Hydrol. Earth Syst. Sci. Discuss., 6, 273–317, 2009
www.hydrol-earth-syst-sci-discuss.net/6/273/2009/
© Author(s) 2009. This work is distributed under
the Creative Commons Attribution 3.0 License.

Hydrology and Earth System Sciences Discussions

Papers published in *Hydrology and Earth System Sciences Discussions* are under open-access review for the journal *Hydrology and Earth System Sciences*

Regional frequency analysis of heavy precipitation in the Czech Republic by improved region-of-influence method

L. Gaál[1,2] **and J. Kyselý**[1]

[1]Institute of Atmospheric Physics, Academy of Sciences of the Czech Republic, Boční II 1401, 141 31 Prague 4, Czech Republic
[2]Department of Land and Water Resources Management, Faculty of Civil Engineering, Slovak University of Technology, Radlinského 11, 813 68 Bratislava, Slovakia

Received: 16 December 2008 – Accepted: 18 December 2008 – Published: 14 January 2009

Correspondence to: L. Gaál (ladislav.gaal@stuba.sk)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

The region-of-influence (ROI) approach is implemented for modelling the probabilities of heavy 1-day and 5-day precipitation amounts in the Czech Republic. Compared to the previous study for Slovakia (Gaál et al., 2008), an improved ROI methodology that makes use of a regional homogeneity test for assigning sites to pooling groups, is applied. Two basically different approaches to forming the pooling groups in which groups of sites are gradually built up and cut down, respectively, are combined. A sensitivity analysis, which examines the performance of the ROI models after changing the set of input site attributes that serve for determining the sites' similarity, is carried out. Finally, several frequency models that include the ROI pooling schemes, a conventional model based on fixed regions, and an at-site analysis are compared by means of Monte Carlo simulations. We conclude that, regardless of the duration of precipitation, the ROI pooling scheme based on the actual proximity of sites (latitude and longitude) outperforms the other frequency models in terms of the root mean square error of the growth curves.

1 Introduction

Frequency analysis, which aims at estimating recurrence intervals of heavy hydroclimatological phenomena, is a specific field of applied statistics, which has been intensively developed over recent decades. Frequency analysis usually benefits from a regional approach, which is applicable if the regional homogeneity criterion is met; that is, the sites that form a given region share the same distribution function of the examined variable apart from a site-specific scaling factor called the index value (Dalrymple, 1960). Different aspects of the regional approach to frequency analysis have been examined in connection with heavy precipitation (e.g. Gellens, 2002; Sveinsson et al., 2002; Fowler and Kilsby, 2003; Boni et al., 2006; Wallis et al., 2007), floods (e.g. Burn, 1997; Madsen and Rosbjerg, 1997; Adamowski, 2000; Kjeldsen et al., 2002; Jingyi

HESSD

6, 273–317, 2009

and Hall, 2004; Solín, 2008), droughts (e.g. Clausen and Pearson, 1999; Chen et al., 2006), extreme sea levels (e.g. van Gelder et al., 2000) and wind speeds (e.g. Sotillo et al., 2006; Modarres, 2008).

The superiority of regional frequency models over the conventional at-site approach (which only utilizes data from the site of interest itself) stems not only from the reduced uncertainty of the estimated high quantiles at the right tails of the distributions (e.g. Lettenmaier et al., 1987; Cunnane, 1988; Stedinger et al., 1993), but also from the fact that the regional methods allow for the estimation of design values at ungauged sites (e.g. GREHYS 1996a, b; Kohnová et al., 2006b).

In the traditional approach to regional frequency analysis, the regions are kept fixed; that is, when changing the focus from one site to another within a given region, the information source for the regional transfer remains unchanged (e.g. Hosking and Wallis, 1997). An alternative to regional frequency estimation, the region-of-influence approach (ROI; Burn, 1990a, b), introduced a basically different concept: the idea of focused pooling. Its main feature is the uniqueness of the "regions" (more precisely, the pooling groups – Reed et al., 1999b): each site under study has its own group of adequately similar sites that form the basis for the transfer of information on extremes to the site of interest. The idea of focused pooling has been adopted in studies of flood flows (e.g. Zrinji and Burn, 1994, 1996; Castellarin et al., 2001; Cunderlik and Burn, 2002; Holmes et al., 2002; Shu and Burn, 2004) and precipitation extremes (Schaefer, 1990; Alila, 1999; Di Baldassare et al., 2006), as well as in complex nationwide projects devoted to the frequency analysis of hydro-climatological extremes (Reed et al., 1999a; Thompson, 2002).

In an analysis of extreme precipitation amounts in Slovakia, Gaál et al. (2008) adopted an original concept of the ROI approach (Burn, 1990b), which, however, had been subjected to criticism due to the need to choose a relatively large number of parameters according to subjective considerations (e.g. Hosking and Wallis, 1997). Zrinji and Burn (1994) revisited the ROI methodology: instead of subjectively selected threshold values, a built-in regional homogeneity test was used for assigning sites to a

HESSD
6, 273–317, 2009

given pooling group. They employed χ^2_R statistics (Chowdhury et al., 1991) for testing the regional homogeneity of the proposed pooling groups. Later, Zrinji and Burn (1996) extended the ROI methodology by a hierarchical feature (Gabriele and Arnell, 1991), which implemented different alternatives to the homogeneity test of Hosking and Wallis (1993). Castellarin et al. (2001) applied the hierarchical pooling methodology of Zrinji and Burn (1996) for a flood frequency analysis in northern central Italy.

In this study we modify the pooling methodology of Castellarin et al. (2001), carry out a sensitivity analysis by examining the performance of the ROI models after changing the set of input site attributes that serve for determining the sites' similarity, and compare the improved ROI model with other frequency models by means of simulation experiments. The methods are applied to modelling probabilities of extreme one-day and multi-day precipitation amounts in the Czech Republic (central Europe).

2 Data

2.1 Precipitation data

Daily precipitation totals measured at 209 stations mostly operated by the Czech Hydrometeorological Institute (CHMI) were used as the input dataset (Fig. 1). The altitudes of the stations range from 150 to 1490 m a.s.l., and the observations at most sites span the period from 1961 to 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 any location changes exceeded 50 m in altitude were excluded from the analysis), and no other sources of inhomogeneities were reported;
- the daily series of precipitation records are uninterrupted (except for the sites discussed below).

CC BY

The data underwent standard quality checking for gross errors (cf. Coufal et al., 1992). A large majority of the station records cover the whole period of 1961–2005; 36 out of the 209 stations have daily data over shorter subperiods of at least 31 consecutive years (mostly between 38 and 43 years; the stations started to operate after 1961 or closed before 2005) and/or minor parts of the records had to be omitted owing to the stations' relocations.

The dataset is superior to the one employed in Kyselý and Picek (2007a), especially since it involves a much larger number of sites with complete daily records, more evenly covers the area of the Czech Republic, extends to the very recent past (December 2005), and a few errors in the original dataset were corrected (a missing month was identified in the records of 3 stations, and the data were supplemented).

At 45 stations, minor gaps in the daily records occurred (a total of up to 1 month over 45 years at 32 sites and not exceeding 3 months at any of the 45 sites). We decided to preserve these stations in the analysis because of their locations in areas that are insufficiently covered by rain-gauges with complete records. The missing daily data were estimated using measurements at 2 to 5 nearest locations available in the climatological database of the CHMI; the methodology is described in Kyselý (2008). (Note that the mean distance to the nearest measuring site was only 15.4 km for the locations where the missing data were estimated, and the percentage of the missing daily records in the entire dataset was only 0.05%.) All other station records with more than 3 months of missing values were excluded from the analysis.

The basic features of the precipitation climate of the Czech Republic, with a focus on extremes, may be found in Kyselý and Picek (2007a) and Kyselý (2008).

Samples of the annual maxima of 1-day and 5-day precipitation amounts were drawn from each station record and are further examined.

2.2 Alternatives of site attributes

The ROI approach as one of the methods of focused pooling techniques is aimed at finding groups of sites that share similar statistical properties of the observed hydro-

CC BY

climatological extremes. It is assumed that the frequency distribution of the extremes at a given site is closely related to its climatological, hydrological, geographical, geomorphological, etc., attributes. Therefore, one of the basic issues of the pooling procedures is the selection of site attributes that are useful for explaining the observed behaviour of the extremes.

The similarity of sites in the pooling process is evaluated using two different sets of site attributes in this study.

The first group of site attributes consists of general climatological characteristics (abbr. ROIcli) that describe a long-term precipitation regime. The following variables are considered to be the basic characteristics of the precipitation climate:

1. mean annual precipitation [mm],
2. mean ratio of the precipitation totals for the warm/cold seasons [–], and
3. mean annual number of dry days [–] (defined as days with a precipitation amount ≥0.1 mm).

The warm (cold) season is defined as April–September (October–March). The basic idea of the choice of the characteristics of the precipitation climate is that the atmospheric mechanisms generating heavy precipitation are similar under similar climatological conditions, particularly when the small extent of the study area is taken into account.

Geographical site characteristics (abbr. ROIgeo) represent the second group of attributes that are employed to define the sites' proximity:

1. latitude [degrees],
2. longitude [degrees] and
3. elevation above sea level [m].

The geographical co-ordinates are chosen since the actual proximity of the sites may also result in similar regimes of extreme precipitation.

HESSD
6, 273–317, 2009

3 Methods

3.1 Description of the pooling scheme

Since the pooling scheme adopted herein originates from that described in the details in Gaál et al. (2008), we confine the description to the cornerstones of the procedure and accentuate the changes and improvements in the methodology.

The similarity of sites in the attribute space is evaluated by means of a weighted Euclidean distance metric:

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

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. Before determining the elements D_{ij} of the distance metric or dissimilarity matrix **D**, the attributes undergo normalization in order to remove any possible bias from the estimation due to different magnitudes. The weight W_m in Eq. (1) is used to express the relative importance of the site attributes. In the current analysis, we use unit weighting coefficients for all the attributes (W_m=1, m=1, …, M), because we did not find any reasons to prefer one site attribute over another.

It is important to point out the difference between two types of the site attributes, which are usually termed as characteristics and statistics. Site characteristics are quantities independent of whether or not daily measurements of precipitation are carried out at a given site. These include geographical co-ordinates, geomorphological attributes and/or descriptors of the long-term (precipitation) climate. On the other hand, site statistics are the results of the statistical processing of the data observed at a given site. It is generally recommended (Hosking and Wallis, 1997; Castellarin et al., 2001) to use site characteristics in the process of the formation of the regions or pooling

HESSD
6, 273–317, 2009

groups, while one should take advantage of site statistics in the process of testing the homogeneity of a proposed group of sites.

