# REGION-OF-INFLUENCE APPROACH TO THE FREQUENCY ANALYSIS OF HEAVY k-DAY PRECIPITATION AMOUNTS

**L. Gaál [1], J. Kyselý [2], P. Štěpánek [3]**

**[1] Czech Hydrometeorological Institute, Prague, ladislav.gaal@chmi.cz**
**[2] Institute of Atmospheric Physics AS CR, Prague, kysely@ufa.cas.cz**
**[3] Czech Hydrometeorological Institute, Brno, petr.stepanek@chmi.cz**

**Abstract**

A regional frequency analysis of heavy k-day precipitation totals (k = 1 to 7 days) has recently been carried out in the Czech Republic by means of the L-moment-based regionalization approach proposed by Hosking and Wallis. This method has generally been considered a kind of standard in the field of the frequency analysis since the early 1990s worldwide. However, there are other frequently referred regionalization techniques in the literature, and one of the most remarkable methods is the region-of-influence (ROI) approach. Its main idea is that there is no need to delineate fixed boundaries between the proposed regions but rather each site can have its own "region" – a group of sites that are sufficiently similar to the site of interest. The ROI method has been proposed in order to overcome possible inconsistencies that may occur on boundaries of fixed homogeneous regions. In such cases, a classical approach of the regional analysis may lead to undesirable step changes of variables and estimated quantiles; such problems are eliminated in the ROI method by defining a unique set of stations (a "region of influence") for each site under study.

The ROI method has up to now been used predominantly in connection with a flood frequency analysis; the current study is one of the first attempts to develop the ROI approach to the frequency analysis of precipitation extremes. A comparison of the results with findings of the "standard" regional frequency analysis methods in the area of the Czech Republic is performed.

**Key words**: regional frequency analysis, region-of-influence approach, extreme precipitation events, Czech Republic

## 1 Introduction

Information on *design values* (quantiles) of extreme precipitation is inevitable at various fields of human activities: in water resources research, designing dam and sewer systems, flood protection, soil and vegetation loss protection in case of intensive convective storms etc. In a traditional *"at-site" approach* to a frequency analysis, the desired precipitation quantiles have long been estimated only using the data sample at the site of interest. There is, however, a fundamental drawback of this single-site approach: in practice, one often needs design values corresponding to return periods that are much larger than lengths of available series of observations. In order to overcome such problems, the so called *regional frequency analysis* came along in the 1960s (Dalrymple, 1960) and has gained wider popularity since the 1980s (e.g. Wiltshire, 1986; Lettenmaier et al., 1987; Hosking and Wallis, 1993). The core idea of the regional approach is a substitution of time extent for space: based on a multi-site analysis one can obtain more reliable quantile estimates than using a classical, at-site approach. The ultimate question of the regional frequency analysis is therefore the following one: *How to draw the regions?*

Recently, the majority of the regional precipitation (or flood) frequency analyses have been based on fixed regions drawn according to political (Gellens, 2002; Pilon et al., 1991; Adamowski et al., 1996), geographical (Sveinsson et al., 2002; Kohnová et al., 2005), hydrological and/or climatological considerations (Smithers and Schulze, 2001; Kjeldsen et al., 2002; Fowler and Kilsby, 2003). A generally accepted guideline to regional frequency analysis is the so-called *regional L-moment algorithm* (Hosking and Wallis, 1997), which is also based on the delineation of fixed regions.

The *region-of-influence* (abbr. ROI) *method* introduced by Burn (1990) is an alternative approach to the regional frequency analysis. The main idea of the technique is that there is no need to delineate fixed boundaries between the proposed regions but rather the regions are defined in a flexible way. It means that each site can have its own "region", that is, a unique set of sufficiently similar stations, from which extreme precipitation information is transferred to the site of interest. The similarity of sites is evaluated by a properly chosen set of site attributes (site characteristics and/or site statistics; see Section 3). Therefore, even two neighboring stations may have very different sets of stations that represent the regions of influence for each site.

The ROI method was proposed in a flood frequency analysis in order to overcome possible inconsistencies that may occur on boundaries of fixed homogeneous regions. In such cases, a classical approach of the regional analysis may lead to undesirable step changes of variables and estimated quantiles; such problems are eliminated in the ROI method by defining a unique set of stations for each site under study.

Burn (1990), analyzing Canadian catchments, concluded that the ROI method is superior to the "standard" regional approaches: the novel ROI technique resulted in improved flood quantile estimates even at "difficult" individual stations. Zrinji and Burn (1994) revisited and extended the ROI approach for ungauged sites. Holmes et al. (2002) applied the ROI approach in regionalization of low flow characteristics within the United Kingdom, while Castellarin et al. (2001) used the ROI technique as a background for comparison of several hydrological similarity measures in northern Italy.

