Vienna Doctoral Programme on **Water Resource Systems**

Doctoral Thesis

Quantifying uncertainty components in flood frequency estimation

submitted in satisfaction of the requirements for the degree of
Doctor of Science in Civil Engineering
of the Vienna University of Technology, Faculty of Civil Engineering

as part of the
Vienna Doctoral Programme on Water Resource Systems

by

Ing. José Luis Salinas Illarena
Matrikelnummer 0827331
Alserbachstrasse 15/13, 1090 Wien

Examiner : Univ.Prof. Dipl.-Ing. Dr. techn. Günter Blöschl
Institut für Wasserbau und Ingenieurhydrologie, TU Wien
Karlsplatz 13, 1040 Wien

Examiner : Univ.Prof. Dipl.-Ing. Dr. techn. Christian Bucher
Institut für Institut für Hochbau und Technologie, TU Wien
Karlsplatz 13, 1040 Wien

Examiner : Prof. Dr. rer. nat. habil. Andreas H. Schumann
Department of Water Resources Management and Environmental Engineering,
Ruhr- University Bochum
Universitätsstraße 150, 44801 Bochum, Germany

Vienna, February 2015

I would like to thank all my colleagues and friends from the Institute of Hydrologic Engineering at the TU Wien, for the nice atmosphere and working environment during this 5 years.

My special gratitude goes to Andrea Kiss, for making the last chapter of this thesis a reality, to my italian roommates Alberto and Maggie, for all the laughs inside and outside the University, and of course to my supervisor Prof. Günter Blöschl, for his patience, sharing his vision about hydrology and about science, and for his extraordinary guidance. Thank you, Günter.

I want to dedicate this thesis to my parents, and my brother,
who have always supported me from the distance.

And last but not least, I wish to thank my partner Mariette,
for being there always for me when I needed her, and for dealing with my stress of the last months
I love you, and I could not have done this without you.

Abstract

During the last decade, a series of large river flooding events (e.g. 2005 in the alpine region, 2013 in central Europe, or 2014 in northwestern Italy) have caused severe damages in Europe. An improved concept of integrated flood risk management in necessary, in order to manage and minimize flooding risks in the future, and to reduce the catastrophic nature of these events. A crucial step in any flood risk assessment is the accurate estimation of extreme flood peak discharges associated with a very low exceedance probability, i.e. with high return periods, which are then used for hydraulic design purposes and risk zone mapping. The uncertainties involved in these estimates need to be quantified, for reliable decision making in these tasks. The aim of this thesis is to better understand the sources and nature of different uncertainty components present in flood frequency estimation, and provide methods for quantifying them at different spatial and temporal scales.

If discharge measurements are not available, flood design values can be computed either from precipitation data by rainfall-runoff modelling, or by transferring the flood regime information to the target site from neighboring donor catchments with statistical regionalisation methods. In the context of the latter, Chapter 2 of this thesis investigates the uncertainties involved in predicting flood frequencies in ungauged catchments as a function of climate, method, and data availability. A global meta-analysis of the existing literature in the last twenty years is performed, involving a total of 3023 catchments worldwide. The reported cross-validation predictive performances of regionalisation methods are used as a surrogate for the total uncertainty involved in the flood frequency estimation when no discharge data is available locally. The results indicate that flood predictions in ungauged catchments are, on average, less accurate in arid than in humid climates and more accurate in large than in small catchments. There is also a tendency towards a lower performance of regressions as compared to other methods when they are applied to the same region, while geostatistical methods generally tend to perform better than other methods. For the particular case of arid catchments, index methods yield significantly lower performances than regression methods or geostatistical approaches.

When flood records are present, design values are usually estimated by flood frequency analysis, i.e., by applying the statistical theory of extreme values to obtain a probability distribution function that describes the flood regime for a certain location. If the flood frequency estimation for an entire region is considered, the choice of a common statistical model for the entire area, also called parent distribution, becomes the first step in approaches such as the index flood method. Chapter 3 deals with the uncertainty associated with the flood frequency model choice, by addressing the question of the existence of a parent flood frequency distribution at a European scale. A simple exploratory analysis of a newly compiled database of L-moment ratios of flood annual maximum series from 4105 catchments suggests the suitability of the *Generalised Extreme Value* (GEV) distribution as

a pan-European flood frequency distribution. However, more detailed Monte Carlo simulations show that the GEV model underestimates the variability in terms of sample skewness and kurtosis present in the data, and particularly fails to represent the kurtosis dispersion for longer sample sizes and medium to high skewness values. Therefore, the GEV distribution was rejected in a statistical hypothesis testing framework as a single pan-European parent distribution for annual flood maxima. The results presented in this chapter indicate that one single statistical model may not be able to fit the entire variety of flood processes present at a European scale.

Chapter 4 further investigates the catchment and climatic factors controlling European flood regimes and their effects on the underlying flood frequency distributions. In particular, the uncertainty in statistical model choice is linked to catchment size and mean annual precipitation (MAP) using flood data from a total of 813 catchments with more than 25 years of record from Austria, Italy and Slovakia. Results shows that the GEV distribution provides a better representation for regionally averaged values of sample L-moment ratios than the other distributions considered, for catchments with medium to high MAP independently of catchment area, while the *three-parameter lognormal* distribution is a more appropriate choice for the drier (lower MAP) intermediate-sized catchments, which exhibit higher skewnesses. The results presented in this chapter could be seen as a first attempt at defining a set of "process-driven" regional parent flood frequency distributions in a European context.

After analysing the uncertainties in flood frequency across different spatial scales in Chapters 2 to 4, Chapter 5 deals with the flood frequency estimation at a particular location in space, but extends the temporal scale of the uncertainties several centuries into the past by introducing historical flood records in the analysis. Knowledge about the historical flood regime is useful because it gives additional information that may improve the estimates of extreme discharges with high return periods and, additionally, may reduce the uncertainty in the estimates. In most practical cases, the information related to historical floods is given in a non-precise manner. This chapter presents a new approach to dealing with the imprecision present in historical floods, that links the descriptions in historical records to fuzzy numbers representing discharges. These fuzzy historical discharges are then introduced in a formal Bayesian inference framework to obtain a fuzzy version of the flood frequency curve, by combining the fuzzy historical flood events and the instrumental data for a given location. Two case studies are selected from the historical literature, representing different facets of the fuzziness typically present in the historical sources. The results are given in the form of the fuzzy estimates of the flood frequency curves together with the fuzzy credibility bounds for these curves. The presented fuzzy Bayesian inference framework provides a flexible methodology to propagate, in an explicit, way the imprecision from the historical records to the flood frequency estimate, which allows assessing the effects of incorporating non-precise historical information in the flood frequency regime estimation.

The findings presented in this thesis help better characterize and quantify different facets of uncertainty involved in the flood frequency estimation process. While these different facets of uncertainty are usually lumped together, the present work aims at throwing light at possible sources of these uncertainties, by analysing the model and data related aspects that constrain them, and by defining the characteristic spatial and temporal scales under which they operate. The results of this thesis have implications for both hydrological understanding, and applied engineering hydrology. On the one hand, linking uncertainties in flood frequency estimation with hydrological and climatological indicators helps identify regions where an improved hydrological process understanding in needed. On the other hand, an improved quantification of the uncertainties helps in obtaining more robust and reliable design decisions and flood risk zones.

Contents

Chapter 1

Introduction

During the last decade, a series of large river flooding events have affected Europe, causing several billions Euro (only the damages of the 2013 floods in central Europe were estimated as 3bn Euro by the reinsurance *Munich Re*). In order to manage and minimize future flooding risks, an improved and updated concept of integrated flood risk management in necessary, so that the catastrophic nature of these events is reduced. The first and most crucial step in any flood risk assessment is the accurate estimation of extreme flood peak discharges associated with a very low exceedance probability, i.e. with high return periods, which are then used for hydraulic design purposes and risk zone mapping. The uncertainties involved in these estimates need also to be quantified, for reliable decision making in these tasks. In this context, the aim of this thesis is to better understand the sources and nature of different uncertainty components present in flood frequency estimation, and provide methods for quantifying them at different spatial and temporal scales. While uncertainties themselves in hydrological modelling can have very different natures (see e.g. *Plate*, 2002; *Merz and Thieken*, 2005; *Schumann*, 2011), the present dissertation will deal with four particular kinds of uncertainty, present in different operational setups of flood frequency estimation:

(a) How big is the uncertainty in predicting design floods in an ungauged basin? How is it related with the catchment attributes, the hydroclimatic characteristics of the region, and the method used to predict the floods?

(b) How big is the uncertainty caused by statistical model choice, when obtaining a regional parent distribution? Does the parent distribution need to exist in all spatial scales?

(c) How can we relate the uncertainty in statistical model choice with catchment and climatic attributes? Are there hydroclimatic regions where a certain model is clearly more appropriate, and therefore the uncertainty caused by model choice will be much reduced?

(d) Can we incorporate in the flood frequency estimation information about historical floods, even if these are given in a non-precise way? How does this imprecision propagates to the flood estimate?

Estimating flood discharges in ungauged basins ist among the most fundamental challenges in catchment hydrology. There is a long track record in statistical hydrology of developing methods to estimate, in an optimal way, these discharges from runoff observations in neighbouring catchments and from catchment characteristics. A classical approach is the index flood method (*Dalrymple,* *1960*) where the flood distribution function scaled by the index flood (e.g. the mean annual flood) is assumed to be homogenous within the region. The procedure consists of first estimating the index flood in the ungauged catchment (e.g. by a regression against catchment characteristics) and then multiplying that index flood with the regional scaled flood distribution function (*IH, 1999*). Flood quantiles regressions against catchment characteristics have also become popular in the last decades (see, e.g. *Cunnane, 1988, and Griffis and Stedinger, 2007*). More recently, geostatistical methods that exploit the spatial correlation of floods either in space (*Merz and Blöschl, 2005*) or along the stream network (see *Skøien et al., 2006*) have been developed, particulary in regions with high gauging desities. For a comprehensive review of methods and studies during the last 20 years, see *Blöschl et al, (2013).*

When flood records are present, design values are usually estimated by flood frequency analysis, i.e., by applying the statistical theory of extreme values to obtain a probability distribution function that describes the flood regime for a certain location. If the flood frequency estimation for an entire region is considered, the choice of a common statistical model for the entire area, also called parent distribution, becomes the first step in approaches such as the index flood method. The question itself of existence of parent distributions at different spatial scales is a topical issue in statistical flood hydrology (see e.g. *Laio et al., 2009*). For example *Matalas et al. (1975), Dawdy and Gupta (1995), Houghton (1978)* and others, found that the variability in terms of sample statistics was always higher for observed data than for simulated flood peaks for a set of considered parent distributions, attributing it to different reasons.

Changing from spatial, regional to temporal scales, the incorporation of historical floods in flood frequency estimation has been perfomed in the literature in several occasions (see e.g. *Leese, 1973; Stedinger and Cohn, 1986; Benito and Thorndycraft, 2005*), but always with very simplified assumptions about the precision of the historical data. In this sense, fuzzy numbers *(Zadeh, 1965)* represent a straightforward model for the non-precise descriptions in historical records, giving indications about flood events in the past.

The objective of this thesis will be to deeply examine the described facets of uncertainty in flood estimation, analyse the context in which they appear, an develop formal methodologies to quantify them. Chapter 2 of this thesis will investigate the uncertainties involved in predicting flood frequencies in ungauged catchments as a function of climate, method, and data availability. Chapter 3 will deal with the uncertainty associated with the flood frequency model choice, by addressing the question of the existence of a parent flood frequency distribution at a European scale. Chapter 4 further investigates the catchment and climatic factors controlling European flood regimes and their effects on the underlying flood frequency distributions. Chapter 5 will deal with the flood frequency estimation at a particular location in space, but extending the temporal scale of the uncertainties several centuries into the past by introducing historical flood records in the analysis by means of a fuzzy model for their imprecission. Finally, Chapter 6 will present an overview of the results, together with the conclusions.

Chapter 2

Comparative assessment of flood predictions in ungauged basins

The present chapter corresponds to the following scientific publication in its original form:

Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies, *Hydrol. Earth Syst. Sci.*, 17, 2637-2652, doi:10.5194/hess-17-2637-2013, 2013.

Hydrol. Earth Syst. Sci., 17, 2637–2652, 2013
www.hydrol-earth-syst-sci.net/17/2637/2013/
doi:10.5194/hess-17-2637-2013

Hydrology and Earth System Sciences
Open Access

Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies

J. L. Salinas[1], G. Laaha[2], M. Rogger[1], J. Parajka[1], A. Viglione[1], M. Sivapalan[3], and G. Blöschl[1]

[1]Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Vienna, Austria
[2]Institute of Applied Statistics and Computing, University of Natural Resources and Life Sciences, BOKU, Vienna, Austria
[3]Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 North Mathews Ave., Urbana, USA

Correspondence to: J. L. Salinas (salinas@hydro.tuwien.ac.at)

Received: 21 December 2012 – Published in Hydrol. Earth Syst. Sci. Discuss.: 14 January 2013
Revised: 28 May 2013 – Accepted: 2 June 2013 – Published: 9 July 2013

Abstract. The objective of this paper is to assess the performance of methods that predict low flows and flood runoff in ungauged catchments. The aim is to learn from the similarities and differences between catchments in different places, and to interpret the differences in performance in terms of the underlying climate-landscape controls. The assessment is performed at two levels. The Level 1 assessment is a meta-analysis of 14 low flow prediction studies reported in the literature involving 3112 catchments, and 20 flood prediction studies involving 3023 catchments. The Level 2 assessment consists of a more focused and detailed analysis of individual basins from selected studies from Level 1 in terms of how the leave-one-out cross-validation performance depends on climate and catchment characteristics as well as on the regionalisation method. The results indicate that both flood and low flow predictions in ungauged catchments tend to be less accurate in arid than in humid climates and more accurate in large than in small catchments. There is also a tendency towards a somewhat lower performance of regressions than other methods in those studies that apply different methods in the same region, while geostatistical methods tend to perform better than other methods. Of the various flood regionalisation approaches, index methods show significantly lower performance in arid catchments than regression methods or geostatistical methods. For low flow regionalisation, regional regressions are generally better than global regressions.

1 Introduction

Estimating flood and low flow discharges in ungauged basins are among the most fundamental challenges in catchment hydrology. There is a long track record in statistical hydrology of developing methods to estimate, in an optimal way, these discharges from runoff observations in neighbouring catchments and from catchment characteristics. Common to these statistical methods is the idea of catchment grouping, i.e. the notion that extreme events that have not been observed in a particular location could already have been observed somewhere else. Therefore runoff data (on floods or low flows) from many sites are pooled in order to obtain a representative sample of what could happen in a particular location. One of the key aspects of the methods consists of exactly how this pooling is performed.

There are a number of options. The classical approach consists of subdividing the study domain into a number of fixed, contiguous regions which are used to regionalise floods or low flows for all catchments in the area (e.g. as used in the index flood method, Dalrymple, 1960). The assumption of this method is that areas close to each other are characterised by similar climate, topography, geology, soils and land use, which gives rise to similar catchment hydrological response and therefore to similar floods or low flows. The grouping is usually found by geographical boundaries, by combining maps of the catchment characteristics in some way (Beable and McKerchar, 1982) or by a diverse set of statistical methods. These include cluster analysis using catchment characteristics (Nathan and McMahon, 1990), residuals from a

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

regression model (Wandle, 1977; Hayes, 1992), regression trees (Laaha and Blöschl, 2006a), and pattern identification on the basis of the seasonality of runoff as an indicator of flood and low flow processes in the catchment (Laaha and Blöschl, 2006b; Piock-Ellena et al., 1999). An alternative is the region of influence (ROI) approach (Burn, 1990) which assigns a different pooling group to each catchment of interest. Similarity between catchments is usually measured by the root mean square difference of all the catchment and climate characteristics in a pair of catchments. A typical application of the ROI approach is given in the UK Flood Estimation Handbook (IH, 1999). The catchments characteristics for the grouping usually include mean annual rainfall, catchment area and soil characteristics.

Once the pooling group has been identified there are again a number of options of how to estimate the flood or low flow discharges. Again a classical one is the index flood method (Dalrymple, 1960) where the flood distribution function scaled by the index flood (e.g. the mean annual flood) is assumed to be homogenous within the region. The procedure consists of first estimating the index flood in the ungauged catchment (e.g. by a regression against catchment characteristics) and then multiplying that index flood with the regional scaled flood distribution function (IH, 1999) or by multiplying that index low flow with the regional scaled low flow distribution function (Clausen and Pearson, 1995; Madsen and Rosbjerg, 1998). With the advent of geographic information systems, alternative methods of using the flood quantiles or low flow quantiles directly in regressions against catchment characteristics have become popular (see, e.g. Cunnane, 1988, and Griffis and Stedinger, 2007, for the case of floods, and Gustard et al., 1992, and Engeland and Hisdal, 2009, for the case of low flows). More recently, geostatistical methods that exploit the spatial correlation of floods (or low flows) either in space (Merz and Blöschl, 2005) or along the stream network (see Skøien et al. (2006) for the case of floods and Laaha et al. (2012) for the case of low flows) have become popular. One of the strengths of the geostatistical approach is that it directly exploits the spatial correlations of the discharges and there is no need for defining pooling groups explicitly, but a relatively dense stream gauge network is needed. There are also methods that estimate flood statistics in ungauged catchments from rainfall (e.g. Moretti and Montanari, 2008).

When reviewing the rich literature on estimating extreme discharges in ungauged basins it is interesting that many of the statistical methods for floods and low flows are similar if not identical. Given this similarity, it is quite surprising that there are very few studies that directly compared the estimation methods for floods and low flows. Another interesting finding is that the predictive performance for ungauged basins strongly depends on the hydrological or climatological setting of the region (Meigh et al., 1997; Farquharson et al., 1992). There is no consensus in the literature on whether one method always outperforms another. This is because there have been few attempts in generalising the findings on the predictive performance of estimation methods beyond individual case studies. Yet, it would be very interesting to understand whether there are general patterns of performance, i.e. whether particular methods generally perform better than others in a given environment. These are the issues, this paper is concerned with. Specifically, in this paper we perform a meta-analysis of the literature on predictive performance of flood and low flow estimation methods in ungauged basins. In a second step we analyse a number of more detailed datasets, again focusing on the performance of the methods. The aim is to learn from the similarities and differences between catchments in different places, and to interpret the differences in predictive performance in terms of the underlying climate–landscape controls. The following research questions are addressed:

i. How good are the predictions of hydrological extremes in different climates?

ii. Which regionalisation method performs best?

iii. How does data availability impact performance?

iv. To what extent does runoff prediction performance depend on climate and catchment characteristics?

This paper is part of a set of three papers that are all concerned with assessing the performance of estimating runoff characteristics in ungauged basins. The two companion papers (Parajka et al., 2013; Viglione et al., 2013) deal with estimating runoff hydrographs in ungauged basins and estimating a set of different runoff characteristics in Austria, respectively.

2 Method of comparative assessment

For the comparative assessment of both flood and low flow predictions in ungauged basins, the same two step process as in Parajka et al. (2013) has been adopted in this paper and is presented below.

Level 1 assessment: in a first step, a literature survey was performed. Publications in the international refereed literature were scrutinised for results of the predictive performance of both floods and low flows. The Level 1 assessment is a meta-analysis of prior studies performed by the hydrological community. The advantage of this type of meta-analysis is that a wide range of environments, climates and hydrological processes can be covered that go beyond what can be reasonably achieved by a single study. It is a comparative assessment that synthesises the results from the available international literature. However, the level of detail of the information provided is often limited. The results in the literature were almost always reported in an aggregated way, i.e. as average or median performance over the study region or part of the study region.

Level 2 assessment: to complement the Level 1 assessment, a second assessment step was performed, termed Level 2 assessment. In this step, some of the authors of the publications from Level 1 were approached to provide data on their floods and low flow predictions *for individual basins*. The data they provided included information on the catchment and climate characteristics, on the method used, the data availability, and predictive performance. The overall number of catchments involved was smaller than in the Level 1 assessment, so the spectrum of hydrological processes covered in the assessment could be potentially narrower. However, the amount and detail of information available in particular catchments was much higher. As in Level 1, the cross-validation performance for ungauged basins was analysed; however, information on individual catchments was now available. The cross-validation performance was estimated by a leave-one-out strategy, where each gauged catchment was in turn considered as ungauged and the estimated low flow or flood index was compared with the observed one.

The comparative assessment conducted in this paper stratifies the analyses into three main groups:

1. *Analysis of process controls on the predictive performance*. A number of climate and catchment characteristics have been identified. A large number of catchments and modelling studies around the world have then been organised according to these climate and catchment characteristics, with the objective of learning from their differences and similarities in performance in a general way.

2. *Analysis of predictive performance for different types of methods*. The methods for estimating flood and low flow indexes in ungauged basins have been grouped into the classes discussed in Sect. 3. Rather than evaluating specific methods the focus has been on types of method, so to be able to generalise beyond individual studies.

3. *Analysis of data availability*. The quality of predictions of extremes in ungauged basins not only depends on the hydrological setting and the regionalisation method but also, importantly, on the data that are available for the information transfer. The comparison therefore also examines the number of stream gauges available in a particular study as an index to characterise data availability.

3 Studies and datasets used

3.1 Low flow studies

Table 1 lists the 14 low flow prediction studies used in this paper. It includes summary information about the study region, regionalisation method applied and the predictive performance in terms of the coefficient of determination (R^2), defined as follows:

$$R^2 = 1 - \frac{\sum \left(Q_{i,\text{pred}} - Q_{i,\text{obs}}\right)^2}{\sum \left(Q_{i,\text{obs}} - \overline{Q}_{\text{obs}}\right)^2}, \tag{1}$$

where

$Q_{i,\text{pred}}$: predicted specific discharge in cross-validation at gauge i,

$Q_{i,\text{obs}}$: observed specific discharge at gauge i,

$\overline{Q}_{\text{obs}}$: spatial mean of the observed specific discharge.

In the great majority of the papers considered, the performance is given in terms of the described coefficient of determination in cross validation, which reports the amount of explained variance by the model, and is also affected by both bias and dispersion of the estimators. The target low flow index, on which this performance is reported, is mainly the q_{95} specific discharge quantile, i.e. the discharge value exceeded 95 % of the time divided by the catchment area, but there were studies presenting performances on other low flow indicators including $q_{7,10}$ (7 days 10 yr specific runoff), $q_{\text{mon},5}$ (monthly 5 day minimum), q_{96}, q_{97} (96–97 % specific runoff quantiles), q_{95}/q_{A} (q_{95} specific runoff quantile normalised by the mean annual specific runoff q_{A}) and baseflow index (BFI). Both the performance measure and the low flow index used in the analysis represent a trade-off between the amount of studies potentially to be included in the analysis and their need to be comparable; the same applies to the flood studies. Several studies compare different regionalisation approaches and/or subsets of data which results in a total of 28 assessments of predictive performance. These results are the base for the Level 1 assessment which represents at total of 3112 catchments (Table 2). Geographically, most of the cross-validation assessments were performed in Europe and North America and only a few studies cover Australia and Asia (Fig. 1, top and Table 1). Six study authors out of the Level 1 assessment provided detailed information about climate and catchment characteristics in a consistent way and reported the regionalisation performance for each catchment (Level 2 assessment). In this sense, the potential of learning from the catchment-by-catchment errors in contrast to the aggregated, regional measures of Level 1 represents a motivation for the Level 2 assessment. Predictive performance on a catchment basis was given as the absolute normalised error (ANE), defined as

$$\text{ANE}_i = \left| \frac{Q_{i,\text{pred}} - Q_{i,\text{obs}}}{Q_{i,\text{obs}}} \right|. \tag{2}$$

The dataset for Level 2 assessment combines data from 1895 catchments. Three catchment characteristics are analysed: aridity index, mean elevation and catchment area.

Table 1. Summary assessment of studies for low flow estimation in ungauged catchments used in Level 1 assessment. Performance indicates the leave-one-out assessment of model efficiency in terms of the coefficient of determination R^2. Low flow regionalisation methods include: process based (PB), global regression (GR), regional regression (RR), geostatistics (G) and short records (SR). Predicted variable indicates the low flow index estimated in the study and includes: 7 days 10 yr specific runoff ($q_{7,10}$), monthly 5 day minimum specific runoff ($q_{\mathrm{mon},5}$), 95–97 % specific runoff quantiles (q_{95}, q_{96}, q_{97}), normalised q_{95} specific runoff quantile (q_{95}/q_{A}) and baseflow index (BFI). Ranges or various values for R^2 represent variations of the methods or the same method applied on different subsamples from the same region.

Study	Region	Climate	Number of catchments	Regionalisation method	Predicted variable	Performance (R^2)	Used in Level 2
Eng et al. (2011)	eastern USA	Humid	516, 125, 422	SR	$q_{7,10}$	0.96, 0.99, 0.97	X
Castiglioni et al. (2011)	central Italy	Humid	51	G	q_{97}	0.89	
Plasse and Sauquet (2010)	France	Humid	1003	GR, RR, G, G	$q_{\mathrm{mon},5}$	0.43, 0.53–0.74, 0.61, 0.63–0.73	X
Vezza et al. (2010)	northwest Italy	Cold	41	GR, RR	q_{95}	0.57, 0.53–0.69	
Engeland and Hisdal (2009)	southwest Norway	Cold	51	RR, PB	q_{96}	0.82, 0.32	X
Laaha and Blöschl (2007)	Austria	Cold	325	RR	q_{95}	0.75	
Laaha et al. (2007)	Austria	Cold	298	G	q_{95}	0.75	X
Laaha and Blöschl (2006a, b)	Austria	Cold	325	GR, RR	q_{95}	0.57, 0.59–0.70	X
Laaha and Blöschl (2005)	Austria	Cold	325	SR	q_{95}	0.62, 0.93	X
Rees et al. (2002)	Himalayas, Nepal and India	Humid	40	GR	q_{95}/q_{A}	0.45, 0.53	
Aschwanden and Kan (1999)	Switzerland	Cold	143	GR, RR	q_{95}	0.51, 0.59-0.84	
Demuth and Hagemann (1994)	Germany (Baden-Württemberg)	Humid	54	GR	BFI	0.86	
Demuth (1993)	Germany (Baden-Württemberg)	Humid	54	GR	BFI	0.81, 0.84	
Nathan and McMahon (1990, 1992)	Australia (New South Wales, Victoria)	Arid	184	RR, GR	BFI	0.75–0.83, 0.71	

These characteristics represent a trade-off between the data availability of the studies, and the literature reports on the main controls of flood and low flow regimes. Aridity (the ratio of potential evaporation E_{PA} and precipitation P_{A} on a long-term basis, averaged across the catchment) is an indicator of the competition between energy and water affecting the water balance. Elevation (average topographic elevation within the catchment) is a composite indicator including a range of processes, such as long-term precipitation and hence soil moisture availability, and air temperature. In some environments there is a relationship between elevation and aridity and elevation and snow processes. Catchment area is an indicator of the degree of aggregation of catchment processes related to scale effects (Skøien et al., 2003); an indicator of storage within the catchment. Catchment size also acts as an indicator of the quality of rainfall data that is available for runoff estimation in ungauged basins, as for a constant raingauge density, the mean areal rainfall estimation variance decreases with increasing the catchment area. This areal rainfall might also be biased by increasing the number of stations located in lower parts of the catchment (Lebel et al., 1987) The low flow regionalisation methods have been classified into the following groups.

