MEAN ERGODICITY IN REGIONAL RAINFALL−FREQUENCY ANALYSES

MILAN ONDERKA[1]. KONSTIANTYN SOKOLCHUK [2]. RATTAYOVÁ VIERA [1] . JOZEF PECHO [1] ✉ milan.onderka@shmu.sk ☎+421 259 415 347

[1] SLOVAK HYDROMETEOROLOGICAL INSTITUTE. JESÉNIOVA 17. BRATISLAVA. SLOVAKIA. SK − 833 15

[2] INSTITUTE OF HYDROLOGY SAS. DÚBRAVSKÁ CESTA 9. BRATISLAVA. SLOVAKIA. SK − 841 04

3 Institute of Landscape Engineering. Slovak University of Agriculture in Nitra. Tulipánová 7. Nitra. SK– 949 76. Slovakia

This study integrates quasi-ergodicity with regional frequency analysis to enhance rainfall frequency estimates for cases when only short or spatially uneven rainfall records are available. Since extreme rainfall events are rare and rainfall observations are often short, extending records at a single site does not necessarily improve frequency estimates. Instead, regional frequency analysis is required, particularly for ungauged locations, to ensure more reliable rainfall frequency estimations. By applying fuzzy clustering alongside with the mean ergodic assumption, we provide a methodological framework that enhances estimation accuracy, particularly for regions with short rainfall records. Our findings suggest that quasi-ergodicity can effectively reduce random errors, improve spatial consistency of regional estimates of rainfall frequencies.

We demonstrate our approach on real-world rainfall data. The Fuzzy C-Means (FCM) clustering was deployed to classify rainfall durations and determine the optimal number of clusters. The Generalized Extreme Value (GEV) distribution was fitted to module coefficients, and Bayesian inference was used to estimate 90% credible intervals, improving the reliability of frequency estimates.

To assess regional homogeneity, we performed analysis of variance across the identified clusters. The results confirmed that clustering effectively reduced variance, validating the appropriateness of the selected number of clusters and enhancing the estimates of rainfall frequencies. At the end, depth-duration-frequency (DDF) curves were generated for recurrence intervals from 2 − 1000 years, revealing that credible intervals widen with increasing recurrence intervals. The overlap of credible intervals beyond 50 years suggests that regional frequency analysis is most effective for shorter recurrence intervals, while at higher recurrence intervals, the study region behaves as a single homogeneous unit.

OBJECTIVES:

The primary objective was to perform regional frequency analysis of short-duration rainfall intensities for duration 5−240 minutes by applying the concepts of quasi-ergodicity as a test of cluster homogeneity. The consequent goal was to estimate regional IDF curves for short-duration rainfall quantiles at a regional scale.

LOCAL FREQUENCY ESTIMATES

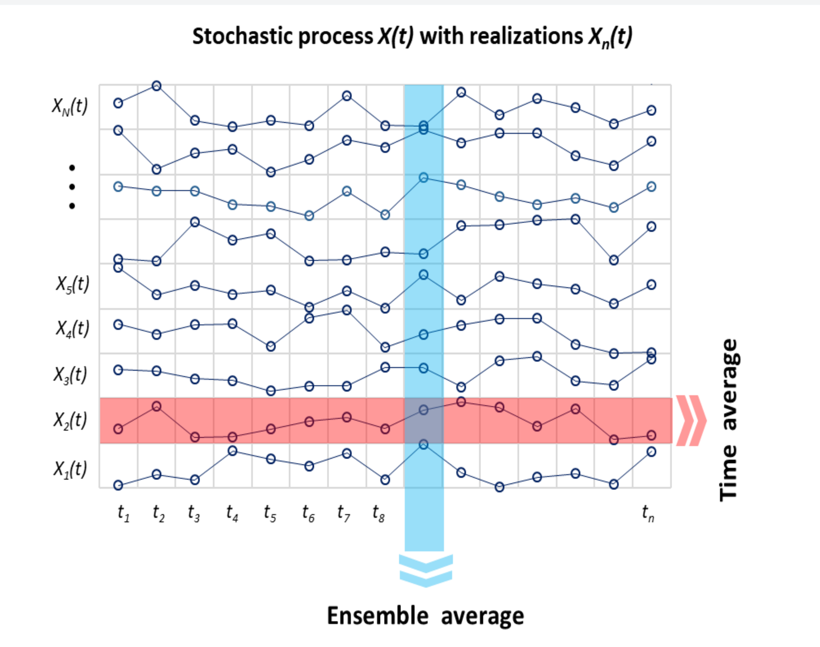
For each location (rain gauge) a Generalized Extreme Value distribution was fitted to annual maximum series (AMS) to estimate qunatiles corresponding annual exceedance probabilities p = 0.5 – 0.001.

