European Geosciences Union | General Assembly 2015
Vienna | Austria | 12 - 17 April 2015
Session: HS7.1/AS1.11/NH1.7/NP9.4 | Precipitation: from measurement to modelling and application in catchment hydrology

Contact info: smp@itia.ntua.gr

Explorations on the Hershfield Factor

Simon Michael Papalexiou[1] | Yannis Dialynas[2] | Salvatore Grimaldi[3]
[1]Department of Water Resources and Environmental Engineering | National Technical University of Athens | Greece
[2]School of Civil and Environmental Engineering | Georgia Institute of Technology | Atlanta | GA | USA
[3]Università degli Studi della Tuscia | Viterbo | Italy

1 What is all about…

The Hershfield factor (*H*) essentially constitutes a multiplier aiming to correct the error between fixed time interval maxima (*F*-maxima) and sliding maxima (*S*-maxima) as a direct consequence of temporal discretization of hydrometeorological time series. Rainfall is typically recorded as an accumulated value in fixed non-overlapping time intervals, e.g., in daily intervals, and thus the annual maximum value expresses the maximum value of these fixed recordings over a year period. Yet if measurements at a finer time scale are available, e.g., hourly, then the annual daily *S*-maximum, i.e., the annual maximum value resulting by sliding a 24-hour time interval starting at any hour of the year, in general, is different than the *F*-maximum value. The *H* factor attempts to correct for this error. Multiplying the *F*-maximum, which can be considered as a random variable, with the *H* factor, theoretically should result in the *S*-maximum random variable. This implies that the location and scale characteristics of the *S*-maximum distribution are explicitly related to the value of *H* and to the characteristics of the *F*-maximum random variable, while its shape characteristics will be exactly the same as those of the *F*-maximum distribution. This study further explores the validity of this well-accepted assumption. In order to verify or discard this assumption we perform an unprecedentedly large empirical analysis based on thousands of hourly rainfall records across the USA.

2 A quick review

- **Hershfield and Wilson (1958):** They report a multiplier of 1.13 for converting "...observation-day rainfall for a particular frequency to the maximum 1440-minute rainfall for the same frequency". They find the same multiplier for converting clock- hour to 60-minute extremes, which they consider to be coincidental (they neglect to give clear details on the data and methods used to estimate Hershfield's factor).
- **Hershfield (1961a) -** U.S. Weather Bureau Technical Paper 40 (TP40)**:** also a popular reference for the 1.13 (H) factor. They concluded that: "It was found that 1.13 times a rainfall value for a particular return period based on a series of annual maximum clock-hour rainfalls was equivalent to the amount for the same return period obtained from a series of 60-minute rainfalls. By coincidence, it was found that the same factor can be used to transform observational-day amounts to corresponding 1440-minute return-period amounts." Thus, they have exactly the same empirical conclusions with the 1958 paper. The approach is empirical, and they appropriately call the 1.13 factor: an "average index relationship".
- However, the data used are obsolete (Faiers et al., 1994, Huff and Angel (1992)). Huff and Angel (1992) suggested that the spatial distribution of the TP40 data is poor, and showed that the TP40 data do not appropriately depict extreme values for several states. Furthermore, over half of the data used in TP40 were derived from station records of less than 15 yr. (Faiers er al., 1994).
- **Hershfield (1961b):** also graphically displays (slope) the 1.13 value.
- **Weiss (1964):** model study: he used a simple probabilistic model, based on uniform rainfall intensity for sliding maxima. He found $H=n/(n-1/8)$ for maxima with sliding starting point, with constant intensity and with a random length between (n - 1) and n time units (a correction to the approach was suggested by Dwyer and Reed (1995a)).
- **Kerr et al. (1970):** Empirical approach. Data from 45 gages (state of Pennsylvania, USA), with at least 17 years of length, each, for the 24 hour H factor, and 60 min. H factor. Gumbel distributions are fitted at each. The data quality is poorly described.
- **Harihara and Tripathi (1973):** similar approach with Kerr et al. (1970) for daily H factor. Data from 67 gages in India. Only 5 gages have length greater than 25 years. Significant bias from data.
- **Natural Environment Research Council (1975)**:Empirical approach. 50 stations from around the UK - data quality not clearly specified. 24h H value: 1.11, 60min H value: 1.15.
- **van Montfort (1990):** Only **1 station (New Zealand) of 58 years** of data was used - the analysis is sample-dependent. He examined the problem of estimating extreme value distribution parameters for sliding maxima, given that only fixed maxima are available. Denoting, for year i, the sliding 24-hour maximum by A_{1i} and the fixed 1-day and 2-day maxima by F_{1i} and F_{2i}, respectively, van Montfort utilizes the inequality $F_{1i} \leq A_{1i} \leq F_{2i}$ to estimate GEV parameters for the sliding maxima using only the fixed maxima. The maximum likelihood method is used. Mean annual rainfall correction: 1.15. 50-year rainfall: 1.14.
- **Huff and Angel (1992):** Empirical approach. "Analysis verified the earlier findings that 1.13 represented the average ratio of maximum 24-hour to calendar-day rainfall in heavy rainstorms." **Data:** using all recording raingage data for the period 1948-1987 in Indiana (**41 gages**) and Illinois (**61 gages**). The 1.13 factor was used accordingly, to develop frequency distributions for storm maxima for the USA midwest.
- **Faiers et al. (1994):** Empirical approach. Data: **14** stations in the state of Louisiana (USA), length of greater that 20 yrs each gage. They agree that 24h H=1.13.
- **Dwyer and Reed (1994)**: they attempted to examine the problem from different theoretical perspectives; they show that many climate variables (rainfall, wind speed and air temperature) exhibit simple scaling behavior. Their method is relatively simple. **Data:** less than ten records for each variable. They come up with H-Duration relationships (from regression) for each variable.
- **Dwyer and Reed (1995a):** The Dwyer and Reed (1994) approach is followed here. Quite few (~20 for rainfall, and less than 10 for other variables (i.e., wind speed and air temperature)) sites (UK,Australia, South Africa) for each variable. They come up with H-Duration relationships (based on regressions) for each variable (same as Dwyer and Reed (1994)).
- **Dwyer and Reed (1995b):** examined the relation of *H* to the fraction of wet hours.
- **van Montfort (1997)**: same type of theoretical distribution was assigned to maxima over fixed intervals and to maxima over sliding intervals. The factor *H* turns out to be associated with location, positively correlated with autocorrelation and positively correlated with the fraction of wet days. **Data**: A Data set (China) with 130 stations over 30 years.

