HYBRID APPROACH TO DELINEATION OF HOMOGENEOUS REGIONS FOR REGIONAL PRECIPITATION FREQUENCY ANALYSIS

LADISLAV GAÁL¹⁾, JÁN SZOLGAY¹⁾, MILAN LAPIN²⁾, PAVOL FAŠKO³⁾

¹⁾Department of Land and Water Resources Management, Faculty of Civil Engineering, Slovak University of Technology, Slovakia; mailto: ladislav.gaal@stuba.sk

²⁾Department of Astronomy, Physics of the Earth and Meteorology, Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava, Slovakia.

³⁾Slovak Hydrometeorological Institute, Bratislava, Slovakia.

Regional frequency analysis of heavy precipitation amounts based on the estimation of the parameters of a regional distribution function using L -moments is adopted for the specific geographical-climatological settings of Slovakia. The paper focuses on the first step of the regional L -moment algorithm (*Hosking, Wallis*, 1997), which is the delineation of homogeneous regions. Objective and process-based logical pooling techniques are used to form homogeneous pooling groups of rainfall gauging stations for regional frequency analysis of k -day precipitation amounts ($k = 1$ to 5 days). Even though the delineation of homogeneous regions by means of objective methods is generally accepted and recommended in the literature, it is concluded here that such a pooling of similar sites should not be carried out automatically in precipitation analysis. Instead, a combination of physical/geomorphological considerations and objective methods should be preferred.

KEY WORDS: Slovakia, Maxima of 1-to-5 Day Precipitation Amounts, Calendar Year, Warm/Cold Season, Regional Frequency Analysis, Homogeneous Regions, L -moments, Cluster Analysis.

Ladislav Gaál, Ján Szolgay, Milan Lapin, Pavol Faško: HYBRIDNÝ PRÍSTUP K VYČLENENIU HOMOGÉNNYCH REGIÓNŮ PRE REGIONÁLNU FREKVENČNÚ ANALÝZU ÚHRNOV ZRÁŽOK. J. Hydrol. Hydromech., 57, 2009, 4; 65 lit., 6 obr., 8 tab.

Článok sa zaoberá regionálnou frekvenčnou analýzou mimoriadnych úhrnov zrážok, ktorá je založená na odhade parametrov regionálneho rozdelenia pravdepodobnosti pomocou L -momentov a ktorá sa aplikuje v špecifických geograficko-klimatických podmienkach Slovenska. Článok je užšie zameraný na prvý krok tzv. regionálneho L -momentového algoritmu (*Hosking, Wallis*, 1997), ktorým je vyčlenenie homogénnych regiónov pre k -denné úhrny zrážok ($k = 1$ až 5). Na formovanie homogénnych zoskupení klimatologických a zrážkomerných staníc sa použila objektívna aj subjektívna (logická) metodika. Napriek tomu, že odborná literatúra všeobecne uznáva a odporúča použiť objektívne postupy na vyčlenenie homogénnych regiónov, v štúdiu sme usúdili, že by sa vo frekvenčnej analýze úhrnov zrážok navzájom podobné stanice nemali vyčleňovať automaticky. Namiesto toho odporúčame, aby sa k tomuto účelu použila kombinácia objektívnych postupov, resp. úvah založených na fyzicko-geografických charakteristikách krajiny.

KLÚČOVÉ SLOVÁ: Slovensko, maximálne 1- až 5-denné úhrny zrážok, kalendárny rok, teplý/chladný polrok, regionálna frekvenčná analýza, homogénne regióny, L -momenty, zhluková analýza.

1. Introduction

Frequency analysis, which is aimed at estimating the design values of extreme hydrological or climatological events (extraordinary floods or precipitation amounts), has intensively been discussed in the environmental sciences for several decades. Statisticians, hydrologists, climatologists and engineers keep developing sophisticated methods

(mostly statistical models) for providing estimates of T -year values (or design values) of extreme hydro-climatological events in order to evaluate the hazard of such natural phenomena and reduce their possibly disastrous impact on society. After catastrophic floods in Central Europe, including Slovakia, in July 1997, June 1999 and August 2002, which caused considerable damages (*Lešková*, 2002; *Mikuličková*, *Babiaková*, 2003;

Pekárová et al., 2005; Ulbrich et al., 2003a,b; Kundzewicz et al., 2005), the need to focus on precipitation and/or flood risk assessment in this region has been stressed (Hlavčová et al., 2008). The enhanced frequency and/or severity of extreme events, which are likely to occur in the future due to global climate change (e.g. Christensen, Christensen, 2004) also underpins the necessity to conduct such studies.

The research presented in the paper was motivated by the fact that the last complex nation-wide frequency studies of heavy precipitation in Slovakia were presented in the 1970s and mid-1980s (Šamaj, Valovič, 1978; Šamaj et al., 1985). These studies relied on at-site frequency analysis. Since then, much broader and reliable records of daily precipitation amounts have become available and accessible in the digital database of the Slovak Hydrometeorological Institute (SHMI). In the last couple of years, several studies of precipitation frequency studies applying new regional statistical methods (Stehlová et al., 2001, 2002; Jurčová et al., 2002; Parajka et al., 2002; Kohnová, Parajka, 2002; Kohnová et al., 2004, 2005; Gaál et al., 2008a) have been published, mostly focusing on larger catchments in Slovakia. Regional frequency analysis (RFA) and progressive statistical methods have mostly been used in connection with floods in Slovakia, see e.g. Kohnová, Podolinská (2004), Kriegerová, Kohnová (2005), Bača, Bačová Mitková (2007), Solín (2008). This paper further explores approaches presented in the dissertation thesis of Gaál (2006) and is aimed at a complex precipitation frequency analysis focusing on Slovakia as a whole.

Cluster analysis is a common method of identifying homogeneous regions for precipitation frequency analysis (e.g. Guttman, 1993; Gong, Richman, 1995). Nevertheless, although this mathematical tool is an objective method of pooling, the selection of various methods/options of clustering and, more importantly, the decision about the acceptability of individual results is always influenced by subjective decisions. Consequently, a fundamental question emerges: why not also delineate regions according to fully subjective considerations? A subjective, process-based pooling may take advantage of the knowledge and long-term experience of climatologists about the precipitation regime of a given area. The current paper is aimed at discussing these two, basically different, approaches to delineating homogeneous regions and, at the end, tries to merge their results

by keeping the beneficial aspects of both methods in the region of the Western Carpathians.

The paper is structured in the following manner: Sect. 2 is devoted to the presentation of the methods applied herein, namely, the regional L -moments algorithm and two tests of regional homogeneity. After a short description of the selected stations and their data in the next section, specific indices of the precipitation regime of Slovakia, which are used in the regional analysis, are described. Sect. 4, which is divided into four subsections, presents and analyzes the different approaches to the delineation of homogeneous regions for a regional frequency analysis of 1-to-5 day heavy precipitation amounts in Slovakia. Finally, Sect. 5 summarizes and discusses the results and outlines the main directions for further extension of the work.

2. Methods

2.1 The regional concept in frequency analysis

At-site (single-site) frequency analysis has been the usual statistical technique for assessing probabilities of extraordinary events in the past. These methods are widely reviewed in the literature (e.g. Stedinger et al., 1993). However, data records in meteorology or hydrology are usually shorter (about several decades) than the desired return period of a really rare event. In climatology, a target return period of 100 years is commonly used; however, often hundreds of years are required in engineering design (e.g. in dam safety applications or spillway design; Stedinger et al., 1993; Hosking, Wallis, 1997). As a consequence, the idea of a RFA that substitutes the temporal dimension of the problem with a spatial one has appeared as a rational alternative to an at-site analysis (Dalrymple, 1960). In a spatial extension of the data sample one needs to pool sites with similar statistical properties of their at-site probability distributions (homogeneous region) to conduct a multi-site analysis. In other words, one of the basic assumptions of a regional approach is that the frequency distributions at different sites are identical apart from a scale factor:

$$Q_i^T = \mu_i q^T, \quad i = 1, \dots, N, \quad (1)$$

where i is the site index, N – the total number of sites in the region, Q_i^T – the T -year quantile (i.e. design value corresponding to the return period T)

at the i -th site, q^T – the regional (dimensionless) T -year quantile and μ_i is the at-site scale factor. The scale factor is also called an index flood (in hydrology) or index storm (in climatology), and it is usually estimated by the at-site sample mean of the selected variable (e.g. *Hosking, Wallis, 1997*).

A considerable impetus to RFA was given by introducing the use of L -moments (*Hosking, 1990*). L -moments are statistical characteristics analogous to conventional (product) moments – they describe the scale and shape properties of the probability distributions. The prefix L denotes that they are computed from linear combinations of the order statistic. For definition formulae and a detailed description of the L -moments, see e.g. *Hosking (1990)*. The superiority of L -moments over conventional moments mostly lies in their lower degree of sensitivity to the presence of outliers, less biased estimation and ability to describe a wider range of distributions (e.g. *Vogel, Fenessey, 1993; Sankarasubramanian, Srinivasan, 1999; Peel et al., 2001*). Simulation experiments focused on extreme temperatures in the Czech Republic (*Kysely, 2005*) also confirmed favourable properties of the L -moment based parameter estimation.

Hosking, Wallis (1997) combined the index flood method and L -moments-based frequency analysis into a widely used guideline to RFA, which is called the *regional L-moment algorithm*. There are examples of the successful adoption of modifications of this method in various regions of the world, e.g. in Canada (*Pilon et al., 1991; Adamowski et al., 1996; Alila, 1999*), USA (*Schaefer, 1990; Guttman, 1993; Guttman et al., 1993; Sveinsson et al., 2002*), Great Britain (*Fowler, Kilsby, 2003*), South Africa (*Smithers, Schulze, 2001*), Belgium (*Gellens, 2002*), Italy (*Boni et al., 2006*) and the Czech Republic (*Kysely, Picek, 2007*).

2.2 Regional L-moments algorithm

In this study, we decided to follow the regional L -moments algorithm as proposed by *Hosking, Wallis (1997)*. The methodology consists of four major steps:

1. Screening of the data;
2. Identification of homogeneous regions;
3. Choice of a regional frequency distribution;
4. Estimation of the parameters of the regional frequency distribution, at-site quantiles and their confidence intervals.

The data that appear in the process of delineating the regions should be separated into two distinct categories: site characteristics and site statistics (*Hosking, Wallis, 1997*). Site characteristics are either quantities that are known even before any measurement or observation is obtained at the site (for example, the location, the elevation or other physiographic properties associated with the site) or quantities that are supposed to be closely related to the long-time climate of the site (such as the mean annual precipitation). On the other hand, site statistics are simply the measurements or any results of the statistical processing of the data observed at the site. The basic difference between these quantities lies in their role in the grouping process. It is strongly recommended to base the formation of regions on site characteristics and to use the at-site statistics in subsequent testing of the homogeneity of the proposed set of regions.

The identification of homogeneous groups of sites (*regions* or *pooling groups*, depending on whether or not sites form geographically contiguous units) is usually the most complex step of a RFA. The initial formation of the groups is usually done by means of objective mathematical methods (mostly cluster analysis); however, most of these also require some subjective decisions or subsequent refinements, e.g. to preselect the number of clusters, to decide to move a site from one region to another, to subdivide a region, to merge one or more regions into a larger region, or to add or delete a site from the data set, etc. Moreover, at the end, it is the researcher who will decide which results of the cluster analysis will be considered acceptable. This choice should be supported by testing the homogeneity of the regions (*Hosking, Wallis, 1997*) and can be supplemented by subjective judgments, e.g. by analyzing which arrangement of regions is consistent with the common physical, climatological, hydrological, etc., knowledge base of the analyzed area.

Therefore it is usual, once the composition of the regions is finished, to check their regional homogeneity, i.e. to statistically test whether the at-site distribution functions are acceptably similar to each other within a given region. For this purpose, regional homogeneity tests have been defined by different authors (cf. *Viglione et al., 2007; Castellarin et al., 2008*). For testing the homogeneity of the proposed regions in a RFA of heavy precipitation amounts in Slovakia, we have chosen the H -test from *Hosking, Wallis (1997)* and the $X10$ -test from *Lu, Stedinger (1992)*. These tests, both taking

the advantage of L -moments, are considered the two most powerful regional homogeneity tests (Fill, Stedinger, 1995). Viglione et al. (2007) indirectly confirmed our selection stating that ‘ L -moments based tests are more powerful when the samples are slightly skewed’, which is the case of the data sets of precipitation in Slovakia. The mathematical description of the selected homogeneity tests is given in the Appendix.

