1) models while the effect of the current value is not negligible, it is not as important as the 10-years case.
- For the Hurst-Kolmogorov model the probability of encountering a maximum over threshold in a 50 year period is highly dependent on the current year value as it ranges from 15% to 80%.

11. Importance of the Hurst exponent in clustering



Probability of encountering a maximum over threshold in a 10 year period, conditional on the current year value being in each of the indicated classes



Probability of encountering a maximum over threshold in a 50 year period, conditional on the current year value being in each of the indicated classes

The importance of the Hurst exponent can be seen clearly in the above charts as well, as it can greatly influence the probability of encountering a maximum both in a 10 year and 50 year period according to the current year value.

A high Hurst exponent can lead to a more intense clustering with more extremes in a given time frame, and it can increase the density of clusters in periods where high values are encountered, and drastically decrease it when they are not (cf. Bunde *et al.* 2005).

It is important to note that a large amount of data is required to estimate the Hurst exponent accurately and that even small increases can have an important effect in the clustering properties of a time series.

12.Conclusions

- The standard deviation of a process at aggregate time scales, provides a simple indication of clustering. This standard deviation generally decreases with time scale, but the rate of decrease is reduced for ascending Hurst exponent.
- For ascending Hurst coefficient, the clustering of extreme is more intense. Even the case where all the particular values in certain period (of length 2-10 time steps) are extremes holds a non-negligible probability.
- The HK model results in stronger clustering effect compared to other models (with equal autocorrelation at small lags); the difference becomes even more evident at higher time scales.
- Except for white noise, an extreme value in the current year decreases the distance to the next extreme value, while an absence of an extreme value increases that distance; this is amplified if the Hurst exponent is high.

References

- Beran, J., *Statistics for Long-Memory Processes*, Chapman and Hall, New York, 1994.
- Bunde, A., J. F. Eichner, J. W. Kantelhardt, S. Havlin, Long-term memory: A natural mechanism for the clustering of extreme events and anomalous residual times in climate records, *Physical Review Letters*, 94, 048701, 2005.
- Koutsoyiannis, D., Climate change, the Hurst phenomenon, and hydrological statistics, *Hydrological Sciences Journal*, 48 (1), 3–24, 2003.