3 The data

- We analyzed a very large database from the National Climatic Data Center (NCDC) that comprises hourly precipitation data from thousands of stations from all over the Unites States. The original precipitation data were given in hundreds of inches and were transformed to mm.
- The records are of variable length ranging from just a few years to more than 50 years.
- All data used at larger than the hourly time scale were constructed by aggregating the original hourly time series.
- Particularly, we studied 7 127 records of hourly precipitation and we estimated the sliding and fixed-interval maximum precipitation for every year and for every record. The result was an unprecedentedly large number of estimated values, i.e., more than 100 000 years for each time scale.

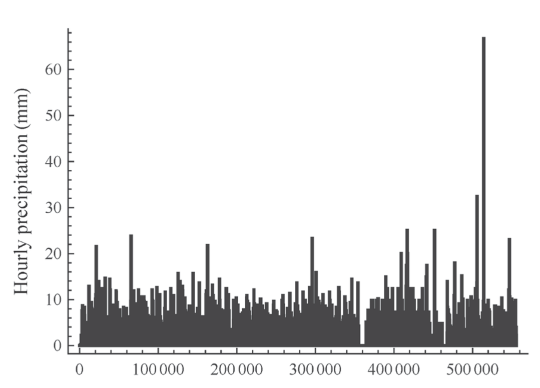

Fig. 1. A randomly selected time series of hourly precipitation values.

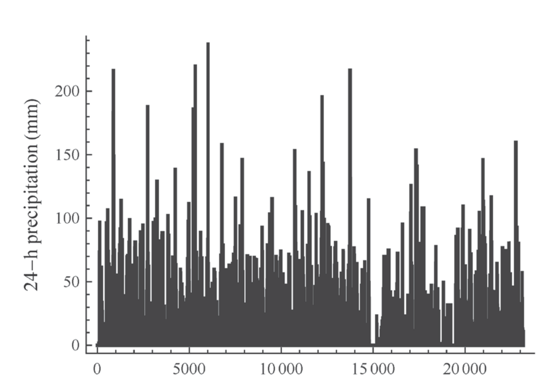

Fig. 2. The time series of Fig. 1 aggregated in the 24 h time scale.

4 Methodology

- For each year and for all records we estimated the missing values (MV's), percentage, i.e., as the ratio of the number of hourly values missing within the the year to the total number of values. We assumed as valid years only those having less than 1/12 of MV's, i.e., on total less than a month missing within the year.
- In these "valid" years the MV's were replaced with the record's median precipitation value.
- For valid years we estimated:
  i. The sliding maximum value at several time scales *k* by sliding a *k*-hour moving window over the year and extracting the maximum of the resulted values.
  ii. The fixed-interval maximum value by aggregating the hourly values over the year in *k*-hour non-overlapping intervals and extracting the maximum of the resulted values.
- We estimated the *H*-factors as the ratio of the *S*-maximum value to the *F*-maximum value.
- We performed this analysis for the following time scales given in hours: {2, 4, 6, 8, 12, 16, 20, 24, 30, 36, 42, 48, 56, 64, 72}.

