Session HS7.4:
Hydroclimatic change and unchange: exploring the mysteries of variability, nature and human impact

COMPARING ESTIMATORS FOR INFERRING DEPENDENCE FROM RECORDS OF HYDROLOGICAL EXTREMES

THEANO ILIOPOULOU AND DEMETRIS KOUTSOYIANNIS

Department of Water Resources and Environment, School of Civil Engineering, National Technical University of Athens
www.itia.ntua.gr

1 INTRODUCTION

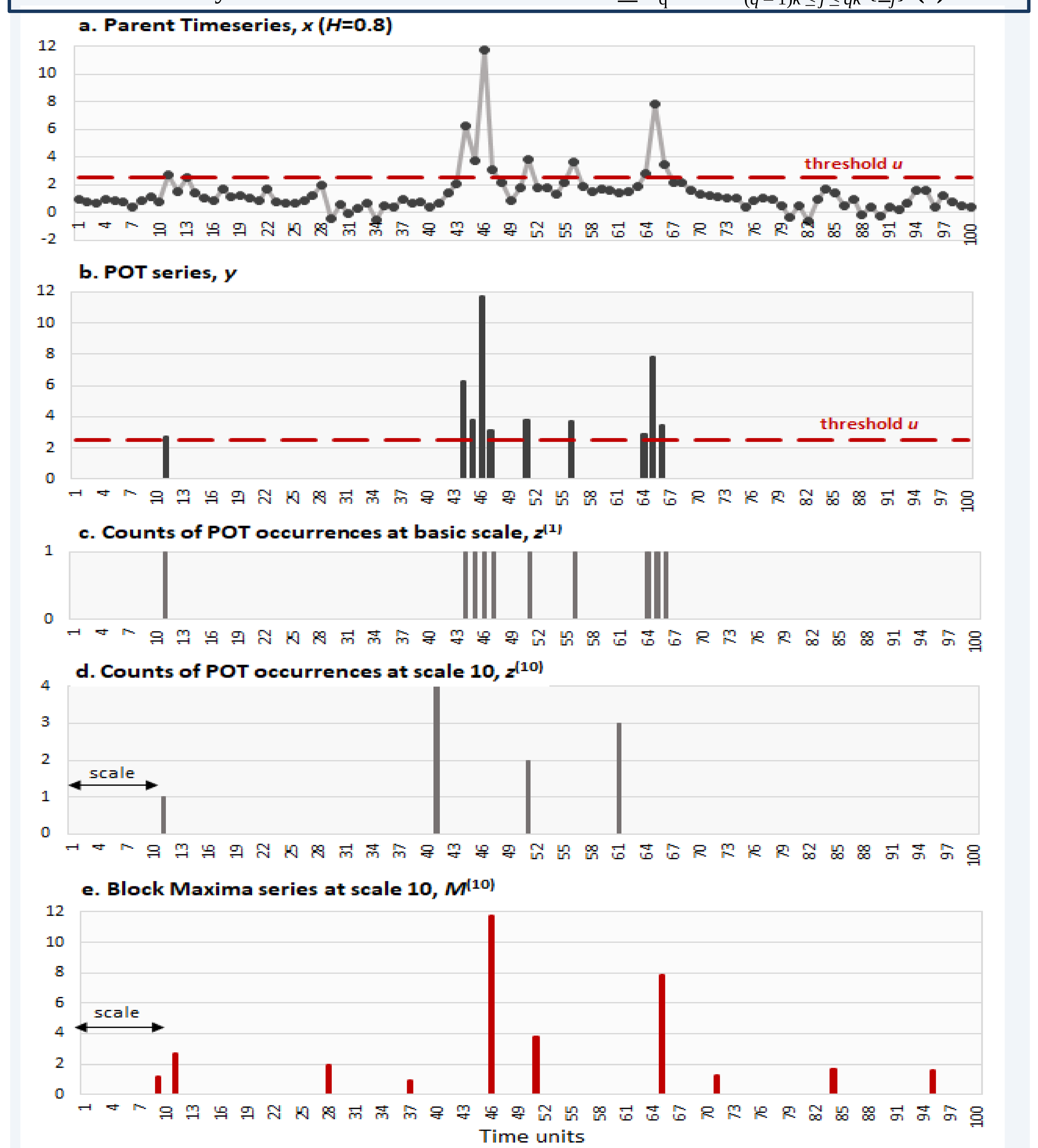
Hydrological extremes are regularly assumed independent in most practical and theoretical applications. The latter is indeed a convenient assumption as temporal independence is usually a prerequisite for the application of the widely used classical statistics.

