Contributions to Geophysics and Geodesy Vol. 37/3, 2007

Process-based delineation of regions for a regional frequency analysis of multi-day precipitation totals in the cold season in Slovakia

L. Gaál, J. Szolgay
Faculty of Civil Engineering, Slovak University of Technology[1]

M. Lapin
Faculty of Mathematics, Physics, and Informatics, Comenius University[2]

P. Faško
Slovak Hydrometeorological Institute[3]

Abstract: Heavy multi-day precipitation totals in the cold season are studied using 35 to 53 years long series of observations from 56 climatological stations in Slovakia. Maxima of three to five-day precipitation totals in the period from October till March were chosen as data sets to be analyzed, since multi-day precipitation events in the cold season that are predominantly of frontal origin involve a considerable potential for flood risk. The selected data are processed by means of the L-moments based regional frequency analysis of Hosking and Wallis. The aim of the paper is to examine, whether it is possible to identify homogeneous regions for the regional frequency analysis based purely on subjective considerations, using knowledge about the long-term regime of precipitation in the country. Several methods of process based-delineation of geographically contiguous regions are proposed and compared. At the end, regional and at-site estimates of design values for the return periods $T = 20$ and 100 years, respectively, are calculated and compared for selected stations from different regions, in order to assess the order of uncertainties associated with the subjective selection of the regions.

Key words: Slovakia, maximum 3 to 5-day precipitation totals, cold season, regional frequency analysis, L-moments, homogeneity testing

---

[1] Radlinského 11, 813 68 Bratislava, Slovak Republic; e-mail: ladislav.gaal@stubas.sk, jan.szolgay@stuba.sk

[2] Mlynská dolina, F-1, 842 48 Bratislava, Slovak Republic; e-mail: lapin@fmph.uniba.sk

[3] Jeséniova 17, 833 15 Bratislava, Slovak Republic; e-mail: pavol.fasko@shmu.sk

1. Introduction

Estimation of design values is of a great importance in engineering, hydrology, or applied climatology. Computation of values that are expected to occur with a given probability is the most essential prerequisite of reliable designing of dams, reservoirs or sewer systems, etc.

For a long period, the at-site (single-site) frequency analysis was the only technique to assess the rarity of extraordinary events. Its methods are widely reviewed (e.g. *Stedinger et al., 1992*). Approximately till the 1990's, no study of precipitation frequency analysis in Slovakia existed but the monographies by *Šamaj and Valovič (1978)* and *Šamaj et al. (1985)*, based only on the at-site analysis of the precipitation data. However, the at-site samples of data from observations within the environmental sciences like meteorology or hydrology are not long enough (usually about several decades) to form a basis of a reliable estimation of the magnitude of a really rare event with a relatively large return period (sometimes even several hundreds of years). Thus, the regional frequency analysis came along as an alternative to the at-site analysis (*Dalrymple, 1960*), and gained wide popularity since about the 1980's. The regional frequency analysis uses data simultaneously from several sites within a given region (multi-site analysis). Such a spatial extension of the data sample requires to group together sites with similar statistical properties of their at-site probability distributions. Having a much wider dataset to analyze from a given region, one is able to obtain a more accurate estimation of the design values of the underlying variable, mostly at the tail of the selected distribution function.

Several-day extreme precipitation totals have been studied mostly by hydrologists worldwide as an input to hydrological modeling of flood risk assessment. Extensive, nation-wide projects, for example the *Flood Estimation Handbook* in Great Britain (*FEH, 1999*), the *KOSTRA project* in Germany (*Malitz, 1999*), the *Precipitation Frequency Atlas of the United States* (*Bonnin et al., 2006*) or the *HIRDS system* in New Zealand (*Thompson, 2002*) offer sophisticated methods, supported by special software for the estimation of the design precipitation of a given duration, often from several minutes to several days. *Fowler and Kilsby (2003)* focused on regional changes of the precipitation regime in Great Britain, analyzing time series of the L-moments ratios and the shape of the regional distribution curves. *Gellens (2002)* examined k-day precipitation totals ($k = 1 \ldots 30$) over Belgium. *Kyselý and Picek (2007)* presented methods and principles of the delineation of homogeneous pooling groups for the annual maxima of precipitation totals of duration 1-, 3-, 5- and 7 days. In Slovakia, various aspects of frequency estimation of multi-day precipitation extremes have been examined, mostly in the upper Hron River catchment (*Jurčová et al., 2002; Kohnová et al., 2004; 2005a,b*).

The aim of the presented paper is to compare the suitability of logical (in other words subjective or expert) methods for identification of homogeneous regions for a regional precipitation frequency analysis. Maxima of three to five-day precipitation totals are examined, mainly due to the fact that several-day precipitation events in the cold season involve a remarkable potential for flood risk.

The paper is structured in the following manner: in the next section, a short overview of the methodology of the regional frequency analysis is presented. It is followed, in Section 3, by a description of the selected sites and the precipitation climate of Slovakia, including the so called Lapin's index of the Mediterranean effect, which is a specific characteristic of the inter-annual variability of precipitation in Slovakia. Section 4 that presents the results of the paper is comprised of four sub-sections: the first one takes a look at the country as a single region, the next two sub-sections present different ways of a logical (process-based) regionalization, while the last one examines some issues related to an application of the regionally estimated design values in the engineering practice. Finally, Section 5 summarizes and discusses the results, and outlines the main avenues for a further extension of the work.

2. Methods

In the estimation of the design precipitation in Slovakia, we decided to implement the L-moments based regional frequency analysis of *Hosking and Wallis (1997)*. This methodology consists of several steps; the following sections introduce its basic principles in more details.

2.1. L-moments

L-moments are statistical characteristics analogous to the conventional (product) moments – they describe the location, the scale, the skewness, the kurtosis, or other statistical properties of the probability distributions. However, they are computed from the linear combination of the order statistic (hence the prefix L). For the definition formulas and a detailed description of the L-moments, see *Hosking (1990)* or *Hosking and Wallis (1997)*. The superiority of the L-moments over the conventional moments has been proved in a number of studies. For example, L-moments do not have sample size related bounds and are less sensitive to the presence of extraordinary values (outliers) in the data sample than the conventional moments (*Hosking, 1990*). *Vogel and Fenessey (1993)* and *Peel et al. (2001)* showed that compared to the conventional moment ratio diagrams, the L-moments ratio diagrams are more efficient tools to identify the parent statistical distribution from which the data sample might have originated, because the L-moments ratios are nearly unbiased.

2.2. The index value method

A great variety of the methods of the regional frequency analysis, mainly from the aspect of hydrology are summarized in *Cunnane (1988)*. Recently, the *index value method*, introduced by *Dalrymple (1960)*, is the most frequently used one in the environmental sciences. It assumes that in a homogeneous region

- observations at any given site are independent and identically distributed;
- observations at different sites are independent;
- frequency distributions at different sites are identical apart for a scale factor.

In other words, in a homogeneous region there is a common dimensionless regional growth curve for each site; the at-site quantile (percentile) corresponding to the return period T X_i^T is then obtained by the product of the dimensionless regional T-year quantile x^T and the at-site scale factor μ_i:

$$X_i^T = \mu_i x^T, \quad i = 1, \ ..., \ N, \tag{1}$$

where the index i denotes a given site and N is the total number of the sites in the region. The scale factor μ_i is generally termed as the *index value*; especially, in hydrology (climatology), it is called the *index flood* (*index storm*). The index value is usually estimated by the at-site sample mean of the selected variable.

