Regional frequency analysis of rainfall in Germany
Regionálna frekvenčná analýza zrážok v Nemecku

Bc. Eva Uhliarová

Abstrakt

V príspevku je opísaná aplikácia regionálnej frekvenčnej analýzy založenej na L-momentoch na priemerných zrážkových údajoch v Nemecku. Analýza bola vykonaná podľa procedúry navrhnutej Hoskingom & Wallisom (1997) a bola vypracovaná v nasledujúcich krokoch: kontrola dát, identifikácia homogénnych regiónov a definovanie funkcie pravdepodobnostnej distribúcie. Analýza bola zameraná najmä na vytvorenie homogénnych regiónov, ktoré charakterizujú priebeh zrážok v regióne a môžu byť ďalej použité ako súčasť iných analýz, prípadne modelov v hydrologickej praxi.

Poskytnuté dáta pochádzali z 81 zrážkomerných staníc rozprestierajúcich sa v severo-západnom Nemecku s dĺžkou záznamov 6 rokov pre každú stanicu. Samotné zrážkové údaje boli analyzované v neštandardnej forme celkového množstva a trvania dažďovej udalosti, podľa vzoru vstupov do takzvaných ARM modelov (Haberlandt, 1998). Analyzované boli teda dve premenné oddelene a výsledky analýzy boli priebežne porovnávané.

Pre účely kontroly dát bol použitý test sériovej korelácie a D-test ako indikátor sporných hodnôt. Na identifikáciu homogénnych regiónov bola aplikovaná klastrová analýza, pričom boli porovnané jej dva prístupy: 1. Kombinácia Ward a K-means klastrovania a 2. Rozdeľovanie okolo medoidov. Ukázalo sa, že pre množstvo zrážok vytvára homogénnejšie regióny prvý spomínaný algoritmus, zatiaľ čo pre trvanie bola metóda rozdeľovania okolo medoidov vhodnejšia. Ako problematický sa ukázal krok priraďovania pravdepodobnostých distribúcií, kde kvôli priestorovej korelácii dát nebolo možné napasovať žiadnu zo skúšaných rozložení pre množstvo zrážok. Keďže, kvôli neštandardnému formátu dát nebolo možné zadefinovať korelačnú maticu a zakomponovať ju do výpočtov, zrážkovému množstvu bolo priradené komplexné 5-parametrálne rozloženie Wakeby. Pre trvanie zrážok najlepšie vyhovovalo generalizované normálne rozloženie. Vzhľadom na rozsahové obmedzenie príspevku nezaraďujem posledný krok analýzy: odhad kvantilov.

Anotácia: V príspevku je zhrnutá aplikácia regionálnej frekvenčnej analýzy (RFA) na zrážky v severo-západnom regióne Nemecka o približnej rozlohe 47,6 tis. km^2. Hlavným účelom analýzy bolo vytvorenie regiónov, čím sa pokryla celá oblasť známou krivkou, z ktorej je neskôr možné odvodiť návrhové hodnoty presnejšie ako pri použití regionálneho vzorca.

Kľúčové slová: zrážky, región, regionálna frekvenčná analýza, homogeneita

Annotation: This paper reports an application of the regional frequency analysis (RFA) on the precipitation data in the north-western Germany with area ca. 47,6 thd. km^2. The main purpose of the analysis was the creation of homogeneous regions from the studied area and the definition of functions that describe the variable in order to later derive reliable quantile estimates.

Key words: precipitation, region, regional frequency analysis, homogeneity

1. Introduction

The regional frequency analysis (RFA) is nowadays probably the most often used method for estimating design values of different hydrological variables worldwide applied on national scales Čunderlík (2000). It has been developed as a modification of the classical frequency analysis for sites with short or no available recordings, when Dalrymple (1960) proposed to "trade space for time" or to merge the observations from similar stations into one sample and treat them as a region instead of individually. The main task of RFA is thus the identification of regions into which sites with samples originating in the same population are merged. The method is mainly used for defining the growth or intensity-duration-frequency (IDF) curves from which design values are derived for example (Fowler & Kilsby, 2003 or Al Mamoon et al., 2014). Regarding the studied hydrological variable, the method was most frequently applied on extreme discharges (Mossaffaie, 2015), droughts (Modarres, 2010) or extreme precipitation (Yang et al., 2010), but interesting are also studies of extreme surges in coastal engineering (Weiss et al., 2012) or groundwater levels (Furst et al., 2011).

The *crucial skill* for a successful execution of the analysis is the knowledge of the hydrological regime in the studied area. Consequently, one can define a function which would describe the variance of the the studied variable for each region which further provides a reliable estimation possibilty for various needs but mainly for the extremes with high return periods.

The RFA was also used in Slovakia. First, Čunderlík (2000) studied annual maximum discharge based on the L-moments and divided Slovakia in 11 regions. Collective of Kohnová, Szolgay, Parajka and others worked in years 2000-2004 with annual maximum **discharge** and tested the application of different pooling methods. They divided the area of Slovakia into 4 regions and concluded that the RFA outperforms the previously used local estimations. Further in 2009 and 2010 they applied a different regionalization principle for mapping 100y quantiles of annual maximum of **daily precipitation**. They found that the so-called ”ROI” method in which each station is center of its own region suits better for complex terrain such as Slovakia than the more classical methods. Solín (2002, 2005 and 2006) studied T-year maximum **discharge** and formed 4 regions out of 156 small basins by means of logical principles. Gaál et al. (2004, 2008 and 2009) studied seasonal and annual maxima of 1-5 days **precipitation** and compared the subjective, objective and ROI approach of forming regions with better results from ROI and subjective repartitioning by a climatologist. Remiášová et al. (2011) focused on mapping maximum annual 1-day **precipitation** for ungauged areas and finally Portela et al. (2015) used RFA for a characterization of **droughts** in Slovakia. Gaál also aimed to perform an analysis of short-term heavy precipitation for whole Slovakia. In this paper, the results of one of the few application of RFA in Germany will be presented.