Motivated by the existence of dependence mechanisms in hydrological processes, i.e. Hurst-Kolmogorov dynamics or long-term persistence, we investigate the propagation of persistence from the parent processes into the series of extremes by focusing especially on the opportunity of inferring the former (persistence) from the latter (records of extremes).

To this aim, we examine relevant stochastic indices such as the Hurst parameter and the Dispersion Index, and discuss their strengths and limitations. Additionally**, we explore a new probabilistic characterization of clustering for extremes which is found to provide new insights into the identification and modeling of extremal dependence.**

2 CONCEPTUALIZATION OF CLUSTERING

- Let $\underline{x}_i$ be a stationary stochastic process in discrete time i and x a single realization of the latter, i.e. a timeseries **(a)**. For u being a threshold, $u \in \mathbb{R}$, we define the process of **Peaks Over the Threshold** (POT) of events surpassing the threshold u, i.e, $\underline{y}_i$ **(b)**.
- Let also $\underline{N}(t)$ be a counting process of POT occurrences in time. We define the process $\underline{z}^{(k)}_q := \underline{N}(qk) - \underline{N}((q-1)k)$ as the number of occurrences of POT at timescales k and at discrete time $q = 1,..,n/k$ **(c,d)**.
- We additionally define the **block maxima** series as $\underline{m}^{(k)}_q := \max_{(q-1)k \le j \le qk} \{\underline{x}_j\}$ **(e).**

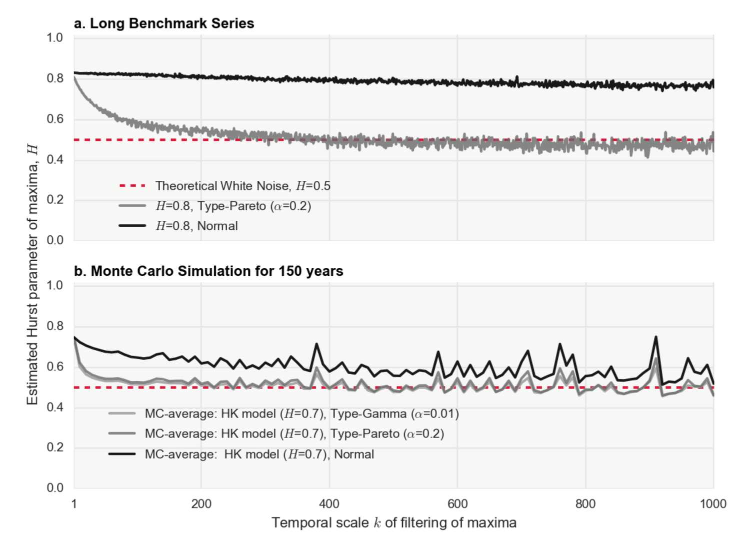


For the time series generation scheme see Dimitriadis and Koutsoyiannis (2018).

