Originally published as:

Guse, B., Thieken, A. H., Castellarin, A., Merz, B. (2010): Deriving probabilistic regional envelope curves with two pooling methods. - Journal of Hydrology, 380, 1-2, 14-26,

DOI: 10.1016/j.jhydrol.2009.10.010

Deriving probabilistic regional envelope curves with two pooling methods

Björn Guse[1,2], Annegret H. Thieken[1,3], Attilio Castellarin[4] and Bruno Merz[2]

[1]Center for Disaster Management and Risk Reduction Technologies (CEDIM), D-76187 Karlsruhe, Germany

[2]Deutsches GeoForschungsZentrum GFZ, Section 5.4 – Hydrology, Telegrafenberg, D-14473 Potsdam, Germany

[3]alpS - Centre for Natural Hazards and Risk Management, University of Innsbruck, A-6020 Innsbruck, Austria

[4]DISTART, School of Civil Engineering, Viale Risorgimento, 2, University of Bologna, I-40136 Bologna, Italy

Corresponding author:

Björn Guse

Deutsches GeoForschungsZentrum GFZ

Section 5.4 – Hydrology

Telegrafenberg

D-14473 Potsdam

Germany

Tel: +49-331-288-1896

Fax: +49-331-288-1570

Abstract:

A probabilistic regional envelope curve (PREC) assigns a recurrence interval to a regional envelope curve. A central point of this method is the determination of homogeneous regions according to the index flood hypothesis. A flood discharge associated with the recurrence interval (PREC flood quantile) is estimated for each gauge of a homogeneous region. In this study, the influence of two pooling methods on PREC for a large group of catchments located in the south-east of Germany is investigated. Firstly, using cluster analysis, fixed homogeneous regions are derived. Secondly, the Region of Influence (RoI) approach is combined with PREC. The sensitivity of PREC flood quantiles with respect to pooling groups is evaluated. Different candidate sets of catchment descriptors are used to derive pooling groups for both pooling methods. Each pooling group is checked by a homogeneity test. PRECs are then constructed for all homogeneous regions. The ensemble of PREC realisations reveals the sensitivity of the PREC flood quantiles. A comparison with the traditional index flood method ascertains the suitability of the pooling methods. A leave-one-out jackknifing procedure points out a similar performance of cluster analysis and RoI. Furthermore, a comparison of different degrees of heterogeneity for deriving pooling groups reveals that the performance of PREC for ungauged catchments decreases in more heterogeneous pooling groups.

Keywords: Probabilistic regional envelope curves, Pooling methods, Region of Influence, Cluster analysis, Sensitivity analysis, Saxony - Germany

1 Introduction

For flood risk analyses and estimations of design floods it is fundamental to accurately quantify the discharges of rare events, i.e. flood events with recurrence intervals of 100 years or more. The well-established methods of flood frequency analysis (FFA) are hampered by the uncertainty that occurs particularly for estimates of high recurrence intervals due to limited observation data (e.g. Robson and Reed, 1999; Merz and Thieken, 2005). Regional Flood Frequency Analysis (RFFA) is widely employed in the estimation of design floods when dealing with data record lengths that are too short compared to the recurrence interval of interest (e.g. Hosking and Wallis, 1997). Still, most methods of FFA and RFFA do not consider an upper bound of the flood discharges.

Regional envelope curves (RECs) are a traditional, deterministic method for representing the upper bound of the maximum floods observed in a distinct region. A REC bounds the largest floods of each gauge, termed floods of record, of a region. Since their first introduction (Jarvis, 1925), RECs have been applied to different regions and scales. Traditionally, they refer to administrative units (e.g. China and USA (Costa, 1987), Europe and World (Herschy, 2002)). RECs have also been constructed for hydro-meteorological regions with different climatic conditions and, consequently, different flood regimes (e.g. 17 regions in the USA (Crippen and Bue, 1977); north-western and western Greece (Mimikou, 1984)).

A main criticism on RECs relates to their deterministic view and their need to be checked routinely for being exceeded by recent events (e.g. Crippen and Bue, 1977; Castellarin et al., 2005). The applicability of RECs to engineering problems, such as flood design, is limited by the lack of an exceedance probability (or a recurrence interval) that can be assigned to the envelope curves. To overcome this deficiency, Castellarin et al. (2005, 2007), Castellarin (2007), and Vogel et al. (2007) proposed a probabilistic interpretation of RECs which, besides the magnitude, also considers the frequency of a REC.

Probabilistic regional envelope curves (PRECs) are based on the well-known index flood method (Dalrymple, 1960), which is often applied in flood regionalisation studies (e.g. GREHYS, 1996; Hosking and Wallis, 1997; Robson and Reed, 1999). Only if a region is homogeneous as defined by the index flood hypothesis, a PREC can be constructed and an exceedance probability can be assigned to the curve. A flood discharge associated with the exceedance probability, termed PREC flood quantile, was derived for each site of a homogeneous region.

According to Castellarin et al. (2005), the estimation of the exceedance probability of a PREC further requires the evaluation of the overall sample years of the underlying data which in turn depends on the intersite or cross correlation amongst the annual maximum series (AMS) of flood flows observed at different gauges. It is important to emphasise that the exceedance probability of a PREC always differs from zero which highlights the difference between PRECs and Probable Maximum Floods. A PREC provides one recurrence interval without an extrapolation and, in principle, enables one to estimate the design flood at ungauged sites as a function of the drainage area (e.g. Castellarin, 2007) or of a set of suitable physiographic and climatic catchment descriptors (e.g. Castellarin et al., 2007). PRECs should be seen as complements to RFFA. They can provide additional information on plausible values of extreme floods and the corresponding exceedance probability in gauged and ungauged basins. A leave-one-out jackknifing approach has shown that PREC flood quantiles have a similar reliability as the traditional index flood method (Castellarin, 2007).

Similarly to RFFA, the construction of a PREC requires the identification of hydrologically homogeneous regions or pooling groups (GREHYS, 1996; Castellarin et al., 2001). Catchments with similar hydrological behaviour can be classified into one group, and the hydrometric information collected at all gauges that belong to the pooling group can be used to improve the accuracy of the design flood estimates for all gauges of the group. The homogeneity of a pooling group can be assessed by statistical tests (e.g. Viglione et al., 2007; Castellarin et al., 2008).

The requirement of homogeneity and the need for sufficient data within a group are often conflictive. On the one hand, a larger number of observations reduces the uncertainty in estimating high recurrence intervals (Robson and Reed, 1999). On the other hand, a larger number of gauges in the pooling group generally results in a higher hydrological heterogeneity of the group. Several studies highlight the relevance of regional homogeneity for RFFA (e.g. Lettenmaier et al., 1987; Stedinger and Lu, 1995) and, more recently, for PRECs (Castellarin, 2007). Therefore, an appropriate classification technique is required for the identification of pooling groups.

Flood regionalisation studies propose two approaches for deriving pooling groups: the delineation of a subdivision of the study area into fixed homogeneous regions and the neighbourhood approach or Region of Influence approach (RoI) (Burn, 1990; GREHYS, 1996; Ouarda et al., 2001). In fixed homogeneous regions, each gauging station definitely belongs to one and only one region. A traditional approach to identify fixed homogeneous regions is a separation in administrative units, where all gauging stations are geographically connected, e.g. in adjacent sub-catchments. This method has been replaced by others that enhance the hydrological similarity within a fixed region (Acreman and Sinclair, 1986). Cluster analysis is an objective procedure that can be applied to subdivide the study area into clusters of catchments (fixed regions) on the basis of a suitable set of climatic and physiographic catchment descriptors (predictor variables). The goal of the procedure is to maximise the similarity within a cluster and the dissimilarity between the clusters (e.g. Mosley, 1981). The catchments of one cluster are not necessarily geographically connected.

The RoI approach identifies a pooling group separately for each gauging station (site of interest) without explicit spatial connection within the RoI (Burn, 1990). Gauging stations for a RoI are selected according to their similarity to the site of interest using a suitable set of predictor variables (Zrinji and Burn, 1994). In a hybrid RoI approach, the RoI is derived by considering the geographical distance between the sites in addition to the predictor variables (Eng et al., 2007).

Up to now, PRECs were applied in northern Italy with a relatively limited number of gauging stations grouped into three different fixed homogeneous regions (Castellarin, 2007). This paper presents the application of the PREC approach in Germany, considering a rather large number of sites. The main aim of the study is to verify, whether the utilisation of the RoI approach in the formation of homogeneous pooling groups may improve the reliability of the design flood estimates that can be retrieved from PRECs for ungauged sites. To address this issue, we construct PRECs for the study area using fixed homogeneous regions and RoIs. In particular, we form several PRECs for each gauging site on the basis of the data collected in homogeneous fixed regions and RoIs with different sizes and catchment descriptors. A sensitivity analysis enables us to consider the sensitivity of PREC flood quantiles to different constitutions of the pooling group. By means of "leave-one-out" cross-validation procedure, we simulate the ungauged conditions at all considered sites during the construction of each PREC as proposed by Castellarin (2007). All flood estimates are compared with the corresponding estimates (i.e. flood quantiles associated with the same values of the recurrence interval) obtained by applying a traditional regionalisation approach. The comparison enables us to better understand and quantify (1) the suitability of the two different pooling methods (i.e. cluster analysis and RoI) in the context of probabilistic regional envelope curves, and (2) the accuracy of flood quantiles retrieved from PRECs for ungauged basins.

2 Methods

Since the construction of pooling groups is a prerequisite for the application of PREC, it is advisable to quantify the sensitivity of PREC to the formation of pooling groups. For both pooling methods (cluster analysis and RoI), the sensitivity of PREC results was determined by considering several variations of pooling groups derived in a three-step-procedure.

1. Formation of candidate sets of catchment descriptors.
2. Construction of homogeneous regions using two pooling methods.
3. Test on homogeneity of each pooling group.

Finally a specific PREC was constructed for each homogeneous region. To compare the different results some performance measures were analysed. Each step of the procedure is described in the remainder of this section.

2.1 Candidate set of catchment descriptors

Different catchment descriptors were used as predictor variables to derive homogeneous regions. In a first step all catchment descriptors were standardised to a mean value of zero and a standard deviation of one. This standardisation allows a comparison between the predictor variables and avoids the influence of different value scales (see e.g. Nathan and McMahon, 1990).

The catchment descriptors were combined by summing up the standardised values for each site. This approach is only applicable, if all standardised variables have a positive correlation with the unit index flood, i.e. the index flood normalised by the catchment size. In order to get only positive correlations, standardised variables with a negative correlation to the unit index flood were multiplied with -1. This implies, for instance, that the fraction of the area, which is not covered by arable land, was used instead of the fraction of arable land for selecting candidate sets of catchment descriptors.

A full enumeration approach was used to consider all possible subsets of the catchment descriptors with one to three predictor variables. A larger number of catchment descriptors within one candidate set could provide small additional information, but could also lead to multi-collinearity (Merz and Blöschl, 2005). Thus variants with more than three predictor variables were not taken into account.

With regard to the selection of suitable sets of predictor variables, it is worth noting that we were interested in assessing the sensitivity of PRECs and of flood quantiles derived from these PRECs with respect to different pooling groups. To this aim, we looked for several good combinations of predictor variables rather than the optimal set. It was assumed that, next to the best subset of catchment descriptors, other ‘good subsets’ have a similar explained variance. Since PREC is based on the assumption of a scaling of the index flood (mean of the annual maxima series), it seemed reasonable to perform a preliminary identification of candidate sets of catchment descriptors by looking at the explained variance of the empirical index flood values. Therefore, candidate sets of catchment descriptors were identified on the basis of this criterion.