PROBLEMS OF LOCAL ESTIMATES

Short records (< 20 years) ⇒ sampling errors ⇒ high uncertainty in quantile estimates especially for $p < 0.1$

REGIONAL FREQUENCY ESTIMATES

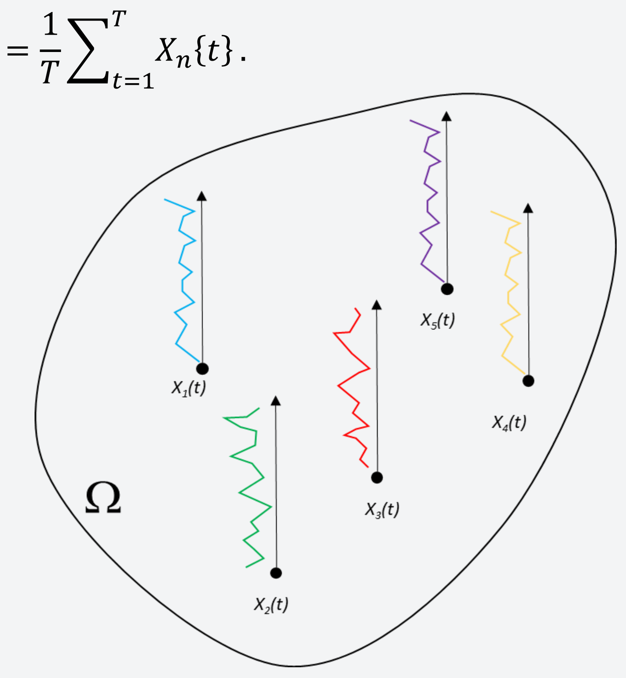
The question whether the global records into a cluster are homogeneous may be answered in a statistical sense by applying the **theoreme of ergodicity**.



Schematized concept of the mean ergodicity theorem. The horizontal axis represents time and the vertical axis represents N realizations of a stationary stochastic process $X(t)$.

$$\bar{X}_T = \frac{1}{T}\int_0^T X_n\{t\}dt \qquad \langle X \rangle = \frac{1}{N}\sum_{n=1}^{N} X_n\{t\}$$

$$\bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_n\{t\}.$$



Distribution of process realizations $X_i(t)$ in space Ω (a statistically homogeneous region). where each dot represents a hypothetical rain gauge with a series of annual maxima.

AMS series (from the individual rain gauges) were standardized to module coefficients

$$k_i = \frac{x_i}{\bar{x}}$$

Station-year method
The probability of exceedance can be determined on the global data of length n_j as

$$P = \frac{m-0.3}{\sum n_j + 0.4}$$

The length of the global series is:

$$n^* = n_1 + n_2 + n_3 + \ldots + n_N = \sum n_j .$$

where $n_1 . n_2 . n_3 \ldots n_N$ are the length of the individual series.

To eliminate the effect of violating the independence (correlation between adjacent stations) we get a reduced length of the global series:

$$n^* = n_1 + n_2(1-r_2) + n_3(1-r_3) + \ldots + n_N(1-r_N).$$

where r_i are the maxima correlation coefficients between pairs of stations.

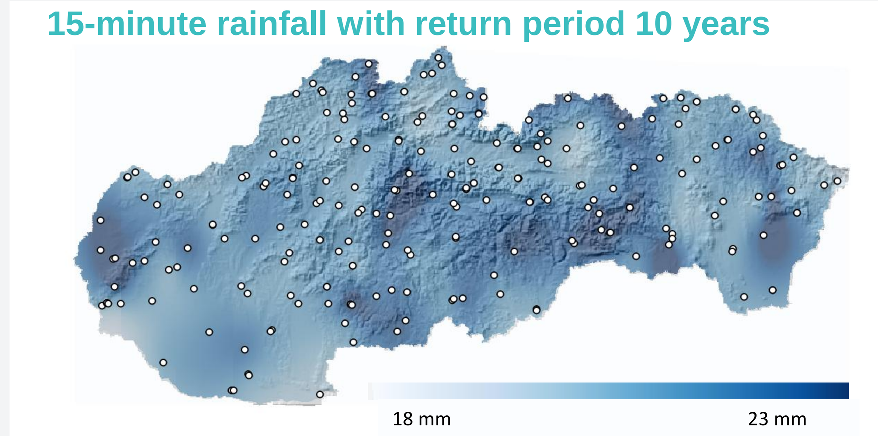
The global series is divided into groups of λ elements:

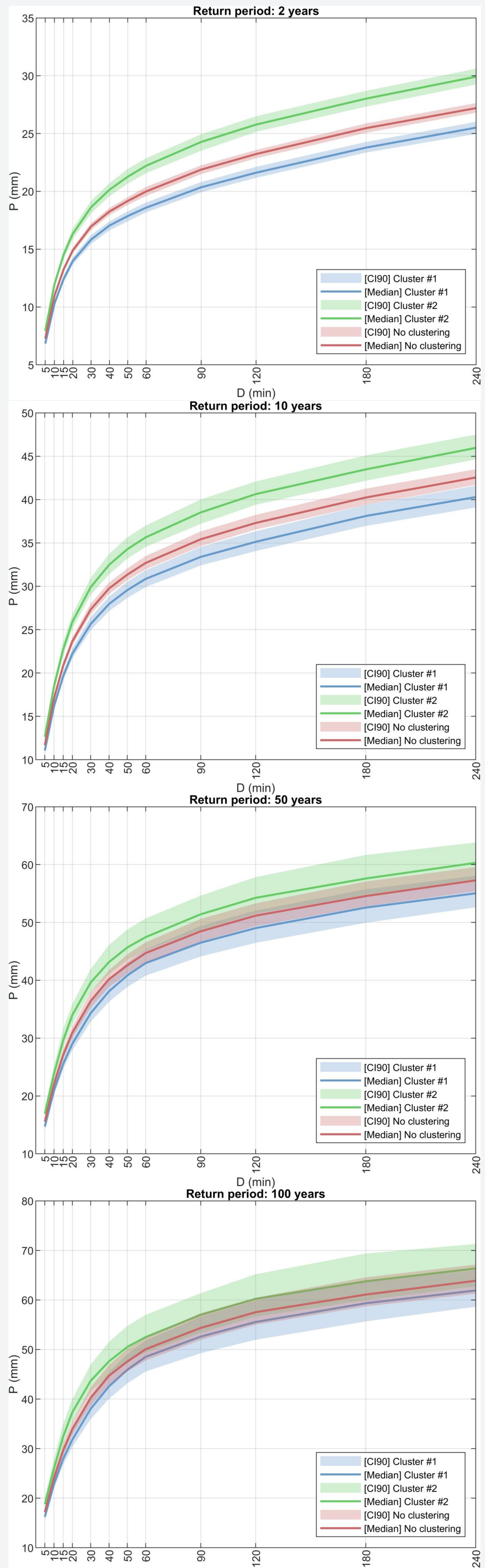
$$\lambda = \frac{\sum n_j}{n^*} = \frac{n_1 + n_2 + .. + n_N}{n_1 + n_2(1-r_2) + + n_N(1-r_N)}$$

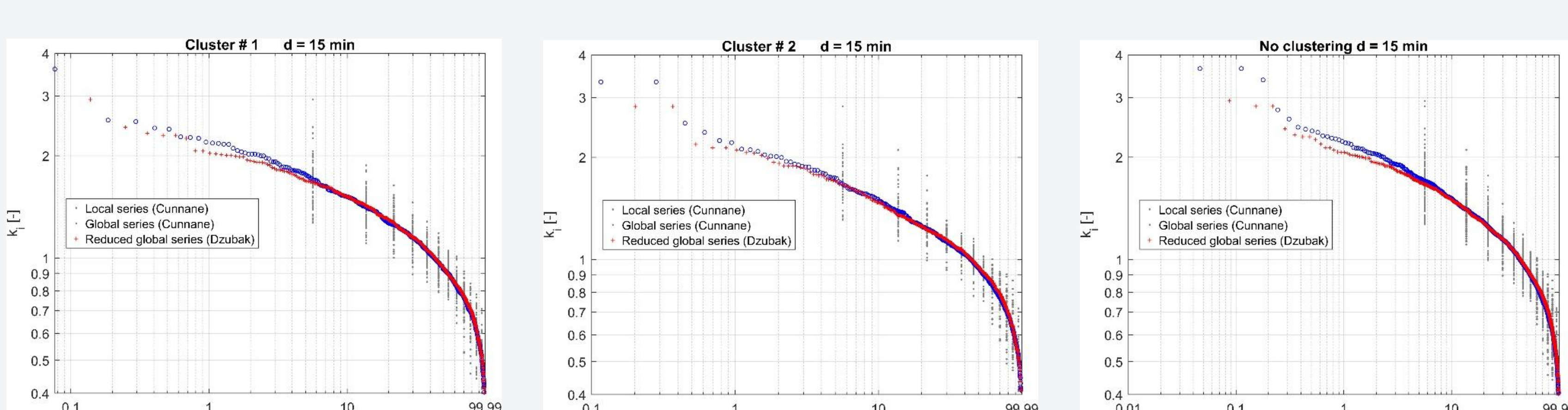
The plotting position (empirical probability of exceedance) for the reduced series has been defined by Dzubák (1969) as:

$$P_m = \frac{1}{n^* + 0.4}\left(0.7 - \frac{\lambda - 1}{2\lambda}\right) + \frac{m-1}{n+0.4}$$

15-minute rainfall with return period 10 years







Probability exceedance curves calculated for module koefficients $k_i = x_i/\bar{x}$. The reduced global series reflect are sed to eliminate the effects of mutual correlations between adjacent locations [1.3.4].

HOMOGENEITY TESTING

The total spatial variance σ_s^2 is considered as the sum of random σ_r^2 and geographical σ_g^2 components [5]:

$$\sigma_s^2 = \sigma_r^2 + \sigma_g^2$$

The random component is considered as the variance of the global series

$$\sigma_g^2 = \frac{\sigma_x^2}{n}$$

The spatial variance of the AMS means was estimated as

$$\sigma_s^2 = \frac{1}{k-1}\sum_{j=1}^{k}\left(\bar{\bar{\mu}} - \bar{x}_j\right)^2$$

where $\bar{\bar{\mu}}$ the mean of means calculated from the individual k series; and x_j is determined for each series as the quantile corresponding to probability of exceedance $p = 0.5$ using the GEV distribution.

The question whether the records in a grouping of stations (clusters) are homogeneous (quasi-ergodic) may be answered in a statistical sense by applying the criterion [3. 4. 5]:

$$\frac{\sigma_s^2}{\left(\frac{\sigma_x^2}{n}\right)} < F_{critical}$$

$\sigma_{spatial}/(\sigma_{rand}/\sqrt{n})$			
D (minutes)	Cluster # 1	Cluster # 2	No clustering
5	1.41	1.28	1.45
10	1.25	1.14	1.34
15	1.20	1.07	1.30
20	1.10	1.04	1.26
30	1.03	1.11	1.33
40	1.15	1.10	1.34
50	1.20	1.15	1.38
60	1.12	1.16	1.37
90	1.00	1.18	1.36
120	1.03	1.22	1.38
180	1.04	1.25	1.40
240	1.08	1.27	1.41

The ratio $\sigma_{\bar{x}_k}/(\sigma_x/\sqrt{n})$ was evaluated for each cluster separatelly. and the whole dataset (no clustering).

CLUSTERING

Fuzzy C-Means (FCM) clustering was applied to a normalized dataset to reduce outlier effects. The algorithm initialized cluster centroids and a fuzzy membership matrix. iteratively updating membership values and centroids based on point distances. Optimization continued until changes in centroids or memberships fell below a threshold. The final outputs included cluster centroids and a membership matrix. reflecting the degree of association of each data point to multiple clusters.

SLOVAK HYDROMETEOROLOGICAL INSTITUTE

REFERENCES

[1] Dzubák M. 1969. Stanovenie zákona rozdelenia maximálnych denných zrážkových úhrnov pomocou globálnych čiar prekročenia. Vodohospodársky časopis 3. 209-223. (in Slovak)
[2] Hosking J.R. M. & Wallis J.R. (1997). Regional Frequency Analysis: An Approach Based on L-Moments. Journal of the American Statistical Association 93. 1233.
[3] Kupčo M. 1985. Estimation of exceedance curves of maximum discharges from short-term forecasts. Collection of papers of the Slovak Hydrometeorological Institute. Bratislava. No. 25. 9-87. (in Slovak)
[4] Kupčo M.. Nejedlík P.. Onderka M. 2001. Problems of short-term rainfall intensities in Slovakia. Meteorological Journal. 24(1). ISSN 1335-339X (in Slovak)
[5] Loboda N.S. et al. 2005. Using meteorological data for reconstruction of annual runoff series over an ungauged area: Empirical orthogonal function approach to Moldova–Southwest Ukraine region . Atmospheric Research 77 (2005) 100– 113 . doi:10.1016/j.atmosres.2004.10.020