Up-to-now, the ROI method has been used predominantly in connection with the flood frequency analysis. The present study focuses on development of the region-of-influence approach to the frequency analysis of precipitation extremes, and presents its comparison with the "standard" regional frequency analysis methods in climatological conditions of the Czech Republic.

The paper is structured in a following manner: after a short description of selected stations and their data, a detailed overview of the mathematical background of the ROI method is presented. Due to its complexity, the analysis has not been finished at the time of the paper submission deadline, therefore results of the analysis are missing and will be presented at the conference.

## 2 Data

Daily precipitation totals measured at 145 stations operated by the Czech Hydrometeorological Institute (CHMI) were used as an input dataset (Fig. 1). The altitudes of stations range from 158 to 1322 m a.s.l., and the observations span the period 1961-2005. Three main criteria were applied when selecting the stations and forming the dataset:

- the stations approximately evenly cover the area of the Czech Republic;
- there were no significant station moves during 1961-2005 (all sites where location changes exceeded 50 metres in altitude were excluded from the analysis);
- the daily series of precipitation records are uninterrupted.

At about 10% of sites, gaps in daily records appeared (in total not exceeding 2 months over 45 years). These missing daily data were estimated using measurements at 3 to 5 nearest locations available in the climatological database of the CHMI, according to a methodology proposed by P. Štěpánek (note that the CHMI operates a dense network of gauging sites; mean distance to the nearest measuring site was only 16 km for the locations where missing data were estimated). All other potential station records with more than 2 months of missing values were excluded from the analysis at this stage; however, they might be used in future when the ROI method will be applied to construct maps of design value estimates of precipitation extremes.

Samples of maximum annual 1-, 3-, 5- and 7-day precipitation amounts were drawn from each station records and are examined as extreme precipitation events. Summer and winter half-years are analyzed separately. Such an approach allows, at least in a rough approximation, to differentiate between extreme precipitation amounts of various origin: convective events that dominate in summer and long-lasting stratiform events of frontal origin that are typical for autumn and winter.

