Regional analysis of extreme precipitation events in the Czech Republic
Regionální analýza extrémních srážkových jevů v ČR

Jan Kyselý[1], Jan Picek[2], Radan Huth[1]

[1] Institute of Atmospheric Physics AS CR, Prague, [2] Technical University, Liberec, Czech Republic

Abstract

Extreme high precipitation amounts are among environmental events with the most disastrous consequences for human society. Estimates of their return periods and design values are of great importance in hydrologic modelling, engineering practice for water resources and reservoirs design and management, planning for weather-related emergencies, etc. The L-moment based method of the regional frequency analysis of maximum annual 1- to 7-day precipitation totals is currently being utilized for the area of the Czech Republic; daily precipitation amounts over 1961-2000 measured at 78 stations are used as an input dataset. The first step of the regional analysis consists in an identification of homogeneous regions. This contribution deals with a comparison of methods used in statistical testing for homogeneity of regions; they include the L-moment X10 test, tests based on L-moment ratios and on the variation of L-moment statistics. Candidate regions were formed by the cluster analysis of site characteristics (including longitude, latitude, elevation, mean annual precipitation, and mean ratio of summer half-year [May to October] to winter half-year [November to April] precipitation), using the average-linkage clustering and Ward's method. The final homogeneous regions enter the next steps of the regional frequency analysis which concern selection of the most appropriate distribution, and estimation of parameters and quantiles of the fitted distribution together with their uncertainty.

Key words: extreme precipitation events – regional analysis – L moments – tests for homogeneity

1. Introduction

Extreme environmental events, such as high precipitation amounts and floods, have severe consequences for human society. How frequently an event of a given magnitude may be expected to occur is of great importance, e.g., in planning for weather-related emergencies, reservoir management, design of structures such as bridges etc. An estimation of frequencies of extreme events is difficult due to the fact that extremes are rare and data records usually short.

The main motivation for the planned research was a recent occurrence of massive floods (in 1997 and particularly 2002) in central Europe, and a development of a new method of the estimation of probabilities of extremes based on a regional analysis combined with an L-moment approach during the 1990s (Hosking, 1990; Pilon and Adamowski, 1992; Hosking and Wallis, 1997; Alila, 1999). The L-moment based method of regional frequency analysis is computationally simpler and for small samples superior to more traditional approaches (based mostly on methods of conventional moments or maximum likelihood, and 'at-site' analysis) and has not been applied in studies dealing with return periods of hydrological extremes in the Czech Republic.

This paper focuses on a description of the methodology and datasets used, and findings concerning the formation of homogeneous regions.

2. Regional frequency analysis based on L-moments

Regional frequency analysis

In a regional frequency analysis, data from several sites are used in estimating frequencies at any one site. The 'index-flood'/'index storm' procedure is an example; the assumption is that frequency distributions at N sites from a homogeneous region are identical apart from a site-specific scaling factor, usually termed the 'index flood' in a streamflow analysis and the 'index storm' in a precipitation analysis. The advantage of the regional over 'at-site' estimation is greater at distribution tails which are focused by practical applications. Many methods recommended by national organizations for general use by hydrologists have a regional component; for example, in the U.S. the annual maximum streamflow (Q) is assumed to have a log-Pearson type III distribution, and the skewness of the distribution of log Q is estimated by combining a data-based estimate with a value read from a map (U.S. Water Resources Council, 1981).

Different methods of regional frequency analysis were reviewed by Cunnane (1988) who rated the algorithm based on probability-weighted-moments (PWMs; Greenwood et al., 1979) as the best. L-moments are statistical quantities that are derived from PWMs and increase the accuracy and ease of use of the PWM-based analysis (Hosking and Wallis, 1997).

L-moments

L-moments are a recent development in mathematical statistics which facilitates the estimation process in the frequency analysis; they represent an alternative set of scale and shape statistics of a data sample or a probability distribution. Their main advantages over conventional (product) moments are that they are able to characterize a wider range of distributions, and (when estimated from a sample) are less subject to bias in estimation and more robust to the presence of outliers in the data (e.g. Royston, 1992; Sankarasubramanian and Srinivasan, 1999; Ulrych et al., 2000). The latter is because ordinary moments (unlike L-moments) require involution of the data which causes disproportionate weight to be given to the outlying values. The identification of a distribution from which the sample was drawn is more easily achieved (particularly for skewed distributions) using L-moments than conventional moments (Hosking, 1990). The method of L-moments is also more efficient in estimating parameters of a fitted distribution compared to the maximum likelihood method (Hosking et al., 1985). (See Appendix I for a formal definition of L-moments.)