Pooling groups in the ROI approach are generally constructed using elements D_{ij} arranged in ascending or descending order; however, there are basically two different ways as to how to accomplish this. The core idea of the first method lies in the gradual building up of the pooling groups (termed as the "forward" approach herein). Starting with the target site i, which represents a single-site pooling group at the very beginning of the process, the next closest site (i.e. the site j with the next lowest value of D_{ij}, $j=1, \ldots, N$) is appended to the existing ROI in each turn, until a given condition for forming the ROI is met. The process of building up the ROI is terminated (i) at a given point, which is defined as a function of the selected quantiles of the dissimilarity matrix **D** (Burn, 1990b) or (ii) when the measure of the regional homogeneity of the proposed group of sites reaches or exceeds an unacceptable level (Castellarin et al., 2001). A reversed "backward" procedure is adopted in the second method of pooling: in its initial stage, all the sites in the analysis are supposed to form a superregion, so step by step, the most dissimilar sites are removed from the bulk of the sites until the remaining group of sites is homogeneous (Zrinji and Burn, 1994).

We implemented the regional homogeneity test proposed by Lu and Stedinger (1992) when forming the regions, for two reasons: (i) its application is computationally straightforward, and (ii) according to the comparative study of Fill and Stedinger (1995), it is one of the most powerful homogeneity tests. A brief description of Lu and Stedinger's homogeneity test, also called the X10 test, is given in the Appendix B.

We tested both the "forward" and "backward" approaches to forming homogeneous ROIs and decided to form the ROIs primarily by gradual building up ("forward") and using a homogeneity test for finding the cutoff point for the inclusion of the sites. The main deficiency of the "backward" procedure was that in some cases, it tended to produce very large homogeneous regions, which did not vary much from site to site; this resulted in the undesirable spatial smoothing of the estimated growth curves in the regional analysis (cf. Castellarin et al., 2001).

CC BY

As mentioned above, the basic idea of the "forward" pooling procedure is an iterative building up of the ROI for the site of interest. In each step of the iteration, the next similar site is added to the existing ROI, and the homogeneity of the proposed pooling group is tested. In the event the proposed ROI is homogeneous, the procedure goes on with the next loop; otherwise, the procedure is stopped, and the formation of the given ROI is finished.

After a detailed scrutiny of the preliminary outcomes of the analysis, we found it useful to slightly modify this pooling procedure. In a number of cases, the heterogeneity had been reached relatively early, i.e. after a few (<4–6) iterations. Neither was the adoption of the idea suggested by Castellarin et al. (2001) helpful, to stop the iteration procedure when the heterogeneity of the ROI is detected for the second time (instead of the first time). In both cases, the difficulty is that a relatively high number of pooling groups consists of a small number of sites and do not meet the "5T rule" (Jakob et al., 1999), which suggests that for a reliable estimation of a design value corresponding to the return period T, one needs 5 times T station-years of data. Considering the fact that the average length of observations at the selected stations is ~44 years (see Sect. 2.1), it is desirable to have at least 10–11 sites included in a pooling group for a reliable estimation of the 100-year precipitation quantiles.

We modified the pooling procedure of Castellarin et al. (2001) as follows:

- At the very beginning, the ROI of the target site consists of 11 stations (i.e. it comprises the site itself and the 10 closest sites), regardless of whether this initial pooling group is homogeneous or not.
- The iteration procedure of testing the homogeneity and adding the next closest site to the pooling group goes on until the first heterogeneous pooling group is found preceded by at least one homogeneous pooling group. In this case, the last homogeneous stage defines the final composition of the pooling group for the site.
- In the case when no homogeneous stage is reached by successively adding sites

to the initial ROI, the program code returns to the initial stage with 11 sites and starts looking for a homogeneous composition by removing the least similar sites from the pooling group. The first homogeneous stage then defines the final composition of the pooling group for the site. In the worst case, when the building nor the removing procedure leads to a homogeneous stage, the ROI consists of nothing but the target site (i.e. a "single-site pooling group").

3.2 Frequency model of Hosking and Wallis

The classical regional analysis of Hosking and Wallis (1993, 1997) consists in delineating fixed regions that are homogeneous according to the statistical characteristics of the probability distributions of the extremes, i.e. the at-site distributions are identical except for a site-specific scaling factor.

The delineation of homogeneous regions for 1-day and 5-day precipitation extremes in the Czech Republic has been updated and modified with respect to the original one presented in Kyselý and Picek (2007a). The reason for revisiting the original results of the homogeneity tests was a new dataset of daily precipitation amounts, which consists of 209 stations (compared to 78 in Kyselý and Picek, 2007a), extends to a more recent past (2005), and altogether covers 9197 station-years (compared to the original 3120). The two largest regions of the original regionalization became heterogeneous when the new data were considered; they have been split into 3 and 2 smaller (homogeneous) areas. The new regionalization recognizes 9 regions (Fig. 1) that are homogeneous according to the X10 test of Lu and Stedinger (1992) as well as the H1 test of Hosking and Wallis (1993) with respect to the statistical distributions of the annual maxima of the 1-day to 7-day precipitation amounts.

3.3 Estimation of growth curves and quantiles

For the construction of regional (pooled) growth curves and the estimation of the precipitation quantiles, the generalized extreme value (GEV) distribution was applied (see

HESSD
6, 273–317, 2009

Appendix A). The GEV distribution is widely used for modelling hydro-climatological extremes worldwide (e.g. Alila, 1999; Smithers and Schulze, 2001; Castellarin et al., 2001; Kharin and Zwiers, 2000), and frequency analyses of precipitation extremes have also confirmed its applicability in central Europe (the Czech Republic – Kyselý and Picek, 2007a; Slovakia – Kohnová et al., 2006a).

The growth curves and precipitation quantiles are estimated using the L-moment-based index storm procedure (Hosking and Wallis, 1997). In the initial step, dimensionless data are calculated by rescaling by the sample mean μ_j (index storm):

$$x_{j,k} = \frac{X_{j,k}}{\mu_j}, \quad j = 1, \ldots, N, \quad k = 1, \ldots, n_j, \tag{2}$$

where $X_{j,k}$ ($x_{j,k}$) denotes the original (dimensionless, rescaled) data; N is the number of sites; and n_j denotes the sample size of the j-th site.

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

$$t^{(j)} = l_2^{(j)} \Big/ l_1^{(j)} \tag{3}$$

and

$$t_r^{(j)} = l_r^{(j)} \Big/ l_2^{(j)}, \; r = 3, 4, \ldots, \tag{4}$$

where $t^{(j)}$ is the sample L-coefficient of variation (L-CV) and $t_r^{(j)}$, r=3, 4, ... are the sample L-moments ratios at site j (Hosking, 1990).

The pooled (regional) L-moment ratios $t^{(i)R}$ and $t_r^{(i)R}$, r=3, 4, ... for the target site i are derived from the at-site sample L-moment ratios as their weighted averages:

$$t^{(i)R} = \frac{\sum_{j=1}^{N} W_{ij} t^{(j)}}{\sum_{j=1}^{N} W_{ij}}, \tag{5}$$

CC BY

where W_{ij} are the weights associated with the j-th site in the analysis. The relationships analogous to Eq. (5) also hold true for $t_r^{(i)R}$, r=3, 4, … .

From a mathematical point of view, the most important difference between the traditional regionalization and the focused pooling consists in the way the weighting coefficients W_{ij} in Eq. (5) are defined. In the traditional regional analysis, W_{ij} are proportional only to the record length n_j for all the sites j within a given region:

$$W_{ij} = \begin{cases} n_j & \forall j \in R_i \\ 0 & \forall j \notin R_i \end{cases}, \quad (6)$$

where R_i denotes the region to which the target site i belongs. Equation (6) is in accordance with the concept of Hosking and Wallis (1997): sites with longer observations provide more information in the regionally averaged statistics. Note that the regional weighting coefficients in Eq. (6) do not change when changing the focus from one site to another within a given region.

In the focused pooling, the length of observations n_j is retained in the relationship for W_{ij}; however, compared to Eq. (6), the reciprocal value of the distance metric element D_{ij} is introduced as an additional factor (Castellarin et al., 2001):

$$W_{ij} = \begin{cases} n_j \Big/ D^*_{ij} & \forall j \in \mathrm{ROI}_i \\ 0 & \forall j \notin \mathrm{ROI}_i \end{cases}, \quad (7)$$

where ROI_i stands for the region of influence of the site i, and

$$D^*_{ij} = \begin{cases} D_{ij} & \text{if } D_{ij} \neq 0 \\ D_{ij,\min} & \text{if } D_{ij} = 0 \end{cases}, \quad (8)$$

where $D_{ij,\min}$ is the lowest non-zero value of the distance metric between the target site i and all the other sites j (Castellarin et al., 2001). (The expression for D^*_{ij} in Eq. (8) is more complex in order to avoid inconsistent results for $i=j$: in this special case $D_{ii}=0$, which would lead to $W_{ii}=\infty$.) Using the reciprocal value of the distance metric element

HESSD
6, 273–317, 2009

D_{ij} as the pooled weighting factor is equivalent to assigning higher weights to sites that lie in the proximity of the target site in the attribute space: the smaller D_{ij} for a given site j is, the greater the amount of information it brings to the procedure for the growth curve estimation at site i.

The regionally weighted (pooled) L-moment ratios $t^{(i)R}$ and $t_r^{(i)R}$, r=3, 4, ... are then used to estimate the parameters of the GEV distribution and the dimensionless cumulative distribution function (the growth curve). A quantile corresponding to the return period T is calculated as a product of the dimensionless T-year growth curve value x_i^T and the index storm μ_i:

$$X_i^T = \mu_i x_i^T . \quad (9)$$

3.4 Framework for an inter-comparison of the frequency models

In this paper, ROI pooling schemes based on climatological and geographical site attributes (Sect. 2.2) are compared with other frequency models, which include (i) the conventional regionalization approach of Hosking and Wallis (1997; abbr. HWreg), and (ii) the at-site frequency analysis. The performance of the different frequency models is assessed by means of Monte Carlo simulation procedures.

The essential issue of the Monte Carlo simulation is the way the unknown parent distribution (the "true" distribution) of the extremes is estimated. We decided to estimate the "true" at-site distribution by adopting the region-of-influence approach in which the similarity of sites is determined according to the statistical properties of the at-site data samples (abbr. ROIsta), as in Castellarin et al. (2001) and Gaál et al. (2008). Three site statistics were selected (cf. Burn, 1990b; Gaál et al., 2008):

1. the coefficient of variation: $c_v = \sigma / \mu$, where $\mu(\sigma)$ is the sample mean (standard deviation);
2. Pearson's 2nd skewness coefficient: $PS = 3(\mu - m)/\sigma$, where m is the sample median (Weisstein, 2002);

CC BY

3. the normalized 10-year precipitation quantile, estimated using the GEV distribution (x^{10y}).

