Hydrol. Earth Syst. Sci. Discuss., 4, S1530–S1542, 2007
www.hydrol-earth-syst-sci-discuss.net/4/S1530/2007/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Region-of-influence approach to a frequency analysis of heavy precipitation in Slovakia" *by* L. Gaál et al.

L. Gaál et al.

Received and published: 28 November 2007

We would like to thank both reviewers (Attilio Castellarin and Martin Pfaundler) for their helpful, insightful and constructive comments on the submitted manuscript. We acknowledge the suggestions of both referees for improving the paper, which will be incorporated in the upcoming revised manuscript. A detailed list of replies to the comments of the referees is given in the following paragraphs.

Reply to the comments of A. Castellarin

General comments:

A. Castellarin: "My major comments concern: 1) previous literature that is particularly relevant to the study and that has not been considered" (Reed et al., 1999; Schaefer,

1990; Alila, 1999; Faulkner, 1999; Di Baldassare et al., 2006). "[...] The authors should consider all these studies in the manuscript and assess the degree of novelty of their analysis with respect to this relevant literature."

Reply #1: The manuscript aims at implementing the region-of-influence (ROI) approach into the frequency analysis of precipitation extremes; the ROI method has thus far been related predominantly to the frequency analysis of floods and low flows (e.g. Burn, 1990a, b, 1997; Zrinji and Burn, 1994, 1996; Tasker et al., 1996; Holmes et al., 2002). We agree that instead of focusing strictly on the term "region of influence", the concept of "flexible regions" should have been seen in a much broader context, i.e. in the sense of the focused pooling. That is the reason why essential works related to the regionalization of precipitation such as Schaefer (1990), Alila (1999) and Di Baldassare et al. (2006) remained beyond our field of view even though we are acquainted with all of them. The relationship to these papers will be addressed in the revised version of the manuscript, and the degree of innovation of the study will be reconsidered.

A. Castellarin: "My major comments concern: [...] 2) the way in which the Authors designed the Monte Carlo simulation experiments.

Reply #2: Similarly to the Referee who described his remarks in three paragraphs, we address the comments concerning the structure of the Monte Carlo experiment one by one.

A. Castellarin: "2.1 The analysis shows that the pooling schemes that adopt the at-site statistics to identify homogeneous groups of sites [...] outperform all other alternatives. In my opinion, an additional regional model should be considered in order to have a fair comparison. The performance of the "a1" pooling schemes should be assessed against a traditional regionalization model in which the fixed regions are identified using the same information (i.e., at-site statistics of rainfall extremes), for instance by using cluster analysis."

Reply #3: The proposed hypothetical model would be indeed interesting; however, we

HESSD

4, S1530–S1542, 2007

are not going to carry out this experiment, since the Referee suggested a complex solution to eliminate the deficiencies of the Monte Carlo experiment (see below, Reply #6).

A. Castellarin: "2.2 More importantly, over the last decade a consensus emerged that focused-pooling techniques and, more in general, regionalization should not use at-site statistics for identifying homogeneous groups of sites. [...]"

Reply #4: Our study aimed at examining whether the variability of growth curve estimates based on the ROI methodology is smaller compared to the HW models. Provided this holds true, then regional (pooled) information may be transferred to the site of the interest more effectively by means of pooling approaches. Although the construction of pooling groups is generally not acceptable based on site statistics as similarity measures, we applied them in the frequency modeling - the "a1" group of models served in order to verify the presumption whether the ROI approach is a reasonable alternative in comparison with the traditional regionalization methods. We are aware of the fact that site statistics as similarity measures do not represent a good perspective in the pooled frequency analysis since this alternative is not a proper one mainly in the case of "ungauged" issues. A viable method is to pool sites according to the site characteristics even though those alternatives exhibit a bit greater variability of the growth curve estimates.

A. Castellarin: "2.3 The study selects the GEV at-site quantiles as reference measures ("true" quantiles, pp. 2379, 2380). The sampling variability of these measures is rather high, and this may impact the significance of the Monte Carlo experiments."

