European Geosciences Union
General Assembly 2013

Vienna, Austria,
7 - 13 April 2013

Session: HS7.5/NP8.4
Hydroclimatic stochastics

Assessment of the dependence structure of the annual rainfall based on a large data set

T. Iliopoulou, S.M. Papalexiou, and D. Koutsoyiannis

Department of Water Resources and Environmental Engineering,

National Technical University of Athens, Greece

1. Abstract

Natural processes are considered to be influenced by long-term persistence, the so-called Hurst effect. A variety of studies have been conducted to identify the Hurst behavior in different data sets and different scientific disciplines ranging from geophysics to economics and to social sciences. In this study we try to test the hypothesis of the existence of long-range dependence in annual rainfall by applying the aggregated variance method in a large set of annual rainfall time series from more than a thousand stations worldwide. In addition, we figure out a simple statistical test in order to assess the hypothesis that the dependence structure of annual rainfall is Markovian.

2. Motivation - Methodology

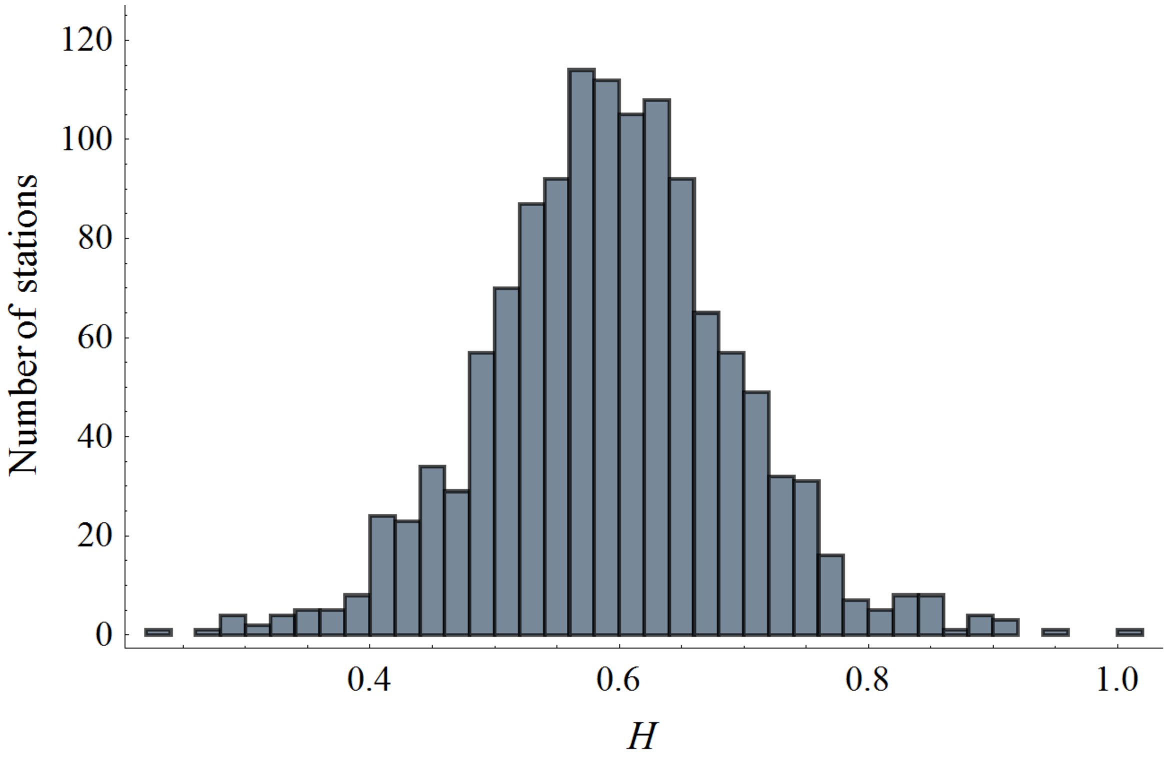
- High scientific interest is raised around the hypothesis testing of long range dependence in geophysical processes.
- We examine the hypothesis of long range dependence in annual rainfall by using a global data set, namely the GHCN-Daily.
- In order to estimate the Hurst coefficient we apply the aggregated variance method as well as the LSSD (Least Squares based on Standard Deviation) method. (The difference of the two is that the former uses the classical estimator of variance, which is biased, while the latter uses an approximately unbiased estimator of the standard deviation).
- We also test whether a common Hurst coefficient can be assumed for all records studied.
- In addition, we study the autocorrelation structure and compare it to a simple AR(1) structure.

3. Data

- From a total of 50 000 stations with daily rainfall records from GHCN-Daily (Global Historical Climatology Network http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/) we choose 3477 stations that satisfy the following quality criteria:
  - Record length over 100 years;
  - Missing values less than 20% of record length.
- For better quality control we delete daily values assigned with quality flags indicating unrealistically large daily values.
- We construct the annual average of daily rainfall time series by estimating the annual average value only for years with more than 345 daily records.
- We infill individual gaps with the average of the previous and the following annual rainfall value (Παππάς 2010).
- Finally, a sample of 1265 stations with 100 non-missing values and missing values less than 15% of the total record length, is chosen for the analysis.