- Process-based methods (PB): there is only a single cross-validation study we encountered in the literature (Engeland and Hisdal, 2009) of this type. The procedure consisted of regionalising the parameters of a conceptual rainfall–runoff model from gauged to ungauged catchments in the region. The low flow characteristics were then derived from the simulated daily hydrographs at the ungauged location of interest.

- Global Regression (GR): in the global regression approach a single relationship between the low runoff statistic of interest, such as q_{95}, and catchment/climate characteristics is established. Both additive and

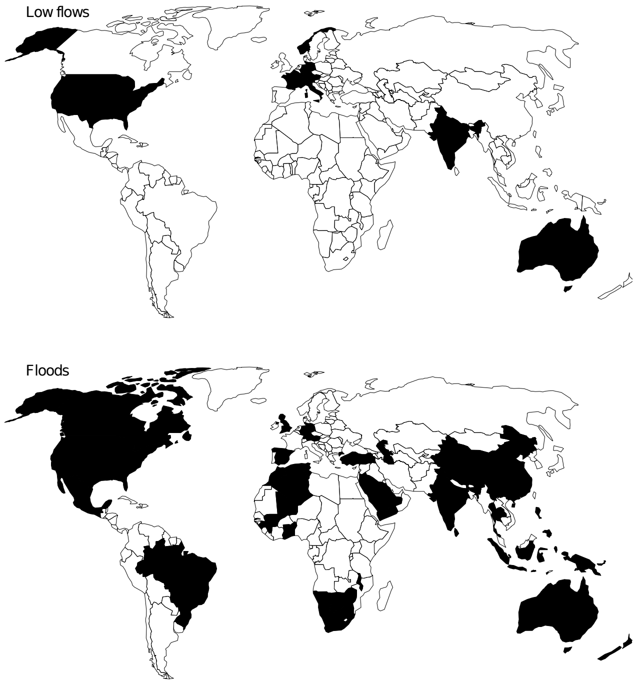


Fig. 1. Map indicating the countries included in the meta-analysis of low flow studies (top) and flood studies (down) reported in the literature (Level 1 assessment).

multiplicative regression models were used. A critical issue in the regression is the choice of the catchment/climate characteristics which include mean annual precipitation and geologic characteristics in the literature. It has been noted that it is important to interpret the catchment/climate characteristics that are found to be significant during a regression analysis from a hydrological perspective, i.e. to link the statistical analysis to the hydrological processes operating at the catchment scale.

– Regional regression (RR): here the procedure is similar, however the entire domain is subdivided into regions and a regression model is applied to each region separately. The main rationale of regional regression is that different processes may operate in the regressions so the catchment/climate characteristics will control low flows in different ways. A number of methods exist for identifying the regions or pooling groups, including cluster analysis of catchment/climate characteristics, residuals from a regression model and pattern identification on the basis of the seasonality of runoff.

– Geostatistical methods (G): geostatistical methods exploit the spatial correlations of low flows based on the rationale that catchments that are geographically close to each other may exhibit similar processes. While some approaches use Euclidean distance as a similarity measure, other approaches use the correlations along the river network. To account for spatially heterogeneous regions, the geostatistical method has been extended to combine it with multiple regressions by using the residuals of the regression for the spatial geostatistical estimation.

– Short records (SR): in some instances there may be short runoff records available for a catchment that is otherwise ungauged. These runoff records may not be representative of the longer time period that is normally used for the estimation of low flows. Methods are therefore used that relate the low flow estimates from the short runoff records to the longer hydrological history of the basin on the basis of regional information, usually involving some element of correlation analysis (Laaha and Blöschl, 2005).

Table 2. Number of studies (in brackets number of results) and number of catchments used. Level 1 refers to an assessment of the average performance of studies, Level 2 to an assessment of the performance for individual catchments.

	Level 1		Level 2	
	No. of studies	No. of catchments	No. of studies	No. of catchments
Low flows	14 (28)	3112	6	1895
Floods	20 (57)	3023	5	1422

3.2 Flood studies

Table 3 lists the 20 flood prediction studies used in this paper. It includes summary information about the study region, regionalisation method applied and the predictive performance in terms of the root mean square normalised error (RMSNE), defined as follows:

$$\mathrm{RMSNE} = \sqrt{\frac{1}{n} \cdot \sum \left(\frac{Q_{i,\mathrm{pred}} - Q_{i,\mathrm{obs}}}{Q_{i,\mathrm{obs}}} \right)^2}. \qquad (3)$$

The cross-validation performance is given, in the great majority of the papers considered, as the defined root mean squared normalised error, a very common error measure for estimators, combining both the bias and the dispersion component of the error. The target flood index, on which this performance was mainly reported, was the 100 yr specific flood quantile q_{100}, i.e. the peak discharge value that occurs on average every 100 yr divided by the catchment area. There are three exceptions, namely Srinivas et al. (2008), Cunderlik and Burn (2002), Jingyi and Hall (2004), where the predictive performance is calculated on volumes and not on specific discharges (Table 3). These studies are plotted as crosses in Figs. 2–4. It is worth mentioning, that the quantities defined as observed discharges $Q_{i,\mathrm{obs}}$, are actually the flood quantiles estimated from local data and are subject to a certain

Table 3. Summary assessment of studies for flood estimation in ungauged catchments used in Level 1 assessment. Error measure indicates the leave-one-out assessment of model efficiency in terms of the root mean square normalised error RMSNE. Flood regionalisation methods include: regression methods (R), index methods (IM) and geostatistics (G). Predicted variable indicates the flood discharge estimated in the study and includes: 100 yr specific flood runoff (q_{100}), 100 yr flood runoff (Q_{100}) and 100 yr flood runoff standardised by the mean annual flood (Q_{100}/Q_m). Ranges or various values for RMSNE represent variations of the methods or the same method applied on different subsamples from the same region.

Study	Region	Climate	Number of catchments	Regionalisation method	Predicted variable	Error measure (RMSNE)	Used in Level 2
Jimenez et al. (2012)	Spain	Arid	217	R	q_{100}	0.54	X
Walther et al. (2011)	Germany (Saxony)	Cold	170	G, IM	q_{100}	0.46, 0.49	X
Kjeldsen and Jones (2010)	United Kingdom	Humid	602	IM	q_{100}	0.51, 0.50	X
Guse et al. (2010)	Germany (Saxony)	Cold	90	R	q_{max}	0.81, 0.88	
Saf (2009)	Turkey	Arid	47	IM	Q_{100}/Q_m	0.43	
Chebana and Ouarda (2008)	Canada (southern Quebec)	Cold	151	R	q_{100}	0.44–0.45, 0.49, 0.64	
Srinivas et al. (2008)	USA (Indiana)	Cold	245	IM	q_{100}, Q_{100}	0.69, 0.27	X
Ouarda et al. (2008)	Mexico	Tropical	29	R, R, IM, IM, G, G	q_{100}	0.74, 0.66, 0.67, 0.67, 0.51, 0.52	
Leclerc and Ouarda (2007)	Canada, USA	Cold	29	R	q_{100}	0.61	
Ouarda et al. (2008)	Canada (southern Quebec)	Cold	63	IM	q_{100}	0.40	
Merz and Blöschl (2005)	Austria	Cold	575	G, R, IM	q_{100}	0.30, 0.46, 0.43	X
Jingyi and Hall (2004)	China (Gan-Ming River)	Humid	86	IM	Q_{20}, Q_{50}, Q_{100}, Q_{200}	0.31	
Chokmani and Ouarda (2004)	Canada (southern Quebec)	Cold	151	R	q_{100}	0.70, 0.51	
Cunderlik and Burn (2002)	United Kingdom	Humid	424	IM	Q_{100}/Q_m	0.29	
Javelle et al. (2002)	Canada (Quebec, Ontario)	Cold	158	IM	q_{100}	0.50	
Pandey and Nguyen (1999)	Canada (Quebec)	Cold	71	R	q_{100}	0.64, 0.81	
Madsen et al. (1997)	New Zealand (South Island)	Humid	48	IM	q_{100}	0.41, 0.39	
Meigh et al. (1997)	Brazil, Ivory Coast, Mali, Guinea, Ghana, Togo, Benin, Malawi, Namibia, Zimbabwe, South Africa and Botswana, Saudi Arabia, Iran, India	Tropic, Humid, Arid	59, 35, 86, 41, 16, 46, 28, 40, 234, 109, 28, 24, 75	IM	q_{100}	0.42, 0.47, 0.50, 0.53, 0.59, 0.42, 0.69, 0.63, 0.52, 0.69, 0.73, 0.65, 0.58	
GREHYS (1996)	Canada (Quebec, Ontario)	Cold	33	IM	q_{100}	0.45	
Farquharson et al. (1992)	Arid and semi-arid basins worldwide	Arid	162	IM	q_{100}	0.73	

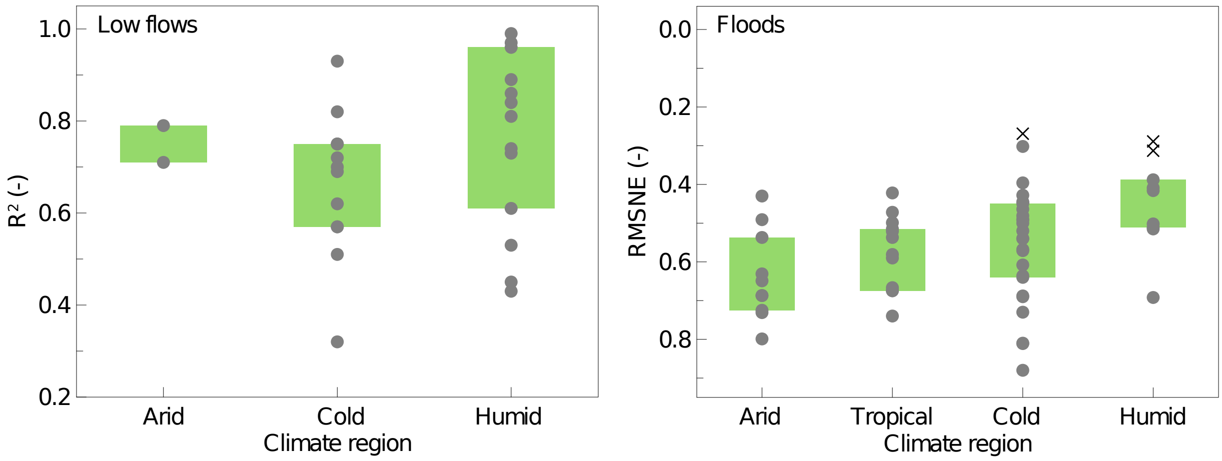


Fig. 2. Coefficient of determination of predicting low flows in ungauged basins (left) and root mean squared normalised error of predicting floods in ungauged basins (right), stratified by climate (Level 1 assessment). Each symbol refers to a result from the studies in Tables 1 and 3. Circles represent performances calculated on specific discharges ($m^3\,s^{-1}\,km^{-2}$), crosses represent performances calculated on discharges ($m^3\,s^{-1}$). Boxes show 25–75 % quantiles.

degree if uncertainty and the same applies to the observed 95 % low flow quantiles. Several studies compare different regionalisation approaches and/or subsets of data which results in a total of 57 assessments of predictive performance. These results are the base for the Level 1 assessment which represents at total of 3023 catchments (Table 2). Figure 1 (bottom) and Table 3 show that the studies are rather evenly spread around the world. Five study authors out of the Level 1 assessment provided detailed information about climate and catchment characteristics in a consistent way and reported the regionalisation performance for each catchment in terms of the absolute normalised error ANE (Level 2 assessment). This dataset combines data from 1422 catchments. As in the case of low flows, three catchment characteristics are analysed: aridity index, mean elevation and catchment area (see Sect. 3.1). The flood regionalisation methods have been classified into the following groups:

– Regression methods: the regression methods for flood discharges are similar to those of low flows where the flood runoff is related to catchment/climate characteristics such as catchment area and mean annual precipitation. As is the case of low flows, it is important to interpret the regression coefficients obtained from a hydrological perspective (Merz and Blöschl, 2008a, b).

– Index methods: the index methods consist of a group of approaches where the flood distribution function is scaled by the index flood (e.g. the mean annual flood or the median annual flood) and assumed to be homogenous within the region. One first estimates the index flood in the ungauged catchment (e.g. by a regression against catchment characteristics) and then multiplies that index flood with the regional-scaled flood distribution function. The methods usually differ in terms of how the homogeneous groups are obtained.

– Geostatistical methods: geostatistical methods are analogous to those in use for regionalising low flows (see Sect. 3.1).

4 Results and discussion

4.1 How good are the predictions of hydrological extremes in different climates?

Figure 2 (left) shows the Level 1 results of estimating low flows in ungauged basins. The distribution of the studies by climatic region is as follows: 2 are considered as arid, 12 as cold and 14 as humid. The highest performance is obtained for humid catchments, but there are also studies in humid climates that report a significantly lower performance. In arid climates, the performance is never very high, but more studies are needed to clearly show this behaviour. The most likely reason for this finding is that arid regions tend to be very heterogeneous with a high variability of low flow producing processes, and low flows generally tend to be lower and more variable, and therefore harder to predict. Cold environments exhibit the largest performance range. This could be because this class contains sub-polar and mountainous environments which may be hydrologically very complex with many different storage types that complicate low flow behaviours (ice/groundwater).

The results for the flood regionalisation (Fig. 2, right) present 10 studies from arid regions, 12 from tropical, 26 from cold and 9 from humid regions. They show that the predictions in humid regions exhibit the smallest errors and arid regions have the largest errors. This means that the predictive

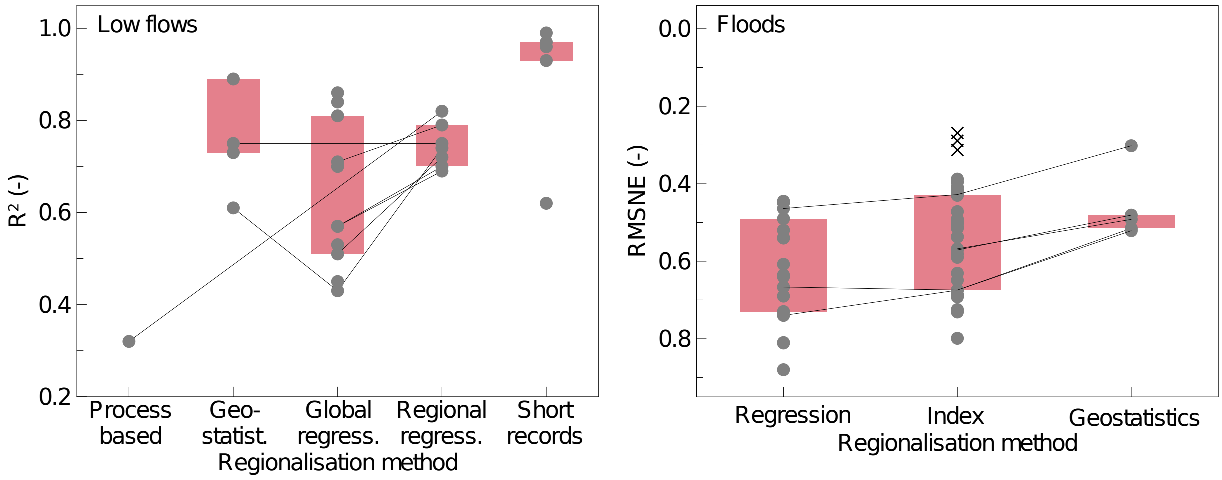


Fig. 3. Coefficient of determination of predicting low flows in ungauged basins (left) and root mean squared normalised error of predicting floods in ungauged basins (right), stratified by regionalisation method (Level 1 assessment). Each symbol refers to a result from the studies in Tables 1 and 3. Circles represent performances calculated on specific discharges ($m^3\,s^{-1}\,km^{-2}$), crosses represent performances calculated on discharges ($m^3\,s^{-1}$). Lines indicate studies that compared different methods for the same set of catchments. Boxes show 25–75 % quantiles.

performance clearly decreases with increasing aridity. There are a number of factors that may contribute to this dependence. The interannual variability (e.g. in terms of coefficient of variation of the annual peak runoff time series) of floods in arid regions is usually bigger than in other climates, due to the associated stronger non-linearities and threshold effects in drier regions and the the larger interannual variability and skewness of rainfall intensities more typical for arid climates. This means that floods are more difficult to estimate from short records. The stronger non-linearity also implies that the spatial hydrological variability in the flood producing processes will impact more strongly the flood frequency curve, so catchments that are close to each other may exhibit quite different flood frequency curves, which reflects poorly on the regionalised predictions. A possible explanation for this non-linearity in arid catchments is given in Goodrich et al. (1997), where the increasingly non-linear response is attributed to the increasing importance of ephemeral channel losses and partial storm area coverage. In contrast, humid catchments tend to be more linear, so the predictability is larger. The biggest range of performances is found in cold climates. This may be partly related to the larger number of studies available for these regions. Also, in cold regions a wide variety of flood producing processes may exist, including snow and rain-on-snow which may lead to different performance, depending on the prevailing processes. For example, snow melt floods tend to be more predictable than rain-on-snow floods (e.g. Sui and Koehler, 2001).

4.2 Which regionalisation method performs best?

The low flow regionalisation methods represented in the assessment included 1 result from the process-based methods group (continuous runoff models); 4 results from the geostatistical group of methods where runoff at the target site was estimated as a weighted mean of runoff at the surrounding gauges; 11 global regression and 7 regional regression results from the regression methods group; and 5 results from the short records group that used various methods. The assessments in each group are not based on exactly the same regionalisation approach, but the methodology is similar. There are also differences in the low flow indices used. They include q_{95} (95 % exceedance probability specific runoff), $q_{7,10}$ (7 days 10 yr specific runoff), and $q_{\mathrm{mon},5}$ (monthly 5 day minimum), all standardised by catchment area or mean flow, and the dimensionless BFI. In particular q_{95} low flows are usually closely correlated to $q_{7,10}$ so that a comparison across the various indices should provide consistent results at the level of detail used for the comparisons. Figure 3 (left) shows a large performance range across the regionalisation methods. Overall, it is clear that low flow predictions from short records ($R^2 = 0.62$ to 0.99) perform best. The method performs significantly better than all other methods, provided continuous runoff measurements from at least 3–5 yr of observations at the site of interest are used. A lower performance (0.62) is obtained when using a single flow measurement during the low flow period. The performance of global regression ranges from 0.43 to 0.86. Studies from high-mountain environments have a lower performance (Austria: 0.57, Switzerland: 0.51, Nepal: 0.53, India: 0.45) perhaps because the heterogeneity of the low flow process in the landscape (including snow) pose difficulties for applying one single regionalisation model for the entire domain, so division into subregions may be necessary. Global regression is better suited to smaller regions (e.g. the German region of

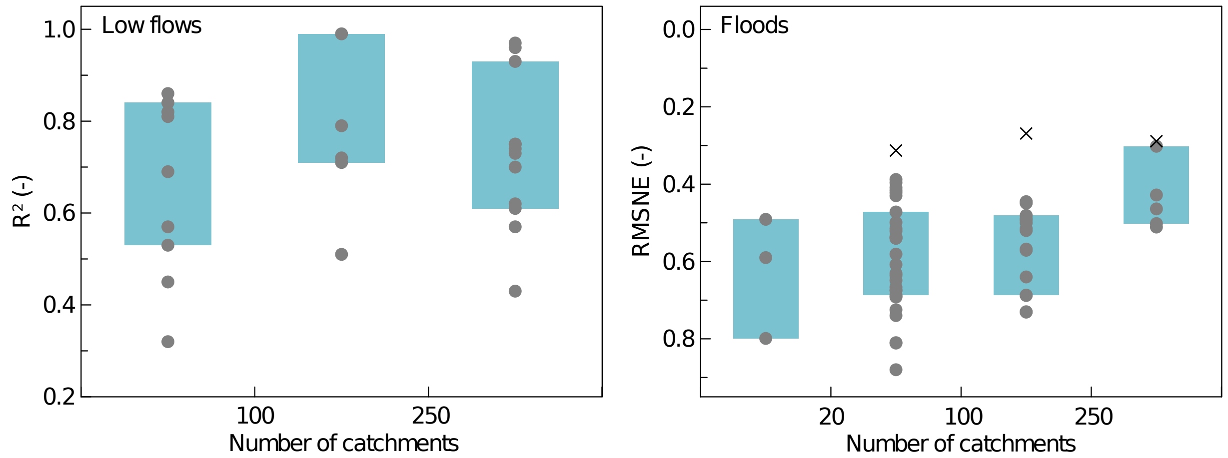


Fig. 4. Coefficient of determination of predicting low flows in ungauged basins (left) and root mean squared normalised error of predicting floods in ungauged basins (right), stratified by the number of catchments within each study (Level 1 assessment). Each symbol refers to a result from the studies in Tables 1 and 3. Circles represent performances calculated on specific discharges ($m^3\,s^{-1}\,km^{-2}$), crosses represent performances calculated on discharges ($m^3\,s^{-1}$). Boxes show 25–75 % quantiles.

Baden-Württemberg) and studies in climates less controlled by snow seasonality (e.g. New South Wales and Victoria in Australia). The four results from geostatistical models give on average the highest performances between 0.61 and 0.89. A continuous runoff model, tested in only one study used in the meta-analysis, gave lower performance than the statistical methods. The studies examined differ in terms of the hydrological characteristics and data availability, so a comparison of methods for different regions will involve some uncertainty. It is therefore useful to apply each different method to the same catchment. A number of studies are available in the literature that have performed such a comparison and the results are indicated as grey lines in Fig. 3 (left). Most of the studies compare global and regional regressions. The comparisons clearly show that the regional regressions always perform better than the global regressions. The studies that conduct this comparison show that the average performance of global regressions is around 0.5 and increases to 0.7 for regional regression. It should be noted that the performance reported is cross-validation performance for ungauged basins, so better performance is related to better predictions rather than to improved goodness of fit of the regressions. There are also a few studies that compared geostatistical methods with regional regression methods. In one study from France (Plasse and Sauquet, 2010) the geostatistical method was based on distance between the catchment centres of gravity. The performance was larger than for global regression and lower than that of regional regression. If the stream network structure is taken into account, the performance of geostatistical methods can in fact be higher than that of regional regression as illustrated in the Austrian case studies (Laaha et al., 2007, 2012). Finally, one study (Engeland and Hisdal, 2009) compared process-based methods with regional regressions and found that the regressions gave better results. Clearly, application of process-based methods does not *per se* include the performance of low flow estimation but their value depends on the amount of information available for careful parameterisation of the model. However, process-based methods have more potential to explore the impact of environmental change than statistical methods.

The flood regionalisation methods represented in the assessment included (i) regression methods, 18 results from different regression models where the flood quantiles or the distribution parameters had been transferred to ungauged basins; (ii) index methods, 34 results where a regional growth curve had been defined for homogeneous regions; (iii) geostatistical methods, 5 results where runoff at the target site was estimated as a weighted mean of runoff at the surrounding gauges. While the assessments made by each group are not based on exactly the same regionalisation approach, the methodology is similar. Figure 3 (right) shows that the geostatistical methods perform best (RMSNE of 0.30–0.52) across the studies analysed, although the number of studies is small compared to the other groups. For example, Merz and Blöschl (2005) in Austria and Walther et al. (2011) in Saxony (Germany), provide the combination of the necessary stream network density and non-arid climate that causes their lower RMSNE values (0.30 and 0.46 respectively). The regression methods have the lowest performance, i.e. the largest predictive errors (median RMSNE of 0.62), and the index methods fall in between. As an illustrative example, we find the low performances (average RMSNE of 0.57) of the index flood method in the arid and semi-arid regions of Meigh et al. (1997) and even lower (RMSNE between 0.81–0.88) for the regression approaches in a cold climate in Guse et al. (2010). The result of the overall ranking of methods is confirmed by studies that compared different approaches in the same region (grey lines in Fig. 3, right). It appears that it

Table 4. Methods with the highest and lowest cross-validation performance of runoff predictions in ungauged basins. Arid relates to catchments with an aridity index > 1. Level 1 refers to an assessment of the average performance of studies, Level 2 to an assessment of the performance for individual catchments. For the number of studies and catchments see Table 2.

	Level 1		Level 2	
	Highest cross-validation performance	Lowest cross- validation performance	Highest cross-validation performance	Lowest cross-validation performance
Low flows	Short records, Geostatistics	Global regressions	Short records, Geostatistics (arid)	Global regressions
Floods	Geostatistics, Index methods	Regression methods	Geostatistics	Index methods (arid), Regression methods

may be difficult to find catchment characteristics that are representative of the flood generating processes. For example, subsurface characteristics are an important control for flood generation and these are difficult to capture unless detailed field surveys are available. Index methods and geostatistical methods are less dependent on the catchment characteristics as they usually take advantage of both spatial proximity (either through spatial correlations or homogeneous regions) and correlations to catchment characteristics. It is also the case that the geostatistical studies in Table 3 have been performed in data-rich environments, which may partly explain their better performance. It is interesting to note that the number of studies applying regression and index methods is much larger than those applying geostatistical methods, which is because they have a longer tradition in hydrology. The first two columns in Table 4 present a summary of the methods with the highest and lowest predictive performances in the Level 1 assessments of low flow and flood studies.