Reply #5: The Monte Carlo simulation experiments were designed according to the considerations of Burn (1990, p. 2260) and Zrinji and Burn (1996, p. 250) who, for a given site, used the GEV distribution function as a parent with the parameters estimated locally. On the other hand, Castellarin et al. (2001) assessed the "true" quantiles according to a pooling scheme based on the similarity of the site statistics (at-site

HESSD
4, S1530–S1542, 2007

L-moments). We accept the remark of the Referee; the at-site quantiles as "true" quantiles will not be used in the simulation experiments of the revised manuscript.

A. Castellarin: "A possible way to address points 2.1-2.3 simultaneously would be to test the performances of the six pooling schemes a2o1-a2o3 and a3o1-a3o3 against the alternatives: "at-site", HW3r and HW1r (see section 4.2) by adopting one of the "a1" pooling alternatives [...] to compute the "true" quantiles, instead of the at-site quantiles."

Reply #6: We acknowledge the constructive criticism of the Referee; his comments also proposed a possible solution as to how to overcome the problems that occurred in the general concept of the analysis and the Monte Carlo simulation experiments. We accept the suggested modifications regarding the assessment of the "true" quantiles. The code of the ROI procedure has already been modified, and processing of the new results is in progress. The preliminary results, which involve growth curve estimates for the annual maxima of 1-day and 5-day precipitation totals, may be summarized as follows (see Replies #8 and #13 for a detailed explanation of the change made in the selection of the datasets):

1) annual maxima of 1-day precipitation:

- according to the averaged statistics, the best model is the ROI approach a3o3 (i.e. the similarity measure based on geographical proximity and all the sites included into a given site's pooling group);

- the ROI approach, regardless of the definition of the similarity measures and the pooling options, outperforms the traditional (HW) frequency models at most stations;

2) annual maxima of 5-day precipitation:

- according to the averaged statistics, the best model is the ROI approach a2o3 (i.e. the similarity measure based on descriptors of the precipitation climate and all the sites included into a given site's pooling group);

HESSD

4, S1530–S1542, 2007

- the ROI approach, regardless of the definition of the similarity measures and the pooling options, clearly outperforms the traditional (HW) frequency models.

Technical corrections:

A. Castellarin: "p. 2363, l. 5 - A citation of the 5T guideline (Jakob et al., 1999) would fit nicely here"

Reply #7: A reference to the 5T guideline will be placed at the suggested position in the text. Moreover, a brief analysis of the size of the individual pooling groups according to the different ROI models will be added to the revised manuscript.

A. Castellarin: "pp. 2367 and 2368 - Are the 2-day rainfall depth maxima suitable for characterizing the frequency of convective storms? Shouldn't we refer to shorter duration (hourly, sub-hourly)?"

Reply #8: We agree with the Referee. However, the lack of digitized sub-daily data in Slovakia does not permit the accomplishing of an experiment in the way proposed by the Referee (see also below, Reply #10). The 2-day precipitation maxima in the warm season were originally involved in the analysis mainly in order to preserve the variety of the datasets of the different seasons and/or durations. This aim, however, did not bring about the desired effect, and the transparency of the analysis was criticized by the other Referee. Therefore, in the revised manuscript, the 2-day precipitation maxima will be omitted.

A. Castellarin: "p. 2369, l. 19 - "only the Euclidean distance metric is used", I disagree, see Cunderlik and Burn (2006), also cited in the manuscript"

Reply #9: The Referee is right. The Euclidean distance metric is used in a majority of similar studies, but the Canadian group of hydrologists also reviewed the methodology of defining the proximity of sites, and introduced the Mahalanobis distance instead (Cunderlik and Burn, 2006).