The correlation coefficient between a subset of catchment descriptors and the unit index flood was used as goodness-of-fit criterion, as in other studies (e.g. Burn, 1990; Uhlenbrook et al., 2000) under the assumption that a high correlation is a good indicator for a sufficient explained variance of the selected subset (Merz and Blöschl, 2004).

All subsets of catchment descriptors were selected that showed a correlation coefficient of more than 0.60. This threshold was assumed as sufficient, because the correlation coefficient was only used for a pre-selection of subsets of catchment descriptors.

All selected subsets were checked for multi-collinearity between the catchment descriptors using the variance inflation factor (VIF) (Hirsch et al., 1992) (Eq. (1)).

$$VIF_k = \frac{1}{1 - {r_k}^2} \qquad (1)$$

${r_k}^2$ stands for a multiple correlation coefficient, which was calculated by a regression of variable k using all other variables as predictor variables. To avoid multi-collinearity, all subsets with VIF > 5 were omitted. Montgomery et al. (2001) and Eng et al. (2005) recommended a threshold between 5 and 10.

2.2 Formation of homogeneous regions

To assess the influence of homogeneous regions on PREC, two different approaches for the derivation of pooling groups were applied. These methods were fixed homogeneous regions derived by a cluster analysis and the Region of Influence (RoI) method. To ensure an appropriate comparison of both methods, the same candidate sets of catchment descriptors were used.

Fixed homogeneous regions using cluster analysis

Fixed homogeneous regions were derived by cluster analysis with the K-means algorithm, which had already been used in flood frequency analysis (e.g. Burn, 1989; Burn and Goel, 2000) and very recently in a flood seasonality study (Beurton and Thieken, 2009). The cluster analysis was performed allowing three to seven clusters, and was therefore applied five times to each subset of predictor variables. The different number of clusters considers the trade-off between the homogeneity within a cluster and the number of sites within one group.

Region of Influence (RoI)

The approach "Region of Influence" (Burn, 1990) constructs an individual region (group of gauging sites) for each gauge by finding stations that are similar to the characteristics of the station under study (site of interest). The RoI was determined by the similarity of gauging stations in the physiographical space of the selected catchment descriptors. Similarity was assessed by the Euclidean distance between each site and the site of interest in the physiographical space. The Euclidean distance has been used in several RoI approaches (e.g. Zrinji and Burn, 1994; Castellarin et al., 2001; Gaál et al., 2008), although other similarity measures are possible (see e.g. Cunderlik and Burn, 2006).

All gauging stations which are closer to the site of interest than a specific threshold of the Euclidean distance in the physiographical space were assigned to the RoI of the site of interest. The higher the threshold, the larger is the number of sites within a region (Burn, 1990). Different similarity measure thresholds to derive RoIs were investigated by Gaál et al. (2008). To account for the sensitivity of the results to the threshold, three thresholds for the similarity measure (0.5, 1 and 2) were applied in this study. In contrast to RoI approaches in frequency analysis (Burn, 1990), the sites were not weighted according to their closeness to the site of interest in the physiographical space. The original RoI method was varied, because the intercept of PREC is only determined by one pair of unit flood of record and drainage area (see "Probabilistic regional envelope curve"). Consequently, a weighting scheme would not affect the magnitude of the regional envelope curve.

Traditionally, a fixed number of sites is targeted at when deriving a RoI (Burn, 1997). This target number is a function of the aspired return period. In our case a target number of sites cannot be determined, since the recurrence interval *T* associated with the PREC is not known a priori. Therefore, the maximum number of sites in the RoI was identified on the basis of the hydrological affinity with the site of interest.

2.3 Homogeneity test

Each pooling group was checked for homogeneity by applying the heterogeneity measure of Hosking and Wallis (1997) (Table 1). The H_1-test calculates the variability of the L-coefficient of variation (L-CV). The sample L-CV is compared with an expected value for a homogeneous region obtained by a Monte-Carlo simulation. The second and third heterogeneity measures H_2 and H_3 consider the L-CV and the L-skewness as well as the L-skewness and the L-kurtosis, respectively. A more detailed explanation of L-moments and the heterogeneity measure is given by Hosking and Wallis (1997).

Since the homogeneity test for the L-CV (H_1) is a more significant test than the tests with higher moments (H_2 and H_3) (Castellarin et al., 2001, 2007; Hosking and Wallis, 1997), this study focused on the H_1-test using the hw.test (Viglione, 2008, implemented in R). All regions with a H_1 value lower than 2 were used for deriving a PREC.

2.4 Probabilistic regional envelope curve

The method of probabilistic regional envelope curves (PREC) is based on two principles. In the first place, all gauging stations of a region have to be homogeneous in terms of the index flood hypothesis. Secondly, the index flood μ_X (mean of the annual maxima series) is related to the drainage area A (Eq. (2), adopted from Castellarin, 2007). Under these assumptions the index flood scales with the drainage area and depends only on the drainage area (Castellarin, 2007):

$$\mu_X = C * A^{b+1} \tag{2}$$

To derive a regional envelope curve, all floods of record Q_{FOR} of a region are normalised by their corresponding catchment area A to the unit flood of record q_{FOR} and are related to A in a double-logarithmic scale (Eq. (3), adopted from Castellarin et al., 2005). The regional envelope curve bounds all unit floods of record of a region and is defined by its slope b and the intercept a:

$$\log\left(\frac{Q_{FOR}}{A}\right) = a + b * \log(A) \tag{3}$$

The slope b is derived by a regression of the unit index flood against the drainage area (Fig. 1). The intercept a is determined by a parallel upshift of the regression until the envelope curve bounds all unit floods of record (Castellarin et al., 2005). In a homogeneous region the index floods of all gauges are close to the regression line. In this study, a PREC was determined for each region with at least four sites. It was assumed that a lower number of sites is not representative for a regression analysis.

An exceedance probability is assigned to that particular data pair of unit flood of record and its drainage area that determines the intercept of the envelope curve. This exceedance probability is valid for the range of catchment sizes covered in the pooling group. For this, the AMS of all gauging stations of that region were considered. The total number of sample years of data was reduced to an effective number of sample years of data, by accounting for cross-correlated sites (Castellarin, 2007). Several studies have shown that the correlation of annual maximum series decreases with the distance of the catchments (see e.g. Hosking and Wallis, 1988; Troutman and Karlinger, 2003). Under these assumptions, a regional cross-correlation function by Tasker and Stedinger (1989) (Eq. (4), from Castellarin, 2007) was optimised using the distances between catchment centroids, the correlation coefficients between the AMS and the lengths of overlapping time series.

$$\rho_{i,j} = \exp\left(-\frac{\lambda_1 d_{i,j}}{1+\lambda_2 d_{i,j}}\right) \quad (4)$$

d is the distance between catchment centroids, ρ the correlation coefficient, λ_1, λ_2 the parameters, and i,j are the index denoting pairs of catchments.

In comparison to Castellarin (2007), the method for considering intersite correlations was changed in this paper due to the larger number of catchments available and the presence of numerous nested catchments, i.e. gauging stations along the same river. Troutman and Karlinger (2003) emphasised that the correlation between the AMS of nested catchments was higher than for unnested catchments. Guse et al. (2009) showed that distinct parameter sets for nested and unnested catchments led to a reduction of the recurrence interval of PRECs due to larger correlations between nested catchments. Hence, specific parameters of the cross-correlation function were used for nested and unnested catchments.

Considering the intersite correlation, the overall effective sample years of data n_{eff} were calculated by an empirical relationship, which was determined by Castellarin et al. (2005) and Castellarin (2007) in Monte-Carlo simulations (Eq. (5)). This approach is based on the average correlation coefficient ρ (see Eq. (4)). Castellarin (2007) proposed an algorithm that can be applied for real world datasets with Y years, in which the record lengths of the gauges varies. In the first step of the algorithm, the number of years n_1 was identified in which only one gauging station had a measured discharge. These observations n_1 were reasonably effective. The remaining years Y-n_1 were divided in $Y_{sub} \le (Y - n_1)$ subsets with the same gauging stations L_s and the length l_s. Next, for each subset s of l_s years, the effective number of observations $n_{eff,s}$ was calculated separately. Finally, the effective samples for all subsets were summed up. The number of effective sample years of data for the whole regional data set n_{eff} includes n_1, the years with one observations, and the sum of $n_{eff,s}$ (Eq. (5), adopted from Castellarin, 2007).

$$n_{eff} = n_1 + \sum_{s=1}^{Y_{sub}} n_{eff,s} = n_1 + \sum_{s=1}^{Y_{sub}} \frac{L_s l_s}{1 + \left[\overline{\rho^{\beta}}\right]_{L_s} (L_s - 1)} \quad \text{with} \quad ß := 1.4 \left[\overline{\frac{(L_s l_s)^{0.176}}{(1-\rho)^{0.376}}}\right]_{L_s} \qquad (5)$$

In this way the effective sample years of data is equivalent to the number of independent observations. This reduction of the regional plotting position determines the information content of the collected data (Castellarin, 2007).

The next step is a selection of an appropriate plotting position depending on an adequate distribution function to estimate the recurrence interval of the PREC. Castellarin (2007) recommended the use of the Hazen plotting position (Eq. (6), from Castellarin (2007)) in order to get unbiased flood quantiles, when the Generalised Extreme Value (GEV) distribution is a suitable parent distribution. Its suitability for the case study is reported in "Study area and data". As a result, the recurrence interval T_{PREC} is twice as high as the number of effective observations n_{eff}.

$$T_{PREC} = 2 * n_{eff} \qquad (6)$$

The exceedance probability is greatly influenced by the formation of homogeneous regions. Adding or removing only one gauging station to/from a homogeneous group modifies the effective sample years of data and hence the exceedance probability of the PREC.

The discharge associated with the exceedance probability for a specific site is determined by the intercept of the drainage area and the regional envelope curve. It is worth noting that the gauging stations within a region have a different influence on the exceedance probability of the PREC. Due to the fact that the intercept of the PREC is determined by the data pair of the highest unit flood of record and its drainage area, this gauging station is the most decisive.

This aspect highlights the particular importance of a consistent assignment of gauging stations to pooling groups.

A discharge Q_{PREC} and a recurrence interval T_{PREC} were derived for all gauging stations of a region. T_{PREC} is constant for all gauging stations in the region. Since the PREC was only calculated for homogeneous regions, the number of PREC realisations is different for the gauging stations. It depends on the number of homogeneous regions in which the specific gauging station is included.

2.5 Sensitivity analysis

The effect of pooling groups on PREC flood quantiles (Q_{PREC}, T_{PREC}) was examined by a sensitivity analysis. Pooling groups of both pooling methods were derived for all candidate sets of catchment descriptors with a correlation coefficient to the unit index flood >0.60. For each candidate set of catchment descriptors, cluster analysis was applied five times (allowing three to seven clusters) and the Region of Influence approach three times (with different thresholds in the physiographical space) (see "Formation of homogeneous regions"). These predefined number of clusters and thresholds in the physiographical space led to several candidate solutions of pooling groups. Ultimately all pooling groups with a heterogeneity measure $H_1 < 2$ were used to derive a PREC. Each PREC realisation led to a pair of Q_{PREC} and recurrence interval T_{PREC} (PREC flood quantile) for each gauge of the pooling group.