L-moments may be applied in four steps of the regional frequency analysis (Hosking and Wallis, 1997; Alila, 1999; Adamowski, 2000):

i. *Screening of the data.* L-moments are used to construct a discordancy measure which identifies unusual sites with sample L-moment ratios markedly different from the other sites. These unusual sites merit close examination.

ii. *Identification of homogeneous regions.* L-moments are used to construct a summary statistics in testing heterogeneity of a region.

iii. *Choice of a frequency distribution.* L-moment ratio diagram and/or regional average L-moments are used in testing whether a candidate distribution gives a good fit to the region's data.

iv. *Estimation of the frequency distribution.* Regional L-moments are used to estimate parameters of the chosen distribution.

The L-moment based methods of regional frequency analysis are now being adopted by many organizations worldwide (Hosking and Wallis, 1997). Recent findings indicate that appropriate modifications of the 'standard' procedure can improve its performance and the reliability of design values (e.g. Sveinsson et al., 2001). Future directions and challenges in the regional analysis involve development of a more rigorous statistical methodology, explanation of the superiority of L-moments for small samples, incorporation of covariates into regional extreme value models, and dealing with the spatial dependence of extremes.

3. Data

Daily precipitation amounts measured at 78 stations covering the Czech Republic (area of 78 864 square km, with complex orography), with altitudes ranging from 158 to 1324 m a.s.l., were provided by the Czech Hydrometeorological Institute (Figure 1). The data span the period of 1961-2000; there are no missing values in this dataset. Samples of maximum annual 1- to 7-day precipitation amounts were drawn from each station data, and are examined as extreme precipitation events.

4. Results

Screening of the data

Firstly, the data underwent standard validity checking for gross errors, and suspicious values have been checked using data from nearby stations. Secondly, the sites that are grossly discordant with the group as a whole were identified using the discordancy measure (D) based on L-moments. The formal definition of the discordancy measure can be found in Hosking and Wallis (1993); it yields a value of D_i for each measuring site. Critical values for the discordancy statistic are tabulated; for the number of sites ≥ 15, the critical value is 3.

Sites recognized as discordant at this stage were examined for errors or for sources of unreliability in data; however, all values of D_i>3 have originated from real observed outliers, mostly extraordinarily high 1997 precipitation amounts at a few stations in the northeast part of the Czech Republic.

Scatter-plots of L-skewness against L-CV for maximum annual 1- to 7-day precipitation amounts are shown in Figure 2; the largest outliers in upper right parts of the scatter-plots for 3-, 5- and 7-day amounts are due to the record high totals in 1997, particularly at mountainous station Lysá hora (1324 m a.s.l.), and do not reflect unreliability in measurements.

Identification of regions

The formation of regions was based on the cluster analysis of six 'site characteristics': longitude, latitude, elevation, mean annual precipitation, mean ratio of summer half-year (May to October) to winter half-year (November to April) precipitation, and mean annual number of dry days (defined as days with precipitation amount <=0.1 mm). Using 'at-site statistics' (quantities calculated from the at-site values of the analyzed variables) instead of/together with the 'site characteristics' would compromise results since there would be a tendency to group together all sites that have high outliers, even though these outliers result from random fluctuations, and testing for the homogeneity of the formed regions by a statistic calculated from the 'at-site statistics' would be misleading (Smithers and Schulze, 2001).

The average-linkage clustering (which tends to form clusters with equal within-cluster variance) as well as Ward's method (which tends to form clusters with equal number of sites) were applied as clustering algorithms (Guttmann, 1993). The latter yields slightly superior results, particularly because of the undesirable 'snowball effect' (Kalkstein et al. 1987; Huth et al. 1993) present in the average-linkage clustering outputs (one big cluster is produced to which smaller clusters are stuck which are more and more dissimilar from the mean). Reasonable numbers of clusters are 8, 4 and 3 for Ward's method, and 5 and 3 for the average-linkage (with 5 and 4 sites unclassified in the latter case); homogeneity tests for all sites taken as one region were performed as well. Examples of unadjusted partitionings based on both methods are given in Figure 3. However, subjective adjustments (mainly according to the site location and its climatological characteristics) are necessary in all cases to improve the geographical and climatological coherence of regions and to avoid heterogeneity.

Testing for homogeneity of regions

Tests for the homogeneity of regions are usually based on a quantity that measures some aspect of the frequency distribution, e.g. the 10-yr event (Lu and Stedinger, 1992), the combination of the L-coefficient of variation L-CV and the L-skewness τ_3 (Chowdhury et al., 1991) or the combination of L-CV, τ_3 and the L-kurtosis τ_4 (Hosking and Wallis, 1993; Adamowski, 2000), and compare the 'at-site' estimates with the regional estimate of this quantity. (See Appendix I for a formal definition of L-CV, τ_3 and τ_4). The mean and standard deviation of the chosen dispersion measure are obtained by a simulation of a homogeneous region with sites having record lengths the same as the observed data (Monte Carlo method).