3 EXISTING INDICES FOR CLUSTERING IDENTIFICATION

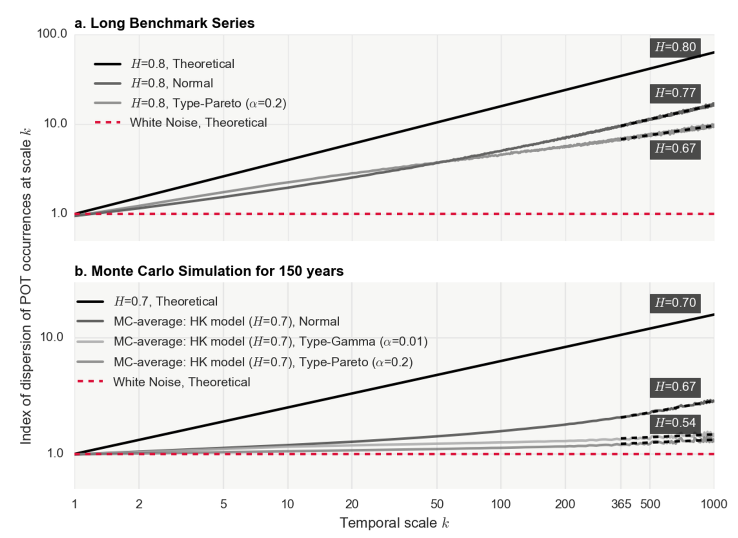
The **Hurst** parameter, a well-established measure of persistence, can be estimated from the slope of the double logarithmic plot of the standard deviation of the averaged process versus the averaging timescale, i.e. the climacogram (Koutsoyiannis, 2010).

- We compute the H parameter for extremes extracted from windows (scales) of length 1 to N/10 where N is the timeseries length. The first H value (k = 1) is the value for the original data and as the scale increases progressively the time series is filtered to show only the most 'extreme' data, e.g. if the basic timescale is daily, the estimated H at timescale k=365 corresponds to the H parameter of the annual maxima.



The **Dispersion Index,** a well-known measure of clustering of events, is defined as the ratio of the variance of the counts of events versus their mean number at a specific timescale k, i.e $\underline{d}^{(k)} = \frac{E[\underline{z}_k^2] - E[\underline{z}_k]^2}{E[\underline{z}_k]}$.

- The dispersion index exhibits power-law scaling behavior which is linked to the underlying persistence structure (Thurner et al., 1997) as: $\underline{d}^{(k)} \approx c k^{2H-1}, k \ge k_0$, where c a real parameter and k_0 denotes the scaling onset timescale. It follows that the exponent $2H - 1$ can be obtained as the slope of the double logarithmic plot of the dispersion index versus the timescale for $k \ge k_0$.
- **Both indices mask persistence for non-Gaussian timeseries!**



4 A NEW PROBABILISTIC CHARACTERIZATION

We form the Peaks Over Threshold series, y_i and the series of counts of the POT events for each scale, $z^{(k)}$. We additionally, define the binary process $\underline{r}_q^{(k)}$ to denote the event of exceedance of the threshold at each time interval q of size k, $q = 1, \dots, \lfloor n/k \rfloor$:

$$\underline{r}_q^{(k)} := \begin{cases} 1, & \underline{z}_q^{(k)} > 0 \\ 0, & \underline{z}_q^{(k)} = 0 \end{cases}$$

Then, **the probability of exceedance of the threshold for timescale k** is obtained as the frequency of exceedances estimated from all $\lfloor n/k \rfloor$ intervals:

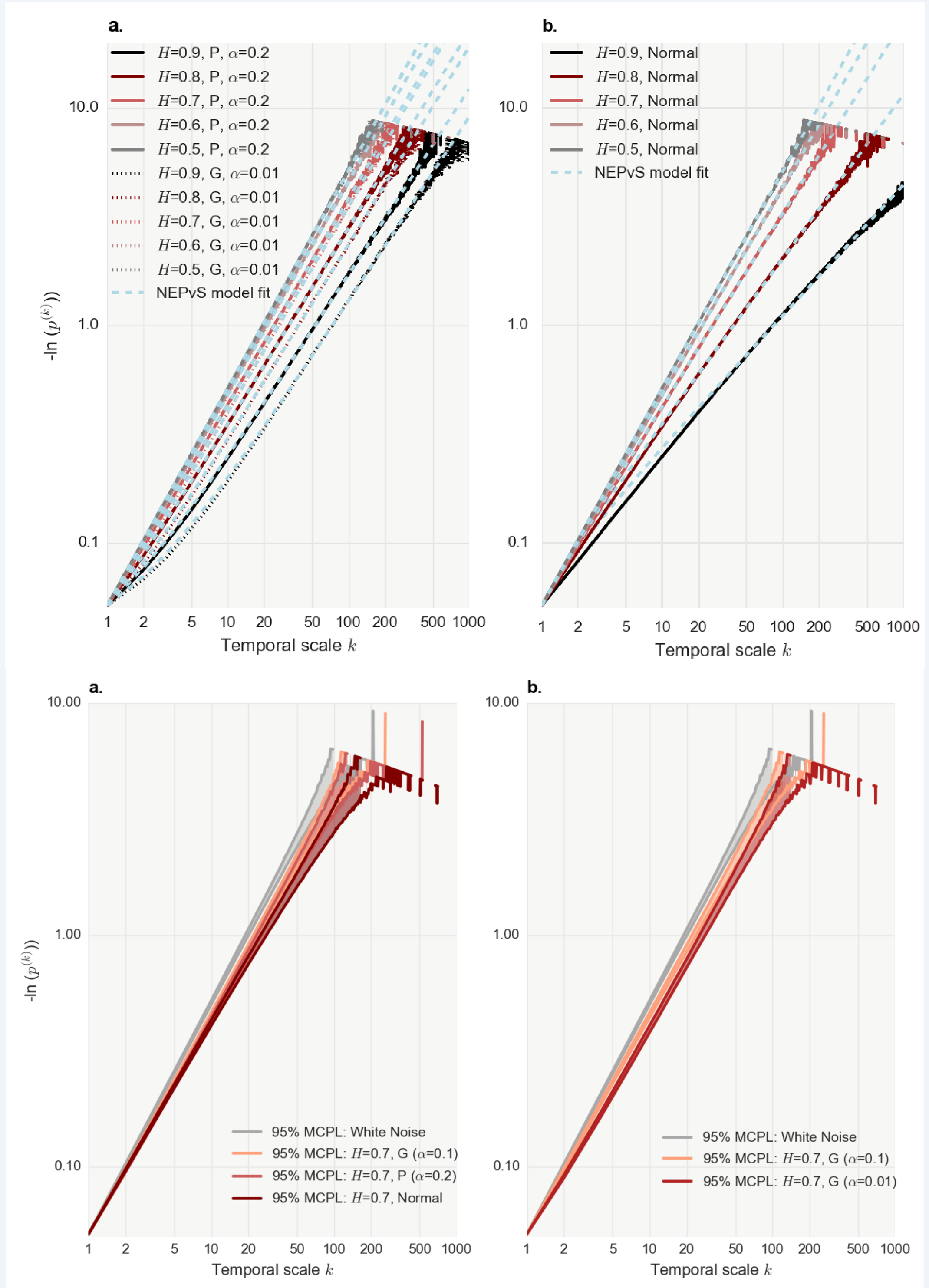
$$\bar{p}^{(k)} = \frac{\sum_1^{\lfloor n/k \rfloor} \underline{r}_q^{(k)}}{\lfloor n/k \rfloor}$$

The latter is **the Exceedance Probability of the threshold versus the Scale (EPvS)** and its complement $p^{(k)} = 1 - \bar{p}^{(k)}$ is the Non-exceedance Probability versus Scale (NEPvS).

- **For a purely random process, the NEPvS is $p^{(k)} = p^k$**

where p is the probability of non-exceedance at the basic scale k = 1

- **For HK (persistent) processes** though, a different behaviour is revealed, with **the probabilities of non-exceedance of the threshold being larger than those obtained under independence**.



- To model the NEPvS, we revisit a probabilistic model proposed by Koutsoyiannis (2006) to describe the clustering behaviour of dry spells in rainfall timeseries.

$$p^{(k)} = p^{[1+(\xi^{-1/\eta}-1)(k-1)]^{\eta}}, \; p = 1 - F(u)$$

For η = 1 and ξ = 0.5, this equation describes the white noise process.