The rationale behind this scheme is that different constitutions of the regions lead to different realisations of PREC. The application of several candidate sets of catchment descriptors allows a quantification of the sensitivity of the PREC results in terms of the pooling method and the selected subset of catchment descriptors. However, it is worth noting that the uncertainty of the ensemble of PRECs results is not estimated by this procedure.

2.6 Performance criteria

The performance of PREC flood quantiles was evaluated by comparing them with a traditional index flood approach.

The index flood method is based on the assumption that a regional growth curve is valid for all sites of a pooling group. For this, the AMS was normalised by the index flood μ_X. To calculate the T-year flood X(T), a regional quantile x_T was scaled to at-site conditions by the index flood μ_X (Eq. (7)).

$$X(T) = \mu_X * x_T \tag{7}$$

The GEV was also used for the index flood approach. The parameters were estimated with regional L-moments, by weighting at-site L-moments of all gauges according to the data length (Robson and Reed, 1999).

In order to assess the accuracy of PREC for ungauged catchments, a cross-validation procedure was applied. The PREC was recalculated following a leave-one-out jackknifing algorithm (Castellarin, 2007; Castellarin et al., 2007), termed PREC-JK: (1) A pooling group with M sites, which fulfilled the homogeneity criteria, was selected. (2) A site m was excluded from this pooling group. (3) For the remaining M-1 stations the PREC-JK was calculated and the recurrence interval of PREC-JK ($T_{PREC\text{-}JK}$) was determined. (4) The discharge of PREC-JK $Q_{PREC\text{-}JK}$ was evaluated for the given drainage area of the site m. Since site m was not included in the calculation, the PREC-JK result was considered as ungauged. (5) The index flood method was applied for the same pooling group. In this case the site m was included. The flood quantile for the given recurrence interval $T_{PREC\text{-}JK}$ was calculated by the index flood method (Q_{IF} ($T_{PREC\text{-}JK}$)). In this context the index flood method was assumed as the 'true' result. To get a perfect estimator for ungauged conditions, $Q_{PREC\text{-}JK}(T_{PREC\text{-}JK})$ was compared with Q_{IF} ($T_{PREC\text{-}JK}$) (Eq. (8), adopted from Castellarin, 2007).

$$\varepsilon_{PREC-JK} = \frac{Q_{PREC-JK}(m, T_{PREC-JK}) - Q_{IF}(m, T_{PREC-JK})}{Q_{IF}(m, T_{PREC-JK})} \tag{8}$$

The cross-validation was performed for all homogeneous regions. It was repeated M-times for all sites within a cluster. In the case of a RoI, the jackknifing approach was only applied once for the site of interest. The relative error of PREC-JK in comparison to the index flood method enables us to compare the two pooling methods.

3 Study area and data

The study area is the federal state of Saxony in the south-east of Germany (Fig. 2). Saxony is characterised by the mountain range of the *Erzgebirge* in the south-west with elevation up to 1214 m above sea level *(Fichtelberg)* and a mean annual precipitation up to 1244 mm (at the synoptic station *Carlsfeld*). The highest monthly precipitation occurs in summer (Flemming, 2001). The river Elbe with a drainage area of about 52,000 km² at the gauge Dresden is the biggest river in Saxony. Several feeder rivers originating in the *Erzgebirge* flow into the Elbe, the most important one is the river Mulde (Fig. 2). The mountain range east of the Elbe has a lower elevation than the *Erzgebirge*. Towards the north the elevation flattens. The north-western and north-eastern parts of Saxony are influenced by mining activities.

In Saxony, several severe floods occurred in the past. Ulbrich et al. (2003) distinguished between flash floods along the tributaries of the rivers Elbe and Mulde and slowly rising river floods along the Elbe. The *Erzgebirge* was affected by local (e.g. in 1927, 1957) and regional floods (e.g. in 1954, 1958, 2002) (Pohl, 2004; Thieken et al., 2007). Among the regional floods, especially the recent destructive flood of 2002 along the rivers Elbe and Mulde and their tributaries from the *Erzgebirge* is still present in people's minds. During this event a record-breaking daily precipitation of 312 mm/day was measured at the synoptic station *Zinnwald-Georgenfeld*, which is located in the upper stream of the Müglitz (Ulbrich et al., 2003). For the 2002 flood, IKSE (2004) estimated recurrence intervals up to 200 – 500 years at some tributaries of the Elbe river, e.g. at the rivers Mulde, Müglitz and Weisseritz.

One hundred and seventeen discharge gauging stations from all over Saxony with the maximum discharges for each month were provided by Saxon authorities. For the catchment of the *Weisse Elster,* which is only partly located in Saxony, additional data was provided by authorities of Thuringia and Saxony-Anhalt. The gauging stations are evenly distributed throughout the area of this study (Fig. 2). All major rivers are included in the data set. Observation periods range from 20 to 150 years with a mean length of 50 years. This data set includes extreme floods with local as well as regional spatial extent. The highest unit discharges were observed in the western tributaries of Elbe (i.e. at the rivers Gottleuba and Müglitz) and in the river Pliessnitz, a tributary of the Lausitzer Neisse near the German-Polish border (Fig. 2). Due to a few very extreme floods, the series of annual maximum floods show a high skewness, especially in the *Erzgebirge* (Petrow et al., 2007).

Since the index flood hypothesis requires a strong homogeneity within a region, only gauging stations were used that represented the regional hydrological situation. Thus, the available data set was reduced, i.e. gauges of heavily influenced rivers due to mining activities (four sites), gauging stations directly downstream of a dam (two sites) and very small catchments (<10 km^2) (four sites) were discarded. Furthermore, only gauging stations with at least 30 years of data were used. Due to these restrictions the number of gauging stations was reduced to 95.

The construction of pooling groups (see “Formation of homogeneous regions”) requires the derivation of different catchment descriptors. These predictor variables were pre-selected based on a literature review (e.g. Wiltshire, 1986; Pitlick, 1994; GREHYS, 1996; Castellarin et al., 2004; Merz and Blöschl, 2005). Those catchment descriptors were applied, which have yielded good results in flood regionalisation studies (Table 2). All catchment descriptors are catchment averages.

Precipitation data with a daily resolution in and around Saxony was provided by the German Weather Service (DWD). Precipitation indices were derived on the basis of 453 stations with at least 30 years of data in order to ensure a sufficient sample size. The second constraint was that the time series endured at least up to 2002. This year was selected because of the severe flood event in August 2002. In order to optimise the spatial distribution of precipitation stations, 23 stations with an observation period of less than 30 years were additionally used to derive the maximum daily precipitation and the 5-day-precipitation sum. These stations were added, because the year of the maximum daily precipitation coincided with the flood of record of the downstream gauging station. In these cases, it was assumed that the maximum precipitation was representative for this catchment. The precipitation values were interpolated for the different precipitation indices using ordinary kriging. In the next step the catchment boundaries were superimposed on the precipitation map and the mean value was derived for each catchment.

The mean elevation of the catchments was derived from a digital elevation model for Saxony with a grid size of 25 m. Outside Saxony the SRTM-DEM (Jarvis et al., 2008) with a grid size of 90 m was resampled to 25 ms. A mean slope was derived from this combined DEM. The DEM also provided the catchment centroids, from which the distances between the catchments were calculated, which were then used to optimise the theoretical cross-correlation function (see Eq. (4)). The digital landscape model ATKIS (BKG GeoDataCentre, 2005) was used to derive landscape parameters such as the fraction of urban area. The hydrogeological map HÜK200 (1:200,000) of the Saxon State Agency of Environment and Geology provided the fraction of bedrock and low permeability area. The hydrogeological map HÜK200 distinguished between bedrock and unconsolidated rock. Permeability was classified in eleven classes. Low permeability was assessed for all rocks with permeability $<10^{-7}$ (AG Boden, 1994).

Soil parameters were not used in this study, since for example Merz (2006) has emphasised the low performance of soil parameters in multiple regressions without a hydrological soil classification such as the Hydrology of Soil Types (HOST) classification in the United

Kingdom (Boorman et al., 1995). The drainage area itself was not used as variable, because it is already included in the concept of regional envelope curves.

Among the available data for the catchment descriptors only the DEM covered the catchments outside of Saxony. Therefore, catchments with insufficient information for the other catchment descriptors were omitted. This led to a further reduction of the data set. In total, all thirteen catchment descriptors listed in Table 2 were determined for 89 gauging stations shown in Fig. 2. Their catchment size varies between 13 (Rennersdorf 2/ Pliessnitz river) and 6170 km² (Bad Düben/ Mulde river).

For each of the 89 gauges the flood of record Q_{FOR} was determined. In a further step, the annual maximum series (AMS), which contain the highest discharge for each hydrological year (1st November to 31th October), was calculated. Independence between flood events in the AMS was ensured by a time gap of at least 7 days between consecutive annual maxima (GREHYS, 1996). L-moment ratio diagram (see e.g. Vogel and Fennessey, 1993; Peel et al., 2001) clearly indicates that the GEV is a suitable parent distribution function for the whole study area.

4 Results

4.1 Suitable candidate sets of catchment descriptors

Considering the 13 catchment descriptors listed in Table 2, 13 subsets with one, 78 with two and 286 with three catchment descriptors resulted. Among the 377 possible subsets of one, two and three catchment descriptors, 39 subsets have a correlation coefficient to the unit index flood higher than 0.6. All subsets with three catchment descriptors were checked for redundancy compared with the subsets with two catchment descriptors. The rationale behind this approach was that an additional parameter ought to lead to a higher proportion of explained variance. Consequently, subsets with three catchment descriptors were only used (a) if they did not include two catchment descriptors, which formed one of the selected subsets with two catchment descriptors, or (b) if the correlation coefficient was higher than this subset with two catchment descriptors. This procedure reduced the number of subsets from 39 to 20. The test of multi-collinearity by the VIF-test resulted in no further reduction.

Table 3 illustrates that the correlation coefficient to the unit index flood of the 20 subsets differed between 0.60 and 0.70. All 20 subsets were considered as candidate set and were used to form homogeneous regions and to derive a PREC. The selected subsets contain two or three catchments descriptors. Among the catchment descriptors precipitation and topographic indices have a higher explanatory power than land use and geologic parameters. The maximum of the 5-day-precipitation sum (MAX5DAY), the range of elevation within the catchment (RANGE_NORM) and the fraction of urban land coverage (URBAN) were most often included.

4.2 Results for the best subset of catchment descriptors

The best subset of predictor variables contains MAX5DAY, the mean elevation (ELEV) and RANGE_NORM with a correlation coefficient of 0.70 (Table 3). The pooling groups derived by cluster analysis are illustrated in Table 4, using the solution with seven clusters as an example. The heterogeneity measure of the cluster analysis shows that there are four ($H_1 < 2$) homogeneous regions (clusters 1, 2, 4, and 6) (Table 4). Clusters 3 and 7 are strongly heterogeneous. The H_1-test was not applied for cluster 5, because there are only two sites in this cluster. For these three regions the assumptions of PREC are not fulfilled. Thus a PREC was only calculated for the clusters 1, 2, 4 and 6.