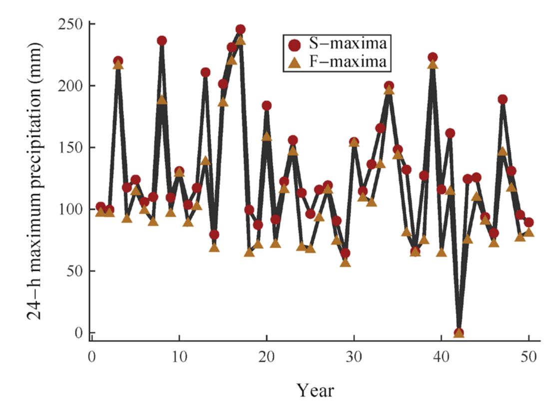

Fig. 3. Sliding and fixed interval maxima of a randomly selected time series.

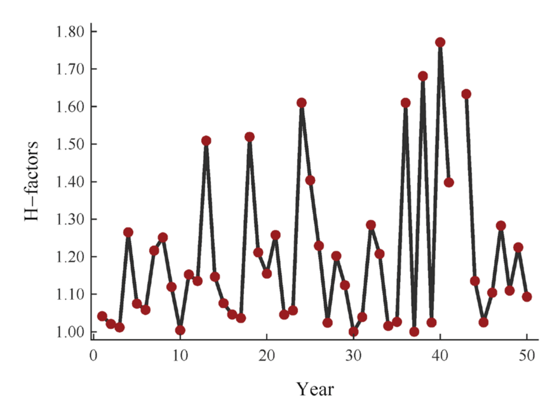

Fig. 4. Estimated 24 h *H*-factors from Fig. 3 as the ratio of *S*- to *F*-maxima.

5 Statistics of *S*-maxima

Table 1. Basic summary statistics of sliding maxima at various time scales

Scale	No of yrs	Median	Mean	SD	C_v	C_s	C_k	τ_2	τ_3	τ_4
2-h	121165	32.00	35.65	23.50	0.66	4.51	89.86	0.33	0.21	0.15
4-h	121165	39.37	44.32	28.36	0.64	5.51	200.08	0.31	0.22	0.17
6-h	121165	44.45	49.86	31.11	0.62	4.64	141.51	0.31	0.22	0.18
8-h	121165	48.26	54.05	33.31	0.62	4.17	110.73	0.31	0.22	0.18
12-h	121165	53.34	60.14	36.67	0.61	3.60	78.91	0.30	0.22	0.19
16-h	121165	57.66	64.61	39.35	0.61	3.27	61.93	0.31	0.23	0.19
20-h	121165	60.96	68.28	41.71	0.61	3.07	51.42	0.31	0.23	0.19
24-h	121165	63.50	71.46	43.72	0.61	2.90	44.06	0.31	0.23	0.19
30-h	121165	66.04	75.42	46.34	0.61	2.75	37.06	0.31	0.23	0.19
36-h	121165	68.58	78.40	48.51	0.62	2.68	32.79	0.31	0.24	0.19
42-h	121165	71.12	80.87	50.36	0.62	2.65	30.40	0.31	0.24	0.19
48-h	121165	72.90	83.20	52.28	0.63	3.45	82.11	0.31	0.24	0.19
56-h	121165	75.18	86.03	54.58	0.63	4.81	198.23	0.31	0.24	0.19
64-h	121165	76.71	88.32	56.19	0.64	4.66	179.80	0.31	0.24	0.19
72-h	121165	78.74	90.62	57.74	0.64	4.53	164.56	0.31	0.24	0.19

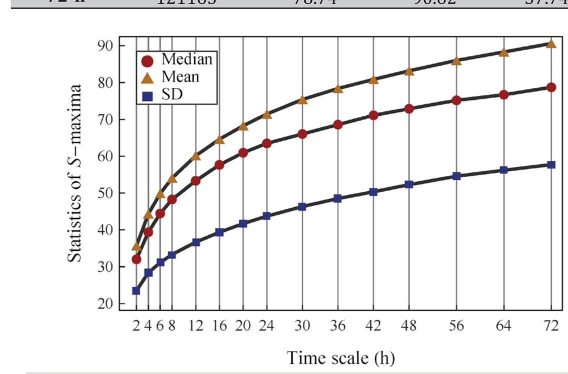

Fig. 5. Summary statistics of sliding maxima *vs.* time scale.

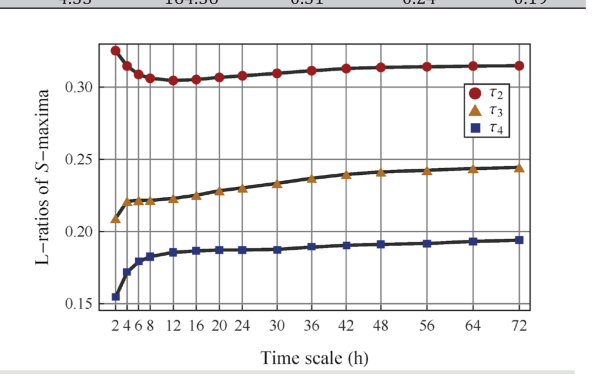

Fig. 6. L-ratios τ_2, τ_3 and τ_4 of sliding maxima *vs.* time scale.