A. Castellarin: "p. 2373, l. 12 - "The spatial distribution of the mean annual precipitation

HESSD

4, S1530–S1542, 2007

(MAP) exhibits strong variability"; concerning this point, the Authors should comment on regional approach proposed by Schaefer (1990) and Alila (1999) and its possible application to the Slovakian conditions"

Reply #10: Schaefer (1990) regionalized precipitation extremes (of durations of 2, 6, and 24 hours from recording gauges, and 1 day from non-recording ones, respectively) in Washington State (US) using a pooling scheme in which growth curves were estimated within flexible pooling groups, defined only in terms of the mean annual precipitation (MAP). The coefficients of variation and skewness are considered as smooth variables, which depend only on the value of MAP. Such an approach offers a relatively simple and effective methodology to estimate the design precipitation at both gauged and ungauged sites. Alila (1999) adopted a similar concept for short duration precipitation extremes over Canada; using an L-moment-based hierarchical regionalization approach, he identified the relationships between the L-coefficient of variation and MAP analogous to those of Schaefer (1990). Di Baldassare et al. (2006) confirmed the validity of the aforementioned findings (i.e. significant relationships between the statistical properties of the precipitation extremes and MAP) in northern central Italy.

An implementation of these regionalization concepts in the physical-geographical conditions of Slovakia would indeed be an interesting task. The topography of the country is rather complex, varying from relatively vast lowland areas (about 100 m a.s.l.) in the south to the mountain ranges of the Tatras (up to 2600 m a.s.l.) in its central parts. The rugged topography significantly modifies the patterns and results in the strong spatial variability of MAP. In principle, we would be able to cope with tasks like that with no major problems; the limitation that prevents us from applying this approach is the lack of available data (as the method relies on high-quality and dense datasets). The network of about 700 climatological stations and non-recording raingauges operated by the Slovak Hydrometeorological Institute (SHMI) forms an acceptable basis for accomplishing the proposed analysis for the 1-day precipitation data; however, we would not be able to generalize the results in the domain of sub-daily durations, which is usually of par-

HESSD

4, S1530–S1542, 2007

ticular interest. Only a minority of the stations operated by SHMI (25-30) have longer continuous recordings, and even though some of these stations have really long series of observations (in some cases since 1961), the key problem is the lack of digitized data.

A. Castellarin: "Section 3.4 - Is the reference to confidence intervals of estimated quantiles really useful in the context of this analysis?"

Reply #11: Confidence intervals are mentioned in connection with the Monte Carlo simulations as a potentially useful derivative. Regarding the fact that the simulations are carried out in order to compare the performance of different frequency models, the reference to the confidence intervals is not necessary, and it will be omitted in the revised manuscript. Sections 3.4 and 3.5 will then be joined with the common title "Evaluation of the frequency models".

A. Castellarin: "p. 2386, l. 21 - "The results demonstrate that the at-site approach to frequency analysis is the least suitable method for the estimation of heavy precipitation quantiles". It does not seem so for small recurrence intervals (T $<$ 10-20 years) from Tables 1-3 and Figures 3-5, but this is a consequence of the selected "true" quantiles."

Reply #12: As mentioned above (i.e. Reply #6), the experiment has been redesigned. According to the new findings, the at-site method of estimation shows better behavior than the regional models only in the case of the bias characteristics and for low return periods (T $<$= 20 years). Nevertheless, in terms of the RMSE, the at-site model is the least suitable one even in the case of small recurrence intervals.

Other comments are of a technical nature, and the suggested changes will be incorporated in the revised manuscript.

Reply to the comments of M. Pfaundler

General comments:

M. Pfaundler: "[...] there were so many models (combinations of alternatives and op-

HESSD
4, S1530–S1542, 2007

tions) analyzed and commented on, that the sheer wealth of statistical analyses was hiding to me at a certain point of reading the paper the question and aim of the paper"

Reply #13: Following the suggestions of the other Referee (see above, Reply #6), the revised manuscript will involve a reduced number of frequency models: alternative #1 will play a role only as the model for estimating the "true" quantiles in the Monte Carlo simulation experiment. Moreover, in order to enhance the transparency of the analysis, only two datasets (the annual maxima of 1-day and 5-day precipitation totals, respectively) will be considered, for the following reasons:

- the dataset of the annual maxima of 1-day totals remains unchanged - it is a standard, frequently analyzed dataset in similar studies as well as in practical applications;

- the dataset of the 2-day maxima in the warm season will be omitted - for details see above, Reply #8;

- the dataset of the 5-day maxima in the cold season will be replaced by the annual maxima of 5-day precipitation totals - this makes a compromise between not dramatically changing the structure of the original manuscript and making the analysis more transparent.