The RoI approach provides one region for each of the 89 gauging stations. As outlined in “Formation of homogeneous regions”, three different thresholds of the similarity measure were applied. The total number of PREC realisations is lower than 89, because in several cases the number of sites in the RoI is lower than four (Table 5). Only for 50 sites, there are at least four sites in the physiographical space with a Euclidean distance lower than 0.5. It becomes apparent that, also for the RoI approach, the method of PREC is not applicable for all gauging stations.

In summary, with both pooling methods heterogeneous regions were constructed, for which it was impossible to calculate a PREC. As mentioned before, this deficiency could partly be compensated by the use of different subsets of catchment descriptors.

4.3 Analysis of homogeneous regions for different candidate sets of catchment descriptors

Since 20 subsets of catchment descriptors were selected and the cluster analysis was performed five times (number of clusters from 3 to 7), altogether 500 regions were constructed and checked for homogeneity by the Hosking-Wallis test. The fraction of homogeneous regions ($H_1 < 2$) is in the range between 43% (3 cluster) and 54% (7 cluster) for the different numbers of clusters (Table 6).

With the RoI approach, one region was formed for each gauging station and each subset of catchment descriptors. The fraction of homogeneous regions is strongly influenced by the threshold of the Euclidean distance in the physiographical space. The number of homogeneous regions decreases from 54% for a threshold of 0.5 to 12% for a threshold of 2. As expected, both methods reveal that the fraction of homogeneous regions increases with a decreasing number of gauging stations (higher number of clusters, lower RoI-threshold).

The distribution of the relative number of homogeneous regions shows a spatial pattern for both pooling methods (Fig. 3). The gauging stations in the *Erzgebirge* are mostly grouped in homogeneous regions. In contrast, there are no or only a low number of homogeneous regions for several gauges in the Weisse Elster subbasin and east of the Elbe. The relative number of homogeneous regions is larger for the cluster analysis than for the RoI approach. This can be explained by the low number of homogeneous regions that were constructed for a threshold of two in the RoI approach (Table 6).

4.4 PREC results for candidate sets of catchment descriptors

Due to the fact that one PREC is provided for each homogeneous region, it is not possible to show all PREC realisations for all sites. All PREC realisations for the gauging station Dohna/Müglitz are shown as an example in Fig. 4. In addition, the pairs of the unit flood of record and the drainage area, which determine the intercept of PREC, are highlighted by black circles. The site itself is indicated separately. Both figures illustrate the influence of different subsets of catchment descriptors and pooling methods on the results of PREC.

Besides the slope and the intercept, also the range of the catchment size that is covered by the PREC depends on the constitution of the pooling group. As expected, the slope decreases with catchment size with two exceptions for RoI. In the example shown in Fig. 4 four sites govern the intercept of PREC including the selected site itself for both pooling methods.

As illustrated in Fig. 5, the results of PREC for the gauge Dohna differ in discharge (400–630 m^3/s) and recurrence interval (300–1200 years) for the two pooling schemes, as well as for different subsets of catchment descriptors. As expected, the discharge augments with increasing recurrence interval. The site itself has only a minor influence on the recurrence interval, because all AMS of the region are collected together (overall sample years of data).

Both pooling methods show the influence of the pair of the unit flood of record and drainage area, which determines the intercept of PREC. All discharges are at least 400 m^3/s, which is the flood of record at the gauge Dohna. In this example the PREC results of both methods are scattered in three groups. In the first group, the gauge Dohna itself determines the intercept of PREC. The gauge Dippoldiswalde and Rehefeld or Hainsberg 1 in the case of the cluster analysis or RoI, respectively, have the highest unit flood of record for the PREC realisations of the second group, where the discharge varies between 400 and 480 m³/s (Figs. 4 and 5). In the third group, the discharge of PRECs for the gauge Dohna is between 580 and 630 m³/s. The intercept of these PRECs is determined by the gauge Neundorf and in two cases for the cluster analysis also by Rehefeld. The range is caused by the different slopes of the PRECs, which were derived for pooling groups with different combinations of gauges. The higher the difference in the catchment size (e.g. Rehefeld (15 km²) and Dohna (198 km²), (see Fig. 4)), the larger is the PREC discharge affected by a variation of the slope.

The three groups of PREC realisations show that the inclusion of a gauge with a high unit flood of record (here: Neundorf) results in an upshift of the PREC. The extent of the upshift depends on the difference between the unit flood of record of the site of interest and the highest unit flood of record in the homogeneous group. It is important to highlight that Dohna and Neundorf have a relatively high unit flood of record. For a gauging station with a lower unit flood of record, the difference between the unit flood of record and the regional envelope curve discharge might be significantly higher, if the PREC is also determined by Neundorf.

4.5 Performance evaluation of PREC

The reliability of the PREC was evaluated by a leave-one-out jackknifing procedure (PREC-JK). The relative error of the PREC-JK to the index flood method was calculated for each gauging station (see Eq. (8)). In Fig. 6, only those gauging stations were considered, which had at least eight PREC-JK realisations. This criterion was fulfilled for 68 (Cluster analysis) and 61 sites (RoI), with on average 44 and 21 PREC-JK realisations, respectively.

The PREC-JK approach for both pooling methods illustrates that the median of the relative error is in most cases positive (Fig. 6). A high positive relative error indicates a high over-estimation of the discharge of PREC-JK for this recurrence interval in comparison to the index flood method. A negative relative error occurs for the gauging stations which determine the intercept of REC or which are close to the REC (see Fig. 7). Comparing the pooling methods, the relative errors (median of the box) as well as the scatter (size of the box) are similar for cluster analysis and RoI (Fig. 6).

The relative error between PREC-JK and the index flood method depends on the position of the gauging station in the 'unit discharge-area plot' (Fig. 7). If the unit flood of record q_{FOR} of a gauging station is close to the regional envelope curve, the unit discharge $q_{PREC\text{-}JK}$ derived from the regional envelope curve for this station is similar to or lower than that of the index flood method. In contrast, the higher the difference between the regional envelope curve q_{PREC} and the flood of record discharge q_{FOR} for a gauging station, the higher the relative error of PREC-JK in comparison to the index flood method. This relationship has a correlation coefficient of 0.73 (see Fig. 7).

4.6 Assessing the effect of the threshold of the heterogeneity measure

The homogeneity of a pooling group is a fundamental assumption of PREC. The influence of the degree of homogeneity on PREC was determined by varying the threshold of the heterogeneity measure. In order to consider the influence of the threshold on PREC, the sensitivity analysis was repeated for stronger ($H_1 < 1$) and weaker ($H_1 < 4$) thresholds of the Hosking-Wallis test. Following the classification of Hosking and Wallis (1997), a threshold of $H_1 < 1$ means that 'possibly homogeneous regions' ($1 < H_1 < 2$) are excluded (Table 1). By increasing the threshold to four, also 'slightly heterogeneous regions' ($2 < H_1 < 4$) are included. In this case only 'strong heterogeneous regions' ($H_1 > 4$) are excluded. The influence of the relative number of homogeneous regions for different thresholds of the Hosking-Wallis test has been discussed by Cunderlik and Burn (2002). An increase of H_1 from 2 to 4 results in a larger number of homogeneous regions (Fig. 8). This is especially relevant for those gauging stations, which were only seldom grouped in a homogeneous region when applying the strict definitions of homogeneity.

A comparison of the mean absolute relative error for the three thresholds illustrates that an increase in the degree of heterogeneity leads to a higher mean absolute relative error for most of the gauging stations and for both pooling methods (Fig. 9, Table 7). In addition, there are more results of the mean absolute relative error for $H_1 < 4$ because of the higher number of PREC realisations.

Considering that the relative error was calculated with the index flood method as reference, it is necessary to mention that the index flood estimate is subject to a higher uncertainty due to the higher degree of heterogeneity.

An overall performance indice was calculated as follows. All sites were selected which had at least four realisations for both pooling methods (see Table 7). The mean and the standard deviation of the absolute relative errors were calculated for all PREC realisations of these sites (n in Table 7). Both were averaged over the n sites. These performance indices increase with a higher degree of heterogeneity (Table 7). The result emphasises the relevance of the homogeneity criteria for PREC. The two performance indices are similar for the cluster analysis and RoI for the three thresholds of heterogeneity.

5 Discussion

The method of probabilistic regional envelope curves (PREC) derives a flood discharge and its recurrence interval for a homogeneous group of discharge gauges. One main assumption is its applicability in a homogeneous region in terms of the index flood method.

By using different subsets of catchment descriptors and two pooling methods (cluster analysis and RoI), a large number of homogeneous regions, which fulfilled the heterogeneity measure of Hosking and Wallis (1993), was derived for the mountainous catchments in Saxony. In contrast, the gauges located in the lowlands were mostly grouped in heterogeneous regions, which mean that the method of PREC could not be applied.

The reliability of PREC was assessed by a cross-validation procedure and a comparison with the index flood method. For a better understanding of the cross-validation results, it is worth emphasising an important difference between the index flood method and the PRECs. The index flood method represents the mean flood behaviour in a homogeneous region by a regional growth curve. Under this assumption it is expected that there are very small differences between the at-site flood behaviour and the regional distribution function in a homogeneous region. In contrast, the regional envelope curve is governed by the highest flood of record in a homogeneous region. Under the assumption that the estimation of the flood of record is more uncertain than the estimation of the index flood, the PREC is more sensitive to gauging stations with a high difference of an at-site flood of record to PREC than the index flood estimation.

The results of the PREC can be compared with a traditional at-site flood frequency analysis. The example of Dohna shows that most of the PREC realisations are close to the GEV distribution function (Fig. 10). This fact enhances the accuracy of the flood quantile estimates for high recurrence intervals. If there were large deviations between PREC and at-site flood frequency analysis, a more detailed consideration of the hydrologic situation at this gauge would be required.

It is important to highlight an essential difference of the PREC in comparison to other regionalisation methods. The magnitude of the recurrence interval of a PREC is mainly governed by one data point, i.e. the pair of the maximum unit flood of record and its drainage area. Castellarin et al. (2005) emphasised that a discordant site might reduce the use of the PREC method, since the recurrence interval is governed by the largest standardised maximum flood.

In other flood regionalisation methods (e.g. index flood, multiple regressions) commonly all sites have the same influence or their influence is weighted according to a selected weighting scheme. Sites, which are closer to other stations in a real or physiographical space, have higher weights. Consequently, the effect of a discordant site could be reduced by weighting the sites according to their similarity to the considered site or by averaging the values for all sites of a region. However, in the PREC concept weighting or averaging of sites is not possible when deriving the intercept of the PREC. Thus, in the PREC concept, the site that determines the intercept, plays an exceptional role. Because of that, an appropriate construction of homogeneous pooling groups is extremely important for PRECs.

The explicit estimation of a recurrence interval in the PREC scheme is another difference to traditional regional flood frequency methods. Whereas a target recurrence interval might be predefined in traditional approaches, the recurrence interval of PREC could only be approximately approached by the number of sites within a pooling group.

6 Conclusion

In this study the method of probabilistic regional envelope curves (PREC) was applied for the first time outside the original study area in Italy. It was shown that the transfer of this method to another region with different geographical conditions is possible. The goal of this paper was to quantify the influence of the pooling methods on PREC and to determine the sensitivity of PREC flood quantiles within different pooling groups. A combination of PREC and the RoI approach was introduced and compared with fixed homogeneous regions.

The main outcomes of this study are:

(1) The number of homogeneous regions strongly depends on the physiographic conditions of the catchment. The application of both pooling methods with different

candidate sets of catchment descriptors leads to a high number of homogeneous regions for the mountainous catchments and to a lower number for gauges in the lowlands and the eastern part of Saxony.