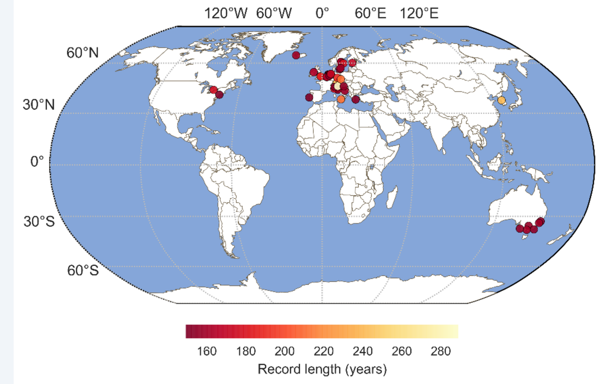
- For threshold 5%, sample size 150 years and H=0.7, we estimate the NEPvS index for benchmark series characterized by different marginal distribution tails, i.e. the Standard Normal, the Gamma (G) and the Pareto (P) distribution, in order to explore the impact of skewness and kurtosis on the NEPvS index.

We find that:

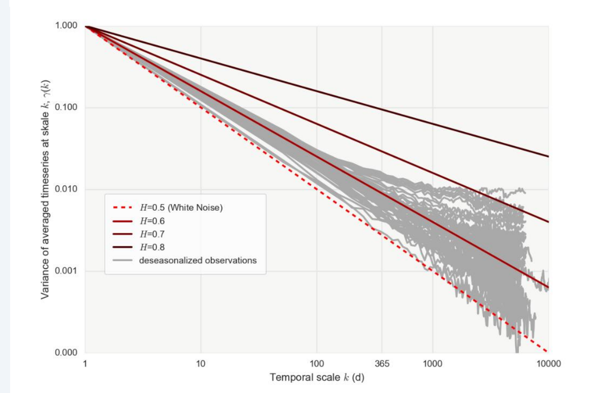
- clustering of extremes and its identifiability is greater for the normal distribution
- for a specified non-Gaussian distribution, clustering is greater and also more visible for increasing skewness and kurtosis.

5 LONGEST RAINFALL RECORDS

We analyze the longest available daily rainfall records, i.e. 60 rainfall records surpassing 150 years of daily values (collected from the Global Historical Climate Network Daily database, European Climate Assessment and Dataset and third parties)



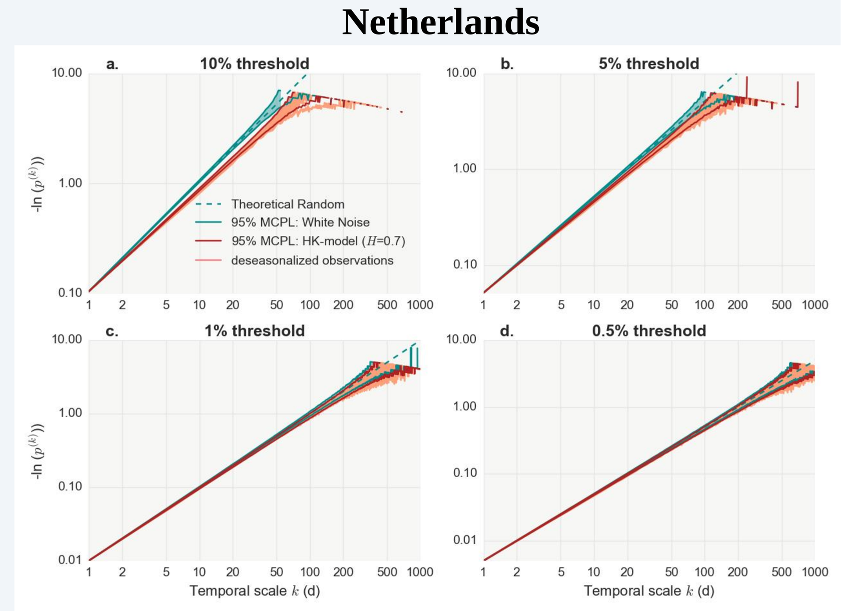
- To test the hypothesis of persistence in the parent process, we estimate the H parameter of the deseasonalized daily rainfall series. The estimated average persistence is even larger than the global estimate (H≈0.6) of Iliopoulou et al. (2016) for annual rainfall.
- To theoretically validate the empirical results we perform the following case-specific MC simulations:



28 stations in the Netherlands

- ✓ It is evident that the assumed model is consistent with the majority of the observed records, with only a few stations located at the South-West of Netherlands exhibiting even stronger clustering.
- As the threshold increases evidence of persistence is progressively 'lost' and the probabilistic behavior of POT approaches randomness.

