

Record-breaking properties for typical autocorrelation structures

Eirini Anagnostopoulou, Andriani Galani, Panagiotis Dimas, Alexandros Karanasios, Theodoros Mastrotheodoros, Eleni-Maria Michaelidi, Dionisios Nikolopoulos, Stamatis Pontikos, Fani Sourla, Anna Chazapi, Simon Michael Papalexiou and Demetris Koutsoyiannis, Department of Water Resources and Environmental Engineering, National Technical University of Athens (itia.ntua.gr/1331)

1. Abstract

Record-breaking occurrences in hydrometeorological processes are often used particularly in communicating information to the public and their analysis offers the possibility of better comprehending extreme events. However, the typical comprehension depends on prototypes characterized by pure randomness. In fact the occurrence of record breaking depends on the marginal distribution and the autocorrelation function of the process as well as the length of available record. Here we study the influence of the process of autocorrelation structure on the statistics of record-breaking occurrences giving emphasis on the differences from those of a purely random process. The particular stochastic processes, which we examine, are the AR(1), AR(2) and ARMA(1,1), as well as the Hurst-Kolmogorov process.

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. Literature Review

Record-breaking theory was first introduced for identical and independent distributed (iid) variables⁽¹⁾, for example Gaussian white noise. Early works⁽²⁾ proved that record breaking statistics is non-parametric and produced general theoretical results. These include the probability (p_n) that the n th observation of a series is a record-breaking event and the expected value ($E[R(n)]$) of the number of record-breaking events occurred in a time-series of length n :

$$p_n = \frac{1}{n}$$

$$E[R(n)] = \sum_{i=1}^n \frac{1}{i} = \ln(n) + \gamma$$

where γ is the Euler-Mascheroni constant, $\gamma=0.5772156649$

Since the iid assumption is the only assumption required for these results, they can be used in testing the assumption⁽³⁾. It was observed that the average global temperature series has a strong autocorrelation structure and cannot be adequately represented by the iid assumption. On the contrary, there is an agreement of the results in maximum monthly 24h precipitation in Nordic countries with the results of iid assumption⁽⁴⁾.

3. Literature Review (continued)

It was confirmed that the iid assumption could fail in the presence of autocorrelation⁽⁵⁾. The influence of long range correlations and linear trends has been examined recently⁽⁶⁾. Also, Bassett⁽⁵⁾ briefly presents some mechanisms that are responsible for the presence of serial correlation in global annual temperature. Since the presence of autocorrelation in climatic and hydro-meteorological variables is quite common, there is need for thorough comprehension of the effect of short term serial persistence on record-breaking statistics.

There is also room for a comparative approach of stochastic models, in order to examine possible dissimilarities.