(2) A sensitivity analysis illustrates that PREC flood quantiles change in discharge as well as in the assigned recurrence interval depending on the constitution of the pooling group. It is thus recommended to compare different subsets as demonstrated in this study instead of using only the best subset of predictions.

(3) A leave-one-out jackknifing approach for ungauged conditions emphasises a similar relative error of the PREC results for both pooling methods (cluster analysis, RoI). An overall performance indice also affirms an increasing absolute relative error for different degrees of heterogeneity.

Acknowledgements

This work is part of the Center for Disaster Management and Risk Reduction Technologies (CEDIM) (http://www.cedim.de), a joint venture between the Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences, the Technical University of Karlsruhe (TH) and the Forschungszentrum Karlsruhe. We thank CEDIM and the GFZ for the financial support. Attilio Castellarin acknowledges the Italian Ministry of Education for the grant no. 2007HBTS85_003.

We thank the State Agency of Environment and Geology of the Free State of Saxony for the permission to use the discharge data and the hydrogeological map (HÜK 200) (http://www.umwelt.sachsen.de/umwelt/geologie/7703.htm, 05.02.2007). Furthermore we thank the State Agency of Flood Prevention and Water Management of Saxony-Anhalt and the State Agency for Environment of Thuringia for additional discharge data. We also thank the Potsdam Institute for Climate Impact Research (PIK) and the German Weather Service (DWD) for the precipitation data as well as the Federal Agency for Cartography and Geodesy

of Germany (BKG) for the ATKIS-Basis-DLM and the digital elevation model for Saxony. We also thank the European Commission – Joint Research Centre (JRC) for the CCM River and Catchment Database. The SRTM Digital Terrain Model was downloaded from (http://srtm.csi.cgiar.org/SELECTION/inputCoord.asp, 19.05.2008).
We acknowledge the very helpful recommendations of an anonymous reviewer and Alberto Montanari.

REFERENCES

Acreman, M.C., Sinclair, C.D., 1986. Classification of drainage basins according to their physical characteristics: an application for flood frequency analysis in Scotland. J. Hydrol. 84 (3), 365-380.

AG Boden, 1994. Bodenkundliche Kartieranleitung, Schweizerbart'sche Verlagsbuchhandlung, Hannover, 392 pp.

Beurton, S., Thieken, A.H., 2009. Seasonality of floods in Germany. Hydrolog. Sci. J. 54 (1), 62-76.

BKG Geodatenzentrum (Federal Agency for Cartography and Geodesy - GeoDataCentre), 2005. Digital Landscape Model ATKIS Basis DLM, Frankfurt/ Main.

Boorman, D.B., Hollis, J.M., Lilly, A., 1995. Hydrology of Soil Types: A Hydrologically Based Classification of the Soils of the United Kingdom, IAHS Report 126. IAHS Press, Wallingford, UK.

Burn, D.H., 1989. Cluster analysis as applied to regional flood frequency analysis. J. Water Res. Pl.-ASCE 115 (5), 567-582.

Burn, D.H., 1990. Evaluation of Regional Flood Frequency Analysis with a Region of Influence Approach. Water Resour. Res. 26 (10), 2257-2265.

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

Burn, D.H., Goel, N.K., 2000. The formation of groups for regional flood frequency analysis. Hydrolog. Sci. J. 45 (1), 97-112.

Castellarin, A., 2007. Probabilistic envelope curves for design flood estimation at ungauged sites. Water Resour. Res. 43 (4), W04406. doi:10.1029/2005WR004384.

Castellarin, A., Burn, D.H., Brath, A., 2001. Assessing the effectiveness of hydrological similarity measures for flood frequency analysis. J. Hydrol. 241 (3), 270-285.

Castellarin, A., Galeati, G., Brandimarte, L., Montanari, A., Brath, A., 2004. Regional flow-duration curves: reliability for ungauged basins. Adv. Water Resour. 27, 953-965.

Castellarin, A., Vogel, R.M., Matalas, N.C., 2005. Probabilistic behaviour of a regional envelope curve. Water Resour. Res. 41, W06018. doi: 10.1029/2004WR003042.

Castellarin, A., Vogel, R.M., Matalas, N.C., 2007. Multivariate probabilistic regional envelopes of extreme floods. J. Hydrol. 336 (3-4), 376-390.

Castellarin, A., Burn, D.H., Brath, A., 2008. Homogeneity testing: How homogeneous do heterogeneous cross-correlated regions seem? J. Hydrol. 360 (1-4), 67-76.

Costa, J.E., 1987. A comparison of the largest rainfall-runoff floods in the United States with those of the People's Republic of China and the world. J. Hydrol. 96 (1-4), 101-115.

Crippen, J.R., Bue, C.D., 1977. Maximum Flood Flows in the Conterminous United States. Geological Survey Water-Supply Paper 1887, United States Printing Office, Washington DC.

Cunderlik, J.M., Burn, D.H., 2002. The use of flood regime information in regional flood frequency analysis. Hydrolog. Sci. J. 47 (1), 77-92.

Cunderlik, J.M., Burn, D.H., 2006. Switching the pooling similarity distances: Mahalanobis for Euclidean. Water Resour. Res. 42, W03409. doi:10.1029/2005WR004245.

Dalrymple, T., 1960. Flood frequency analyses. US Geol. Surv. Water Supply Pap.: 1543-A, Reston, VA.

Eng, K., Tasker, G.D., Milly, P.C.D., 2005. An analysis of Region of Influence methods for flood regionalization in the Gulf Atlantic Rolling Plains. J. Am. Water Resour. Assoc. 41 (1), 135-143.

Eng, K., Milly, P.C.D., Tasker, G.D., 2007. Flood Regionalization: A Hybrid Geographic and Predictor-Variable Region of Influence Regression Method. J. Hydraul. Eng.-ASCE 12 (6), 585-591.

Flemming, G., 2001. Angewandte Klimatologie von Sachsen. Tharandter Klimaprotokolle, Band 4, Technical University of Dresden, Tharandt.

Gaál, L., Kyselý, J., Szolgay, J., 2008. Region of Influence approach to a frequency analysis of heavy precipitation in Slovakia. Hydrol. Earth Syst. Sci. 12 (3), 825-839.

GREHYS, 1996. Presentation and review of some methods for regional flood frequency analysis. J. Hydrol. 186 (1-4), 63-84.

Guse, B., Castellarin, A., Thieken, A.H., Merz, B., 2009. Effects of intersite dependence of nested catchment structures on probabilistic regional envelope curves. Hydrol. Earth Syst. Sci. 13 (9), 1699-1712.

Herschy, R., 2002. The world's maximum observed floods. Flow Meas. Instrum., 13, 231-235.

Hirsch, R.M., Helsel, D.R., Cohn, T.A., Gilroy, E.J., 1992. Statistical analysis of hydrological data. In: Maidment, D.A. (Ed.), Handbook of Hydrology. McGraw-Hill, New York, pp. 17.1 - 17.55.

Hosking, J.R.M., Wallis, J.R., 1988. The effect of intersite dependence on regional flood frequency analysis. Water Resour. Res. 24 (4), 588-600.

Hosking, J.R.M., Wallis, J.R., 1993. Some statistics useful in regional frequency analysis. J. Hydrol. 29 (2), 271-281.

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

IKSE, 2004. Dokumentation des Hochwassers vom August 2002 im Einzugsgebiet der Elbe, IKSE (Internationale Kommission zum Schutz der Elbe), Magdeburg.

Jarvis, C.S., 1925. Flood Flow Characteristics. Trans. Am. Soc. Civil Eng. 88, 985-1032.

Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., 2008. Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database. <http://srtm.csi.cgiar.org>.

Lettenmaier, D.P., Wallis, J.R., Wood, E.F., 1987. Effect of regional heterogeneity on flood frequency estimation. Water Resour. Res. 23 (2), 313-323.

Merz, R., 2006. Regionalisierung von statistischen Hochwasserkenngrößen. In: Gutknecht, D. (Ed.), Methoden der hydrologischen Regionalisierung, Wiener Mitteilungen, Wasser-Abwasser-Gewässer, Band 197, Institut für Wasserbau und Ingenieurhydrologie, Technische Universität Wien, pp. 109-130.

Merz, R., Blöschl, G., 2004. Regionalisation of catchment model parameters. J. Hydrol. 287 (1-4), 95-123.

Merz, R., Blöschl, G., 2005. Flood frequency regionalisation - spatial proximity vs. catchment attributes. J. Hydrol. 302 (1-4), 283-306.

Merz, B., Thieken, A.H., 2005. Separating natural and epistemic uncertainty in flood frequency analysis. J. Hydrol. 309 (1-4), 114-132.

Mimikou, M., 1984. Envelope curves for extreme flood events in North-Western and Western Greece. J. Hydrol. 67 (1-4), 55-66.

Montgomery, D.C., Peck, E.A., Vining, G.G., 2001. Introduction to Linear Regression Analysis, third ed. John Wiley and Sons, Inc., New York, NY.

Mosley, M.P., 1981. Delimination of New Zealand hydrologic regions. J. Hydrol. 49 (1-2), 173-192.

Nathan, R.J., McMahon, T.A., 1990. Identification of homogeneous regions for the purposes of regionalization. J. Hydrol. 121 (1-4), 217-238.

Ouarda, T.B.M.J., Girard, C., Cavadias, G., Bobée, B., 2001. Regional flood frequency estimation with canonical correlation analysis. J. Hydrol. 254 (1-4), 157-173.

Peel, M.C., Wang, Q. J., Vogel, R.M.. McMahon, T.A., 2001. The utility of L-moment ratio diagrams for selecting a regional probability distribution. Hydrolog. Sci. J. 46 (1), 147-156.

Petrow, T., Merz, B., Lindenschmidt, K.-E., Thieken, A.H., 2007. Aspects of seasonality and flood generating circulation patterns in a mountainous catchment in south-eastern Germany. Hydrol. Earth Syst. Sci. 11 (4), 1455-1468.

Pitlick, J., 1994. Relation between peak flows, precipitation, and physiography for five mountainous regions in the western USA. J. Hydrol. 158 (3-4), 219-240.

Pohl, R., 2004. Historische Hochwasser aus dem Erzgebirge. Wasserbauliche Mitteilungen, Heft 28, Fakultät Bauingenieurwesen, Institut für Wasserbau und Technische Hydromechanik, Technische Universität Dresden, Dresden.

Robson, A., Reed, D., 1999. Flood Estimation Handbook 3: Statistical Procedures of Flood Frequency Estimation. Institute of Hydrology, Wallingford/UK. 338 pp.

Stedinger, J.R., Lu, L., 1995. Appraisal of regional and index flood quantile estimators. Stoch. Hydrol. Hydraul. 9 (1), 49-75.

Tasker, G.D., Stedinger, J.R., 1989. An operational GLS model for hydrologic regression. J. Hydrol. 111 (1-4), 361-375.

Thieken, A.H., Kreibich, H., Müller, M., Merz, B., 2007. Coping with floods: preparedness, response and recovery of flood-affected residents in Germany in 2002, Hydrolog. Sci. J. 52 (5), 1016-1037.

Troutman, B.M., Karlinger, M.R., 2003. Regional flood probabilities. Water Resour. Res. 39 (4), 1095. doi:10.1029/2001WR001140.