2.3 Cluster analysis

Cluster analysis is an exploratory mathematical tool which classifies objects into groups (clusters) so that objects from the same cluster are more similar to each other than objects from different clusters. Several monographs or chapters of handbooks are devoted to various aspects of cluster analysis and its applications in the environmental sciences (e.g. Wilks, 2005; Xu, Wunsch, 2008). In this paper, we only deal with selected options of cluster analysis that are relevant to the current study. These options are as follows:

- *Transformation of input parameters.* Since the magnitudes of the input variables are different, they have to be rescaled before entering the cluster analysis in order to have the final results influenced in an equal way. Usually, two methods of transformation are used. The first one is standardization, which results in a variable with a zero mean and a unit variance:

$$x \rightarrow (x - \bar{x}) / \sigma_x, \quad (2)$$

where x denotes a variable of an arbitrary magnitude and range, and $\bar{x}$ (σ_x) – its mean (standard deviation). The second method of transformation is by means of extremes that result in variables in a range $\langle 0; 1 \rangle$:

$$x \rightarrow (x - x_{\min}) / (x_{\max} - x_{\min}), \quad (3)$$

where $x_{\max}$ ($x_{\min}$) is the maximum (minimum) value of the variable x . A drawback of transformations like these, however, is that ‘information on the variability of a particular measure is lost’ (Nathan and McMahon, 1990).

- *Distance metric.* We considered only the Euclidean distance metric, the most intuitive and most frequently used one in cluster analysis (e.g. Wilks, 2005):

$$D_{ij} = \sqrt{\sum_{k=1}^K (x_{ik} - x_{jk})^2}, \quad i, j = 1, \dots, N, \quad (4)$$

where D_{ij} is the distance between the sites i and j , K – the number of site attributes, x_{ik} – the k -th attribute at site i , and N is the total number of sites.

- *Linkage methods.* There are several different alternative linkage techniques in cluster analysis (Wilks, 2005); their advantages and/or drawbacks are discussed in Gong, Richman (1995). In a RFA, it is recommended to use Ward’s method, which has a tendency to form clusters containing similar numbers of sites, or average-link clustering, which tends to form clusters with an equal within-cluster variance of site attributes (Hosking, Wallis, 1997). During preliminary experiments with cluster analysis, the complete linkage (furthest neighbour) method also seemed to produce realistic results, so we took this linkage method into consideration as well.

3. Data

3.1 Selection of sites

Daily precipitation series from 56 climatological stations from the whole area of Slovakia were selected (Fig. 1) from the digital database of SHMI. The selected sites may be divided into two groups. The first one (29 stations) has records of daily precipitation amounts with no interruption in the period from 1961 (in some cases, even since 1951) to 2003 (when the data was prepared). The second group consists of 27 sites that have minor gaps in their daily rainfall records (breaks of up to several months). It was necessary to use sites with (slightly) incomplete records to ensure a more or less uniform spatial distribution of sites throughout the country (Fig. 1).

The basic data set at the 56 sites selected consists of 2464 station-years. The longest (shortest) observation records are 53 (35) years long. The most common value of the length of the observation records is 43 years (15 occurrences).

For the regional analysis, k -day precipitation amounts were computed as the sums of the daily precipitation amounts for k consecutive days ($k = 1, 2, 3, 4$ or 5). Maximum k -day precipitation amounts were determined for 3 periods: the calen-

Fig. 1. Fifty six selected climatological stations in Slovakia. 29 sites with at least 43 complete years of the observation of daily precipitation amounts are depicted with solid circles. The other 27 sites with small gaps in their daily records of precipitation amounts (35–42 complete years of observation) are symbolized with blank squares.

Obr. 1. Päťdesiatšesť vybraných klimatologických staníc na Slovensku. Plný kruh označuje 29 staníc s kompletným, prinajmenšom 43-ročným radom pozorovania denných úhrnov zrážok. Ďalších 27 staníc, ktoré majú minimálne prerušenia vo svojich radoch (35–42 kompletných rokov pozorovania) je označených prázdny štvorčekom.

dar year, the warm season (April–September) and the cold season (October–March). Consequently, 15 data sets were prepared for each station as a combination of 5 durations and 3 periods of the year.

The data underwent standard quality checking for gross errors using the discordancy test of *Hosking, Wallis* (1997) and temporal homogeneity of the observation series using the standard normal homogeneity test (*Alexandersson*, 1986) and the non-parametric method of trend detection of *Wald, Wolfowitz* (1943). Except for a single case (the Lom nad Rimavicou station), no significant inhomogeneities (shift or trend changes) were detected; therefore, the data set is regarded as sufficiently homogeneous for further climatological analyses.

3.2 Specific indices of the precipitation regime of Slovakia used in the regionalization

Due to Slovakia's geographical position in Europe, its precipitation regime is influenced by different climatic factors; the most dominant ones are the effect of the Mediterranean, the western circulation and that of the continental climate (*Lapin, Tomlain*, 2001). The influence of the Medi-

terranean area has a pronounced role in the intra-annual variability of monthly precipitation amounts, mainly in the southern parts of the country. In general, the annual cycle of precipitation has a typical continental regime with a sharp maximum in June (on average 95 mm, especially in the south) and a sharp minimum in February (on average 43 mm). However, at the majority of stations of South Slovakia, (i) the annual maximum is usually shifted towards spring (June or occasionally May), and (ii) due to cyclones moving from the area of the Ligurian Sea (the Mediterranean area), a secondary autumn maximum (in October/November) appears. *Lapin's index of the Mediterranean effect* L_M (*Lapin, Melo*, 2005) is a quantitative characteristic of the magnitude of this influence. It is defined using 3 ratios of certain mean monthly precipitation amounts:

$$L_M = \frac{R_{\text{Max}}}{R_{\text{VII}}} + \frac{R_V}{R_{\text{VII}}} + \frac{R_{\text{Max2}}}{R_{\text{Min2}}} - 2.5, \quad (5)$$

where the indices denote May (V), July (VII) and months with the maximum (Max), secondary maximum (Max2) and secondary minimum (Min2)

precipitation amounts in the annual cycle. In the case of a smooth annual course of monthly precipitation amounts, the secondary extremes are missing; therefore, the third ratio of Eq. (5) should be taken as 1.0. The number 2.5 in Eq. (2) is a correction factor.

Lapin's index of continentality L_C (Lapin, Melo, 2005) is another descriptor of the intra-annual variability of precipitation in Slovakia. In general, a continental regime of precipitation may be characterized by enhanced precipitation amounts in summer and, on the other hand, by relatively low amounts in winter. Usually, the range or proportion between these two extremes forms the basis of the measure of continentality. *Lapin's index of continentality* is defined using 2 ratios of certain mean monthly precipitation amounts:

$$L_C = \frac{R_{\text{Max}}}{R_{\text{Min}}} + \frac{R_{\text{Max}}}{R_{\text{Max}2}} - 2.0, \quad (6)$$

where the indices correspond to the ones in Eq. (5) and, similarly, the number 2.0 in Eq. (6) is a correction factor. For a more detailed description and interpretation of the indices L_M and L_C , refer to Gaál (2005).

3.3 Further physiographic site characteristics used in the regionalization

Besides L_M and L_C (Sect. 3.2), the following 7 variables, which describe the long-term precipitation regime of Slovakia, were prepared for each selected site:

- φ – latitude,
- λ – longitude,
- h – elevation above sea level,
- R_{MAP} – mean annual precipitation,
- POT_0 – mean annual number of wet days (daily precipitation amounts ≥ 0.0 mm),
- POT_1 – mean annual number of days with daily precipitation amounts ≥ 1.0 mm,
- $R_{W/C}$ – ratio of the mean precipitation amounts for the warm and cold seasons.

The first 3 variables (φ , λ , h) are geographical characteristics; the next 3 variables (R_{MAP} , POT_0 , POT_1) are descriptors of the long-term precipitation regime. The last variable $R_{W/C}$ along with

L_M and L_C are indicators of the intra-annual variability of precipitation amounts.

4. Results

The following section, which presents the results of the study, is comprised of four main subsections: the first one takes a look at Slovakia as a single region and examines its spatial homogeneity; the second one describes the settings and the results of the cluster analysis as an objective method of pooling; the third one describes considerations of the process-based (subjective) pooling; and the last one gives an overview of the final delineation of the selected sites into 3 regions along with an analysis of the possible reasons for their inhomogeneities.

4.1 Homogeneity of all of Slovakia

In the first stage of our analysis, the homogeneity of the whole area of Slovakia was examined. This step is argued for as follows: if Slovakia were found to be a single homogeneous region, there would be no need to deal with a further division of the area. Therefore, the basic question is the following: *can we consider Slovakia as a one-and-only compact homogeneous region* regardless of the duration of precipitation and/or seasons analyzed?

The results of the homogeneity tests are summarized in Tab. 1. They suggest that *Slovakia could obviously be treated as a single homogeneous region* for annual maxima of k -day precipitation amounts. Nevertheless, according to the H -test, 4 possible heterogeneous data sets were found for the warm/cold seasons; 3 of them were marked as heterogeneous by the $XI0$ -test as well. Note that the H -test herein relies on the initial criteria (Hosking, Wallis, 1997) where possible heterogeneity is concluded in the case of $1 \leq H < 2$, and not on the updated criteria (Wallis et al., 2007), where the critical values are shifted, and possible heterogeneity is indicated if $2 \leq H < 3$.

The possible inhomogeneities detected can be explained by the general spatial and temporal properties of multi-day precipitation events. The heterogeneity in the case of the 2-day summer maxima arises from the possibility of intensive convective storms occurring in 2 consecutive days (or simply influencing data for 2-day maxima). Such heavy storms are less likely to occur for 3 or more consecutive days; that might be the reason for the homogeneity of the other several-day precipitation

Table 1. Results of the homogeneity tests of 15 precipitation data sets in a case when the whole area of Slovakia was treated as a single region (56 stations). Two homogeneity tests were used: the *H*-test (Hosking, Wallis, 1997) and the *X10*-test (Lu, Stedinger, 1992). The critical value of the *X10*-test for $N = 56$ is $\chi^2_{0.95, N-1} = \chi^2_{0.95, 55} = 73.31$. Figures marked in italic indicate possible heterogeneity (*H*-test) or heterogeneity (*X10*-test).

Tabuľka 1. Výsledky analýzy testov homogenity 15 dátových súborov pre špeciálny prípad, v ktorom sa územie SR považovalo za jediný, komplexný región (56 staníc). Použili sa dva testy homogenity: *H*-test (Hosking, Wallis, 1997) a *X10*-test (Lu, Stedinger, 1992). Kritická hodnota *X10*-testu pre $N = 56$ je $\chi^2_{0.95, N-1} = \chi^2_{0.95, 55} = 73,31$. Čísla označené kurzívou znamenajú možnú heterogenitu (*H*-test) alebo heterogenitu (*X10*-test).

'Slovakia as one region'	1-day prec.		2-day prec.		3-day prec.		4-day prec.		5-day prec.	
	<i>H</i>	<i>X10</i>	<i>H</i>	<i>X10</i>	<i>H</i>	<i>X10</i>	<i>H</i>	<i>X10</i>	<i>H</i>	<i>X10</i>
Calendar year	0.35	66.58	0.10	66.88	-0.47	63.34	-0.90	54.35	-0.08	64.53
Warm season	0.13	65.10	<i>1.13</i>	<i>75.51</i>	0.73	72.98	0.19	64.61	0.78	71.93
Cold season	-0.12	60.45	0.58	64.52	<i>1.47</i>	<i>74.96</i>	<i>1.29</i>	<i>71.18</i>	<i>1.83</i>	<i>74.97</i>

amounts in the warm season. On the other hand, several-day precipitation amounts in the cold season that are mostly of a frontal origin may reach an extraordinary depth. Such events may modify the properties of the at-site frequency distribution fundamentally, so a given (outlying) site's sample may deteriorate the homogeneity of the whole region.

Another interesting aspect of the heterogeneity analysis presented in Tab. 1 is that none of the data sets may be considered as definitively heterogeneous ($H > 2.0$). One of the basic principles of the regional approach is that even though a region may be moderately heterogeneous, regional analysis may still yield much more accurate quantile estimates than at-site analysis (Hosking, Wallis, 1997). If we therefore strictly insisted on this principle, we might choose the simplest approach to the problem, i.e. to ignore each critical value from Tab. 1 and, referring to the *H*-test only, treat the area of Slovakia as a one-and-only homogeneous region regardless of the duration of heavy precipitation and/or its seasonal occurrence. Such a view, although it fits in the statistical concepts of RFA, does not seem acceptable from a climatological point of view. Long-term experience indicates that even though the area of Slovakia is relatively small, it is unreasonable to treat the country as a single region (Faško, Lapin, 1996). The precipitation regime of the country is influenced mostly by the effects of the Atlantic Ocean, the Mediterranean area and the continental inland, which are further differentiated due to the complex topography (Lapin, Tomlain, 2001). Therefore, several regions with different long-term precipitation regimes could be distinguished for a RFA. That is the main reason we also decided to delineate and test sub-regions for the regional precipitation frequency analysis in Slovakia.