2.3. Regional approach by Hosking and Wallis

Hosking and Wallis (1997) amalgamated the index value method and the toolbox of L-moments into a generally accepted guideline to the regional frequency analysis. This procedure consists of four major steps:

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

Note that geographers distinguish between *regions* and *pooling groups* depending on whether the objects form geographically contiguous units or not. In the regional frequency analysis of heavy precipitation in Slovakia, and at the recent stage of the work, we prefer geographically contiguous regions, and the terms "region", "regional" and "regionalization" will be used herein.

2.3.1. Screening of the data

A thorough scrutiny of the selected data set should be the very first step of any climatological analysis, in order to check whether the data are appropriate for the analysis. Such a procedure may reveal gross errors like outliers, repeated values, trends, shift changes or other irregularities in the observation series. *Hosking and Wallis (1993)* proposed the so-called *discordancy test*, which is aimed at identifying "sites that are grossly discordant with the group as a whole". When individual stations are represented as points in a 3-dimensional space of at-site L-moments, they form a rather scattered cluster of points. Hosking's discordancy measure then identifies points (sites), which significantly differ from the bulk of the points, i.e., lie relatively far from the imaginary center of the cluster. Properties of stations like that are very likely to stem from different climatological conditions or rough (human) errors in the stage of the data preparation (*Hosking and Wallis, 1993*).

2.3.2. Identification of homogeneous regions and testing their homogeneity

The identification of homogeneous regions is usually the most complex step of a regional frequency analysis and it requires the greatest proportion of subjective decisions. There are objective techniques (cluster analysis, principal component analysis etc.), as well as subjective methods for delineating groups of sites with similar properties. The current paper examines only some alternatives of subjective methods of defining the regions.

As soon as the regions are formed, it is necessary to assess whether their composition is meaningful. For this purpose, homogeneity tests are used. *Fill and Stedinger (1995)* compared the statistical power of some of the most frequently used regional homogeneity tests, and they concluded that the two most powerful ones are the H-test from *Hosking and Wallis (1997)* and the $X10$-test from *Lu and Stedinger (1992)*. These two tests were therefore selected as the basis for testing the homogeneity of the proposed regions in the regional frequency analysis of heavy precipitation totals in Slovakia.

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 the 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:

$$V = \sqrt{\frac{\sum\limits_{i=1}^{N} n_i \left(t^{(i)} - t^R\right)^2}{\sum\limits_{i=1}^{N} n_i}}, \quad t^R = \frac{\sum\limits_{i=1}^{N} n_i t^{(i)}}{\sum\limits_{i=1}^{N} n_i}, \tag{2}$$

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

$$H = \frac{V - \mu_V}{\sigma_V}, \tag{3}$$

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

The region is *acceptably homogeneous*, if $H < 1$, *possibly heterogeneous*, if $1 \leq H < 2$ and *definitely heterogeneous*, if $H \geq 2$ *(Hosking and Wallis, 1997)*.

The H-test has two other alternatives, which differ in the definition of the measure V in Eq. (2). In the first case, instead of the L-CV, a combination of the L-CV and the L-skewness is used; in the second one, the L-CV is replaced by a combination of the L-CV and the L-kurtosis. However, these two alternative homogeneity tests have not been applied in our analysis because of their restricted discriminatory power to identify homogeneity or heterogeneity of the candidate regions (*Hosking and Wallis, 1997*).

The X10-test by Lu and Stedinger

The test is based on the sampling variance of the normalized 10-year precipitation x_{10} in a homogeneous region. It assumes that the precipitation extremes follow the *generalized extreme value* (GEV) distribution (e.g. *Coles, 2001*). The value of $x_{10}^{(i)}$ (the 10-year precipitation at i-th site) is estimated as follows (according to *Fill and Stedinger, 1995*):

$$x_{10}^{(i)} = 1 + \frac{t^{(i)}}{1 - 2^{-k}} \left(1 - \frac{\left(-\ln 0.9\right)^k}{\Gamma\left(1 + k\right)} \right) \; if \;\; k \neq 0, \tag{4}$$

$$x_{10}^{(i)} = 1 + 2.4139 t^{(i)} \; if \;\; k = 0, \tag{5}$$

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

$$k = 7.8590C + 2.9554C^2 \text{ and } C = \frac{2}{t_3^{(i)} + 3} - \frac{\ln 2}{\ln 3}. \tag{6}$$

The heterogeneity measure is then

$$\chi_R^2 = \sum_{i=1}^{N} \frac{\left(x_{10}^{(i)} - x_{10}^{R}\right)^2}{\operatorname{var} x_{10}^{(i)}}, \quad x_{10}^{R} = \frac{\sum\limits_{i=1}^{N} n_i x_{10}^{(i)}}{\sum\limits_{i=1}^{N} n_i}, \tag{7}$$

where x_{10}^{R} is the weighted regional average of $x_{10}^{(i)}$, and $\operatorname{var} x_{10}^{(i)}$ is the asymptotic variance of $x_{10}^{(i)}$ [it is usually determined from a large number of simulations; however *Lu and Stedinger (1992)* provide tables for the asymptotic variance $\operatorname{var} x_{10}^{(i)}$].

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

2.3.3. Choice of the frequency distribution

Choosing the most appropriate distribution function is a specific issue of a frequency analysis. Various goodness-of-fit tests are used to assess whether a given frequency distribution fits the observed data acceptably closely (e.g. *Hosking and Wallis, 1997; Stedinger et al., 1992*). In the current analysis, however, only the *generalized extreme value* (GEV) *distribution* (e.g. *Coles, 2001*) is applied without any testing of the fit. There are two main reasons for it:

a) GEV is a suitable model for the 1-day, as well as the multi-day precipitation extremes in Central Europe, including the area of Slovakia (*Kyselý and Picek, 2007; Kohnová et al., 2005a*).

b) A selection of the most suitable distribution function is beyond the scope of the recent work; the design values in Section 4.4 are only estimated in order to illustrate some methodological problems related to the regional frequency analysis.

2.3.4. Estimation of the design values

The *regional L-moments algorithm* (*Hosking and Wallis, 1997*) is used in order to estimate the at-site quantiles, i.e., the design values for a given return period T.

Suppose that the region consists of N sites, where the site i is characterized by its length of observation n_i, sample mean μ_i (*index value*) and the sample L-moments ratios $t^{(i)}$, $t_3^{(i)}$, $t_4^{(i)}$, ... Regional L-moments ratios t^R and t_r^R, $r = 3,\ 4,\ ...$, are derived from the at-site sample L-moments ratios as weighted regional averages, where weights are proportional to the sites' record length n_i:

$$t^R = \frac{\sum\limits_{i=1}^{N} t^{(i)} n_i}{\sum\limits_{i=1}^{N} n_i} \tag{8}$$

and

$$t_r^R = \frac{\sum\limits_{i=1}^{N} t_r^{(i)} n_i}{\sum\limits_{i=1}^{N} n_i}, \quad r = 3,\ 4,\ ... \tag{9}$$

The regionally weighted L-moments ratios t^R and t_r^R, $r = 3,\ 4,\ ...$ are then used to estimate the parameters of the selected distribution function (it is the GEV distribution – see Section 2.3.3 above) in order to get the dimensionless cumulative distribution function (*growth curve*). Finally, the precipitation quantiles for a given return period T at the i-th site of the region are obtained by multiplying the dimensionless T-year quantile x^T with the index value μ_i (Eq. (1)).

3. Data

3.1. Selection of the sites

Altogether 56 climatological stations from the whole Slovakia have been selected (Fig. 1). The basis of the selection forms 29 climatological stations, from which observations of daily precipitation totals with no gaps since the year 1961 (in some cases since 1951) are available in the digital database of the Slovak Hydrometeorological Institute. At the time of data preparation, the last complete calendar year was the year 2003. Considering the fact that the selected 29 sites do not cover the area of Slovakia evenly, the basic selection has been extended by additional 27 sites having minor gaps in their daily rainfall records (breaks of one to several-months). Nevertheless, all the additional 27 sites have at least 35 complete years of observations. Fig. 1 shows that mainly in the central and in the northern parts of the Western Slovakia there are areas, where it was necessary to look for sites even with some gaps in order to ensure a more or less uniform spatial distribution of sites.