6 Statistics of *F*-maxima

Table 2. Basic summary statistics of fixed-interval maxima at various time scales

Scale	No of yrs	Median	Mean	SD	C_v	C_s	C_k	τ_2	τ_3	τ_4
2-h	121165	29.46	32.93	21.91	0.67	4.73	91.89	0.32	0.21	0.16
4-h	121165	35.56	40.04	25.77	0.64	4.98	129.35	0.31	0.22	0.17
6-h	121165	40.13	44.81	28.27	0.63	5.27	191.29	0.31	0.22	0.18
8-h	121165	43.18	48.29	29.91	0.62	3.79	74.23	0.31	0.22	0.18
12-h	121165	48.26	53.65	33.11	0.62	4.18	111.73	0.31	0.22	0.19
16-h	121165	50.80	57.45	35.40	0.62	3.75	87.65	0.31	0.22	0.19
20-h	121165	53.34	60.51	37.37	0.62	3.54	74.17	0.31	0.23	0.19
24-h	121165	55.88	62.80	38.89	0.62	3.32	63.70	0.31	0.23	0.19
30-h	121165	58.42	66.40	41.10	0.62	3.06	52.14	0.31	0.23	0.19
36-h	121165	60.96	69.28	43.00	0.62	2.97	46.29	0.31	0.23	0.19
42-h	121165	63.25	71.56	44.91	0.63	2.95	41.83	0.31	0.24	0.19
48-h	121165	64.26	73.52	46.32	0.63	2.84	37.51	0.31	0.24	0.19
56-h	121165	66.04	75.98	47.93	0.63	2.77	34.07	0.32	0.24	0.19
64-h	121165	68.58	78.28	49.53	0.63	2.77	32.52	0.32	0.24	0.19
72-h	121165	69.60	80.04	50.63	0.63	2.71	30.56	0.32	0.24	0.19

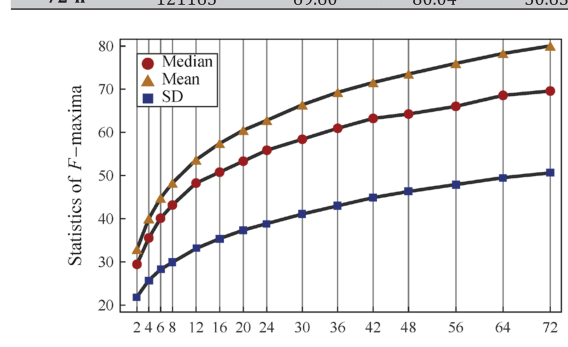

Fig. 7. Summary statistics of fixed-interval maxima *vs.* time scale.

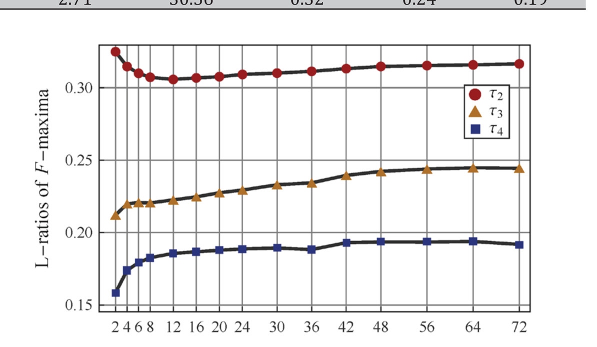

Fig. 8. L-ratios τ_2, τ_3 and τ_4 of fixed-interval maxima *vs.* time scale.