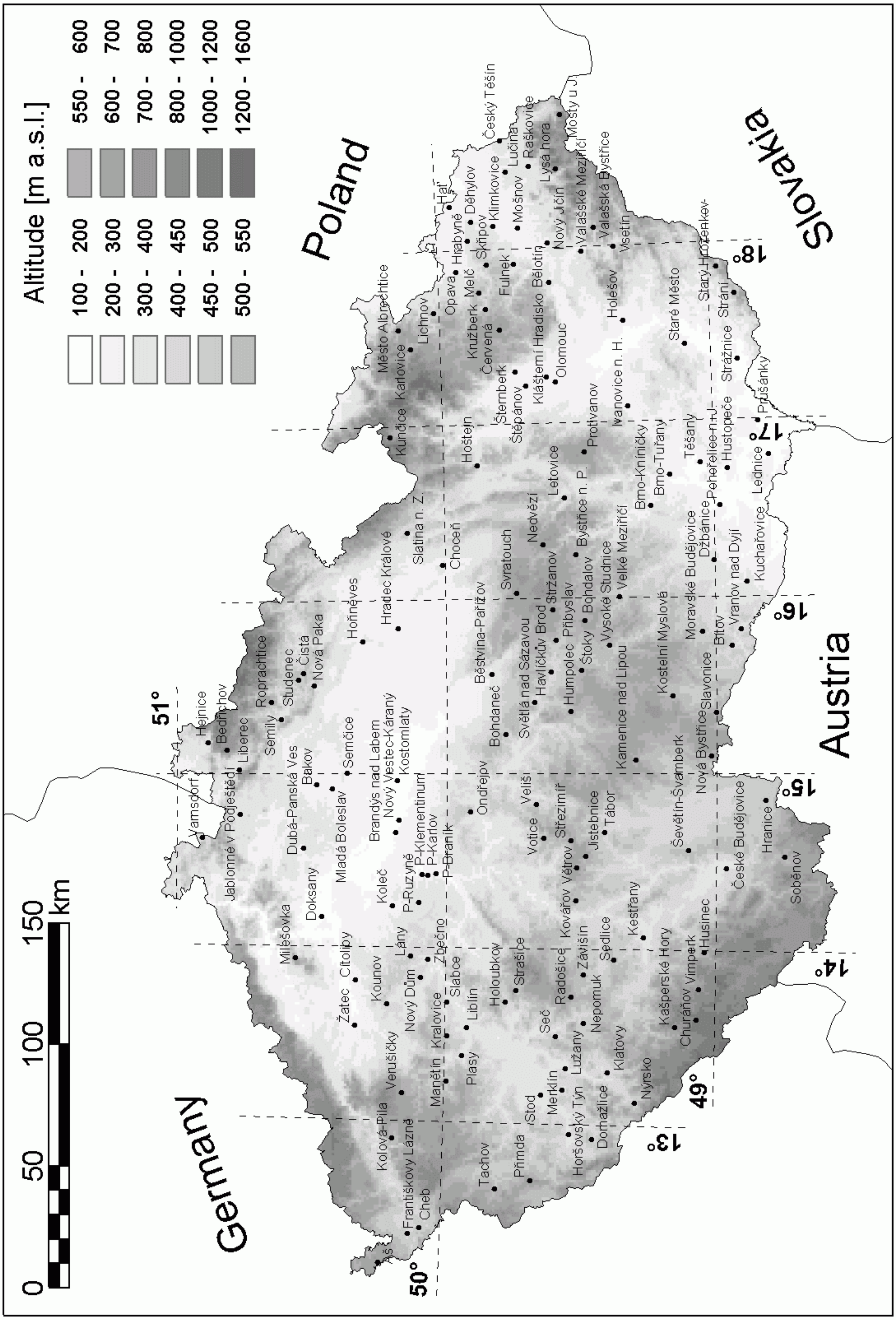


Fig. 1  Location of stations used in the analysis.

The data underwent standard quality checking for gross errors as well as checking in terms of a discordancy measure based on L-moments (Hosking and Wallis, 1993).

## 3 Mathematical model of the region-of-influence method

The region of influence for a given site consists of a group of sites that are sufficiently similar to the site of interest. The similarity of sites is judged according to site characteristics and/or site statistics (Hosking and Wallis, 1997). *Site characteristics* are quantities that are known before any measurement or observation is obtained at a site and include, for example, the location, the elevation and other physical-geographical properties associated with a site. Some long-term characteristics of the precipitation regime of sites (such as mean annual precipitation or long-term averages of monthly/seasonal precipitation totals) are also frequently classified here: although they are set from the at-site measurements, their values can also be estimated at ungauged sites from climatological maps with a relatively smaller degree of uncertainty compared to extremes. On the other hand, *site statistics* are simply the measurements or results of statistical processing of the observed data at a given site. Hereafter, site characteristics and site statistics will be termed generally as *site attributes*.

### *The distance metric*

The closeness of each site to every other site in the attribute space is evaluated by means of the weighted Euclidean distance metric that is known mainly from cluster analysis:

$$D_{ij} = \left[ \sum_{m=1}^{M} W_m \left( X_m^i - X_m^j \right)^2 \right]^{\frac{1}{2}}, \tag{1}$$

where $D_{ij}$ is the weighted Euclidean distance between sites *i* and *j*, $W_m$ is the weight associated with the *m*-th site attribute, $X_m^i$ is the value of the *m*-th attribute at site *i*, and *M* is the number of attributes.

The distance metric matrix *D* is symmetrical ($D_{ij} = D_{ji}$) with zeros on its diagonal ($D_{ii} = 0$).

The region of influence for a given station is constructed according to the following scheme. Firstly, the site with the lowest value of $D_{ij},\ j = 1, ..., N$ is added to the ROI for site *i*. Particularly, in the very first step it is the site *i* itself, for which the distance metric $D_{ii} = 0$ is always the one with the lowest value. Following this, a further site with the next lowest value of $D_{ij}$ is added into the ROI for site *i*. Sites are successively added into the ROI until a given option (see below: "*Pooling a station's ROI*") for forming of the ROI is fulfilled.

As the site attributes $X_m$ may have substantially different magnitudes, a transformation of the initial site attributes before calculating (1) should be applied. A possible alternative of such operation is normalization of the variables:

$$X \rightarrow \frac{X - \overline{X}}{\sigma_X}, \tag{2}$$

where $\overline{X}$ is the mean and $\sigma_X$ is the standard deviation of the attribute *X*. As a result of normalization (2), each of the site attributes $X_m$ are of a comparable magnitude, i.e. they have zero mean and unit variance.

Selection of site attributes has a key role in the ROI method: the success of the whole procedure lies in finding the right number and combination of proper site characteristics and/or statistics. In the first attempts with the ROI technique for regional precipitation analysis in the Czech Republic, three different alternatives of selection of site attributes are discussed:

a) <u>Alternative #1</u>: **site statistics**

The whole set of site attributes consists exclusively of statistical characteristics that are related to the examined data sample at each of the sites. The following site statistics are considered:

1) *coefficient of variation* ($c_v$) - a traditional characteristics of the scale of the data sample:

$$c_v = \frac{\sigma}{\mu}, \tag{3}$$

where $\mu$ is the mean and $\sigma$ is the standard deviation of the data sample.

2) *Pearson's second skewness coefficient* (*PS*) - a less traditional characteristics of the skewness of the data sample (Weisstein, 2006):

$$PS = \frac{3(\mu - m)}{\sigma}, \tag{4}$$

where *m* is the median of the data sample.