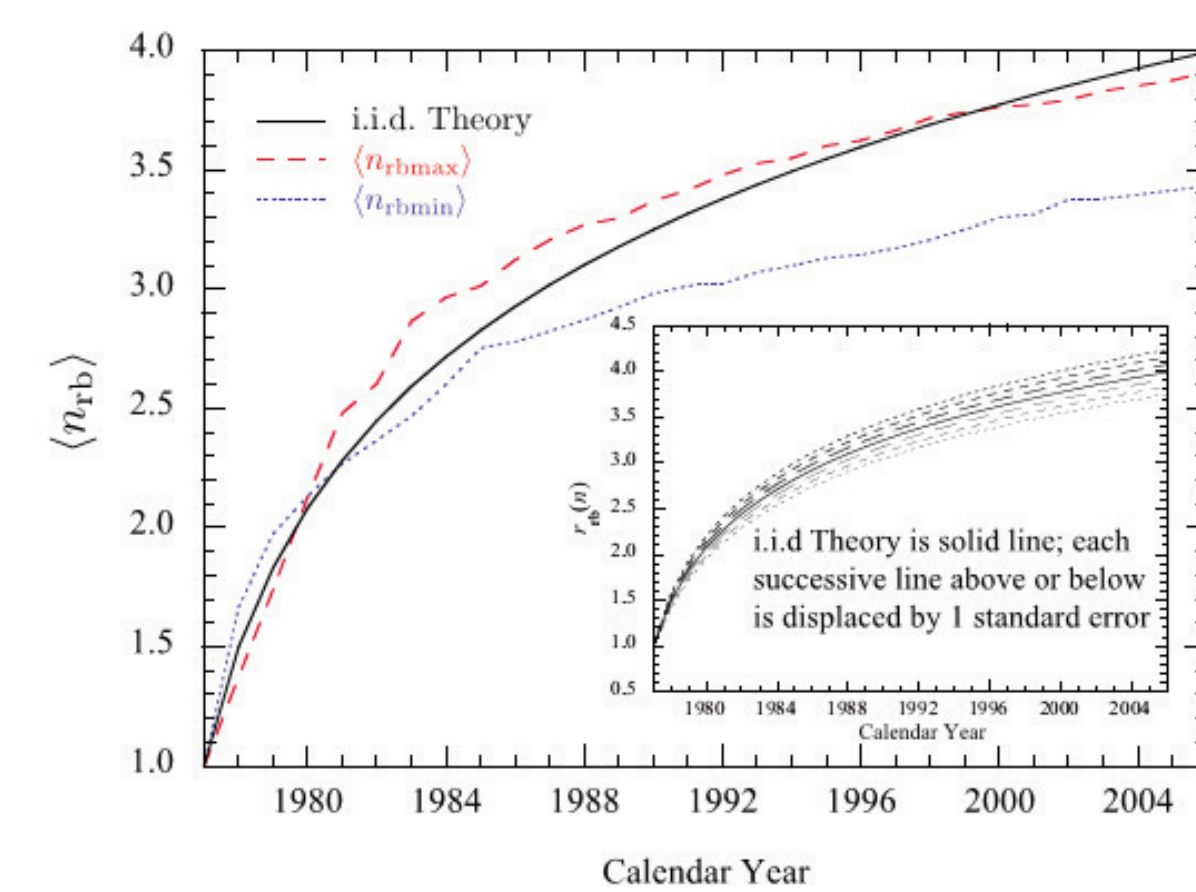


FIG. 9. (Color online) The average number of record-breaking maximum and minimum temperatures, (n_{max}) (dashed line) and (n_{min}) (dotted line), as a function of time measured forward from January 1, 1977. The i.i.d. theory is also shown (solid line). The average is over the 365 days of the year. Also included as an inset is the number expected for an i.i.d. random process from Eq. (5).

Source: Newman et al. (2011).

4. Methodology

We adopt a Monte Carlo approach. We generate 10000 time series with a sample size of 200 years each, using 2 major cases of different autocorrelations, and we compare the results with the pure random numbers with normal distribution (White Noise). The models we used are:

- Autoregressive of order 1 (AR(1)), with autocorrelation for lag j equal to $\rho_j = \rho_1^j$;
- Autoregressive of order 2 (AR(2)), with $\rho_j = a_1 \rho_{j-1} + a_2 \rho_{j-2}$, where a_1 and a_2 are parameters;
- First-order autoregressive - first-order moving average (ARMA(1, 1)), with $\rho_j = \rho_1 (\rho_2 / \rho_1)^{j-1}$;
- Hurst-Kolmogorov (aka FGN), with $\rho_j = (1/2) [|j+1|^{2H} + |j-1|^{2H}] - |j|^{2H}$ where H is the Hurst coefficient