7 Comparing the *S*- and *F*-maxima (I)

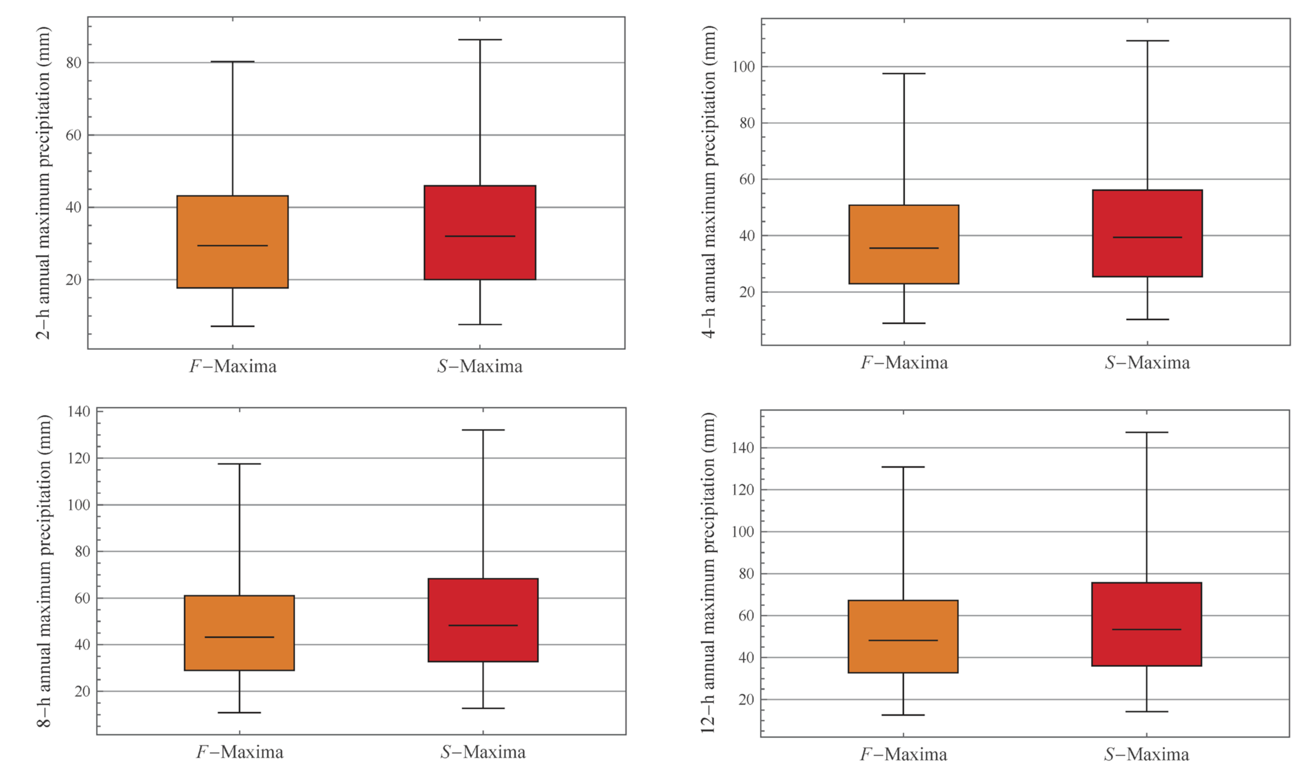

Fig. 9. Box plots of the estimated annual sliding and fixed-interval maxima for the following time scales: 2 h, 4 h, 8 h and 12 h. The outer fences of the box plots express the 95% empirical confidence interval.

8 Comparing the *S*- and *F*-maxima (II)

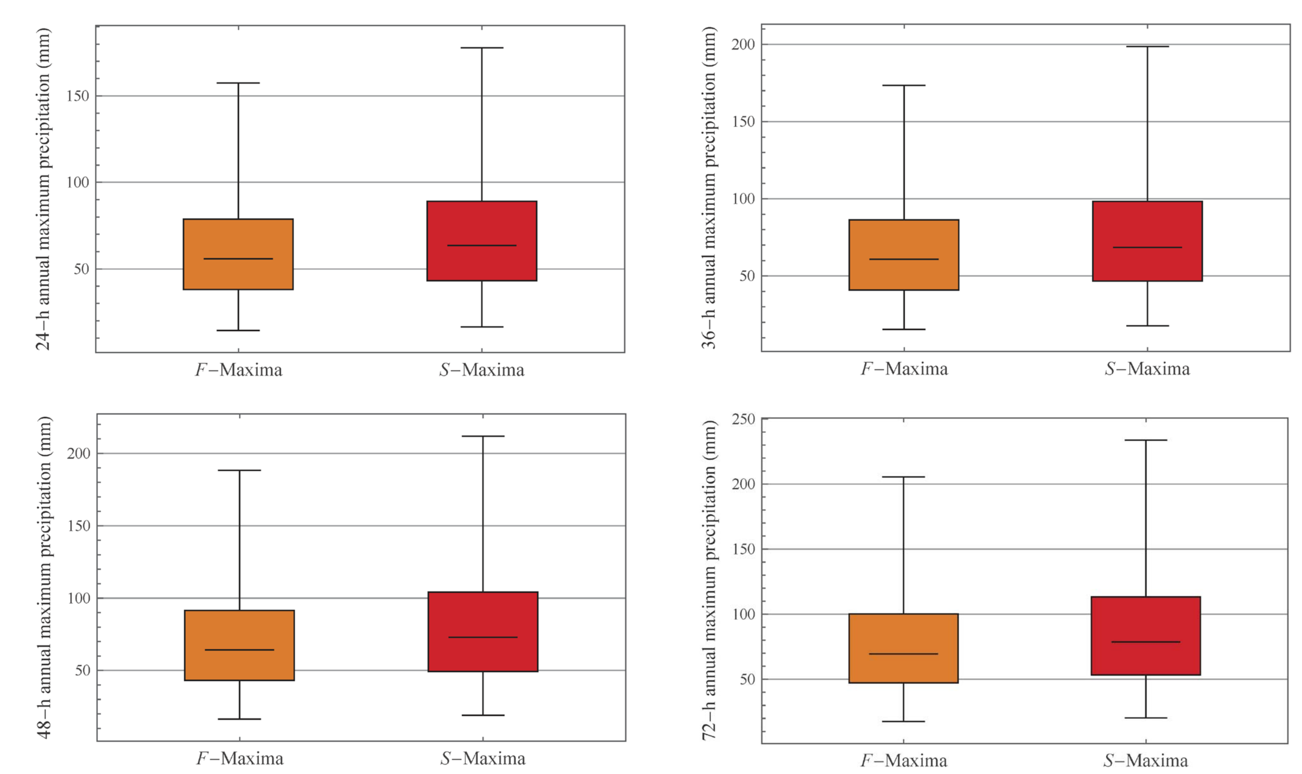

Fig. 10. Box plots of the estimated annual sliding and fixed-interval maxima for the following time scales: 24 h, 36 h, 48 h and 72 h. The outer fences of the box plots express the 95% empirical confidence interval.