The tests employed in the present study were those of Lu and Stedinger (1992), Hosking and Wallis (1993), and Alila (1999); see Appendix II for their description.

As expected, none of the partitionings based on the cluster analysis yields only homogeneous regions in any of the variables examined (maximum annual 1-, 3-, 5- and 7-day precipitation amounts) and any of the tests. 'More regions' do not necessarily mean 'more homogeneity' owing to different sample sizes, different parameters of the distribution used in the simulations, and different means and variances of the dispersion measure in the simulated homogeneous region. Since Hosking-Wallis tests and Alila tests yield very similar results, only results of Hosking-Wallis tests are further evaluated.

Tests based on L-CV (H1) are more frequently indicating a heterogeneity or a potential heterogeneity than tests based on the L-skewness (H2) and the L-kurtosis (H3); H2 and H3 tests have a very small discriminatory power and should be avoided. Lu-Stedinger test indicates a heterogeneity more frequently than the other tests, and there is no general agreement between results of Lu-Stedinger and Hosking-Wallis tests: the heterogeneous regions according to Lu-Stedinger test are acceptably homogenous in 78% according to both H2 and H3, and in 33% according to H1.

The most promising partitioning is obtained by Ward's method of the cluster analysis with 4 clusters (Figure 3 / left): two of them form large regions (comprising 83% of sites; clusters 1 and 3 in Figure 3) which are climatologically reasonable and homogenous according to large majority of the tests and tested variables.

Final formation of homogeneous regions

Several subjective adjustments were necessary; the final partitioning recognizes 4 (homogeneous) regions ranked with respect to the number of sites (Figure 4):

- region 1: lowland stations in the area stretching from northwest Bohemia to southeast Moravia (30 stations; the elevation range from 158 to 468 m a.s.l., the mean elevation 276 m a.s.l.; corresponds approximately to cluster 3 of Ward's method with 4 clusters - see left panel of Figure 3); the region is homogeneous according to all tests and variables except for Lu-Stedinger test for maximum annual 1-day precipitation amounts
- region 2: higher-elevated stations in the west and central parts of the Czech Republic (27 stations; the elevation range from 429 to 1118 m a.s.l., the mean elevation 553 m a.s.l.; corresponds approximately to cluster 1 of Ward's method with 4 clusters); the region is homogeneous according to all tests and variables
- region 3: northeast Moravia (12 stations; the elevation range from 220 to 750 m a.s.l., the mean elevation 391 m a.s.l.; typical for the region are considerably enhanced mean maximum annual k-day precipitation amounts as well as mean annual precipitation [relative to the elevation of sites], and the region covers the area most affected with the 1997 record-high precipitation amounts); the region is homogeneous according to all tests and variables
- region 4: north Bohemia (7 stations; the elevation range from 370 to 495 m a.s.l., the mean elevation 413 m a.s.l.; corresponds approximately to the west part of cluster 2 of Ward's method with 4 clusters; typical for the region are enhanced mean annual precipitation, low number of dry days, and low ratio of summer to winter precipitation); the region is homogeneous according to all tests and variables except for the H1 test for maximum annual 5-day and 7-day precipitation amounts

Two stations (Lysá hora and Bedřichov) are unclassified; their inclusion in any of the regions leads to a considerable distortion of its homogeneity.

5. Conclusions

The disastrous consequences of extreme high precipitation events, resulting in floods, may become more pronounced in a future climate since an increase in their frequency and severity is expected and/or observed in large parts of Europe (IPCC, 2001). This research was motivated by the recent occurrence of severe summer floods (in 1997 and particularly 2002) in central Europe. It makes use of the development in environmental sciences, the L-moment based method of the regional frequency analysis, which has not been applied in studies dealing with return periods of hydrological extremes in the Czech Republic.

The study is currently in an early stage. The area of the Czech Republic has been divided into 4 homogeneous regions, based on the cluster analysis of site characteristics and tests for the homogeneity of the regions. The regions will enter next steps of the regional frequency analysis which concern (i) the selection of the most appropriate distribution of extreme precipitation events, and (ii) the estimation of parameters and quantiles of the fitted distribution together with their uncertainty, with an emphasis on return periods of the 1997 and 2002 extreme precipitation events which caused massive floods in central Europe. Benefits of the regional frequency analysis of precipitation extremes compared to the at-site analysis will be evaluated.