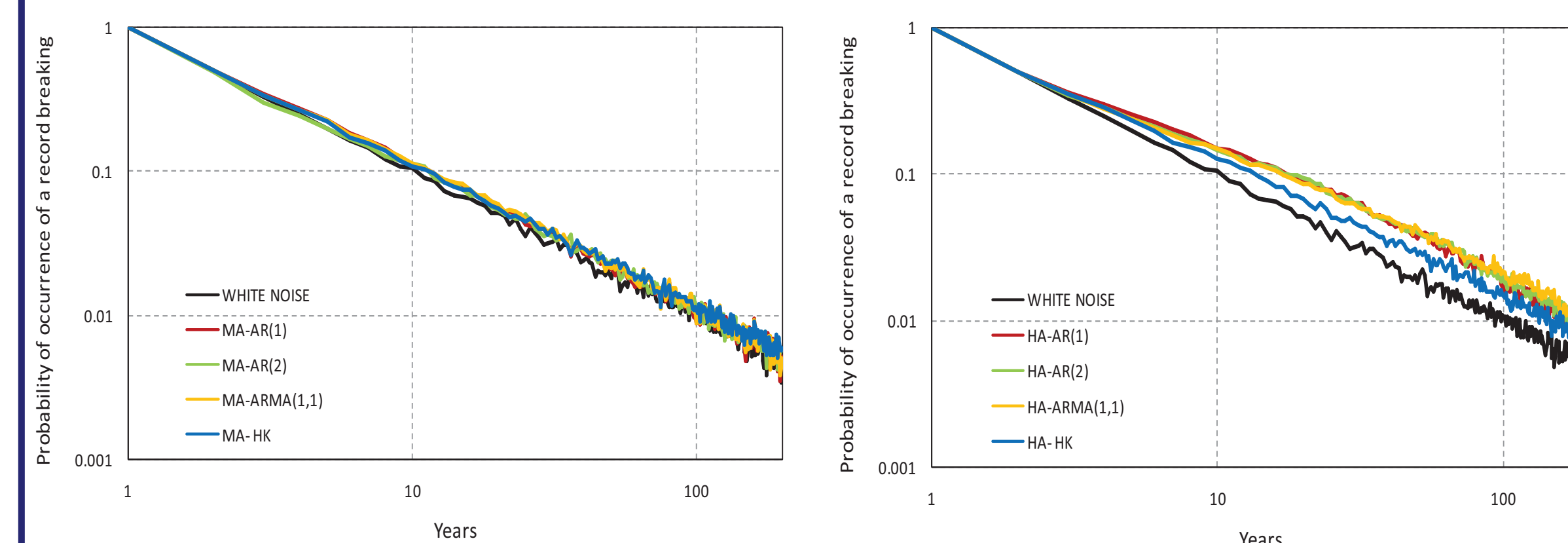
The two major cases of autocorrelations are:

- High Autocorrelation (HA), with $\rho_1=0.87$, $\rho_2=0.80$, $H=0.95$.
- Medium Autocorrelation (MA), with $\rho_1=0.41$, $\rho_2=0.27$, $H=0.75$.

The results of the Monte Carlo analyses are presented in graphical form in the following graphs. Among other things we compare:

- The probability of occurrence of a record breaking,
- The average time distance from the previous record,
- The average number of records per decade,
- The 99th percentile of records per decade.