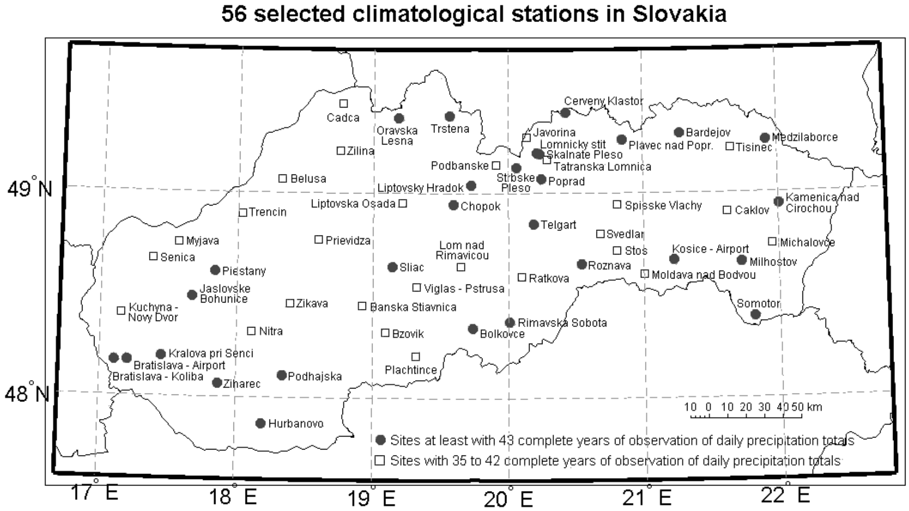

Fig. 1. 56 selected climatological stations in Slovakia.

The basic data set at the selected 56 sites makes up 2464 station-years. The shortest observation records are 35 years long (in 2 cases); on the other hand, the longest ones consist of 53 years (at 13 sites). The most common value of the length of the observation records is 43 years (15 occurrences). The missing values in the observation series have not been replaced.

We have decided to examine the maxima of three to five-day precipitation totals in the cold season, mainly for the following two reasons. Firstly, precipitation events in the cold season are predominantly of frontal (cyclonic) origin. Although they are usually not accompanied by high intensities, they involve a remarkable flood potential due to the fact that the mean duration of cyclonic synoptic situations is several days, during which considerable precipitation totals may be accumulated. Secondly, the cold season is an interesting period in Slovakia, since the precipitation regime is influenced by several different circulation factors (mainly by the Atlantic and the Mediterranean area, respectively; see Section 3.2 below) which are supposed to result in spatially heterogeneous data sets of extremes.

The k-day precipitation totals have been computed as the sums of the daily precipitation totals for k consecutive days ($k = 3$, 4 or 5). Maximum k-day precipitation totals have been determined for the cold season that lasts from October 1st till March 31st.

The data underwent standard quality checking for gross errors using Hosking's discordancy test (*Hosking and Wallis, 1993*); however, no such errors were identified at the selected 56 stations in Slovakia. Temporal homogeneity of the observation series was examined using several homogeneity tests. Except for a single case (station Lom nad Rimavicou), no significant inhomogeneities (shift or trend changes) have been detected; therefore, the data series are regarded as sufficiently homogeneous for further climatological analyses.

3.2. The precipitation regime in Slovakia

Slovakia lies in the temperate climate zone at the border between the Atlantic and the continental part of Europe. The precipitation regime of the country is therefore influenced by different factors; the most dominant ones are the effect of a) the Mediterranean area, b) the western circulation, and c) the European continent (*Lapin and Tomlain, 2001*).

A typical continental regime of precipitation has a smooth annual course: it only has a single sharp maximum in July and a single sharp minimum in winter (February). In such a clear pattern, the continental regime is discernible in the northeastern parts of the country. This basic scheme, however, is superimposed by the effect of the Mediterranean area: the absolute maximum is shifted towards the spring (June, at some places May), and, at the same time, a secondary autumn maximum appears (in October/November). *Lapin's index of the Mediterranean effect* L_M is a quantitative characteristic of the magnitude of this influence. It is defined using three ratios of certain monthly precipitation totals:

$$L_M = \frac{R_{Max}}{R_{VII}} + \frac{R_V}{R_{VII}} + \frac{R_{Max2}}{R_{Min2}} - 2.5, \tag{10}$$

where indices denote May (V), July (VII), and months with the maximum (Max), secondary maximum (Max2) and secondary minimum (Min2) of the monthly precipitation totals in the annual cycle. In the case of a smooth annual course of the monthly precipitation totals, the secondary extremes are missing; therefore the third ratio of Eq. (10) should be taken as being 1.0. The number 2.5 in Eq. (10) is a correction factor. For a more detailed line of the reasoning, see *Gaál (2005).*

The spatial pattern of the index L_M is highly influenced by the topographical conditions of the country. The mountain ranges in Central Slovakia are arranged predominantly in an east-west direction, and the height of these hills and ranges ascends generally from south to north. Moist air masses that approach Slovakia in late autumn from the Mediterranean area (from western and southwestern directions) produce enhanced precipitation in lowlands and on the windward slopes of the hills and mountains (L_M = 0.8 or higher, at some places about 1.0; Fig. 2). However, the range of Low Tatras represents a real barrier for those air masses: beyond this boundary (and in Eastern Slovakia), the effect of the Mediterranean

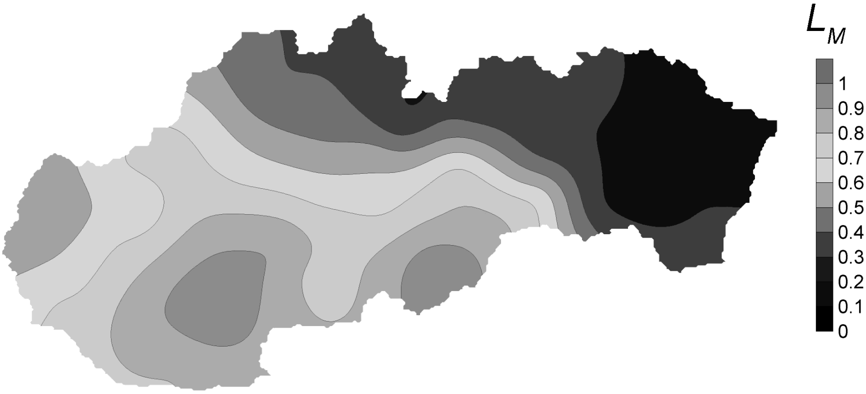

Fig. 2. Map of Lapin's index of the Mediterranean effect.

cyclones becomes insignificant, expressed by relatively low values of L_M ($L_M < 0.4$).

At some stations in the northwestern part of Slovakia, the November / December secondary maximum in the annual course of monthly precipitation is influenced by the late-autumn / early-winter precipitation increase in the North Atlantic area (*Lapin and Tomlain, 2001*).

4. Results

The results of the analysis have been divided into four sub-sections. In the first one, no regionalization method is applied, only the homogeneity of the whole country that is regarded as a compact region is examined. The next two sub-sections are devoted to different methods of a process-based division of Slovakia into several geographically contiguous regions. In the last sub-section, selected design values of five-day precipitation totals are estimated and mutually compared at a couple of stations from different regions.