3) *10-year design precipitation estimated using the generalized extreme value (GEV) distribution* ($R_{10y}$) - a characteristics of extreme value magnitudes of the data sample.

Selection of the aforementioned site statistics ensures that the ROI for a given site consists of sites for which characteristics of the probability distribution functions are similar enough to the one of the target site.

The set of attributes in alternative #1 follows Burn's concept, which has been applied in the introduction of the ROI method (Burn, 1990). There are, however, small differences between the original and the current analysis. Burn used a modification of the Pearson's skewness coefficient, where the factor "3" in the numerator of eq. (4) has been neglected. Considering the fact that the attribute *PS* in the current analysis underlies the transformation (2), it can be concluded that there would be no difference between results based either on the "original" or the "modified" Pearson's skewness coefficient. Another difference can be found in selection of the third site attribute. Burn used $Q_{10y}$, which is "a plotting position estimate of the 10-year flood event, interpolated from the available annual flow series". Instead of it a GEV estimation of the 10-year value has been used herein, mainly for two main reasons: (1) "The normalized 10-year flood estimator has very little bias and approximately a normal distribution in small samples" (Lu and Stedinger, 1992). As the sites in the precipitation frequency analysis for the Czech Republic has 45 years of observations, the design precipitation with return period $T = 10$ years can reliably be estimated from each of the at-site data samples. (2) As shown in previous studies on regional approaches to the precipitation frequency analysis, the GEV is a generally acceptable distribution in each of the 4 homogeneous regions within the Czech Republic (Kyselý et al., 2006; Kyselý and Picek, 2006). As a consequence, we believe that the uncertainty of choosing the GEV-estimated 10-year design precipitation can be at most comparable with the uncertainty of the Burn's method of the 10-year value estimation.

b) <u>Alternative #2</u>: **general climatological site characteristics**

The set of site attributes is made up by characteristics that describe the long-term precipitation regime of the site, regardless of the observed precipitation extremes. The following quantities are taken into consideration:

1) *mean annual precipitation* ($\overline{R}$);

2) *the ratio of the mean precipitation totals for the summer and winter half-year* ($R_S / R_W$);

3) *a characteristics of inter-annual variability of precipitation amounts* (exact definition of such a site characteristics is not readily available at the time of preparing this contribution; however, the basic idea is that this site attribute is similar to the *Lapin's index of the Mediterranean effect*, defined for climatological conditions of the Slovak Republic; Lapin and Melo, 2005; Gaál, 2005).

Using the site characteristics of alternative #2 in the distance metric would, in principle, result in groups of sites with similar climatological conditions that may be to some extent related also to mechanisms that generate heavy precipitation. In practice, however, there is no guarantee that proximity of sites in the *M*-dimensional space of climatological site characteristics implies similarity in terms of extreme precipitation regime.

c) Alternative #3: **geographical site characteristics**

This set of site attributes consists of the well-known and widely used geographical characteristics of the site location:

1) *latitude* ($\varphi$);
2) *longitude* ($\lambda$);
3) *elevation* above the sea level ($H$).

Having alternative #3 selected, there emerge groups of sites which members are close to the site of interest in a geographical sense. Nevertheless, it cannot be strictly understood as the ordinary geographical proximity between two points. For example, a higher-elevated site is usually grouped together with other higher-elevated sites and is not necessarily joined with other sites, which are located really close to the target site in the traditional sense of latitude and longitude.

Again, concerning alternative #3, the same drawback can be mentioned as in the previous case: pooling sites according to geographical characteristics does not necessarily imply similarity in terms of precipitation climate.

Alternatives #1-#3 also take into consideration the question of the design value estimation at gauged / ungauged sites. When the regional analysis is aimed to obtain quantiles at ungauged sites as well, alternative #1 is not feasible as data sample to determine site statistics are not available. In such cases, much attention should be paid on finding the most appropriate climatological characteristics.

Eq. (1) incorporates a further factor: a weighting coefficient associated with the attributes. In practice, weighting coefficients reflect the relative importance of the attributes within each of the alternatives. Weights can be computed, for example, by means of multivariate analysis, where independent variables can be the selected site attributes and the dependent variable could be a suitably chosen site characteristics. Burn (1990) used a simple correlation analysis between each of the selected attributes and the 100-year flood event, respectively. The resulting correlation coefficients were then set as the weights for the attributes. Such a multivariate or correlation analysis is also a useful tool in reducing the number of candidate attributes according to objective criteria.