4. Aggregated Variance method results

- 85% of stations have $H > 0.5$
- 50%, $H > 0.59$
- 25%, $H > 0.653$
- 2.5%, very strong dependence structure with $H > 0.8$
- For 95% confidence interval the estimates are in the range 0.4 – 0.8

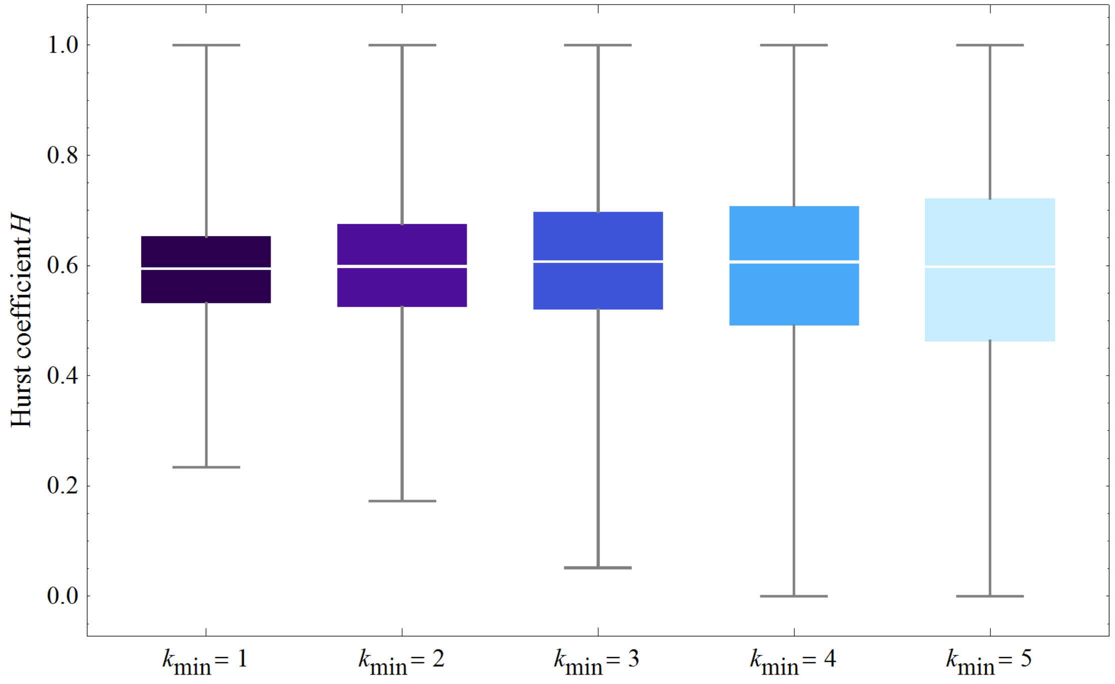


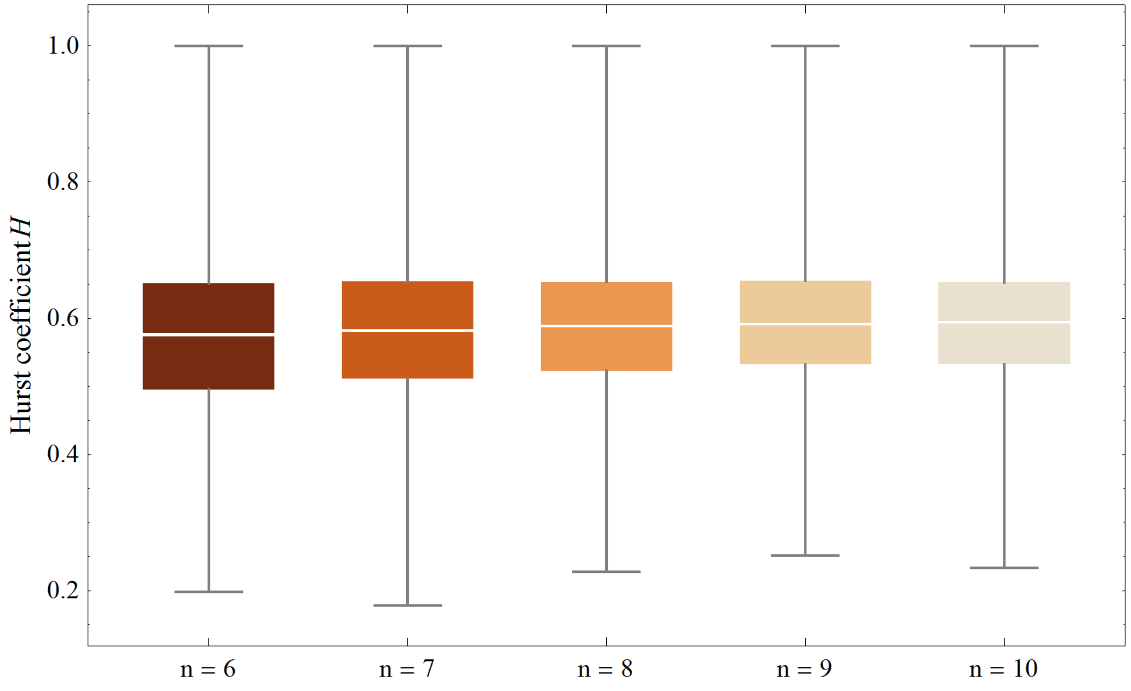
Algorithmic choices: minimum scale k, minimum number of values n in the maximum scale **min k =1, n = 10**