4.1. Slovakia as a single region

In the first stage of our analysis, we decided to examine the homogeneity of the whole area of Slovakia (labeled as Region SR). If the results of the homogeneity tests showed that such a region is homogeneous, then there would not be necessary to deal with a further division of the country into smaller sub-regions. The statistics of two homogeneity tests applied to Region SR are summarized in Table 1. H and $X10$ measures demonstrate that there is a questionable homogeneity of the analyzed data sets in all the cases: the H-test indicates possible heterogeneity three times, which is confirmed by the $X10$-test in two cases (what is more, in the case of the maxima of 4-day precipitation totals, the $X10$ measure is relatively close to the critical value).

The L-moments diagrams (Fig. 3) of the analyzed data sets show scattered clusters of points, where it is rather difficult to identify individual sites or groups of sites that might be considered as a clear reason of the inhomogeneities. However, a possible alternative of explanation is depicted

Table 1. The results of the homogeneity tests, applied on the case when the whole area of Slovakia is treated as a single region (Region SR), consisting of 56 sites. Two homogeneity tests have been used: the H-test (*Hosking and Wallis, 1997*) and the $X10$-test (*Lu and Stedinger, 1992*). The critical value of the $X10$-test is $\chi^2_{0.95,N-1} = \chi^2_{0.95,55} = 73.31$. Figures marked in bold indicate possible heterogeneity (H-test) or heterogeneity ($X10$-test)

	Region SR ($\chi^2_{0.95,55}$ = 73.31)	
	H-test	*X10*-test
3-day precipitation	**1.47**	**74.96**
4-day precipitation	**1.29**	71.18
5-day precipitation	**1.83**	**74.97**

in Fig. 3. The ellipses highlight four stations (Beluša, Žilina, Čadca and Medzilaborce), which show similar behavior on each L-moments diagrams. In all the cases, they form an isolated cluster of points on the very left side of the graphs; therefore, they might be the reason for higher values of homogeneity tests. As a result of excluding these sites from the analysis, the test statistics drop considerably below the critical values, and indicate homogeneity of the Region SR*, consisting of 52 stations (Table 2). The very low values of L-CV at these stations indicate an especially stable regime of the cold season precipitation. Beluša, Žilina and Čadca are neighboring stations in the northwestern parts of the country, and their performance might be explained by their different regime of the winter precipitation due to the influence of the Atlantic area. Windward slopes in this micro-region receive enhanced precipitation totals in the period November to January as a result of the dominating westerly circulation *(Lapin et al., 1995; Lapin and Faško, 1998; Lapin and Tomlain, 2001).* The reason of the behavior of the station Medzilaborce, located in the northeastern corner of Slovakia, might be considered as the effect of the vast Euro-Asian continent. Another explanation is based on the fact that higher altitudes of the Eastern Carpathians ridge have in Slovakia considerably more precipitation in December and January, than the remaining parts of eastern Slovakia. It is possibly a remote influence of the North Atlantic precipitation regime also in this small region. Note that the described explanation of the observed heterogeneity of the Region SR is one of the several other possible solutions: excluding other sites, for example from the right top corner of the cluster

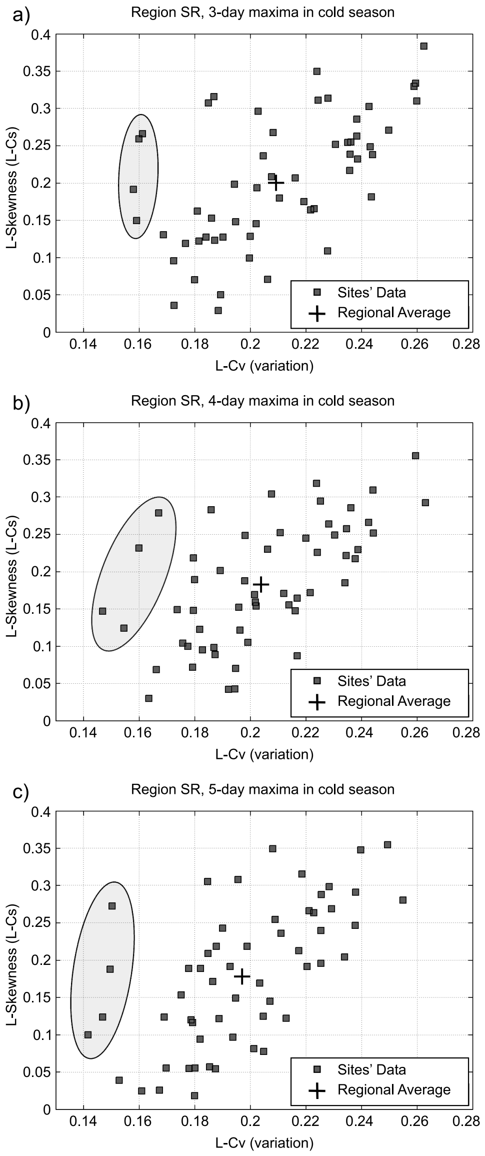

Fig. 3. *L*-moments diagrams for the maxima of a) 3-day, b) 4-day, and c) 5-day precipitation totals in the cold season when the whole country is treated as a single region (Region SR). The ellipses mark 4 stations (Beluša, Žilina, Čadca and Medzilaborce) that are supposed to be the reason of the inhomogeneities of the data sets.

Table 2. The results of the homogeneity tests, applied on the case when the whole area of Slovakia is treated as a single region with 4 stations (Beluša, Žilina, Čadca and Medzilaborce) excluded, i.e. consisting of 52 sites (Region SR*)

	Region SR* ($\chi^2_{0.95,51}$ = 68.67)	
	H-test	*X10*-test
3-day precipitation	0.45	55.76
4-day precipitation	0.18	52.45
5-day precipitation	0.40	53.24

of points in Fig. 3, might lead to improved heterogeneity measures as well.

As a consequence of the detected inhomogeneities, we decided to delineate sub-regions for Slovakia by means of different methods of a subjective, process-based (in other words logical) regionalization.

4.2. Process-based regionalization A

A simple process-based regionalization has been accomplished based on Lapin's index of the Mediterranean effect L_M. As described in Section 3.2., this index expresses the influence of the cyclones of Mediterranean origin mostly in the southern parts of Slovakia in the late autumn; that is, exactly in the cold season. Using Fig. 2 as a guide, the area of the country has been divided into two parts: *Region A#1* (= region#1 from the process-based regionalization A) that is affected by this phenomena, and *Region A#2* that is not or only slightly influenced (Fig. 4). The threshold value of L_M between two groups is about 0.45, since there is a gap in the ascendingly ordered L_M values: 0.426 is the maximum in *Region A#2*, while 0.494 is the minimum in the other region.