4.2 Objective pooling: cluster analysis

Some precipitation frequency studies do not deal with the issue of clustering in detail; usually only the homogeneity of larger geographical or political units is checked (e.g. Adamowski et al., 1996; Pilon et al., 1991; Sveinsson et al., 2002; Gellens, 2002). The issues of clustering were discussed in the course of analyses that were set in rather complex geographic conditions, e.g. South Africa (Smithers, Schulze, 2001) or the Czech Republic (Kyseľý, Pícek, 2007).

One of the pitfalls of any objective method of regionalization is the selection of the variables that should serve to evaluate the similarity of the analyzed objects. In principle, any set of variables is able to generate clusters (however, the physical, hydrological and/or climatological interpretation of the resulting groups of objects may be disputable), so it is necessary to select variables according to their relevance to the problem (Nathan, McMahon, 1990).

The more clustering variables are available, the better chance for a successful cluster analysis; however, in the case of a larger database of variables, the redundancy of information may negatively influence the results of the cluster analysis. Some of the variables may be closely correlated, as they may contain similar climatological information. The input variables should therefore be as independent as possible, since the technique of the clustering (particularly the Euclidean distance, Eq. (4)) is very sensitive to the redundancy of information encompassed in the variables (Guttman, 1993; Gong, Richman, 1995).

Since the number of selected variables here is relatively small, we did not attempt to apply sophisticated methods to find the minimum number of

variables with the maximum amount of independent information, such as a principal component analysis or factor analysis. We decided that for the purposes of the current study, an examination of cross-correlations among the selected 9 variables would be sufficient. Tab. 2 demonstrates that the majority of the selected variables are significantly correlated; the correlation matrix plot on Fig. 2 provides further details about the dependence of the variables. Based on the correlation structure of the variables, we made the following restrictions for the input variables of the cluster analysis:

- geographical co-ordinates (φ and λ) would not be omitted in any case, regardless of their relationship to the other variables, as we prefer geographically contiguous regions;
- characteristics R_{MAP} , POT_0 and POT_1 are strongly correlated with the elevation h , so in a given clustering procedure, one should not use more than one of these 4 variables;
- the same conclusion also holds true for the strongly correlated variables $R_{W/C}$ and L_C .

In spite of these restrictions, the selected variables offer a wide variety of combinations as to how to form the input parameter sets that enter the cluster analysis.

We have conducted a large number of cluster analysis experiments with various settings, i.e. with different combinations of input variable sets, transformation methods of the variables, and alternative linkage methods (cf. Sect. 2.3). Furthermore, we examined the results at various reasonable stages of clustering, usually from the 2-cluster stage up to the stage with 7–8 clusters. The homogeneity of each individual region was examined as well for all the durations and seasons of the year. Our goal at this stage of the analysis was to arrive at a set of variables yielding a pooling scheme that would be consistent in the following senses across all the analyzed precipitation durations:

- comparable number of sites in a region;
- groups of sites not scattered in the geographical space, i.e. groups that can be regarded as contiguous geographical regions;
- agreement with climatological theories about the precipitation regime (Lapin et al., 1995; Lapin, Faško, 1996, 1998);
- one pooling scheme for all the durations;
- homogeneity in all the groups, durations and seasons;
- no subjective intervention into the composition of clusters and/or no need to shift sites from region to region.

T a b l e 2. Pearson correlation coefficients among the selected variables. Figures marked in italic are significant at the confidence level $p < 0.05$. φ – latitude, λ – longitude, h – elevation, R_{MAP} – mean annual precipitation, POT_0 – mean annual number of wet days (daily precipitation amounts ≥ 0.0 mm), POT_1 – mean annual number of days with daily precipitation amounts ≥ 1.0 mm, $R_{W/C}$ – the ratio of the mean precipitation amounts for the warm and cold seasons, L_M – Lapin's index of the Mediterranean effect, L_C – Lapin's index of continentality.

T a b u l k a 2. Pearsonove korelačné koeficienty medzi vybranými charakteristikami. Kurzíva označuje korelácie signifikantné na úrovni $p < 0.05$. φ – zemepisná šírka, λ – zemepisná dĺžka, h – nadmorská výška, R_{MAP} – priemerný ročný úhrn zrážok, POT_0 – priemerný počet dní v roku so zrážkami (tj. denný úhrn zrážok $R \geq 0,0$ mm), POT_1 – priemerný počet dní v roku s denným úhrnom zrážok nad prahovou hodnotou $R \geq 1,0$ mm, $R_{W/C}$ – podiel dlhodobého úhrnu zrážok za teplý a chladný polrok, L_M – Lapinov index vplyvu Stredomoria, L_C – Lapinov index kontinentality.

	φ	λ	h	R_{MAP}	POT_0	POT_1	$R_{W/C}$	L_M	L_C
φ	–	<i>0.453</i>	<i>0.457</i>	<i>0.602</i>	<i>0.718</i>	<i>0.719</i>	<i>0.468</i>	<i>–0.691</i>	<i>0.519</i>
λ	<i>0.453</i>	–	0.143	0.190	0.204	0.205	<i>0.584</i>	<i>–0.561</i>	<i>0.619</i>
h	<i>0.457</i>	0.143	–	<i>0.881</i>	<i>0.762</i>	<i>0.856</i>	0.035	–0.199	0.082
R_{MAP}	<i>0.602</i>	0.190	<i>0.881</i>	–	<i>0.845</i>	<i>0.953</i>	0.041	<i>–0.332</i>	0.094
POT_0	<i>0.718</i>	0.204	<i>0.762</i>	<i>0.845</i>	–	<i>0.921</i>	0.121	<i>–0.526</i>	0.165
POT_1	<i>0.719</i>	0.205	<i>0.856</i>	<i>0.953</i>	<i>0.921</i>	–	0.043	<i>–0.428</i>	0.093
$R_{W/C}$	<i>0.468</i>	<i>0.584</i>	0.035	0.041	0.121	0.043	–	<i>–0.556</i>	<i>0.988</i>
L_M	<i>–0.691</i>	<i>–0.561</i>	–0.199	<i>–0.332</i>	<i>–0.526</i>	<i>–0.428</i>	<i>–0.556</i>	–	<i>–0.598</i>
L_C	<i>0.519</i>	<i>0.619</i>	0.082	0.094	0.165	0.093	<i>0.988</i>	<i>–0.598</i>	–

Fig. 2. Correlation matrix plot of 9 selected variables for cluster analysis. Lat. – latitude, Lon. – longitude, H – elevation, R_{MAP} – mean annual precipitation, POT_0 – mean annual number of wet days (daily precipitation amounts ≥ 0.0 mm), POT_1 – mean annual number of days with daily precipitation amounts ≥ 1.0 mm, $R_{W/C}$ – the ratio of the mean precipitation amounts for the warm and cold seasons, L_M – Lapin's index of the Mediterranean effect, L_C – Lapin's index of continentality.

Obr. 2. Maticové zobrazenie 9 charakteristík vybraných do zhlukovej analýzy. Lat. – zemepisná šírka, Lon. – zemepisná dĺžka, H – nadmorská výška, R_{MAP} – priemerný ročný úhrn zrážok, POT_0 – priemerný počet dní v roku so zrážkami (t. j. denný úhrn zrážok $\geq 0,0$ mm), POT_1 – priemerný počet dní v roku s denným úhrnom zrážok nad prahovou hodnotou $\geq 1,0$ mm, $R_{W/C}$ – podiel dlhodobého úhrnu zrážok za teplý a chladný polrok, L_M – Lapinov index vplyvu Stredomoria, L_C – Lapinov index kontinentality.

Even though a large number of station groupings were created and tested, none of them met the relatively strict criteria listed above perfectly. In the end, the 5-cluster pooling that mostly satisfied the conditions described above (cluster analysis based on 5 input variables [φ , λ , H , L_M , $R_{W/C}$], transformation of variables using the extremes, and Ward's linkage method of clustering) was selected. The fact that the majority of groupings failed to pass our conditions was the basis for our conclusion that hybrid grouping based on combined subjective and objective selection of sites is to be preferred for such an analysis.

The particular results can be described as follows: Fig. 3 presents the composition and the loca-

tion of each cluster (region), and Tab. 3 gives a brief overview of the most relevant characteristics of the regions.

Region CA#1 (= region #1 from cluster analysis, CA) consists of 19 sites in the lowlands and moderate altitudes of the southern and central parts of Western Slovakia (up to about 600 m a.s.l.). The region can be characterized by relatively low mean annual precipitation ($R_{MAP} \approx 550\text{--}650$ mm), by the rather considerable influence of the Mediterranean area ($L_M \approx 0.6\text{--}0.8$) and by the lowest value of the variable $R_{W/C}$ (1.4) among the regions (Tab. 3).

Region CA#1 is homogeneous regarding both tests and all the precipitation data sets (Tab. 4).

Fig. 3. Delineation of 56 stations into 5 regions based on a cluster analysis (objective method of pooling).

Obr. 3. Rozdelenie 56 staníc do 5 regiónov na základe zhľukovej analýzy (objektívna metóda zgrupovania).

Table 3. Overview of the physical-climatological characteristics of 5 regions delineated objectively as a result of a cluster analysis (CA). IQR stands for inter-quartile range.

Tabuľka 3. Prehľad fyzikálno-klimatologických charakteristík 5 regiónov vyčlenených na základe objektívnych kritérií zhľukovej analýzy (CA). Skratka IQR znamená interkvartilové rozpätie.

	Region CA#1	Region CA#2	Region CA#3	Region CA#4	Region CA#5
Number of stations	19	10	10	7	10
Range of H [m]	111–605	365–2635	214–1018	380–1030	100–305
Average of H [m]	249	1160	448	633	191
IQR of R_{MAP} [mm]	553–665	817–1103	595–773	640–775	609–653
Median of R_{MAP} [mm]	602	964	641	739	633
IQR of L_M	0.63–0.84	0.30–0.41	0.75–0.80	0.28–0.39	0.23–0.36
Median of L_M	0.71	0.40	0.79	0.32	0.27
IQR of $R_{W/C}$	1.33–1.46	1.46–1.74	1.56–1.69	2.15–2.29	1.64–1.95
Median of $R_{W/C}$	1.38	1.61	1.59	2.19	1.81

Region CA#2, which is located in the northwestern parts of Slovakia, is made up of 10 stations from the Orava, Kysuca and Liptov regions, and partly from the High and Low Tatras. The cluster analysis separated these sites predominantly due to the highest altitudes of the stations and the relatively low values of the index L_M (Tab. 3).

The homogeneity of *Region CA#2* is, however, questionable, mostly in the case of multi-day precipitation extremes (3 to 5 days) in the cold season (Tab. 4). This is mostly due to the fact that the cluster analysis pooled sites with markedly different precipitation regimes. In the northwestern corner of

Slovakia, the effect of the Atlantic Ocean is noticeable, which results in enhanced precipitation in the cold season (Lapin et al., 1995; Lapin, Faško, 1996, 1998; Lapin, Tomlain, 2001). The reason for the heterogeneity is the fact that the influence of the Atlantic becomes considerably weaker in a relatively short distance, thus it can be observed to a lesser degree near the area of the Tatras.

For annual 1-day maxima (Tab. 4), the possible heterogeneity of *Region CA#2* stems from other causes. The high values of the H and X_{10} statistics occurred due to the relatively skewed 1-day precipitation data at the Podbanské climatological station

Table 4. Statistics of the homogeneity tests of 15 precipitation data sets for 5 regions delineated objectively by means of a cluster analysis (CA). Figures marked in italic indicate possible heterogeneity (*H*-test) or heterogeneity (*X10*-test); an asterisk (*) indicates definitive heterogeneity in the case of the *H*-test.

Tabuľka 4. Výsledky analýzy testov homogenity 15 dátových súborov pre 5 regiónov vyčlenených na základe objektívnych kritérií zhlukovej analýzy (CA). Čísla označené kurzívou znamenajú možnú heterogenitu (*H*-test) alebo heterogenitu (*X10*-test); hviezdička (*) označuje definitívnu heterogenitu v prípade *H*-testu.