min	$Q_{2.5}$	Q_{25}	Q_{50}	Q_{75}	$Q_{97.5}$	max	mean	Sd
0.23	0.402	0.532	0.594	0.653	0.799	0.99	0.594	0.01

5. Effects of algorithmic choices

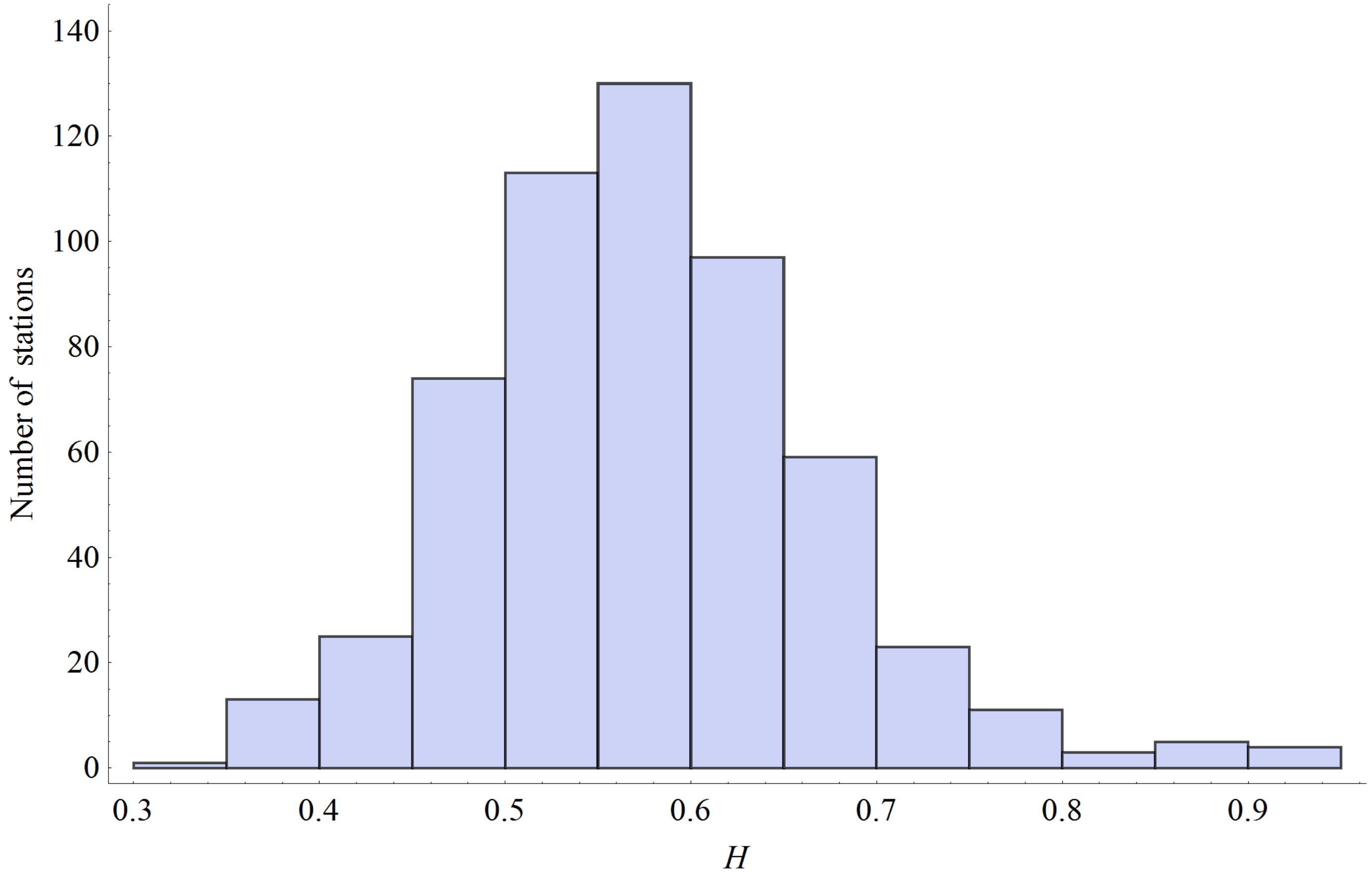
- We repeat the estimation for different choices in the algorithm of estimating *H* by the Aggregated Variance method.
- It is shown that the increase in the value of the minimum scale as well as the decrease in the number of values in the maximum scale do not affect the value of the median in the Hurst parameter estimate but they increase the variance of the estimation.
- The preservation of the same median although the change in the minimum scale is a positive fact towards the existence of long range dependence as the method would show differences in the estimation in the case of short term dependence (due to curvature in the latter case).