Acknowledgement: The study is supported by the Grant Agency of the Academy of Sciences of the Czech Republic under project B3042303. Thanks are due to J.Hošek for assistance in drawing the figures.

Appendix I: Definition of L-moments

Derivation of L-moments is based on order statistics which are obtained simply by sorting the sample $\{X_1, X_2, ..., X_n\}$ of n independent realizations of variable X in ascending order $\{X_{1:n}, X_{2:n}, ..., X_{n:n}\}$; the subscript $k{:}n$ denotes the k-th smallest number in the sample of length n. L-moments λ_k are defined as expectations of linear combinations of these order statistics,

$$\lambda_1 = E(X_{1:1}), \quad \lambda_2 = \frac{1}{2}E(X_{2:2} - X_{1:2}), \quad \lambda_3 = \frac{1}{3}E(X_{3:3} - 2X_{2:3} + X_{1:3}),$$

and generally for the k-th L-moment

$$\lambda_k = \frac{1}{k}\sum_{j=0}^{k-1}(-1)^j \binom{k-1}{j} E(X_{k-j:k}),$$

where E denotes expectation operator (Hosking, 1990; von Storch and Zwiers, 1999). The first L-moment is the expected smallest value in a sample of one, i.e. the conventional first moment. The second L-moment is the expected absolute difference between any two realizations, multiplied by 1/2 (i.e., analogue to the conventional second moment). The third and fourth L-moments are shape parameters. L-moment ratios (sometimes termed standardized L-moments) are the L-coefficient of variation $\frac{\lambda_2}{\lambda_1}$ (L-CV), the L-skewness $\frac{\lambda_3}{\lambda_2}$ (τ_3) and the L-kurtosis $\frac{\lambda_4}{\lambda_2}$ (τ_4); they take values between -1 and +1 (except for some special cases of small samples).

Hosking (1990) showed that the k-th L-moment λ_k ($k \le n$) can be estimated as

$$l_k = \sum_{l=0}^{k-1}(-1)^{k-l-1}\binom{k-1}{l}\binom{k+l-1}{l} b_l ,$$

where

$$b_l = \frac{1}{n}\sum_{i=1}^{n}\frac{(i-1)(i-2)\ldots(i-l)}{(n-1)(n-2)\ldots(n-l)} X_{i:n}, \; l \ge 1, \text{ and } b_0 = \frac{1}{n}\sum_{i=1}^{n} X_{i:n}$$

For the first three L-moments, estimators can be expressed in much simpler form as

$$l_1 = \frac{\sum_i X_i}{n}, \; l_2 = \frac{\sum_{i>j}(X_{i:n} - X_{j:n})}{2\binom{n}{2}}, \text{ and } l_3 = \frac{\sum_{i>j>k}(X_{i:n} - 2X_{j:n} + X_{k:n})}{3\binom{n}{3}}.$$

Appendix II: Description of tests for homogeneity of regions

Suppose that the proposed region has N sites, with site i having record length n_i and sample L-moment ratios $t^{(i)}$ (L-CV), $t_3^{(i)}$ (L-skewness) and $t_4^{(i)}$ (L-kurtosis) of maximum annual k-day (k=1, 3, 5, 7) precipitation amounts.

Test 1 (Hosking and Wallis, 1993):

The test statistic is

$$H_1 = \frac{V_1 - \mu_V}{\sigma_V}, \text{ where } V_1 = \sqrt{\frac{\sum_{i=1}^{N} n_i \left(t^{(i)} - t^R\right)^2}{\sum_{i=1}^{N} n_i}}, \; t^R = \frac{\sum_{i=1}^{N} n_i t^{(i)}}{\sum_{i=1}^{N} n_i}$$

and μ_V, σ_V are determined from simulations (500 realisations of a homogeneous region with N sites, each having a four-parameter kappa distribution with L-moment ratios equal to t^R, t_3^R and t_4^R and the at-site mean equal to 1) as the mean and standard deviation of the simulated values of V_1.

Two other analogous tests are based on L-skewness t_3 (test statistic H_2) and L-kurtosis t_4 (test statistic H_3) instead of L-CV t.

The region is regarded as ‘acceptably homogeneous’ if $H < 1$, ‘possibly heterogeneous’ if $1 \le H < 2$, and ‘definitely heterogeneous’ if $H \ge 2$ (Hosking and Wallis, 1993).