The results of the homogeneity testing are presented in Table 3, and the L-moments diagrams for all the six data sets are shown in Fig. 5. According to both tests, *Region A#1* is homogeneous; moreover, the test statistics would drop when excluding station Beluša from the region (its position on the L-moments diagrams is indicated by a circle, Fig. 5a-c). On the other hand, *Region A#2* is certainly heterogeneous. This fact is not surprising, since *Region A#2* is a mixture of sites from areas that are under different dominant climatological influences. In the central parts of *Region A#2*, the

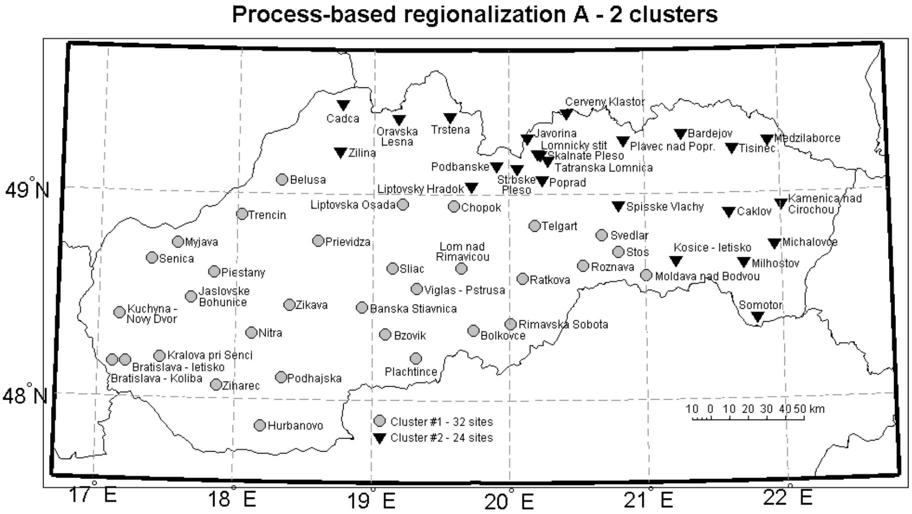

Fig. 4. Delineation of two regions based on the spatial pattern of Lapin's index of the Mediterranean effect (process-based regionalization A).

Table 3. The results of the homogeneity tests, applied on the case when Slovakia is divided into two regions, based on the spatial pattern of Lapin's index of the Mediterranean effect (process-based regionalization A). The critical values of the $X10$-test are placed in the column headers. Figures marked in bold indicate possible heterogeneity (H-test) or heterogeneity ($X10$-test); star (*) indicates definitive heterogeneity (H-test)

Process-based regionalization A	Region A#1 ($\chi^2_{0.95,31}$ = 44.99)		Region A#2 ($\chi^2_{0.95,23}$ = 35.17)	
	H-test	*X10*-test	*H*-test	*X10*-test
3-day precipitation	0.48	34.59	**1.74**	**38.10**
4-day precipitation	0.33	33.79	**1.88**	**37.24**
5-day precipitation	0.45	31.20	**2.44***	**43.88**

High Tatras dominate; in east, the effect of the continent and the rainfall shadow of the mountains (e.g. the High and Low Tatras, the Slovenské Rudohorie) have leading role in formation of the precipitation regime; and in northwest, there is the influence of the Atlantic area, described in details in the previous sections. Excluding stations Čadca, Žilina and Medzilaborce

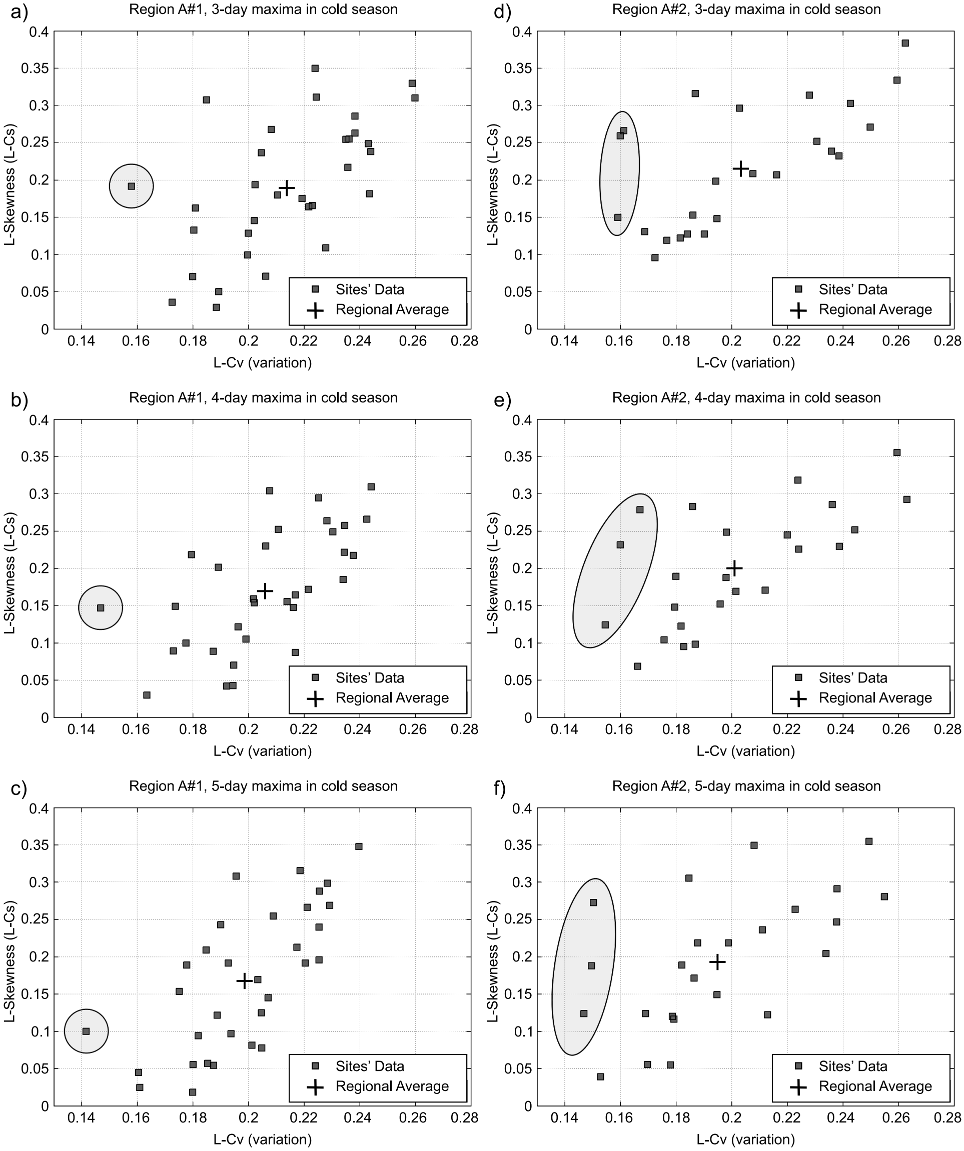

Fig. 5. *L*-moments diagrams for the maxima of 3-day (top), 4-day (middle), and 5-day (bottom) precipitation totals in the cold season when Slovakia is divided into two regions (process-based regionalization A). Figures a)-c) correspond to Region A#1 (with the station Beluša highlighted by a circle), and figures d)-f) correspond to Region A#2 (with the stations Čadca, Žilina and Medzilaborce highlighted by an ellipse).

from *Region A#2* (highlighted by an ellipse in Fig. 5d-f) would result in more acceptable statistics of the H-test (0.91; 0.98 and 1.13 for 3-, 4- and 5-day events, respectively); however, the new H measures still indicate that it is hard to consider the remainder of *Region A#2* as homogeneous.

We conclude that the regionalization based on Lapin's index of the Mediterranean effect reached its goal: the region, which is noticeably influenced by the dominating phenomenon, is homogeneous. The other region has been formed in a passive way, since it contains all the sites NOT included in the first region. Consequently, the detected heterogeneity of *Region A#2* fully meets our expectations. Further climatological processes have to be identified and analyzed in this region, in order to divide it into homogeneous sub-regions.

4.3. Process-based regionalization B

A more elaborated process-based delineation of regions has been proposed by P. Faško from the Slovak Hydrometeorological Institute (*Faško, 2006*). Taking into consideration the topography of the country, and the effects of different patterns of general air-mass circulation, four regions of the extreme precipitation have been identified (Fig. 6). The homogeneity measures of the individual regions are summarized in Table 4, and the L-moments diagrams for selected data sets (maxima of 5-day precipitation totals in each region) are presented in Fig. 7. The L-moments diagrams for the maxima of 3-day and 4-day precipitation totals are not shown herein; however, they are to a great extent similar to the demonstrated ones.