In the current analysis, equal (unit) weights $W_m = 1, m = 1, ..., M$ have been selected for each of the alternatives. What is more, $M = 3$ for each of the three alternatives, and therefore no further analyses have been carried out in order to reduce the number of the candidate attributes. In the current study, there have been no attempts to evaluate the relative importance and/or the number of the site attributes because (1) the set of attributes in each of the alternatives reflects reasonable considerations; (2) we believe that $M = 3$ is the minimum number of attributes which is sufficient for calculation of the matrix of distance metric; (3) our attention was focused on other - possibly more important - aspects of the ROI approach. However, such a task is an excellent challenge for the upcoming investigations within the problems of a regional frequency analysis by the technique of ROI.

***Pooling a station's ROI***

As soon as the appropriate site attributes are selected and the distance metric matrix is calculated there remain two main tasks so that each site's ROI can be found. The first one is to determine a cutoff-point of the distance metric for the *i*-th site: all sites below the selected threshold value will be included into the site's ROI, and, on the other hand, any station with a distance value in excess of the threshold will not:

$$I_i = \left\{ j : D_{ij} \le \theta_i \right\}, \tag{5}$$

where $I_i$ is the set of stations in the ROI for site *i*, and $\theta_i$ is the threshold distance value for site *i*.

The next important feature of the ROI is a determination of the weighting parameters associated with each of the sites included in the ROI. These weighting coefficients should reflect the relative closeness of any site of the ROI to the site of interest. The substance of this step is evident: the closer a given site is to the site of interest in sense of the distance metric, the bigger amount of information it provides in the regional precipitation frequency analysis. In a mathematical form:

$$\eta_{ij} = f\left(D_{ij}, \Psi\right) \quad \forall\ j \in I_i \tag{6}$$

and

$$\eta_{ij} = 0 \quad \forall\ j \notin I_i, \tag{7}$$

where $\eta_{ij}$ is the weight for site *j* in the ROI for site *i*; *f* ( ) is a functional relationship defining the weight, and $\Psi$ is parameter vector for the function *f* ( ).

Following Burn's (1990) framework, the threshold distance $\theta_i$ and the weights $\eta_{ij}$ in the current analysis are determined according to 3 different options. These options reflect 3 different philosophies of pooling information from the sites of the ROI in order to get more accurate quantile estimates at the site of interest.

A) Option #1: "Less sites with high values of weights"

The basic idea of option #1 is that the ROI for a given site entails only a limited number of stations; however, all of the selected stations are given weights substantially different from zero while computing regional relationships.

The threshold value $\theta_i$ (5) is defined as follows:

$$\theta_i = \theta_L \qquad \text{if } NS_i \ge NST, \tag{8}$$

where $\theta_L$ is a lower threshold value defining a desired proximity for stations to be included in the ROI for site *i*, $NS_i$ is the number of stations in the ROI for site *i* with the threshold at $\theta_L$, and *NST* is the target number of stations for a ROI.

If the number of sites in the ROI for the threshold at $\theta_L$ is lower than a desired minimum number *NST* then a less restrictive threshold $\theta_i$ should apply:

$$\theta_i = \theta_L + \left(\theta_U - \theta_L\right)\left(\frac{NST - NS_i}{NST}\right) \qquad \text{if } NS_i < NST, \tag{9}$$

where $\theta_U$ is an upper threshold value for sites with fewer than *NST* stations in the ROI.

The weighting function for this option is defined as:

$$\eta_{ij} = 1 - \left(\frac{D_{ij}}{TP}\right)^n, \tag{10}$$

where $\Psi = \{TP, n\}$ are the parameters of the weighting function.

There are 5 parameters to be initialized for option #1: *NST*, $\theta_L$, $\theta_U$, *TP* and *n*.

B) Option #2: "More sites with different values of weights"

In option #2, a relatively large number of sites are included in the ROI for a given site. Stations which are sufficiently similar to the site of interest have unit weights, while stations that are less similar have appropriately low values of weighting functions.

The threshold value $\theta_i$ is defined as follows:

$$\theta_i = \theta_U, \tag{11}$$

where $\theta_U$ is a constant threshold value.

The weighting function for this option is defined as:

$$\eta_{ij} = 1 \qquad \text{if } D_{ij} \le \theta_L, \tag{12}$$

and

$$\eta_{ij} = 1 - \left(\frac{D_{ij} - \theta_L}{TN - \theta_L}\right)^n \qquad \text{if } \theta_L < D_{ij} \le \theta_U, \tag{13}$$

where $\theta_L$ is a lower threshold value for the distance matrix, and *TN* and *n* are the parameters of the weighting function. *TN* is defined in a specific way:

$$TN = \max\left(TL_i, TPP\right), \tag{14}$$

where

$$TL_i = \max_{\{j\}}\left(D_{ij}\right). \tag{15}$$

$\Psi = \{\theta_L, \theta_U, TPP, n\}$ are the four parameters of the weighting function for option #2 to be initialized.