Test 2 (Alila, 1999):

The test statistic is

$$S_1 = \frac{\sigma_1^2 - \mu_{\sigma^2}}{\sigma_1^2}, \quad \text{where } \sigma_1^2 = \frac{\sum_{i=1}^{N} n_i \left(t^{(i)} - t^R\right)^2}{\sum_{i=1}^{N} n_i}$$

and μ_{σ^2} is determined from simulations (500 realisations of a homogeneous region with *N* sites, each having a three-parameter GEV distribution with L-moment ratios equal to t^R, t_3^R and the at-site mean equal to 1) as the mean of the simulated values of σ_1^2.

Two other analogous tests are based on L-skewness t_3 (test statistic S_2) and L-kurtosis t_4 (test statistic S_3) instead of L-CV *t*.

The test yields a heterogeneity measure analogous to Hosking and Wallis (1993).

Test 3 (Lu and Stedinger, 1992):

The test statistic is

$$\chi_R^2 = \sum_{i=1}^{N} \frac{\left(\xi_{0.9}^{(i)} - \xi_{0.9}^R\right)^2}{Var\xi_{0.9}^{(i)}},$$

$$\text{where } \xi_{0.9}^R = \frac{\sum_{i=1}^{N} n_i \xi_{0.9}^{(i)}}{\sum_{i=1}^{N} n_i}, \quad \xi_{0.9}^{(i)} = 1 + \frac{t^{(i)}}{1-2^{-k}}\left(1 - \frac{(-\ln 0.90)^k}{\Gamma(1+k)}\right),$$

$$\text{and } k = 7.8590C + 2.9554C^2, \quad C = \frac{2}{t_3^{(i)} + 3} - \frac{\ln 2}{\ln 3}.$$

Var $\xi_{0.9}^{(i)}$, *i*=1,…*N* was again determined from simulations (500 realisations of a region consisting of *N* sites, each having a three-parameter GEV distribution with L-moment ratios equal to $t^{(i)}$, $t_3^{(i)}$ and the at-site mean equal to 1) as the variance of the 90% sample quantiles.

If $\chi_R^2 < \chi_{0.95,N-1}^R$ (where $\chi_{0.95,N-1}^R$ is the 95%-quantile of χ^2 distribution with N-1 degrees of freedom) we do not reject the null hypothesis (the region is homogeneous) at the 5% significance level; if $\chi_R^2 \geq \chi_{0.95,N-1}^R$ the null hypothesis is rejected (the region is heterogeneous).

References

Adamowski K. 2000. Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods. *J. Hydrol.,* **229**, 219-231.

Alila Y. 1999. A hierarchical approach for the regionalization of precipitation annual maxima in Canada. *J. Geophys. Res. - Atmospheres,* **104**, 31645-31655.

Chowdhury J.U., Stedinger J.R., Lu L.-H. 1991. Goodness-of-fit tests for regional generalized extreme value distributions. *Water Resour. Res.,* **27**, 1765-1776.

Cunnane C. 1988. Methods and merits of regional flood frequency analysis. *J. Hydrol.,* **100**, 269-290.

Greenwood J.A., Landwehr J.M., Matalas N.C., Wallis J.R. 1979. Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. *Water Resour. Res.,* **15**, 1049-1054.

Guttman N.B. 1993. The use of L-moments in the determination of regional precipitation climates. *J. Climate*, **6**, 2309-2325.

Hosking J.R.M., Wallis J.R., Wood E.F. 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics,* **27**, 251-261.

Hosking J.R.M. 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. Stat. Soc.,* **52B**, 105-124.

Hosking J.R.M., Wallis J.R. 1993. Some statistics useful in regional frequency analysis. *Water Resour. Res.,* **29**, 271-281.

Hosking J.R.M., Wallis J.R. 1997. *Regional Frequency Analysis. An Approach Based on L-moments.* Cambridge University Press, Cambridge, New York, Melbourne, 224 pp.

Huth R., Nemešová I., Klimperová N. 1993. Weather categorization based on the average linkage clustering technique: An application to European mid-latitudes. *Int. J. Climatol.,* **13**, 817-835.

IPCC. 2001. *Climate Change 2001. The Scientific Basis.* Cambridge University Press, Cambridge, 881 pp.

Kalkstein L.S., Tan G., Skindlov J.A. 1987. An evaluation of three clustering procedures for use in synoptic climatological classifications. *J. Clim. Appl. Meteorol.,* **26**, 717-730.

Lu L.-H., Stedinger J.R. 1992. Sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test. *J. Hydrol.,* **138**, 223-245.

Pilon P.J., Adamowski K. 1992. The value of regional flood frequency analysis using the method of L-moments. *Can. J. Civ. Engin.,* **19**, 137-147.

Royston P. 1992. Which measures of skewness and kurtosis are best? *Statistics in Medicine,* **11**, 333-343.