Region B#1 (= region #1 from the process-based regionalization B) is located in the western parts of the country, completely along the boundaries with the Czech Republic, and includes 11 stations. The region is homogeneous, since all the statistics of H and $X10$-tests, respectively, lie below the critical values (Table 4). Stations Beluša, Čadca and Žilina that might have caused inhomogeneities in the previous arrangements (Sections 4.1. and 4.2.) do not show strange behavior any longer in this region. Higher H statistic ($H = 0.89$) in case of the maxima of 5-day precipitation totals is caused by the station Kuchyňa - Nový Dvor, symbolized by a separated point in the left top corner in Fig. 7a. The reason for this fact is probably an outlying value of 106.1 mm observed at the station Kuchyňa - Nový

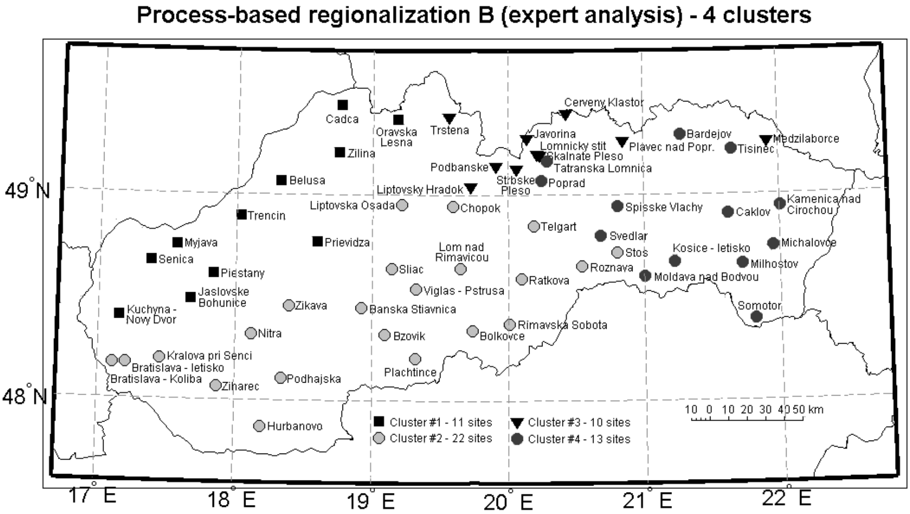

Fig. 6. Delineation of four regions based on an expert analysis (process-based regionalization B).

Table 4. The results of the homogeneity tests, applied on the case when Slovakia is divided into four regions, based an expert analysis (process-based regionalization B). The critical values of the $X10$-test are placed in the column headers. Figures marked in bold indicate possible heterogeneity (H-test) or heterogeneity ($X10$-test)

Process-based regionalization B	Region B#1 ($\chi^2_{0.95,10}$ = 18.31)		Region B#2 ($\chi^2_{0.95,21}$ = 32.67)		Region B#3 ($\chi^2_{0.95,9}$ = 16.92)		Region B#4 ($\chi^2_{0.95,12}$ = 21.03)	
	H-test	*X10*-test	*H*-test	*X10*-test	*H*-test	*X10*-test	*H*-test	*X10*-test
3-day precipitation	-0.08	10.52	-0.12	19.02	0.25	13.86	**1.14**	13.41
4-day precipitation	0.42	11.78	-0.55	17.80	0.38	12.34	**1.47**	15.80
5-day precipitation	0.89	12.85	-0.29	17.20	0.81	14.17	**1.64**	19.60

Dvor, while the other 5-day extremes in this region usually do not reach 70-80 mm.

Region B#2 is the largest region from the process-based regionalization B: it is comprised of 22 stations from the southern parts of West and Central Slovakia, respectively. The very low values of the test statistics in Table 4, as well as a relatively compact cluster of points in Fig. 7b indicate a high

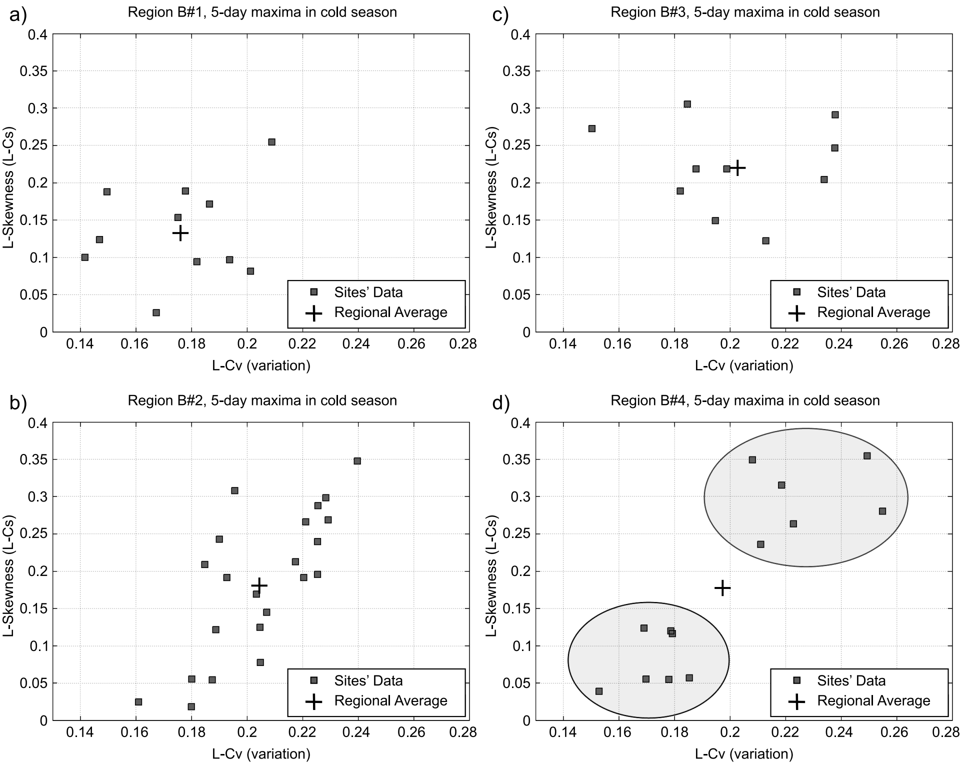

Fig. 7. *L*-moments diagrams for the maxima of 5-day precipitation totals in the cold season when Slovakia is divided into four regions (process-based regionalization B). a) Region B#1, b) Region B#2, c) Region B#3, and d) Region B#4.

degree of homogeneity in this region. Negative values of the H-test may be explained by an excessive regularity in the precipitation data, i.e., the sites within a region are inordinately cross-correlated due to the same extreme precipitation events (*Hosking and Wallis, 1997*).

Region B#3 consists of 10 stations, located mostly in the northern parts of Central Slovakia, mainly in the area of the High Tatras and its proximity. Homogeneity of the data sets within this region is acceptable (Table 4). A higher value of the H-test ($H = 0.81$) for the maxima of 5-day precipitation totals is observed due to the station Medzilaborce, which is the only station in this region that does not belong to the area of the High Tatras and its neighborhood. In Fig. 7c, the station Medzilaborce is represented by separated a point on the very left side of the diagram. If this station was excluded from *Region B#3* the test statistic would drop to $H = -0.17$.