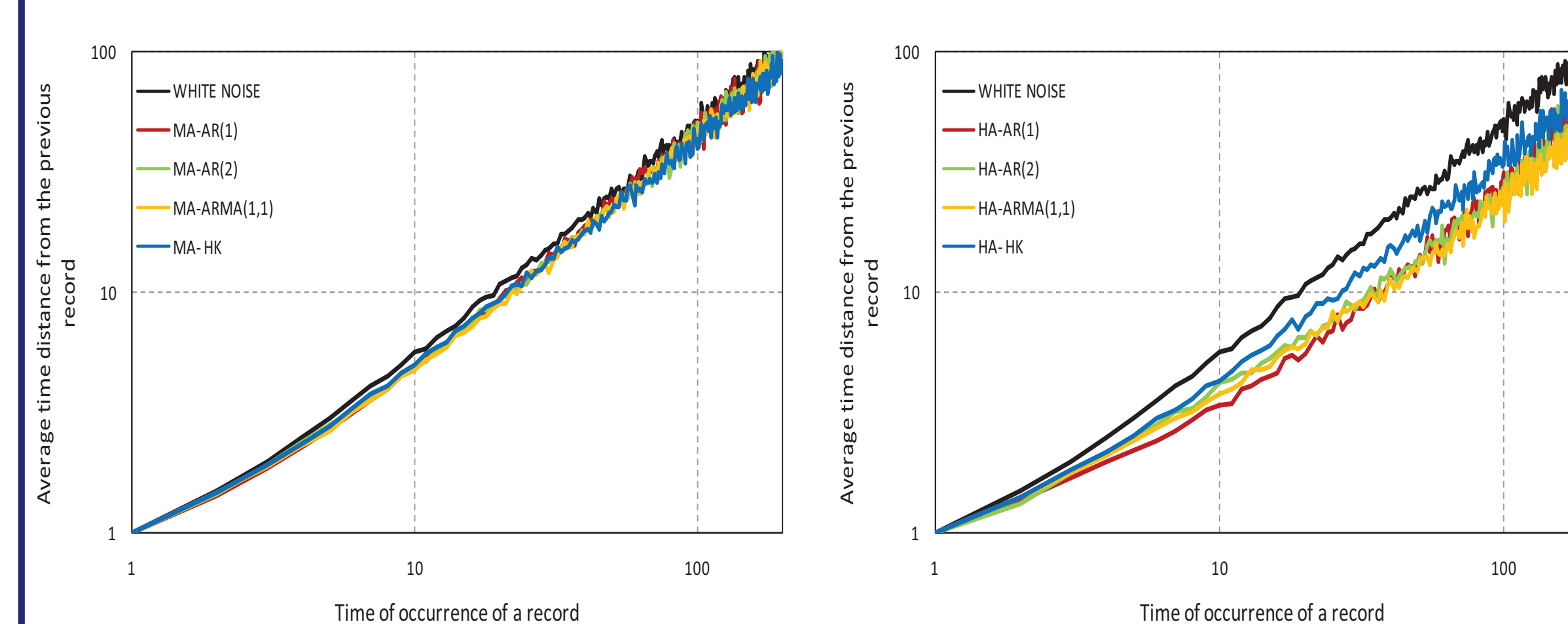
5. Probability of occurrence of a record breaking



The probability of occurrence of a record-breaking decreases almost with the same rate for all MA structures and the White noise case (the different models give almost the same probability of record-breaking).

For the HA time series, the stochastic models with an autocorrelation structure give a higher probability of record breaking than the White Noise.