The selected statistics characterize the scale (c_v), shape (PS) and location (x^{10y}) of the empirical distribution of the samples. A pooling scheme based on the site statistics is supposed to result in groups of sites that have a frequency distribution of extremes similar to the target site.

In each loop of the Monte Carlo procedure, samples of the annual maxima that resemble the real world (the actual number of sites, the length of the observations, and the spatial correlations between the sites) are simulated. At each site, the parent distribution is the GEV; its parameters correspond to the pooled L-moments according to the ROIsta pooling scheme. Having simulated the at-site samples, the pooling schemes and frequency models described above are applied to estimate the T-year quantiles of the precipitation extremes, which are then compared with the "true" quantiles obtained by the ROIsta pooling scheme. The loops of the Monte Carlo simulations are repeated 5000 times.

The different frequency models are compared by means of the bias and the root mean square error (RMSE) statistics. For a given return period T,

$$\mathrm{RMSE}^T = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{M}\sum_{m=1}^{M}\left(\frac{\hat{x}_{i,m}^T - x_i^T}{x_i^T}\right)^2\right]^{\frac{1}{2}} \tag{10}$$

and

$$\mathrm{BIAS}^T = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{m=1}^{M}\left(\frac{\hat{x}_{i,m}^T - x_i^T}{x_i^T}\right), \tag{11}$$

where i (m) is the index over the sites (repetitions); N (M) is the number of sites (repetitions); x_i^T is the "real" T-year value at site i; and $\hat{x}_{i,m}^T$ is the estimated T-year value at site i from the m-th sample of the Monte Carlo simulation.

HESSD
6, 273–317, 2009

The Monte Carlo simulation procedure and some related considerations are described in more detail in Gaál et al. (2008, Sect. 4).

4 Results

4.1 Sensitivity analysis

A sensitivity analysis was performed in order to explore the role of different site characteristics and site statistics entering the dissimilarity matrix **D** (Eq. 1) and to identify the optimum setting of the ROI pooling schemes. The basic ROI schemes were analogous to those used in Gaál et al. (2008); the models were based on 3 climatological (geographical) site characteristics (Sect. 2.2), and labelled as ROIcli3 (ROIgeo3); both were connected with the model ROIsta based on 3 site statistics (ROIsta3) used for the estimation of the "true" quantiles during the simulation procedures (Sect. 3.4). The sensitivity analysis examined the performance of the ROI models after removing one or two site attributes from the basic ROI pooling schemes (ROIcli3, ROIgeo3) or from the "true" frequency model (ROIsta3). The analysis was divided into two parts: (i) to examine the effects of changes made to the basic ROI schemes, while keeping the "true" model unchanged, and (ii) to examine the effects of changes in the "true" frequency model, while using the basic ROI schemes with 3 parameters. The different alternatives of the newly constructed ROI pooling schemes and the modified "true" frequency models are summarized in Table 1. Note that the number of alternatives to the ROIgeo models is reduced: While the ROIcli alternatives make use of all 6 possible combinations of the 3 available climatological attributes into singles (labelled as ROIcli1a, b and c) or pairs (labelled as ROIcli2a, b and c), there are no reasons for using other simplified ROIgeo models than those based purely on elevation (ROIgeo1) or the pair of geographical co-ordinates (ROIgeo2). Furthermore, the modified "true" frequency models are based only on pairs of possible combinations of the site statistics defined in Sect. 3.4 (labelled with the suffixes "2Sa", "2Sb" and "2Sc") since it is unreasonable

HESSD
6, 273–317, 2009

to construct "true" models based purely on one statistic. The sensitivity analysis was performed for both data sets of the 1-day and 5-day annual maxima.

First, we focus on the consequences of the changes made to the basic pooling schemes ROIcli3 and ROIgeo3. The summary statistics of the models' performance in terms of the average RMSE and the average bias for the quantiles of the estimated distributions of the 1-day (5-day) maxima corresponding to T=5, 10, 20, 50, 100 and 200 years are given in Table 2 (Table 3); the box-and-whisker plots of the RMSE in Fig. 2 (Fig. 3) illustrate the spread statistics. In general, growth curves of the various models do not differ much in terms of the bias: that is why the box plots of the bias are not shown herein.

It is obvious that the ROI models based purely on a single site attribute show a very poor degree of performance: the average RMSE is approximately 50% higher (in some cases, especially for the 5-day duration, by ~100% – Table 3) than for the best ROI models based on more than one site attribute. These results are in accordance with one's expectation: the frequency behaviour of the precipitation extremes cannot be explained by a single climatological characteristic. The ROIcli2 models based on two site attributes clearly perform better; however, the best average and spread characteristics of the RMSE are obtained for the basic ROIcli3 model for both data sets (for 5-day durations, there are three models based on ≥2 attributes with a comparable performance; the odd one is the ROIcli2b model, which suggests that the missing site characteristic of the warm/cold season precipitation ratio (Sect. 2.2) plays an important role in the other ROIcli models).

While for the ROIcli models, more site attributes improve performance, a similar conclusion cannot be drawn for the ROIgeo pooling schemes: the ROIgeo2 model always outperforms the basic ROIgeo3 model (Tables 2–3, Figs. 2–3). Such behaviour is accounted for by the role of the elevation in ROIgeo3: while in ROIgeo2, sites are pooled according to the geographical distance from the site of interest, ROIgeo3 gives preference to sites that are located in similar altitudes as the target one.

We demonstrate this with an example of two selected stations, Červená and Olo-

HESSD
6, 273–317, 2009

mouc, which are located not far from each other (40 km) but in different altitudes (750 and 225 m a.s.l., respectively). Pooling groups for the 1-day annual maxima at these stations according to both models are shown in Fig. 4, and some basic statistical characteristics of the composition of the pooling groups are summarized in Table 4.

Figure 4a (Fig. 4b) illustrates the pooling group for the Červená station according to ROIgeo2 (ROIgeo3). There is a relatively clear pattern of the distribution of the sites belonging to the ROI around the target site in ROIgeo2 (Fig. 4a): the further apart they are located, the less weight is assigned to them, and the weights are obviously independent on the elevation. A larger spread of sites in a broader neighbourhood of the target station appears in ROIgeo3 (Fig. 4b): sites located at similar altitudes are preferred over those in the very proximity of the target site, which is also underpinned by the fact that even sites from over 200 km are included in the ROI. Similar spatial patterns are observed for the Olomouc station (Figs. 4c–d) when comparing ROIgeo2 and ROIgeo3, although the differences between the two pooling groups are not that large, mainly because many neighbouring stations are also located at similar altitudes. Generally, when comparing the pooling groups in both models, the pooling groups of the ROIgeo3 model exhibit a greater average distance between the target site and the other sites within the ROI and a narrower elevation range (Table 4). Analogous findings hold true for the other pairs of stations, which are located not far from each other and at different altitudes.

The effects of the changes made to the "true" frequency model are summarized in Tables 5–6 and Figs. 5–6. Both the average values and spread statistics of the RMSE indicate that the best "true" model is the one based on all three statistical characteristics, regardless of the data set (1-day/5-day precipitation maxima) and the pooling scheme (ROIcli3/ROIgeo3). Such a conclusion is also underpinned by the spread characteristics of the bias: the narrowest box-and-whisker plots are always related to the models based on the basic version of the "true" ROISsta3 model.

The performance of the modified "true" models with the suffixes "2Sa", "2Sb" and "2Sc" allows for the ranking of the individual site statistics according to their impor-

CC BY

tance. The ROIcli2Sc/ROIgeo2Sc models show the worst statistical properties, which suggests that the missing coefficient of variation plays the most important role among the selected site statistics in the "true" frequency models. On the other hand, the Pearson's 2nd coefficient is the least important statistic, since when it is ignored, the ROIcli2Sb/ROIgeo2Sb models perform nearly as well as the best "true" model.

Similar Monte Carlo simulations (examining changes made to the "true" frequency model) were also carried out for the ROIgeo2 pooling scheme (not shown); the general features of its performance in the light of "true" frequency models based on different input statistics are analogous to those for the basic ROIcli3 and ROIgeo3 pooling schemes.

4.2 Inter-comparison of the frequency models

Tables 7–8 and Figs. 7–8 summarize the performance of the frequency models for 1-day and 5-day precipitation amounts, which correspond to different concepts: the two superior ROI models (based on three climatological and two geographical characteristics, ROIcli3 and ROIgeo2, respectively – Sect. 4.1), the Hosking-Wallis approach based on fixed regions (HWreg – Sect. 3.2), and the at-site (local) estimates. The average values of the root mean square error in Tables 7–8 reveal that the at-site approach, regardless of the duration, is clearly inferior, since the RMSE statistics of the local model are worse nearly by a magnitude compared to the other models. The poor performance of the at-site approach is explained by the enhanced effects of the sampling fluctuations, which are reduced by the multi-site approach in the regional models/pooling schemes (cf. Hosking and Wallis, 1997; Gaál et al., 2008). That is why we focus on a comparison of the regional/pooling approaches hereafter.

The ROIgeo2 pooling scheme outperforms the other models in terms of the average RMSE for both durations (Tables 7–8). The spread statistics of the RMSE in Fig. 8 clearly confirm this fact for the 1-day maxima. For the 5-day maxima, the box-and-whisker plots for the ROIgeo2 and HWreg models show much more balanced performance; however, the lower values of the 75% and 95% quantiles of the RMSE for all

HESSD
6, 273–317, 2009

the return levels suggest that ROIgeo2 is superior for the 5-day maxima, too (the percentage of sites at which the RMSE is large is reduced more efficiently compared to the other models). Another disadvantage of the HWreg model is a slight tendency towards a negative bias for both 1-day and 5-day durations (Fig. 7).

The results suggest that the Hosking-Wallis regional analysis may compete with the ROI method based on geographical characteristics only for multi-day precipitation extremes, the spatial variability of which is less affected by random (sampling) variations and more closely linked to some regional patterns in central Europe, which are related to atmospheric circulation and orographic features. The regional differences in the distributions of the multi-day extremes reflect, for example, the varied influences of Mediterranean cyclones (which often produce heavy multi-day precipitation) between the eastern and western parts of the country (e.g. Kyselý and Picek, 2007b). For one-day precipitation extremes, which are mostly related to convective phenomena in the warm season (88% of one-day maxima occurs in April-September), the ROI method based on geographical characteristics is clearly superior to all other frequency models, including the Hosking-Wallis regional analysis.