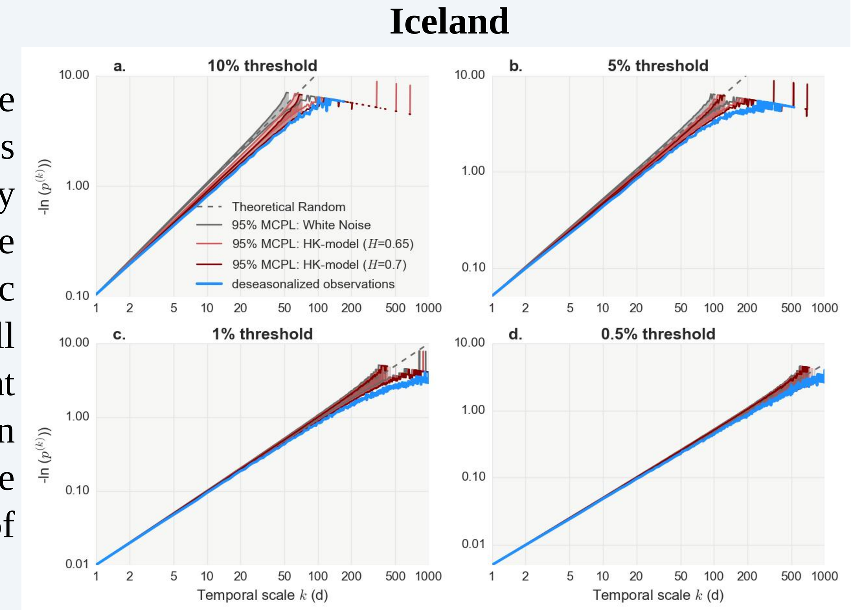
Netherlands



Stykkishólmur, Iceland

- The 'discrepancy' between the persistence of the parent process and the stronger one implied by the extremes might be due to the impact of large-scale atmospheric circulation patterns on rainfall extremes (NAO), which might need even longer record lengths in order to be summarized by the second-order characterization of the H parameter.

Iceland



6 CONCLUSIONS

- Both the second-order properties (H parameter) and the high-order moments of the parent process impact the temporal properties of the generated extremes, and therefore characterizations of clustering of extremes need to account for both.
- Identifiability of persistence from records of maxima is in general limited and weakens as the threshold for extremes increases.
- The estimates from the Hurst parameter and the dispersion index are severely downward biased for extremes originating from non-Gaussian processes.
- ✓ A new probabilistic index is proposed to represent clustering based on the probability of non-exceedance of a given threshold across scales, the NEPvS (Non-Exceedance Probability vs Scale) index. The index can be directly used for statistical testing of departures from independence. Case-specific Monte Carlo simulations are needed to validate models coupling persistence with different marginal properties.

REFERENCES

Iliopoulou, T., Papalexiou, S.M., Markonis, Y. and Koutsoyiannis, D., 2018. Revisiting long-range dependence in annual precipitation. Journal of Hydrology, 556, pp.891-900., https://doi.org/10.1016/j.jhydrol.2016.04.015

Dimitriadis, P. and Koutsoyiannis, D., 2018. Stochastic synthesis approximating any process dependence and distribution. Stochastic environmental research and risk assessment, 32(6), pp.1493-1515. https://doi.org/10.1007/s00477-018-1540-2

Koutsoyiannis, D., 2006. An entropic-stochastic representation of rainfall intermittency: The origin of clustering and persistence. Water Resources Research, 42(1)., https://doi.org/10.1029/2005WR004175

Koutsoyiannis, D., 2010. HESS Opinions" A random walk on water". Hydrology and Earth System Sciences, 14(3), pp.585-601., https://doi.org/10.5194/hess-14-585-2010

Thurner, S., Lowen, S.B., Feurstein, M.C., Heneghan, C., Feichtinger, H.G. and Teich, M.C., 1997. Analysis, synthesis, and estimation of fractal-rate stochastic point processes. Fractals, 5(04), pp.565-595.