4.3 How does data availability impact performance?

While the information on the data used was never very detailed in the studies examined, some inferences on data availability can be drawn from the number of catchments used in the studies. These are usually those catchments used both for the cross validation and for regionalising runoff to neighbouring catchments. Figure 4 (left) shows the predictive performance (R^2) for the case of low flows as a function of the number of catchments analysed in each study. It is clear that the studies with less than 100 catchments have, on average, the lowest performance and performance increases with the number of catchments used in analysis. Possibly, this is due to the lower stream gauge density in studies with a smaller number of stream gauges, but more detailed analyses on the precise geographic extent of the studies would be needed to ascertain the data controls on performance. The performance decreases for very large datasets (> 250 catchments). This decrease is related to the higher heterogeneity of larger study areas and to the fact that a number of the studies used global regression methods that did not perform very well in these regions.

Figure 4 (right) shows the RMSNE for the case of floods as a function of the number of catchments analysed in each study. The errors clearly decrease and the performance increases with the number of catchments included in the analysis. This is possibly because of the higher stream gauge density in the larger studies with a bigger number of stations involved, which makes the transfer of floods across the landscape more accurate, in particular if there is a stream gauge upstream or downstream of the target site. Also, the regionalisation methods may be more robust if the total number of stations is larger.

4.4 To what extent does runoff prediction performance depend on climate and catchment characteristics?

The assessment of the predictive performance of the low flow regionalisation methods with respect to three climate and catchment characteristics (Level 2 assessment) is presented in Fig. 5. The lines indicate the median runoff prediction performance of catchments belonging to the same study. Overall, the absolute normalised error (ANE, see Sect. 3.1), clearly increase with increasing aridity. This means that the performance is consistently lower in drier, and more arid environments. These are regions that tend to be particularly heterogeneous, with a high presence of intermittent rivers (Jacobson and Jacobson, 2013) and where low flows may be small, which makes them particularly hard to predict.

Figure 5 also indicates that there is a tendency for performance to increase with catchment elevation. The average of all methods shows that errors decrease from 0.37 for lowland catchments (mean elevation < 200 m a.s.l.) to 0.16 for high mountain catchments. This may be partially due to the higher specific discharges of mountainous catchments compared to lowland catchments which may increase predictability. Also, in the high mountains, low flows may be of a winter low flow type, so low flows may depend on frost strength which is closely related to catchment elevation. The bottom panels in the figure show the performance as a function of catchment scale. For all methods the performance increases with catchment scale. This may be related to both data availability and space–time aggregation of runoff processes in the

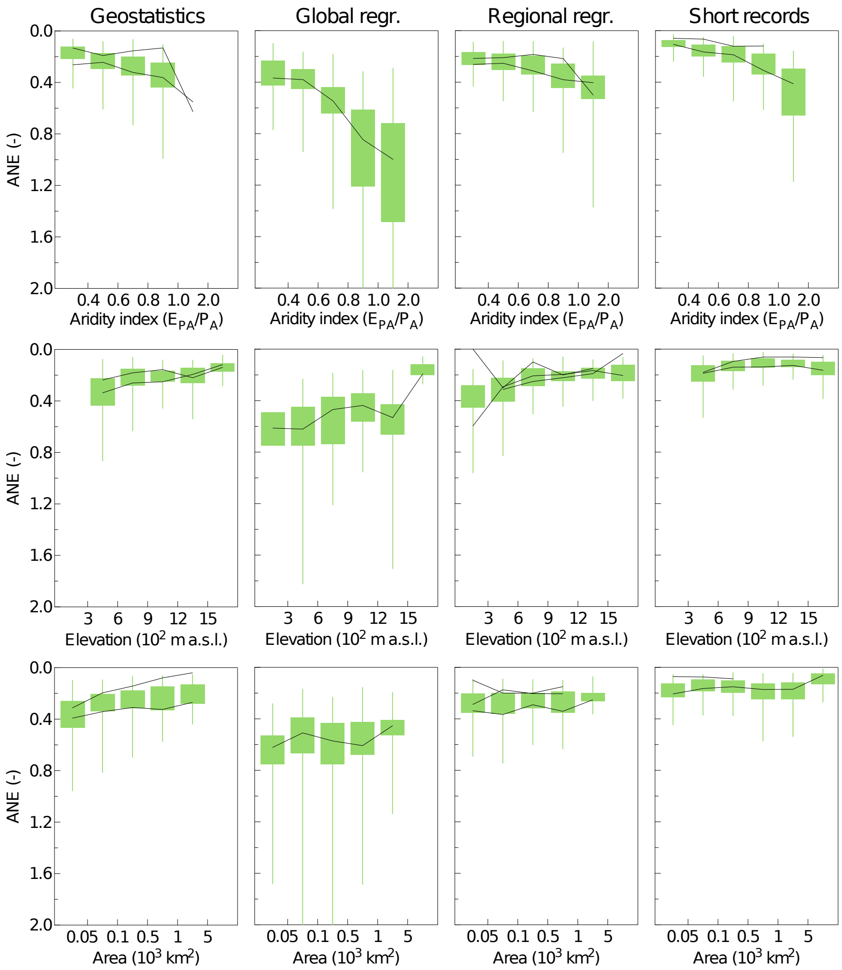


Fig. 5. Absolute normalised error of predicting q_{95} low flows ($m^3\,s^{-1}\,km^{-2}$) in ungauged basins as a function of aridity (E_{PA}/P_A), mean elevation and catchment area for different regionalisation methods (Level 2 assessment). Lines connect median errors for the same study. Boxes are 40–60 % quantiles, whiskers are 20–80 % quantiles.

catchments, which will increase the predictability. The exceptions are methods that use short runoff records at the site of interest. In these cases, the performance dependence on catchment size is less pronounced than for the other methods. These types of methods may be more dependent on the representativeness of the short runoff record to the temporal variability of low flows, so the dependence on the spatial variability and therefore catchment size may be lower.

The left panels in Fig. 6 summarise the performance for different regionalisation approaches, stratified by the aridity index. The left-top, left-middle and left-bottom panels show the performance for all catchments, catchments with an aridity index below and catchments with an aridity index above 1, respectively. Overall, for all catchments the performance of the global regression is much lower than that of any other method. This is consistent with the Level 1 assessment. In the arid catchments the performance of the global regression is particularly low and the absolute normalised errors are on average around 1.1. In the humid regions the short records perform better than any other method. This is, again, consistent with the Level 1 assessment. However, this is no longer the case for the arid catchments. For the arid catchments, the performance of the short records is in fact lower than those of the geostatistical methods and regional regression. It appears that, in arid regions, the variability of the low flows between years may be larger than in other climates, what makes the method more dependent on an appropriate donor site. The appropriateness of a donor depends on gauging density which is often lower in the more arid countries. Methods may be needed in arid regions that specifically account for the runoff generation processes in the region, and preferably are based on proxy data that account for these processes.

The Level 2 assessment for flood prediction studies, i.e. the assessment of the ANE measure with respect to the three climate and catchment characteristics is presented in Fig. 7. The lines indicate again the median runoff prediction performance of catchments belonging to the same study. The top panel shows that the errors clearly increase with

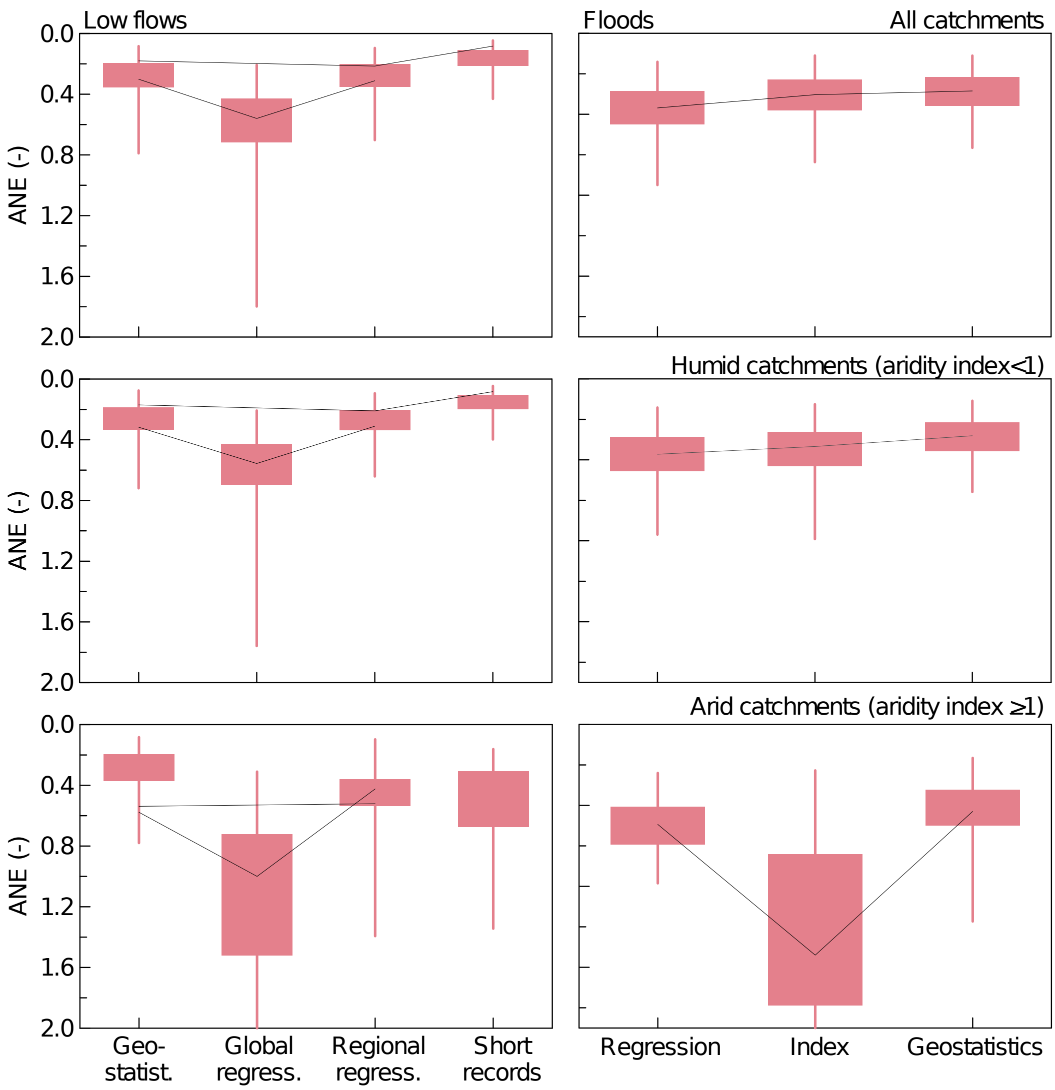


Fig. 6. Absolute normalised error of predicting q_{95} low flows ($m^3\,s^{-1}\,km^{-2}$), left panels, and q_{100} floods ($m^3\,s^{-1}\,km^{-2}$), right panels, in ungauged basins for different regionalisation methods, stratified by aridity (Level 2 assessment). Top: all catchments. Centre: humid catchments (aridity index < 1). Bottom: arid catchments (aridity index ≥ 1). Lines connect median efficiencies for the same study. Boxes are 40–60 % quantiles, whiskers are 20–80 % quantiles.

increasing aridity, i.e. there is a decrease in performance with aridity for all three methods. This is also supported by the lines representing comparative studies. This clear trend is in line with the Level 1 assessment for floods, but also with both assessment levels for low flows. Arid regions tend to be more heterogeneous than humid regions and runoff processes are more non-linear, which makes the predictions for both floods and low flows more difficult. There is a slight increase in performance with elevation but, in contrast to aridity, the errors do not change much with elevation. In the studies examined here, the highest elevation catchments are influenced by snowmelt, so there is a tendency for the flood predictions to improve if snow melt is involved in the flood generation processes.

The results stratified by catchment area (Fig. 7, bottom panels) indicate a clear increase in performance (decrease of ANE) with increasing catchment area for all methods. The increasing performance with catchment size is likely related to two factors. The first is related to the data availability. As the catchment size increases the likelihood that gauged subcatchments are available as donor stations increases. This will lead to more reliable transfer of the flood characteristics. Additionally, for larger catchments, there are aggregation effects on the flood generating processes, so floods tend to be less flashy and therefore easier to predict.

The right panels in Fig. 6 summarise the runoff prediction performance of different regionalisation approaches, stratified by the aridity index. Again, the right-top, right-middle

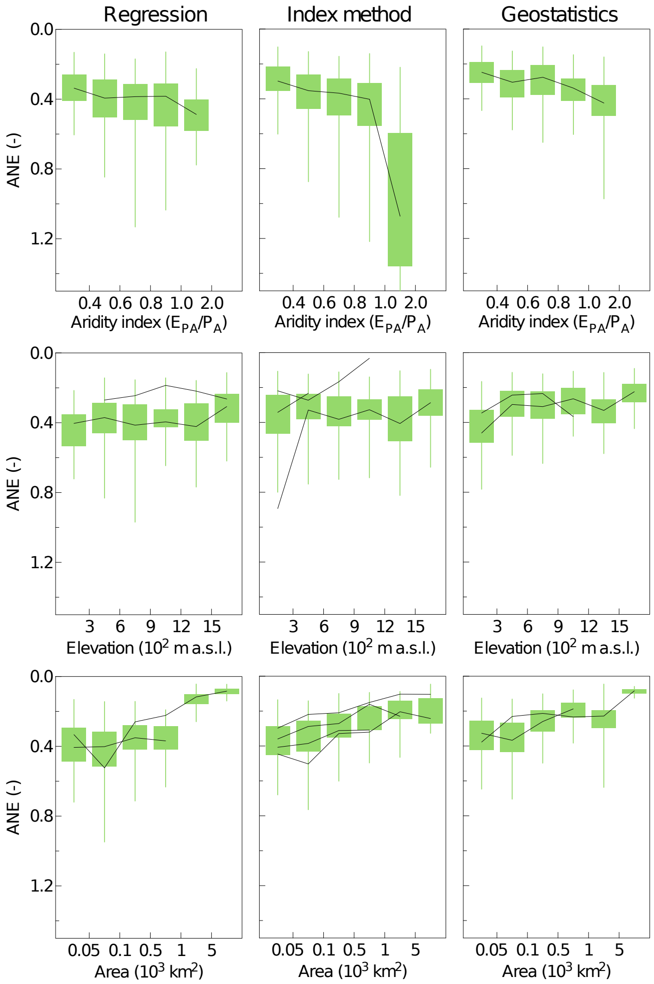


Fig. 7. Absolute normalised error of predicting q_{100} floods ($\mathrm{m}^3\,\mathrm{s}^{-1}\,\mathrm{km}^{-2}$) in ungauged basins as a function of aridity ($E_{\mathrm{PA}}/P_{\mathrm{A}}$), mean elevation and catchment area for different regionalisation methods (Level 2 assessment). Lines connect median errors for the same study. Boxes are 40–60 % quantiles, whiskers are 20–80 % quantiles.

and right-bottom panels show the performance for all catchments; catchments with an aridity index below and catchments with an aridity index above 1, respectively. Analysis of the overall performance of the three methods shows that performance is similar for geostatistical and index methods, which have a slightly better performance than the regression methods. For humid catchments, again, the performance of geostatistical methods is slightly better than index methods, and the performance of the regression methods is slightly lower. For dry catchments, however, the index methods performs significantly worse than the other two methods. The low performance of the index flood method in arid regions may be related to the underlying assumption of using the same non-dimensional flood frequency curve (i.e. growth curve) in the entire regions. Arid regions may be spatially more heterogeneous, leading to lower performance. More importantly, most arid catchments have the larger errors for the index methods, as the result of the prediction overestimate on the 100 yr floods (Fig. 7, top centre). The median absolute normalised error is 1.0, and the errors were in the vast majority positive (presented in Blöschl et al., 2013), indicating that typically the methods predict around twice the floods actually observed. If a homogeneous region contains both arid catchments with relatively lower floods and wetter catchments with higher floods, the homogeneity assumption will tend to lead to an overestimation in those catchments with the lower floods. The last two columns in Table 4 present a summary of the methods with the highest and lowest predictive performances in the Level 2 assessments of low flows and floods.

5 Conclusions

This paper has compared the performance of predicting low flow and flood discharges in ungauged basins using different regionalisation methods. Two kinds of assessments were performed; a Level 1 assessment which constitutes a meta-analysis from the literature; and a Level 2 assessment which analyses individual catchments in more detail. The results indicate that the Level 1 and Level 2 assessments are consistent while shedding light on different aspects of the prediction problem. The assessment of flood and low flow estimation methods in this paper represents the largest existing meta-analysis of regionalisation studies of hydrological extremes. However, it is clear that the analysis cannot cover all facets of hydrological variability worldwide. Arid and tropical climates are missing in the case of low flows. Arid climates are especially prone to droughts, so it would be of worth to pursue more detailed research on assessing predictions of extreme low flows in these areas. Also, some of the methods, e.g. process-based methods, are under-represented in the literature and a more detailed analysis of these would be of interest. For the flood regionalisation studies, the coverage of climates is more uniform, but there is a clear dominance of the regression-type and index-flood methods over geostatistical approaches. The increasing trend in the application of the latter group of methods is likely to lead to a sizeable sample of studies in the literature which will allow more comprehensive tests of their performance in the near future.

The Level 1 analysis suggests that in humid regions the performance of predicting both low flows and floods in ungauged basins tends to be better than in other climates. For the case of floods the performance tends to be lowest in arid regions. For the case of low flows, geostatistical methods can perform better than regional regressions in regions with medium to high stream gauge density if the stream network structure is taken into account. Regional regressions that divide a domain into subregions and apply regression models separately always perform much better than global regressions. For the case of floods, geostatistical methods tend to perform better than the other methods, regressions tend to have the lowest performance, and index methods lie between geostatistic and regression methods. This suggests that it may be difficult to find catchment characteristics that are suitable for regression methods, both for low flows and floods. Again, for both low flows and floods the performance tends to increase with number of stations in a region highlighting the value of stream gauge data in the region of interest, even for the case of ungauged basins.

The results of the more detailed analysis (Level 2) are mostly consistent with those of the meta-analysis from the literature (Level 1). For the case of low flows the predictive performance tends to decrease with increasing aridity (both Level 1 and Level 2 assessments). The performance improves with increasing catchment area (Level 2 assessment), apparently because of the presence of longer water flow pathways that accompany increasing catchment size. The availability of short records is particularly useful to improve performance of low flow predictions (both Levels 1 and 2), especially in humid regions, and are perhaps not as useful in arid regions because of the strong interannual variability together with the usually low stream gauge density in arid regions (Level 2). Of the various methods, regional regressions have been shown to be better than global regressions (from Level 1 and Level 2 assessments). For the case of floods, the predictive performance also tends to decrease with increasing aridity (both Level 1 and Level 2 assessments). As expected, predictive performance increases with increasing catchment area (Level 2 assessment). Both Level 1 and Level 2 assessments indicated that the geostatistical methods have the best performance (especially when data availability is high), index methods work next best, and regression methods have the relatively lowest performance. In arid conditions the index methods are significantly biased and significantly overestimate the 100 yr floods in the catchments analysed. The Level 2 assessment also indicated that index methods do not work well in arid regions. Arid regions would therefore need more gauges to capture the temporal and spatial variability, but achieving this is unrealistic in many arid parts of the world where (due to economic reasons) data density is typically lower than in humid regions. Methods that are able to exploit the specifics of the region would be needed here. Use of readily available landscape information, such as erosional patterns, based on the idea of reading the landscape, may assist in improving the predictions of runoff extremes. More research on arid hydrology is urgently needed. Scale, uncertainty, and choice of proxy data are likely important considerations in this body of research (e.g. Blöschl, 2006; Koutsoyiannis et al., 2009).

The meta-analysis of the literature highlighted that the results on predictive performance of low flows and floods are presented in widely diverse ways, using different performance measures, different ways of aggregating the information of the regions of interest, and different levels of details on the hydrological characteristics of the regions. It appears that, to make the results more useful to the hydrological community, it would be essential to adjust the reporting of results and make them more comparable. This would assist in generalising the findings from individual case studies. We need techniques to exploit information from individual catchment studies, as well as the compilation of all studies from around the world. As a community collectively we need to go beyond that, and find systematic ways to generate knowledge, in terms of the patterns that connect across the multitude of studies and thereby provide a higher level of predictability as to what will happen next and understanding that will enable extrapolation to new situations. This points to the importance of hydrological synthesis as a vehicle for creating these connections.

Acknowledgements. We are very grateful to K. Eng, E. Sauquet, H. Hisdal, L. Mediero, R. Merz, T. Kjeldsen and V. V. Srinivas for providing the hydrologic data and results of their studies. Without their very useful support the Level 2 analysis of this study would not have been possible. We would also like to thank the Austria Science Funds (FWF) as part of the Vienna Doctoral Programme on Water Resource Systems (DK-plus W1219-N22), the FWF Project P23723-N21, the Innovative Ideas project of the TU Wien, and the Austrian Climate and Energy Fund (Project Nr K10AC0K00003, CILFAD) for financial support.

Edited by: H. H. G. Savenije

References

Aschwanden, H. and Kan, C.: Die Abflussmenge Q347, Eine Standortbestimmung, Hydrologische Mitteilungen Communications hydrologiques, Nr. 27, Le débit Landeshydrologie und geologie, Bern, 1999.

Beable, M. E. and McKerchar, A. I.: Regional flood estimation in New Zealand, Technical Report No. 20, National Water and Soil Conservation Organisation, Water and Soil Division, Christchurch, NZ, 132 pp., 1982.

Blöschl, G., Grayson, R. B., and Sivapalan, M.: On the representative elementary area (REA) concept and its utility for distributed rainfall-runoff modelling, Hydrolog. Process., 9, 313–330, 1995.

Blöschl, G.: Hydrologic synthesis: Across processes, places, and scales, Water Resour. Res., 42, W03S02, doi:10.1029/2005WR004319, 2006.

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H. (Eds.): Runoff Prediction in Ungauged Basins – Synthesis across Processes, Places and Scales, Cambridge University Press, Cambridge, United Kingdom, 2013.

Burn, D. H.: An appraisal of the "region of influence" approach to flood frequency analysis, Hydrolog. Sci. J., 35, 149–165, doi:10.1080/02626669009492415, 1990.

Castiglioni, S., Castellarin, A., Montanari, A., Skøien, J. O., Laaha, G., and Blöschl, G.: Smooth regional estimation of low-flow indices: physiographical space based interpolation and top-kriging, Hydrol. Earth Syst. Sci., 15, 715–727, doi:10.5194/hess-15-715-2011, 2011.

Chebana, F. and Ouarda, T. B. M. J.: Depth and homogeneity in regional flood frequency analysis, Water Resour. Res., 44, 879–887, doi:10.1029/WR024i006p00879, 2008.

Chokmani, K. and Ouarda, T. B. M. J.: Physiographical space-based kriging for regional flood frequency estimation at ungauged sites, Water Resour. Res., 40, 1–13, doi:10.1029/2003WR002983, 2004.

Clausen, B. and Pearson, C. P.: Regional frequency analysis of annual maximum streamflow drought, J. Hydrol., 173, 111–130, doi:10.1016/0022-1694(95)02713-Y, 1995.

Cunderlik, J. M. and Burn, D. H.: Analysis of the linkage between rain and flood regime and its application to regional flood frequency estimation, J. Hydrol., 261, 115–131, 2002.

Cunnane, C.: Methods and merits of regional flood frequency analysis, J. Hydrol., 100, 269–290, 1988.

Dalrymple, T.: Flood frequency analysis, Water Supply Paper 1543A, US Geological Survey, Reston, Virginia, 1960.

Demuth, S.: Untersuchungen zum Niedrigwasser in West-Europa (European low flow study). Freiburger Schriften zur Hydrologie, Band 1, Freiburg, Germany, p. 205, 1993.

Demuth, S. and Hagemann, I.: Estimation of flow parameters applying hydrogeological area information, in: FRIEND: Flow Regimes from International Experimental and Network Data, IAHS Publ. No. 221, 151–157, 1994.

Eng, K., Kiang, J. E., Chen, Y., Carlisle, D. M., and Granato, G. E.: Causes of systematic over- or underestimation of low streamflows by use of index-streamgage approaches in the United States, Hydrol. Process., 25, 2211–2220, doi:10.1002/hyp.7976, 2011.

Engeland, K. and Hisdal, H.: A Comparison of Low Flow Estimates in Ungauged Catchments Using Regional Regression and the HBV-Model, Water Resour. Manage., 23, 2567–2586, doi:10.1007/s11269-008-9397-7, 2009.

Farquharson, F. A. K., Meigh, J. R., and Sutcliffe, J. V.: Regional flood frequency analysis in arid and semi-arid areas, J. Hydrol., 138, 487–501, doi:10.1016/0022-1694(92)90132-F, 1992.

Goodrich, D. C., Lane, L. J., Shillito, R. M., Miller, S. N., Syed, K. H., and Woolhiser, D. A.: Linearity of basin response as a function of scale in a semiarid watershed, Water Resour. Res., 33, 2951–2965, 1997.

GREHYS (Groupe de Recherche en Hydrologie Statistique): Intercomparison of flood frequency procedures for Canadian rivers, J. Hydrol., 186, 85–103, 1996.

Griffis, V. W. and Stedinger, J. R.: Log-Pearson Type 3 distribution and its application in Flood Frequency Analysis. 1: Distribution characteristics, J. Hydrol. Eng., 12, 482–491, doi:10.1061/(ASCE)1084-0699(2007)12:5(482), 2007.