Objective pooling: cluster analysis	1-day prec.		2-day prec.		3-day prec.		4-day prec.		5-day prec.	
	<i>H</i>	<i>X10</i>	<i>H</i>	<i>X10</i>	<i>H</i>	<i>X10</i>	<i>H</i>	<i>X10</i>	<i>H</i>	<i>X10</i>
<i>Region CA#1 – 19 sites, $\chi^2_{0.95,18} = 28.87$</i>										
Calendar year	–0.58	15.68	–0.11	19.63	–1.19	13.25	–1.16	12.24	–1.21	12.65
Warm season	–0.34	17.41	0.31	21.56	–0.61	14.79	–0.38	14.55	–1.13	10.47
Cold season	–0.42	14.18	–0.77	13.83	–0.43	16.06	–0.36	15.82	–0.32	15.69
<i>Region CA#2 – 10 sites, $\chi^2_{0.95,9} = 16.92$</i>										
Calendar year	<i>1.98</i>	<i>20.23</i>	0.24	12.77	0.26	13.75	–0.17	10.26	–0.44	8.44
Warm season	<i>1.37</i>	15.79	0.20	10.83	0.54	14.17	0.32	11.49	–0.25	9.44
Cold season	0.45	16.33	0.80	15.05	<i>1.40</i>	<i>17.80</i>	<i>1.71</i>	<i>18.35</i>	<i>2.42*</i>	<i>20.61</i>
<i>Region CA#3 – 10 sites, $\chi^2_{0.95,9} = 16.92$</i>										
Calendar year	–1.27	4.70	–1.25	5.44	–1.72	3.55	–0.22	7.43	0.78	11.69
Warm season	–1.20	5.29	–0.64	7.33	–1.24	4.97	–0.80	6.76	–0.65	6.47
Cold season	–0.70	5.24	–1.64	2.16	–1.72	2.43	–1.55	3.27	–1.51	2.93
<i>Region CA#4 – 7 sites, $\chi^2_{0.95,6} = 12.59$</i>										
Calendar year	0.02	6.82	–0.95	3.14	–1.02	2.68	–0.55	3.77	0.03	5.13
Warm season	0.09	6.95	–1.01	2.62	–1.67	1.47	–1.55	1.77	–0.61	3.61
Cold season	–1.62	1.82	–1.04	3.26	–0.39	6.31	–0.20	6.03	–0.50	5.64
<i>Region CA#5 – 10 sites, $\chi^2_{0.95,9} = 16.92$</i>										
Calendar year	0.19	9.35	<i>1.03</i>	13.51	0.53	13.33	–0.46	8.21	–0.25	8.99
Warm season	–0.44	6.26	<i>1.38</i>	14.68	0.78	14.24	0.36	12.20	0.43	11.59
Cold season	<i>1.27</i>	13.22	0.11	8.67	0.75	9.87	0.09	7.32	0.13	8.03

($H = 972$ m a.s.l.). The station has a specific location, which also influences the unique precipitation regime. This is underpinned by 4 extraordinary 1-day events above the 100 mm threshold (126.9; 120.5; 112.2 and 105.4 mm), while none of the other annual maxima of 1-day amounts reached 70 mm.

Region CA#3 comprises of 10 sites in the southern parts of Central Slovakia, which predominantly lie in moderate altitudes. The mean annual precipitation of about 600–700 mm is highly influenced by the circulation of air masses from the area of the Mediterranean Sea. In accordance with this fact, the index L_M reaches the highest values of Slovakia, i.e. generally about 0.8 and at some sites close to 1.0 (Tab. 3). *Region CA#3* is homogeneous in each combination of the respective durations and periods of precipitation (Tab. 4).

Region CA#4 consists of 7 sites located in the northwestern parts of Eastern Slovakia. The sites are rather scattered at various altitudes, but the mean altitude is the second largest one among the 5

regions (Tab. 3). The uniqueness of this group of sites lies in the very low values of L_M (i.e. a small influence of the cyclones of a Mediterranean origin) and the extraordinarily high ratios of the warm/cold season precipitation amounts (i.e. a noticeable continental regime of precipitation in the annual course). *Region CA#4* is homogeneous according to both homogeneity tests used in the analysis (Tab. 4).

Region CA#5, which is comprised of 10 sites, is located in the easternmost parts of the country in the lowlands (up to 300 m a.s.l.). The common features of the sites in this region are the very high values of $R_{W/C}$ and the lowest values of the index L_M , which are accompanied by a relatively stationary mean annual precipitation of about 630 mm (Tab. 3).

The homogeneity of *Region CA#5* is acceptable except for two data sets (Tab. 4). A possible heterogeneity in the case of 1-day maxima in the cold season ($H = 1.27$) originates at the Bardejov, Čak-

lov and Stropkov-Tisinec stations, where extraordinary events were registered on 21/10/1974, 10/10/1980 and 02/10/1999, respectively. A higher H -statistic for the 2-day maxima in the warm season ($H = 1.38$) was influenced by the Somotor station having 2 outstanding precipitation events. The first one occurred on 09-10/08/1970 with a 2-day amount of 127.1 mm, while the second one was observed on 21-22/07/1980 with a 2-day amount of 94.8 mm. These extremes are ranked as the second and the fifth largest ones among the total number of the 444 values of the 2-day maxima at all the sites of Region CA#5.

4.3 Subjective pooling

A logical (process-based but subjective) delineation of sites was proposed by Faško (2006). Taking into consideration the rich topography of Slovakia, the effects of different circulation patterns and other

mechanisms producing heavy precipitation, 4 regions of extreme precipitation were identified (Fig. 4). The homogeneity of these regions was tested, again, for all the durations of precipitation and seasons considered, using both the Hosking and Wallis method and the method of Lu and Stedinger. The homogeneity measures of the regions are summarized in Tab. 5. In the following parts of the paper, a brief description of the regions will be given, focusing mainly on possible sources of spatial inhomogeneities.

Region PB#1 (= region #1 from the process-based pooling, PB) is located in the western parts of Slovakia. It lies completely along the borders with the Czech Republic and includes 11 stations. The region is homogeneous, since all the test statistics H and X_{10} , respectively, lie far below the critical values (Tab. 5).

Fig. 4. Delineation of 56 stations into 4 regions based on an expert analysis (subjective, process-based pooling).

Obr. 4. Rozdelenie 56 staníc do 4 regiónov na základe expertnej analýzy (subjektívna metóda zgrupovania na základe prevládajúcich atmosférických procesov).

Region PB#2 is the largest among the regions from the process-based pooling; it is comprised of 22 stations in the southern parts of Western and Central Slovakia, respectively. The region is homogeneous with the exception of a single data set: in the case of 2-day maxima in the warm season, both tests indicated heterogeneity (Tab. 5). Detailed

scrutiny of the data set revealed that there are 3 stations that observed extraordinary 2-day precipitation amounts above the 100 mm threshold: Kráľová pri Senci (159.7 mm), Bratislava-Koliba (146.1 mm) and Bratislava-Airport (103.1 and 100.4 mm). It is very likely that these events are the reason for the inhomogeneity of the region, since at

Table 5. Statistics of the homogeneity tests of 15 precipitation data sets for 4 regions delineated subjectively as a result of process-based pooling (PB). Figures marked in italic indicate possible heterogeneity (*H*-test) or heterogeneity (*X10*-test).

Tabuľka 5. Výsledky analýzy testov homogenity 15 dátových súborov pre 4 regióny vyčlenených na základe subjektívnych kritérií (zgrupovanie na základe prevládajúcich atmosférických procesov, PB). Čísla označené kurzívou znamenajú možnú heterogenitu (*H*-test) alebo heterogenitu (*X10*-test).

Subjective pooling: expert analysis	1-day prec.		2-day prec.		3-day prec.		4-day prec.		5-day prec.	
	<i>H</i>	<i>X10</i>	<i>H</i>	<i>X10</i>	<i>H</i>	<i>X10</i>	<i>H</i>	<i>X10</i>	<i>H</i>	<i>X10</i>
<i>Region PB#1 – 11 sites, $\chi^2_{0.95,10} = 18.31$</i>										
Calendar year	0.37	12.24	–0.65	8.26	–0.87	8.69	–1.12	6.55	–1.15	6.74
Warm season	0.28	12.03	–1.00	6.12	–0.86	7.27	–0.70	6.84	–0.31	9.64
Cold season	–0.60	9.33	–0.15	12.35	–0.08	10.52	0.42	11.78	0.89	12.85
<i>Region PB#2 – 22 sites, $\chi^2_{0.95,21} = 32.67$</i>										
Calendar year	–0.62	18.93	0.11	25.66	–0.85	17.19	–0.58	17.57	0.12	22.70
Warm season	–0.12	26.21	1.83	39.54	0.94	30.20	0.48	26.71	0.25	23.13
Cold season	–0.40	16.33	0.29	17.63	–0.12	19.02	–0.55	17.80	–0.29	17.20
<i>Region PB#3 – 10 sites, $\chi^2_{0.95,9} = 16.92$</i>										
Calendar year	1.84	17.23	1.14	15.36	1.28	15.61	0.40	11.23	0.26	11.22
Warm season	1.38	13.30	0.93	13.07	1.11	14.32	0.26	9.91	–0.12	9.02
Cold season	–0.25	10.82	0.21	12.59	0.25	13.86	0.38	12.34	0.81	14.17
<i>Region PB#4 – 13 sites, $\chi^2_{0.95,12} = 21.03$</i>										
Calendar year	0.64	16.72	0.90	16.54	0.56	19.33	0.10	15.01	0.88	17.84
Warm season	–0.02	12.19	1.20	15.86	0.58	16.29	0.20	14.35	0.63	15.74
Cold season	0.18	13.59	–1.01	7.02	1.14	13.41	1.47	15.80	1.64	19.60

the other sites of the region, the absolute maxima of 2-day precipitation amounts do not reach 70–80 mm.

Region PB#3 consists of 10 stations, mostly located in the northern parts of Central Slovakia, especially in the area of the High Tatras and its environs. The homogeneity of the region is disputable in the case of the maxima of 1-to-3 day precipitation amounts in the warm season and calendar year, respectively (Tab. 5). The higher values of the *H* and *X10* statistics may be explained by the suspicious behaviour at the Podbanské station, previously discussed in the analysis of the homogeneity measures of *Region CA#2* (Sect. 4.2). In principle, if this site was excluded from *Region PB#3*, no data set would yield heterogeneity. In practice, however, no exclusion was considered. The possible moving of the Podbanské station to another region was also rejected, since the site would spoil the regional homogeneity in any other region.

Region PB#4 with its 13 sites stretches through East Slovakia, approximately from the High Tatras to the southeastern boundaries of the country. There are two stations from the broader neighbourhood of the High Tatras (Tatranská Lomnica and Poprad), assignment of which needs further explanation. These stations are classified to *Region PB#4* (and

not to *Region PB#3* that encompasses the most of the Tatras region) due to (i) the shadowing effect of range of the High Tatras, which causes noticeable decrease in the rainfall amounts (snow cover depth) even in short distances eastward from the Štrbské Pleso station and (ii) the remarkable effect of the continentality in the eastern parts of the Poprad hollow and the Spiš basin (cf. *Region CA#4* in Sect. 4.2, and Gaál et al., 2008b).

The test statistics in Tab. 5 yield the possible heterogeneity of the region mostly in the case of multi-day precipitation amounts in the cold season. The *L*-moments diagrams (Fig. 5) reveal that the region is made up of two distinct groups of sites. What is, however, more interesting is that these groups form two distinct geographical regions. The first group is made up of 7 sites from the southern parts (the lowlands) of *Region PB#4*: 44 (Moldava nad Bodvou), 49 (Čaklov), 50 (Košice-Airport), 53 (Milhostov), 54 (Somotor), 55 (Michalovce) and 56 (Kamenica nad Cirochou). The other group consists of 6 sites from the northern parts of *Region PB#4*; their precipitation regime is considerably influenced by the more complex terrain of their surroundings: 35 (Poprad), 36 (Tatranská Lomnica), 42 (Švedlár), 45 (Spišské Vlachy), 48 (Bardejov) and 51 (Stropkov-Tisinec). According to the *L*-

moments diagrams in Fig. 5, a further subdivision of *Region PB#4* is advised.

As a result we can also conclude that this type of approach does not guarantee that the regions selected will pass ordinary spatial homogeneity tests.

This is a further argument in support of our conclusion that a combination of objective and subjective pooling should be attempted in the course of a regional extreme precipitation analysis.

Fig. 5. Selected *L*-moments diagrams for maxima of a) 3-day, b) 4-day, and c) 5-day precipitation amounts in the cold season in Region PB#4 from the process-based pooling. The numbers beside the squares denote the rank of individual stations in the list of all 56 stations according to their IDs (more detail in Sect. 4.3).