Sankarasubramanian A., Srinivasan, K. 1999. Investigation and comparison of sampling properties of L-moments and conventional moments. *J. Hydrol.,* **218**, 13-34.

Smithers J.C., Schulze R.E. 2001. A methodology for the estimation of short duration design storms in South Africa using a regional approach based on L-moments. *J. Hydrol.,* **241**, 42-52.

Sveinsson O.G.B., Boes D.C., Salas J.D. 2001. Population index flood method for regional frequency analysis. *Water Resour. Res.,* **37**, 2733-2748.

Ulrych T.J., Velis D.R., Woodbury A.D., Sacchi M.D. 2000. L-moments and C-moments. *Stoch. Env. Res. Risk Assessm.,* **14**, 50-68.

U.S. Water Resources Council. 1981. Guidelines for determining flood flow frequency. *Bulletin 17B,* Hydrology Committee, Washington, D.C.

von Storch H., Zwiers F. 1999. *Statistical Analysis in Climate Research.* Cambridge University Press, 484 pp.

RNDr. Jan Kyselý, Ph.D., Institute of Atmospheric Physics AS CR, Boční II 1401, 141 31 Prague, Czech Republic; e-mail: honza@ufa.cas.cz

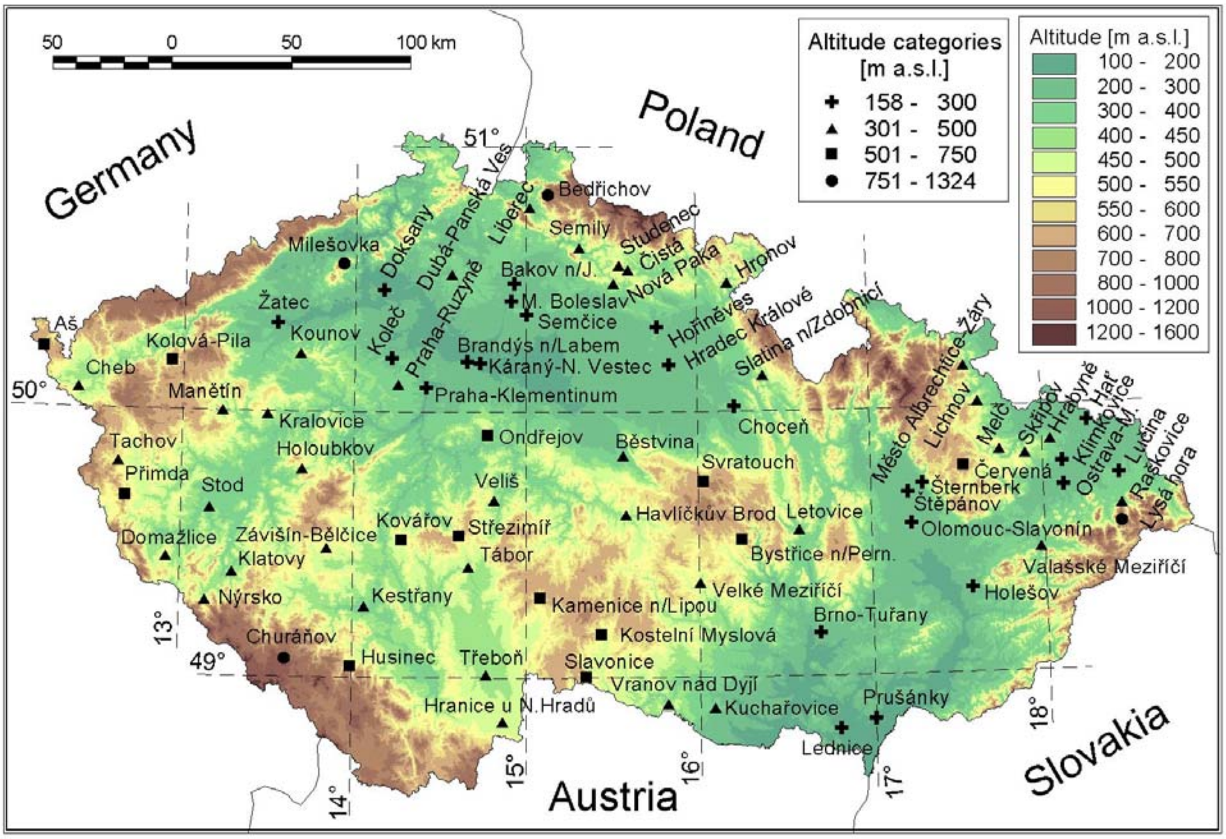

Figure 1. Stations used in the regional frequency analysis of extreme precipitation events in the Czech Republic. Altitude categories (in m a.s.l.) are indicated by symbols.