2. Description of data

The analysis was performed on precipitation data from 81 raingauge stations scattered through north-western Germany (see Fig. 1). Total area of interest is approximately 48 000 km2, which is comparable to the area of Slovak republic.

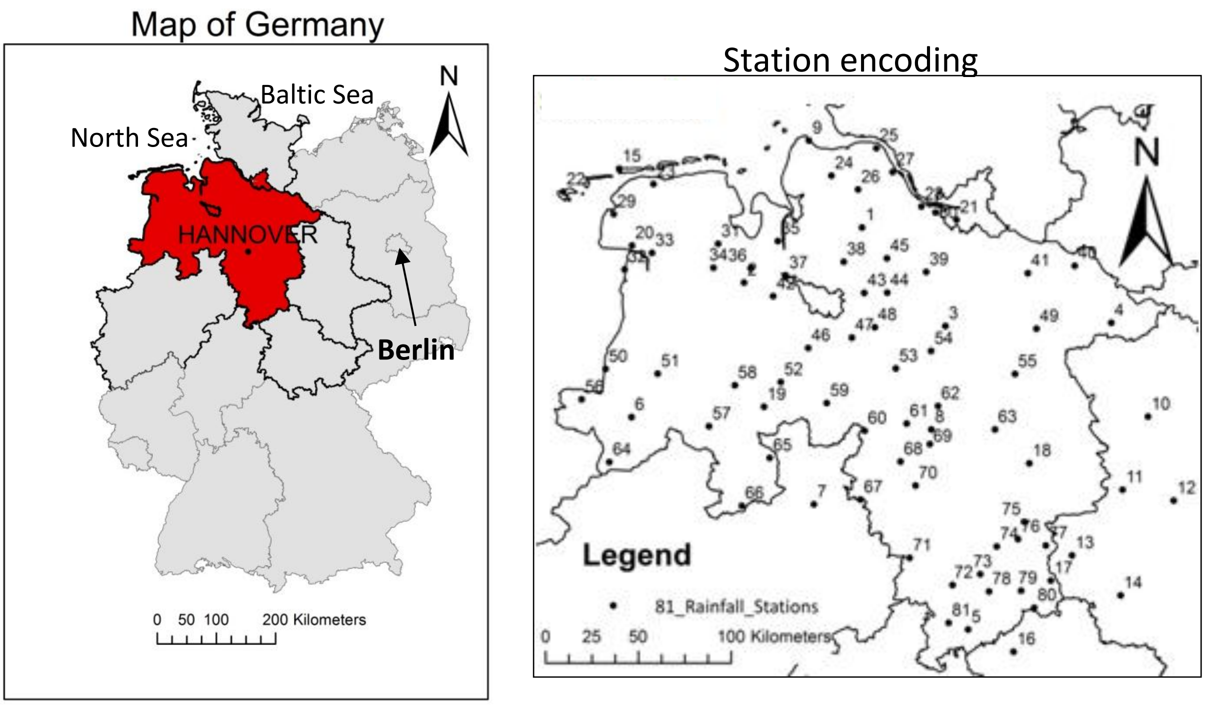

Fig. 1: Left: map of Germany with highlighted studied area, Right: Location of studied stations with the encoding used in the analysis

The studied area is mostly formed by the great North German Plain that is in the south bordered by the Weser, Deister, and the Harz mountains, and in the north by the North Sea. The region is mostly flat with > 80% of total area (61 of studied stations) located on altitude < 100 m a.s.l. The highest point of the area has altitude of 1088,4 m a.s.l. but the highest located station has the elevation of 607 m a.s.l. (see more details in Tab.1).

Tab. 1: Altitudinal coverage of studied area and corresponding number of rainfall stations

Classes [m a.s.l.]	Area [km2]	Percentage [%]	N of stations [-]
sea level	1123	2,30	2
0,1 - 100	38963	79,83	59
100,1 - 200	4782	9,80	8
200,1 - 300	2256	4,62	6
300,1 - 400	885	1,81	4
400,1 - 1088,4	801	1,64	2

As reported (Web: Klima und Wetter in Niedersachsen), the hydrological regime as well as climate in Lower Saxony is influenced: in the north-western part by Atlantic Ocean and has the characteristics of the western European maritime climate and in the south-eastern part by the continental climate coming from the Eastern Europe. In the coastal part, the Atlantic influence causes balanced temperatures over all seasons with mild winters as well as moderate summers and more precipitation. On the other hand, the seasonal differences are more marked in the regions´ capital - Hannover and in other south-eastern cities. The average precipitation amounts in different parts of region can be seen in the Tab.2.