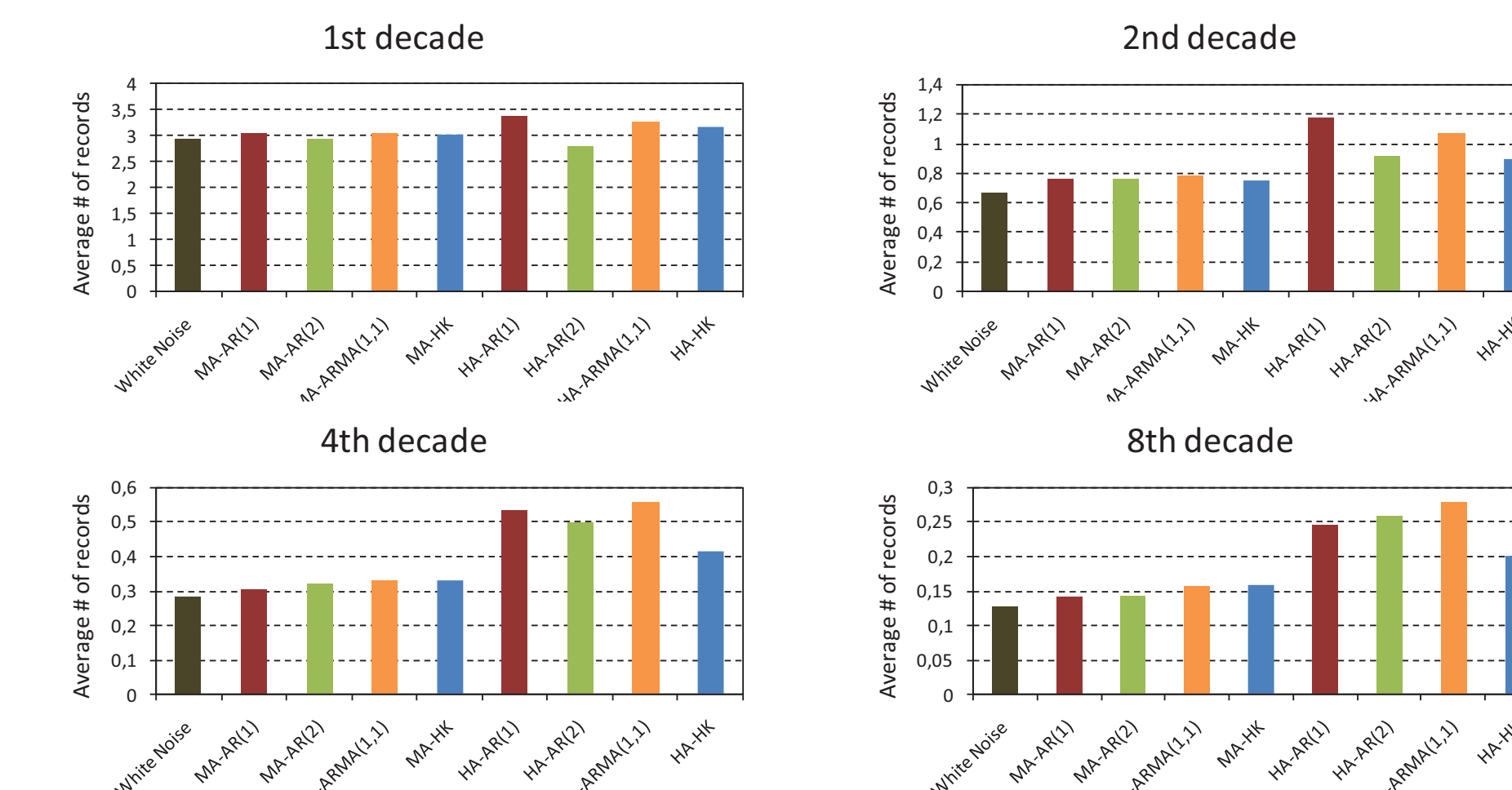
6. Average time distance from the previous record



For the MA time series, all stochastic models give only slightly smaller time distances between 2 record breakings than in the White Noise case.

For the HA time series, the models with autocorrelation give noticeably smaller time distances between 2 record breakings than in the White Noise case. All models with short-term persistence (AR(1), AR(2) and ARMA(1,1)) give virtually the same results, while the HK model results are in between the short-term persistence models and the White Noise.