Guse, B., Thieken, A. H., Castellarin, A., and Merz, B.: Deriving probabilistic regional envelope curves with two pooling methods, J. Hydrol., 380, 14–26, doi:10.1016/j.jhydrol.2009.10.010, 2010.

Gustard, A., Bullock, A., and Dixon, J. M.: Low Flow Estimation in the United Kingdom, Institute of Hydrology Report 108, Wallingford, UK, 1992.

Hayes, D. C.: Low flow Characteristics of Streams in Virginia, US Geological Survey, Water Supply Paper 2374, 1992.

IH: Flood Estimation Handbook, Institute of Hydrology (IH), Wallingford, UK, 1999.

Jacobson, P. J. and Jacobson, K. M.: Hydrologic controls of physical and ecological processes in Namib Desert ephemeral rivers: Implications for conservation and management, J. Arid Environ., 93, 80–93, doi:10.1016/j.jaridenv.2012.01.010, 2013.

Javelle, P., Ouarda, T. B. M. J., Lang, M., Bobée, B., Galéa, G., and Grésillon, J. M.: Development of regional flood-duration-frequency curves based on the index-flood method, J. Hydrol., 258, 249–259, 2002.

Jimenez, A., Garcia, C., Mediero L., Incio, L., and Garrote, J.: Map of maximum flows of intercommunity basins, Revista de Obras Publicas, 3533, 7–32, 2012.

Jingyi, Z. and Hall, M. J.: Regional flood frequency analysis for the Gan-Ming River basin in China, J. Hydrol., 296, 98–117, doi:10.1016/j.jhydrol.2004.03.018, 2004.

Kjeldsen, T. R. and Jones, D. A.: Predicting the index flood in ungauged UK catchments: on the link between data-transfer and spatial model error structure, J. Hydrol., 387, 1–9, 2010.

Koutsoyiannis, D., Montanari, A., Lins, H. F., and Cohn, T. A.: Climate, hydrology and freshwater: towards an interactive incorporation of hydrological experience into climate research, Hydrolog. Sci. J., 54, 394–405, 2009.

Laaha, G. and Blöschl, G.: Low flow estimates from short stream flow records – a comparison of methods, J. Hydrol., 306, 264–286, doi:10.1016/j.jhydrol.2004.09.012, 2005.

Laaha, G. and Blöschl, G.: A comparison of low flow regionalisation methods—catchment grouping, J. Hydrol., 323, 193–214, doi:10.1016/j.jhydrol.2005.09.001, 2006a.

Laaha, G. and Blöschl , G.: Seasonality indices for regionalizing low flows, Hydrolog. Process., 20, 3851–3878, doi:10.1002/hyp.6161, 2006b.

Laaha, G. and Bloschl, G.: A national low flow estimation procedure for Austria, Hydrolog. Sci. J., 52, 625–644, doi:10.1623/hysj.52.4.625, 2007.

Laaha, G., Skøien, J., and Bloschl, G.: A comparison of top-kriging and regional regression for low flow regionalisation, Geophys. Res. Abstr., 9, 07015, 2007.

Laaha, G., Skøien, J. O., and Blöschl, G.: Spatial prediction on river networks: comparison of top-kriging with regional regression, Hydrol. Process., doi:10.1002/hyp.9578, in press, 2012.

Lebel, T., Bastin, G., Obled, C., and Creutin, J. D.: On the accuracy of areal rainfall estimation: a case study (Cevennes, France), Water Resour. Res., 23, 2123–2134, 1987.

Leclerc, M., and Ouarda, T. B. M. J.: Non-stationary regional flood frequency analysis at ungauged sites, J. Hydrol., 343, 254–265, doi:10.1016/j.jhydrol.2007.06.021, 2007.

Madsen, H. and Rosbjerg, D.: A regional Bayesian method for estimation of extreme streamflow droughts, in: Statistical and Bayesian Methods in Hydrological Sciences, edited by: Parent, E., Bobee, B., Hubert, P., and Miquel, J., UNESCO, Paris, PHI Series, 327–340, 1998.

Madsen, H., Rasmussen, P. F., and Rosbjerg, D.: Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling, Water Resour. Res., 33, 747, doi:10.1029/96WR03848, 1997.

Meigh, J., Farquharson, F., and Sutcliffe, J.: A worldwide comparison of regional flood estimation methods and climate, Hydrolog. Sci. J., 42, 225–244, 1997.

Merz, R. and Blöschl, G.: Flood frequency regionalisation-spatial proximity vs. catchment attributes, J. Hydrol., 302, 283–306, doi:10.1016/j.jhydrol.2004.07.018, 2005.

Merz, R. and Bloschl, G.: Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, Water Resour. Res., 44, W08432, doi:10.1029/2007WR006744, 2008a.

Merz, R. and Bloschl, G.: Flood frequency hydrology: 2. Combining data evidence, Water Resour. Res., 44, W08433, doi:10.1029/2007WR006745, 2008b.

Moretti, G. and Montanari, A.: Inferring the flood frequency distribution for an ungauged basin using a spatially distributed rainfall-runoff model, Hydrol. Earth Syst. Sci., 12, 1141–1152, doi:10.5194/hess-12-1141-2008, 2008.

Nathan, R. J. and McMahon, T. A.: Identification of homogeneous regions for the purposes of regionalisation, J. Hydrol., 121, 217–238, 1990.

Nathan, R. J. and McMahon, T. A.: Estimating low flow characteristics in ungauged catchments, Water Resour. Manage., 6, 85–100, doi:10.1007/BF00872205, 1992.

Ouarda, T. B. M. J., Cunderlik, J. M., St-Hilaire, A., Barbet, M., Bruneau, P., and Bobée, B.: Databased comparison of seasonality-based regional flood frequency methods, J. Hydrol., 330, 329–339, 2006.

Ouarda, T. B. M. J., Bâ, K. M., Diaz-Delgado, C., Cârsteanu, A., Chokmani, K., Gingras, H., Quentin, E., Trujillo, E., and Bobée, B.: Intercomparison of regional flood frequency estimation methods at ungauged sites for a Mexican case study, J. Hydrol., 348, 40–58, doi:10.1016/j.jhydrol.2007.09.031, 2008.

Pandey, G. R. and Nguyen, V. T. V.: A comparative study of regression based methods in regional flood frequency analysis, J. Hydrol., 225, 92–101, doi:10.1016/S0022-1694(99)00135-3, 1999.

Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 1: Runoff-hydrograph studies, Hydrol. Earth Syst. Sci., 17, 1783–1795, doi:10.5194/hess-17-1783-2013, 2013.

Piock-Ellena, U., Merz, R., Blöschl G., and Gutknecht, D.: On the regionalization of flood frequencies – Catchment similarity based on seasonality measures. XXVIII IAHR Proceedings, CD-Rom, 434.htm, 1999.

Plasse, J. and Sauquet, E.: Interpolation des débits de référence d'étiage, Rapport d'étude, Cemagref, Mai 2010, p. 166, 2010 (in French).

Rees, G., Croker, K., Zaidman, M., Cole, G., Kansakar, S., Chalise, S., Kumar, A., Saraf, A., and Singhal, M.: Application of the regional flow estimation method in the Himalayan region, IAHS Publ. No. 274, 433-440, 2002.

Saf, B.: Regional Flood Frequency Analysis Using L-Moments for the West Mediterranean Region of Turkey, Water Resour. Manage., 23, 531–551, doi:10.1007/s11269-008-9287-z, 2009.

Skøien, J. O., Bloschl, G., and Western, A. W.: Characteristic space scales and timescales in hydrology, Water Resour. Res., 39, 1304, doi:10.1029/2002WR001736, 2003.

Skøien, J. O., Merz, R., and Blöschl, G.: Top-kriging – geostatistics on stream networks, Hydrol. Earth Syst. Sci., 10, 277–287, doi:10.5194/hess-10-277-2006, 2006.

Srinivas, V. V., Tripathi, S., Rao, A. R., and Govindaraju, R. S.: Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering, J. Hydrol., 348, 148–166, doi:10.1016/j.jhydrol.2007.09.046, 2008.

Sui, J. and Koehler, G.: Rain-on-snow induced flood events in Southern Germany, J. Hydrol., 252, 205–220, doi:10.1016/S0022-1694(01)00460-7, 2001.

Vezza, P., Comoglio, C., Rosso, M., and Viglione, A.: Low Flows Regionalization in North-Western Italy, Water Resour. Manage., 24, 4049–4074, doi:10.1007/s11269-010-9647-3, 2010.

Viglione, A., Parajka, J., Rogger, M., Salinas, J. L., Laaha, G., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 3: Runoff signatures in Austria, Hydrol. Earth Syst. Sci., 17, 2263–2279, doi:10.5194/hess-17-2263-2013, 2013.

Walther, J., Merz, R., Laaha, G., and Büttner, U.: Regionalising floods in Saxonia Regionalisierung von Hochwasserabflüssen in Sachsen, in: Hydrologie & Wasserwirtschaft – von der Theorie zur Praxis, edited by: Blöschl, G. und Merz, R., Forum für Hydrologie und Wasserbewirtschaftung, Heft 30.11, 29–35, ISBN:978-3-941897-7, 2011 (in German).

Wandle, S. W.: Estimating the Magnitude and Frequency of Flood on Natural Streams in Massachusetts, US Geological Survey Water Resources Investigations Report, 77–39, 1977.

Zhang, Z. and Kroll, C. N.: A Closer Look at Base flow Correlation, J. Hydrol. Eng., ASCE, 12, 190–196, 2007a.

Zhang, Z. and Kroll, C. N.: The base flow correlation method with multiple gauged sites, J. Hydrol., 34, 371–380, 2007b.

Chapter 3

Is the GEV model suitable as a pan-European parent?

The present chapter corresponds to the following scientific publication in its original form:

Salinas, J. L., Castellarin, A., Viglione, A., Kohnová, S., and Kjeldsen, T. R.: Regional parent flood frequency distributions in Europe – Part 1: Is the GEV model suitable as a pan-European parent?, *Hydrol. Earth Syst. Sci.*, 18, 4381-4389, doi:10.5194/hess-18-4381-2014, 2014.

Hydrol. Earth Syst. Sci., 18, 4381–4389, 2014
www.hydrol-earth-syst-sci.net/18/4381/2014/
doi:10.5194/hess-18-4381-2014

Hydrology and Earth System Sciences
Open Access

Regional parent flood frequency distributions in Europe – Part 1: Is the GEV model suitable as a pan-European parent?

J. L. Salinas[1], A. Castellarin[2], A. Viglione[1], S. Kohnová[3], and T. R. Kjeldsen[4]

[1]Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Vienna, Austria
[2]Department DICAM, School of Civil Engineering, University of Bologna, Bologna, Italy
[3]Department of Land and Water Resources Management, Faculty of Civil Engineering, SUT Bratislava, Bratislava, Slovakia
[4]Department of Architecture and Civil Engineering, University of Bath, Bath, UK

Correspondence to: J. L. Salinas (salinas@hydro.tuwien.ac.at)

Received: 2 May 2013 – Published in Hydrol. Earth Syst. Sci. Discuss.: 22 May 2013
Revised: 11 September 2014 – Accepted: 15 September 2014 – Published: 5 November 2014

Abstract. This study addresses the question of the existence of a parent flood frequency distribution on a European scale. A new database of L-moment ratios of flood annual maximum series (AMS) from 4105 catchments was compiled by joining 13 national data sets. Simple exploration of the database presents the *generalized extreme value* (GEV) distribution as a potential pan-European flood frequency distribution, being the three-parameter statistical model that with the closest resemblance to the estimated average of the sample L-moment ratios. Additional Monte Carlo simulations show that the variability in terms of sample skewness and kurtosis present in the data is larger than in a hypothetical scenario where all the samples were drawn from a GEV model. Overall, the *generalized extreme value* distribution fails to represent the kurtosis dispersion, especially for the longer sample lengths and medium to high skewness values, and therefore may be rejected in a statistical hypothesis testing framework as a single pan-European parent distribution for annual flood maxima. The results presented in this paper suggest that one single statistical model may not be able to fit the entire variety of flood processes present at a European scale, and presents an opportunity to further investigate the catchment and climatic factors controlling European flood regimes and their effects on the underlying flood frequency distributions.

1 Introduction

The first step for any assessment of the flooding potential or flood hazard is the estimation of the design flood associated with a given annual exceedance probability, often quoted in terms of a recurrence interval T measured in years. This information is most commonly obtained using flood frequency estimation techniques based on statistical analyses of series of observed flood peak discharges. Unsurprisingly, extreme flood events are seldom observed locally, and hydrologists have little or no chance of gathering data from an adequate sample of catastrophes for analysis, especially for prediction at ungauged sites, with the exception of post-event surveys (see e.g. Marchi et al., 2010; Gaume et al., 2010). It is therefore important that effective and practical procedures are available, to assist hydrologists in making inferences about flood risk, both at gauged and at ungauged sites (Blöschl et al., 2013; Salinas et al., 2013).

When only an estimate of the peak flow value is needed, at-site and regional statistical extreme value analysis of river flow data can be used, depending on data availability. However, it is well-known that such estimates can be associated with a high degree of uncertainty, and it is therefore important to ensure that decisions are robust and are made based on as much information as possible (Viglione et al., 2013; Merz and Blöschl, 2008a, b; Martins and Stedinger, 2001). The choice of method for flood frequency estimation in any particular situation is often dictated by factors such as national or institutional tradition, modeller expertise, complexity and objective of study, legislative requirements, and data availability (Castellarin et al., 2012). It is usual though, to assume the existence of a parent flood frequency distribution within a certain region. The question of the existence itself of a parent distribution on different spatial scales has puzzled hydrologists for many years and substantial work has been done in order to verify or falsify this hypothesis. Matalas et al. (1975) found that the variance of sample coefficient of skewness was always higher for observed data than for simulated flood peaks for a set of considered parent distributions, calling this phenomenon "skew separation". They further showed that it could not be attributed either to small sample properties of the skewness estimator or to autocorrelation of the flood peak series. Dawdy and Gupta (1995) related the magnitude of this "skew separation" to the scaling structure of flood peaks and the heterogeneity among regions on the study from Matalas et al. (1975). Other authors have also shown that different distributions than those considered in Matalas et al. (1975) were able to reproduce the skewness variability; namely, Houghton (1978) for the *Wakeby* distribution and Rossi et al. (1984) for the TCEV (*two-component extreme value distribution*), while Ashkar et al. (1992) and Bobee et al. (1993) provide criticism on using the separation of skewness for choosing the type of distribution to be used in regional flood frequency analyses and give more importance to the step of the definition of homogeneous regions in order to avoid mixing of different skewness values from different populations.

More recently, the different behaviour of the tails of the distribution functions used in environmental extremes has been investigated by El Adlouni et al. (2008), and Gaál et al. (2009) or Papalexiou et al. (2013) performed studies on daily rainfall, similar to the present one, analysing which kind of extreme value distribution fits best the precipitation extremes

The present paper introduces the first available inventory of data and statistical methods for flood frequency estimation across Europe, compiled with the aim to homogenize and harmonize the current level of knowledge on the approach to flood frequency estimation used across Europe. The inventory has been created as part of COST (European Cooperation in Science and Technology) Action ES0901 (European Procedures for Flood Frequency Estimation – FloodFreq), which is a European-Commission-funded project that develops a network of experts, involved in nationally funded flood frequency estimation research projects. Their main task is to undertake a pan-European comparison and evaluation of methods of flood frequency estimation under the various climatologic and geographic conditions found in Europe, and promoting a synergic approach to flood hazard assessment, as requested by the European Flood Directive (Kjeldsen, 2011; Castellarin et al., 2012).

This study addresses explicitly the question of the existence of a parent flood frequency distribution on a European scale. It presents the results of an assessment based on the analysis of a newly established pan-European database of annual maximum series (AMS) of flood flows and their statistical characteristics compiled for the FloodFreq project, in order to find the most suitable frequency distributions for modelling the flood frequency regimes in Europe. In the companion paper by Salinas et al. (2014), the link between catchment and climate attributes and the choice from a set of potential parent regional flood frequency distributions is investigated on a subset of the database presented in this article.

Table 1. FloodFreq streamflow database: number of sites and station-years of data for the national annual maxima sequences.

Country	No. of sites	Station-years of data	Kind of data
Austria	676	28 592	Instantaneous
Cyprus	9	382	Daily
Germany	415	22 516	Daily
Denmark	43	2789	Daily
France	1172	45 331	Instantaneous
Ireland	215	6708	Instantaneous
Italy	373	8207	Instantaneous
Lithuania	30	1953	Instantaneous
Norway	104	3120	Daily
Poland	39	3426	Instantaneous
Slovakia	174	7995	Instantaneous
Spain	220	8594	Instantaneous
United Kindom	635	23 200	Instantaneous
FloodFreq	4105	162 813	

2 Inventory of data and methods and the European flood database

The first phase of the FloodFreq COST Action focused on the compilation of inventories of data sets and methods for flood frequency estimation at a European scale. An extensive survey was conducted among 15 European countries in order to assess the availability of flood data and catchment descriptors and to investigate the existence of national guidelines for flood frequency estimation. Particularly, if these guidelines existed, related to the issue of large-scale underlying parent distributions, it was of interest if any type of flood frequency distribution was recommended. The main results of the survey relevant to this paper are presented below.

2.1 European flood database

The assessment of flood data availability at national level for the 15 European countries included in the survey showed that the AMS of flood flows are the most common standard. Therefore, it was decided to focus on a collection of AMS of flood flows considering daily flows, as well as instantaneous peak flows where available.

From the 15 surveyed countries, 13 agreed to share flood data in the frame of the FloodFreq COST Action. Due to national policies and regulations that restrict the publication of some of these data, the flood data themselves were summarized into statistical moments. In particular, the AMS for a total of 4105 sites was characterized by the number of observations n and sample L-moment ratios of orders 2–4 (i.e. sample L-coefficient of variation, sample L-coefficient of skewness and sample L-coefficient of kurtosis). Table 1 contains a summary of the national AMS data sets available in the database. The use of L-moments instead of conventional moments offers several advantages such as the possibility of characterizing a wider range of distributions, smaller bias and higher robustness of the parameter estimators when applied to short samples. For more details on L-moments see, for example, Hosking (1990) and Hosking and Wallis (1997).

2.2 National guidelines for flood frequency estimation

National guidelines for flood frequency estimation are available in 9 of the 15 surveyed countries. In Germany, reference studies are available at the level of the federal states, and in Belgium for the Flemish part only. Public agencies and institutions in six countries provide recommendations as to the most suitable parent distributions to be used for flood frequency analysis, but in general such guidance appears to be sparse. In a number of countries, the *generalized extreme value* (GEV) distribution is among the recommended choices (e.g. Austria, Germany, Italy, Spain), but a variety of two-, three- or four-parameter distributions are also used, including the *Gumbel* (GUM) distribution in Finland and Spain, the three-parameter *generalized Pareto* (GPA) distribution in Belgium, the three-parameter *generalized logistic* (GLO) distribution in the UK, or the four-parameter TCEV in Italy and Spain. The Slovenian Environment Agency uses five different distributions (*normal*, *lognormal*, *Pearson type 3*, *log-Pearson type 3* and *Gumbel*) implemented in their own software DIST. In Slovakia, the *gamma*, *three-parameter lognormal*, *log-Pearson type 3* and GEV distributions are often used. Six countries reported that they have no standard parent distribution and the choice of an appropriate model depends mostly on the region where the analysis is undertaken (Castellarin et al., 2012).

The existence of some preferred statistical models, provides a motivation for a further investigation into potential candidates for pan-European flood frequency distributions, taking advantage of the uniquely extensive flood data collection compiled in this study (see Table 1).

3 A pan-European flood frequency distribution?

3.1 L-moment ratio diagram framework

The analyses of the FloodFreq database presented in this study address the issue of probabilistic model choice in the L-moment working environment and, in particular, through the use of L-moment ratio diagrams which enable graphical identification of a suitable regional parent distribution among several two- and three-parameter candidates (see e.g. Hosking and Wallis, 1997). The scientific literature seems to agree on the value of L-moment ratio diagrams for guiding the selection of a regional parent distribution (e.g. Vogel and Fennessey, 1993; Vogel and Wilson, 1996; Peel et al., 2001; Strupczewski et al., 2011). An additional advantage of the diagrams is that L-moments are particularly suitable for short samples often associated with annual flood sequences, as sample L-moments tend to be less biased than the corresponding estimators of conventional moments (Vogel and Fennessey, 1993).

Two types of L-moment ratio diagrams are commonly used in the literature to assess the goodness of fit of regional parent distributions: (i) a diagram plotting the L-coefficient of variation against the L-coefficient of skewness (or L-Cs–L-Cv diagram), and (ii) a plot of the L-coefficient of kurtosis against the L-coefficient of skewness (or L-Cs–L-Ck diagram). The former is used to assess the suitability of various two-parameter distributions, while the latter version of the diagram is more commonly used when three-parameter distributions are considered. The suitability of various candidate parents is assessed by analysing how close the cloud of sample L-moments computed for the study region lies, relative to the lines corresponding to the different theoretical models. This study only presents L-Cs–L-Ck diagrams, as all the parent distributions considered are three-parametric, while in the companion paper by Salinas et al. (2014) both L-Cs–L-Ck and L-Cs–L-Cv diagrams will be used.

3.2 Average behaviour of the FloodFreq database

Figure 1a shows the L-Cs–L-Ck diagram for the entire FloodFreq data set (see also Table 1) and includes the sample L-moment ratios for all of the catchments in the data set (light-grey circles), together with four lines illustrating the theoretical relationship between L-Cs and L-Ck for the three-parameter frequency distributions that, as highlighted by the survey commented in Sect. 2.2, were preferentially recommended in the national guidelines, namely the GEV, GLO and *three-parameter lognormal* (LN3), and *Pearson type 3* (PE3).

To reduce some of the noise that is present in the empirical data due to sampling uncertainty and better determine which of the four three-parameter distributions considered better represents the statistical properties of the sample, a record length weighted moving average (WMA) is included in Fig. 1a. In particular, the WMA is computed by taking the weighted mean of the 50 neighbouring sample L-Cs values, and plotting it against the weighted mean of the corresponding 50 sample L-Ck values. Each sample L-moment ratio is weighted proportionally to the record length to reduce the impact of sampling variability from short records. The WMA values in Fig. 1a follow closer the theoretical relationship between L-Cs and L-Ck of the GEV distribution than that of

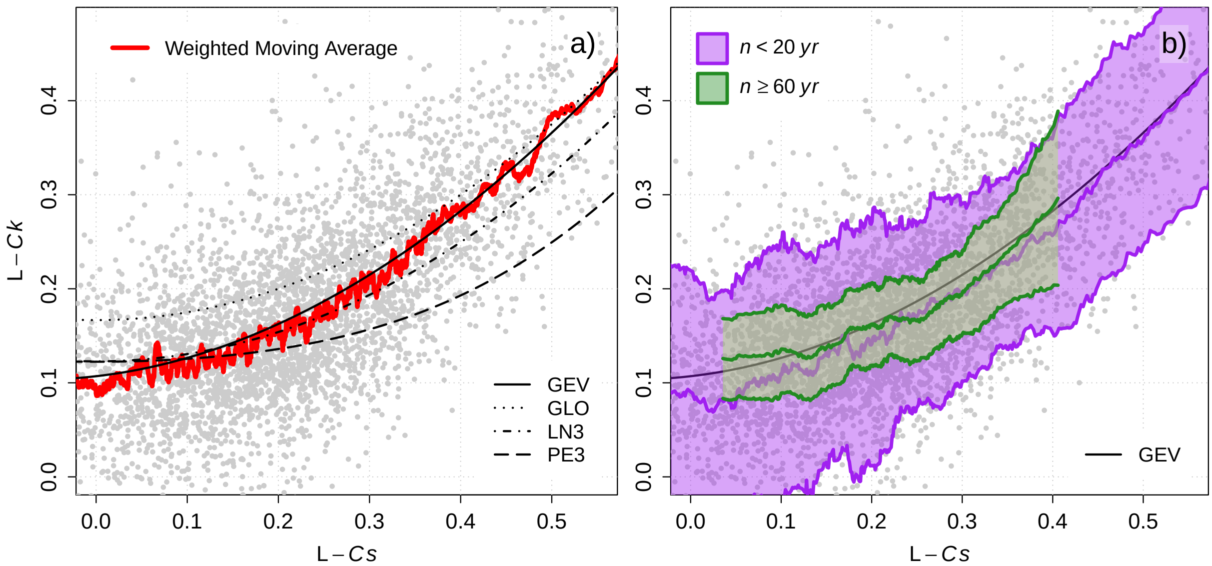


Figure 1. L-Cs–L-Ck diagram for the FloodFreq database with **(a)** the record length weighted moving average over 50 catchments (red line) for the whole database and **(b)** the record length weighted moving averages and standard deviations for stations with more than 60 yr (green lines and polygon) and less than 20 yr of data (purple lines and polygon) separately.

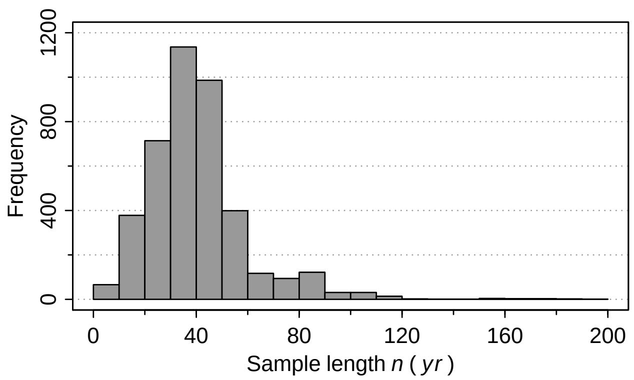


Figure 2. Sample length distribution for the entire FloodFreq database.

any of the other considered distributions. The position of the WMA indicates therefore that the GEV distribution might be a better candidate for describing the frequency regime of annual maximum floods at a pan-European level, compared to the other three extreme value distributions studied.