6. LSSD method results

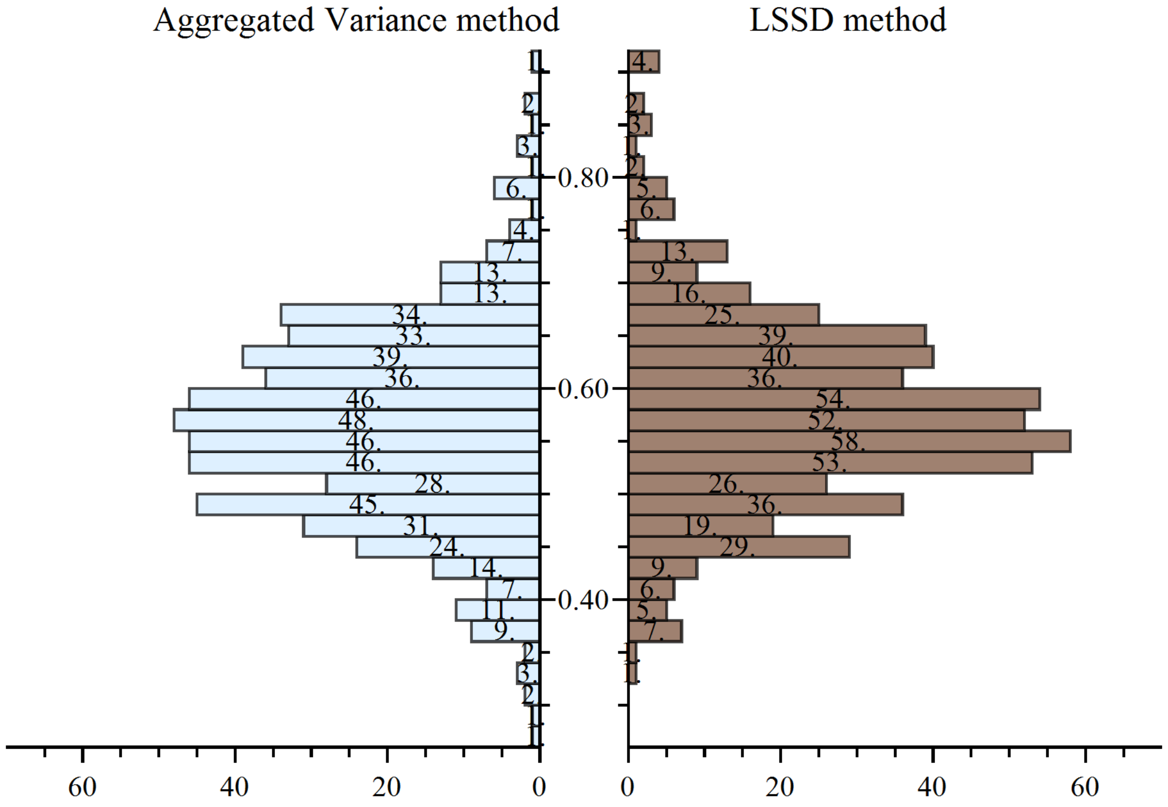
- The LSSD method (Least Squares based on Standard Deviation) is a known algorithm (Koutsoyiannis 2003, Tyralis and Koutsoyiannis 2011) which reduces the negative bias of the aggregated variance method.
- The method is applied to 558 stations (44% of the data) with no missing values.



7. Comparison of the two methods

- We apply the aggregated variance method to the same sample of 558 stations in order to compare the results.
- There exists 1-2% negative bias in the estimation of the Hurst coefficient by the aggregated variance method which is evident in the number of high ($H > 0.8$) and low ($H < 0.4$) estimates that each method produces.