Obr. 5. Vybrané *L*-momentové diagramy a) pre 3-denné, b) pre 4-denné a c) pre 5-denné maximálne úhrny zrážok v chladnom polroku pre región PB#4 z expertnej analýzy. Čísla pri štvorčkoch označujú poradové číslo jednotlivých staníc v zozname 56 staníc, zoradených na základe rastúceho indikativu (ďalšie detaily v sekcii 4.3).

4.4 Hybrid delineation of regions

The final set of regions for a RFA of heavy precipitation amounts in Slovakia was created as a *compromise between the results of the cluster analysis* (objective method) *and the process-based delineation* (subjective method). On the one hand, there was an effort to preserve the integrity of the individual geomorphological units and to consider an objective measure of similarity; on the other hand, the main goal was to obtain test statistics as acceptable as possible. We tried to maintain the best features of both methods of pooling (Sects. 4.2 and 4.3), which resulted in 3 groups of sites. The composition and geographical locations of the regions are presented in Fig. 6, a short summary of the physical-climatological characteristics of the regions is given in Tab. 6, and their heterogeneity measures are shown in Tab. 7.

Region HA#1 (HA stands for “hybrid approach”): 11 sites in Western/Northwestern Slovakia, which cover the western-most areas of the Carpathians. The northernmost sites of **Region HA#1** are under the dominant effects of the Atlantic area in the winter (Lapin and Tomlain, 2001). **Region HA#1** is identical with **Region PB#1** (Sect. 4.3), since there was no reason to change the composition of **Region PB#1** due to its high degree of homogeneity

(Tab. 7) and its geographical compactness, respectively.

Region HA#2: 29 sites in South Slovakia, which includes mainly lowland sites and the Low Tatras; the sites of **Region HA#2** are under the dominant effects of the Mediterranean area in autumn. The basis of **Region HA#2** is formed by **Region PB#2**, which was extended by the southern cluster of sites from **Region PB#4** (Sect. 4.3, Fig. 5). In other words, **Region HA#2** consists of the majority of stations of **Region CA#1**, the highly homogeneous **Region CA#3** and a part of **Region CA#5** (Sect. 4.2). Questions concerning the homogeneity of **Region HA#2** arise only in the case of 2 and 3-day precipitation amounts in the warm season (Tab. 7); the possible inhomogeneities have already been discussed in previous sections (e.g. Somotor, Sect. 4.2, **Region CA#5** or Kráľová pri Senci, Sect. 4.3, **Region PB#2**).

Region HA#3: 16 sites in the north of the country; it is a mixture of high and medium-elevated sites (including the High Tatras) and is under the dominant effects of continentality. In practice, **Region HA#3** was created from all the sites that were not included in the first two regions. Tab. 7 demonstrates that there is no problem with the homogeneity of the data sets in the cold season at all; however, in the warm season and calendar year, some

Fig. 6. Final (hybrid) delineation of 56 stations into 3 regions for frequency analysis of heavy precipitation amounts in Slovakia.
Obr. 6. Finálne (hybridné) rozdelenie 56 staníc do 3 regiónov pre frekvenčnú analýzu mimoriadnych úhrnov zrážok na Slovensku.

T a b l e 6. Overview of the physical-climatological characteristics of the final set of 3 regions as a result of a hybrid approach (HA). IQR stands for inter-quartile range.

T a b u l k a 6. Prehľad fyzikálno-klimatologických charakteristík finálneho usporiadania 3 regiónov, vyčlenených na základe hybridného prístupu (HA). Skratka IQR znamená interkvartilové rozpätie.

	Region HA#1	Region HA#2	Region HA#3
Number of stations	11	29	16
Range of H [m]	165–780	100–2008	216–2635
Average of H [m]	313	361	828
IQR of R_{MAP} [mm]	590–740	565–713	668–954
Median of R_{MAP} [mm]	646	610	775
IQR of L_M	0.45–0.69	0.51–0.80	0.25–0.40
Median of L_M	0.55	0.73	0.31
IQR of R_{WC}	1.40–1.52	1.38–1.72	1.70–2.18
Median of R_{WC}	1.47	1.55	2.03

T a b l e 7. Statistics of the homogeneity tests of 15 precipitation data sets for the final set of 3 regions as a result of a hybrid approach (HA). Figures marked in *italic* indicate possible heterogeneity (H -test) or heterogeneity ($X10$ -test).

T a b u l k a 7. Výsledky analýzy testov homogenity 15 dátových súborov pre finálne usporiadanie 3 regiónov, vyčlenených na základe hybridného prístupu (HA). Čísla označené kurzívou znamenajú možnú heterogenitu (H -test) alebo heterogenitu ($X10$ -test).

Final pooling: hybrid approach	1-day prec.		2-day prec.		3-day prec.		4-day prec.		5-day prec.	
	H	$X10$	H	$X10$	H	$X10$	H	$X10$	H	$X10$
<i>Region HA#1 – 11 sites, $\chi^2_{0.95,10} = 18.31$</i>										
Calendar year	0.37	12.24	–0.65	8.26	–0.87	8.69	–1.12	6.55	–1.15	6.74
Warm season	0.28	12.03	–1.00	6.12	–0.86	7.27	–0.70	6.84	–0.31	9.64
Cold season	–0.60	9.33	–0.15	12.35	–0.08	10.52	0.42	11.78	0.89	12.85
<i>Region HA#2 – 29 sites, $\chi^2_{0.95,28} = 41.34$</i>										
Calendar year	–0.46	28.19	0.36	35.78	–0.28	31.28	–0.44	27.56	0.16	32.84
Warm season	–0.32	32.81	<i>1.88</i>	<i>48.41</i>	<i>1.22</i>	<i>43.32</i>	0.59	37.43	0.73	37.02
Cold season	0.55	29.10	0.89	27.81	0.92	31.65	0.16	28.10	0.66	30.21
<i>Region HA#3 – 16 sites, $\chi^2_{0.95,15} = 25.00$</i>										
Calendar year	<i>1.66</i>	24.42	0.99	22.73	0.91	21.90	0.46	19.23	<i>1.01</i>	22.54
Warm season	<i>1.23</i>	20.00	0.93	19.98	0.83	20.38	0.41	18.08	0.66	19.87
Cold season	–1.11	13.91	–0.13	14.51	–0.13	18.32	0.00	17.43	0.05	18.50

T a b l e 8. Statistics of the homogeneity tests of 15 precipitation data sets in a hypothetical case, when the dissimilar Podbanské station would be excluded from the *Region HA#3*. Figures marked in *italic* indicate possible heterogeneity (H -test) or heterogeneity ($X10$ -test).

T a b u l k a 8. Výsledky analýzy testov homogenity 15 dátových súborov pre hypotetický prípad, v ktorom sa stanica Podbanské vylúčila z *Regiónu HA#3*. Čísla označené kurzívou znamenajú možnú heterogenitu (H -test) alebo heterogenitu ($X10$ -test).

	1-day prec.		2-day prec.		3-day prec.		4-day prec.		5-day prec.	
	H	$X10$	H	$X10$	H	$X10$	H	$X10$	H	$X10$
<i>Region HA#3 – 15 sites, $\chi^2_{0.95,14} = 23.69$</i>										
Calendar year	0.27	16.52	0.37	18.15	0.53	18.48	0.21	16.69	<i>1.03</i>	21.66
Warm season	–0.05	14.29	0.34	16.47	0.52	17.62	0.23	16.00	0.58	18.23
Cold season	–1.08	13.13	–0.62	13.53	–0.02	18.17	0.09	16.92	0.19	18.24

of the data sets are possibly heterogeneous or lie very close to the threshold value of the H -test 1.0. Again, the reason for the enhanced heterogeneity in *Region HA#3* is the Podbanské station. The proof of this fact is given by the results of the H and $X10$ -tests that have been applied to *Region HA#3* with

the Podbanské station excluded (Tab. 8). The heterogeneity measures in Tab. 8 dropped considerably compared to the corresponding values in Tab. 7, mostly in the case of the warm season and annual data.

5. Discussion

Even though some of the data sets in *Region HA#2* and *Region HA#3* are possibly heterogeneous based on the strict (original) testing criteria (*Hosking, Wallis, 1997*), we did not make any effort to perform a further re-grouping or excluding of slightly dissimilar sites, for the following reasons:

- it is hard to find sets of regions where the ‘problematic’ sites would fit in perfectly (i.e. no higher homogeneity measures for any of the 15 data sets), and, at the same time, the arrangement of the regions is consistent with the geomorphological considerations;
- excluding the dissimilar sites from the analysis would result in a loss of useful information on the precipitation extremes;
- keeping the dissimilar sites in the analysis, but excluding their extraordinary events, is inconceivable, since we consider it to be an intolerable intervention into the process of independent evaluation;
- there is only a minority of ‘problematic’ sites in the analysis; what is more, their influence is significantly reduced when computing regionally averaged values of *L*-moments ratios to get the parameters of the regional distribution function.

Note that according the newly released criteria for homogeneity testing (*Wallis et al., 2007*), all the regions including ‘problematic’ sites could be considered as acceptably homogeneous, thus the above mentioned interpretations would lose their justification.

The precipitation frequency analysis in Slovakia could also be approached in alternative ways. The methodology of *Hosking and Wallis (1997)* applied herein is based on firmly separated regions. *Burn (1990)*, however, proposed the region-of-influence approach to frequency analysis, which makes use of ‘flexible’ regions. This means that each site has its own ‘region’, that is, a unique set of sufficiently similar stations from which information on extreme precipitation is transferred to the site of interest. The region-of-influence method has already been applied in Slovakia, and has been found superior to the conventional regional approach (*Gaál et al., 2008a*). These alternative methods are, however, sometimes not accepted by practitioners, who prefer the notion of contiguous regions and process-based subjective reasoning. Moreover, in this group of experts there is a tendency towards a unified classification of sites into regions based on generally accepted climatological regime theories. Here

we attempted to maintain continuity in approaching the problem on the one hand, but on the other hand, to introduce a less subjectively-based classification for extreme precipitation analysis.

The hybrid method of pooling resulted in the delineation of three homogeneous regions in Slovakia. Nevertheless, the number, the composition and the homogeneity of the final regions are highly dependent on the availability of the precipitation data. Therefore, as soon as a denser network of rain-gauges with adequately long precipitation records, or at least precipitation data from the network of the 56 selected stations from the year 2003 onward are available, it is strongly advised to check the homogeneity of the regions again (cf. *Kysely, 2008* vs. *Kysely and Picek, 2007*).

5. Conclusions

Spatially homogeneous regions, i.e. groups of sites that exhibit similar frequency distributions of extremes are the cornerstone of a regional frequency analysis. The present paper focused on this issue in the climatological conditions of Slovakia in order to lay the foundations for a frequency analysis of the maxima of *k*-day precipitation amounts (*k* = 1 to 5) in 3 different periods of the year (calendar year, warm/cold seasons). The paper was based on the classical *Hosking, Wallis* methodology (1997). Except for the delineation of homogeneous regions, it does not deal with the other steps of the regional *L*-moments algorithm, i.e. finding the most appropriate regional distribution function and deriving the at-site quantiles with their confidence intervals. These tasks, however, are much easier to perform compared to the identification of homogeneous regions.

Cluster analysis as an objective method and process-based pooling as a subjective method have been separately used to obtain homogeneous groups of sites. The final delineation of the selected 56 climatological stations into 3 regions appeared as a solution that takes into consideration the beneficial aspects of both methods of pooling. The main aim was to test the possibility of arriving at a unifying classification for all the durations of precipitation and seasons considered. It was shown that such a goal could be achieved only partially and by a combination of objective and subjective pooling methods.

Cluster analysis is an objective method for forming groups with similar inter-cluster properties. Nevertheless, the whole process of clustering is

influenced by a lot of subjective judgments, starting with the selection of proper clustering variables and ending with a decision as to whether the resulting set of clusters meets actual climatological and/or geomorphological considerations. Therefore we conclude that in a regional precipitation frequency analysis, it is not advisable to delineate groups of sites automatically and to base the pooling of sites on objective methods only. Objective methods of pooling such as cluster analysis should serve as an initial tool to obtain a first look at the possible configuration and composition of regions.

Acknowledgements. This work was supported by the Slovak Research and Development Agency under Contract No. APVV-0443-07. Parts of the research were also supported by VEGA Agency Grant 1/0496/08. The authors gratefully acknowledge this support.

Appendix

Let us suppose a region comprising of N sites, the i -th of them characterized by n_i (record length), $t^{(i)}$ (sample L -CV), $t_3^{(i)}$ (sample L -skewness) and $t_4^{(i)}$ (sample L -kurtosis).