Tab. 2: Annual precipitation in Lower Saxony, source: Elsholz & Berger (1989)

Annual Precipitation in Lower Saxony	**[mm/a]**
Regions average	730
North – eastern	550
Hannover	600
Coastal area	800
Western slopes of Weser mountains	800 - 900
In middle altitude of Harz mountains	1000 - 1300
At top elevations of Harz mountains	< 1500

The analysed data were provided by the German weather service – "Deutscher Wetterdienst (DWD)". The data encompass only the summer events and the common record length at each station was six years (01/2007 – 12/2012). The measurements were performed by two types of devices: 1. Drop Counter with resolution 0.01 mm/min and 2. Tipping Bucket with resolution 0.1 mm/min. The majority of data stems from the device with the higher resolution. For the purpose of a more complex study in this region, the original time series format of the data was for this analysis transformed into the form of **wet spell amount (wsa) and wet spell duration (wsd)**. This type of input data origins in the application of so-called "alternating renewal models (ARM)" used for precipitation synthesis analysis (Haberlandt et al., 2008). In these models, the data are simplified into the form of two successive intervals: the wet- and the dry- spell of precipitation (Whitt, 2011). Furthermore, to each wet spell, a precipitation amount is attributed as a multiple of the wet spell duration by the events´ intensity. For the purposes of this work, all intensities < 0,01 mm/min were discarded and only rainfall events with total rainfall amount > 1 mm were included in the analysis. In addition, the initial high resolution was simplified and the duration is given in 5 minutes intervals. Two consecutive events were distinguished by minimal dry spell duration of 5 minutes. The advantage of ARM models is thus that they are straightforward as their external structure is characterized as a point information (Haberlandt et al., 2008). The construction of data e.g. the external structure of an ARM model is illustrated in the Fig. 2 - left. Its internal structure that describes the distribution of precipitation within a wet spell can be seen in the Fig. 2 -right.

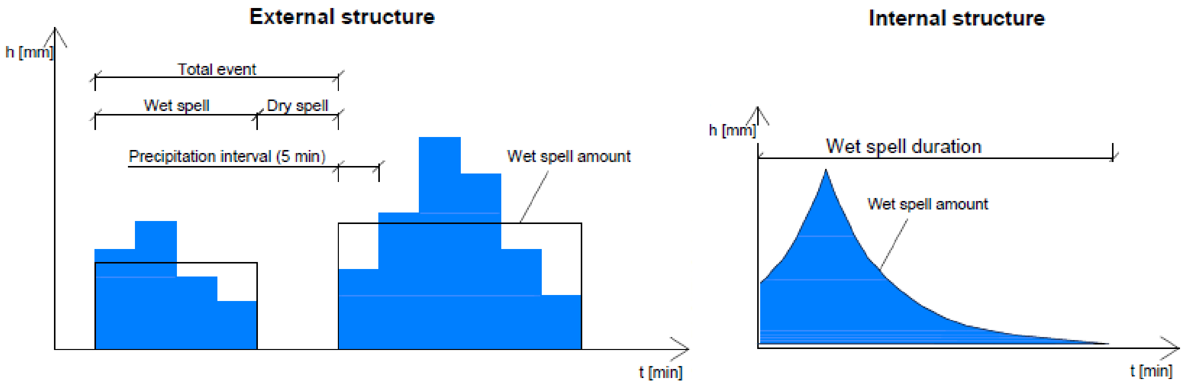

Fig. 2: left - rainfall histogram with 5 min time step and description of the external structure of an ARM model, right - internal structure of ARM model (distribution of a wet spell event), source: (Haberlandt, 1998)

A general description of data is also presented in Tab. 3, showing the maximum, minimum, average and median values for each studied dataset are presented.

Tab. 3: Summary statistics of data with peak values, minima´s, mean and median

Variable	max	min	mean	median
Wsa [mm]	102.16	1	3.76	2.27
Wsd [min]	1285	5	70.09	45

3. Methods

In this analysis, the procedure proposed by Hosking and Wallis (1997) was followed which is divided into these steps:

- Screening of data
- Identification of homogeneous regions
- Fitting distributions
- Quantile estimation

3.1 Screening of data

At first, the data had to be checked for: gross errors, homogeneity, consistency and stationarity. In addition, according to the assumptions of the index-flood method, the serial and spatial independency of sites is required. For the detection of: incorrect values, outliers, trends and shifts in the mean of a sample designed Hosking and Wallis a so-called discordancy measure. In order to check the serial correlation the correlograms were used. Finally, the spatial correlation was not checked by any test but it was expected given the fact that the studied phenomenon was rainfall and the density of stations was high.

3.1.1 Discordancy measure

Hosking (2015) defined the discordancy measure D_i as an indicator, for site i, as the discordancy between the site's L-moment ratios and the (unweighted) regional average L-moment ratios. Mathematically Hosking & Wallis (1997) defined the discordancy for RFA as follows:

A vector $$u_i = [t_2^{(i)} \quad t_3^{(i)} \quad t_4^{(i)}]^T \tag{3.1}$$

is supposed, where t, t_3 and t_4 are L-moment ratios (see chapter 2 in Hosking and Wallis (1997) and T indicates a transposition of the vector. Let $\bar{u}$ be the average of u_i defined as:

$$\bar{u} = N^{-1}.\sum_{i=1}^{N} u_i \tag{3.2}$$

Then, the matrix of sums of squares and cross-products A is:

$$A = \sum_{i=1}^{N} (u_i - \bar{u})\,(u_i - \bar{u})^T \tag{3.3}$$

Finally the discordancy measure can be derived:

$$D_i = \frac{1}{3} N\,(u_i - \bar{u})^T A^{-1} (u_i - \bar{u}) \tag{3.4}$$

The final value of D_i is considerably influenced also by the number of sites contained in the studied region. Hosking & Wallis (1997) stated that D_i has only an informational character if the number of sites is lower than seven. They defined the critical values presented in the Tab. 4. Even though they added that the critical value rises with the number of sites, they recommend to use as critical value Di = 3 for groups containing N > 15. Finally, if a site is flagged discordant, a closer investigation of the L-moments ratios, their scatter plots and trend is recommended.