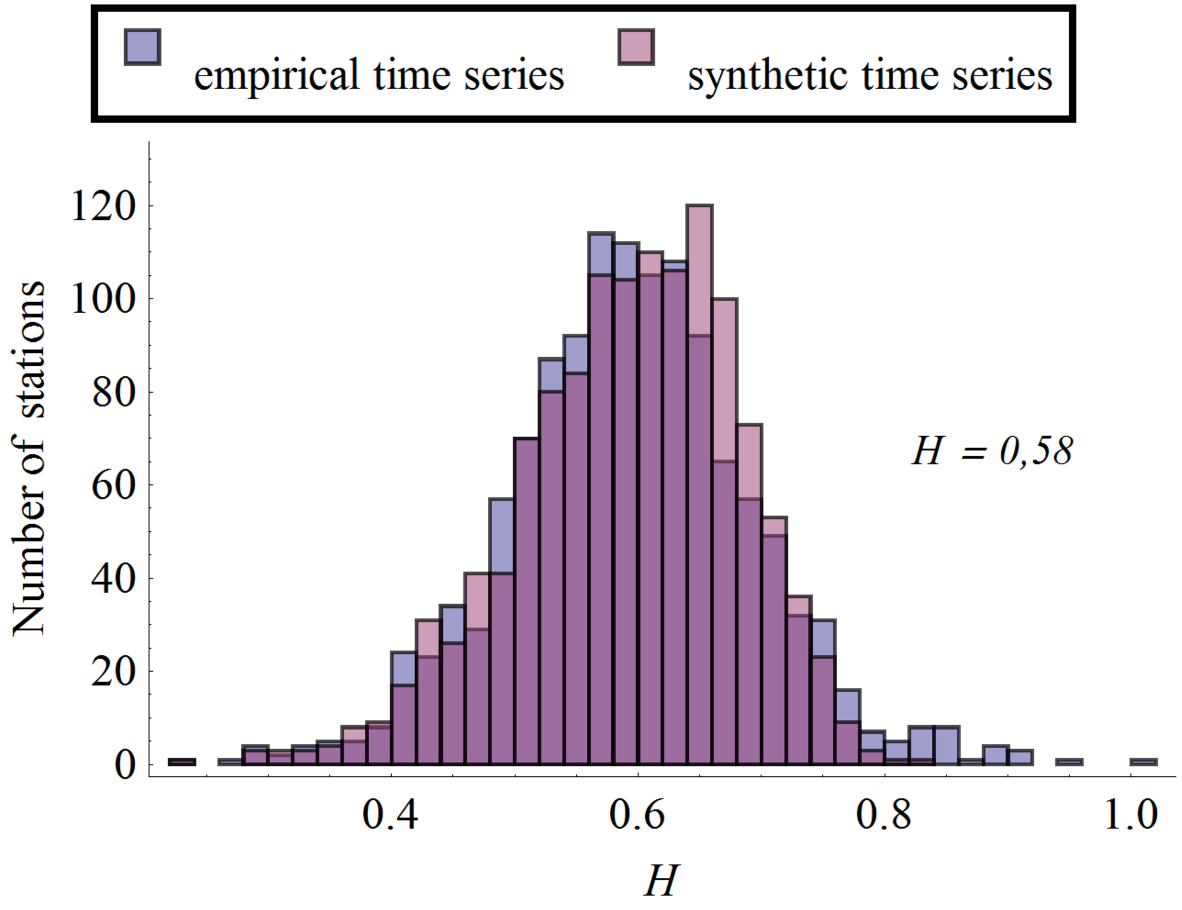
	Aggregated Variance method	LSSD method
Mean value	0.564	0.576
Standard Deviation	0.099	0.094
Min	0.276	0.334
$Q_{2.5}$	0.369	0.398
Q_{25}	0.496	0.52
Median	0.564	0.572
Q_{75}	0.63	0.635
$Q_{97.5}$	0.782	0.794
Max	0.904	0.916