The H -test by Hosking and Wallis

The H -test compares the between-site variation in sample L -moments for a group of sites with the variation that would be expected in the case of a homogeneous region. The test is based on the weighted standard deviation V of the at-site sample L -CVs:

$$V = \sqrt{\frac{\sum_{i=1}^N n_i (t^{(i)} - t^R)^2}{\sum_{i=1}^N n_i}}, \quad t^R = \frac{\sum_{i=1}^N n_i t^{(i)}}{\sum_{i=1}^N n_i}, \quad (\text{A.1})$$

where t^R is the weighted regional average of the sample L -CVs (the weights are proportional to the record length n_i). The heterogeneity measure is then

$$H = \frac{V - \mu_V}{\sigma_V}, \quad (\text{A.2})$$

where μ_V and σ_V are obtained by Monte Carlo-simulations (500 simulations with the same N and n_i as in the 'real world' are executed; at each site of the region, a 4-parameter kappa distribution is fitted with parameters that correspond to the regional

average of the sample L -moments ratios $[1, t^R, t_3^R, t_4^R]$; for each simulated region, the parameter V is calculated; from all of the simulations, the mean μ_V and the standard deviation σ_V of V are determined).

According to the initial definition of the homogeneity test (Hosking, Wallis, 1997), the critical H values are set as follows: the region is *acceptably homogeneous* if $H < 1$, *possibly heterogeneous* if $1 \leq H < 2$, and *definitely heterogeneous* if $H \geq 2$. Note that even though the authors later released the test criteria (regions with $H < 2$ may be considered as acceptably homogeneous, while $H \geq 3$ is an indicator of the definitive heterogeneity; Wallis et al., 2007), we applied the test in the original manner.

The X_{10} -test by Lu and Stedinger

The test is based on the sampling variance of a normalized 10-year precipitation x_{10} in a homogeneous region. It assumes that the precipitation amounts follow the *generalized extreme value* (GEV) distribution. The value of $x_{10}^{(i)}$ (the dimensionless 10-year precipitation at the i -th site) is estimated as follows (according to Fill, Stedinger, 1995):

$$x_{10}^{(i)} = 1 + \frac{t^{(i)}}{1 - 2^{-k}} \left(1 - \frac{(-\ln 0.9)^k}{\Gamma(1+k)} \right) \text{ if } k \neq 0, \quad (\text{A.3})$$

$$x_{10}^{(i)} = 1 + 2.4139 t^{(i)} \text{ if } k = 0, \quad (\text{A.4})$$

where $\Gamma(\cdot)$ is the gamma function and k – the shape parameter of the GEV distribution

$$k = 7.8590C + 2.9554C^2 \text{ and } C = \frac{2}{t_3^{(i)} + 3} - \frac{\ln 2}{\ln 3}. \quad (\text{A.5})$$

The heterogeneity measure is then

$$\chi_R^2 = \sum_{i=1}^N \frac{(x_{10}^{(i)} - x_{10}^R)^2}{\text{var } x_{10}^{(i)}}, \quad x_{10}^R = \frac{\sum_{i=1}^N n_i x_{10}^{(i)}}{\sum_{i=1}^N n_i}, \quad (\text{A.6})$$

where x_{10}^R is the weighted regional average of $x_{10}^{(i)}$, and $\text{var } x_{10}^{(i)}$ – the asymptotic variance of $x_{10}^{(i)}$ (it is

usually determined from a large number of simulations; however, *Lu, Stedinger* (1992) provide tables for the asymptotic variance $\text{var } x_{10}^{(i)}$.

The test statistic χ_R^2 has an approximate chi-square distribution with $N-1$ degrees of freedom. If $\chi_R^2 < \chi_{0.95, N-1}^2$, one would accept the null hypothesis; that is, the region can be considered as homogeneous at a confidence level of 95%. In the opposite case, one would reject the null hypothesis, so the region can be regarded as heterogeneous (*Lu and Stedinger*, 1992).

List of symbols

C	– constant in the approximation estimation of the shape parameter k ,
D_{ij}	– distance metric between the sites i and j [–],
h	– elevation of the station above the sea level [m],
H	– heterogeneity measure of Hosking and Wallis,
i	– station index,
j	– station index,
k	– aggregation time of precipitation amounts [days],
k	– shape parameter of the distribution function,
k	– index for site attributes,
K	– number of site attributes in the cluster analysis,
L_C	– Lapin's index of continentality [–],
L_M	– Lapin's index of the Mediterranean effect [–],
N	– total number of stations in the region,
n_i	– record length at the i -th site,
POT_0	– mean annual number of wet days (daily precipitation amounts ≥ 0.0 mm) [–],
POT_1	– mean annual number of days with daily precipitation amounts ≥ 1.0 mm [–],
q^T	– regional (dimensionless) T -year quantile [–],
Q_i^T	– T -year quantile (design value) at the i -th site [mm],
R_{MAP}	– mean annual precipitation [mm],
R_{Max}	– maximum of the mean monthly precipitation amounts in the annual cycle [mm],
R_{Max2}	– secondary maximum of the mean monthly precipitation amounts in the annual cycle [mm],
R_{Min}	– minimum of the mean monthly precipitation amounts in the annual cycle [mm],
R_{Min2}	– secondary minimum of the mean monthly precipitation amounts in the annual cycle [mm],
R_V	– mean monthly precipitation amount in May [mm],
R_{VII}	– mean monthly precipitation amount in July [mm],
$R_{W/C}$	– ratio of the mean precipitation amounts for the warm and cold seasons [–],
$t^{(i)}$	– sample coefficient of L -variation (L -CV) at the i -th site,
t^R	– regional coefficient of L -variation (L -CV),
$t_3^{(i)}$	– sample L -skewness at the i -th site,

t_3^R	– regional L -skewness,
$t_4^{(i)}$	– sample L -kurtosis at the i -th site,
t_4^R	– regional L -kurtosis,
T	– return period [years],
V	– weighted standard deviation of the at-site sample L -CVs,
x	– general symbol for a variable (site attribute),
x_{ik}	– the k -th attribute at the i -th site,
x_{max}	– maximum of the variable x ,
x_{min}	– minimum of the variable x ,
$\bar{x}$	– mean of the variable x ,
$x_{10}^{(i)}$	– 10-year dimensionless precipitation at the i -th site [–],
x_{10}^R	– regional average of $x_{10}^{(i)}$,
Γ	– gamma function,
λ	– longitude [°],
μ_i	– at-site scale factor (index value) [mm],
μ_V	– mean of the statistics V ,
σ_V	– standard deviation of the statistics V ,
σ_x	– standard deviation of the variable x ,
φ	– latitude [°],
χ_R^2	– test statistic of the of the regional homogeneity test of Lu and Stedinger,
$\text{var } x_{10}^{(i)}$	– asymptotic variance of $x_{10}^{(i)}$,
$\chi_{0.95, N-1}^2$	– critical value of the of the regional homogeneity test of Lu and Stedinger.

REFERENCES

- ADAMOWSKI K., ALILA Y., PILON P.J., 1996: Regional rainfall distribution for Canada. *Atmospheric Research*, 42, 1–4, 75–88.
- ALEXANDERSSON H., 1986: A homogeneity test applied to precipitation data. *J. of Climatology*, 6, 661–675.
- ALILA Y., 1999: A hierarchical approach for the regionalization of precipitation annual maxima in Canada. *J. of Geophysical Research*, 104 (D24), 31,645–31,656.
- BAČA P., BAČOVÁ MITKOVÁ V., 2007: Analysis of seasonal extreme flows using peaks over threshold method. *J. Hydrol. Hydromech.*, 55, 1, 16–22.
- BONI G., PARODI A., RUDARI R., 2006: Extreme rainfall events: Learning from raingauge time series. *J. of Hydrology*, 327, 3–4, 304–314.
- BURN D.H., 1990: Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research*, 26, 10, 2257–2265.
- CASTELLARIN A., BURN D.H., BRATH A., 2008: Homogeneity testing: How homogeneous do heterogeneous cross-correlated regions seem? *J. of Hydrology*, 360, 67–76.
- CHRISTENSEN O.B., CHRISTENSEN J.H., 2004: Intensification of extreme European summer precipitation in a warmer climate. *Global and Planetary Change*, 44, 107–117.
- DALRYMPLE T., 1960: Flood frequency analyses. *Water Supply Paper 1543-A*. U.S. Geological Survey, Reston.
- FAŠKO P., 2006: Slovak Hydrometeorological Institute. March, 2006. Personal communication, paval.fasko@shmu.sk.

- FAŠKO P., LAPIN M., 1996: Relation of daily and monthly precipitation totals in Slovakia and atmospheric circulation types in the period 1874–1993. In *Universitas Iagellonica Acta Scientiarum Litterarumque. Proc. of the Int. Conf. on Climate Dynamics and the Global Change Perspective*, Cracow, 175–182.
- FILL H.D., STEDINGER J.R., 1995: Homogeneity test based upon Gumbel distribution and a critical appraisal of Dalrymple's test. *J. of Hydrology*, 166, 1–2, 81–105.
- FOWLER H.J., KILSBY C.G., 2003: A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. *Int. J. of Climatology*, 23, 11, 1313–1334.
- GAÁL L., 2005: Introduction of Lapin's indices into the cluster analysis of maximum k-day precipitation totals in Slovakia. *Meteorological J.*, 8, 2, 85–94.
- GAÁL L., 2006: Metódy výpočtu štatistických charakteristík návrhových hodnôt krátkodobých až viacdenných úhrnov zrážok. (Estimation methods of the statistical properties of short-term to several-day design precipitation in Slovakia.) (In Slovak.) [PhD thesis.] Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava. 228 pp.
- GAÁL L., KYSELÝ J., SZOLGAY J., 2008a: Region-of-influence approach to a frequency analysis of heavy precipitation in Slovakia. *Hydrology and Earth System Sciences*, 12, 825–839.
- GAÁL L., SZOLGAY J., LAPIN M., 2008b: Regional frequency analysis of heavy precipitation totals in the High Tatras region in Slovakia. *Contributions to Geophysics and Geodesy*, 38, 3, 327–355.
- GELLENS D., 2002: Combining regional approach and data extension procedure for assessing of extreme precipitation in Belgium. *J. of Hydrology*, 268, 1–4, 113–126.
- GONG X., RICHMAN M.B., 1995: On the application of cluster analysis to growing season precipitation data in North America East of the Rockies. *J. of Climate*, 8, 4, 897–931.
- GUTTMAN N.B., 1993: The use of L-moments in the determination of regional precipitation climates. *J. of Climate*, 6, 12, 2309–2325.
- GUTTMAN N.B., HOSKING J.R.M., WALLIS J.R., 1993: Regional precipitation quantile values for the continental United States computed from L-moments. *J. of Climate*, 6, 12, 2326–2340.
- HLAVČOVÁ K., SZOLGAY J., HALMOVÁ D., PARAJKA J., KOHNOVÁ S., 2008: Zmeny hydrologického režimu slovenských tokov a základné adaptačné opatrenia na zmenu klímy vo vodnom hospodárstve. (Changes in hydrologic regime of Slovak rivers and basic adaptation measures to reduce climate change impacts on water management.) (In Slovak.) In *Národný klimatický program SR (National Climate Program of the Slovak Republic)*, Vol. 12. Slovak Ministry of the Environment, Slovak Hydrometeorological Institute, Bratislava, 61–86.
- HOSKING J.R.M., 1990: L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. of the Royal Statistical Society, Series B*, 52, 1, 105–124.
- HOSKING J.R.M., WALLIS J.R., 1997: Regional frequency analysis: an approach based on L-moments. Cambridge University Press, Cambridge; New York; Oakleigh. 224 pp.
- JURČOVÁ S., KOHNOVÁ S., SZOLGAY J., GAÁL L., 2002: K výberu vhodnej distribučnej funkcie maximálnych 5-denných úhrnov zrážok. (On the selection of a distribution function of maximum 5-day precipitation amounts.) (In Slovak.) *Acta Hydrologica Slovaca*, 3, 2, 165–173.
- KOHNOVÁ S., GAÁL L., SZOLGAY J., HLAČOVÁ K., 2005: Analýza maximálnych úhrnov zrážok v povodí horného Hrona. (Analysis of heavy precipitation amounts in the Upper Hron region.) (In Slovak.) Faculty of Civil Engineering, Slovak University of Technology, Bratislava, Slovakia. 1st edition. 120 pp.
- KOHNOVÁ S., PARAJKA J., 2002: Estimation of design maximum daily precipitation depths in a mountain region of Slovakia. In: *Geophysical Research Abstracts*, Vol. 4. Abstracts of the Contributions of the 27th General Assembly of the European Geophysical Society, EGS, Nice, 2002.
- KOHNOVÁ S., PODOLINSKÁ J., 2004: Lokálny odhad N-ročných maximálnych prietokov a regionálna typizácia povodí pre účely regionálnej frekvenčnej analýzy maximálnych prietokov na Slovensku. (At-site estimation of the discharge maxima and a regional typization of watersheds for the purposes of a regional frequency analysis of the discharge maxima in Slovakia.) (In Slovak.) In: *Proceedings of the seminar "Hydrologické modelovanie ako súčasť integrovaného manažmentu"*. (Hydrological modeling as a part of the integrated management.) SHMÚ, VÚVH and KVHK SvF STU Bratislava, pp. 29–41.
- KOHNOVÁ S., SZOLGAY J., GAÁL L., 2004: Estimation of design 5-day maximum precipitation totals in the upper Hron region. *Meteorological J.*, 7, 2, 79–83.
- KRIEGEROVÁ I., KOHNOVÁ S., 2005: Seasonality analysis of flood occurrence in mid-sized catchments in Slovakia. *J. Hydrol. Hydromech.*, 53, 3, 154–163.
- KUNDZEWICZ Z.W., ULBRICH U., BRÜCHER T., GRACZYK D., KRÜGER A., LECKEBUSCH G.C., MENZEL L., PIŃSKWAR I., RADZIEJEWSKI M., SZWED M., 2005: Summer floods in Central Europe – climate change track? *Natural Hazards*, 36, 1–2, 165–189.
- KYSELÝ J., 2005: Odhad parametrov rozdelení extrémnych hodnôt metódou L momentů. (Estimation of parameters of extreme value distributions using the method of L moments.) (In Czech.) *Meteorologické zprávy*, 58, 6, 161–167.
- KYSELÝ J., 2008: Trends in heavy precipitation in the Czech Republic over 1961–2005. *Int. J. of Climatology* (accepted), doi: 10.1002/joc.1784.
- KYSELÝ J., PICEK J., 2007: Regional growth curves and improved design value estimates of extreme precipitation events in the Czech Republic. *Climate Research*, 33, 243–255.
- LAPIN M., FAŠKO P., 1996: Úhrny zrážok na Slovensku a zmeny atmosférickej cirkulácie v období 1874–1993. (Precipitation amounts in Slovakia and changes in the atmospheric circulation in the period 1874–1993.) (In Slovak.) *Meteorologické zprávy*, 49, 1, 1–11.
- LAPIN M., FAŠKO P., 1998: Change of precipitation variability in Slovakia in the 1881–1997 period. In: *Proceedings of the 25th International Conference of Alpine Meteorology – ICAM98*, Torino, Italy, September 14–19, 1998, 126–131.
- LAPIN M., MELO M., 2005: Priestorová interpretácia výstupov klimatických scenárov v povodí Hrona a Váhu geoštatistickými metódami. (Spatial interpretation of climate scenario outputs in catchments of the Hron and Váh rivers by means of geostatistical methods.) (In Slovak.) In *Pekárová, P. et al. (eds.), 2005: Scenáre zmien vybraných zložiek hydrosféry a biosféry v povodí Hrona a Váhu v dôsledku klimatickej zmeny. (Scenarios of selected components of the hydrosphere and biosphere in catchments of the Hron and Váh rivers due to climatic change.) VEDA, Bratislava, 49–80.*