Given the known relationship between sampling uncertainty of the L-moment ratios and sample length n (see e.g. Viglione, 2010; Hosking and Wallis, 1997), it is necessary at this point to describe more in detail the distribution of n in the database. Figure 2 shows a histogram of the record lengths. The mean and median values are 40 and 38 yr, respectively. Later in the analysis, the database is subdivided into four n classes with the limits at 20, 40 and 60 yr. The percentage of stream gauges within each class is 10.8 % (444 stations with $n < 20$ yr), 45.0 % (1850 stations with $20\,\text{yr} \leq n < 40$ yr), 33.7 % (1385 stations with $40\,\text{yr} \leq n < 60$ yr), and 10.4 % (426 stations with $n \geq 60$ yr), respectively. This classification will ease the interpretation of the posterior analyses, trying to reduce the heterogeneity introduced by mixing stations with different sample sizes, as can be acknowledged in Fig. 1b. Here, the intervals corresponding to the moving averages plus-minus the moving standard deviations of the L-Ck values (both weighted proportionally to the record length over 50 elements) for the stations with $n < 20$ yr and $n \geq 60$ yr are depicted. One can see that, while the moving averages continue to stay closer to the GEV line in both cases, the spread of the data increases considerably with sample size, and therefore should be taken into account explicitly in the more detailed investigation performed in the following section.

3.3 Monte Carlo simulations

The GEV is the statistical model that best represents the averaged statistical properties of the entire database, if compared to the other three-parameter distributions. In order to falsify or verify the hypothesis that the GEV could actually be a valid underlying pan-European flood frequency distribution, the spread of the observed data has to be compared with the theoretical scenario where all stations represent random samples drawn from GEV distributions with a variety of sample lengths, skewness and kurtosis values. This reference scenario was set up via Monte Carlo simulations. Specifically, a total of 50 000 European-scale simulations are carried out as follows. As commented in the previous section, the effect of sample size n should be taken into account if the dispersion of the data in the L-moment ratio diagram needs to be investigated. Therefore, the database is sorted

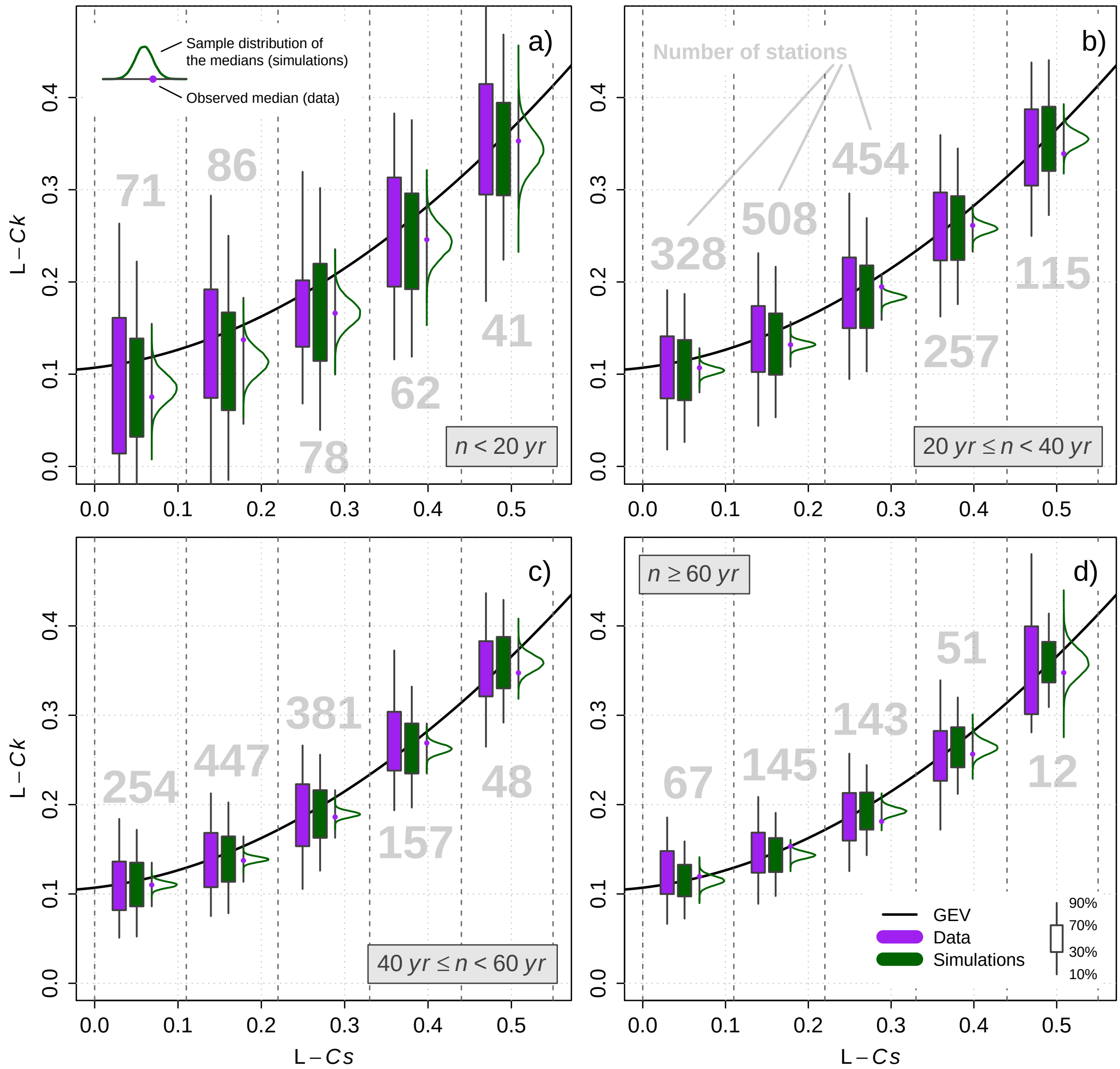


Figure 3. Monte Carlo simulations comparing the distribution of sample L-Ck from 50 000 realisations at the European scale, all randomly drawn from GEV distributions (green box plots) and the distribution of sample L-Ck from the FloodFreq database (purple box plots). Results are sorted by station sample length n in **(a)** $n < 20$ yr, **(b)** $20\,\text{yr} \leq n < 40$ yr, **(c)** $40\,\text{yr} \leq n < 60$ yr, and **(d)** $n \geq 60$ yr.

into the record length classes described in Sect. 3.2, i.e. 444 stations with $n < 20$ yr, 1850 stations with $20\,\text{yr} \leq n < 40$ yr, 1385 stations with $40\,\text{yr} \leq n < 60$ yr, and 426 stations with $n \geq 60$ yr. In each class, L-Cs values of flood flow sequences are assumed to vary randomly between sites as described by a *normal* distribution with means ranging from 0.20 to 0.22 and standard deviations ranging from 0.14 to 0.17, as in all four cases the sample distributions passed a normality test. The record length distributions inside each n class are modelled by a triangular ($n < 20$ yr), uniform ($20\,\text{yr} \leq n < 40$ yr) and *three-parameter gamma* distributions ($40\,\text{yr} \leq n < 60$ yr and $n \geq 60$ yr). All parameters from the described distributions were estimated from the characteristics of the observed distributions of the database with the method of L-moments. The types of distribution were chosen based on their respective probabilistic plots. Record length and L-Cs are assumed independent, as no significant correlation is found between the sample values. Then, for each one of the 50 000 simulations and each n class, two samples of the corresponding length (444, 1850, 1385 or 426 depending on sample-size class) are generated from the previously defined distributions for the record lengths and population L-Cs values, representing the properties of each of the synthetic European samples. The population L-Ck values are then obtained from the functional relationship between skewness and kurtosis for the GEV model and, without loss of generality, the mean and L-Cv values are both set to 1. For each of the virtual stations, a sample from a GEV distribution is generated, with each of the previously simulated population properties. Finally, the sample L-moment ratios are computed for each of the generated GEV stations, which will not necessarily be located on the theoretical GEV line on the L-Cs–L-Ck diagram, due to both sample variability (finite record lengths) and biases in the sample estimators of the L-moment ratios (see e.g. Hosking and Wallis, 1997, p. 29).

To ease the interpretation of the results, the L-Cs–L-Ck values for both the observed data and the simulations are binned into five equidistant classes between the L-Cs values of 0 and 0.55, and the spread of the L-moment ratios is represented by the 10, 30, 50, 70 and 90 % quantiles of the sample L-Ck values for each bin and shown in the box plots of Fig. 3, together with the number of stations that are inside each of the bins. For the observed data, every bin is associated with only one value of the 10, 30, 50, 70 and 90 % percentiles, respectively, but in the case of the simulations we obtain 50 000 values for each of the percentiles, one for each of the European-scale simulations. This means that we obtain sample distributions for all of the quantiles, as is shown for the case of the medians in Fig. 3. The 10, 30, 70 and 90 % percentiles for the simulations have their corresponding sample distributions, but in Fig. 3 only their averaged values are represented as limits of the box plots for the sake of clarity.

In the Monte Carlo simulations, the explicit assumption that the underlying parent distribution of all stations in Europe is given by the GEV model is made, and this assumption will be verified or falsified with the classical approach of statistical hypothesis testing. We define the null hypothesis H_0 such that, for the ith bin, the j% quantile of the sample L-Ck distribution from the data is not significantly different to the j% quantile in the same bin, for all stations randomly drawn from GEV distributions. For each of the percentiles there is a sample distribution (from the simulations), and an observed value (from the observed data) available, so we can calculate the associated p values, set a significance level and reject or accept the null hypothesis for all the quantiles.

Table 2 shows these calculated p values for the plotted quantiles in Fig. 3. For almost all the cases, and independently of sample size, we cannot reject the null hypothesis that the medians are equal to those from a randomly generated GEV of Europe. This result from Monte Carlo simulations, i.e. that the mean or median L-Cs–L-Ck behaviour can be consistently explained by the GEV distribution, is more robust than the one described in Sect. 3.2 and shown in Fig. 1 on the weighted moving averages. According to, for example, Hosking and Wallis (1997, p. 29), the L-moment ratios in Fig. 1 should be shifted towards the top-right corner of the diagram if one takes into account the sample bias. This would probably cause the averages to differ slightly more from the theoretical GEV curve. In the simulations presented in this section, sample L-moments from the database are compared to sample L-moments from GEV generations and are therefore subject to the same sample bias.

The results for the dispersion of the L-Ck values show a more interesting pattern than the median values. For $n <$ 20 yr, the spreads in the data and the simulations are not significantly different from each other for any of the L-Cs bins in almost all the cases (no rejections with a significance level α of 5 % and three rejections with an α of 10 %). For increasing sample size, the rate of rejection increases, especially for L-Cs values larger than 0.22. Except for two cases, the GEV distribution always fails to explain the dispersion above the median and for larger record lengths also the dispersion below the median, which is significantly higher for the observed data. It is worth noting at this point that the robustness of these results is affected by two factors. First, the number of stations inside each bin after the record length and L-Cs sorting. For example, one cannot draw any strong conclusion from the last bin in Fig. 3d, as it contains only 12 elements, while for other bins the results can be considered more consistent. Second, the biases in the L-moment ratios already commented (see e.g. Hosking and Wallis, 1997, p. 29), which are particularly large in short to medium sample sizes and higher L-Cs values, have the potential to shift some of the stations, whose L-Cs population lies in one bin, into another bin because their sample L-Cs is lower. However, checks of the number of stations per bin in the data and the simulations resulted in a difference of less than 5 %.

3.4 Discussion

Even though the results of the Monte Carlo simulations point out that selecting the GEV distribution as a pan-European parent cannot fully describe the observed variability of sample L-moments, there are some aspects that deserve a deeper discussion. In fact, it is remarkable that for all the European geographical areas considered, including catchments with very different sizes, climatic conditions, and geomorphologic characteristics covered in the FloodFreq database, there is not enough statistical evidence to reject the hypothesis that the GEV distribution is a suitable parent distribution for describing the median behaviour in terms of sample L-moment ratios. From a purely statistical point of view, this could be explained by the fact that the GEV distribution is the theoretical extreme value distribution that expresses in a closed analytical way the three possible asymptotic distributions derived from any kind of parent population, as described in the Fisher–Tippett–Gnedenko theorem (Fisher and Tippett, 1928; Gnedenko, 1943). Therefore, it offers a theoretical justification for using it to reproduce the sample frequency distribution of annual maxima series from many different hydrological and geological extreme phenomena (precipitation depths, flood flows, earthquake magnitudes, wind speeds and others) observed in different geographical contexts around the world (e.g Robson and Reed, 1999; Castellarin et al., 2001; Thompson et al., 2007; Grimaldi et al., 2011).

Table 2. Empirical p values for the observed quantiles of L-Ck given the sample distributions generated from the GEV simulations shown in Fig. 3. In bold, values rejecting a 5 % significance test. In bold and italic, values rejecting a 10 % significance test.

	Range of L-Cs values ($n < 20$ yr)					Range of L-Cs values ($20\,\text{yr} \leq n < 40$ yr)				
L-Ck quantile	0–0.11	0.11–0.22	0.22–0.33	0.33–0.44	0.44–0.55	0–0.11	0.11–0.22	0.22–0.33	0.33–0.44	0.44–0.55
90 %	0.949	***0.968***	0.788	0.626	***0.965***	0.746	**0.996**	**1.000**	***0.969***	0.413
70 %	0.908	0.945	0.129	0.825	0.816	0.806	**0.978**	***0.975***	0.757	0.375
50 %	0.271	***0.956***	0.498	0.562	0.669	0.762	0.478	**0.997**	0.730	***0.025***
30 %	0.121	0.822	0.846	0.559	0.515	0.680	0.784	0.489	0.469	***0.028***
10 %	0.275	0.172	0.926	0.433	0.064	0.069	***0.028***	***0.057***	***0.026***	**0.015**
	Range of L-Cs values ($40\,\text{yr} \leq n < 60$ yr)					Range of L-Cs values ($n \geq 60$ yr)				
L-Ck quantile	0–0.11	0.11–0.22	0.22–0.33	0.33–0.44	0.44–0.55	0–0.11	0.11–0.22	0.22–0.33	0.33–0.44	0.44–0.55
90 %	**0.990**	**0.989**	**0.978**	**1.000**	0.731	**0.999**	**0.999**	**0.976**	***0.966***	**0.999**
70 %	0.627	0.890	***0.961***	**0.989**	0.316	**0.996**	0.928	0.457	0.302	0.847
50 %	0.482	0.342	0.194	0.893	0.117	0.813	**0.995**	**0.005**	0.163	0.238
30 %	0.119	***0.026***	**0.005**	0.728	0.170	0.684	0.420	**0.003**	**0.020**	**0.010**
10 %	0.385	0.202	**0.000**	0.324	**0.010**	0.191	***0.046***	**0.001**	**0.000**	***0.047***

Nevertheless, the results for the Monte Carlo simulations regarding the spread of the L-moment ratios around the GEV line show that the dispersion is bigger than expected from random sampling, particularly for the deviations above the median for non-short samples. In the terminology used by Matalas et al. (1975), a "kurtosis separation" appears, if only the GEV distribution is considered as the underlying parent across Europe. Intersite correlation is most probably present in the original annual flood flow series, from which the sample L-moment ratios have been computed, and this correlation should reduce the observed L-moment variability. In the Monte Carlo generations, this intercorrelation has been entirely neglected and still the variability of the generated L-moments is lower than the observed one. Also, sample estimation uncertainty, particularly for high values of L-Ck, could also play a role by augmenting the variability in the observed L-moments, but the systematic underestimation of the dispersion points out to the fact that the GEV distribution alone is not complex enough to fully describe the variability of sample L-Cs and L-Ck values estimated for the FloodFreq database, being these L-moment ratios surrogates for the entire spectrum of flood generation processes occurring across Europe responsible for the diversity of shapes of flood frequency distributions. It is therefore necessary to further investigate the links of hydrological processes to the L-moment ratios, and in particular to high values of skewness and kurtosis, in order to try to explain these discrepancies.

4 Conclusions

The issue of existence of underlying parent flood frequency distributions across different processes, places and scales is directly addressed in this study. One of the most applied and recommended statistical models, the GEV distribution, has proven to capture the mean and median statistical properties of a pan-European database annual maximum flood series, but the observed variability in the data is larger than what this model can randomly reproduce. This implies that the GEV alone cannot be considered as a candidate for a pan-European flood frequency distribution, as it is not complex enough to reproduce the entire variety of hydrological processes leading to the different shapes of flood frequency curves. This fact rises the more fundamental question if we actually need a pan-European flood frequency distribution, which does not need to be necessarily the case. The investigation of the GEV as a single model for all European catchments was performed based on a very basic inspection of the flood database, but there are many examples that show that one single analytic expression across large scales and more important, across different processes, is not valid for describing all possible local characteristics at once. Rogger et al. (2012) proved that step changes appear in the flood frequency curve when local runoff generation mechanisms are influenced by threshold processes, especially for small mountainous catchments, and these are not captured by any traditional statistical model so far. The case of several Mediterranean regions which are characterized by two distinct flood populations is also very common. These populations are referred by Rossi et al. (1984) as "ordinary floods", generated by frontal-type rainfalls, and "extraordinary floods", generated by highly convective rainstorms. For the modelling of these flood regimes it could be appropriate to use the TCEV model (see e.g. Francés, 1998) as there is a mixture of populations, while it could not be suitable in other regional contexts if there were not such a mixing. Merz and Blöschl (2003) showed for an Austrian data set that different typologies of floods classified after their generation mechanisms may have very different statistical properties and can therefore lead to distinct flood frequency distributions. In particular, if many flood generation processes take place in one catchment, possibly depending on rainfall or snowfall regimes, the overall flood frequency distribution is a result of the combination of the distributions associated with the single mechanisms and is not necessarily expressed in terms of a single analytical model.

The inclusion of information on the underlying hydrological processes in the model choice is therefore of high importance. The companion paper by Salinas et al. (2014) focuses precisely on defining the controls of catchment and climate indicators on the averaged L-moment ratios and the regional flood frequency distributions.

Acknowledgements. The authors are grateful for the financial contribution made by the COST Office grant ES0901 European procedures for flood frequency estimation (FloodFreq) and for contributions made by members of Working Group 2 of the COST Action. Funding from the ERC FloodChange project (ERC Advanced Grant FP7-IDEAS-ERC-AG-PE10 291152) and from the Austrian Science Funds (FWF) as part of the Vienna Doctoral Programme on Water Resource Systems (DK-plus W1219-N22) is also acknowledged. The compilation of the Italian data set was partially supported by the Italian Government through its national grant to the programme on Characterisation of Ungauged Basins by Integrated uSe of hydrological Techniques (CUBIST Project) and the compilation of the Slovakian data set was partially supported by the Slovak VEGA Grant Agency under contract no. 1/0776/13. Both supports are greatly acknowledged. Referees F. Laio, R. Romanowicz, S. Grimaldi, and V. Iacobellis, as well as Günter Blöschl, are thankfully acknowledged for their constructive comments and suggestions.

Edited by: R. Woods

References

Ashkar, F., Bobee, B., and Bernier, J.: Separation of skewness: reality of regional artifact?, J. Hydraul. Eng., 118, 460–475, doi:10.1061/(ASCE)0733-9429(1992)118:3(460), 1992.

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H.: Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales, Cambridge University Press, Cambridge, UK, 2013.

Bobee, B., Cavadias, G., Ashkar, F., Bernier, J., and Rasmussen, P.: Towards a systematic approach to comparing distributions used in flood frequency analysis, J. Hydrol., 142, 121–136, 1993.

Castellarin, A., Burn, D., and Brath, A.: Assessing the effectiveness of hydrological similarity measures for flood frequency analysis, J. Hydrol., 241, 270–285, doi:10.1016/S0022-1694(00)00383-8, 2001.

Castellarin, A., Kohnová, S., Gaál, L., Fleig, A., Salinas, J. L., Toumazis, A., Kjeldsen, T. R., and Macdonald, N.: Review of applied statistical methods for flood frequency analysis in Europe, Milestone Report for WG2 of COST Action ES0901, Tech. rep., Centre for Ecology & Hydrology, Wallingford, UK, on behalf of COST, available at: http://www.cost.eu/media/publications/, 2012.

Dawdy, D. R. and Gupta, V. K.: Multiscaling and skew separation in regional floods, Water Resour. Res., 31, 2761–2767, doi:10.1029/95WR02078, 1995.

El Adlouni, S., Bobée, B., and Ouarda, T. B. M. J.: On the tails of extreme event distributions in hydrology, J. Hydrol., 355, 1–4, doi:10.1016/j.jhydrol.2008.02.011, 2008.

Fisher, R. A. and Tippett, L. H. C.: Limiting forms of the frequency distribution of the largest and smallest member of a sample, Math. Proc. Cambridge Phil. Soc., 24, 180–190, 1928.

Francés, F.: Using the TCEV distribution function with systematic and non-systematic data in a regional flood frequency analysis, Stoch. Hydrol. Hydraul., 12, 267–283, doi:10.1007/s004770050021, 1998.

Gaál, L., Szolgay, J., Lapin, M., and Faško, P.: Hybrid approach to delineation of homogeneous regions for regional precipitation frequency analysis, J. Hydrol. Hydromechan., 57, 226–249, doi:10.2478/v10098-009-0021-1, 2009.

Gaume, E., Gaál, L., Viglione, A., Szolgay, J., Kohnová, S., and Blöschl, G.: Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites, J. Hydrol., 394, 101–117, doi:10.1016/j.jhydrol.2010.01.008, 2010.

Gnedenko, B. V.: Sur la distribution limite du terme maximum d'une serie aleatoire, Ann. Math., 44, 423–453, 1943.

Grimaldi, S., Kao, S. C., Castellarin, A., Papalexiou, S. M., Viglione, A., Laio, F., Aksoy, H., and Gedikli, A.: Treatise on Water Science, chap. Statistical hydrology, 479–517, Elsevier, Oxford, UK, available at: http://www.sciencedirect.com/science/referenceworks/9780444531995, 2011.

Hosking, J. R. M.: L-moments: analysis and estimation of distributions using linear combinations of order statistics, J. R. Stat. Soc. B, 52, 105–124, 1990.

Hosking, J. R. M. and Wallis, J. R.: Regional Frequency Analysis: an Approach Based on L-Moments, Cambridge University Press, Cambridge, UK, 1997.

Houghton, J. C.: Birth of a parent: the Wakeby Distribution for modeling flood flows, Water Resour. Res., 14, 1105–1109, doi:10.1029/WR014i006p01105, 1978.

Kjeldsen, T. R.: COST Action ES0901: European procedures for flood frequency estimation (FloodFreq), Geophys. Res. Abs., 13, EGU2011–10327, 2011.

Marchi, L., Borga, M., Preciso, E., and Gaume, E.: Characterisation of selected extreme flash floods in Europe and implications for flood risk management, J. Hydrol., 394, 118–133, doi:10.1016/j.jhydrol.2010.07.017, 2010.

Martins, E. and Stedinger, J. R.: Historical information in a generalized maximum likelihood framework with partial duration and annual maximum series, Water Resour. Res., 37, 2559–2567, doi:10.1029/2000WR000009, 2001.

Matalas, N. C., Slack, J. R., and Wallis, J. R.: Regional skew in search of a parent, Water Resour. Res., 11, 815–826, doi:10.1029/WR011i006p00815, 1975.

Merz, R. and Blöschl, G.: A process typology of regional floods, Water Resour. Res., 39, 1340, doi:10.1029/2002WR001952, 2003.

Merz, R. and Blöschl, G.: Flood frequency hydrology: 1. temporal, spatial, and causal expansion of information, Water Resour. Res., 44, W08432, doi:10.1029/2007WR006744, 2008a.

Merz, R. and Blöschl, G.: Flood frequency hydrology: 2. combining data evidence, Water Resour. Res., 44, W08433, doi:10.1029/2007WR006745, 2008b.

Papalexiou, S. M., Koutsoyiannis, D., and Makropoulos, C.: How extreme is extreme? An assessment of daily rainfall distribution tails, Hydrol. Earth Syst. Sci., 17, 851–862, doi:10.5194/hess-17-851-2013, 2013.

Peel, M. C., Wang, Q. J., Vogel, R. M., and McMahon, T. A.: The utility of L-moment ratio diagrams for selecting a regional probability distribution, Hydrol. Sci. J., 46, 147–156, doi:10.1080/02626660109492806, 2001.

Robson, A. J. and Reed, D. W.: Flood Estimation Handbook (FEH), chap. Statistical procedures for flood frequency estimation, p. Vol. 3, Institute of Hydrology, Wallingford, UK, 1999.

Rogger, M., Pirkl, H., Viglione, A., Komma, J., Kohl, B., Kirnbauer, R., Merz, R., and Blöschl, G.: Step changes in the flood frequency curve: process controls, Water Resour. Res., 48, W05544, doi:10.1029/2011WR011187, 2012.

Rossi, F., Fiorentino, M., and Versace, P.: Two-component extreme value distribution for flood frequency analysis, Water Resour. Res., 20, 847–856, doi:10.1029/WR020i007p00847, 1984.

Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies, Hydrol. Earth Syst. Sci., 17, 2637–2652, doi:10.5194/hess-17-2637-2013, 2013.

Salinas, J. L., Castellarin, A., Kohnová, S., and Kjeldsen, T. R.: Regional parent flood frequency distributions in Europe – Part 2: Climate and scale controls, Hydrol. Earth Syst. Sci., 18, 4391–4401, doi:10.5194/hess-18-4391-2014, 2014.

Strupczewski, W. G., Kochanek, K., Markiewicz, I., Bogdanowicz, E., Weglarczyk, S., and Singh, V. P.: On the tails of distributions of annual peak flow, Hydrol. Res., 42, 171–192, doi:10.2166/nh.2011.062, 2011.

Thompson, E. M., Baise, L. G., and Vogel, R. M.: A global index earthquake approach to probabilistic assessment of extremes, J. Geophys. Res., 112, B06314, doi:10.1029/2006JB004543, 2007.

Viglione, A.: Confidence intervals for the coefficient of L-variation in hydrological applications, Hydrol. Earth Syst. Sci., 14, 2229–2242, doi:10.5194/hess-14-2229-2010, 2010.