8. Can a common Hurst coefficient be assumed?

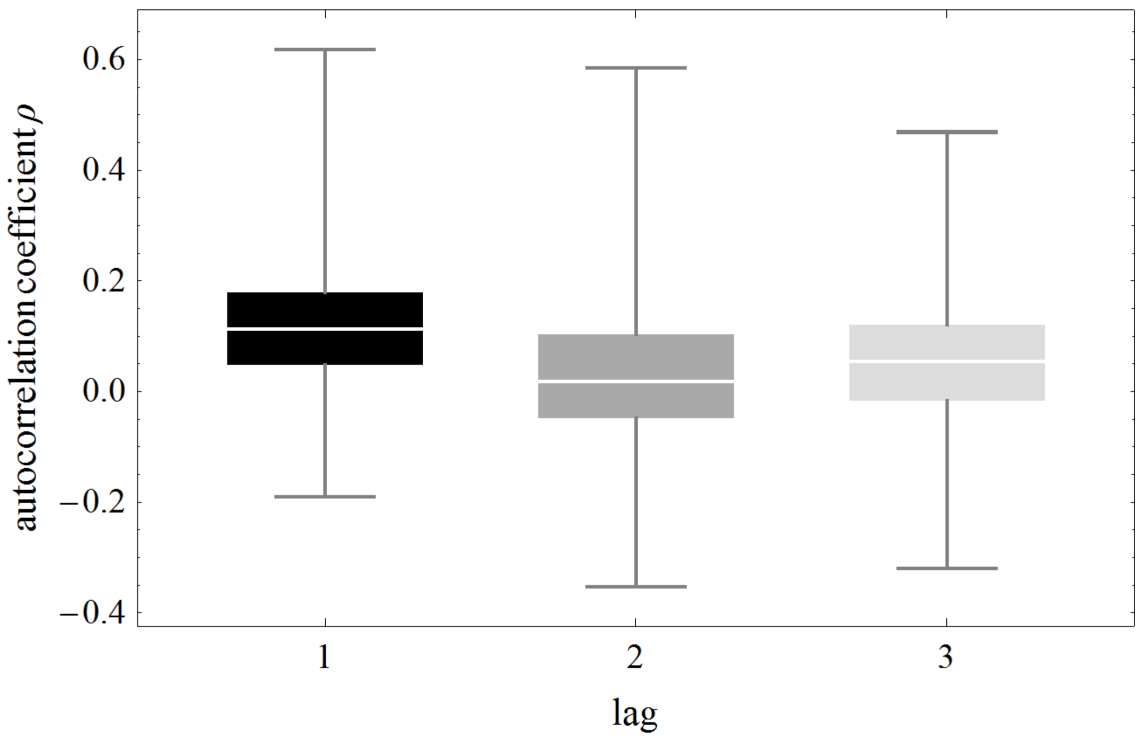
- We produce 1265 synthetic time series with a simple algorithm generating Fractional Gaussian Noise based on a multiple timescale fluctuation approach (Koutsoyiannis 2002).
- For each station, the sample size, the mean and the standard deviation are preserved.
- We repeat the same procedure for different theoretical values of *H* and then estimate the sample values *H*.
- The distribution of the sample estimates for the synthetic time series is compared to the distribution of the sample estimates for the empirical time series.
- Multiple trials are performed in order to find the Hurst coefficient *H* for which the match is best.

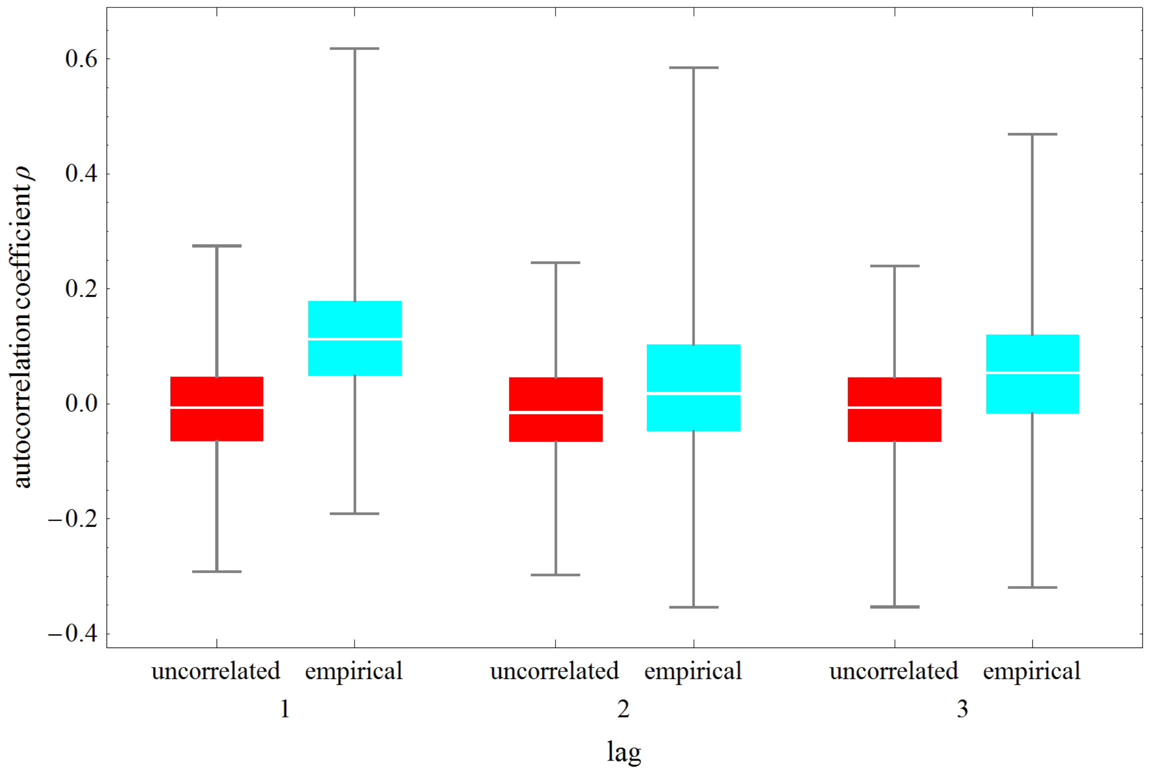
✓ Best distribution match for *H* = 0.58