Tab. 4: Critical values for discordancy by different number of units in a group.

Number of sites in region (N)	5	6	7	8	9	10	11	12	13	14	≥ 15
Critical value (D_i)	1,333	1,648	1,917	2,14	2,329	2,491	2,632	2,757	2,869	2,971	3

The discordancy measure is used in RFA twice. First, it to the whole sample at once in order to reveal the sites that need detailed investigation and proof of reliability. Next, it is used to assess the correctness of classification of sites to regions.

3.2 Identification of homogeneous regions

The core part of RFA is the formation of homogeneous regions. It is based on grouping sites to regions in which the site samples can be considered as coming from the same population and their frequency distributions are identical apart from a scaling factor (Hosking & Wallis, 1997). This task is a multi-objective optimization problem and can be accomplished via different approaches. It can be performed subjectively especially if the solver has appropriate experience and knowledge about the characteristics and the hydrological regime of studied area for ex.: Solín (2002) or Gáal et al. (2009) or one can choose from a wide range of objective methods. Nevertheless, the best results are achieved when the two approaches are coupled. For this analysis the most classical approach for pooling regions was chosen - the cluster analysis. More specifically it was two combinations of different clustering algorithms:

1. Ward´s hierarchical clustering method coupled with K-means clustering
2. Partitioning around medoids (PAM)

Concerning the input data for pooling regions, Hosking & Wallis (1997) recommended to use so-called "site characteristic". These include all available data describing the studied area other than the actual measurements. For the purposes of environmental engineering these are the most frequent: site´s location, elevation and other physical features of the site associated with the studied variable. In this work, it was besides the location characteristics also the distance to sea, mean annual precipitation, correlation between the wet-spell amount and wet-spell duration, proportion of time steps with rainfall related to total time series and total seasonal rainfall.

The formation of clusters is based on comparing the dissimilarities or distances between sites. For this problem, the "Euclidean distance" was used. The general formula can be written as follows:

$$D_{(x,\,y)} = \sqrt{\Sigma(x_i - y_i)^2} \qquad (3.5)$$

However, its disadvantage is that the resulting distances can be noticeably affected by differences in scale of descriptors. Therefore, the data must be transformed to a comparable scale.

3.2.1 Heterogeneity measure

The heterogeneity measure is a tool used for assessment of initially defined clusters. It compares the variations of L-moments of the sites inside a cluster with values that would be expected for a homogeneous cluster. For this purpose, a simulation of a homogeneous series of samples with the same record length as the observed sites are generated and the mean with the standard deviation of simulation are used to compute its dispersion and to compare it with the dispersion of observation as in formula (3.6).

$$H = \frac{(\text{observed dispersion}) - (\text{mean of simulations})}{(\text{standard deviation of simulations})} \qquad (3.6)$$

The dispersion of observations is a weighted standard deviation of sites´ coefficient of L-variation, L-CV. The regional average L-CVs are first weighted proportionally to the sites´ record lengths as in (3.7) and next, the standard deviation V of at-site sample L-CVs is computed according to formula (3.8):

$$t_R = \sum_{i=1}^{N}(n_i t_i) \,/\, \sum_{i=1}^{N}(n_i) \quad (3.7)$$

where t_R is the regional average L-CV, t_i is the sites´ L-CV and n_i is the sites´ record length.

$$V = \sqrt[2]{\sum n_i (t_i - t_R)^2 / \sum n_i} \quad (3.8)$$

The same statistics are calculated from each simulation and subsequently the mean and standard deviation of simulated statistics are calculated: μ_V, σ_V. Finally the formula for heterogeneity measure can be defined as in (3.9):

$$H = (V - \mu_V) / \sigma_V \quad (3.9)$$

Hosking and Wallis further performed a series of Monte Carlo simulation experiments and defined limit values of heterogeneity as:

- if $H < 1$ the region can be declared acceptably homogeneous,
- if $1 \leq H < 2$ the region behaves as possibly heterogeneous
- if $H > 2$ the region is definitely heterogeneous

3.3 Identification of regional distributions

The main aim of fitting distributions to data is being able to reproduce data with the same characteristics and extract unobserved extreme events (Haan, 1977). In general, the most frequently used distribution for describing the precipitation data is generalized extreme value (GEV) distribution and the Gumbel distribution. However it is always recommended to fit several distributions and compare. For this purpose, Hosking & Wallis (1993) developed "Z –statistics" as a goodness of fit test that is based on L-moments.

In fact, the distributions are fitted based on equality of L-skewness and the goodness of fit is judged by comparison of L-kurtosis of the distribution τ_4^{DIST} and the regional average L- kurtosis t_4^R. In addition, to measure the goodness of fit, the difference between the observed and the simulated L-kurtosis is calculated and divided by the standard deviation σ_4 as in the formula:

$$Z^{DIST} = \frac{(t_4^R - \tau_4^{DIST})}{\sigma_4} \quad (3.10)$$

However, the simulation may, according to Hosking & Wallis (1997), cause a bias B_4 of L-kurtosis (eq 3.11). Then

$$B_4 = \frac{\sum_{m=1}^{Nsim}(t_4^m - t_4^R)}{N_{sim}} \quad (3.11)$$

$$Z^{DIST} = \frac{(\tau_4^{DIST} - t_4^R + B_4)}{\sigma_4} \quad (3.12)$$

4. Results

4.1 Screening of data

First step in the data screening was the test of serial correlation. The autocorrelation displays the inner behavior of a variable, thus the dependence of one event on a previous one. The test was evaluated visually from correlograms at lags 1-5. Chosen examples are presented in the Fig. 3.

