APPLICATION OF EXTREMAL THEORY TO THE PRECIPITATION SERIES IN NORTHERN MORAVIA

DANIELA JARUŠKOVÁ

Department of Mathematics, Czech Technical University, Prague; jarus@mat.fsv.cvut.cz

1. Introduction

The Moravian floods in July 1997 killed 49 people, affected 536 settlements, 29 000 houses and cost 2.5 billion USD. The water discharges of Moravian rivers reached almost hundred times their average. According to public opinion the flood was caused by unusual weather conditions when severe storms occurred almost simultaneously in the basin drained by the rivers Opava and Opavice.

The main flood risks in the Czech Republic are usually classified as:

1. Snow melt period of December to April in combination with precipitation (typical of lowland and hilly parts of the Labe, Vltava and Morava catchments).
2. Longer duration of rainfall lasting 10 to 72 hours (rainfall intensified and prolonged by elevation enhancement).
3. Intensive short duration of summer rainfall exceeding (flood peak occurring within hours of rainfall peak, small areas affected, flash flood are characteristic of small short streams after April – September thunderstorm.)

2. Data

We have obtained a data set from the Czech Hydro-Meteorological Institute (CHMI) detailing daily precipitation values for 10 meteorological stations in the Northern Moravia – Heřmanovice (HE), Karlovice (KA), Karlova Studánka (KS), Krnov (KR), Lichnov (LI), Opava (OP), Praděd (PR), Rejvíz (RE), Vidly (VI), Albrechtice – Žáry (ZY). We have excluded station (KA) due to the short duration of measurements and worked with 9 stations only. Station (KR) has missing records for the period of the 1997 flood and is also often irrelevant for our study. The longest record spans 45 years (1/1/1960 – 6/2/2005), but most of the records are shorter with missing data.

By studying the relationship between rainfalls and water discharges we would like to avoid problems with snowfalls and melting and that is why we work only with "summer days" records of the months May – October. However, it is true that results of a statistical inference for the daily records of all months and of "summer months" are very similar. Table 1 presents several basic descriptive characteristics.

station	number of all obs.	number of pos. obs.	mean	90 % quantile	max
HE	7697	3473	7.3	18.2	196.5
KA	8250	3258	6.4	16.2	124.2
KR	8187	3265	5.7	14.9	59.2
LI	8280	3332	5.6	15.0	110.0
OP	8096	3385	5.3	14.4	62.0
PR	6931	3563	7.6	18.9	139.4
RE	7452	3297	8.5	21.6	214.2
VI	7605	3869	7.2	18.6	199.3
ZY	8280	3899	5.6	15.2	125.0

Table 1. Basic extremal descriptive characteristics (mean and 90% quantile are calculated from the positive precipitation values).

3. One – variate POT method

Authors of the books on the extreme value theory, see Beirlant (2004), Embrechts (1997) recommend to use a mean excess function to find a right threshold value. This seems to us that it is not an easy task for an unexperienced researcher and it is probably the reason why the block maxima method is more popular within the meteorological community. We have observed that estimates of a generalized Pareto distribution were more stable to a change of threshold for the "longer – tailed" data than for the "short – tailed data". Finally, for all studied series we decided to choose a threshold near the 90 % quantile calculated from the respective positive values so that the number of observations that exceeded it was between 320 – 420.

Statistical inference showed that the observation sites form three clusters. First, the sites in mountains (HE, PR, RE, VI) are wet, their averages are larger and they are long – tailed. Second, the sites in the valleys (KR, LI, OP, ZY) are short – tailed and their averages are smaller. The third cluster consists of the site (KA) only, which is situated on the half way between the mountains and the sites in the valleys. The estimated parameters of a generalized Pareto distribution are given in Table 2.

station	treshold	β	ξ	2 %	1 %	0.1 %	max	p
HE	18.4	11.72	0.25	179.9	217.7	405.8	196.2	1.8%
KA	16.0	8.85	0.13	94.5	108.6	165.2	124.2	0.8%
KR	9.0	8.67	0.00	60.4	64.0		59.2	
LI	15.0	8.00	0.07	73.8	82.6	114.9	110.0	0.2%
OP	15.0	9.18	0.00	69.0	75.4	96.6	62.0	5.9%
PR	18.4	9.77	0.24	157.5	190.2	346.6	139.4	3.0%
RE	18.4	12.24	0.19	165.4	195.5	329.1	214.2	0.8%
VI	18.4	10.41	0.21	150.8	179.8	313.3	199.3	0.6%
ZY	15.0	9.48	0.05	82.3	91.4	124.0	125.0	0.1%

Table 2. Threshold values and the parameters of a GPD fitted to the values and the 2 %, 1 % and 0.1 % upper quantiles (in years) of a generalized Pareto model and the maximal values (column 5) with respective exceedence probabilities.