On the other hand, it is worth noting that the Hosking-Wallis regional analysis outperforms the ROI model based on climatological characteristics for both durations of precipitation extremes, which contrasts with the results for Slovakia (Gaál et al., 2008). It is likely related to the choice of the climatological characteristics in Slovakia, particularly the availability of the Lapin's index of Mediterranean influence (Gaál, 2005), which is closely linked to the occurrence of heavy precipitation (see also Sect. 5 below); no analogous index is available for the area of the Czech Republic. Another reason may be a different approach to the delineation of the homogeneous regions for the Hosking-Wallis analysis in the two countries, with 3 contiguous regions in Slovakia compared to 9 regions in the Czech Republic, which also take into account the altitudinal zonality, and therefore not all of them are contiguous.

HESSD
6, 273–317, 2009

5 Discussion and conclusions

The paper deals with the estimation of growth curves of the annual maxima of 1-day and 5-day precipitation amounts in the Czech Republic by improved region-of-influence (ROI) methodology. The improvements consist in the way the pooling groups are constructed. The regional homogeneity test of Lu and Stedinger (1992) is incorporated in order to avoid subjective decisions concerning the parameters involved in the ROI methodology, and to avoid forming heterogeneous pooling groups for the estimation. The remaining parameter in the improved ROI method, the "baseline" number of sites in a pooling group, is chosen according to the "5T rule" (Jakob et al, 1999), i.e. a rule of thumb for the minimum number of sites within a pooling group needed for a reliable estimation of a T-year quantile (set to 100 years herein). The proposed ROI methodology combines two different concepts of constructing the pooling groups described previously in hydrological studies (the "backward" approach of Zrinji and Burn, 1994, and the "forward" approach of Castellarin et al., 2001), and preserves some of their beneficial features:

1. Reasonably large numbers of sites in the pooling groups typical for a pooling scheme based on the strategy of gradual building up (Castellarin et al., 2001); too large pooling groups tend to be formed by the "backward" approach, which may smooth spatial details. For example, applying the ROIgeo3 scheme for the 1-day maxima in the present study, the average number of stations in the pooling groups is 49.0, 49.3, and 70.5 according to the "forward" approach of Castellarin et al. (2001), our modified method, and the "backward" approach of Zrinji and Burn (1994), respectively.

2. Small numbers of the pooling groups with insufficient number of sites typical for a pooling scheme based on the strategy of cutting down (Zrinji and Burn, 1994). Using the ROIgeo3 scheme for the 1-day maxima again, the number of ROIs that do not meet the "5T rule" (i.e. small pooling groups of a size ≤11) is 40, 20, and 14, and the number of single-site pooling groups is 8, 1, and 1, according to

HESSD
6, 273–317, 2009

the "forward" approach of Castellarin et al. (2001), our modified method, and the "backward" approach of Zrinji and Burn (1994), respectively.

A sensitivity analysis, which examined the consequences of the changes made to the input attribute sets of the pooling schemes, confirmed a simple principle "the more input variables – the better performance" in the case of climatological site characteristics (used in the ROIcli models) and site statistics (used in the ROIsta pooling scheme as the "true" frequency model in the simulation procedure). On the other hand, in the case of geographical site characteristics (used in the ROIgeo models), the pooling scheme based on two co-ordinates (latitude and longitude) was found superior compared to the one that makes use of all three co-ordinates (including altitude). In general, however, both alternatives to the ROIgeo models have their own pros and cons. The main drawback of the ROIgeo3 pooling scheme is the tendency to pool sites from considerable distances away from the target site, while the disadvantage of the ROIgeo2 pooling scheme is that it pools sites regardless of their altitudinal zonality. However, the drawbacks of the ROIgeo2 model are less pronounced; therefore, it always outperforms the ROIgeo3 pooling scheme in terms of the RMSE of the estimated quantiles in the present application. An open question is whether some combination of the geographical and climatological characteristics would result in a model that outperforms the ROI schemes based on either geographical or climatological attributes, and whether there are other useful site attributes in addition to those examined; these issues go outside the scope of the study and deserve further investigation.

The main finding concerning the inter-comparison of various regional frequency models is that the ROI pooling scheme based on the actual proximity of sites (latitude and longitude) outperforms the other models (including the Hosking-Wallis regional analysis), regardless of the duration of the precipitation extremes. Such a conclusion, in general, is in good agreement with the findings of Gaál et al. (2008) who also showed the superiority of the pooling approach over the other frequency models, when analyzing the precipitation data in Slovakia. Nevertheless, they pointed out that different durations need different ROI approaches: while the pooling scheme based on geographical

attributes is preferable for 1-day maxima, the ROI model based on climatological attributes shows better statistical properties for 5-day durations. The difference between the main findings of the two studies may be related to the following causes:

- The more rugged terrain of Slovakia (the mountain range of the West Carpathians belt, including the High and Low Tatras, prolonged from west to east), which may enhance the role of climatological characteristics in the identification of similar patterns of the precipitation extremes;
- A different suite of climatological characteristics used in the analysis for Slovakia, with Lapin's index of Mediterranean influence (Gaál, 2005); no similar index of precipitation climate related to extremes is available for the Czech Republic;
- The poorer performance of the pooling schemes based on three geographical characteristics in Slovakia, which is likely due to the higher altitudinal variability of the selected sites in the country (a much larger percentage of higher-elevated sites at altitudes >1000 m a.s.l.);
- The denser network of the sites available in the Czech Republic (about one site per 400 km^2) compared to Slovakia (about one site per 900 km^2), which tends to give preference to the similarity of sites based on geographical proximity rather than climatological characteristics.

One may argue that in the present study, pooling schemes based on climatological characteristics show a poorer performance due to failing to choose the right attributes for the analysis. We used the same set of site characteristics (Sect. 2.2) that had constituted the basis for the identification of homogeneous regions in the traditional regionalization (Sect. 3.2; Kyselý and Picek, 2007a), in order to preserve consistency. Furthermore, the selected climatological attributes are among the most appropriate descriptors of the precipitation climate. Therefore, it may be more reasonable to expend efforts on finding a suitable combination of both climatological and geographical attributes for a single pooling scheme rather than searching for some more descriptive

climatological characteristics. An objective method for identifying the proper weights W_m for the site attributes in the dissimilarity matrix (Eq. 1) would have to be implemented for this purpose.

Appendix A

The generalized extreme value distribution

The cumulative distribution function of the generalized extreme value (GEV) distribution is

$$F(x;\xi,\alpha,k) = \begin{cases} \exp\left\{-\left[1-k\left(\frac{x-\xi}{\alpha}\right)\right]^{-1/k}\right\} & \text{if } k\neq 0 \\ \exp\left\{-\exp\left[-\left(\frac{x-\xi}{\alpha}\right)\right]\right\} & \text{if } k=0 \end{cases}, \quad \text{(A1)}$$

where ξ, α and k is the location, scale and shape parameter, respectively (Hosking and Wallis, 1997). The parameters satisfy $-\infty<\xi<\infty$, $\alpha>0$ and $-\infty<k<\infty$. For an estimation of the parameters, we use the approximation of Hosking et al. (1985):

$$k \approx 7.8590c + 2.9554c^2, \quad c = \frac{2}{3+t_3} - \frac{\log 2}{\log 3}, \quad \text{(A2)}$$

$$\alpha = \frac{l_2 k}{\left(1-2^{-k}\right)\Gamma(1+k)}, \quad \text{(A3)}$$

and

$$\xi = l_1 - \frac{\alpha}{k}\left[1-\Gamma(1+k)\right], \quad \text{(A4)}$$

where t_3 is the sample L-skewness, l_1 and l_2 are the first two sample L-moments (cf. Hosking and Wallis, 1997) and $\Gamma(\cdot)$ denotes the gamma function

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt. \quad \text{(A5)}$$

HESSD
6, 273–317, 2009

Appendix B

Regional homogeneity test by Lu and Stedinger

The regional homogeneity test by Lu and Stedinger (1992), which is known as the X10 test, is based on the sampling variance of the dimensionless 10-year precipitation x_{10} in a homogeneous region. It is assumed that the precipitation extremes follow the GEV distribution (Eq. A1). According to Fill and Stedinger (1995), the 10-year quantile of the growth curve of precipitation at the i-th site $x_{10}^{(i)}$ is estimated by means of L-moments as follows:

$$x_{10}^{(i)} = \begin{cases} 1 + \frac{t^{(i)}}{1-2^{-k}} \left(1 - \frac{(-\ln 0.9)^k}{\Gamma(1+k)}\right) & \text{if } k \neq 0 \\ 1 + 2.4139 t^{(i)} & \text{if } k = 0 \end{cases}. \quad \text{(B1)}$$

where $t^{(i)}$ is the sample L-Cv (Eq. 3 in Sect. 3.3) at the i-th site, and the shape parameter k is estimated by Eq. (A2).

The heterogeneity measure of the X10 test is then

$$\chi_R^2 = \sum_{i=1}^{N} \frac{\left(x_{10}^{(i)} - x_{10}^{R}\right)^2}{\mathrm{var} x_{10}^{(i)}}, \quad \text{(B2)}$$

where N is the total number of sites in the region,

$$x_{10}^{R} = \sum_{i=1}^{N} n_i x_{10}^{(i)} \Big/ \sum_{i=1}^{N} n_i \quad \text{(B3)}$$

is the weighted regional average of $x_{10}^{(i)}$ (with the weights proportional to the record length n_i), and $\mathrm{var} x_{10}^{(i)}$ is the asymptotic variance of $x_{10}^{(i)}$. The asymptotic variance is

CC BY

usually determined by means of simulations; however, Lu and Stedinger (1992) provide tables and small-sample correction factors for $\mathrm{var} x_{10}^{(i)}$.

The test statistic χ_R^2 has an approximate chi-square distribution with $N-1$ degrees of freedom. If $\chi_R^2 < \chi^2_{0.95,\,N-1}$, the null hypothesis is not rejected at the 0.05 level and the region may be considered homogeneous. In the opposite case, one rejects the null hypothesis and the region is considered heterogeneous (Lu and Stedinger, 1992).

Acknowledgements. The study was supported by the young-scientists' research project B300420801 of the Grant Agency of the Academy of Sciences of the Czech Republic. Thanks are due to P. Štěpánek, Czech Hydrometeorological Institute, for preparing the precipitation datasets; J. Picek, Technical University of Liberec, for providing the results of regional homogeneity tests for the Hosking-Wallis analysis; and O. Halásová, Czech Hydrometeorological Institute, for drawing Figs. 1 and 4.