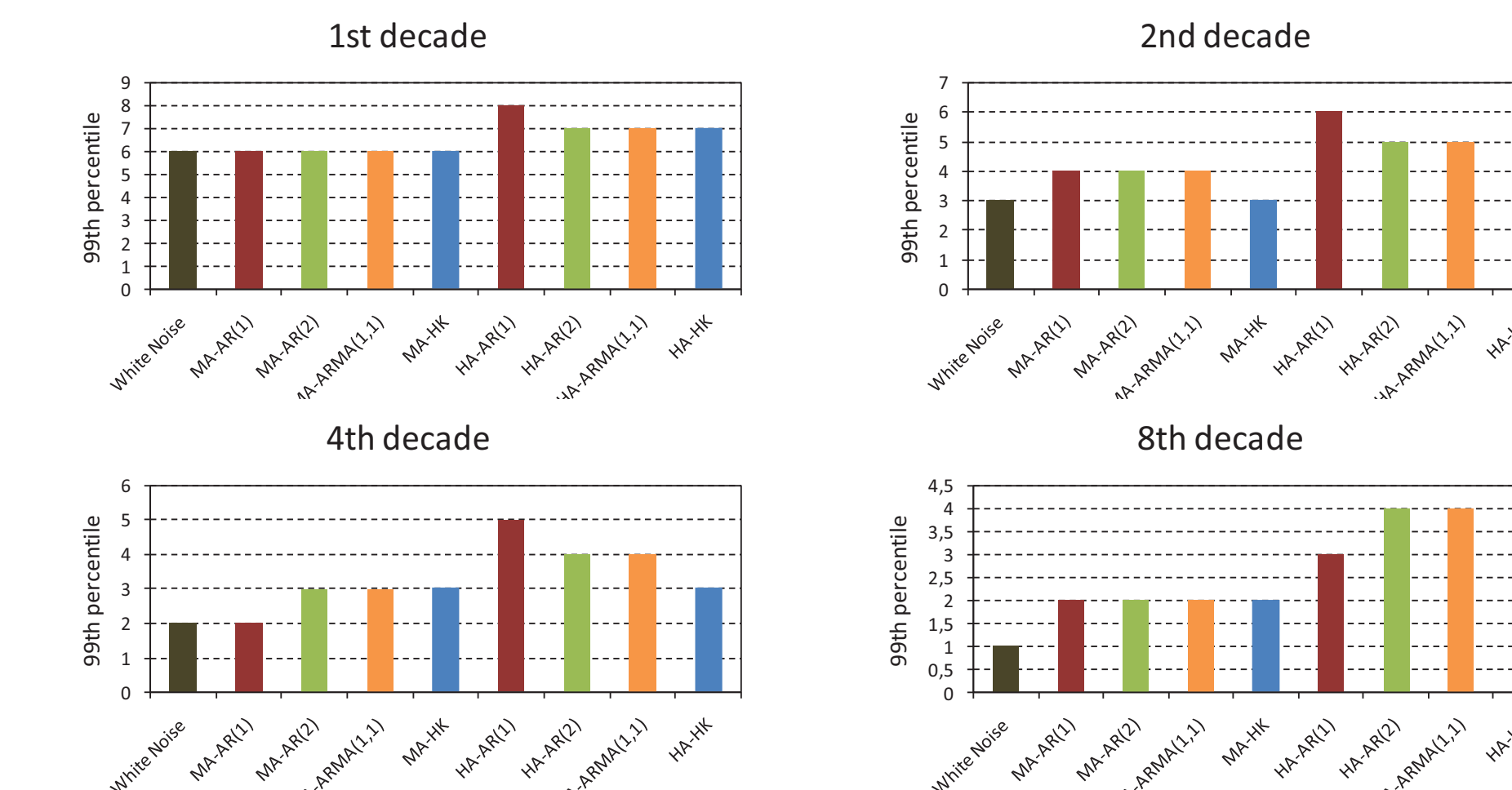
7. Average number of records per decade



For the first decade, the average number of records has no difference in all models, including White Noise.

However, several decades later, the High Autocorrelation structures exhibit large differences from White Noise and from Medium Autocorrelation structures.