Q – Q plots for the data and the chosen model are presented by Figures 1 – 3. We see that the model sometimes overestimated and sometimes underestimated the observed extremal values.

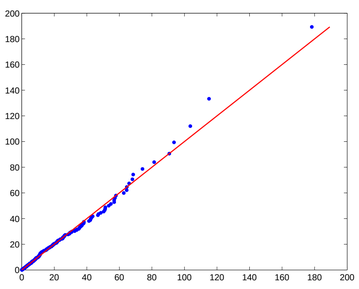


Figure 1. Q-Q plot for HE.

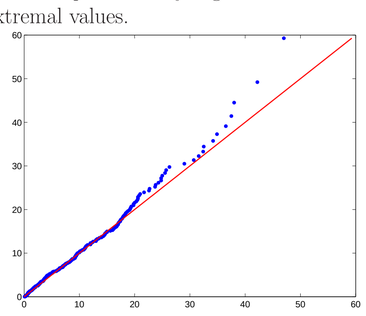


Figure 2. Q-Q plot for RE.

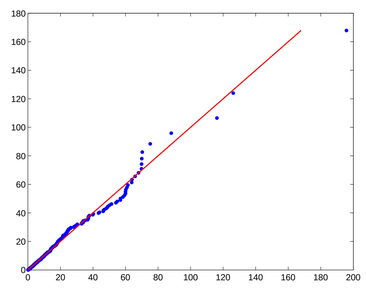


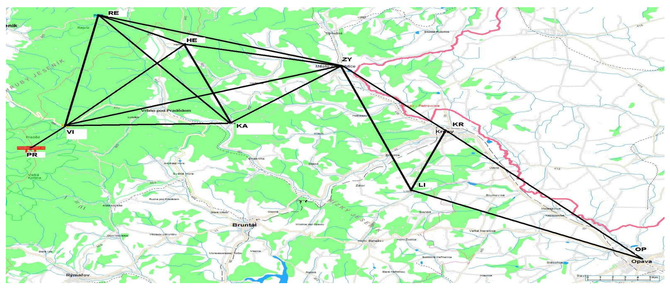
Figure 3. Q-Q plot for OP.

Table 2 presents the 2 %, 1 % and 0.1 % upper quantiles (in years) of estimated model together with the exceedence probabilities for the observed maximal values. Five exceedence probabilities are smaller than 1%. The site (OP) is situated in a larger distance from the rest of sites and the event of 1997 was not so severe there. The sites (HE) and (PR) belong to the "wet stations" so that their maximal values corresponding to the year 1997 were not so much extremal as at other sites.

4. Bivariate analysis

Our series were obtained from the sites that are relatively very close to each other and the dependence between daily precipitation values is very strong. The correlation coefficients for all observation sites calculated for the days with non – zero rainfall values are presented in the following matrix:

	KA	KR	LI	OP	PR	RE	VI	ZY
HE	0.70	0.56	0.57	0.56	0.68	0.83	0.75	0.79
KA		0.65	0.65	0.59	0.68	0.70	0.73	0.78
KR			0.74	0.70	0.55	0.57	0.57	0.76
LI				0.72	0.50	0.56	0.55	0.70
OP					0.50	0.57	0.55	0.67
PR						0.69	0.81	0.68
RE							0.73	0.76
VI								0.72



It is interesting to notice that the magnitudes of the correlation coefficients correspond very accurately to the distances between the sites. If we take a schematic map of the region and connect the sites with the correlation coefficients larger than 0.7 we get an interesting graph.

A natural summary of extremal dependence is the coefficient $\chi = \lim_{u\to 1} P(V > u|U > u)$ with (U, V) representing transformed versions of the studied variables with uniform margins, see Coles (1999). The statistical inference showed that the value of χ is positive for all pairs of sites. Even more informative is the function

$$\chi(u) = 2 - \frac{\log P(U < u, V < u)}{\log P(U < u)} \quad \text{for} \quad 0 < u < 1.$$

Figure 4 shows the function $\chi(u)$ for the pair HE – RE (close sites with correlation coefficient 0.83) while Figure 5 for HE – OP (more distant sites with correlation coefficient 0.56). It is striking that the functions $\{\chi(u)\}$ are always almost constant in the range $0.1 < u < 0.8$ but they start to decrease for $u > 0.9$ (sometimes even earlier). This is true for all pairs of gauges and the decrease is larger when the sites are more distant.