The most important point, however, is that the reduction in the analyzed datasets has no particular effect on the general conclusions of the analysis, which supports the superiority of the pooling approaches over the traditional regionalization methods. The seasonal differences that are also related to different circulation patterns and prevailing precipitation mechanisms in the warm and cold seasons lie outside the scope of the present paper and will be analyzed in a separate study.

M. Pfaundler: "I would put it rather at the lower end of the originality scale"

Reply #14: The present work is a kind of case study, i.e. an application of the ROI approach to the frequency analysis of precipitation extremes in Slovakia. The main innovation in the ROI methodology is the incorporation of a specific index (Lapin's index

HESSD

4, S1530–S1542, 2007

of Mediterranean influence) as a measure of the long-term precipitation regime. The applicability of the pooling approach in the given specific climatological and physical-geographical settings was the main goal of the analysis.

In the context of the state-of-the-art in Slovakia, the proposed paper is a step forward from the generally accepted HW methodology. It could also be seen as a pilot study for the Carpathian region: after a refinement of the methodology and dataset extension from a much broader region, the methodology has a great potential in hydrometeorological applications in other countries as well.

M. Pfaundler: "What limits a bit the practical value is twofold: First the best performing approach relies on statistical at-site characteristics [...]; thus, it is of no use for ungauged sites"

Reply #15: This is the general conclusion in the comments of both Referees (see above, Replies #5 and #6), and in the revised manuscript, we are going to follow the proposed suggestions and pool the sites based exclusively on at-site characteristics.

M. Pfaundler: "What limits a bit the practical value is twofold: [...] Second, the analysis addresses only the growth curve, but does not enter into the question of how to estimate the index storm."

Reply #16: No attention has been paid to the estimation of the index storm, mainly for the following reasons:

- The ROI methodology has been examined only at sites with direct meteorological observations, where the value of the index storm is known. The issue of the design value estimation at ungauged sites, which naturally implies a need for the assessment of the index value, has not been addressed thus far.

- The index storm is usually estimated with a considerable degree of uncertainty, mainly in a low density network of sites such as the one employed in the manuscript. Therefore, we aimed at not making the interpretation of the results of the simulation experi-

HESSD

4, S1530–S1542, 2007

ments more difficult by considering the uncertainty of the index storm estimates.

There are several methods for estimating the index storm at ungauged places (e.g. mapping techniques, regional regression methods, application of neural networks etc.). The study of this issue is in progress, and a separate paper will be devoted to it. Following this, a joint analysis of the uncertainties of the growth curve and the index storm estimates becomes possible.

M. Pfaundler: "[...] the "true" at-site quantiles are identical with the at-site sample quantiles. Of course this is an unavoidable and wide-spread procedure in that kind of analysis but it is somewhat critical [...] and has implications on the significance of the findings and conclusions"

Reply #17: The strategy suggested by the other Referee will be applied in the revised simulation experiments (see above, Replies #5 and 6), which eliminates the integration of the at-site quantiles as reference quantities.

M. Pfaundler: "The analyses associate equal weights to the used attributes for the definition of similarity. [...] I would add some further comments that attention has to be paid when selecting the attributes such as examining that the chosen attributes are independent and not correlated or carry redundant information."

Reply #18: Burn (1990a) in his original framework presented a simple methodology (correlation analysis) for the selection of site attributes and estimation of their weights. On the other hand, Castellarin et al. (2001) used equal (unit) weights for the selected attributes emphasizing that it allows for a better understanding of the effects of the differences between various attribute sets. Anyway, a comment on the selection of site attributes will appear in the revised manuscript:

The wider the database of the site attributes the greater the potential for a successful pooling / clustering / etc. However, one should be aware of redundant information involved among the selected attributes. They are often cross-correlated, carrying the

HESSD
4, S1530–S1542, 2007

same information about the analyzed phenomena. Clustering techniques are sensitive to the redundancy of information (Guttman, 1993; Gong and Richman, 1995); in practice, the selected attributes should be as independent as possible. There are several multivariate statistical methods for filtering redundant information, such as principal component analysis, factor analysis or Procrustes analysis - Dinpashoh et al. (2004) is an example of a thorough examination of a wide database of site attributes. The risk of having unduly correlated characteristics becomes greater with increasing the number of attributes, and that is why we confined the analysis to three attributes in all the alternatives.