Uhlenbrook, S., Steinbrich, A., Tetzlaff, D., Leibundgut, C., 2000. Zusammenhang zwischen extremen Hochwassern und ihren Einflussgrößen. In: Arbeitskreis KLIWA (Eds.), KLIWA-Heft 1, KLIWA-Symposium - Klimaveränderung und Konsequenzen für die Wasserwirtschaft, Karlsruhe, pp. 187-203.

Ulbrich, U., Brücher, T., Fink, A.H., Leckebusch, G.C., Krüger, A., Pinto, J. G., 2003. The central European floods of August 2002: part 1 - Rainfall periods and flood development. Weather, 58, 371-377.

Viglione, A., 2008. The nsRFA Package, r package version 0.4-8.

Viglione, A., Laio, F., Claps, P., 2007. A comparison of homogeneity tests for regional frequency analysis. Water Resour. Res. 43, W03428, doi:10.1029/2006WR005095.

Vogel, R.M., Fennessey, N.M., 1993. L moment diagrams should replace product-moment diagrams. Water Resour. Res. 29 (6), 1745-1752.

Vogel, R.M., Matalas, N.C., England Jr., J.F., Castellarin, A., 2007. An assessment of exceedance probabilities of envelope curves. Water Resour. Res. 43 (7), W070403. doi:10.1029/2006WR005586.

Vogt, J.V., Soille, P., de Jager, A., Rimavičũité, E., Mehl, W., Foisneau, S., Bódis, K., Dusart, J., Paracchini, M.L., Haastrup, P., Bamps, C., 2007. A pan-European River and Catchment Database, European Commission, EUR 22920 EN – Joint Research Centre.

Wiltshire, S.E., 1986. Identification of homogeneous regions for flood frequency analysis. J. Hydrol. 84 (3-4), 287-302.

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

TABLES:

Table 1

Interpretation of the heterogeneity measure (Hosking and Wallis, 1993; Robson and Reed, 1999).

Heterogeneity measure	Interpretation	Review
< 1	Homogeneous	Not required
1 – 2	Possibly heterogeneous	Optional
2 – 4	Heterogeneous	Desirable
> 4	Strongly heterogeneous	Essential

Table 2

List of catchment descriptors.

Abbreviation	Catchment descriptors
MAP	Mean annual precipitation (mm)
MAXDAY	Maximum daily precipitation (mm)
P50	Annual frequency of days with precipitation of more than 50 mm/d (%)
MAX5DAY	Maximum precipitation in 5 days (mm)
PAMS	Mean of the annual maximum series of daily precipitation (mm)
ELEV	Mean elevation of the catchment (m asl)
SLOPE	Mean slope of the catchment (%)
RANGE_NORM	Range of catchment elevation, normalised with the catchment size ($10^{-3}m^{-1}$)
ARABLE	Fraction of arable land coverage (%)
URBAN	Fraction of urban land coverage (%)
MINING	Fraction of mining activities (%)
BEDROCK	Fraction of bedrock areas (%)
KF_LOW	Fraction of low permeability areas (%)

Table 3

Selected subsets of catchment descriptors (CD) and the correlation coefficient (COR) to the unit index flood of all gauging stations.

CD1	CD2	CD3	COR
MAX5DAY	ELEV	RANGE_NORM	0.70
MAX5DAY	RANGE_NORM	URBAN	0.69
MAP	MAX5DAY	RANGE_NORM	0.69
MAX5DAY	RANGE_NORM		0.68
MAX5DAY	ELEV	URBAN	0.68
ELEV	RANGE_NORM	URBAN	0.66
PAMS	RANGE_NORM	URBAN	0.64
MAX5DAY	ELEV		0.64
ELEV	RANGE_NORM		0.64
MAP	MAX5DAY	URBAN	0.64
MAP	MAX5DAY		0.62
MAP	RANGE_NORM		0.62
PAMS	RANGE_NORM		0.62
P50	RANGE_NORM	URBAN	0.61
MAX5DAY	ARABLE	URBAN	0.61
MAXDAY	RANGE_NORM	URBAN	0.61
MAX5DAY	URBAN	BEDROCK	0.61
MAX5DAY	PAMS	URBAN	0.61
RANGE_NORM	URBAN	BEDROCK	0.60
RANGE_NORM	BEDROCK		0.60

Table 4

Results of heterogeneity measure and of PREC method for the best subset of catchment descriptors, derived by cluster analysis for the seven-cluster solution.

Cluster	1	2	3	4	5	6	7
Number of gauges	7	24	10	18	2	8	20
H_1	0.6	0.8	7.8	1.5		-1.4	9.3
Number of observations	277	1471		1326		498	
Effective number of observations	160	483		403		202	
Recurrence interval [a]	320	966		805		403	

Table 5

Number of sites below and above the threshold (H_1=2) of the heterogeneity measure for the best subset of catchment descriptors constructed by the Region of Influence approach using different thresholds of the Euclidean distance.

Threshold	$H_1 < 2$	$H_1 > 2$	Sum
0.5	28	22	50
1	27	49	76
2	14	74	88

Table 6

Number of homogeneous regions derived by cluster analysis and Region of Influence (RoI).

Number of clusters	$H_1 < 2$	$H_1 > 2$	$H_1 < 2$ [%]
3	26	34	43.3
4	35	44	44.3
5	48	47	50.5
6	58	55	51.3
7	67	58	53.6
RoI-threshold			
0.5	575	493	53.8
1	628	1002	38.5
2	212	1539	12.1

The H_1-test was not applied for pooling groups with less than four sites.

Table 7

Overall performance indices of the jackknifing procedure for both pooling methods and the different thresholds of the heterogeneity measure.

	Cluster analysis	Region of Influence
$H_1 < 1$: n = 57		
Mean of the mean absolute relative error	0.54	0.54
Mean of the standard deviation of absolute relative error	0.21	0.26
$H_1 < 2$: n = 70		
Mean of the mean absolute relative error	0.81	0.69
Mean of the standard deviation of absolute relative error	0.36	0.40
$H_1 < 4$: n = 75		
Mean of the mean absolute relative error	1.12	0.88
Mean of the standard deviation of absolute relative error	0.56	0.53

n = Number of sites with at least four PREC realisations

FIGURES

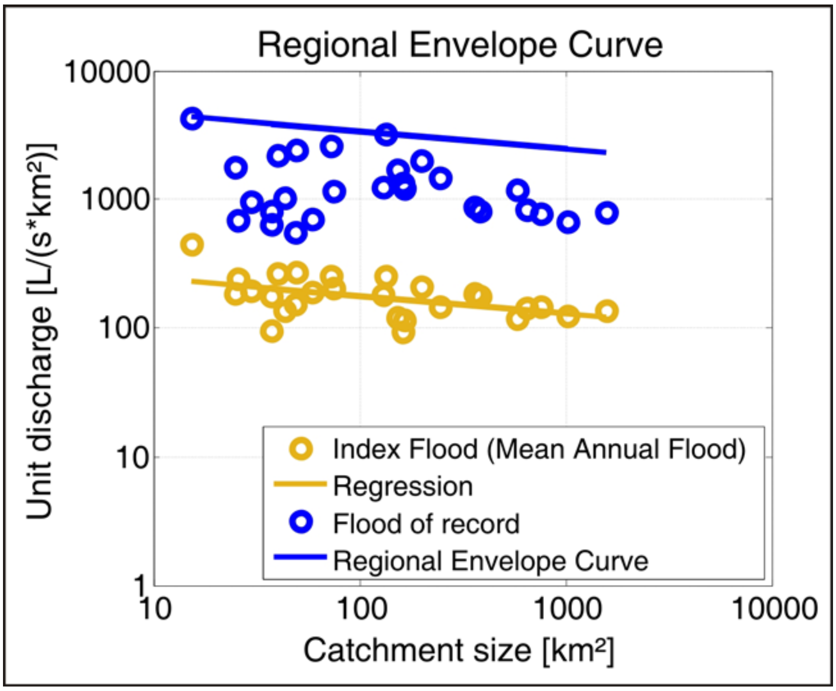

Fig. 1: Example of a Regional Envelope Curve.

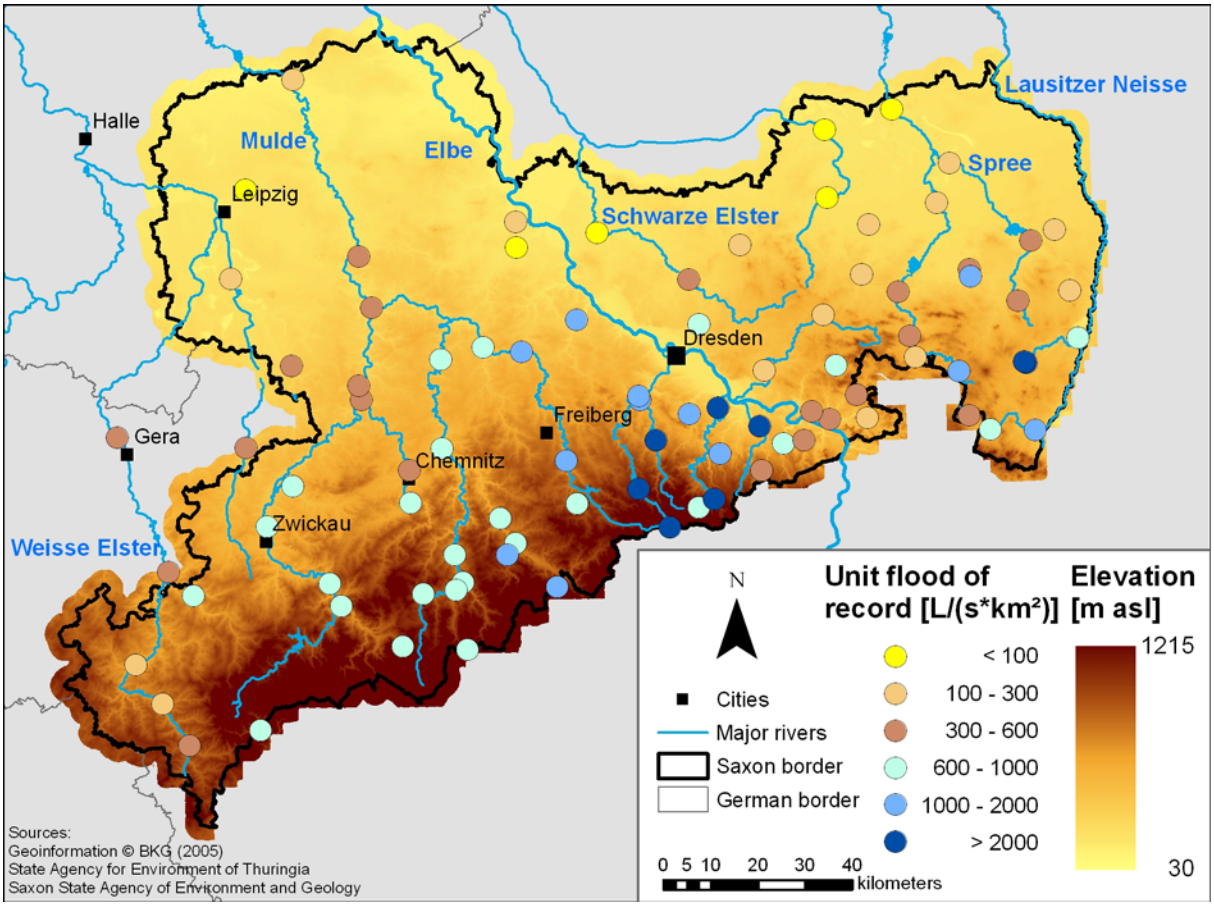

Fig. 2: Study area: Elevation above sea level in the federal state of Saxony, Germany, and available discharge gauging stations coloured by the unit flood of record.