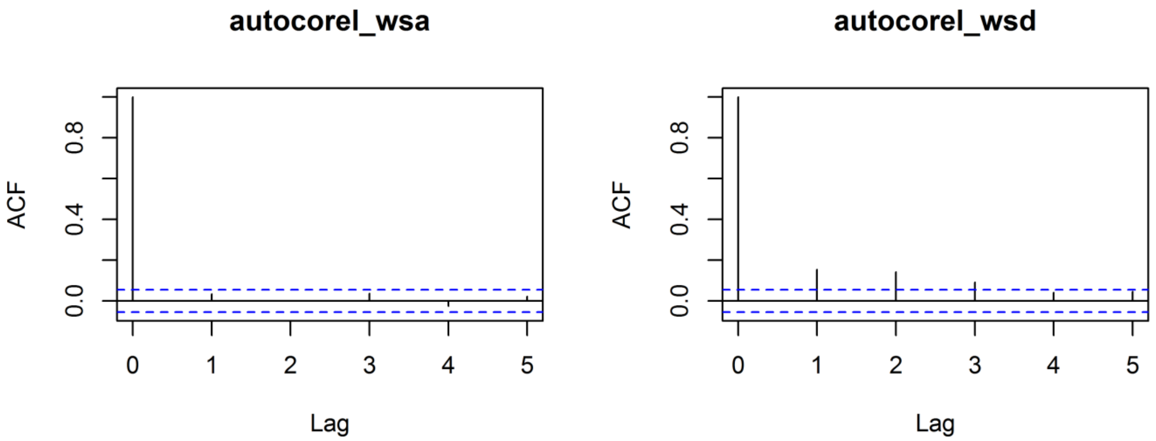

Fig. 3: Correlogram of data at station N.9: left – wsa, right - wsd

The test showed that the wsa variable data are not serially correlated, exceeding threshold only at 4 stations. On the other hand, the duration variable, wsd cannot be considered independent as there, the threshold was exceeded in most of the stations. However, a small amount of serial dependence has according to Hosking & Wallis (1997) only little influence on the results of RFA. The possible intersite correlation was also taken into account.

Next, the discordancy test which draws attention on stations that have samples considerably different from the rest was executed. The threshold for discordancy measure, when the sample is greater than 15 is D=3. Tab. 5 lists the sites for which D > 3 along with the results from their additional investigation.

Tab. Chyba! Dokument neobsahuje žiadny text so zadaným štýlom.**: Discordant sites with additional investigation statistics: D= discordancy, p = significance level of Mann-Kendall trend test, ρ= Spearmans correlation coefficient.**

Variable			trend		outliers		rank according to sample size
	site Nb.	D	p	ρ	sample size	percentage of outliers	
wsa	11	3.12	0.5606	-0.0187	434	0.46	69
	24	3.33	0.7394	-0.0094	567	2.47	9
	52	3.53	0.4788	0.0220	465	1.29	53
wsd	18	3.26	0.2248	-0.0377	466	1.29	51

The sites flagged as discordant were further analyzed for containing trend, outliers and shift in median. The results of the Mann Kendall test showed no significant trend for all analyzed data. The outliers were to be considered mainly at st. 24. The original dataset was controlled but since no error was discovered, all the values were retained in the sample as a characterization of the natural process. Generally, all suspicious data were investigated but no error was discovered. Therefore, all the data were included for the analysis.

4.2 Identification of homogeneous regions

After the data were transformed to have approximately the same probability distribution and rescaled to the range of [0;10], the two chosen clustering algorithms were applied and compared. The decision process between the two methods was supported by the heterogeneity measure. It was found that for wsa, more suitable method is the combination of Ward´s and k-means clustering (WAKM) while for wsd the partitioning around medoids method (PAM). However, after the automatic clustering no option with all regions being acceptably homogeneous was found. Therefore, the clusters had to be manually readjusted. The focus was oriented towards the discordancy of sites inside a region. Such sites were moved to another region that seemed to be more appropriate regarding the site characteristics. If such a site was discordant also in the altered region and it caused an augmented heterogeneity, it should be examined and eventually deleted from the sample.

The results of heterogeneity before and after the manual readjustment are summarized in the Tab. 6. The number of regions is marked with the same color as it can be seen in the graphical representation of created regions in the Fig. 4.

Tab. 6: Results of homogeneity testing of each region for variables wsa and wsd B- before the manual readjustment and A- after it.

var	dataset	Number of region					Absolute average H
		1	2	3	4	5	
wsa	B	-1.561	-0.515	-0.839	-1.548	-0.069	-0.907
	A	-1.140	-0.541	-0.811	-1.106	-0.670	**-0.854**
wsd	B	-0.655	-1.049	-0.620	-0.588	-1.825	-0.947
	A	-0.634	-1.043	-1.030	-0.984	-0.877	**-0.913**