Region B#4 with its 13 sites is stretched in the East Slovakia, approximately from the High Tatras to the southeastern boundaries of the country. The H-test in Table 4 yields possible heterogeneity of the region in all the cases; however, the statistics of the $X10$-test do not exceed critical values. The L-moments diagram (Fig. 7d) reveals that the region is made up of two separate groups of sites. The most interesting thing is, however, that these groups form two distinct geographical regions. The first group is made up of 7 sites from the southern parts (lowlands) of *Region B#4*: Moldava nad Bodvou, Čaklov, Košice - Airport, Milhostov, Somotor, Michalovce and Kamenica nad Cirochou (the left bottom cluster in Fig. 7d). The other group consists of 6 sites from the northern parts of *Region B#4*; their regime of precipitation is considerably influenced by a more complex terrain in their surroundings: Poprad, Tatranská Lomnica, Švedlár, Spišské Vlachy, Bardejov and Stropkov - Tisinec (the top right cluster in Fig. 7d). The L-moments diagrams of the maxima of 3-day and 4-day precipitation totals show similar results (not included here). According to these results, a further subdivision of *Region B#4* should be considered.

In spite of the slight inhomogeneities detected in the forth region, the logical (expert) delineation of regions proposed by Faško seems to be a very reasonable and acceptable starting point for a regional frequency analysis of heavy multi-day precipitation totals in the cold season.

4.4. Estimation of design values

Regionalization in the field of the frequency analysis is of a primary importance. Benefits of the regional approach compared to the traditional at-site estimation methods have been proved in a multitude of papers in the scientific literature (e.g. *Acreman and Sinclair, 1986; Wiltshire, 1986; Lettenmaier et al., 1987*). It is no longer questionable, whether to prefer the regional approach over the at-site one; the question is how to do the regionalization. Nevertheless, there is no golden rule or universal guideline for it. Both subjective and objective methods (the latter not discussed herein) are widely used in practice.

However, regardless of the methods, if the proposed regions seems to be logical, and are in accordance with the common climatological, geographical, etc. knowledge base, there is no reason to reject them. Certainly, there are always several acceptable sets of regions, since the process of regionalization is never a unique one. Consequently, this fact has a serious impact on the engineering practice. If there are several, mutually equivalent and acceptable sets of the regions, then the design values for a given site may differ, too, as a result of using different regional information. Therefore, different design values from different but acceptable sets of regions must also be acceptable. In this sub-section, we are going to give an illustration of these problems.

Altogether 8 stations (Čadca, Hurbanovo, Kamenica nad Cirochou, Kuchyňa - Nový Dvor, Liptovský Hrádok, Medzilaborce, Švedlár, Telgárt) from different parts of the country have been selected to demonstrate the differences between the design value estimation based on various delineation of regions, and the at-site analysis, respectively. All the regions from the process-based regionalization B are represented by two stations, and, at the same time, 4-4 stations have been taken from both regions of the process-based regionalization A. The only aim of such a selection was to include stations that are either representative for a given region (e.g. Hurbanovo, Liptovský Hrádok), or problematic in a certain sense (e.g. Čadca or Medzilaborce - see previous sections). The design values of the maxima of the 5-day precipitation totals in the cold season for the return periods $T = 20$ and 100 years, estimated by the generalized extreme value distribution are presented in Tabs. 5 and 6.

The at-site estimates are only shown to illustrate relatively large differences between the regional and the at-site approaches, respectively. In Tabs. 5 and 6, there are considerably overestimated (stations Kuchyňa - Nový Dvor, Liptovský Hrádok, Švedlár, Telgárt), as well as underestimated (stations Hurbanovo, Kamenica nad Cirochou) at-site quantiles with respect to the regional ones. The overestimation may exceed 20% (21-29% for the return period $T = 100$ years in the case of Liptovský Hrádok station, Table 6), while the rate of underestimation is obviously less, up to 11-12% for $T = 100$ years.

The inter-comparison among the regional estimates reveals that they are more balanced than the locally estimated quantiles. There are only minor differences between the 100-year quantiles based on the Region SR and the process-based regionalization A (Table 6); moreover, no difference is observed for the return period $T = 20$ years (Table 5). The process-based regionalization B, however, shows an ambivalent behavior. While the design values for the stations in *Regions B#2* and *B#4* are reasonably similar to the other quantiles estimated regionally, for the stations in *Region B#1* and *Region B#3*, there are marked differences. It is due to the fact that a great portion of stations in *Region B#2* (*B#4*) is identical with *Region A#1* (*A#2*), while, for example, in the case of Čadca station, the regional information is pooled from regions *B#1* and *A#2*, respectively, with almost completely different compositions.

Table 5. Design values of the maxima of 5-day precipitation totals in the cold season for the return period $T = 20$ years, for 8 selected stations. Abbreviations: AS – at-site estimation lacking regional approach, SR – Region SR, PA – process-based regionalization A, PB – process-based regionalization B. Figures marked in italic indicate that the design values have been estimated using regional information within an inhomogeneous region

	Belongs to the region		R_T [mm], T = 20 years				Relative difference [%]		
			AS	SR	PA	PB	(AS-SR)/ SR	(AS-PA)/ PA	(AS-PB)/ PB
Kuchyňa – Nový Dvor	A#1	B#1	78.2	*74.7*	74.7	70.5	4.7	4.7	10.9
Čadca	A#2	B#1	73.6	*81.4*	*81.4*	76.8	-9.6	-9.6	-4.2
Hurbanovo	A#1	B#2	68.6	*71.1*	71.1	72.2	-3.5	-3.5	-5.0
Telgárt	A#1	B#2	105.7	*96.6*	96.6	98.1	9.4	9.4	7.7
Liptovský Hrádok	A#2	B#3	86.7	*77.5*	*77.5*	79.4	11.9	11.9	9.2
Medzilaborce	A#2	B#3	68.8	*74.6*	*74.6*	76.4	-7.8	-7.8	-9.9
Švedlár	A#1	B#4	102.2	*94.6*	94.6	*94.7*	8.0	8.0	7.9
Kamenica nad Cirochou	A#2	B#4	70.7	*74.8*	*74.8*	*74.9*	-5.5	-5.5	-5.6

5. Summary and conclusions

The presented paper focuses on one of the most essential issues of a regional frequency analysis, namely on the identification of homogeneous regions, and testing their spatial homogeneity. The analysis has been performed on the maxima of three to five-day precipitation totals in the cold

Table 6. Design values of the maxima of 5-day precipitation totals in the cold season for the return period $T = 100$ years, for 8 selected stations. Abbreviations: AS – at-site estimation lacking regional approach, SR – Region SR, PA – process-based regionalization A, PB – process-based regionalization B. Figures marked in italic indicate that the design values have been estimated using regional information within an inhomogeneous region

	Belongs to the region		R_T [mm], T = 100 years				Relative difference [%]		
			AS	SR	PA	PB	(AS-SR)/ SR	(AS-PA)/ PA	(AS-PB)/ PB
Kuchyňa – Nový Dvor	A#1	B#1	109.2	*96.0*	95.2	85.9	13.8	14.7	27.1
Čadca	A#2	B#1	92.0	*104.6*	*105.9*	93.7	-12.0	-13.1	-1.8
Hurbanovo	A#1	B#2	83.5	*91.3*	90.6	93.5	-8.5	-7.8	-10.7
Telgárt	A#1	B#2	152.1	*124.1*	123.1	127.0	22.6	23.6	19.8
Liptovský Hrádok	A#2	B#3	128.6	*99.6*	*100.8*	106.5	29.1	27.6	20.8
Medzilaborce	A#2	B#3	92.6	*95.8*	*97.0*	102.5	-3.3	-4.5	-9.7
Švedlár	A#1	B#4	153.4	*121.6*	120.5	*121.6*	26.2	27.3	26.2
Kamenica nad Cirochou	A#2	B#4	85.5	*96.1*	*97.3*	*96.2*	-11.0	-12.1	-11.1

season. In the first approach, Slovakia has been examined as a compact region with no subdivision into regions. Following this, two alternatives of a subjective, process-based regionalization have been accomplished. The first one divided the country into two parts based only on the influence of the cyclones of Mediterranean origin. The second, expert analysis resulted in the delineation of four regions according to the effects of the general atmospheric circulation in the complex topography of Slovakia. In all the cases, homogeneous as well as possibly or definitively heterogeneous regions have been identified by two independent homogeneity tests. A common feature of the homogeneity testing is that the inhomogeneities are usually caused by the same stations – it is really hard to find for them an appropriate region, where they would not spoil the regional homogeneity.