M. Pfaundler: "I would recommend trying to enlarge the "information-pool" [...] by gathering and adding stations from beyond the Slovak borders"

Reply #19: We entirely agree with the Referee; however, there are administrative problems in getting data from the national weather services of neighboring countries. Although some datasets are available within the ECA&D project (European Climate Assessment and Dataset, http://eca.knmi.nl/), most neighboring countries (Hungary, Austria, Poland, the Czech Republic) provide either no data or only a limited number (1-2) of freely available series within the ECA&D project, and the direct exchange of climatological data (with daily resolutions) between the meteorological services does not operate in Central Europe. Another issue is the mutual (in)consistency of the datasets of different origins (e.g. temporal homogeneity).

All other technical corrections suggested by the Referee will be addressed, and corrections will be made in the revised manuscript.

References:

Alila, Y.: A hierarchical approach for the regionalization of precipitation annual maxima in Canada, J. Geophys. Res., 104(D24), 31,645-31,656, 1999.

Burn, D. H.: An appraisal of the "region of influence" approach to flood frequency

HESSD

4, S1530–S1542, 2007

analysis. Hydrolog. Sci. J., 35(2), 149-165, 1990a.

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

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

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

Cunderlik, J. M. and Burn, D. H.: Switching the pooling similarity distances: Mahalanobis for Euclidean. Water Resour. Res., 42(3), W03409, 2006.

Di Baldassare, G., Castellarin, A. and Brath, A.: Relationships between statistics of rainfall extremes and mean annual precipitation: an application for design-storm estimation in northern central Italy. Hydrol. Earth Syst. Sci., 10, 589-601, 2006.

Dinpashoh, Y., Fakheri-Fard, A., Moghaddam, M., Jahanbakhsh, S. and Mirnia, M.: Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. J. Hydrol., 297, 109-123, 2004.

Faulkner, D.: Rainfall frequency estimation. In: Flood Estimation Handbook, Vol. 2, Institute of Hydrology, Wallingford, UK, 1999.

Gong, X. and Richman, M. B.: On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. J. Climate, 8(4), 897-931, 1995.

Guttman, N. B.: The use of L-moments in the determination of regional precipitation climates. J. Climate, 6(12), 2309-2325, 1993.

Holmes, M. G. R., Young, A. R., Gustard, A. and Grew, R.: A region of influence approach to predicting flow duration curves within ungauged catchments. Hydrol. Earth Syst. Sc., 6(4), 721-731, 2002.

HESSD

4, S1530–S1542, 2007

Hosking, J. R. M. and Wallis, J. R.: Regional frequency analysis: an approach based on L-moments. Cambridge University Press, Cambridge, 1997. 224 pp.

Jakob, D., Reed, D. W. and Robson, A. J.: Selecting a pooling-group. In: Flood Estimation Handbook, Vol. 3. Institute of Hydrology, Wallingford, UK, 1999.

Lu, L.-H. and Stedinger, J. R.: Sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test. J. Hydrol., 138, 223-245, 1992.

Reed, D. W., Jakob, D., Robinson, A. J., Faulkner, D. S. and Stewart, E. J.: Regional frequency analysis: a new vocabulary. Hydrological Extremes: Understanding, Predicting, Mitigating, Proc. IUGG 99 Symposium, Birmingham, IAHS Publ. no. 255, 237-243, 1999.

Schaefer, M. G.: Regional analyses of precipitation annual maxima in Washington State. Water Resour. Res., 26(1), 119-131, 1990.

Tasker, G. D., Hodge, S. A. and Barks, C. S.: Region of influence regression for estimating the 50-year flood at ungaged sites. Water Resour. Bull., 32(1), 163-170, 1996.

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

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

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 4, 2361, 2007.