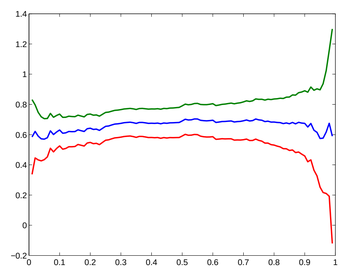


Figure 4. Function $\chi(u)$ for HE – RE.

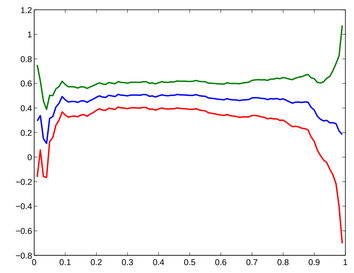


Figure 5. Function $\chi(u)$ for HE – OP.

Further, we transformed the studied variables to variables $\{Z_i\}$ with standard Fréchet margins and tried to model their dependence structure. As our data were relatively in agreement with a symmetric model we applied a logistic model. To find a threshold for a bivariate case was even more difficult than for one – variate. Figures 6 – 8 show the histograms of $\omega_i = Z_{1i}/r_i$ where $r_i = Z_{1i} + Z_{2i}$ for $\{\omega_i : r_i > r\}$ where $r = \exp(2)$, $r = \exp(3)$ and $r = \exp(4)$ for the sites HE – RE while Figures 9 – 11 the corresponding histogram for the sites HE – OP.

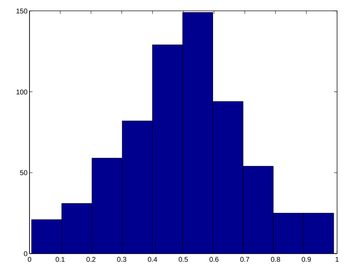


Figure 6. Histogram of ω_i where $r_i > \exp(2)$ for HE – RE.

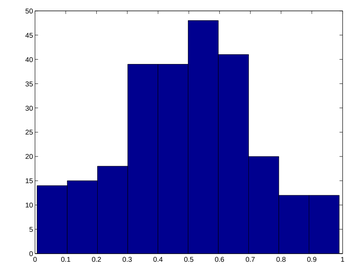


Figure 7. Histogram of ω_i where $r_i > \exp(3)$ for HE – RE.

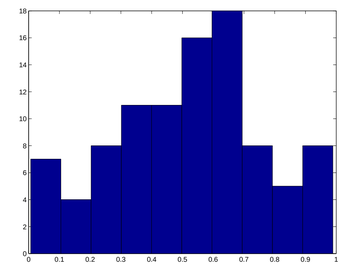


Figure 8. Histogram of ω_i where $r_i > \exp(4)$ for HE – RE.

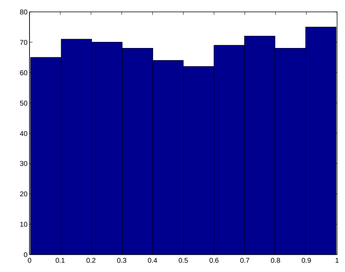


Figure 9. Histogram of ω_i where $r_i > \exp(2)$ for HE – OP.

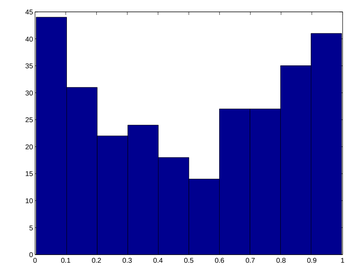


Figure 10 Histogram of ω_i where $r_i > \exp(3)$ for HE – OP.

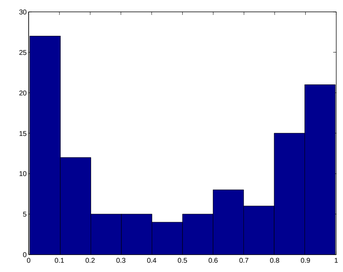


Figure 11. Histogram of ω_i where $r_i > \exp(4)$ for HE – OP.

Finally, we decided to take $r = \exp(3.5)$ for all pairs. For the sites HE – RE the estimate $\widehat{\alpha}_{4,11} = 0.44$ and for HE – OP $\widehat{\alpha}_{4,9} = 0.66$. Figure 12 shows the density contours for HE – RE, Figure 13 shows the same contours for HE – OP.

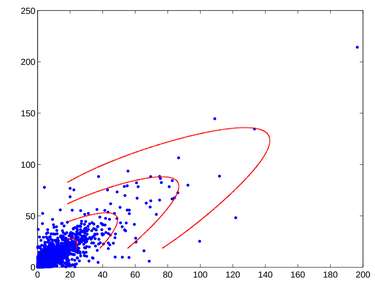


Figure 12. Scatter plot and density contours for HE – RE.