- LAPIN M., NIEPLOVÁ E., FAŠKO P., 1995: Regionálne scenáre zmien teploty vzduchu a zrážok na Slovensku. (Regional air temperature and precipitation change scenarios for Slovakia.) (In Slovak.) In: Národný klimatický program SR (National Climate Program of the Slovak Republic), Vol. 3. Slovak Ministry of the Environment, Slovak Hydrometeorological Institute, Bratislava, 17–57.
- LAPIN M., TOMLAIN J., 2001: Všeobecná a regionálna klimatológia. (General and regional climatology.) (In Slovak.) Comenius University Edition, Bratislava. 184 pp.
- LEŠKOVÁ D., 2002: Povodňová situácia na Dunaji v auguste 2002. (Flood situation on the Danube river in August 2002.) (In Slovak.) Vodohospodársky spravodajca, 45, 10, 8–10.
- LU L.-H., STEDINGER J.R., 1992: Sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test. *J. of Hydrology*, 138, 1–2, 223–245.
- MIKULÍČKOVÁ M., BABIAKOVÁ G., 2003: Zhodnotenie povodní na Slovensku za posledné desaťročie. (Evaluation of floods in Slovakia in the last decade.) (In Slovak.) In: Proceedings of 'Strategy of flood prevention in Slovakia'. INCHEBA, Bratislava. pp. 40–50.
- NATHAN R.J., MCMAHON T.A., 1990: Identification of homogeneous regions for the purposes of regionalisation. *J. of Hydrology*, 121, 1–4, 217–238.
- PARAJKA J., KOHNOVÁ S., SZOLGAY J., 2002: Prieštorská interpolácia *N*-ročných maximálnych denných zrážkových úhrnov v povodí horného Hrona pomocou stochastických interpolačných metód. (Spatial interpolation of *N*-year maximum daily precipitation amounts in the upper Hron catchment by means of stochastic interpolation methods.) (In Slovak.) *Acta Hydrologica Slovaca*, 3, 1, 35–45.
- PEEL M.C., WANG Q.J., VOGEL R.M., MCMAHON T.A., 2001: The utility of L-moment ratio diagrams for selecting a regional probability distribution. *Hydrological Sciences J.*, 46, 1, 147–155.
- PEKÁROVÁ P., HALMOVÁ D., MITKOVÁ V., 2005. Simulation of the catastrophic floods caused by extreme rainfall events – Uh River basin case study. *J. Hydrol. Hydromech.*, 53, 4, 219–230.
- PILON P.J., ADAMOWSKI K., ALILA Y., 1991: Regional analysis of annual maxima precipitation using L-moments. *Atmospheric Research*, 27, 1–3, 81–92.
- SANKARASUBRAMANIAN A., SRINIVASAN K., 1999: Investigation and comparison of sampling properties of L-moments and conventional moments. *J. of Hydrology*, 218, 1–2, 13–34.
- SCHAEFER M.G., 1990: Regional analyses of precipitation annual maxima in Washington State. *Water Resources Research*, 26, 1, 119–131.
- 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. of Hydrology*, 241, 1–2, 42–52.
- SOLÍN L., 2008: Analysis of floods occurrence in Slovakia in the period 1996–2006. *J. Hydrol. Hydromech.*, 56, 2, 95–115.
- STEDINGER J.R., VOGEL R.M., FOUFOULA-GEORGIU E., 1992: Frequency analysis of extreme events. In *Handbook of Hydrology*, Chapter 18. D.R. Maidment (ed.). McGraw-Hill, New York.
- STEHLIOVÁ K., KOHNOVÁ S., SZOLGAY J., 2001: Analýza dvojdnových úhrnov zrážok v oblasti horného Hrona. (Analysis of 2-day precipitation amounts in the upper Hron region.) (In Slovak.) *Acta Hydrologica Slovaca*, 2, 1, 167–174.
- STEHLIOVÁ K., KOHNOVÁ S., SZOLGAY J., 2002: Regionálna analýza maximálnych 2-dňových úhrnov zrážok v oblasti horného Hrona. (Regional analysis of maximum 2-day precipitation amounts in the upper Hron region.) (In Slovak.) *Acta Hydrologica Slovaca*, 3, 2, 174–182.
- SVEINSSON O.G.B., SALAS J.D., BOES D.C., 2002: Regional frequency analysis of extreme precipitation in North-eastern Colorado and Fort Collins Flood of 1997. *J. of Hydrologic Engineering*, 7, 1, 49–63.
- ŠAMAJ F., VALOVIČ Š., 1978: Dlhodobé priemery úhrnov zrážok na Slovensku za obdobie 1901–1970. (Long-term averages of precipitation amounts in Slovakia in the period 1901–1970.) (In Slovak.) Proceedings of Works of the Hydrometeorological Institute, Vol. 14. Alfa, Bratislava, 413 pp.
- ŠAMAJ F., VALOVIČ Š., BRÁZDIL R., 1985: Denné úhrny zrážok s mimoriadnou výdatnosťou v ČSSR v období 1901–1980. (Daily precipitation amounts with extraordinary substantiality in Czechoslovakia in the period 1901–1980.) (In Slovak.) In Proceedings of Works of the Slovak Hydrometeorological Institute, Vol. 24. Alfa, Bratislava, 9–113.
- ULBRICH U., BRÜCHER T., FINK A.H., LECKEBUSCH G.C., KRÜGER A., PINTO J.G., 2003a: The Central European floods in August 2002, Part I: Rainfall periods and flood development. *Weather*, 58, 371–376.
- ULBRICH U., BRÜCHER T., FINK A.H., LECKEBUSCH G.C., KRÜGER A., PINTO J.G., 2003b: The Central European floods in August 2002, Part II: Synoptic causes and considerations with respect to climatic change. *Weather*, 58, 434–441.
- VIGLIONE A., LAIO F., CLAPS P., 2007: A comparison of homogeneity tests for regional frequency analysis. *Water Resources Research*, 43, W03428, doi: 10.1029/2006WR005095.
- VOGEL R.M., FENESSEY N.M., 1993: L-moments diagrams should replace product moment diagrams. *Water Resources Research*, 29, 6, 1745–1752.
- WALD A., WOLFOWITZ J., 1943: An exact test for randomness in the nonparametric case based on serial correlation. *Annals of Mathematical Statistics*, 14, 4, 378–388.
- WALLIS J.R., SCHAEFER M.G., BARKER B.L., TAYLOR G.H., 2007: Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State. *Hydrology and Earth System Sciences*, 11, 1, 415–442.
- WILKS D., 2005: Statistical methods in the atmospheric sciences. 2nd ed. Academic Press, San Diego. 648 pp.
- XU R., WUNSCH D.C., 2008: Clustering. IEEE Press Series on Computational Intelligence. John Wiley and Sons, Hoboken, New Jersey. 358 pp.

Received 19 February 2009
Scientific paper accepted 9 June 2009

HYBRIDNÝ PRÍSTUP K VYČLENENIU HOMOGÉNNYCH REGIÓNŮV PRE REGIONÁLNU FREKVENČNÚ ANALÝZU ÚHRNOV ZRÁŽOK

Ladislav Gaál, Ján Szolgay, Milan Lapin, Pavol Faško

Vyčlenenie homogénnych regiónov je jedným z najdôležitejších a zároveň aj najzložitejších krokov regionálnej frekvenčnej analýzy. V predloženej práci sme sa usilovali o vytvorenie homogénnych regiónov pre regionálnu frekvenčnú analýzu mimoriadnych 1- až 5-denných úhrnov zrážok na základe pozorovaní 56 klimatologických staníc na Slovensku. Finálne zoskupenie 3 regiónov sme našli pomocou hybridnej metodiky, ktorá sa snaží zachovať kladné aspekty objektívnych metód (klastrová analýza), resp. subjektívnych úvah (expertný prístup). Hlavným cieľom práce bolo otestovať možnosť, či sa dá dopracovať k jednotnej klasifikácii staníc do regiónov pre všetky uvažované trvania, resp. sezóny výskytu mimoriadnych úhrnov zrážok. Napokon sa ukázalo, že takýto cieľ sa dá dosiahnuť iba čiastočne, a to kombináciou objektívnych a subjektívnych metód vytvárania homogénnych regiónov.

Štatistická analýza mimoriadnych hodnôt hydrologických alebo klimatologických prvkov (kulminálnych prietokov, úhrnov zrážok atď.), nazývaná ako frekvenčná analýza alebo analýza početnosti výskytu mimoriadnych javov, je intenzívne diskutovanou témou v rámci environmentálnych vedných disciplín. Štatici, hydrológovia, klimatológovia alebo inžinieri neustále vyvíjajú novšie metódy (štatistické modely) so zámerom poskytnúť spoľahlivé odhady návrhových hodnôt hydrologických prvkov, čiže oceniť pravdepodobnosť výskytu mimoriadnych javov tak, aby sa na základe týchto poznatkov mohlo znížiť riziko katastrofálnych dôsledkov extrémnych prejavov počasia na ľudstvo. Ničivé povodne v strednej Európe v júli 1997, v júni 1999 či v auguste 2002 výrazne podčiarkli dôležitosť protipovodňových opatrení. Predpokladaná zvýšená početnosť a/alebo intenzita mimoriadnych poveternostných situácií v dôsledku globálnej klimatickej zmeny taktiež potvrdzuje nutnosť zaoberať sa štúdiami frekvenčnej analýzy týchto javov.