Viglione, A., Merz, R., Salinas, J. L., and Blöschl, G.: Flood frequency hydrology: 3. a Bayesian analysis, Water Resour. Res., 49, 1–18, doi:10.1029/2011WR010782, 2013.

Vogel, R. M. and Fennessey, N.: L moment diagrams should replace product moment diagrams, Water Resour. Res., 29, 1745–1752, doi:10.1029/93WR00341, 1993.

Vogel, R. M. and Wilson, I.: The probability distribution of annual maximum, minimum and average streamflow in the United States, J. Hydrol. Eng., 1, 69–76, doi:10.1061/(ASCE)1084-0699(1996)1:2(69), 1996.

Chapter 4

Climate and scale controls on regional flood frequency distributions

The present chapter corresponds to the following scientific publication in its original form:

Salinas, J. L., Castellarin, A., Viglione, A., Kohnová, S., and Kjeldsen, T. R.: Regional parent flood frequency distributions in Europe – Part 2: Climate and scale controls, *Hydrol. Earth Syst. Sci.*, 18, 4391-4401, doi:10.5194/hess-18-4391-2014, 2014.

Hydrol. Earth Syst. Sci., 18, 4391–4401, 2014
www.hydrol-earth-syst-sci.net/18/4391/2014/
doi:10.5194/hess-18-4391-2014

Hydrology and Earth System Sciences
Open Access

Regional parent flood frequency distributions in Europe – Part 2: Climate and scale controls

J. L. Salinas[1], **A. Castellarin**[2], **S. Kohnová**[3], **and T. R. Kjeldsen**[4]

[1]Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Vienna, Austria
[2]Department DICAM, School of Civil Engineering, University of Bologna, Bologna, Italy
[3]Department of Land and Water Resources Management, Faculty of Civil Engineering, SUT Bratislava, Slovak Republic
[4]Department of Architecture and Civil Engineering, University of Bath, Bath, UK

Correspondence to: J. L. Salinas (salinas@hydro.tuwien.ac.at)

Received: 2 May 2013 – Published in Hydrol. Earth Syst. Sci. Discuss.: 22 May 2013
Revised: 11 September 2014 – Accepted: 15 September 2014 – Published: 5 November 2014

Abstract. This study aims to better understand the effect of catchment scale and climate on the statistical properties of regional flood frequency distributions. A database of L-moment ratios of annual maximum series (AMS) of peak discharges from Austria, Italy and Slovakia, involving a total of 813 catchments with more than 25 yr of record length is presented, together with mean annual precipitation (MAP) and basin area as catchment descriptors surrogates of climate and scale controls. A purely data-based investigation performed on the database shows that the *generalized extreme value* (GEV) distribution provides a better representation of the averaged sample L-moment ratios compared to the other distributions considered, for catchments with medium to higher values of MAP independently of catchment area, while the *three-parameter lognormal* distribution is probably a more appropriate choice for drier (lower MAP) intermediate-sized catchments, which presented higher skewness values. Sample L-moment ratios do not follow systematically any of the theoretical two-parameter distributions. In particular, the averaged values of L-coefficient of skewness (L-Cs) are always larger than *Gumbel*'s fixed L-Cs. The results presented in this paper contribute to the progress in defining a set of "process-driven" pan-European flood frequency distributions and to assess possible effects of environmental change on its properties.

1 Introduction

The companion paper by Salinas et al. (2014) presents a newly established database of flood L-moments from 13 European countries. Based on a preliminary visual inspection and some basic averaging, the *generalized extreme value* (GEV) distribution appeared to be a potential pan-European flood frequency distribution. However, Monte Carlo simulations showed that there is not enough statistical evidence for the existence of a single three-parameter distribution suitable for representing flood frequency regimes all over Europe. This supports the fact of statistical model selection being a topical issue in hydrology and flood frequency analysis in particular (see e.g. Laio et al., 2009). Literature shows recent advances on how to combine different theoretical models together to improve the representation of the local flood frequency regime through multimodel approaches when the reproduction provided by a single theoretical model is not satisfactory (Bogdanowicz, 2010; Kochanek et al., 2012).

Concerning probabilistic model selection, Hosking and Wallis (1997), Vogel and Fennessey (1993) and Peel et al. (2001), among others, recommend using the L-moment ratio diagrams to guide the selection of the most suitable parent flood frequency distribution. L-moment ratio diagrams have been used for detecting suitable parents for a wide spectrum of geohydrological extremes (precipitation depths, flood flows, earthquake magnitudes and others) observed in different geographical contexts around the world (see e.g. Vogel and Wilson, 1996; Robson and Reed, 1999; Thompson et al., 2007). In the hydrological application of L-moment

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

ratio diagrams, and referring more specifically to flood frequency analysis, usage of these kind of diagrams requires sample statistics, which are unavailable or highly uncertain for ungauged or poorly gauged regions. This is one of the reasons why many authors have performed data-based analyses trying to find relationships between sample moments of the flood series and catchment descriptors. Among many others, Schaefer (1990), Farquharson et al. (1992), Meigh et al. (1997), Blöschl and Sivapalan (1997), Iacobellis et al. (2002), Brath et al. (2003), Merz and Blöschl (2003), Di Baldassarre et al. (2006), Merz and Blöschl (2009), Padi et al. (2011), and Viglione et al. (2012) have found regional relationships between sample moments (mainly mean annual flood and coefficient of variation) and catchment area, mean annual rainfall, and other lumped climatic indicators such as aridity. While this correlation may change due to local processes, there is more or less a consensus in the literature that the mean annual specific flood and coefficient of variation of peak annual discharges increase with decreasing catchment size, as well as with decreasing mean annual precipitation (MAP), although this last effect is more clear in arid climates. The current paper tries to go a step further and relate the catchment descriptors to the probabilistic model selection through their controls on the flood moments.

On an applied level, existing guidelines give recommendations on which statistical model, i.e. regional or local parent distribution, to use. This choice could have an important effect on the estimation of high return period flood quantiles due to the different behaviour of the tails of the distribution functions (see e.g. El Adlouni et al., 2008). In some occasions, these recommendations are not justified by any evidence from the local data, or are simply inspired or adapted from analogue guidelines in other countries. Keeping in mind the need for a more effective use of existing data, a key scientific and practical challenge for improved risk assessment is a pan-European comparison and evaluation of the consistency of estimates across methods, physiographic regions and a variety of spatial scales in order to ensure comparable flood frequency estimates and safety measures across Europe, as requested by the Directive 2007/60/EC. In fact, it is of utmost importance for the implementation of the Flood Directive that state-of-the-art and harmonized methods are used to estimate extreme flood frequencies to obtain consistent values for locations where rivers cross national borders.

On the basis of these considerations, this paper addresses the two main scientific questions: can we quantify the main physical controls on the shape of the flood frequency distribution at a regional scale? Can we represent these controls graphically on L-moment ratio diagrams to guide the selection of suitable parent distribution on the basis of fundamental catchment descriptors? For this purpose, a subset of the data presented in Salinas et al. (2014) with a total of 1132 catchments from Austria, Italy, and Slovakia is used to study the control of commonly available physiographic and climatic characteristics (catchment size and mean annual precipitation) on the properties of the underlying probability distribution of flood flow.

2 Description of the data set

The analysis presented in this paper focuses on annual maximum series (AMS) of peak flow from three national databases, namely Austrian, Italian and Slovakian, and it addresses the control of catchment size and climate, respectively, on the flood frequency regime. In particular, the analysis considers catchment area and MAP as catchment descriptors for the three data sets. The flood data was shared in the frame of the FloodFreq COST Action ES0901 (Kjeldsen, 2011; Castellarin et al., 2012) and constitutes a subset of the database presented in Salinas et al. (2014). The countries selected for this study were able to share not only the discharge data but the two catchment descriptors mentioned as well, while the rest of the countries had some kind of limitation in this sense.

On the choice of catchment area and mean annual precipitation, previous studies have proven them to exert significant control on the frequency regime of hydrological extremes (see e.g. Schaefer, 1990; Blöschl and Sivapalan, 1997; Brath et al., 2003; Di Baldassarre et al., 2006; Padi et al., 2011). They can be regarded as lumped catchment descriptors used as surrogate covariates representing the spatially distributed and complex hydrological processes controlling the catchment flood response. Precisely, the area of the basin is an indicator of the scale interplays between catchment processes and rainfall (Blöschl and Sivapalan, 1995), while mean annual precipitation acts as control of probabilistic behaviour of floods through its effect on antecedent soil moisture conditions (Sivapalan et al., 2005), and also provides an indication about other local and atmospheric process.

When combined, the three national data sets consist of AMS from a total of 1132 catchments (Austria, 676 gauges; Italy, 282 gauges; and Slovakia, 174 gauges). Table 1 describes the data set in terms of catchment area, MAP, record length of annual maximum series (n), sample L-coefficient of variation (L-Cv), L-coefficient of skewness (L-Cs) and L-coefficient of kurtosis (L-Ck) for the case study. The geographical locations of the considered stream gauges are shown on the map in Fig. 1. From a purely visual analysis, due to the lack of detailed information on the hydrological regimes, the database is dominated by catchments with humid and continental climates, with a big proportion of mountainous catchments (where snow is supposed to play a significant role, but could not be included in the analysis due to the lack of this kind of data), and only a small percentage of catchments that could be considered Mediterranean or arid located in Italy.

As illustrated in the scatterplot of Fig. 2, the data set includes a range of values for catchment area and MAP, and does not show any statistically significant correlation

Table 1. Summary of the Austrian, Italian and Slovakian national data sets. Information on the distribution of catchment area, MAP, record length (n) and sample L-moment ratios of the annual flood sequences is given.

	Area (km^2)	MAP ($mm\,yr^{-1}$)	n (yr)	L-Cv (–)	L-Cs (–)	L-Ck (–)
Min.	4.6	501.7	9	0.0152	−0.1209	−0.1583
1st quartile	64.9	902.8	22	0.2194	0.1777	0.1268
Median	157.0	1112.0	34	0.2763	0.2705	0.1905
Mean	2096.5	1163.6	38	0.2945	0.2782	0.2074
3rd quartile	534.0	1369.3	47	0.3558	0.3733	0.2730
Max.	131 488.0	2312.3	182	0.7691	0.7737	0.7132

between the two (i.e. sample Pearson coefficient is equal to −0.010, and the null hypothesis of zero correlation is associated with a p value of 0.732). It is worth noting here that, as illustrated in Fig. 2, very large catchments (catchment areas larger than 10 000 km^2) are associated with medium MAP values (about 1200 $mm\,yr^{-1}$). Therefore, very large catchments in the study area are neither "wetter" nor "drier" catchments ("wetter" and "drier" as defined in Sect. 3.1), and this is an important element for the analyses described in Sects. 3.1 and 3.2.

As in the companion paper by Salinas et al. (2014), the framework used to analyse the flood data is the L-moment environment, and, in particular, through the use of L-moment ratio diagrams. Both L-Cs–L-Cv and L-Cs–L-Ck diagrams are used, as several two- and three-parameter candidate distributions are investigated.

3 Results

3.1 Area and MAP control on sample L-moments

A detailed exploration was undertaken to better understand the controls on the flood frequency regime exerted by physiographic and climatological factors, represented here by area and MAP. To minimize the possible effects of sampling variability associated with short records when estimating higher-order sample L-moments (see e.g. Viglione, 2010), the minimum record length was set to 25 yr of data, reducing the data set to a total of 813 catchments (Austria, 493 gauges; Italy, 151 gauges; and Slovakia, 169 gauges). The combined data set was divided into six smaller subsets based on thresholds defined as the 20 and 80 % quantiles of the catchment descriptor values. For convenience, the following subsets are defined to characterize the catchments according to size: smaller catchments (area $< 55\,km^2$), intermediate catchments ($55\,km^2 <$ area $< 730\,km^2$) and larger catchments (area $> 730\,km^2$). Analogously, the catchments were classified based on MAP as drier catchments (MAP $< 860\,mm\,yr^{-1}$), medium catchments ($860\,mm\,yr^{-1} <$ MAP $< 1420\,mm\,yr^{-1}$) and wetter catchments (MAP $> 1420\,mm\,yr^{-1}$) (see also Fig. 2). The adjectives drier and wetter, and smaller and larger are relative to the distribution of wetness and sizes of the study data set. The 20 and 80 % quantiles were selected after a set of preliminary trials as they enabled us to enhance the representation of the peculiarities in the flood frequency regimes for drier against wetter and for larger against smaller catchments for the considered data set.

For each of the wetness and size subsets, the record length weighted moving average (WMA) values of samples L-Cv, L-Cs and L-Ck were computed using a window of 70 catchments, where the 70 neighbouring catchments were selected by taking the closest catchments in terms of the considered descriptor (area or MAP). For each sample of the 70 catchments, the associated WMA value was plotted against the corresponding mean of catchment descriptor (area or MAP) as shown in Fig. 3 for each of the six subsets. Note that in this context, each individual WMA value has a regional validity, as it is derived from a pooling group of 70 sites, defined based on the similarity in terms of catchment size and rainfall regime (MAP). For example, the first point of the yellow line in Fig. 3a represents a non-contiguous region with MAP $< 860\,mm\,yr^{-1}$ (drier catchments) and the 70 smallest sizes of the subset (in this case, catchment areas from 36 to 103 km^2). The minimum amount of information for each point, assuming serial and spatial independence of the stations, would correspond to $70 \times 25 = 1750$ station-years of data. The width of the window (i.e. 70 sites) provides a trade-off between the desire to effectively identify and visualize larger-scale structures in the data set and local deviations from the averaging process, and the conflicting need to work on larger samples to reduce the effects of sampling uncertainty.

Considering the WMA values plotted in Fig. 3, an observable feature is a general tendency for all the L-moment ratios to decrease with increasing area and MAP values. This is a result already reported in the literature for the case of conventional product moments, with special focus on scale effect on the coefficient of variation of the flood distribution (see e.g. Schaefer, 1990; Blöschl and Sivapalan, 1997; Brath et al., 2003; Merz and Blöschl, 2003; Di Baldassarre et al., 2006; Merz and Blöschl, 2009; Padi et al., 2011; Viglione

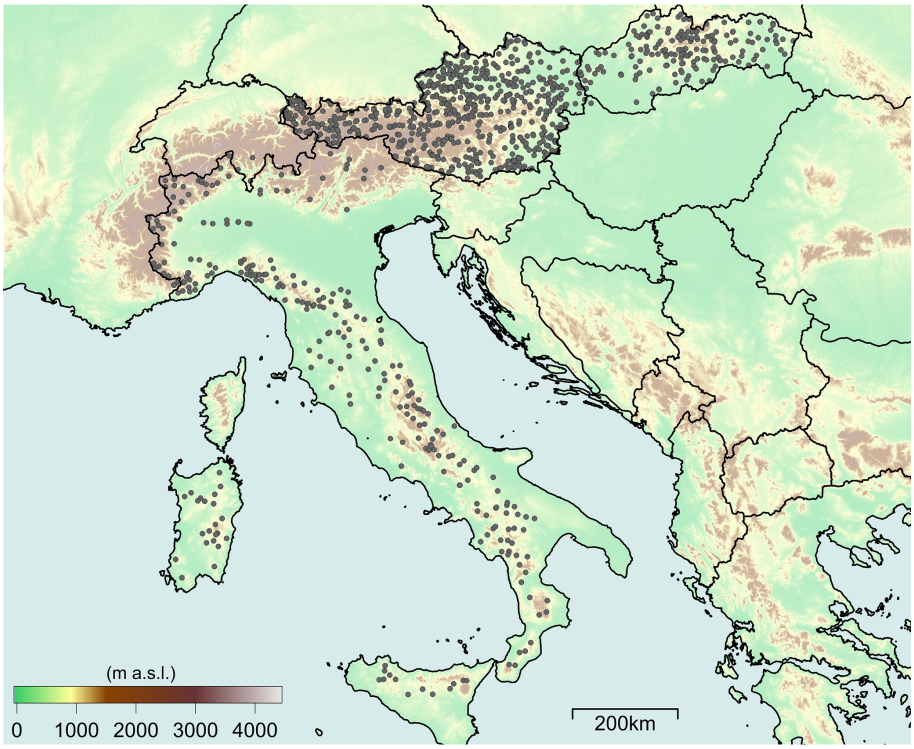


Figure 1. Map showing the location of the 1132 considered Austrian, Italian and Slovakian gauging stations (points). Colour scale in the background represents terrain elevation in metres a.s.l. (above sea level).

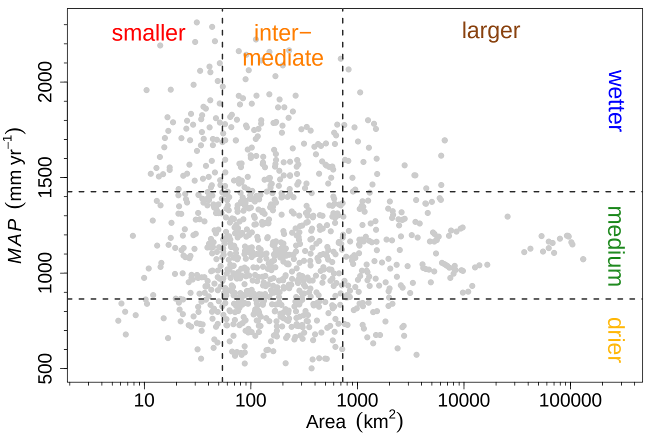


Figure 2. Catchment characteristics of the Austrian, Slovakian and Italian data sets. For each catchment, MAP is plotted against catchment area (grey circles). Black, dashed lines represent the 20 and 80 % quantiles for each catchment descriptor, defining the subsets as smaller (area $< 55\,\mathrm{km}^2$), intermediate ($55\,\mathrm{km}^2 <$ area $< 730\,\mathrm{km}^2$), larger (area $> 730\,\mathrm{km}^2$), drier (MAP $< 860\,\mathrm{mm\,yr}^{-1}$), medium ($860\,\mathrm{mm\,yr}^{-1} <$ MAP $< 1420\,\mathrm{mm\,yr}^{-1}$), and wetter catchments (MAP $> 1420\,\mathrm{mm\,yr}^{-1}$).

et al., 2012). Farquharson et al. (1992) and Meigh et al. (1997) also found an increase of coefficient of variation and skewness with increasing aridity, which is consistent with this study, if one takes the lower MAP values as an indicator for a higher aridity. The largest gradients in Fig. 3 are observed for L-Cv, followed by L-Cs and, finally, L-Ck, confirming the lower variability of higher-order L-moments in space (see e.g. Hosking and Wallis, 1997), and also showing that the lower-order L-moment ratios have a stronger link to catchment and climate properties than higher-order L-moment ratios. It is also noticeable that the WMA lines are not evenly spaced, which indicates a degree of non-linearity between the flood characteristics and the catchment properties. This is particularly evident when considering L-Cv plotted against both catchment area and MAP in Figs. 3a and b, for L-Cs plotted against MAP (subsets defined based on area) in Fig. 3d and, to some extent, for L-Ck plotted against MAP (again, subsets defined based on area) in Fig. 3f. In addition to the general tendency of decreasing averaged L-moment ratios with increasing area and MAP values, an interesting feature is observed in the L-Cv versus area diagram in Fig. 3a. If one looks at the single stations (grey dots) there is an ascending–descending relationship of the L-Cv values with increasing area, with a maximum located around $100\,\mathrm{km}^2$. This was also found, among others, by Blöschl and Sivapalan (1997) and Iacobellis et al. (2002) in more regional contexts, and it is relevant to find it as well in the present study.

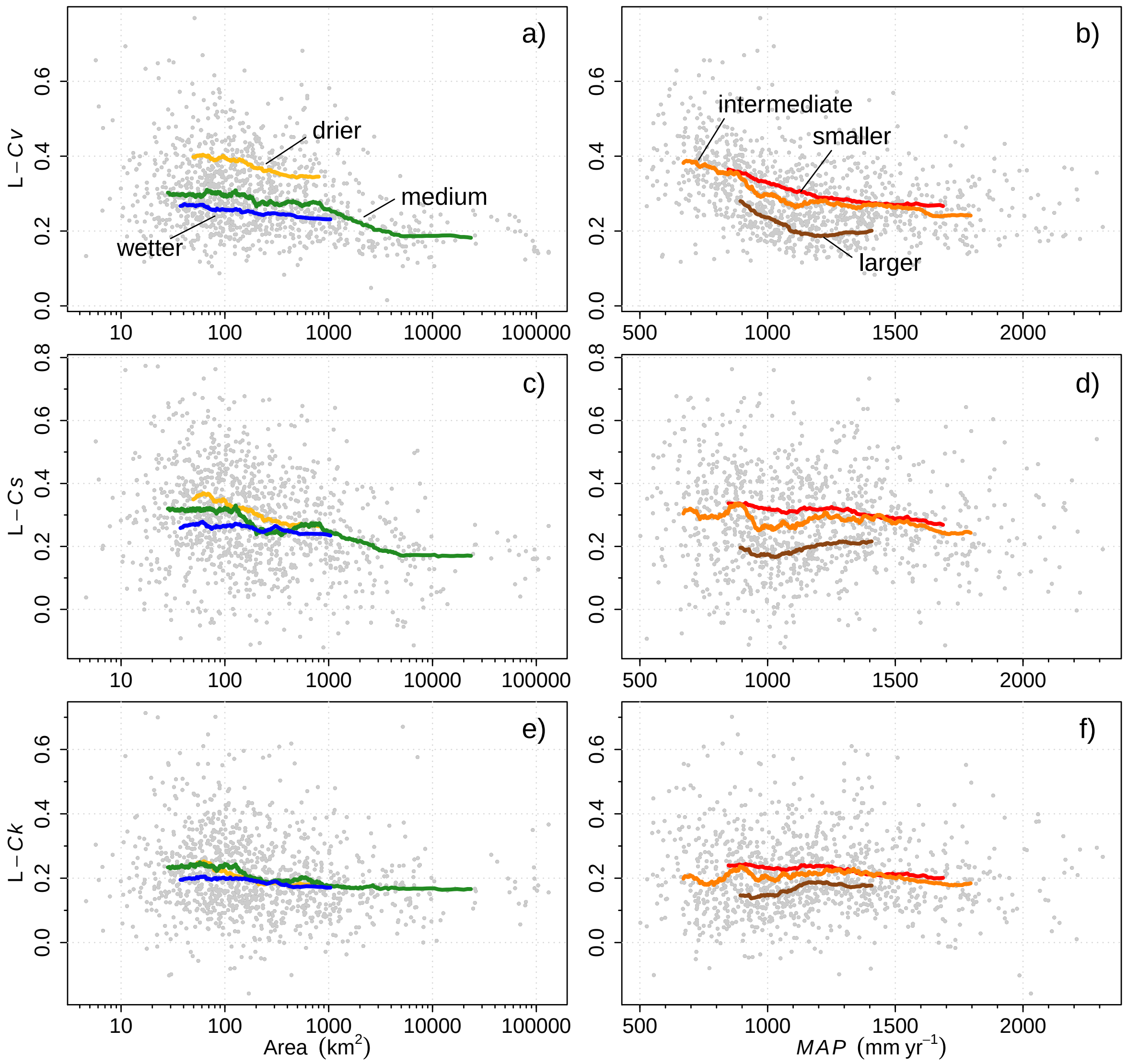


Figure 3. Sample L-Cv, L-Cs and L-Ck values for each catchment (grey points) plotted against catchment area and MAP. Lines show the record length WMAs over 70 catchments for the subsets smaller, intermediate, larger, drier, medium, and wetter defined in Fig. 2.

3.2 Area and MAP control on regional flood frequency distribution

Acknowledging the influence of both catchment size and mean annual precipitation on the L-moment ratios, the next step in the analysis is to assess the impact of this influence on the underlying regional parent distribution of annual flood sequences. This investigation is based on the novel use of L-moment ratio diagrams, which, in this context, are used to analyse the sensitivity of the choice of a parent distribution to the catchment and climate characteristics. The two types of L-moment ratio diagrams mentioned in Sect. 2 are used, namely (i) L-Cv–L-Cs and (ii) L-Cs–L-Ck. The former is used in this study for assessing the suitability of the commonly used two-parameter distributions of *Gumbel* (GUM), *gamma* (GAM), *two-parameter lognormal* (LN2) and *exponential* (EXP). The latter is used in connection with the three-parameter distributions *generalized logistic* (GLO), *generalized extreme value* (GEV), *three-parameter log-normal* (LN3) and *Pearson type 3* (PE3). These distributions, presented in the companion paper by Salinas et al. (2014), were found to be preferentially recommended in national guidelines for flood frequency estimation from several European countries (Castellarin et al., 2012). The assessment of which statistical model fits better the averaged statistical properties of the sample was done visually based on the distance between the averaged sample L-moment ratios and the theoretical lines, as objective goodness of fit measures require the flood peak data (Laio, 2004) and, in this case, only the L-moment ratios were available.

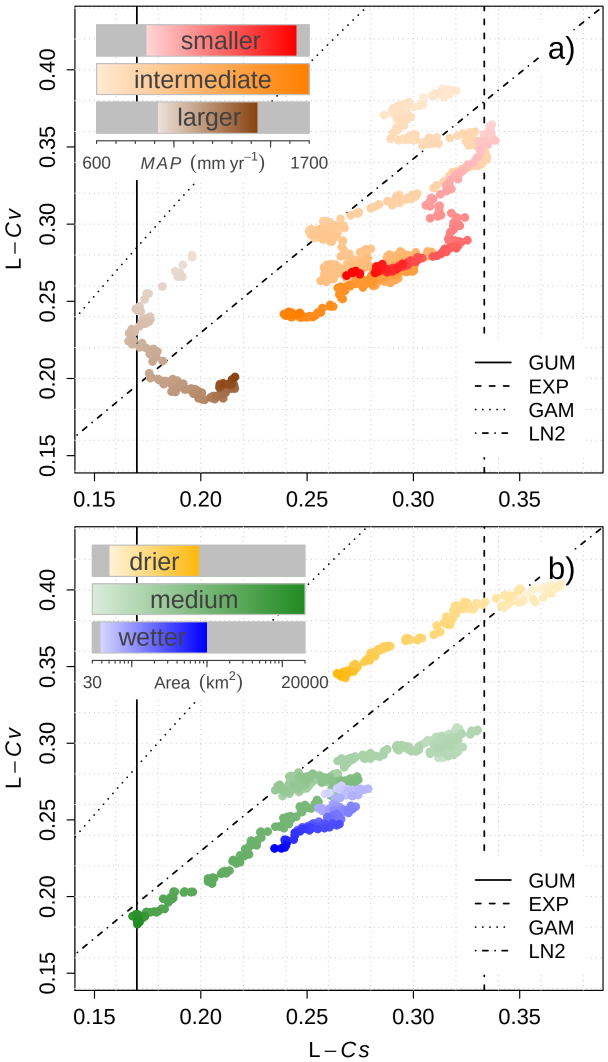


Figure 4. L-moment ratio diagrams for the subsets defined by **(a)** catchment area (smaller, intermediate, larger), and **(b)** MAP (wetter, medium, drier) described in Fig. 3. Each point represents the record length WMA over 70 catchments of L-Cv against corresponding values of L-Cs and the colour intensity is proportional to **(a)** MAP and **(b)** catchment area.