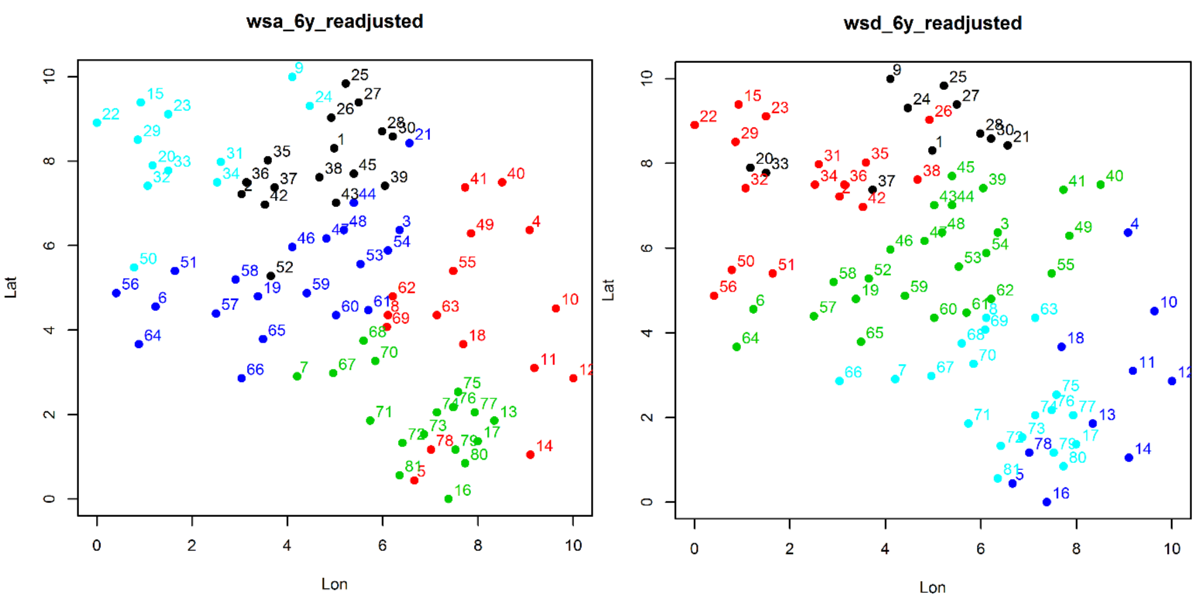

Fig. 4: The final classification of stations into homogeneous regions: left – wsa, right – wsd

4.3 Fitting the distributions

For the fitting of distributions, the goodness of fit test Z^{DIST} was computed. The test was fitting to the data the following set of distributions: generalized logistic (GLO), generalized extreme value (GEV), generalized normal (lognormal) (GNO), Pearson type III (3-parameter gamma) (PE3), generalized Pareto (GPA). A distribution is considered having an acceptable fit when $Z^{DIST} \leq 1,645$. In the Tab. 7 are summarized the results for the best fitting option for each variable. Moreover, the decision of distribution choice was supported by ratio diagrams showed in the Fig. 5 which display: 1. in different colors and legend accordingly, the fixed curves describing course of L-moment ratios of the tested 3-parameter distributions, 2. the black quadruples with labels E and G indicate the L-moment ratios for the two parameters exponential and Gumbel distributions that are often used for hydrological purposes and 3. the crosses indicate the location of L-moment ratios for each region.

Tab. 7: Results of goodness of fit test for the best fitting distribution

variable (distribution)	Z-statistics				
	R1	R2	R3	R4	R5
wsd (GNO)	0.20	0.26	1.69	1.49	1.33
wsa (GPA)	1.84	3.10	3.91	4.61	5.30

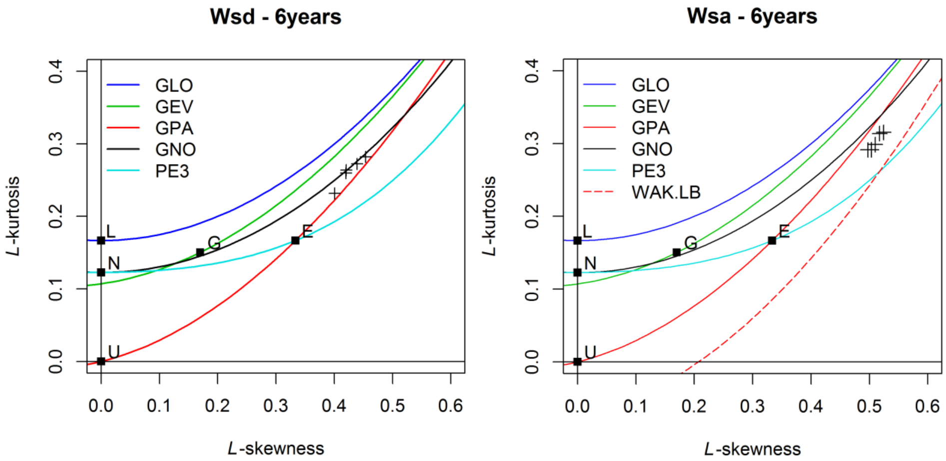

Fig. Chyba! Dokument neobsahuje žiadny text so zadaným štýlom.**: L-moment ratio diagrams: left – wsd, right – wsa.**

It can be seen that for wsd, the generalized normal distribution GNO was found with an acceptable fit for all regions except one. However, according to the literature, in the cases where the majority of the regions follow one distribution, this one should be fitted to the whole studied area. Therefore, for further purposes and analysis of wsd, the generalized normal distribution will be used.

On the other hand, the goodness of fit test for wsa resulted in no acceptable fit for any of the fitted distributions. This case can according to the literature occur in these situations:

1. If the regions are not sufficiently homogeneous
2. If the region is constituted of a large number of stations.
3. If the data are correlated: The correlation in data increases the variability of t_4^R (Hosking & Wallis, 1997) and since, in the goodness of fit test, it is compared with an uncorrelated simulation, the resulting Z^{DIST} falsely indicates that none of the distributions fits the data. This reasoning is considered as the true one in this case. In order to obtain the real Z^{DIST} and more accurate quantile estimates, the

intersite correlation should be specified by a correlation matrix and included into the simulations. However, for the data in form of "wet spell amount" of a rain event, the identification of a correlation coefficient is not possible. Therefore, it was decided to proceed with the wsa as with heterogeneous regions and use a more parametric distribution to fit the data.