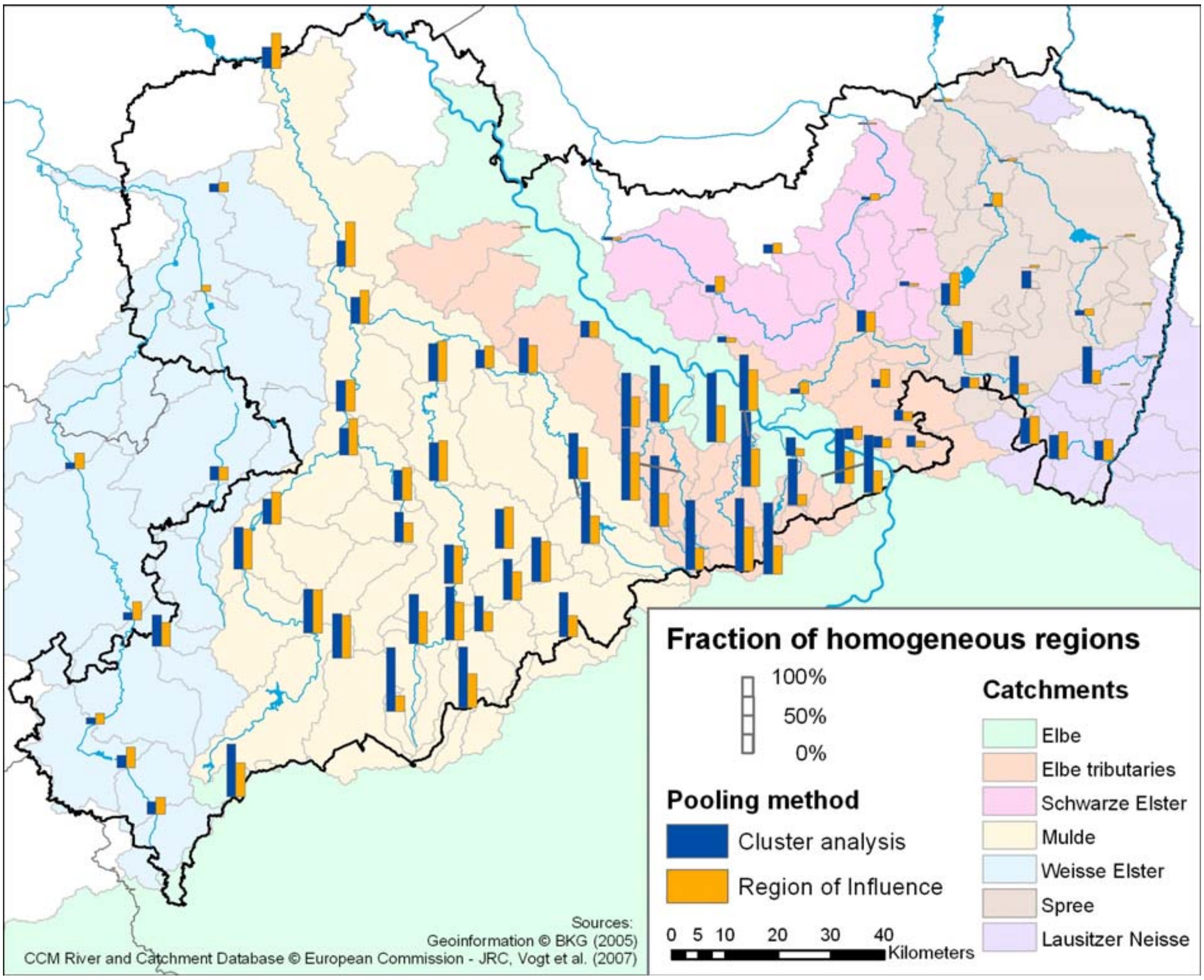

Fig. 3: Fraction of homogeneous regions ($H_1<2$) [%] by cluster analysis and Region of Influence for the gauging stations in the study area.

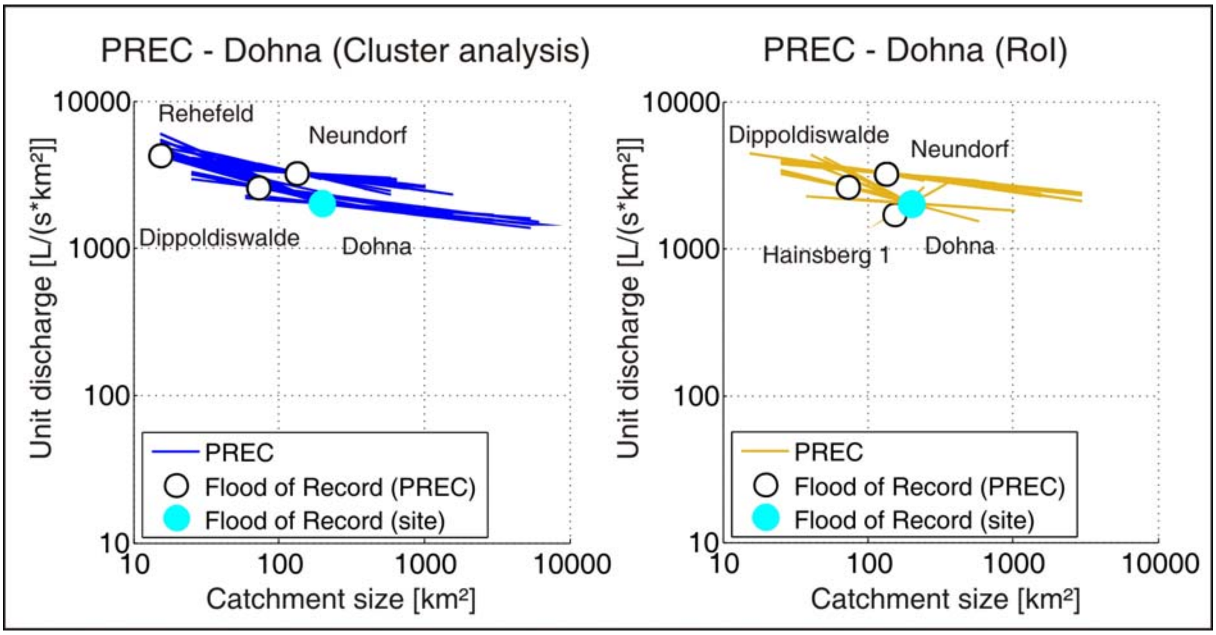

Fig. 4: All PREC realisations for the gauge Dohna in homogeneous regions derived by cluster analysis (left) and RoI (right).

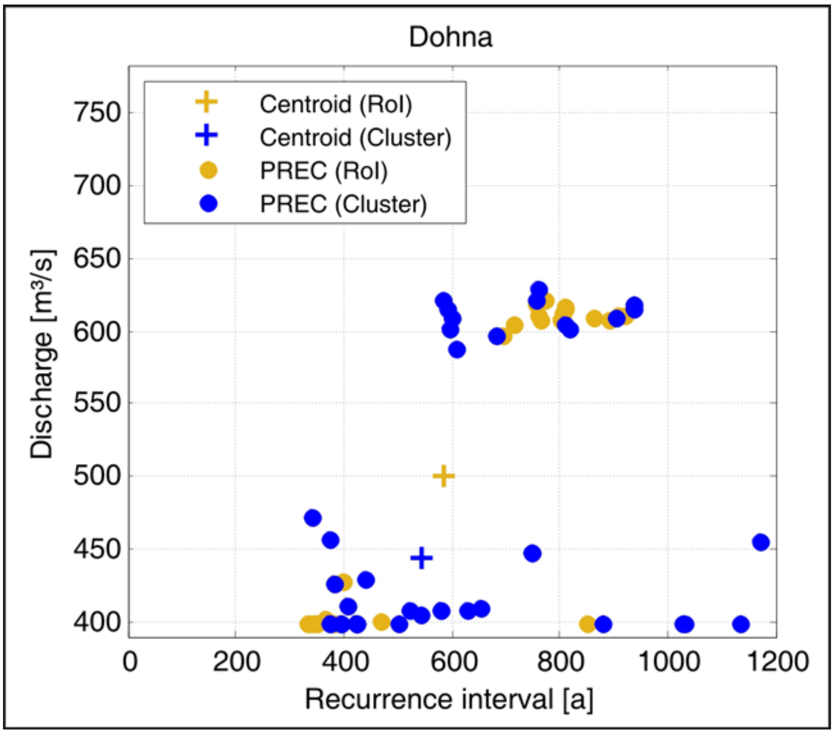

Fig. 5: Pairs of discharges and recurrence intervals for all PREC realisations of the gauge Dohna.

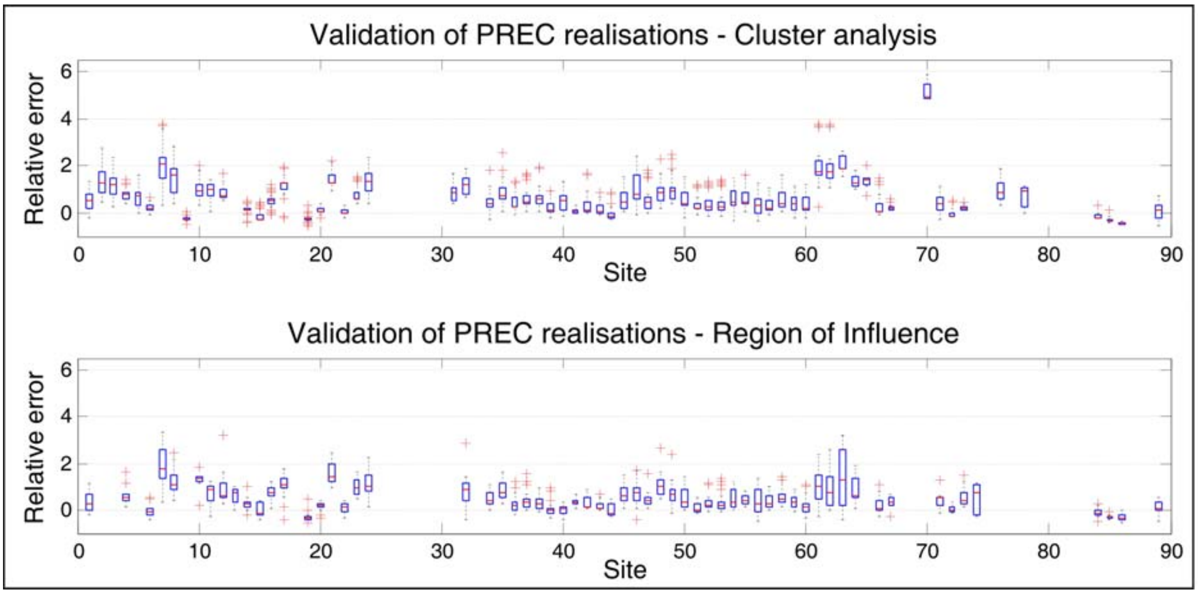

Fig. 6: Relative error of the PREC realisations for the two pooling methods cluster analysis (top) and Region of Influence (bottom) for the 89 sites of the study area. The boxplot edges are formed by the 25th and 75th percentiles. Outliers are illustrated with red crosses.