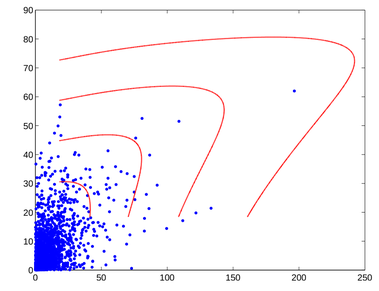


Figure 13. Scatter plot and density contours for HE – OP.

Several bivariate quantiles (in years) for HE – RE are presented in Figure 14 while some bivariate quantiles for HE – OP are presented in Figure 15. We also calculated the exceedence probability for the extremal value of HE – RE (196.5, 139.4) being 0.5 % and of HE – OP (196.5,62.0) being slightly less than 1 %. It is interesting to compare with the exceedence probabilities from Table 2.

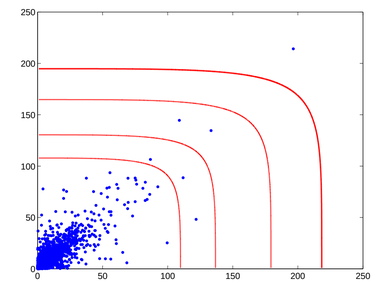


Figure 14. 10%, 5%, 2%, 1% bivariate quantiles (in years) for HE – RE.

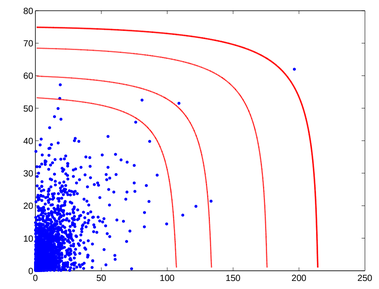


Figure 15. 10%, 5%, 2%, 1% bivariate quantiles (in years) for HE – OP.

5. Conclusion

Application of the theory of extremes to rainfalls is traditional, see Coles (2003) or Engeland (2004). Our goal is to find statistical model for relationship between extremal precipitation and water discharges. We have to admit that we are still at the very beginning. Until now we used POT method to find a model for univariate and bivariate tails of precipitation series measured by nine gauges in a relatively small area. However, we did not find any good model for the set of all nine precipitation series. We tried to use a Dirichlet distribution to describe the dependence structure but the model did not seem to be satisfactory, see Coles (1991) or Coles (1994).

Even in bivariate case we are not sure whether the asymptotic theory may be applied. It seems to us that the behavior of the very extremal data and of the data in a subextremal region slightly differs in a way that the extremal values are more independent that the subextremal values. It is more evident for series measured by more distant gauges that by the closer ones. One possible explanation is an appearance of local thunderstorms that bring a heavy rain to a small area. On the other hand we think that the error we do if we apply the asymptotic theory is not large and leads to a slight overestimation of the exceedence probabilities so that we are on the safety side.

Studying relationship between precipitations and water discharges we have found out that the extremal water discharges are more connected to the sum of three or more preceeding daily precipitation values than to only single daily values. Here, finding a reasonable model is even more complicated because of strong dependence and a good declustering technique is necessary. This all means that our goal is very difficult to achieve and that we have still lot of work ahead.

References.

[1] Beirlant J., Goegebeur Y., Segers J.,Teugels J. *Statistics of Extremes.* John Wiley and Sons, Chichester, 2004.

[2] Coles S.G. and Tawn J.A. *Modelling extreme multivariate events.* J.R. Statist. Soc. B **53**, 377 – 392, 1991.

[3] Coles S.G. and Tawn J.A. *Statistical methods for multivariate extremes: An application to structural design.* Appl. Statist., **43**, No.1, 1 – 48, 1994.

[4] Coles S.G. and Tawn J.A. *Modelling extremes of the areal rainfall process.* J.R. Statist. Soc. B **58**, 329 – 347, 1996.

[5] Coles S.G., Heffernan J. and Tawn J.A. *Dependence measures for extreme value analyses.* Extremes, **2:4**, 339 – 365, 1999.

[6] Coles S.G., Pericchi L.R. and Sisson S. *A fully probabilistic approach to extreme rainfall modelling.*Jour.of Hydr., **273**, 35 – 50, 2003.

[7] Embrechts P., Küppelberg C, Mikosch T. *Modelling extremal events.* Springer-Verlag, Heidelberg, 1997.

[8] Engeland K., Hisdal H., Frigessi A. *Practical extreme value modelling of hydrological floods and droughts: a case study.* Extremes, **7**, 5 – 30, 2004.

[9] Joe H., Smith R.L. and Weissman I. *Bivariate threshold methods for extremes.* J.R.Statist. B, **54**, No.1, 171 – 183, 1992.

Acknowledgement. The work was partially supported by the grant GAČR 201/03/0945.