C) Option #3: "All sites with different values of weights"

Option #3 is nearly the same as option #2 with the only difference that in option #3 all of the analyzed stations are included in the ROI for a given site with proper values of weighting functions.

The threshold value $\theta_i$ is defined as follows:

$$\theta_i = TL_i, \tag{16}$$

and the definition of the weighting function is the same as in case of option #2 – see eqs. (12)-(15). As option #3 entails all of the sites there is no need to deal with selection of the upper threshold $\theta_U$; therefore the number of the parameters to be initialized is 3 ($\Psi = \{\theta_L, TPP, n\}$).

### *Parameter estimation for the ROI options*

In case of options #1-#3, values of several parameters have to be selected following subjective considerations, in accordance with the theoretical background of individual options. The distance metric matrix should serve as the basis for estimation of the threshold values $\theta_L$, $\theta_U$, *TP* and *TPP*. As mentioned (see above "*The distance metric*"), the distance metric matrix is a symmetrical one with zeros on its diagonal. The non-zero values above (below) the diagonal form an upper (lower) triangular matrix that contains the distance metric values between each pair of stations. It is possible to sort the metric values within a triangular matrix in ascending order, to derive their empirical distribution, and then find certain statistical characteristics (maximum, minimum, median or given percentiles) of such distribution.

Burn (1990) suggests choosing the values of the parameters $\theta_L$, $\theta_U$, *TP* and *TPP* right according to the empirical distribution of distance metric values. The lower distance threshold $\theta_L$ should be associated with the 25$^{th}$ percentile distance value (in both options #1 and #2); the upper distance threshold $\theta_U$ should be the 75$^{th}$ percentile (again, in both options #1 and #2), while the parameters *TP* and *TPP* should be the 85$^{th}$ percentile of the distance metric distribution.

The target number of sites *NST* for option #1 was set to 15 (20). Considering the fact that the total number of sites appearing in the current analysis is 145, the choice *NST* = 15 (20) seems to be a reasonable compromise between having too many or too few sites, respectively, within a ROI. Small *NST* does not necessarily have to represent a remarkable improvement of the regional approach in comparison with a pure at-site analysis. On the other hand, too high value of *NST* may lead, in principle, to creating potentially inhomogeneous sets of stations within a ROI for a given site. In the Burn's original framework, *NST* was set to 15, i.e. approximately to one third of the total number (46) of the available hydrometric stations.

Exponent $n$ [eqs. (10) and (13), respectively] also significantly influences the magnitudes of the weighting coefficients $\eta_{ij}$. According to Burn (1990), the value of $n$ in option #1 was set to 2.5. Such a choice corresponds with the conception of option #1: any ROI consists of a relatively small number of sites, with each site within the ROI having a high value of weighting coefficient that is not unduly different from 1.0. On the other hand, exponent $n$ in case of options #2 and #3 was set to 0.1. This choice enables one to make a clear distinction between two types of sites in a ROI: it assigns relatively low values of weights to sites which are "further" from the site of interest (between the threshold values $\theta_L$ and $\theta_U$), while the information from "closer" sites is pooled by weighting coefficients 1.0.

### *Estimation of at-site precipitation quantiles using information from the ROI*

As soon as each station's ROI and the appropriate weighting coefficients are known, it is possible to estimate the at-site precipitation quantiles using information from the sites of the ROI. The design values of extreme precipitation may be computed by the *L-moment-based index storm procedure*.

The number of sites in the analysis is $N$ with the $j$-th site having sample size $n_j$. In the index storm procedure, at-site data $X_{j,k}$, $j = 1, \ldots, N$, $k = 1, \ldots, n_j$ are rescaled by the so-called *index storm* (usually the sample mean $\mu_j$, as well as in the current analysis) in order to get dimensionless data:

$$x_{j,k} = \frac{X_{j,k}}{\mu_j}, \quad k = 1, \ldots, n_j . \tag{17}$$

These dimensionless data $x_{j,k}$ at site $j$ are then used to compute the sample L-moments $l_1^{(j)}$, $l_2^{(j)}$, … and L-moment ratios:

$$t^{(j)} = \frac{l_2^{(j)}}{l_1^{(j)}} \tag{18}$$

and
$$t_r^{(j)} = \frac{l_r^{(j)}}{l_2^{(j)}}, \quad r = 3, 4, \ldots, \tag{19}$$

where $j$ denotes the index of the analyzed site, $t^{(j)}$ is the sample L-CV and $t_r^{(j)}, r = 3, 4, \ldots$ are sample L-moments ratios at site $j$ (for a detailed definition and description of the L-moments refer to Hosking and Wallis, 1997).