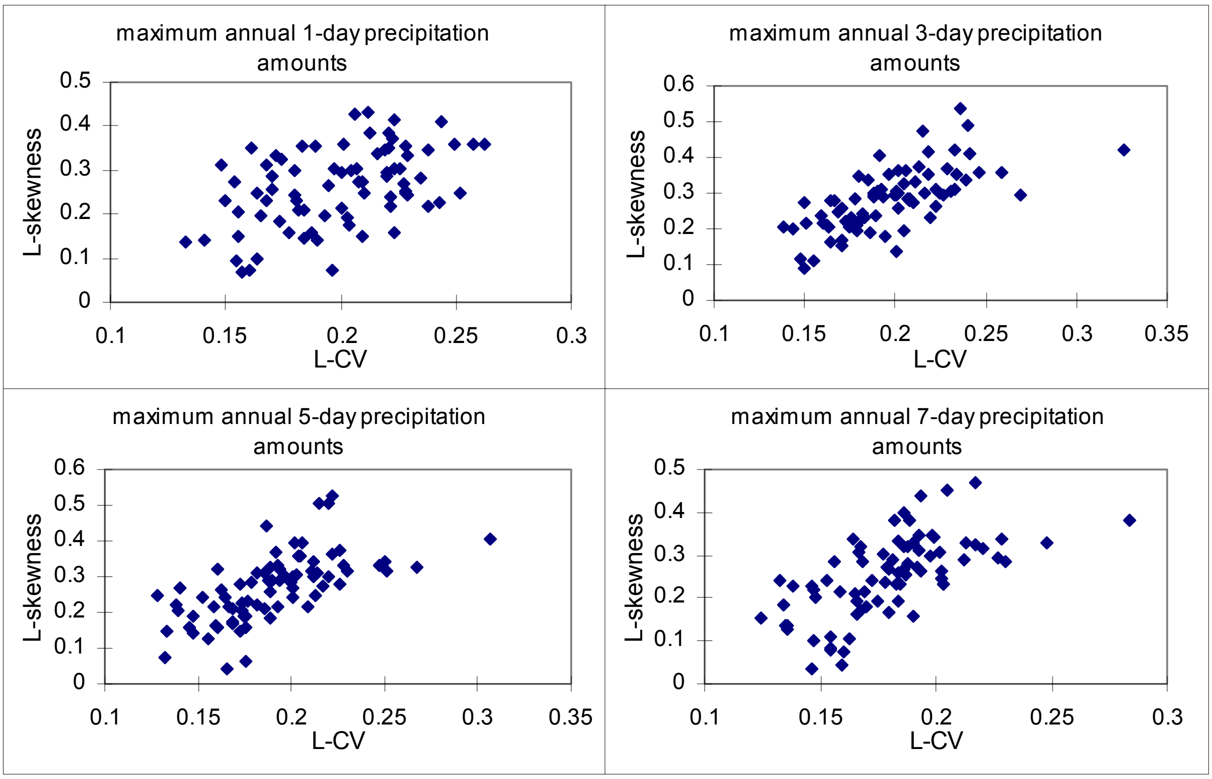

Figure 2. Scatter-plots of L-skewness (τ_3) against L-CV for maximum annual 1-, 3-, 5- and 7-day precipitation amounts. The largest outliers in upper right parts of the scatter-plots for 3-, 5- and 7-day amounts are due to the record high totals in 1997, particularly at mountainous station Lysá hora (1324 m a.s.l.), and do not reflect unreliability in measurements.