Finally, the five parameter **Wakeby** distribution was chosen for amount variable. It should be noticed that the lowest "Z" were achieved for generalized Pareto distribution. This one was applied also in the work of Roth et al. (2014) who analyzed the extreme precipitation in the Netherlands and north-western Germany.

Nevertheless, even though the type of the distribution is common for the whole area, each individual regions´ distribution has different parameters and these are summarized in the Tab. 8. Regarding wsa, it can be seen that three regions have the parameters alpha and beta equal to 0. The Wakeby distribution in this form corresponds according to the literature to the generalized Pareto distribution mentioned above. Moreover, the fitting of distributions enables the visual comparison of their course in the form of probability density functions, see Fig. 6.

Tab. 8: Parameters of distribution for each region

region Nb.	wsd - GNO			wsa - Wakeby				
	xi	alpha	kappa	xi	alpha	beta	gamma	delta
R1	42.677	38.796	-0.983	0.992	-4.316	0.133	5.638	0.129
R2	41.928	37.063	-0.944	0.899	0.000	0.000	1.906	0.354
R3	47.743	42.080	-0.854	0.897	0.000	0.000	1.948	0.377
R4	52.245	46.343	-0.901	0.922	0.000	0.000	1.830	0.330
R5	50.015	43.788	-0.901	1.003	-2.623	0.472	3.769	0.133

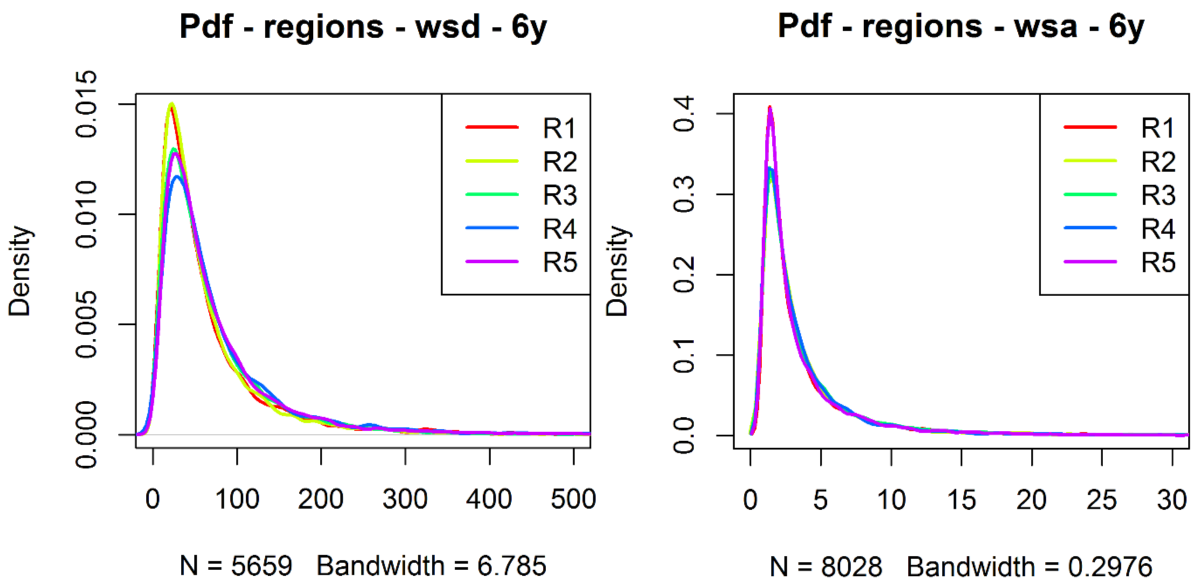

Fig. 6: Probability density function of created regions, the values on x- axis are cut for a better visibility

5. Conclusions

In this work the regional frequency analysis pursuant Hosking & Wallis (1997) was performed based on L-moments. It was applied on rainfall data from 81 stations in north-western Germany. The analysis involved: 1. Screening of the data, 2. Repartitioning of sites into homogeneous regions and 3. Determination of suitable distributions.

The screening of data detected the spatial correlation for both variables and serial correlation only for wsd. In all data, no trend was detected and the discordancy measure indicated: 3 sites in wsa and 1 site in wsd as suspicious. The samples were consequently controlled but no error was found. Therefore, all the available data was included in the analysis.

In the section of repartitioning of sites into the regions, first two clustering methods were compared: 1. the combination of Ward´s and K-means clustering and 2. the partitioning around medoids. It was found that the WAKM method yielded more suitable regions for wsa while for wsd the PAM should be used. For both variables five regions were formed. As these were not sufficiently homogeneous after the objective clustering, the regions required a further manual readjustment.

In the distribution fitting part it was found that none from the tested three-parameter distributions fits wsa data because of the intersite correlation. Since it was not possible to quantify the correlation, it was decided to fit the data a more complex – five parameter Wakeby distribution. The wsd was uniquely determined to follow the generalized normal distribution.

Generally it can be stated that the analysis of wsa yielded better results in all parts of the analysis except for fitting the distribution. The regionalization and readjustment of regions required considerably more time in the case of wsd. I assume that this result can be attributed to its revealed autocorrelation. The herein defined probability distribution functions are ready for next step of the analysis which is the quantile estimation. However, because of the space limitation this was set out of the scope of this report.

References:

Al Mamoon, A., Joergensen, N., Rahman, A., & Qasem, H. (2014, 3). Derivation of new design rainfall in Qatar using L-moment based index frequency approach. *International Journal of Sustainable built environment* , pp. 111-118.