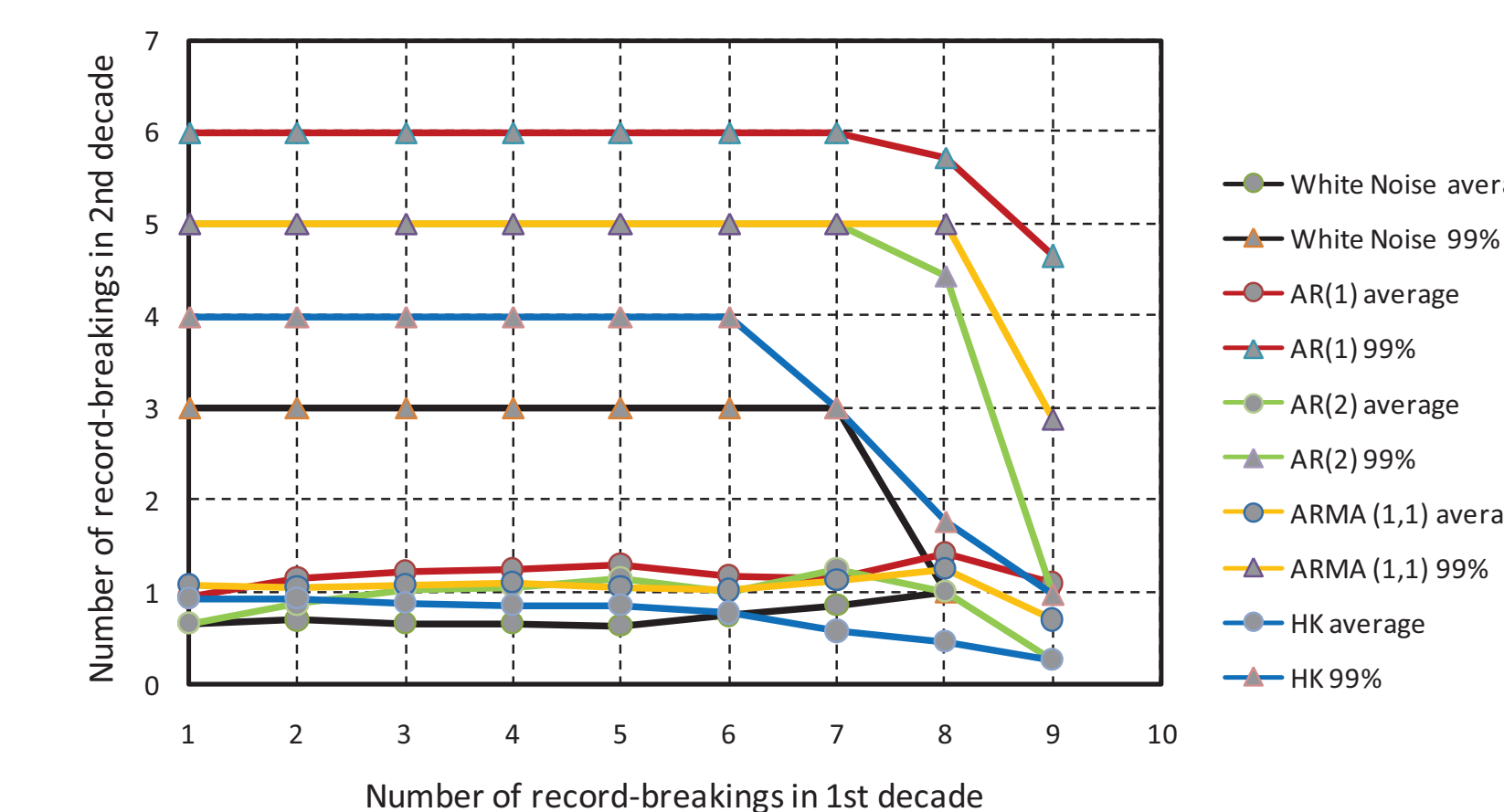
8. 99th percentile of number of records per decade



For the first decade, there is no difference in the 99th percentile of the number of records between the White Noise and the models with autocorrelation.