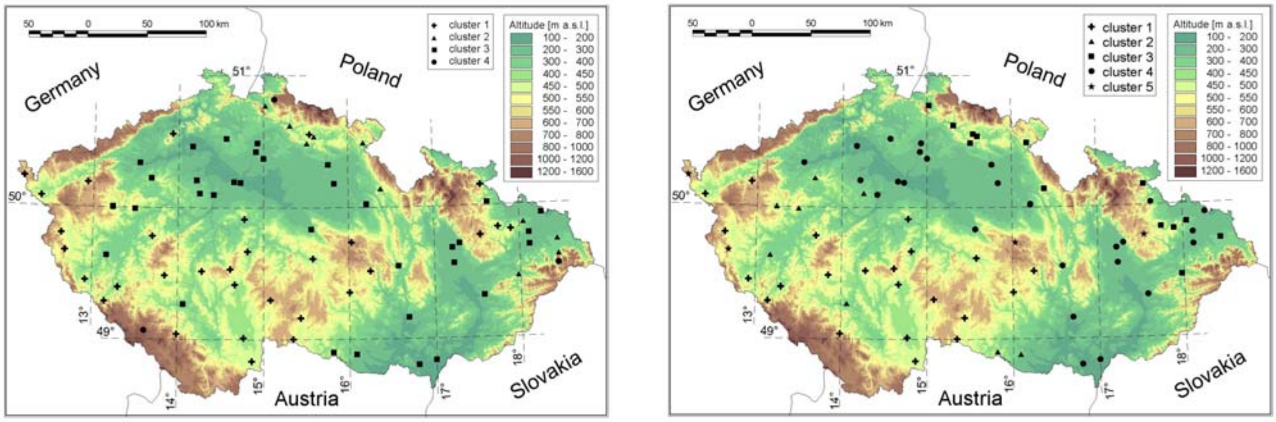

Figure 3. Examples of the unadjusted partitioning based on Ward's method (left, 4 clusters) and the average-linkage method (right, 5 clusters but several sites unclassified) of the cluster analysis.

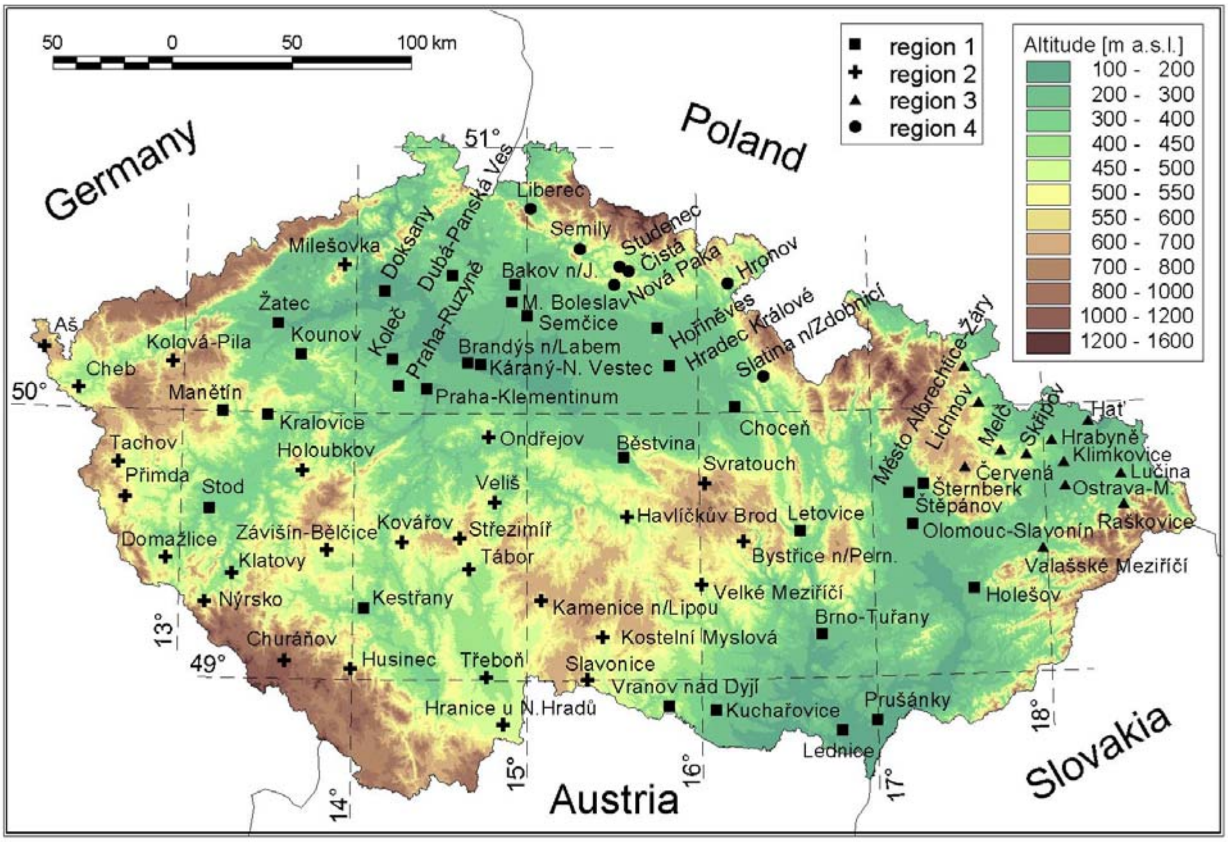

Figure 4. Final formation of the homogeneous regions according to extreme precipitation characteristics. Regions 1 and 2 originate in Ward's method of the cluster analysis of site characteristics; regions 3 and 4 are based mainly on subjective adjustments and testing for the homogeneity.