Regional L-moment ratios $t^{(i)R}$ and $t_r^{(i)R}, r = 3, 4, \ldots$ within the ROI for site $i$ are derived from the at-site sample L-moment ratios as weighted regional averages, where two different weights are applied: the sample size $n_j$ (length of observation) and the weighting function based on the ROI distance metric $\eta_{ij}$:

$$t^{(i)R} = \frac{\sum_{j \in I_i} t^{(j)} n_j \eta_{ij}}{\sum_{j \in I_i} n_j \eta_{ij}} \tag{20}$$

and
$$t_r^{(i)R} = \frac{\sum_{j \in I_i} t_r^{(j)} n_j \eta_{ij}}{\sum_{j \in I_i} n_j \eta_{ij}}, \quad r = 3, 4, \ldots, \tag{21}$$

where $I_i$ is the set of stations included in the ROI for site $i$ (5), for which the weighted regional L-moment ratios are calculated. The regionally weighted values $t^{(i)R}$ and $t_r^{(i)R}, r = 3, 4, \ldots$ are then used to derive the parameters of the selected distribution function in order to get the dimensionless cumulative distribution function, the so-called *growth curve*. Having the growth curve for the site of interest $i$, the precipitation quantiles with the desired return period $T$ can be obtained by multiplying the dimensionless $T$-year growth curve value $x_i^T$ and index storm $\mu_i$:

$$X_i^T = \mu_i x_i^T \tag{22}$$

In the current analysis, the generalized extreme value (GEV) distribution is used as the regional distribution function. It was selected mainly for two main reasons: (1) the analysis is focused on investigation of the results stemming from combinations of various alternatives of site attribute sets and options for defining weighting functions, respectively, and the possibility of choosing among several different distribution functions might have negative influence on clear explanation of the results; (2) the GEV distribution has been proved as a generally acceptable regional distribution function in previous studies on regional frequency distributions of precipitation extremes in the Czech Republic (Kyselý et al., 2006; Kyselý and Picek, 2006).

For each of the sites ($N$ = 145), precipitation durations (k = 1, 3, 5 and 7 days), alternative of set of attributes (a#1-a#3) and options of ROI procedures (o#1-o#3), 7 extreme precipitation quantiles $X_i^T$ are determined, namely the design values with return periods of $T$ = 2, 5, 10, 20, 50, 100 and 200 years.

***Confidence intervals for the estimated quantiles and the evaluation of the ROI approaches***

Monte Carlo simulations are carried out in order to evaluate the uncertainty associated with the estimated quantiles, and at the same time, to evaluate the performance of various ROI approaches. The Monte Carlo experiment consists of *NR* = 1000 repetitions, with each of the realizations made up of the following steps:

1) Use a random number generator to generate artificial samples of precipitation data at each of the analyzed stations, having the same record lengths as their real-world counterparts. The simulated data samples at the *i*-th site of the region have kappa distribution with parameters corresponding to the at-site L-moments [1, $t^{(i)}$, $t_3^{(i)}$, $t_4^{(i)}$]. The kappa distribution as a parent distribution is used in order to avoid too early a commitment to a particular distribution as a parent. In cases when fitting of the kappa distribution fails the GEV or generalized logistic (GLO) distribution is used instead.
2) Calculate the at-site statistics, determine the distance metric matrix, and define the region of influence and weighting function values for each station, each of the above described alternatives of attribute sets and each option of ROI procedures. It should be noted that in case of alternatives #2 and #3, there is no need to recalculate the values of $D_{ij}$, $I_i$ and $\eta_{ij}$ in each repetition of the Monte Carlo simulation as the distance metric is determined from unchanging site characteristics.
3) Determine the at-site estimates of L-moments, the regional L-moments within each station's ROI, and, finally, the "simulated" extreme precipitation quantiles for each station in accordance with the above described alternatives and options.

From the Monte Carlo experiment, it is possible to draw confidence intervals for the estimated extreme precipitation quantiles. For each quantile $X_i^T$ (for all combinations of the alternatives and options), the confidence interval is defined as the interval between the 5$^{\text{th}}$ and the 95$^{\text{th}}$ percentile of the empirical distribution of the simulated quantiles.

The relative performance of various ROI options is also evaluated using the simulated extreme precipitation quantiles. The simulated design values are used to calculate the *root mean square error* (RMSE) and *bias* for each quantiles through:

$$RMSE^T = \left[ \frac{1}{N} \sum_{i=1}^{N} \frac{1}{NR} \sum_{m=1}^{NR} \left( \frac{\hat{X}_{i,m}^T - X_i^T}{X_i^T} \right)^2 \right]^{\frac{1}{2}} \tag{23}$$

and

$$BIAS^T = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{NR} \sum_{m=1}^{NR} \left( \frac{\hat{X}_{i,m}^T - X_i^T}{X_i^T} \right). \tag{24}$$

Eqs. (23) and (24) are summations through each repetition of the Monte Carlo experiment (*m* = 1 to *NR*) and each of the analyzed stations (*i* = 1 to *N*); $RMSE^T$ and $BIAS^T$ are the root mean square error and the relative bias for return period *T*, respectively; $X_i^T$ is the "real" value for the *T*-year event at site *i*, and $\hat{X}_{i,m}^T$ is the simulated value for the *T*-year event at site *i* from the *m*-th sample of Monte Carlo simulation.