References

Adamowski, K.: Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods, J. Hydrol., 229(3–4), 219–231, 2000.

Alila, Y.: A hierarchical approach for the regionalization of precipitation annual maxima in Canada, J. Geophys. Res., 104(D24), 31645–31656, 1999.

Boni, G., Parodi, A., and Rudari, R.: Extreme rainfall events: Learning from raingauge time series, J. Hydrol., 327(3–4), 304–314, 2006.

Burn, D. H.: An appraisal of the 'region of influence' approach to flood frequency analysis, Hydrol. Sci. J., 35(2), 149–165, 1990a.

Burn, D. H.: Evaluation of regional flood frequency analysis with a region of influence approach, Water Resour. Res., 26(10), 2257–2265, 1990b.

Burn, D. H.: Catchment similarity for regional flood frequency analysis using seasonality measures, J. Hydrol., 202(1–4), 212–230, 1997.

Castellarin, A., Burn, D. H., and Brath, A.: Assessing the effectiveness of hydrological similarity measures for flood frequency analysis, J. Hydrol., 241(3–4), 270–287, 2001.

Chen, Y. D., Huang, G., Shao, Q., and Xu, C.-Y.: Regional analysis of low-flow using L-moments for Dongjiang basin, South China, Water Resour. Res., 51(6), 1051–1064, 2006.

Chowdhury, J. U., Stedinger, J. R., and Lu, L.-H.: Goodness-of-fit tests for regional generalized extreme value flood distribution, Water Resour. Res., 27(7), 1765–1776, 1991.
Clausen, B. and Pearson, C. P.: Regional frequency analysis of annual maximum streamflow drought, J. Hydrol., 173(1–4), 111–130, 1995.
Coufal, L., Langová, P., and Míková, T.: Meteorological data in the Czech Republic in the period 1961-1990. National Climatic Programme, Czech Hydrometeorological Institute, Prague, 160 pp., 1992 (in Czech).
Cunderlik, J. M. and Burn, D. H.: Analysis of the linkage between rain and flood regime and its application to regional flood frequency estimation, J. Hydrol., 261(1–4), 115–131, 2002.
Cunnane, C.: Methods and merits of regional flood frequency analysis, J. Hydrol., 100(1–3), 269–290, 1988.
Dalrymple, T.: Flood frequency analyses, Water Supply Paper 1543-A., US Geological Survey, Reston, USA, 1960.
Di Baldassarre, G., Castellarin, A., and Brath, A.: Relationships between statistics of rainfall extremes and mean annual precipitation: an application for design-storm estimation in northern central Italy, Hydrol. Earth Syst. Sci., 10, 589–601, 2006, http://www.hydrol-earth-syst-sci.net/10/589/2006/.
Fill, H. D. and Stedinger, J. R.: Homogeneity tests based upon Gumbel distribution and a critical appraisal of Dalrymple test, J. Hydrol., 166(1–2), 81–105, 1995.
Fowler, H. J. and Kilsby, C. G.: A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000, Int. J. Climatol., 23, 1313–1334, 2003.
Gaál, L., Kyselý, J., and Szolgay, J.: Region-of-influence approach to a frequency analysis of heavy precipitation in Slovakia, Hydrol. Earth Syst. Sci., 12, 825–839, 2008, http://www.hydrol-earth-syst-sci.net/12/825/2008/.
Gaál, L.: Introduction of Lapin's indices into the cluster analysis of maximum k-day precipitation totals in Slovakia, Meteorol. J., 8(2), 85–94, 2005.
Gabriele, S. and Arnell, N.: A hierarchical approach to regional flood frequency analysis, Water Resour. Res., 27(6), 1281–1289, 1991.
Gellens, D.: Combining regional approach and data extension procedure for assessing of extreme precipitation in Belgium, J. Hydrol., 268(1–4), 113–126, 2002.
GREHYS (Groupe de recherche en hydrologie statistique): Presentation and review of some methods for regional flood frequency analysis, J. Hydrol., 186(1–4), 63–84, 1996a.
GREHYS (Groupe de recherche en hydrologie statistique): Inter-comparison of regional flood

frequency procedures for Canadian rivers, J. Hydrol., 186(1–4), 85–103, 1996b.
Holmes, M. G. R., Young, A. R., Gustard, A., and Grew, R.: A region of influence approach to predicting flow duration curves within ungauged catchments, Hydrol. Earth Syst. Sci., 6, 721–731, 2002, http://www.hydrol-earth-syst-sci.net/6/721/2002/.
Hosking, J. R. M. and Wallis, J. R.: Regional frequency analysis: an approach based on L-moments. Cambridge University Press, Cambridge, 224 pp., 1997.
Hosking, J. R. M. and Wallis, J. R.: Some statistics useful in regional frequency analysis, Water Resour. Res., 29, 271–281, 1993.
Hosking, J. R. M., Wallis, J. R., and Wood, E. F.: Estimation of the generalized extreme-value distribution by the method of probability-weighted moments, Technometrics, 27, 251–261, 1985.
Hosking, J. R. M.: L-moments: analysis and estimation of distributions using linear combinations of order statistics, J. Roy. Stat. Soc. B, 52(1), 105–124, 1990.
Jakob, D., Reed, D. W., and Robson, A. J.: Selecting a pooling-group, in: Flood Estimation Handbook, vol. 3, Institute of Hydrology, Wallingford, UK, 1999.
Jingyi, Z. and Hall, M. J.: Regional flood frequency analysis for the Gan-Ming River basin in China, J. Hydrol., 296(1–4), 98–117, 2004.
Kharin, V. V. and Zwiers, F. W.: Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM, J. Climate, 13, 3760–3788, 2000.
Kjeldsen, T. R., Smithers, J. D., and Schulze, R. E.: Regional flood frequency analysis in the KwaZulu-Natal province, South Africa, using the index-flood method, J. Hydrol., 255(1–4), 194–211, 2002.
Kohnová, S., Hlavčová, K., Szolgay, J., and Parajka, J.: On the choice of spatial interpolation method for the estimation of 1- to 5- day basin average design precipitation, in: Flood risk management: Hazards, vulnerability and mitigation measures, edited by: Schanze, J., Zeman, E. and Marsalek, J., NATO Science Series, 67, 77–89, 2006a.
Kohnová, S., Szolgay, J., Solín, L., and Hlavčová, K.: Regional methods for prediction in ungauged basins, Key Publishing, Ostrava, 113 pp., 2006b.
Kyselý, J. and Picek, J.: Regional growth curves and improved design value estimates of extreme precipitation events in the Czech Republic, Climate Res., 33, 243–255, 2007a.
Kyselý, J. and Picek, J.: Probability estimates of heavy precipitation events in a flood-prone central-European region with enhanced influence of Mediterranean cyclones, Adv. Geosci.,

12, 43–50, 2007b, http://www.adv-geosci.net/12/43/2007/.
Kyselý J.: Trends in heavy precipitation in the Czech Republic over 1961–2005, Int. J. Climatol., early view, available at: http://www3.interscience.wiley.com/journal/121566539/abstract, doi:10.1002/joc.1784, 2008.
Lettenmaier, D. P., Wallis, J. R., and Wood, E. F.: Effect of regional heterogeneity on flood frequency estimation, Water Resour. Res., 23(2), 313–323, 1987.
Lu, L.-H. and Stedinger, J. R.: Sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test, J. Hydrol., 138(1–2), 223–245, 1992.
Madsen, H. and Rosbjerg, D.: The partial duration series method in regional index-flood modeling, Water Resour. Res., 33(4), 737–746, 1997.
Modarres, R.: Regional maximum wind speed frequency analysis for the arid and semi-arid regions of Iran, J. Arid Environ., 72(7), 1329–1342, 2008.
Reed, D. W., Faulkner, D. S., and Stewart, E. J.: The FORGEX method of rainfall growth estimation II: Description, Hydrol. Earth Syst. Sci., 3, 197–203, 1999a, http://www.hydrol-earth-syst-sci.net/3/197/1999/.
Reed, D. W., Jakob, D., Robinson, A. J., Faulkner, D. S., and Stewart, E. J.: Regional frequency analysis: a new vocabulary, Hydrological Extremes: Understanding, Predicting, Mitigating, Proc. IUGG 99 Symposium, Birmingham, UK, IAHS Publ. no. 255, 237–243, 1999b.
Schaefer, M. G.: Regional analyses of precipitation annual maxima in Washington State, Water Resour. Res., 26(1), 119–131, 1990.
Shu, C. and Burn, D. H.: Homogeneous pooling group delineation for flood frequency analysis using a fuzzy expert system with genetic enhancement, J. Hydrol., 291(1–2), 132–149, 2004.
Smithers, J. C. and Schulze, R. E.: A methodology for the estimation of short duration design storms in South Africa using a regional approach based on L-moments, J. Hydrol., 241, 42–52, 2001.
Solín, L.: Analysis of floods occurrence in Slovakia in the period 1996–2006, J. Hydrol. Hydromech., 56(2), 95–115, 2008.
Sotillo, M. G., Aznar, R., and Valero, F.: Mediterranean offshore extreme wind analysis from the 44-year HIPOCAS database: different approaches towards the estimation of return periods and levels of extreme values, Adv. Geosci., 7, 275–278, 2006, http://www.adv-geosci.net/7/275/2006/.

Stedinger, J. R., Vogel, R. M., and Foufoula-Georgiou, E.: Frequency analysis of extreme events, in: Handbook of Hydrology, edited by: Maidment, D. R., McGraw-Hill, New York, USA, 1993.

Sveinsson, O. G. B., Salas, J. D., and Boes, D. C.: Regional frequency analysis of extreme precipitation in Northeastern Colorado and Fort Collins Flood of 1997, J. Hydrol. Eng., 7(1), 49–63, 2002.

Thompson, C. S.: The high intensity rainfall design system: HIRDS [Abstract], in: Int. Conf. on Flood Estimation, Bern, Switzerland, 6–8 March 2002.

van Gelder, P. H. A. J. M., de Ronde, J. G., Neykov, N. M., and Neytchev, P.: Regional frequency analysis of extreme wave heights: trading space for time. Proc. 27th ICCE, 1099–1112, vol. 2, Coastal Engineering, Sydney, Australia, 2000.

Wallis, J. R., Schaefer, M. G., Barker, B. L., and Taylor, G. H.: Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State, Hydrol. Earth Syst. Sci., 11, 415–442, 2007, http://www.hydrol-earth-syst-sci.net/11/415/2007/.