Zatiaľ posledné komplexné, celoslovenské štúdie klimatickej zabezpečnosti mimoriadnych úhrnov zrážok, ktoré boli založené na lokálnom odhade radov pozorovania denných úhrnov zrážok, vznikli počas sedemdesiatych-osemdesiatych rokov 20. storočia (Šamaj, Valovič, 1978; Šamaj et al., 1985). V uplynulých desaťročiach sa nielenže rozšírila základňa kvalitných zrážkomerných údajov Slovenského hydrometeorologického ústavu (SHMÚ), ale hlavne sa objavili a presadili nové pohľady na frekvenčnú analýzu, medzi ktoré patria predovšetkým rôzne alternatívy regionálneho odhadu. Lokálna (tzv. *at-site*) frekvenčná analýza je tradičným nástrojom odhadu návrhových hodnôt v mieste pozorovania. Keďže však dĺžka radov pozorovania je väčšinou obmedzená na niekoľko desiatok rokov, použitie lokálneho odhadu nie

je adekvátne v prípadoch, keď centrom záujmu sú odhady návrhových hodnôt s relatívne veľkou dobou opakovania T (pre niektoré hydrotechnické stavby rádovo aj stovky rokov). Regionálna frekvenčná analýza pristupuje k tomuto problému nahradením časového rozmeru priestorovým rozmerom, t.j. namiesto jednej stanice súčasne skúma viacero staníc z daného regiónu s podobným režimom analyzovaných prvkov. Regionálny prístup tak umožňuje získať spoľahlivejšie odhady návrhových hodnôt aj s nižšou pravdepodobnosťou výskytu, avšak iba pri splnení požiadavky, že regióny sú homogénne. To znamená, že normované rozdelenia početnosti všetkých staníc v regióne musia byť identické. Slovo „identické“ platí iba hypoteticky; prakticky sa vyžaduje dostatočná podobnosť empirických distribučných kriviek, čo sa posudzuje na základe testov regionálnej homogenity, napr. pomocou tzv. H -testu (Hosking, Wallis, 1997; Appendix) alebo $X10$ -testu (Lu, Stedinger, 1992; Appendix). Kľúčovou otázkou regionálnej frekvenčnej analýzy a zároveň aj predloženej práce je teda to, ako sa definujú homogénne regióny.

Zhluková analýza je bežným spôsobom identifikácie homogénnych regiónov v regionálnej frekvenčnej analýze. Napriek tomu, že tento matematický nástroj je považovaný za objektívnu metódu vyčleňovania objektov s podobnými atribútmi, rôzne nastavenia zhlukovej analýzy a (čo je ešte dôležitejšie) proces rozhodovania o prijateľnosti výsledných zoskupení sú ovplyvnené subjektívnymi úvahami. Protikladom k takýmto prístupom je možnosť vytvoriť homogénne regióny iba na základe subjektívnych kritérií. Predložená práca pojednáva tieto dva odlišné prístupy (t.j. objektívny a subjektívny) k vymedzeniu homogénnych regiónov pre regionálnu frekvenčnú analýzu ročných a sezónnych maxim jedno- až viacdenných úhrnov zrážok.

Do analýzy bolo vybraných 56 klimatologických staníc z celého územia Slovenska (obr. 1). Rady denných úhrnov zrážok pochádzajú z obdobia 1961–2003, v niektorých prípadoch sú však k dispozícii aj od roku 1951. Súhrnný počet rokov na vybraných 56 staniaciach je 2464. k -denné úhrny zrážok sme počítali ako sumu denných úhrnov zrážok za k po sebe nasledujúcich dní ($k = 1, 2, 3, 4$ alebo 5). Maximálne k -denné úhrny zrážok sme určovali za tri obdobia: za kalendárny rok, za teplý polrok (1. apríl – 30. september), resp. za chladný polrok (1. október – 31. marec).

Ešte predtým než sme sa začali venovať vyčleneniu homogénnych regiónov, chceli sme sa presvedčiť, či je to naozaj potrebným krokom, a tak sme otestovali homogenitu Slovenska ako jediného celku. Výsledky dvoch testov homogenity (tab. 1) naznačili, že územie SR principiálne môžeme považovať za homogénne, nezávisle od trvania maximálnych úhrnov zrážok, resp. od ročného obdobia ich výskytu. Skúsenosti však hovoria, že napriek relatívne malej rozlohe Slovenskej republiky sa v krajine dá vymedziť niekoľko oblastí s rozdielnym dlhodobým režimom zrážok (napr. Faško, Lapin, 1996). Práve na základe tohto argumentu sme

pristúpili k vymedzeniu menších homogénnych regiónov na Slovensku, prvotne pomocou zhlukovej analýzy.

Širší zoznam staničných atribútov pre účely klastrovej analýzy pozostával z nasledujúcich 9 charakteristík: 3 geografické charakteristiky (zemepisná šírka, zemepisná dĺžka, nadmorská výška), 3 charakteristiky opisujúce dlhodobé zrážkové pomery stanice (priemerný ročný úhrn zrážok, priemerný počet dní v roku so zrážkami, priemerný počet dní v roku s denným úhrnom zrážok nad 1,0 mm), resp. 3 charakteristiky opisujúce vnútro-ročnú variabilitu mesačných alebo sezónnych úhrnov zrážok (podiel dlhodobého úhrnu zrážok za teplý polrok a chladný polrok, Lapinov index vplyvu Stredomoria, Lapinov index kontinentality). Po následnej korelačnej analýze sme mierne znížili počet charakteristík, ktoré boli príliš korelované (tab. 2, obr. 2), s cieľom redukovať multikolinearitu atribútov vstupujúcich do zhlukovej analýzy.

Pre zloženie homogénnych regiónov sme si určili nasledujúce požiadavky:

- porovnateľný počet staníc v jednotlivých regiónoch;
- možnosť interpretácie regiónov ako súvislých geografických jednotiek;
- súhlas výsledkov s klimatologickými poznatkami o režime zrážok;
- jednotné zloženie regiónov pre všetky kombinácie trvaní a sezón;
- homogenita vo všetkých regiónoch, pre všetky trvania a sezóny;
- žiadne subjektívne zásahy do zloženia regiónov (t.j. žiadne premiestnenie staníc).

Najpriateľnejším výsledkom zhlukovej analýzy bol súbor 5 regiónov (obr. 3), ktoré sme vytvorili na základe 5 charakteristík (zem. šírka, zem. dĺžka, priemerný ročný úhrn zrážok, podiel dlhodobého úhrnu zrážok za teplý/chladný polrok a Lapinov index vplyvu Stredomoria). Keďže však niekoľko dátových súborov v troch regiónoch naznačovali možnú heterogenitu (tab. 4), rozhodli sme sa pre subjektívne vymedzenie regiónov.

Logické členenie územia Slovenska na regióny sme pripravili v spolupráci s SHMÚ (Faško, 2006). Expertná analýza je založená na poznatkoch orografických pomeroch SR a vplyvov prevládajúceho prúdenia vzduchových hmôt na tvorbu zrážok. Výsledkom subjektívnej analýzy boli 4 regióny (obr. 4), ktoré sú v prevažnej väčšine homogénne pre rôzne trvania a sezóny výskytu zrážok (tab. 5).

Finálne rozdelenie vybraných 56 staníc do regiónov vzniklo ako kompromisné riešenie medzi výsledkami objektívnej zhlukovej analýzy a subjektívnej expertnej analýzy. Na jednej strane bola snaha zachovať si jednotnosť rozdielných geografických jednotiek, na druhej strane zas úsilie, aby aj výsledné miery regionálnej heterogenity boli čo najpriateľnejšie. Finálne členenie SR pozostáva z 3 regiónov (obr. 6). Napriek tomu, že v prípade niekoľko málo dátových súborov vo finálnom zoskupení regiónov ide o možnú heterogenitu (tab. 7), nevykonali sme žiadnu ďalšiu reorganizáciu regiónov,

keďže zvýšenú mieru heterogenity v regiónoch spôsobuje vždy malý počet tých istých staníc s mimoriadnymi hodnotami v ich výberových súboroch. Štatistické vlastnosti takýchto radov sú mierne atypické, čo potom spôsobí vyššiu mieru heterogenity, nezávisle od ich priradenia do hociktorého z regiónov. Takéto „problematické“ stanice však zásadne nevylučujeme z analýzy, lebo týmto krokom by sme odstránili zo základnej databázy zároveň aj množstvo cenných informácií o mimoriadnych úhrnoch zrážok. Posledným argumentom v prospech zachovania výsledného usporiadania regiónov je to, že na základe zrevidovaných kritických hodnôt H -testu (Wallis et al., 2007) všetky „problematické“ regióny sa môžu označiť ako definitívne homogénne.

Po zvážení všetkých argumentov sme sa rozhodli, že zoskupenie staníc získané hybridnou metodikou budeme považovať za finálne rozdelenie klimatologických staníc do homogénnych regiónov, a že tento výsledok bude základom v ďalších krokoch regionálnej frekvenčnej analýzy mimoriadnych úhrnov zrážok na Slovensku, ktorými sú voľba regionálnej distribučnej funkcie a odhad návrhových hodnôt v konkrétnych miestach regiónov.

Zoznam symbolov

C	– konštanta v aproximácii parametra tvaru k ,
D_{ij}	– vzdialenosť medzi objektmi i a j [–],
h	– nadmorská výška stanice [m]
H	– Hoskingova-Wallisova miera regionálnej heterogenity,
i	– index označujúci stanice,
j	– index označujúci stanice,
k	– trvanie zrážkových oddielov [dni],
k	– parameter tvaru distribučnej funkcie,
k	– index označujúci atribúty staníc,
K	– počet atribútov staníc v zhlukovej analýze,
L_C	– Lapinov index kontinentality [–],
L_M	– Lapinov index vplyvu Stredomoria [–],
N	– celkový počet staníc v regióne,
n_i	– dĺžka radu pozorovania na i -tej stanici,
POT_0	– priemerný počet dní v roku so zrážkami (denný úhrn zrážok $\geq 0,0$ mm) [–],
POT_1	– priemerný počet dní v roku s denným úhrnom zrážok $\geq 1,0$ mm [–],
q^T	– T -ročná regionálna (bezrozmerná) čiara prekročenia [–],
Q_i^T	– T -ročný kvantil (návrhová hodnota) na i -tej stanici [mm],
R_{MAP}	– priemerný ročný úhrn zrážok [mm],
R_{Max}	– ročné maximum priemerných mesačných úhrnov zrážok [mm],
R_{Max2}	– podružné ročné maximum priemerných mesačných úhrnov zrážok [mm],
R_{Min}	– ročné minimum priemerných mesačných úhrnov zrážok [mm],

$R_{\text{Min}2}$	– podružné ročné minimum priemerných mesačných úhrnov zrážok [mm],	$\bar{x}$	– priemerná hodnota premennej x ,
R_V	– dlhodobý priemerný úhrn zrážok v mesiaci máj [mm],	$x_{10}^{(i)}$	– 10-ročný (bezrozmerný) úhrn zrážok na i -tej stanici [–],
R_{VII}	– dlhodobý priemerný úhrn zrážok v mesiaci júl [mm],	x_{10}^R	– regionálny priemer $x_{10}^{(i)}$,
$R_{W/C}$	– podiel dlhodobého úhrnu zrážok za teplý polrok a chladný polrok [–],	Γ	– gama funkcia,
$t^{(i)}$	– koeficient L -variácie (L -Cv) na i -tej stanici v regióne,	λ	– zemepisná dĺžka [°],
t^R	– regionálny priemer koeficientu L -variácie (L -Cv),	μ_i	– indexová hodnota na i -tej stanici [mm],
$t_3^{(i)}$	– L -šikmost' na i -tej stanici v regióne,	μ_V	– priemer štatistiky V ,
t_3^R	– regionálny priemer L -šikmosti,	σ_V	– smerodajná odchýlka štatistiky V ,
$t_4^{(i)}$	– L -špicatost' na i -tej stanici v regióne,	σ_x	– smerodajná odchýlka premennej x ,
t_4^R	– regionálny priemer L -špicatosti,	φ	– zemepisná šírka [°],
T	– doba opakovania [roky],	χ_R^2	– štatistika testu regionálnej homogenity Lua a Stedingera,
V	– vážená smerodajná odchýlka koeficientu L -variácie,	$\text{var } x_{10}^{(i)}$	– asymptotický rozptyl $x_{10}^{(i)}$,
x	– všeobecné označenie premennej (atribútu stanice),	$\chi_{0.95, N-1}^2$	– kritická hodnota testu regionálnej homogenity Lua a Stedingera.
x_{ik}	– k -tý atribút i -tej stanice,		
x_{max}	– maximálna hodnota premennej x ,		
x_{min}	– minimálna hodnota premennej x ,		