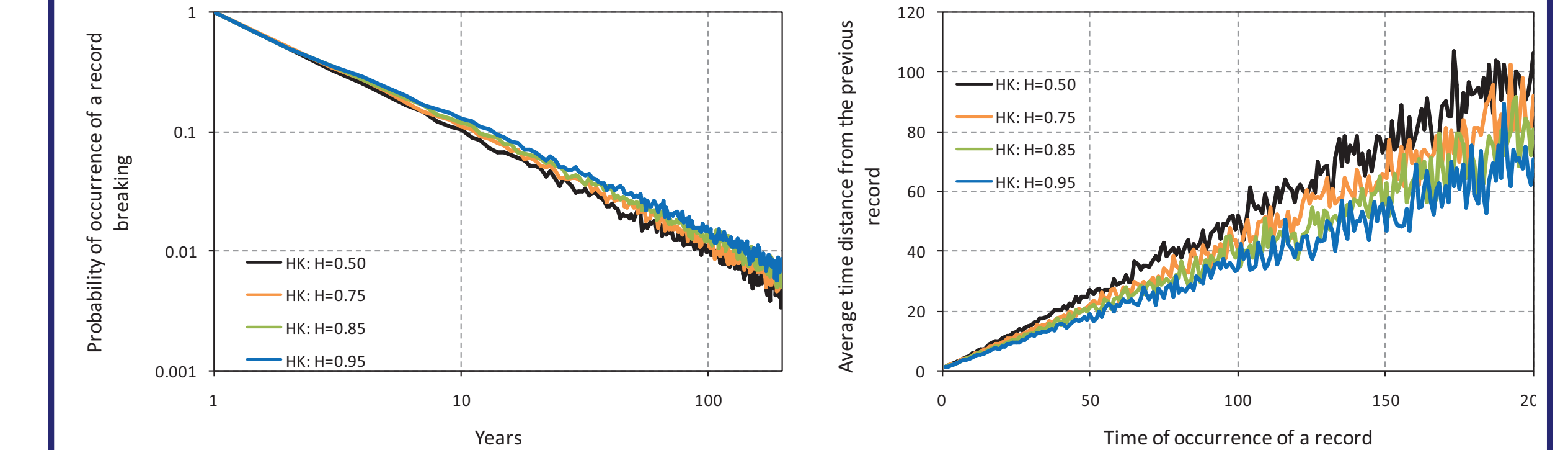
Several decades later, the High Autocorrelation structures exhibit spectacular differences from White Noise and from Medium Autocorrelation structures.

9. Investigation of possible correlations between number of records in consecutive decades



Detailed analyses on all decades showed that the number of record breakings in a specific decade is not related to the number of record breakings in the previous decade(s); that is, all the models give almost the same number of record breakings in a decade, irrespective of the number of records in a previous decade. The graph above illustrates this result for the second decade with reference to the first decade.

10. Effect of the Hurst coefficient on the record-breaking statistics for HK model



As the Hurst coefficient of a time series increases, the probability of a record breaking occurrence is greater when using the Hurst-Kolmogorov model.

Likewise, as the Hurst coefficient of a time series increases, the average time distance from the previous record breaking decreases.

11. Conclusions

- Generally, positive autocorrelation in time series results in greater probability of occurrence of record breakings and smaller time periods between consecutive records.
- However, for small and medium autocorrelations, the differences from White Noise (independent series) are very small.
- On the other hand, models with high autocorrelation give record breakings noticeably more often than the models of medium or no autocorrelation.
- All short-term persistence models behave in a very similar manner and their results are virtually indistinguishable.
- The HK model behaves in a different manner than its short-term persistence counterparts. Strangely, it gives smaller probability of record breaking than the short-term persistence models.
- While for the first decade all models behave similarly, irrespective of the type and strength of autocorrelation, in later decades the average number of record breakings (and even more so the 99th percentile) is greater for highly-autocorrelated models than for medium-autocorrelated ones and White Noise.

12. References

1. Tata, M. N. On outstanding values in a sequence of random variables, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 12 (1), 9-20, 1969.
2. Glick, N., Breaking records and breaking boards, *American Mathematical Monthly*, 85 (2), 2-26, 1978.
3. Foster, F. G., and A. Stuart, Distribution-free tests in time-series based on the breaking of records, *Journal of the Royal Statistical Society Series B*, 26 (1) 1-22, 1954.
4. Benestad, R. E., How often can we expect a record event?, *Climate Research*, 25, 3-13, 2003.
5. Bassett Jr, G. W., Breaking recent global temperature records, *Climatic Change*, 21 (3), 303-315, 1992.
6. Newman, W. I., B. D. Malamud and D. L. Turcotte, Statistical properties of record breaking temperatures, *Physical Review E*, 82, 066111, 2010.
7. Koutsoyiannis, D., *Lecture Notes on Stochastic Methods in Water Resources*, Department of Water Resources and Environmental Engineering - National Technical University of Athens, Athens, 2011.