9. Autocorrelation estimation

	ρ_1	ρ_2	ρ_3
Mean	**0.119**	**0.032**	**0.054**
Standard Deviation	0.113	0.116	0.109
Min	-0.191	-0.353	-0.319
$Q_{2.5}$	-0.095	-0.163	-0.154
Q_{25}	0.048	-0.048	-0.017
Q_{50}	0.113	0.018	0.054
Q_{75}	0.179	0.103	0.121
$Q_{97.5}$	0.37	0.286	0.273
Max	0.618	0.585	0.469





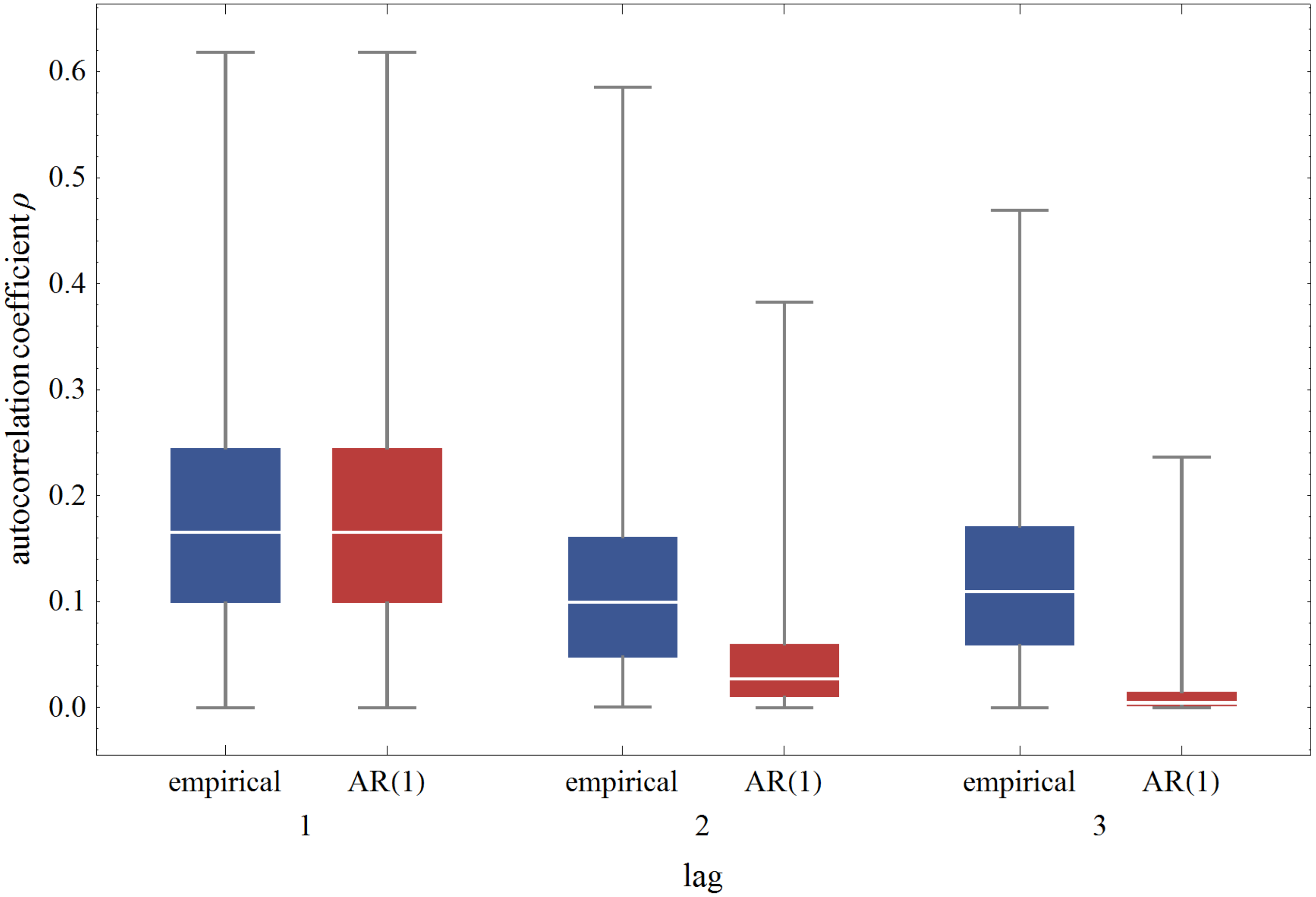
The estimated autocorrelation coefficients are low, still higher from the ones estimated from uncorrelated data.