Weisstein, E. W.: Pearson's skewness coefficients. From MathWorld – A Wolfram Web Resource. Available at: http://mathworld.wolfram.com/PearsonsSkewnessCoefficients.html, last modified: 26 November 2002, access: 10 December 2008.

Zrinji, Z. and Burn, D. H.: Flood frequency analysis for ungauged sites using a region of influence approach, J. Hydrol., 153(1–4), 1–21, 1994.

Zrinji, Z. and Burn, D. H.: Regional flood frequency with hierarchical region of influence, J. Water Res. Pl.-ASCE, 122(4), 245–252, 1996.

Table 1. Summary of site characteristics and site statistics used in different ROI pooling schemes. The sign √(–) indicates that the given site characteristic or statistic is (is not) included in the pooling scheme.

		Climatological site chars.			Geographical site chars.			Site statistics		
		Mean annual precipitation [mm]	Warm/cold season precipitation ratio [–]	Mean annual nr. of dry days [–]	Latitude [°] Longitude [°]		Altitude [m]	Coefficient of variation	Pearson's 2nd skewness	Normalized 10-yr GEV-based quantile
Changes in the ROI pooling schemes	ROIcli1a	√	–	–	–	–	–	√	√	√
	ROIcli1b	–	√	–	–	–	–	√	√	√
	ROIcli1c	–	–	√	–	–	–	√	√	√
	ROIcli2a	√	√	–	–	–	–	√	√	√
	ROIcli2b	√	–	√	–	–	–	√	√	√
	ROIcli2c	–	√	√	–	–	–	√	√	√
	ROIcli3	√	√	√	–	–	–	√	√	√
	ROIgeo1	–	–	–	–	–	√	√	√	√
	ROIgeo2	–	–	–	√	√	–	√	√	√
	ROIgeo3	–	–	–	√	√	√	√	√	√
Changes in the "true" frequency model	ROIcli2Sa	√	√	√	–	–	–	√	√	–
	ROIcli2Sb	√	√	√	–	–	–	√	–	√
	ROIcli2Sc	√	√	√	–	–	–	–	√	√
	ROIgeo2Sa	–	–	–	√	√	√	√	√	–
	ROIgeo2Sb	–	–	–	√	√	√	√	–	√
	ROIgeo2Sc	–	–	–	√	√	√	–	√	√

HESSD
6, 273–317, 2009

Table 2. Performance of ROI pooling schemes based on different combinations of site characteristics as measures of similarity for 1-day annual precipitation maxima. $RMSE^T$ denotes the average root mean square error of the estimated growth curves corresponding to the return period T [years], expressed in %; the smallest values of the statistics are marked in bold.

	$RMSE^T$ [%]									
	Climatological site characteristics							Geographical site characteristics		
T [yrs]	ROIcli1a	ROIcli1b	ROIcli1c	ROIcli2a	ROIcli2b	ROIcli2c	ROIcli3	ROIgeo1	ROIgeo2	ROIgeo3
5	0.0122	0.0130	0.0129	0.0089	0.0084	0.0095	**0.0075**	0.0139	**0.0065**	0.0067
10	0.0492	0.0514	0.0490	0.0380	0.0383	0.0394	**0.0345**	0.0536	**0.0284**	0.0304
20	0.1635	0.1690	0.1628	0.1271	0.1277	0.1309	**0.1159**	0.1780	**0.0933**	0.0997
50	0.4710	0.4852	0.4721	0.3647	0.3635	0.3751	**0.3301**	0.5155	**0.2638**	0.2804
100	0.8496	0.8756	0.8564	0.6551	0.6491	0.6742	**0.5885**	0.9346	**0.4693**	0.4979
200	1.3794	1.4229	1.3983	1.0584	1.0427	1.0905	**0.9432**	1.5261	**0.7508**	0.7955

HESSD
6, 273–317, 2009

Table 3. Performance of ROI pooling schemes based on different combinations of site characteristics as measures of similarity for 5-day annual precipitation maxima. $RMSE^T$ denotes the average root mean square error of the estimated growth curves corresponding to the return period T [years], expressed in %; the smallest values of the statistics are marked in bold.

	$RMSE^T$ [%]									
	Climatological site characteristics							Geographical site characteristics		
T [yrs]	ROIcli1a	ROIcli1b	ROIcli1c	ROIcli2a	ROIcli2b	ROIcli2c	ROIcli3	ROIgeo1	ROIgeo2	ROIgeo3
5	0.0113	0.0093	0.0122	0.0054	0.0076	0.0052	**0.0047**	0.0127	**0.0054**	0.0057
10	0.0417	0.0352	0.0424	0.0221	0.0299	0.0211	**0.0202**	0.0453	**0.0200**	0.0227
20	0.1499	0.1268	0.1509	0.0839	0.1097	0.0808	**0.0784**	0.1610	**0.0711**	0.0812
50	0.4580	0.3858	0.4629	0.2592	0.3356	0.2506	**0.2437**	0.4913	**0.2144**	0.2450
100	0.8502	0.7120	0.8636	0.4779	0.6206	0.4629	**0.4497**	0.9128	**0.3933**	0.4491
200	1.4103	1.1720	1.4413	0.7821	1.0235	0.7583	**0.7354**	1.5173	**0.6424**	0.7338

Table 4. Basic characteristics of the pooling groups constructed according to two different pooling schemes ROIgeo2 and ROIgeo3 for two selected stations, Červená and Olomouc. The target sites are not included in the number of sites in the ROI.

	Červená (750 m a.s.l.)		Olomouc (225 m a.s.l.)	
	ROIgeo2	ROIgeo3	ROIgeo2	ROIgeo3
# of sites in the ROI	35	24	33	45
average distance [km]	59.5	127.8	69.9	85.1
minimum distance [km]	14.5	14.5	4.6	4.6
maximum distance [km]	120.5	244.9	122.8	187.4
average elevation [m]	363	561	376	290
minimum elevation [m]	197	435	197	180
maximum elevation [m]	1490	737	750	435

HESSD
6, 273–317, 2009

Table 5. Performance of ROI pooling schemes ROIcli3 and ROIgeo3 based on different combinations of site statistics in the "true" frequency model for 1-day annual precipitation maxima. $RMSE^T$ ($BIAS^T$) denotes the average root mean square error (average bias) of the estimated growth curves corresponding to the return period T [years], expressed in %; the smallest values of the statistics (in an absolute sense) are marked in bold.

	$RMSE^T$ [%]							
	Climatological site characteristics				Geographical site characteristics			
T [yrs]	ROIcli2Sa	ROIcli2Sb	ROIcli2Sc	ROIcli3	ROIgeo2Sa	ROIgeo2Sb	ROIgeo2Sc	ROIgeo3
5	0.0077	0.0082	0.0137	**0.0075**	0.0070	0.0074	0.0130	**0.0067**
10	0.0396	0.0453	0.0483	**0.0345**	0.0348	0.0411	0.0449	**0.0304**
20	0.1382	0.1420	0.1577	**0.1159**	0.1178	0.1259	0.1426	**0.0997**
50	0.3976	0.3800	0.4562	**0.3301**	0.3350	0.3324	0.4069	**0.2804**
100	0.7095	0.6572	0.8260	**0.5885**	0.5959	0.5709	0.7336	**0.4979**
200	1.1356	1.0293	1.3429	**0.9432**	0.9522	0.8893	1.1898	**0.7955**
	$BIAS^T$ [%]							
	Climatological site characteristics				Geographical site characteristics			
T [yrs]	ROIcli2Sa	ROIcli2Sb	ROIcli2Sc	ROIcli3	ROIgeo2Sa	ROIgeo2Sb	ROIgeo2Sc	ROIgeo3
5	0.4700	**0.4382**	0.5441	0.4482	0.4997	0.4763	0.5473	**0.4730**
10	0.3094	0.2909	0.4005	**0.2906**	0.4483	0.4329	0.4791	**0.4084**
20	0.0591	0.0623	0.1503	**0.0416**	0.3046	0.3048	0.3062	**0.2509**
50	−0.3365	−0.3001	**−0.2597**	−0.3586	0.0433	0.0694	**−0.0038**	−0.0330
100	−0.6517	−0.5911	**−0.5904**	−0.6843	−0.1771	**−0.1310**	−0.2630	−0.2760
200	−0.9606	**−0.8797**	−0.9141	−1.0112	−0.3985	**−0.3347**	−0.5208	−0.5258

HESSD
6, 273–317, 2009

Table 6. Performance of ROI pooling schemes ROIcli3 and ROIgeo3 based on different combinations of site statistics in the "true" frequency model for 5-day annual precipitation maxima. $RMSE^T$ ($BIAS^T$) denotes the average root mean square error (average bias) of the estimated growth curves corresponding to the return period T [years], expressed in %; the smallest values of the statistics (in an absolute sense) are marked in bold.

	$RMSE^T$ [%]							
	Climatological site characteristics				Geographical site characteristics			
T [yrs]	ROIcli2Sa	ROIcli2Sb	ROIcli2Sc	ROIcli3	ROIgeo2Sa	ROIgeo2Sb	ROIgeo2Sc	ROIgeo3
5	0.0093	0.0065	0.0106	**0.0047**	0.0100	0.0075	0.0105	**0.0057**
10	0.0305	0.0281	0.0313	**0.0202**	0.0318	0.0315	0.0336	**0.0227**
20	0.1305	0.1045	0.1309	**0.0784**	0.1243	0.1080	0.1330	**0.0812**
50	0.4238	0.3149	0.4347	**0.2437**	0.3936	0.3150	0.4298	**0.2450**
100	0.7879	0.5713	0.8272	**0.4497**	0.7297	0.5680	0.8108	**0.4491**
200	1.2856	0.9200	1.3838	**0.7354**	1.1933	0.9153	1.3507	**0.7338**
	$BIAS^T$ [%]							
	Climatological site characteristics				Geographical site characteristics			
T [yrs]	ROIcli2Sa	ROIcli2Sb	ROIcli2Sc	ROIcli3	ROIgeo2Sa	ROIgeo2Sb	ROIgeo2Sc	ROIgeo3
5	0.6221	0.5209	0.6489	**0.5137**	0.6175	0.5193	0.5943	**0.5151**
10	0.4069	**0.3393**	0.4633	0.3426	0.5940	0.5146	0.4974	**0.4843**
20	0.0565	**0.0404**	0.1327	0.0487	0.4532	0.4114	**0.2777**	0.3454
50	−0.5155	−0.4563	**−0.4245**	−0.4520	0.1779	0.1960	−0.1091	**0.0682**
100	−0.9832	**−0.8727**	−0.8869	−0.8811	−0.0548	**0.0063**	−0.4294	−0.1799
200	−1.4521	**−1.3026**	−1.3525	−1.3327	−0.2820	**−0.1867**	−0.7446	−0.4415

CC BY

Table 7. Average root mean square error ($RMSE^T$) and average bias ($BIAS^T$) of growth curves of 1-day annual precipitation maxima for return period T [years], expressed in %. The smallest values of the statistics (in an absolute sense) are marked in bold.