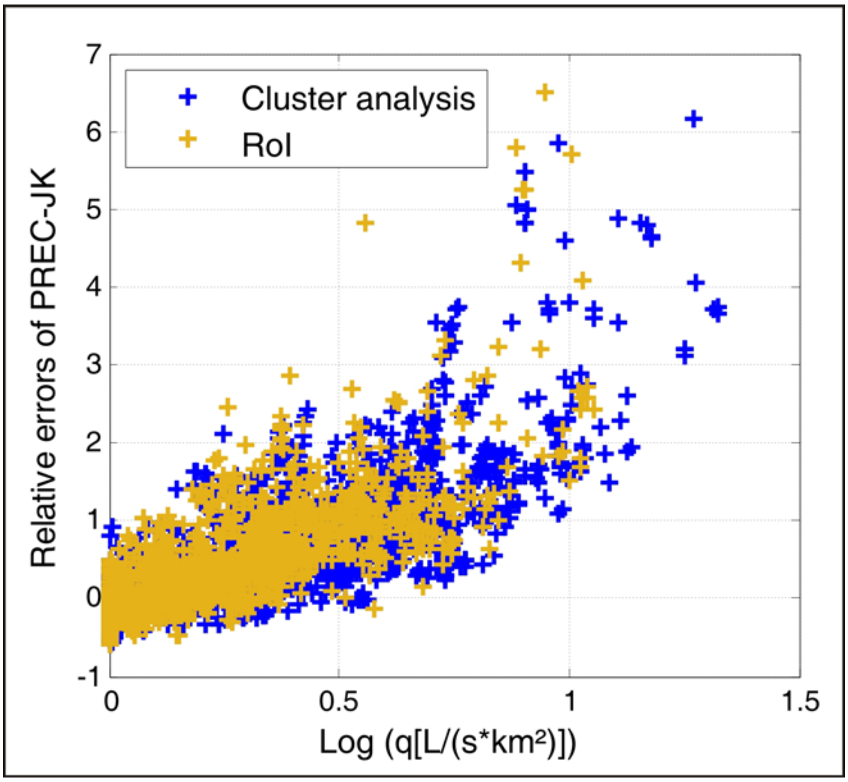

Fig. 7: Relative error of PREC-JK versus the distance of the unit flood of record q_{FOR} to q_{PREC} for pooling groups identified by cluster analysis and the Region of Influence approach.

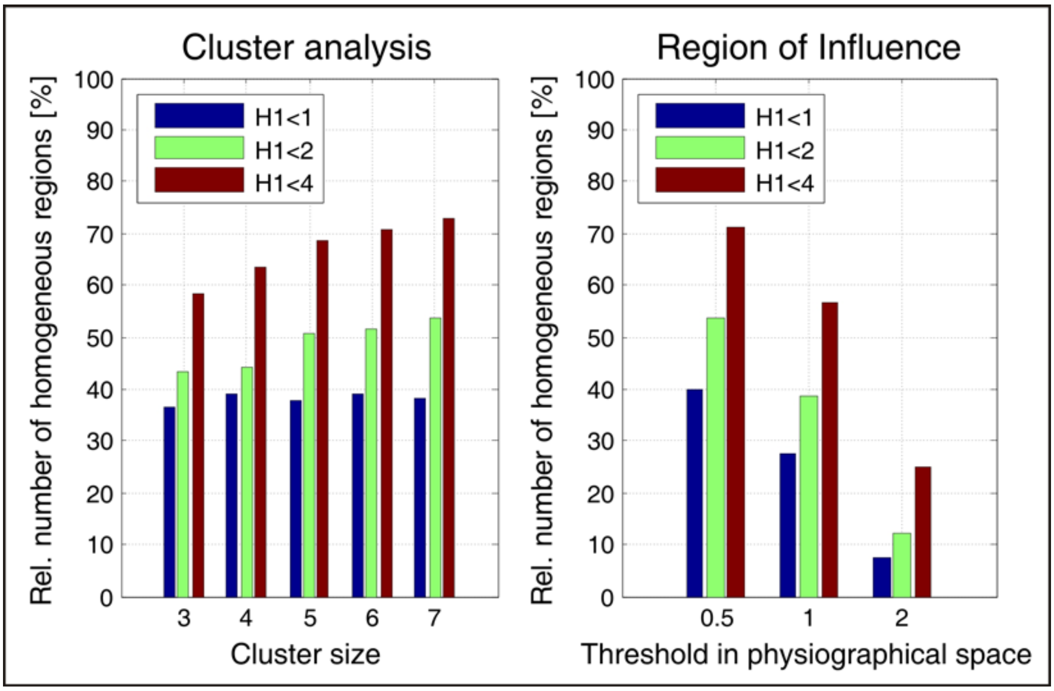

Fig. 8: Relative number of homogeneous regions for different thresholds of heterogeneity for cluster analysis and Region of Influence.

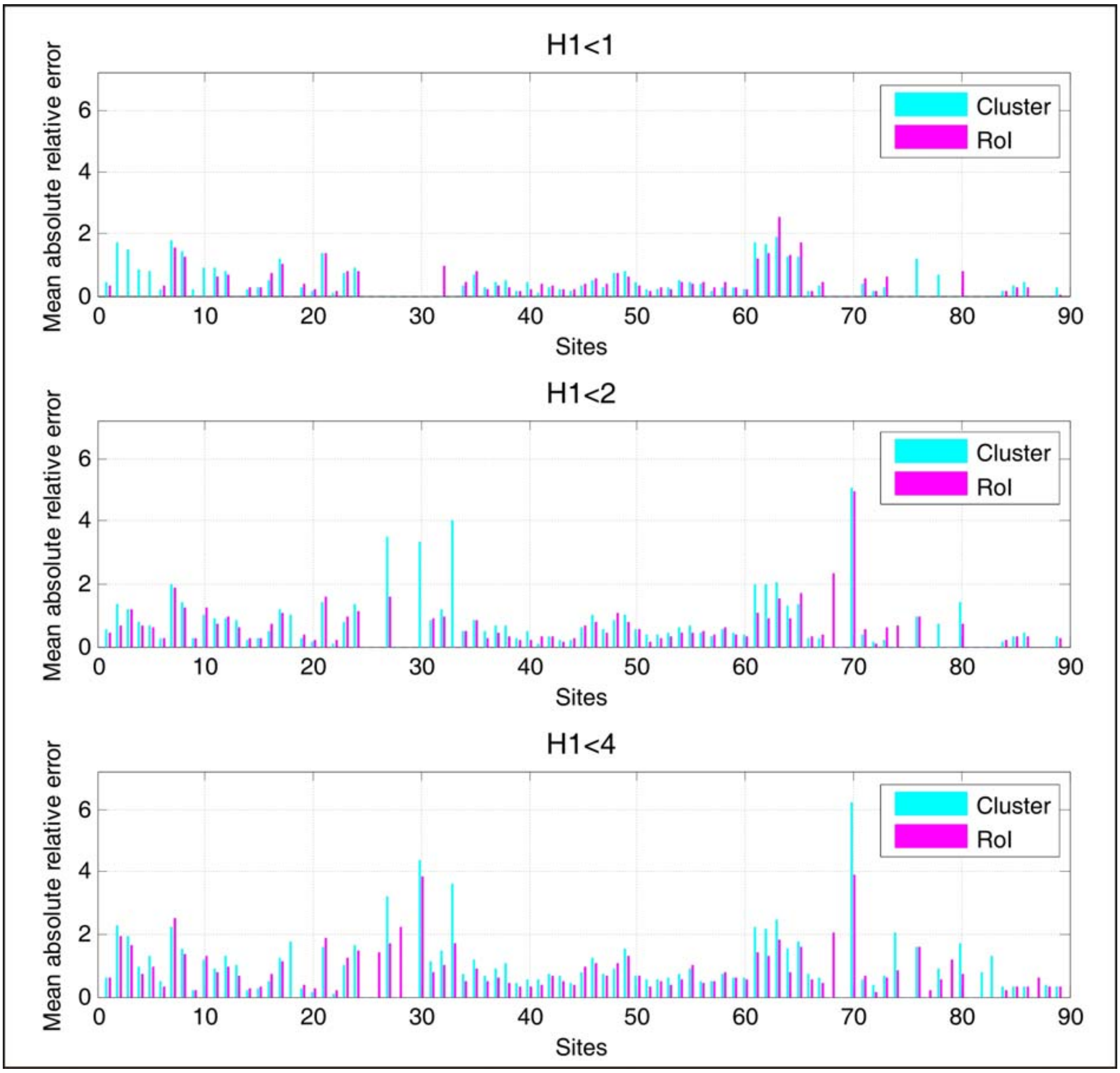

Fig. 9: Mean absolute relative error of PREC-JK for both pooling methods (cluster analysis, Region of Influence (RoI)) using different thresholds of the heterogeneity measure H_1. The mean absolute relative error is illustrated for sites with at least four PREC-JK realisations.

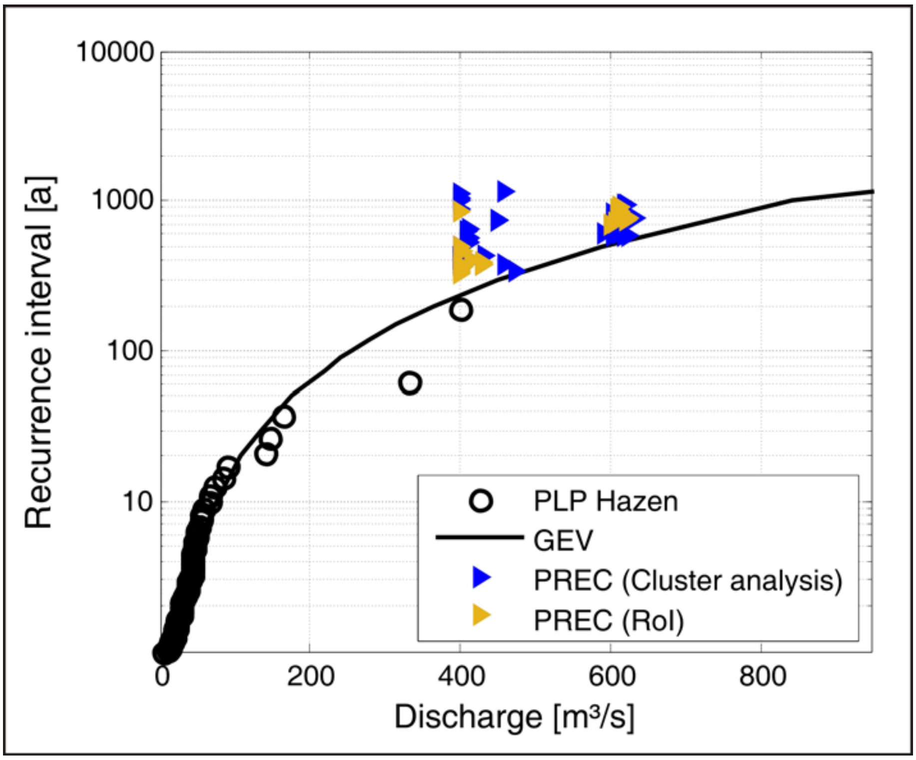

Fig. 10: Comparison of PREC results of both pooling methods with at-site flood frequency analysis (GEV) for Dohna.