3.2.1 Two-parameter distributions

The diagrams in Fig. 4 report the WMA values of L-Cs and L-Cv associated with a given average value of catchment area or MAP for the same moving windows for the 70 catchments defined in the previous section and shown in Fig. 3, again stratified in smaller, intermediate and larger catchments (Fig. 4a) and wetter, medium and drier catchments (Fig. 4b). To emphasize the influence of the catchment descriptors (i.e. area or MAP), the colour intensity of each plotted WMA value has been graded according to the value of the catchment descriptor that has not been used for the stratification, with increasing intensity for increasing descriptor value. For example, Fig. 4a shows the WMA values when the data set is divided by catchment size into smaller, intermediate and larger catchments, and the colour grading of the points reflects the mean value of MAP for each subset of the 70 catchments. More precisely, red WMA values of L-Cv and L-Cs in Fig. 4a correspond to the transect associated with smaller basins in the catchment descriptor space defined by MAP, orange relates to the intermediate-sized-basins transect in the catchment descriptor space defined by MAP and brown points stand for the larger-basins transect in the catchment descriptor space defined by MAP. Analogously, in Fig. 4b, yellow WMA values of L-Cv and L-Cs represent the transect associated with the drier basins in the catchment descriptor space defined by catchment area, green corresponds to the medium MAP basins transect in the catchment descriptor space defined by area, and blue points relate to the wetter basins transect in the catchment descriptor space defined by area.

The position of the WMA values of samples L-Cv and L-Cs relative to the theoretical distributions shown in Figs. 4a and b indicate that none of the considered two-parameter distributions fits the statistical properties of the data sets. In particular, both figures show that the WMA values of sample L-Cs are always larger than that of the *Gumbel* distribution (fixed value of 0.1699) and smaller than that of the *exponential* distribution (fixed value of 0.3333), with the exception of values up to 0.37 for the smallest drier (low MAP) catchments (less intense yellow points in Fig. 4b). Sample values of L-moment ratios do not seem to follow systematically the shape of any of the lines representing the theoretical L-Cv and L-Cs relationships of the considered two-parameter distributions, but some WMA values tend to lie closer to the LN2 curve than to any other. This is the case for the intermediate-sized and medium MAP catchments of the data set (midintensity orange and midintensity green points in Figs. 4a and b, respectively). The only subset for which sample values approach the statistical properties of the *Gumbel* distribution, and they do it towards the intersection with the LN2 curve, is for larger and medium MAP catchments (intense green points in Fig. 4b and, to some extent, midintensity brown points in Fig. 4a). The subset corresponding to drier catchments presents the largest L-Cv and L-Cs values, while the smallest L-moment ratios are found for the subset of larger catchments lying as mentioned before closer to the *Gumbel* line than the rest of WMA values. Inside each subset (i.e. smaller, intermediate, larger, drier, medium or wetter basins) the intensity of gradation increases with decreasing L-Cv and L-Cs values. This means that for larger values of catchment area and MAP, lower regionally averaged values of L-Cv and L-Cs are expected. The gradients are clearer for the smaller and drier catchments (red and yellow points in Fig. 4), while there is a slight increase of the WMA L-Cs values for wetter catchments inside the larger catchments subset. This could be attributed to the fact that, as pointed out in Sect. 2, all larger catchments have a similar averaged

MAP value (between 1000–1400 mm yr^{-1}) belonging to the same intermediate wetness subset and the differences in the rainfall regime are not big enough to draw conclusions about their control on the averaged L-Cs values inside the group.

3.2.2 Three-parameter distributions

Figures 5 and 6 report the L-moment diagrams defined by plotting WMA values of L-Cs and L-Ck in a similar fashion to Fig. 4. In this case, the two intermediate subsets (i.e intermediate-sized and medium MAP values) were plotted separately from the subsets defining smaller, larger, drier and wetter catchments to ease the visual interpretation of the plots (as far as possible avoiding overlapping points). Figure 5a shows the subset of WMA values derived for the smaller and larger catchments, with the colour intensity representing the average MAP value and Fig. 5b illustrates intermediate-sized catchments with the gradation representing again the average MAP value. Similarly, Fig. 6a shows the subset of WMA values representing drier and wetter catchments, with the colour intensity representing catchment size, while Fig. 6b is relative to the subset characterized by intermediate MAP values.

Figure 5a shows that the WMA values associated with larger catchments are located closer to the GEV line than to any other distribution, generally showing slightly higher L-Ck values than expected for a GEV distribution. For smaller catchments, the GEV is the distribution that best represents the statistical properties of the sample, being the scatter of points much closer to the theoretical curve than that from the subset of larger catchments. For intermediate-sized catchments, Fig. 5b highlights a strong control of MAP on the appropriate distribution; medium-sized catchments associated with high MAP values are situated closer the curve of the GEV distribution, while catchments with lower MAP values move towards the LN3 distribution. This implies that, for drier catchments inside the intermediate-sized subset, the distribution type shifts to a more skewed one (for the same L-Ck, the LN3 has higher L-Cs values than GEV).

Figure 6a shows the WMA values for the two subsets including the drier and wetter catchments as defined by the MAP values. The WMA values associated with wetter catchments are located closer to the line defining the GEV distribution, suggesting that GEV is an appropriate distribution for wetter catchments more or less regardless of catchment size (as determined by the blue colour gradation). In contrast, the statistical properties of the drier catchments are better represented by the LN3 distribution, as also illustrated in Fig. 5b. Catchments with intermediate MAP values are associated with a larger range of L-Cs values depending on their size (see Fig. 6b), and lie closer to the GEV line than to any other distribution considered, with a slight tendency for the smallest and largest catchments inside the subset to exhibit higher values of L-Ck than expected from a GEV distribution.

Analogously to Fig. 4, the area and MAP control on the position of the relative WMA values between the subsets remain in Figs. 5 and 6, showing again higher averaged L-moment ratio values when comparing smaller to larger catchments and drier to wetter catchments.

4 Discussion

Previous sections have highlighted the importance of linking the flood generation processes to the observed L-moment ratios of the annual maxima sequences, and the position of the regional averages at the diagrams, in order to understand from the differences between catchments in terms of underlying parent distributions. Two lumped-catchment descriptors are used as surrogate covariates representing the spatially distributed and complex hydrological processes controlling the catchment flood response. Precisely, the area of the basin is an indicator of the scale interplays between catchment processes and rainfall (Blöschl and Sivapalan, 1995), while mean annual precipitation acts as control of probabilistic behaviour of floods through its effect on antecedent soil moisture conditions (Sivapalan et al., 2005), and also provides an indication about other local and atmospheric process.

For example, low MAP values could indicate regions with prevalence for more localized and variable storms, usually flashier in time and with higher rainfall intensities. This higher between-years variability and skewed distribution of rainfall extreme intensities translates into higher L-Cv and L-Cs values of annual floods, as Figs. 4b and 6a suggest. In contrast, long-duration frontal or advective events, associated with larger spatial extensions and lower rainfall intensities, are expected at catchments presenting higher MAP values, more clearly shown in Fig. 6a. These two kinds of precipitation regimes will also have an effect on the co-evolution of landform with hydrological processes (Gaál et al., 2012), in which rainfall plays an important role at multiple timescales. The variability of flood magnitude between years, and the L-coefficient of variation as a measure of this variability, tends to be higher in smaller and intermediate-sized catchments, compared to the larger ones, as shown in Fig. 4. The main reasons are both the spatial heterogeneity of rainfall and the interaction between the spatial and temporal scales of rainfall and catchment size taking place. This interplay causes the catchment to resonate with storms of similar spatio-temporal extension. In the case of smaller basins this corresponds to short duration, high intensity, spatially concentrated storms (i.e. convective events or flash floods), which are also typical of drier climates; while in larger catchments the resonance appears with longer storms, usually associated with lower intensities, with a bigger spatial extension (i.e. advective or frontal events), more typical of wetter environments (see e.g. Blöschl and Sivapalan, 1995; Sivapalan and Blöschl, 1998). These two differentiated regimes for rainfall extremes will

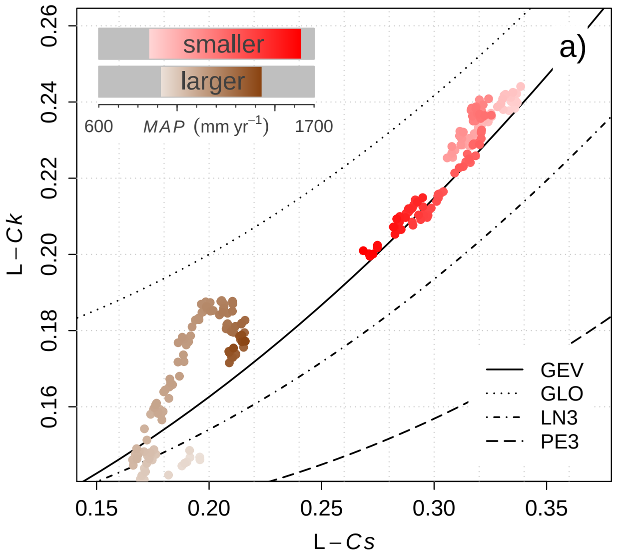


Figure 5. L-moment ratio diagrams for the subsets defined by catchment area: **(a)** smaller, larger, and **(b)** intermediate described in Fig. 3. Each point represents the record length WMA over 70 catchments of L-Ck against corresponding values of L-Cs and the colour intensity is proportional to MAP.

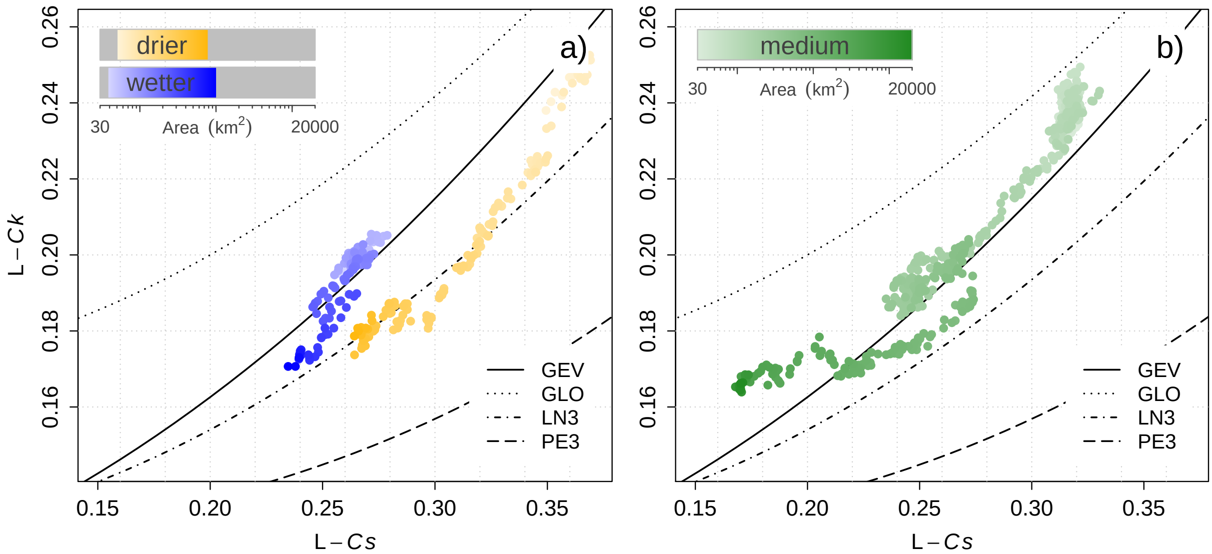


Figure 6. L-moment ratio diagrams for the subsets defined by MAP: **(a)** drier, wetter, and **(b)** medium described in Fig. 3. Each point represents the record length WMA over 70 catchments of L-Ck against corresponding values of L-Cs and the colour intensity is proportional to catchment area.

cause not only a higher L-coefficient of variation but also a higher L-coefficient of skewness in the flood distributions for smaller and drier catchments compared to the larger and wetter ones, as shown in Fig. 5a. Aside from precipitation input, other catchment processes can also play an important role in shaping the properties of the flood distribution. The presence of non-linearities in runoff production and routing in smaller, drier basins (Medici et al., 2008) in contrast to the aggregation of processes in larger catchments (Sivapalan et al., 2002) will translate in decreasing values of L-Cv and L-Cs with increasing values of MAP and, more strongly, catchment size. One visible consequence of the higher dispersion and skewness of the flood frequency distributions with decreasing catchment area and increasing aridity is the fact that predicting flood magnitudes and exceedance probabilities in ungauged basins is more difficult in smaller, more arid catchments, as compared to bigger, less arid ones (see e.g. Salinas et al., 2013).

Therefore, the main findings from the analysis presented in the previous sections need to be interpreted in a hydrological way, instead of in a merely statistical sense. For example, the fact that the GEV distribution is found to be the model representing better the averaged statistical properties of catchments with medium to high values of MAP regardless of size, is probably because few of the catchments in arid regions with highly skewed distributions of rainfall extremes are present in these subclasses. In contrast, there is a clear indication that the LN3 distribution, which has a higher skewness than GEV for a given kurtosis, reproduces better the sample properties of the drier, intermediate-sized subset, representing most likely other flood generation processes than for the data subsets more affine to the GEV distribution. Nevertheless, the limited number of catchments classified simultaneously as smaller and drier, larger and wetter, or larger and drier prevents these conclusions from being extended further.

The LN2 distribution represents in some circumstances a valid alternative to the other commonly used three-parameter distributions, especially for intermediate-sized, medium MAP catchments. The fact of having one parameter less than the GEV or the LN3 allows the LN2 to reproduce only a limited range of hydrological processes maybe not able to capture the extreme cases of the smaller or drier catchments. Sample L-Cs values are shown to be, on a regional average, higher than the ones of the *Gumbel* distribution, being the larger and medium MAP catchments the ones closer to its theoretical curve. This is likely due to the fixed skewness value of the *Gumbel* distribution, relatively low for the regional averages of the given data set and the selected aggregation levels, corresponding substantially to the smoother processes in larger catchments. Also, for the smaller, wetter, and drier catchment subclasses, none of the considered two-parameter distributions is capable of accurately representing the averaged values of the subset.

Recalling the scientific questions presented in the introduction, this study has been able to quantify the controls of area and mean annual precipitation on the sample L-moment ratios of annual maximum flood discharges, and has shown how these controls may guide the selection of suitable parent distributions in an L-moments diagram framework. Additionally, the novel use of traditional L-moment ratio diagrams presented in Figs. 4, 5 and 6 may be very informative, and could help to better understand the changes in flood hazard resulting from different sources of environmental change. By explicitly accounting for the conceptual process controls through catchment descriptors (catchment area and MAP in this study), the sensitivity of the flood frequency distribution to changes in process controls can be determined. For example, Figs. 5b and 6a show that for medium-sized catchments the most appropriate distribution changes from a GEV to a LN3 distribution as MAP decreases. Thus, if future climate projections indicate a reduction of MAP, then the results in Figs. 5b and 6a suggest that the corresponding change in flood distribution is likely to be characterized by a move towards a larger skewness (e.g. from a GEV to a LN3 distribution), assuming that the current relationship between MAP and storm rainfall intensity distributions holds in future climates. This sensitivity analysis could be extended by including additional catchment descriptors representing processes likely to change, e.g. land cover and urbanization, and by weighting each distribution type in case a multimodel approach is selected for representing the regional flood frequency distribution (see e.g. Laio et al., 2009).

5 Conclusions

This study has shown that the inclusion of information on the underlying hydrological processes in the model choice is of high importance. Each catchment has been characterized in terms of size and mean annual precipitation, as these properties have previously been found to be rough surrogates for the different flood generation processes, but also because a survey presented in the companion paper by Salinas et al. (2014) showed that only these, the most elemental catchment properties, are readily available across Europe. Some preliminary conclusions can be drawn, such as the shift from a GEV to a LN3 as a more appropriate distribution for decreasing MAP values in intermediate-sized catchments. However, the robustness of these statements is limited due to the lack of more data from simultaneous extremes of the subclasses (e.g. smaller and drier catchments) but maybe more due to the absence of better catchment attributes that allow us to fully describe the flood generation processes.

Several studies of flood hydrology have also highlighted the potential utility of soil and land-use data for characterizing flood frequency curves in ungauged European catchments. Thus, there are potentially larger benefits associated with future development of consistent pan-European catchment descriptor data sets as a fundamental step in harmonizing methods. In particular for this database, there is a variety of other catchment and climatic descriptors that could potentially improve the analysis. As pointed out in Sect. 2, there is a large number of catchments that could be strongly affected by snow processes. Its control on the flood regimes could be analysed through surrogates such as median (or maximum) catchment elevation, mean annual air temperature or by more specific ones like the fraction of solid to liquid precipitation. Also, having information on a station basis about the dominant event types and the precipitation that most likely triggers the hydrological extremes (convective rainfall, long rain events, etc.) might allow us to have a better understanding of the whole hydrological regime which could be used to more accurately sort the database. For example, a recent study by Gaál et al. (2014) suggested the use of lightning data as an indicator of convectivity in rainfall events. Additionally, there are other climatic indicators related to the energy and water balance, such as the aridity index (ratio between potential evapotranspiration and mean annual precipitation), that could provide together with rainfall and temperature valuable information on the climate type.

The original utilization of the traditional L-moment ratio diagrams presented in this study, in conjunction with a more refined characterization of European catchments based upon a richer catchment descriptor data set, could also contribute to better understanding the modifications in flood hazard resulting from different sources of environmental change, and to move further towards the definition of a set of "process-driven" pan-European flood frequency distributions.

Acknowledgements. The authors are grateful for the financial contribution made by the COST Office grant ES0901 European procedures for flood frequency estimation (FloodFreq) and for contributions made by members of Working Group 2 of the COST Action. Funding from the ERC FloodChange project (ERC Advanced Grant FP7-IDEAS-ERC-AG-PE10 291152) and from the Austrian Science Funds (FWF) as part of the Vienna Doctoral Programme on Water Resource Systems (DK-plus W1219-N22) is also acknowledged. The compilation of the Italian data set was partially supported by the Italian Government through its national grant to the programme on Characterisation of Ungauged Basins by Integrated uSe of hydrological Techniques (CUBIST Project) and the compilation of the Slovakian data set was partially supported by the Agency for Research and Development under contract no. APVV-0303-11. Both supports are greatly acknowledged. Referees F. Laio, R. Romanowicz, S. Grimaldi, and V. Iacobellis, as well as Günter Blöschl, are thankfully acknowledged for their constructive comments and suggestions. Juraj Parajka is greatly acknowledged for preparing the map in Fig. 1.

Edited by: R. Woods

References

Blöschl, G. and Sivapalan, M.: Scale issues in hydrological modelling: A review, Hydrol. Process., 9, 251–290, doi:10.1002/hyp.3360090305, 1995.

Blöschl, G. and Sivapalan, M.: Process controls on regional flood frequency: coefficient of variation and basin scale, Water Resour. Res., 33, 2967–2980, doi:10.1029/97WR00568, 1997.

Bogdanowicz, E.: Multimodel approach to estimation of Extreme Value distribution quantiles (in Polish), Hydrology in Eng. and Water Management (Hydrologia w Inżynierii i Gospodarce Wodnej), edited by: Wiezik, B., Polish Academy of Sciences, Environmental Eng. Committee, Monographs, 68, 57–70, 2010.

Brath, A., Castellarin, A., and Montanari, A.: Assessing the reliability of regional depth-duration-frequency equations for gaged and ungaged sites, Water Resour. Res., 39, 1367, doi:10.1029/2003WR002399, 2003.

Castellarin, A., Kohnová, S., Gaál, L., Fleig, A., Salinas, J. L., Toumazis, A., Kjeldsen, T. R., and Macdonald, N.: Review of applied statistical methods for flood frequency analysis in Europe, Milestone Report for WG2 of COST Action ES0901, Tech. rep., Centre for Ecology & Hydrology, Wallingford, UK, on behalf of COST, available at: (http://www.cost.eu/media/publications/), 2012.

Di Baldassarre, G., Castellarin, A., and Brath, A.: Relationships between statistics of rainfall extremes and mean annual precipitation: an application for design-storm estimation in northern central Italy, Hydrol. Earth Syst. Sci., 10, 589–601, doi:10.5194/hess-10-589-2006, 2006.

El Adlouni, S., Bobée, B., and Ouarda, T. B. M. J.: On the tails of extreme event distributions in hydrology, J. Hydrol., 355, 1–4, doi:10.1016/j.jhydrol.2008.02.011, 2008.

Farquharson, F. A. K., Meigh, J. R., and Sutcliffe, J. V.: Regional flood frequency analysis in arid and semi-arid areas, J. Hydrol., 138, 487–501, 1992.

Gaál, L., Szolgay, J., Kohnová, S., Parajka, J., Merz, R., Viglione, A., and Blöschl, G.: Flood timescales: understanding the interplay of climate and catchment processes through comparative hydrology, Water Resour. Res., 48, W04511, doi:10.1029/2011WR011509, 2012.

Gaál, L., Molnar, P., and Szolgay, J.: Selection of intense rainfall events based on intensity thresholds and lightning data in Switzerland, Hydrol. Earth Syst. Sci., 18, 1561–1573, doi:10.5194/hess-18-1561-2014, 2014.

Hosking, J. R. M. and Wallis, J. R.: Regional Frequency Analysis: an Approach Based on L-Moments, Cambridge University Press, Cambridge, UK, 1997.

Iacobellis, V., Claps, P., and Fiorentino, M.: Climatic control on the variability of flood distribution, Hydrol. Earth Syst. Sci., 6, 229–238, doi:10.5194/hess-6-229-2002, 2002.

Kjeldsen, T. R.: COST Action ES0901: European procedures for flood frequency estimation (FloodFreq), Geophys. Res. Abs., 13, EGU2011–10327, 2011.

Kochanek, K., Strupczewski, W. G., and Bogdanowicz, E.: On seasonal approach to flood frequency modelling. Part II: flood frequency analysis of Polish rivers, Hydrol. Process., 26, 717–730, doi:10.1002/hyp.8178, 2012.

Laio, F.: Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters, Water Resour. Res., 40, W09308, doi:10.1029/2004WR003204, 2004.

Laio, F., Di Baldassarre, G., and Montanari, A.: Model selection techniques for the frequency analysis of hydrological extremes, Water Resour. Res., 45, W07416, doi:10.1029/2007WR006666, 2009.

Medici, C., Butturini, A., Bernal, S., Vázquez, E., Sabater, F., Vélez, J. I., and Francés, F.: Modelling the non-linear hydrological behaviour of a small Mediterranean forested catchment, Hydrol. Process., 22, 3814–3828, doi:10.1002/hyp.6991, 2008.

Meigh, J. R., Farquharson, F. A. K., and Sutcliffe, J. V.: A worldwide comparison of regional flood estimation methods and climate, Hydrol. Sci. J., 42, 225–244, doi:10.1080/02626669709492022, 1997.

Merz, R. and Blöschl, G.: A process typology of regional floods, Water Resour. Res., 39, 1340, doi:10.1029/2002WR001952, 2003.

Merz, R. and Blöschl, G.: Process controls on the statistical flood moments – a data based analysis, Hydrol. Process., 23, 675–696, doi:10.1002/hyp.7168, 2009.

Padi, P. T., Di Baldassarre, G., and Castellarin, A.: Floodplain management in Africa: Large scale analysis of flood data, Phys. Chem. Earth, 36, 292–298, doi:10.1016/j.pce.2011.02.002, 2011.

Peel, M. C., Wang, Q. J., Vogel, R. M., and McMahon, T. A.: The utility of L-moment ratio diagrams for selecting a regional probability distribution, Hydrol. Sci. J., 46, 147–156, doi:10.1080/02626660109492806, 2001.

Robson, A. J. and Reed, D. W.: Flood Estimation Handbook (FEH), chap. Statistical procedures for flood frequency estimation, p. Vol. 3, Institute of Hydrology, Wallingford, UK, 1999.

Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies, Hydrol. Earth Syst. Sci., 17, 2637–2652, doi:10.5194/hess-17-2637-2013, 2013.

Salinas, J. L., Castellarin, A., Viglione, A., Kohnová, S., and Kjeldsen, T. R.: Regional parent flood frequency distributions in Europe – Part 1: Is the GEV model suitable as a pan-European parent?, Hydrol. Earth Syst. Sci., 18, 4381–4389, doi:10.5194/hess-18-4381-2014, 2014.

Schaefer, M. G.: Regional analyses of precipitation annual maxima in Washington State, Water Resour. Res., 26, 119–131, doi:10.1029/WR026i001p00119, 1990.

Sivapalan, M. and Blöschl, G.: Transformation of point rainfall to areal rainfall: intensity-duration-frequency curves, J. Hydrol., 204, 150–167, doi:10.1016/S0022-1694(97)00117-0, 1998.