	$RMSE^T$ [%]				$BIAS^T$ [%]			
T [yrs]	ROIcli3	ROIgeo2	HWreg	At-site	ROIcli3	ROIgeo2	HWreg	At-site
5	0.0075	0.0065	**0.0060**	0.0493	0.4482	0.4670	0.4709	**−0.0225**
10	0.0345	**0.0284**	0.0315	0.1304	0.2906	0.3912	0.3719	**−0.1586**
20	0.1159	**0.0933**	0.1039	0.4402	**0.0416**	0.2215	0.1772	−0.1762
50	0.3301	**0.2638**	0.2891	1.3808	−0.3586	**−0.0808**	−0.1609	0.0814
100	0.5885	**0.4693**	0.5089	2.6683	−0.6843	**−0.3395**	−0.4483	0.5484
200	0.9432	**0.7508**	0.8067	4.6599	−1.0112	**−0.6070**	−0.7460	1.2970

Table 8. Average root mean square error ($RMSE^T$) and average bias ($BIAS^T$) of growth curves of 5-day annual precipitation maxima for return period T [years], expressed in %. The smallest values of the statistics (in an absolute sense) are marked in bold.

	$RMSE^T$ [%]				$BIAS^T$ [%]			
T [yrs]	ROIcli3	ROIgeo2	HWreg	At-site	ROIcli3	ROIgeo2	HWreg	At-site
5	**0.0047**	0.0054	0.0053	0.0465	0.5137	0.5212	0.5128	**0.0347**
10	0.0202	**0.0200**	0.0205	0.1253	0.3426	0.4958	0.4128	**−0.1009**
20	0.0784	**0.0711**	0.0750	0.4301	**0.0487**	0.3588	0.1957	−0.1366
50	0.2437	**0.2144**	0.2287	1.3680	−0.4520	**0.0776**	−0.1969	0.0803
100	0.4497	**0.3933**	0.4203	2.6652	−0.8811	**−0.1790**	−0.5407	0.5101
200	0.7354	**0.6424**	0.6872	4.6882	−1.3327	**−0.4545**	−0.9048	1.2205

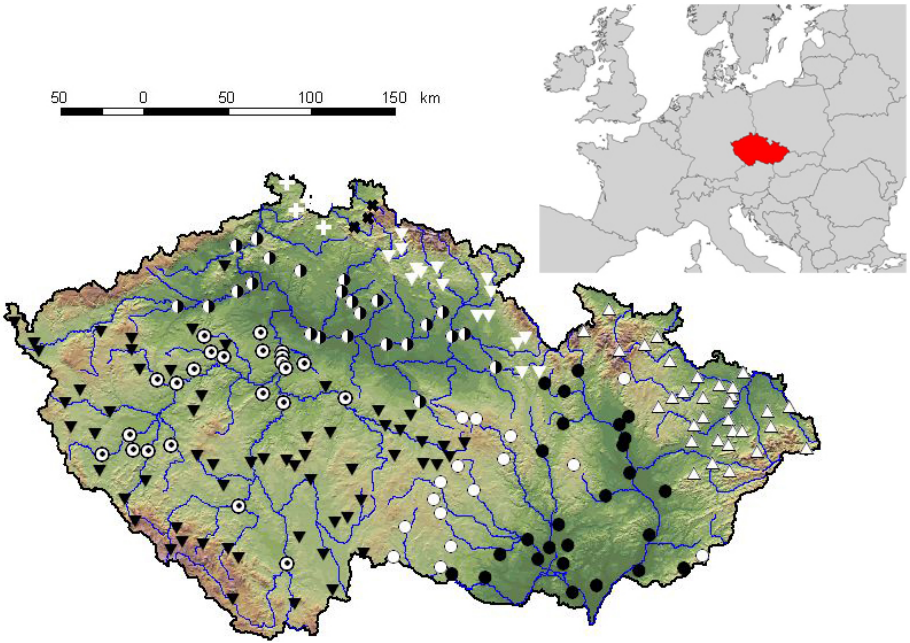

Fig. 1. 209 climatological stations in 9 homogeneous regions available for a regional frequency analysis of heavy precipitation amounts in the Czech Republic.

HESSD

6, 273–317, 2009

CC BY

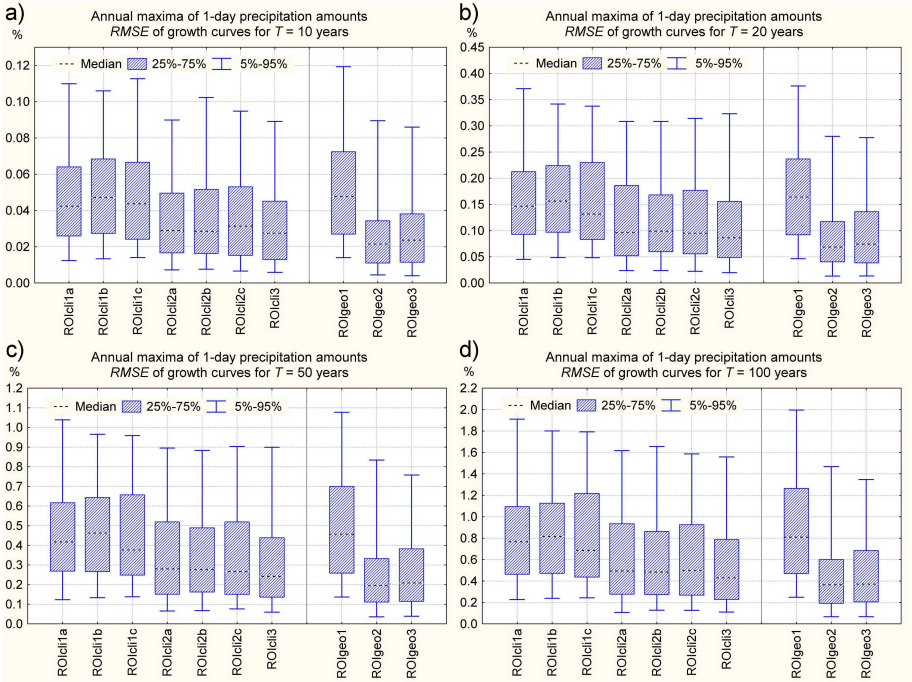

Fig. 2. Root mean square error (RMSE) of growth curves of 1-day annual maxima in a sensitivity analysis when changes made to the basic ROI pooling schemes are examined. T denotes return period. **(a)** T =10 years, **(b)** T =20 years, **(c)** T =50 years, **(d)** T =100 years.

HESSD
6, 273–317, 2009

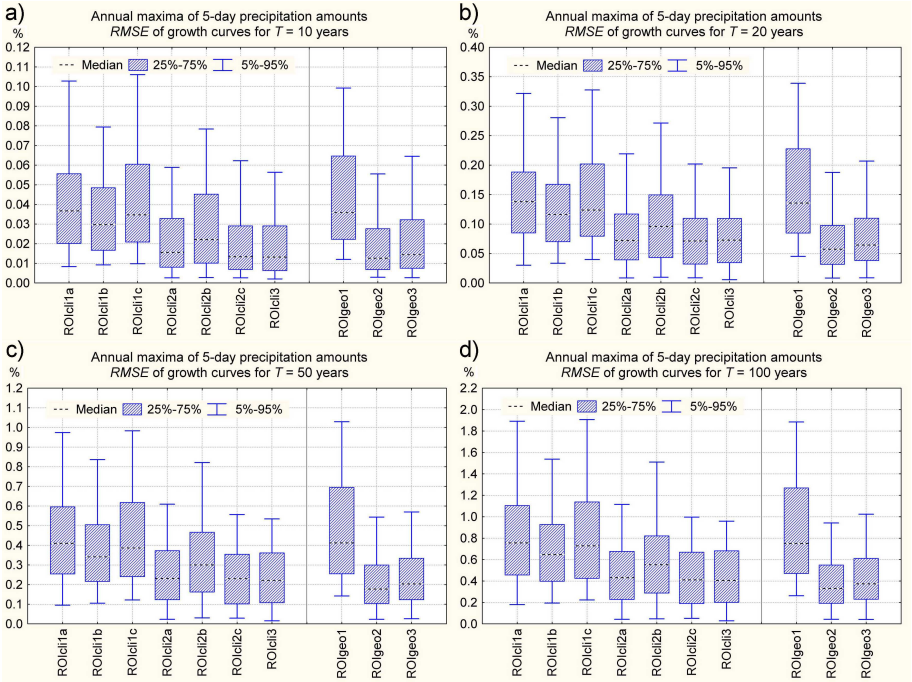

Fig. 3. Root mean square error (RMSE) of growth curves of 5-day annual maxima in a sensitivity analysis when changes made to the basic ROI pooling schemes are examined. T denotes return period. **(a)** T =10 years, **(b)** T =20 years, **(c)** T =50 years, **(d)** T =100 years.

CC BY

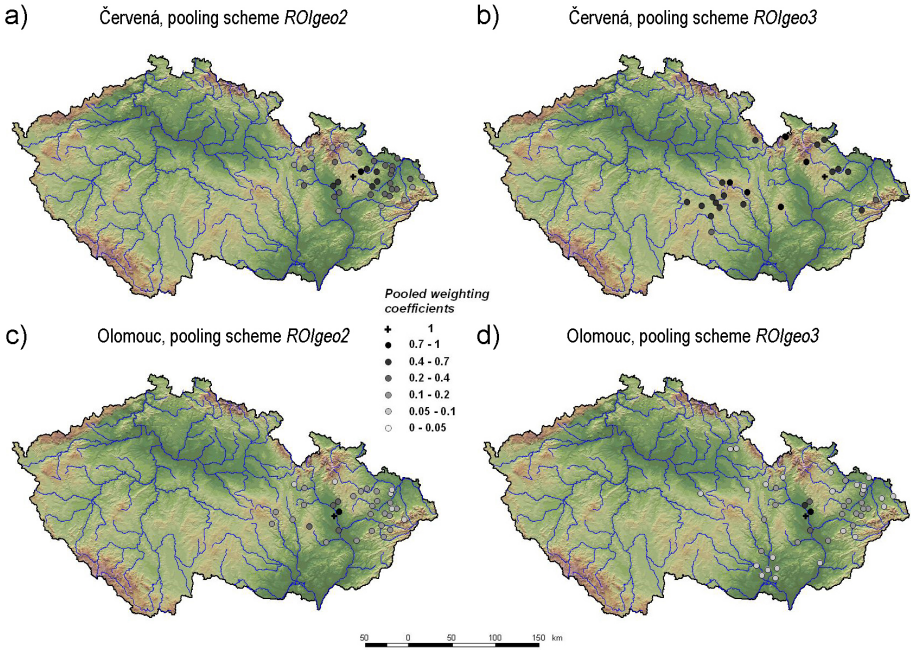

Fig. 4. Pooling groups of Červená (750 m a.s.l.) and Olomouc (225 m a.s.l.) stations according to the ROIgeo2 and ROIgeo3 pooling schemes. Sites of interest are denoted by a "+" sign; the shading of the circles expresses the magnitude of the pooled weighting coefficients of the sites included in the individual pooling groups.