Čunderlík, J. (2000). Flood flow regionalization based on L-moments and its use with the index flood method. In J. M. (eds.), *Flood Issues in Contemporary Water Management* (S. 49 - 59). Kluwer Academic Publishers

Dalrymple, T. (1960). *Flood frequency analyses. Manual of Hydrology: Part 3.* Washington: UNITED STATES GOVERNMENT PRINTING OFFICE.

Elsholz, M., & Berger, H. (1989). *Hydrologische Landschaften im Raum Niedersachsen. Oberirdische Gewässer 6/98.* Hildesheim: NLÖ.

Fowler, H. a., & Kilsby, C. (2003). A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. *Int. J. Climatol 23* , S. 1313–1334.

Furst, J., Bichler, A., & Konecny, F. (2011). Estimation of extreme groundwater tables. *Mitteilungsblatt des hydrographischen Dienstes in Oesterreich Nr. 87 (in german)* , S. 1-44.

Gaál, L., Kyselý, J., & Szolgay, J. (2008). Region-of-influence approach to frequency analysis of heavy precipitation in Slovakia. *Hydrol. Earth Syst. Sci. 12* , pp. 825–839.

Gaál, L., Szolgay, J., Lapin, M., & Fasko, P. (2009). Hybrid approach to delineation of homogeneous regions for regional precipitation frequency analysis. *J. Hydrol. Hydromech.* (57), pp. 226-249.

Haan, C. T. (1977). *Statistical methods in hydrology.* Ames: The Iowa State University press.

Haberlandt, U. (1998, July). Stochastic rainfall synthesis using regionalized model parameters. *Journal of Hydrologic Engineering* (3), pp. 160-168.

Haberlandt, U., Ebner von Eschenbach, A.-D., & Buchwald, I. (2008). A space hybrid hourly rainfall model for derived flood frequency analysis. *Hydrol. Earth Syst. Sci.* (12), pp. 1353 - 1367.

Hosking, J. R., & Wallis, J. R. (1993). Some statistics useful in regional frequency analysis. *Water Resour. Res. , 2*, S. 271–281.

Hosking, J. &., & Wallis, J. (1997). *Regional frequency analysis—an approach.* Cambridge: Cambridge University Press.

Kohnová, S., & Szolgay, J. (2000). Regional estimation of design flood discharges for river restoration for mountainous basins of northern Slovakia. In *Flood Issues in Contemporary Water Management* (S. 41 - 47).

Kohnová, S., Gaál, L., Parajka, J., Szolgay, J., & Hlavčová, K. (2010). Mapping precipitation quantiles by means of pooling approaches at ungauged sites in Slovakia

Modarres, R. (2010, August). Regional Dry Spells Frequency Analysis by L-Moment and Multivariate Analysis. *Water Resources Management* (24), pp. 2365-2380.

Mossaffaie, J. (2015). Comparison of Two Methods of Regional Flood Frequency Analysis by Using L-moments. *Water Resources, No. 3* (42), S. 313 - 321.

NLWKN, H.-H. (n.d.). *Hydrologische Landschaften im Raum Niedersachsen | Nds. Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz*. Retrieved April 11, 2016, from Web: Niedersachsen.de: http://www.nlwkn.niedersachsen.de/wasserwirtschaft/fluesse_baeche_seen/hydrologische_landschaften/niederschlag/hydrologische-landschaften-im-raum-niedersachsen-38694.html

Portela, M., Zelenáková, M., Santos, J., Purcz, P., Silva, A., & Hlavatá, H. (2015). Drought analysis in Slovakia: regionalization, frequency analysis and precipitation thresholds. *River Basin Management , VIII* (197), pp. 237 - 248.

Remiášová, R., Gaál, L., Hlavčová, K., Kohnová, S., & Szolgay, J. (2011). *Priestorová regionalizácia návrhových zrážok na Slovensku.* Bratislava: KEY Publishing.

Roth, M., Buishand, T., Jongbloed, G., Klein Tank, A., & Van Zanten, J. (2014, January 23). Projections of precipitation extremes based on a regional, non-stationary peaks-over-threshold approach: A case study for the Netherlands and north-western Germany. *Weather and Climate Extremes 4* , pp. 1 – 10

Solín, Ľ. (2005, April). Floods – size estimates for small basins of Slovakia by the method of regional frequency analysis (in slovak). *Geografický časopis 57* , pp. 287 - 307.

Web: Klima und Wetter in Niedersachsen. (n.d.). Retrieved April 10, 2016, from wetter.de: http://www.wetter.de/deutschland/wetter-karte-niedersachsen-c49p3.html

Web: Norddeutscher Klimamonitor. (n.d.). Retrieved April 11, 2016, from http://www.norddeutscher-klimamonitor.de/klimaentwicklung/1951-2010/winter/niederschlag/niedersachsen-bremen/coastdat-2/zeitreihe.html

Weiss, J., Bernerdara, P., & Benoit, M. (2012). Assessment of the regional frequency analysis to the estimation of extreme storm surges. *33rd International Conference on Coastal Engineering (ICCE 2012).* Santander: American Society of Civil Engineers (ASCE).

Whitt, P. (2011, April). Alternating Renewal Processes and The Renewal Equation [lecture]. *IEOR 4106* .

Yang, T., Shao, Q., Hao, Z.-C., Chen, X., Zhang, Z., Xu, C.-Y., et al. (2010). Regional frequency analysis and spatio-temporal pattern characterization of rainfall extremes in the Pearl River Basin, China. *Journal of Hydrology 380* , pp. 386 - 405.