10. Testing for Markov hypothesis

For 52% of the stations with positive autocorrelation coefficients the following ad hoc test is applied:

- The Markovian autocorrelation coefficient for lag 2 is estimated as $\rho_2 = \rho_1{}^2$, where ρ_1 the empirical estimation. Likewise, $\rho_3 = \rho_1{}^3$
- The resulting theoretical estimation is compared to the empirical one for the same lag.
- If the empirical value is higher than the theoretical AR(1) one, then given the negative bias existing in the autocorrelation estimation, the Markov hypothesis is invalidated and the data exhibit a stronger autocorrelation structure.

It is evident that the empirical estimates are considerably higher than the theoretical ones resulting from an AR(1) structure. The Markov hypothesis is invalid.



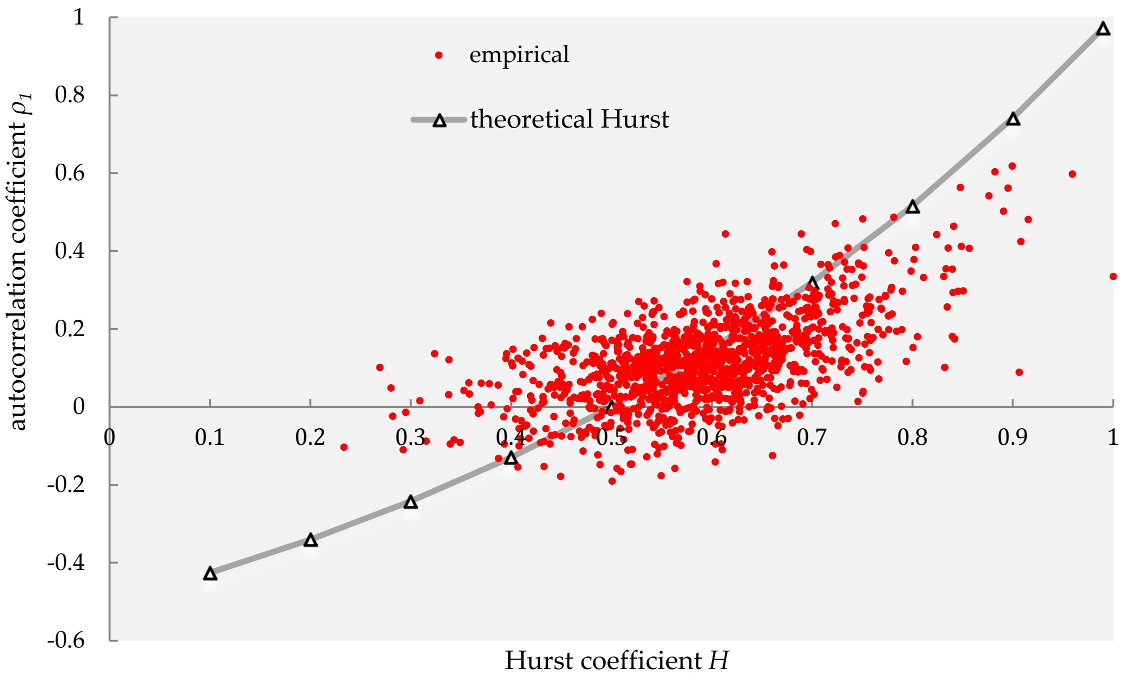
11. Comparison to FGN model

For a Fractional Gaussian Noise model the autocorrelation function for any aggregated timescale *k* is independent of the scale *k* :

$$\rho_j^{(k)} = \rho_j = \frac{1}{2}\left[(j+1)^{2H} + (j-1)^{2H}\right] - j^{2H}, \qquad j > 0$$

The empirical autocorrelation coefficient ρ_1 is plotted against the corresponding estimated Hurst coefficient *H*. The results are compared to the theoretical ones of a FGN model.

Given the existence of considerable negative bias in the estimation of the autocorrelation coefficient, the diagram shows that the autocorrelation structure is consistent with a FGN model.



12. Conclusions

- There are indications of existence of long range dependence in the annual rainfall.
- 85% of the rainfall stations exhibit $H > 0.50$ (presence of the phenomenon) but only 50% exhibit $H > 0.59$, that is to say stronger dependence structure.
- Because the Hurst parameter is not very high, the aggregated variance method induces only 1-2% negative bias in the Hurst coefficient estimation.
- A characteristic Hurst coefficient value $H = 0.58$ is representative for the majority of the stations.
- The study of the autocorrelation function shows that it is consistent with the autocorrelation of a FGN model. For the 52% of the stations the Markov hypothesis is invalidated.

References

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.

Tyralis, H., and D. Koutsoyiannis. 2011. 'Simultaneous Estimation of the Parameters of the Hurst–Kolmogorov Stochastic Process'. *Stochastic Environmental Research and Risk Assessment* 25 (1): 21–33.

Παππάς, Χ. 2010. 'Βέλτιστη συμπλήρωση ελλιπών υδρομετεωρολογικών δεδομένων με χρήση γειτονικών χρονικά παρατηρήσεων'. ΤΥΠΕ. http://www.itia.ntua.gr/el/docinfo/1065/.