9 Statistics of *H*-factors

Table 3. Basic summary statistics of H factors at various time scales

Scale	No of yrs	*H* = 1 (%)	Mode	Median	Mean	SD	C_v	C_s	C_k	τ_2	τ_3	τ_4
2-h	121165	55.87	1	1.00	1.09	0.16	0.14	2.17	8.03	0.06	0.55	0.22
4-h	121165	36.54	1	1.05	1.12	0.16	0.14	1.87	6.83	0.07	0.43	0.14
6-h	121165	31.23	1	1.06	1.13	0.16	0.14	1.80	6.46	0.07	0.41	0.14
8-h	121165	28.29	1	1.07	1.13	0.17	0.15	1.75	6.21	0.07	0.39	0.13
12-h	121165	27.82	1	1.07	1.13	0.17	0.15	1.72	6.01	0.07	0.39	0.13
16-h	121165	26.84	1	1.07	1.14	0.17	0.15	1.67	5.74	0.08	0.39	0.13
20-h	121165	26.02	1	1.07	1.14	0.18	0.16	1.66	5.66	0.08	0.39	0.13
24-h	121165	23.83	1	1.08	1.15	0.19	0.16	1.58	5.31	0.08	0.37	0.12
30-h	121165	25.57	1	1.08	1.15	0.18	0.16	1.61	5.40	0.08	0.38	0.12
36-h	121165	27.49	1	1.07	1.14	0.18	0.16	1.66	5.61	0.08	0.40	0.13
42-h	121165	28.39	1	1.07	1.14	0.18	0.16	1.68	5.67	0.08	0.40	0.13
48-h	121165	27.53	1	1.07	1.15	0.19	0.16	1.65	5.56	0.08	0.40	0.13
56-h	121165	27.88	1	1.07	1.15	0.19	0.16	1.65	5.55	0.08	0.40	0.13
64-h	121165	19.94	1	1.07	1.14	0.18	0.16	1.69	5.69	0.08	0.41	0.13
72-h	121165	17.38	1	1.07	1.15	0.19	0.16	1.64	5.49	0.08	0.39	0.13

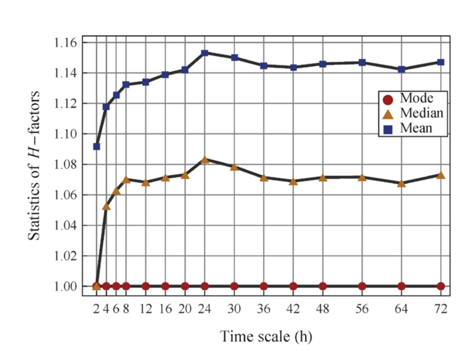

Fig. 10. Summary statistics of *H*-factors *vs.* time scale.

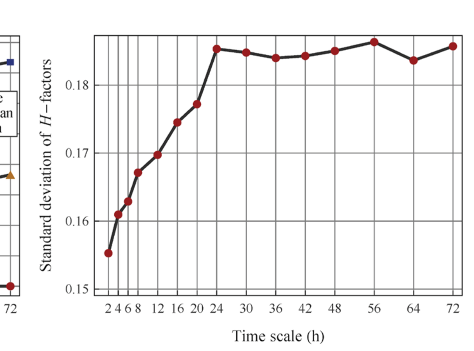

Fig. 11. Standard deviation of *H*-factors *vs.* time scale.

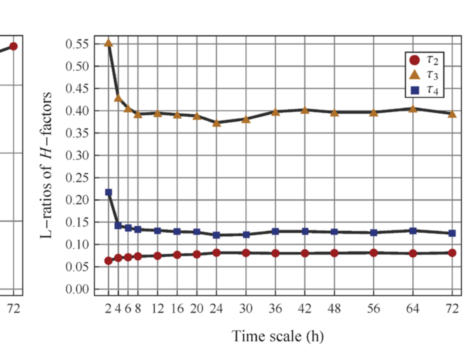

Fig. 12. L-ratios τ_2, τ_3 and τ_4 of *H*-factors *vs.* time scale.