The conclusions of the paper, with an emphasized stress on the engineering practice, may be summarized as follows. If one is interested in calculating design values at a given site with return periods T shorter than the length of the observation n, the estimation may be based both on a regional, as well as on an at-site approach, since the differences between the two approaches are negligible. Nevertheless, it is strongly advised to use regional methods if one desires to extrapolate the return periods T beyond the scope of the available length of the data series n. Tabs. 5 and 6 prove that regardless of the way of delineating the regions, design values estimated using regional methods are relatively stable compared to the quantiles estimated using a traditional at-site approach. The information gathered from the other sites of a given region favorably reduces the uncertainty of the estimated quantiles (*Hosking and Wallis, 1997*). On the other hand, one should be aware of an overly extrapolation of the return periods T. It is recommended to constrain to the relationship $T \leq 3n$: quantiles corresponding to longer return periods cannot be estimated reliably (*Malitz, 1999*).

It is obvious that the magnitude of a design value highly depends on the composition of the regions. Unfortunately, at the recent stage of the analysis it is not possible to decide which of the regional estimates is the most reliable one. The quantiles in Tabs. 5 and 6 are only point estimates of the design values. In order to determine the most suitable delineation of regions (and, in a broader sense, the most acceptable regional frequency model), it is necessary to get information about the spread of the estimates as well, for example, by means of multiple Monte Carlo simulations. This might be one of the main avenues for continuation with the presented analysis in the near future. Another possibility is extending this work by cluster analysis as an objective method of regionalization. The most important issue of clustering is to find the most appropriate descriptors of the precipitation regime as input variables of the cluster analysis.

Acknowledgments. The authors gratefully acknowledge the Slovak VEGA Grant Agency for the support of grants 1/2032/05, 1/4024/07 and 1/1042/04. The support of APVT 51-017804 project is also acknowledged. Thanks are due to J. Parajka (Technische Universität, Wien) for assistance in drawing figures.

References

Acreman M. C., Sinclair C. D., 1986: Classification of drainage basins according to their physical characteristics, and application for flood frequency analysis in Scotland. Journal of Hydrology, **84**, 3, 365–380.

Bonnin G. M., Todd D., Lin B., Parzybok T., Yekta M., Riley D., 2006: Precipitation-frequency Atlas of the United States. NOAA Atlas, **1**, 4.0, NOAA, NNWS, Silver Spring, Maryland. 261 p. `http://hdsc.nws.noaa.gov/hdsc/pfds/docs/NA14Vol1.pdf` [cit. 2007-06-20].

Coles S., 2001: An introduction to statistical modeling of extreme values. Springer-Verlag, London. 222 p.

Cunnane C., 1988: Methods and merits of regional flood frequency analysis. Journal of Hydrology, **100**, (1-3), 269–290.

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, e-mail: pavol.fasko@shmu.sk.

FEH, 1999: Flood estimation handbook, **1-5**, Institute of Hydrology, Wallingford.

Fill H. D., Stedinger J. R., 1995: Homogeneity test based upon Gumbel distribution and a critical appraisal of Dalrymple's test. Journal 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. International Journal 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 Journal, **8**, 2, 85–94.

Gellens D., 2002: Combining regional approach and data extension procedure for assessing of extreme precipitation in Belgium. Journal of Hydrology, **268**, 1-4, 113–126.

Hosking J. R. M., 1990: L-moments: analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society, Series B, **52**, 1, 105–124.

Hosking J. R. M., Wallis J. R., 1993: Some statistics useful in regional fequency analysis. Water Resources Research, **29**, 2, 271–281.

Hosking J. R. M., Wallis J. R., 1997: Regional frequency analysis: an approach based on L-moments. Cambridge University Press, Cambridge, 224 p.

Jurčová S., Kohnová S., Szolgay J., Gaál L., 2002: To the selection of an appropriate distribution function of the maxima of 5-day precipitation amounts. Acta Hydrologica Slovaca, **3**, 2, 165–174 (in Slovak).

Kohnová S., Gaál L., Szolgay J., Hlavčová K., 2005a: Analysis of heavy precipitation amounts in the Upper Hron region. Faculty of Civil Engineering, Slovak University of Technology, Bratislava, 120 p. (in Slovak).

Kohnová S., Lapin M., Szolgay J., Gaál L., 2005b: Methodology for the selection of 10-day maximum precipitation totals and their statistical analysis in the upper Hron region. Contr. Geophys. Geod., **35**, 3, 299–318.

Kohnová S., Szolgay J., Gaál L., 2004: Estimation of design 5-day maximum precipitation totals in the upper Hron region. Meteorological Journal, **7**, 2, 79–83.

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., 1998: Change of precipitation variability in Slovakia in the 1881-1997 period. In: Booklet of Proceedings of the 25th International Conference of Alpine Meteorology - ICAM98, Torino, Italy, September 14-19, 1998, 126–131.

Lapin M., Nieplová E., Faško P., 1995: Regional scenarios of temperature and precipitation change in Slovakia. In: Slovak National Climate Programme, II, 3, MŽP SR and SHMÚ, Bratislava, 17–57 (in Slovak).

Lapin M., Tomlain J., 2001: General and regional climatology. Vydavateľstvo UK, Bratislava, 184 p. (in Slovak).

Lettenmaier D. P., Wallis J. R., Wood E. F., 1987: Effect of regional heterogeneity on flood frequency estimation. Water Resources Research, **23**, 2, 313–323.

Lu L.-H., Stedinger J. R., 1992: Sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test. Journal of Hydrology, **138**, 1-2, 223–245.

Malitz G., 1999: Starkniederschlag in Deutschland - Messergebnisse, statistische Auswertungen, Schätzungen. In: Klimastatusbericht 1999, Deutscher Wetterdienst, 35–41. `http://www.dwd.de/de/FundE/Klima/KLIS/prod/KSB/ksb99/malitz.pdf` [cit. 2007-06-20].

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 Journal, **46**, 1, 147–155.

Stedinger J. R., Vogel R. M., Foufoula-Georgiou E., 1992: Frequency analysis of extreme events. In: Handbook of Hydrology, Chapter 18, D. R. Maidment (ed.). McGraw-Hill, New York.

Šamaj F., Valovič Š., 1978: Long-time averages of precipitation totals in Slovakia in the period 1901-1970. Proceedings of works of HMI 14. Alfa, Bratislava, 413 p. (in Slovak).

Šamaj F., Valovič Š., Brázdil R., 1985: Daily precipitation totals with extraordinary intensity in ČSSR in the period 1901-1980. In: Proceedings of works of SHMI, 24, Alfa, Bratislava, 9–113 (in Slovak).

Thompson C. S., 2002: The high intensity rainfall design system: HIRDS (Abstract). In: Proceedings of the International Conference on Flood Estimation, Bern, Switzerland, March 6-8, 2002.

Vogel R. M., Fenessey N. M., 1993: L-moments diagrams should replace product moment diagrams. Water Resources Research, **29**, 6, 1745–1752.

Wiltshire S. E., 1986: Identification of homogeneous regions for flood frequency analysis. Journal of Hydrology, **84**, 287–302.