HESSD
6, 273–317, 2009

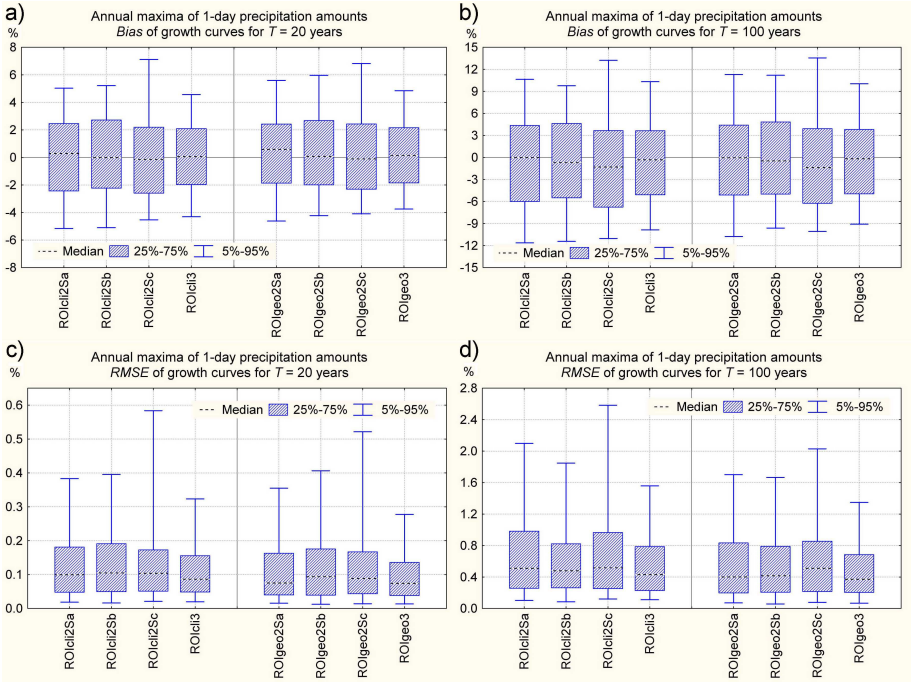

Fig. 5. Bias and root mean square error (RMSE) of growth curves of 1-day annual maxima in a sensitivity analysis when changes made to the "true" frequency models are examined. T denotes return period. **(a)** and **(c)** T=20 years, **(b)** and **(d)** T=100 years.

HESSD
6, 273–317, 2009

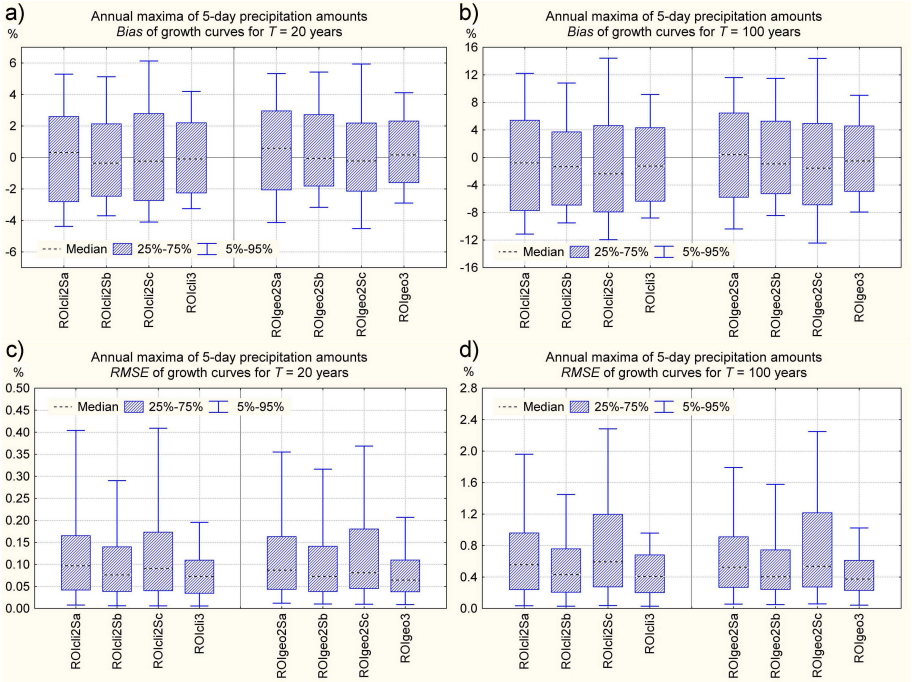

Fig. 6. Bias and root mean square error (RMSE) of growth curves of 5-day annual maxima in a sensitivity analysis when changes made to the "true" frequency models are examined. *T* denotes return period. **(a)** and **(c)** *T* =20 years, **(b)** and **(d)** *T* =100 years.

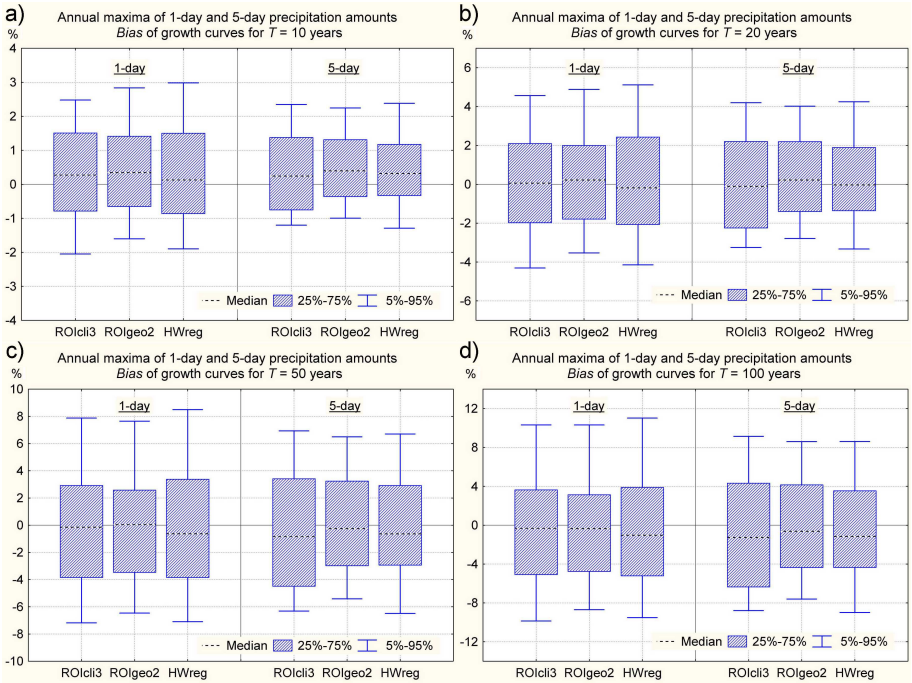

Fig. 7. Bias of growth curves of 1-day and 5-day annual maxima in a comparison of the performance of the preferred pooling schemes (ROIcli3 and ROIgeo2) with the traditional frequency model based on fixed regions (HWreg). T denotes return period. **(a)** T = 10 years, **(b)** T = 20 years, **(c)** T = 50 years, **(d)** T = 100 years.

HESSD
6, 273–317, 2009

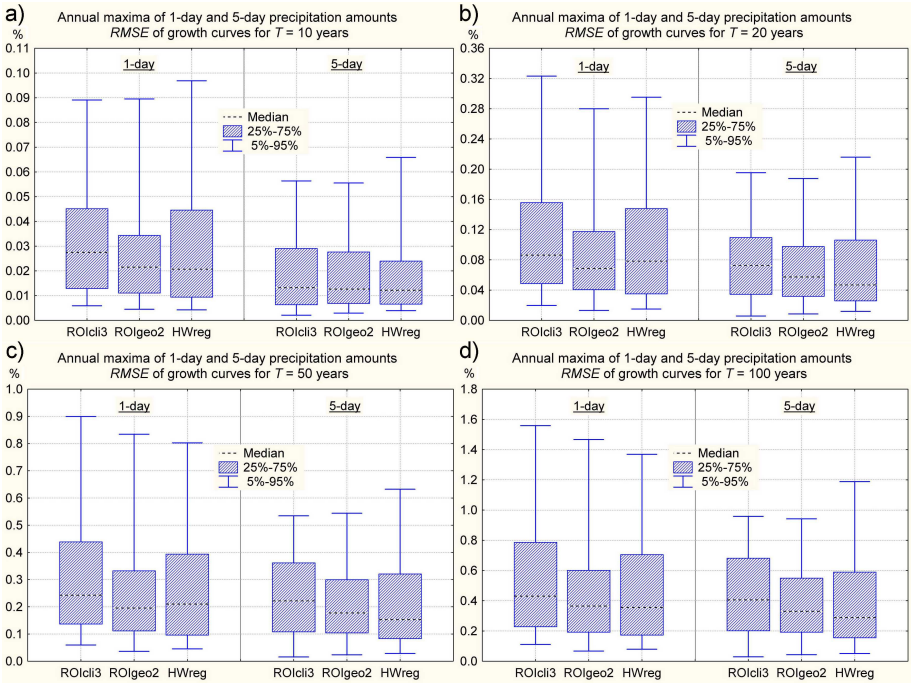

Fig. 8. Root mean square error (RMSE) of growth curves of 1-day and 5-day annual maxima in a comparison of the performance of the preferred pooling schemes (ROIcli3 and ROIgeo2) with the traditional frequency model based on fixed regions (HWreg). T denotes return period. **(a)** T =10 years, **(b)** T =20 years, **(c)** T =50 years, **(d)** T =100 years.

HESSD
6, 273–317, 2009