Furthermore, the performance of different ROI options can be compared (1) with results of the former regional precipitation frequency analysis that has been carried out using the "traditional" regionalization approach of Hosking and Wallis (4 homogeneous regions within the Czech Republic); (2) with results of a regional frequency analysis where the whole country is treated as a one-and-only homogeneous region; and (3) with results of a traditional at-site frequency analysis without a regional approach.

## 4 Results

Results of the analysis will be presented at the conference.


### Acknowledgements

The work described in this paper was supported by a young-scientists' research grant from the Grant Agency of Academy of Sciences of the Czech Republic, under project ID B300420601. This support is gratefully acknowledged.


### References


ADAMOWSKI, K., ALILA, Y., PILON, P.J., 1996. Regional rainfall distribution for Canada. *Atmospheric Research* **42**(1-4): 75-88.

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., 2001. Assessing the effectiveness of hydrological similarity measures for flood frequency analysis. *Journal of Hydrology* **241**: 270-287.

DALRYMPLE, T., 1960. *Flood frequency analyses*. Water Supply Paper 1543-A. U.S. Geological Survey, Reston.

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**: 1313-1334.

GAÁL, L., 2005. Introduction of Lapin's indices into the cluster analysis of maximum k-day precipitation totals in Slovakia. *Meteorologický časopis - 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**: 113-126.

HOLMES, M.G.R., YOUNG, A.R., GUSTARD, A., GREW, R., 2002. A region of influence approach to predicting flow duration curves within ungauged catchments. *Hydrology and Earth System Sciences* **6**(4): 721-731.

HOSKING, J.R.M., WALLIS, J.R., 1993. Some statistics useful in regional frequency analysis. *Water Resources Research* **29**: 271-281.

HOSKING, J.R.M., WALLIS, J.R., 1997. *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press, Cambridge; New York; Oakleigh. 224 p. ISBN 0-521-43045-3.

KJELDSEN, T.R., SMITHERS, J.C., SCHULZE, R.E., 2002. Regional flood frequency analysis in the KwaZulu-Natal province, South Africa, using the index-flood method. *Journal of Hydrology* **255**(1-4): 194-211.

KOHNOVÁ, S., GAÁL, L., SZOLGAY, J., HLAVČOVÁ, K., 2005. *Analýza maximálnych úhrnov zrážok v povodí horného Hrona*. Stavebná fakulta, Slovenská technická univerzita v Bratislave. 1. vyd. 120 s. ISBN 80-227-2339-8.

KYSELÝ J., PICEK J., HUTH R., 2006. Formation of homogeneous regions for regional frequency analysis of extreme precipitation events in the Czech Republic. *Studia Geophysica et Geodaetica*, accepted.

KYSELÝ, J., PICEK, J., 2006. Regional design value estimates of extreme precipitation events in the Czech Republic. *Proc. of Bioclimatological workshop, Strečno, Slovakia*, this volume.

LAPIN, M., MELO, M., 2005. Priestorová interpretácia výstupov klimatických scenárov v povodí Hrona a Váhu geoštatistickými metódami. In: *Pekárová, P. a kol. (2005): Scenáre zmien vybraných zložiek* *hydrosféry a biosféry v povodí Hrona a Váhu v dôsledku klimatickej zmeny*. VEDA, Bratislava, 49-80. ISBN 80-224-0884-0.

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**: 223-245.

PILON, P.J., ADAMOWSKI, K., ALILA, Y., 1991. Regional analysis of annual maxima precipitation using L-moments. *Atmospheric Research* **27**: 81-92.

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. *Journal of Hydrology* **241**: 42-52.

SVEINSSON, O.G.B., SALAS, J.D., BOES, D.C., 2002. Regional frequency analysis of extreme precipitation in Northeastern Colorado and Fort Collins Flood of 1997. *Journal of Hydrologic Engineering* **7**(1): 49-63.

WEISSTEIN, E.W., 2006 [online]. *Pearson's Skewness Coefficients*. From *MathWorld* - A Wolfram Web Resource. http://mathworld.wolfram.com/PearsonsSkewnessCoefficients.html. Last modified 2002-11-26. [cit. 2006-06-20]

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

ZRINJI, Z., BURN, D.H., 1994. Flood frequency analysis for ungauged sites using a region of influence approach. *Journal of Hydrology* **153**(1-4): 1-21.