10 Empirical distribution of *H*-factors (I)

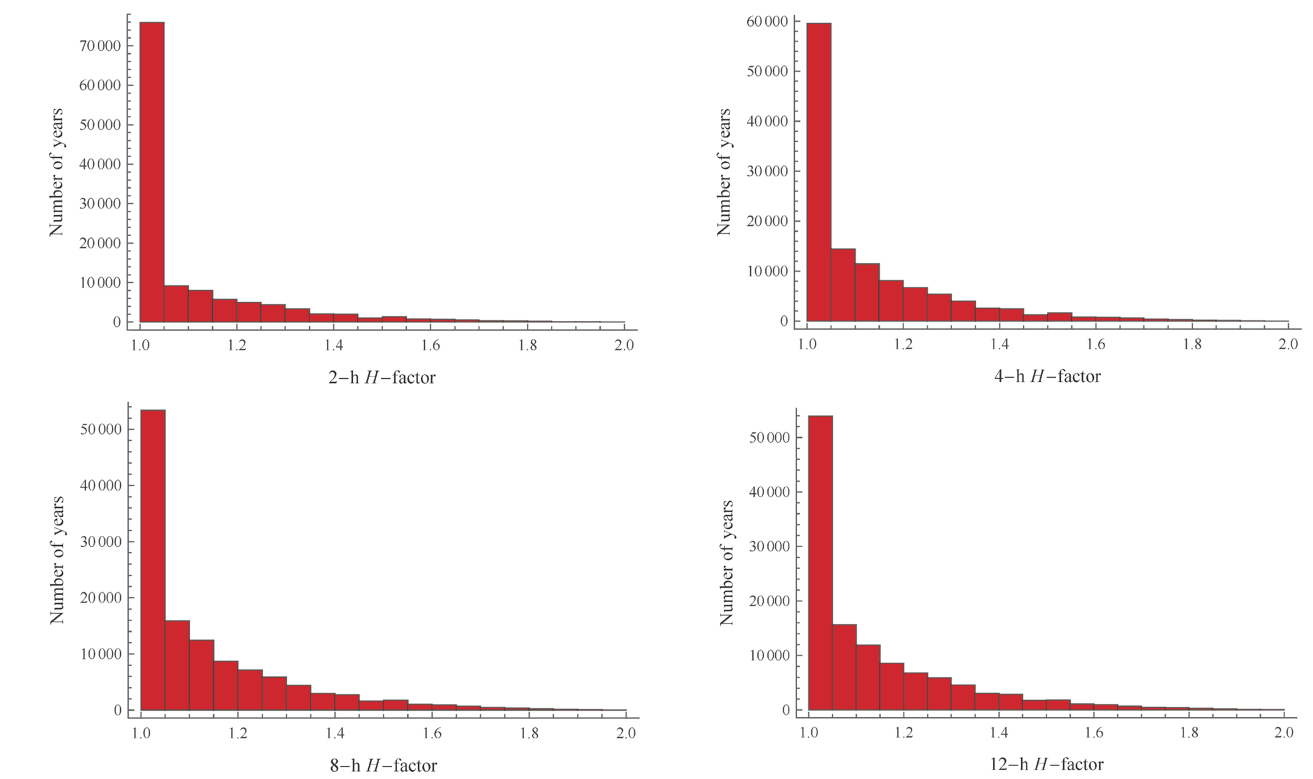

Fig. 14. Empirical distributions of the estimated *H* factors for the following time scales: 2 h, 4 h, 8 h and 12 h.

11 Empirical distribution of *H*-factors (II)

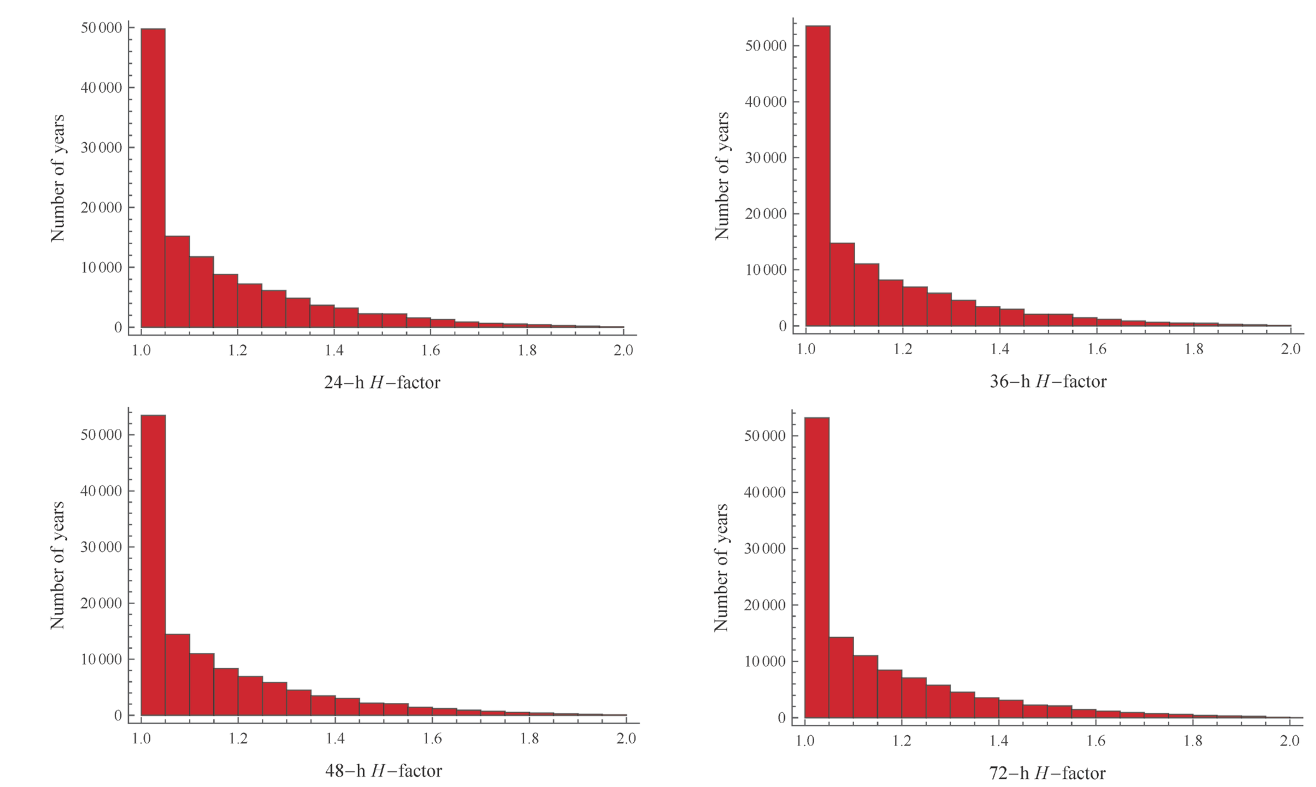

Fig. 15. Empirical distributions of the estimated *H* factors for the following time scales: 24 h, 36 h, 48 h and 72 h.

12 Conclusions

- Hershfield's factor is an important quantity as it corrects for the effects of the temporal discretization and thus can be crucial for quantifying better rainfall extremes and leading to a safer hydrological design.
- Here, we performed an unprecedentedly large analysis of thousands of hourly precipitation records across the USA. We estimated the annual sliding and fixed-interval maxima for more than a 100 000 years and for several time scales in order to estimate the *H*-factors.
- We examined the assumption that the distribution of *S*- and *F*-maxima differs also in its shape characteristics. We found that sample shape measures as the L-ratios (τ_2, τ_3, τ_4) essentially do not significantly differ, implying thus that the distribution in terms of shape remains the same.
- We verify that the difference between the *S*- and the *F*-maxima distributions lies in the mean and in the standard deviation values.
- We estimated thousands of *H*-factors and for several time scales. Our estimates are in agreement with the values given in literature. We note though that our analysis is by far the largest one in terms of number of stations analyzed if compared to other related studies.
- It is worth noting that the values of *H*-factors used in practice so far are mean values. Yet, the mode and the median values (see Table 3), which essentially express more probable values, differ from the mean and are markedly smaller.

References

- Dwyer, I. J., & Reed, D. W. (1994). Effective fractal dimension and corrections to the mean of annual maxima. Journal of Hydrology, 157(1), 13-34.
- Dwyer, I.J. and Reed, D.W., 1995a. Allowance for discretization in hydrological and environmental risk estimation. Institute of Hydrology, Wallingford, UK; Report No. 123, 45 pp.
- Dwyer, I.J. and Reed, D.W., 1995b. Correcting mean annual maxima for data discretization. Preprints 6th Int. Meet. on Statistical Climatology; Galway, Ireland, pp. 447–450.
- Faiers, G.E., J.M. Grymes, III, B.D. Keim, and R.A. Muller. 1994. A Reexamination of Extreme 24
- Hour Rainfall in Louisiana, U.S.A. Climate Research 4:2531.
- Harihara Ayya, P.S. and Tripathi, N., 1973. Relationship of the clock-hour to 60-min and the observational day to 1440-min rainfall. Ind. J. Meteorol. Geophys., 24 (3): 279-282.
- Hershfield, D.M. & Wilson, W.T. 1958. Generalizing of rainfall-intensity-frequency data. *IUGGIIAHS publication no. 43,* 499-506.
- Hershfield, D. M., 1961a. Rainfall frequency atlas of the United States. Weather Bureau Technical Paper 40, U.S. Department of Commerce, Washington. DC
- Hershfield, D. M., Estimating the probable maximum precipitation, Proc. ASCE, J. Hydraul. Div., 87(HY5), 99-106, 1961b.
- Huff, F. A. and J. R. Angel, 1992. Survey Rainfall Frequency Atlas of the Midwest (Bulletin 71). Illinois State Water
- Kerr, R.L., McGinnis, D.F., Reich, B.M. & Rachford, T.M. 1970. Analysis of rainfall-duration-frequency for Pennsylvania. Institute for Research on Land and Water Resources, The Pennsylvania State University, research publication 70.
- Natural Environment Research Council, 1975. Flood Studies Report, 5 vols. Natl. Environ. Res. Counc. London.
- van Montfort, M. A. (1990). Sliding maxima. *Journal of Hydrology, 118*(1), 77-85.
- van Montfort, M. A. (1997) Concomitants of the Hershfield factor.
- Weiss, L. L. (1964). Ratio of true to fixed-interval maximum rainfall. *Journal of Hydraulics Division, 90*, 77-82.