Sivapalan, M., Jothityangkoon, C., and Menabde, M.: Linearity and nonlinearity of basin response as a function of scale: discussion of alternative definitions, Water Resour. Res., 38, 4–1–4–5, doi:10.1029/2001WR000482, 2002.

Sivapalan, M., Blöschl, G., Merz, R., and Gutknecht, D.: Linking flood frequency to longterm water balance: Incorporating effects of seasonality, Water Resour. Res., 41, W06012, doi:10.1029/2004WR003439, 2005.

Thompson, E. M., Baise, L. G., and Vogel, R. M.: A global index earthquake approach to probabilistic assessment of extremes, J. Geophys. Res., 112, B06314, doi:10.1029/2006JB004543, 2007.

Viglione, A.: Confidence intervals for the coefficient of L-variation in hydrological applications, Hydrol. Earth Syst. Sci., 14, 2229–2242, doi:10.5194/hess-14-2229-2010, 2010.

Viglione, A., Castellarin, A., Rogger, M., Merz, R., and Blöschl, G.: Extreme rainstorms: comparing regional envelope curves to stochastically generated events, Water Resour. Res., 48, W01509, doi:10.1029/2011WR010515, 2012.

Vogel, R. M. and Fennessey, N.: L moment diagrams should replace product moment diagrams, Water Resour. Res., 29, 1745–1752, doi:10.1029/93WR00341, 1993.

Vogel, R. M. and Wilson, I.: The probability distribution of annual maximum, minimum and average streamflow in the United States, J. Hydrol. Eng., 1, 69–76, doi:10.1061/(ASCE)1084-0699(1996)1:2(69), 1996.

Chapter 5

Fuzzy Bayesian flood frequency estimation with historic data

The present chapter corresponds to the following scientific publication, to be submitted:

Salinas, J. L., Kiss, A., Viglione, A., Viertl, R., and Blöschl, G.: A Fuzzy Bayesian approach to flood frequency estimation with historical information, to be submitted to *Water Resources Research*, 2015.

Abstract

Efforts of the historical environmental extremes community during the last decades have resulted in the obtention of long time series of historical floods, which in some cases range longer than 500 years in the past. In hydrological engineering, historical floods are useful because they give additional information which improves the estimates of discharges with low annual exceedance probabilities, i.e. with high return periods, and additionally might reduce the uncertainty in those estimates. In order to use the historical floods in formal flood frequency analysis, the precise value of the peak discharges would ideally be known, but in most of the cases, the information related to historical floods is given, quantitatively, in a non-precise manner.

This work presents an approach on how to deal with the non-precise historical floods, by linking the descriptions in historical records to fuzzy numbers representing discharges. These fuzzy historical discharges are then introduced in a formal Bayesian inference framework, taking into accoung the arithmetics of non-precise numbers modelled by fuzzy logic theory, to obtain a fuzzy version of the flood frequency curve combining the fuzzy historical flood events and the intrumental data for a given location. Two case studies are selected from the historical literature, representing different facets of the fuzziness present in the historical sources. The results from the cases studies are given in the form of the fuzzy estimates of the flood frequency curves together with the fuzzy 5% and 95% Bayesian credibility bounds for these curves.

The presented fuzzy Bayesian inference framework provides a flexible methodology to propagate in an explicit way the imprecision from the historical records into the flood frequency estimate,

which allows to assess the effect that the incorporation of non-precise historical information can have in the flood frequency regime.

5.1 Introduction

Due to recent great flood events during the last decades, an increased scientific and media interest turns attention towards historical floods, both in terms of individual past extremes and the understanding of long-term flood behaviour. Following this interest, a rapidly growing amount of scientific literature is available in the subject of historical records of flood events in a diversity of spatial and temporal scales. Examples range from a compilation of flood evidences during the middle-ages in the entire Carpathian basin (*Kiss and Laszlovszky*, 2013), to the reconstruction of the largest floods of the Rhine river at the Swiss city of Basel during the last 750 years (*Wetter et al.*, 2011). European overviews for the main research results applied to specific regions can be found in e.g. *Brázdil et al.* (2006, 2012); *Glaser et al.* (2010); *Hall et al.* (2014). Information about these historical floods often provides a very useful insight for flood risk assessments, given the extended lenght of the period considered, and can and should be used in flood frequency estimation (*Kjeldsen et al.*, 2014).

In the context of flood frequency hydrology (see *Merz and Blöschl*, 2008a,b; *Viglione et al.*, 2013), temporal information expansion refers to the collection of information on the flood behavior outside the period of discharge observations (i.e. outside the systematic or instrumental data period). There are formal statistical methods to combinine information about historical floods with available flood discharge measurements (e.g., *Leese*, 1973; *Stedinger and Cohn*, 1986; *Cohn et al.*, 1997; *O'Connell et al.*, 2002; *England et al.*, 2003; *Reis and Stedinger*, 2005; *Benito and Thorndycraft*, 2005), which would be considered temporal information expansion.

In this paper, the Bayesian framework is chosen, as it combines different sources of information obtaining a probabity distribution for the parameters of a given statistical model, giving therefore information about their predictive uncertainty. In flood hydrology, Bayesian methods have been used in the literature to incorporate additional information such as historic floods (e.g., *Stedinger and Cohn*, 1986; *O'Connell et al.*, 2002; *Parent and Bernier*, 2003; *Reis and Stedinger*, 2005; *Neppel et al.*, 2010; *Payrastre et al.*, 2011) with the aim of fitting a flood frequency distribution, either regional or at-site.

Information about historical floods presents itself almost always, from the quantitative point of view, in a non-precise way, and would require previous treatment to be used in formal statistical methods, such as Bayesian inference for flood frequency estimation. In the present study, the mathematical theory of fuzzy numbers is used to model the unprecision of these historical discharges. Fuzzy numbers originated with the generalized concept of indicator function of *Menger* (1951), and the term "fuzzy sets" was first coined by *Zadeh* (1965) in the field of control theory in electrical engineering. Applicatons of fuzzy numbers in water resources have dealt with fuzzy regression analysis (*Bárdossy et al.*, 1990a; *Chachi et al.*, 2014), fuzzy geostatistical approaches (*Bárdossy et al.*, 1988, 1989, 1990c,b), modelling parts of the hydrological cycle (*Bárdossy and Disse*, 1993; *Bárdossy*, 1996; *Schulz and Huwe*, 1997; *Özelkan and Duckstein*, 2001; *Nasseri et al.*, 2014), and classification purposes (*Bárdossy et al.*, 1995; *Bárdossy and Samaniego*, 2002). But maybe the most common application has been in decision making and optimal control theory for water resources planning or flood forecasting (*Esogbue et al.*, 1992; *Shrestha et al.*, 1996; *Russell and Campbell*, 1996; *Bender and Simonovic*, 2000; *Prodanovic and Simonovic*, 2002; *Simonovic and Verma*, 2008; *Bárdossy*, 2008; *Schumann and Nijssen*, 2011). More examples of fuzzy logic applied in water resources related research is presented in *Bogardi et al.* (2003).

The aim of this paper is to exploit the benefit of the additional information that historical records have to offer for flood frequency estimation, even if they are expressed in a non-precise way, and include them in a Bayesian estimation procedure incorporating the aritmetics of fuzzy numbers, something that up to this date has not been reported in the hydrological literature. Section 2 describes, from a historical point of view, the typical sources and nature of historical floods records. Due to their inherent unprecision, a central point of this paper will be on how to link specific information about historical flood events with their fuzzy model, so they can be incorporated in a formal fuzzy Bayesian inference framework (see e.g *Viertl*, 2008a,b); this is done in section 3. Two case studies are presented in section 4, where the methodology described in section 3 is applied to time series, which feature two different aspects of the typical fuzziness that can be present in historical records. Finally, the last section presents a discussion of the results together with the conclusions.

5.2 Historical Data is Fuzzy

Historical time series refer to the pre-instrumental period, developed based on various written documentary evidence such as narratives (e.g. chronicles, annals, diaries), institutional sources namely economic-administrative (accounts) and legal-administrative (e.g. charters, official/administrative letters, notes, reports) evidence, epigraphic evidence (e.g. flood marks, paintings/drawings), media information (e.g. newspapers, pamphlets) etc. Depending on source availability, quality and the regularity of observations and recording practices, flood series might be built based on one source type (e.g. accounts, flood marks, charters/letters etc.), but most of the flood series are based on the combination of various source types (e.g. narratives or the combination of narratives, epigraphic and institutional sources).

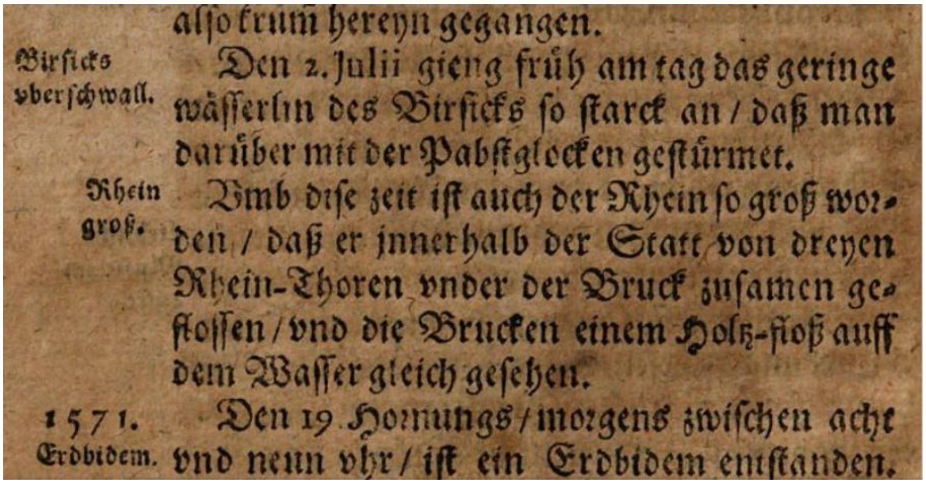
A good example of the fuzzy character of historical documentary evidence is the general terminology applied. In medieval and early modern times texts are often written in Latin where - apart from other important information such as the caused damages, casualties, height/extension of water and duration of flood - in narratives a clear term of an extreme flood of exceptional magnitude is deluge ("diluvium"). This term is not really used before the beginning of the 14th century, and typically used for extreme cases such as 1315, 1342, 1343, 1374 or 1501 (see e.g. *Rohr*, 2004).

While deluge is a term for describing an extreme flood in narrative sources, an extreme flood is sometimes mentioned (for example in narratives and charters) as "inundatio maxima" (very great flood), for example the 1343 Upper-Rhine flood (Generallandesarchiv Karlsruhe Urk. 1345, Sept. 30, GLA. 16/97, Konv. 22), while "nimia inundatio" in the documentation means a great flood: usually great but not extraordinary in magnitude.

In some other cases the word "deluge" (Latin: diluvium) is a clear definition people usually applied in the medieval early modern times (for terminological discussion and other examples, see e.g.

Rohr, 2004)) only for the most exceptional huge flood events. This was, for example, on the Werra and other Thüringen rivers the "Diluvium Thuringiacum" in Latin or the "Thüringishe Sintflut" in German (see *Deutsch and Pörtge*, 2003). Another frequent expression in describing catastrophic floods is "maximae aquarum inundationes" that could be translated as "very great floods of water" (e.g. February 1342, by Franciscus Pragensis: Loserth 1875). A similarly simple but direct way of flood characterisation is when basic damage measures are applied, such as in summer 1275: "Festo Petri et Pauli Rhenus pontem Basiliensem destruxit, submersis plus minus 100 hominibus." (source published in Latin: Pertz 1861) - which reference may define an exceptionally great flood event saying "On the day of Peter and Paul (29 June in Julian Cal.) the bridge of Basel was destroyed, and around 100 people submerged" (thus, with no precise information in maximum water levels). Similarly large floods, suggested to be of extraordinary in magnitude, are described as the one that has not seen or heard about in human memory (i.e. human lifetime): "de a memoria hominum, non visa non audita" for the extreme ice jam flood of February 1775 in Budapest (see *Kiss 2007*; referred in the town protocols of /Buda/Pest: BFL Pest. IV. 1202a. 355a. 22 Feb, 1775). Almost the same expression is used for the catastrophic 1343 summer flood event of the Constance sea (Austrian-German-Swiss border): "quod antea non est visum, ut antiquiores tunc temporis referebant" (reported by J. von Winterthur: *Brun and Baethigen 1924*). Great, but not catastrophic floods may often mentioned as "gross güß" or "nimia/magna/ingens inundatio aquarum" — in both cases in the meaning of "great flood" – and are the most frequent in all document types (e.g. annals, chronicles, accounts, charters, diaries etc. – see e.g. *Rohr 2006, Kiss 2007, 2009, Kiss and Laszlovszky 2013; Wetter et al. 2011*).

Despite describing the extension and duration of the flood, and also its economic-social impacts (e.g. casualties, houses destroyed etc.), other assumptions – reflecting on highest water levels of the flood – like "die Brücken einem Holz-floß auff dem Wasser gleich gesehen" (referring to the 1570 great flood event; see Fig. 1) meaning "the bridge looked like a raft on the water" (Wetter et al. 2011: in the Kurtze Bassler Chronik: Gross 1624, p. 210) or "the town looked like an island in the water" (e.g. *Kiss 2007*: Pressburger Zeitung I March, 1784) often as well appear in documentary evidence, generally helping in estimating the very approximate extension of water and the magnitude of flood. A typical feature, applied for fixing an approximate height of maximum flood levels (of extreme high flood events) in the Basel reconstruction was the impression when standing on the bridge "people could wash hands in the water": "Etiam homines in ponte stantes facile in Reno manus lavare poterant", in this case talking about the extreme flood event of 1424 (mentioned by Kaplan Hieronymus Brilinger, published in: Bernouli 1915).



alſo krum hereyn gegangen.
Birſicks vberſchwall. Den 2. Julii gieng früh am tag das geringe wäſſerlin des Birſicks ſo ſtarck an / daß man darüber mit der Pabſtglocken geſtürmet.
Rhein groß. Vmb diſe zeit iſt auch der Rhein ſo groß worden / daß er innerhalb der Statt von dreyen Rhein-Thoren vnder der Bruck zuſamen gefloſſen / vnd die Brucken einem Holtz-floß auff dem Waſſer gleich geſehen.
1571. Erdbidem. Den 19 Hornungs / morgens zwiſchen acht vnd neun vhr / iſt ein Erdbidem entſtanden.

Figure 1. Original archival record from the "Kurtze Bassler Chronik" containing the description for the 1570 river Rhine flood event in Basel.

When enough information is available, discharge reconstruction is possible via hydraulic modelling. Examples of such reconstructions are *Benito et al.* (2003); *Brázdil et al.* (2005, 2006); *Herget and Meurs* (2010); *Wetter et al.* (2011); *Elleder et al.* (2013). For an overview on the methodology of reconstructing discharges with historical data, see *Herget et al.* (2014). One of such examples (Basel-series: *Wetter et al.* (2011)) will be applied in the analysis and discussed in more detail in a later part of this paper.

Some of the pre-instrumental, documentary-based, historical flood series are classified into magnitude categories, and each magnitude class receives a numerical value (indices). Although the most usual index classification is based on 3 flood intensity categories/values (see Table 1), depending on the detailedness of the available documentary evidence, there exist 4 and 5 index value classifications, too (e.g. *Rohr*, 2006; *Retsö*, 2014).

Level	Classification	Primary Indicators	Secondary Indicators
1	Smaller, regional flood	Little damage, e.g. fields and gardens close to the river, wood supplies that were stored close to the river are moved to another place	Short flooding
2	Above average, or supra-regional, flood	Damage to buildings and constructions related to the water like dams, weirs, footbridges, bridges and buildings close to the river, like mills, etc.; water in buildings	Flood of average duration; severe damage to fields and gardens close to the river, loss of animals and sometimes people
3	Above average, or supra-regional, flood on a disastrous scale	Severe damage to buildings and constructions related to the water, i.e. dams, weirs, footbridges, bridges and buildings close to the river, like mills etc.; water in buildings. In part, buildings are completely destroyed or torn away by the flood	Duration of flood: several days or weeks; severe damage to fields and gardens close to the river, extensive loss of animals and people; morpho-dynamic processes like sand sedimentation cause lasting damages and change the surface structure

Table 1. Criteria used for classifying documentary evidence on historical floods into indices (see e.g. Sturm et al., 2001; Glaser and Stangl, 2004).

Table 1 provides the most widespread method of classifying the information derived from the flood descriptions available in various source types listed above: a good example for the application of the 3 category index classification is the 500-year long flood series of the Werra river (published in *Mudelsee et al.* (2006); discussed and analysed later in more detail). Although in most historical series, the 3-scaled index classification is applied, in some other studies (when detailed source evidence is available) 4 or even 5-scaled classification are in use (see e.g. *Rohr*, 2006). In these cases, the precise wording described at the beginning of this section (i.e. "diluvium", "inundatio maxima", "gross güß", etc.) is of great help in classifying the events.

In building up historical time series, the best-quality evidence can be found in contemporary sources written around the time or short after the event (and location). If sources are non-contemporary (i.e. written decades or even longer after the event), they can be subject to various types of errors, for example, concerning the dating, magnitude, impacts etc. of an actual event. Therefore, their qualitative analysis prior to converting them into numerical values is of crucial importance. Even if based on good-quality sources, pre-instrumental, historical data are fuzzy: sources are made for various purposes other than precisely recording the hydrological parameters of the actual flood event. Therefore, while some parameters such as seasonality can be often precisely provided, considering other parameters they contain description of a flood event that - in hydrological sense - can be only treated as a non-precise information referring to, for example, the magnitude of flood events in terms of peak discharge.

5.3 A mathematical model for non-precise historical floods

Section 2 has explained in detail the nature of historical hydrological data, in particular information about historical floods. These data is inherently non-precise and we need a mathematical model to capture this characteristic. Fuzzy numbers (*Menger*, 1951; *Zadeh*, 1965) are a model for the unprecison of a certain magnitude, for example an analogue measurment, or more interesting for this paper, the linguistic nature of expressions describing the magnitude of historical floods, or the discharge threshold between indexed historical floods. The contents of this section aim to familiarize the reader with the main mathematical aspects of Bayesian inference with fuzzy numbers, and its application to flood frequency estimation with non-precise historical information. For more details on the topic of statistics with fuzzy numbers, see e.g. *Viertl* (2011a).

5.3.1 Fuzzy numbers, samples and functions

In the context of flood frequency modelling in the presence of fuzzy historical information, we will consider a single non-precise historical discharge as a fuzzy number; a collection of historical events as a fuzzy sample; and the operations (mainly of statistical nature) involving these fuzzy numbers as fuzzy-valued function. This subsection will introduce some theoretical properties of these elements. We can understand a fuzzy number x^*, as a mathematical model for representig a certain magnitude which is non-precise by nature, as opposite to a random variable, which has a precise value, but we do not know it with full certainty. Fuzzy numbers are therefore valid for modelling magnitudes occupying a certain range or interval on the real line. Examples of fuzzy quantities could be time of sunset (we can consider it a time interval), linguistic expressions like "very cold" (will cover a range of temperatures, probably dependent on the context), or like in our case, peak discharges corresponding to historical floods descriptions. Mathematically, a non-precise or fuzzy number x^* is uniquely determined by its characterizing function $\xi(\cdot)$, a real-valued function of one real variable x giving values between 0 and 1. The heuristic interpretation of the characterizing function $\xi_{x^*}(x)$, is the degree of membership of each value $x \in \mathbb{R}$ to the fuzzy number x^*, that's why the characterizing function is also denoted as membership function. For each value between 0 and 1, denoted usually as δ, the δ-cuts of the fuzzy number x^* are defined as $C_\delta(x^*) := \{x \in \mathbb{R} : \xi(x) \geq \delta\}$. These δ-cuts must be constructed by a finite union of compact intervals (i.e. bounded intervals including the limit points), but most commonly consist of a single compact interval $[a_\delta; b_\delta]$ for each δ value. Furthermore, it can be demonstrated, that the characterizing function $\xi(\cdot)$ of a fuzzy number x^* can be reconstructed from entire the family of δ-cuts $C_\delta(x^*)$, with $\delta \in (0; 1]$; i.e. they contain the same information. This is important, because on a practical basis all operations are done on the base of δ-cuts, and the membership function for the result is reconstructed afterwards. In this paper, mostly trapezoidal fuzzy numbers (see Fig. 2) will be used for simplicity.

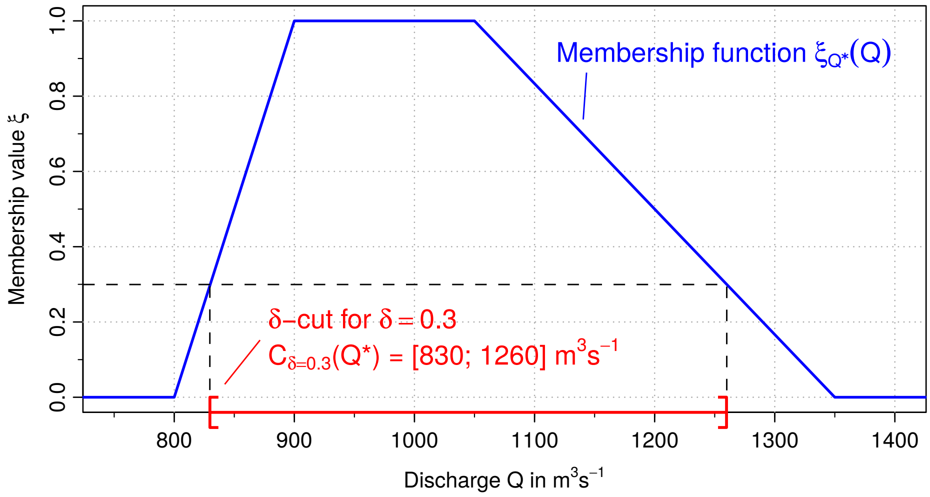


Figure 2. Depiction of a trapezoidal fuzzy discharge Q^*, represented by it membership function $\xi_{Q^*}(\cdot)$. The δ-cut corresponding to $\delta = 0.3$ is shown.

In inferencial statistics, fuzzy numbers usually will not appear isolated. One will deal with a collection of k fuzzy quantities that we can define as fuzzy sample $\mathbf{x}^* = \{x_1^*, x_2^*, \ldots, x_k^*\}$. In order to further operate with the fuzzy sample, it necessary to create a so-called fuzzy valued vector. In practice, this will correspond to establish a mapping between the k characterizing functions $\xi_i(\cdot)$ of the k elements of the fuzzy sample and the vector-characterizing function $\xi(x_1, x_2, \ldots, x_k)$, which is a generalization to k-dimensions of the one dimensional characterizing function defined above. The mapping is done via the so-called triangular norms (or t-norms), which are $\mathbb{R}^k \to \mathbb{R}$ functions that need to fullfill a series of conditions to be well defined, in a similar fashion to the copula functions, or the metric functions in a vector space. It turns out that the most advantageous way, having some desired properties when estimating sample averages and other sample statistics, is to use the minimum t-norm, defined by

$$\xi(x_1, x_2, \ldots, x_k) = \min[\xi_1(x_1), \xi_2(x_1), \ldots, \xi_1(x_k)], \quad \forall (x_1, x_2, \ldots, x_k) \in \mathbb{R}^k$$

One important feature of the minimum t-norm is that the corresponding δ-cuts $C_\delta(\mathbf{x}^*)$ of the fuzzy sample can be computed by simply performing the cartesian product of the δ-cuts $C_\delta(x_i^*)$

of the k fuzzy numbers belonging to the sample, which makes the operations with multiple fuzzy numbers easier.

The mathematical concept of a real-valued function f with k real arguments can also be generalized to a fuzzy-valued function with a k-dimensional fuzzy vector argument via the so-called extension principle, whose exact formulation is not simple to interpret, but has the following consequence for the δ-cuts of the fuzzy value $f(\mathbf{x}^*)$:

$$C_\delta\left[f\left(\mathbf{x}^*\right)\right] = \left[\min_{\mathbf{x}\in C_\delta(\mathbf{x}^*)} f(\mathbf{x})\,;\ \max_{\mathbf{x}\in C_\delta(\mathbf{x}^*)} f(\mathbf{x})\right]$$

Knowing the δ-cuts, one can reconstruct the membership function for $f(\mathbf{x}^*)$, as stated above. Generaliziling, we can say that for any fuzzy-valued function $f^*(\cdot)$, the functional equivalent of δ-cuts are the so-called δ-level functions or curves, defined by $C_\delta\left[f^*(\cdot)\right] = \left[\underline{f}_\delta(\cdot)\,;\bar{f}_\delta(\cdot)\right]$.

Particularizing for monotonic functions with k fuzzy arguments x_i^* (merged into a fuzzy vector $\mathbf{x}^*$ with the minimum t-norm), if we know the δ-cuts $C_\delta\left(x_i^*\right) = \left[a_{\delta,i}; b_{\delta,i}\right]$ for each argument, the δ-level curves of $f^*(\mathbf{x}^*)$ can be evaluated as follows:

$$C_\delta\left[f^*\left(\mathbf{x}^*\right)\right] = \left[\min\left\{f(\mathbf{a}_\delta), f(\mathbf{b}_\delta)\right\};\max\left\{f(\mathbf{a}_\delta), f(\mathbf{b}_\delta)\right\}\right],$$

where $\mathbf{a}_\delta = (a_{\delta,1}, a_{\delta,2}, \ldots, a_{\delta,k})$ and $\mathbf{b}_\delta = (b_{\delta,1}, b_{\delta,2}, \ldots, b_{\delta,k})$. This means, one can calculate the δ-level curves of $f^*(\mathbf{x}^*)$ by evaluating the equivalent real-valued function f at the limits of the δ-cuts for the different elements x_i^* of the fuzzy sample. This property will be used when estimating fuzzy flood frequency distribution with historical data.

Once the main properties for the generalization from real to fuzzy numbers and functions have been explained, the following subsection will deal with the analogous generalization of the Bayesâ theorem.

5.3.2 Fuzzy Bayesian Inference

Bayesian inference is a statistical framework in which the Bayesâ theorem is used to combine prior information with observed data, in order to obtain updated information on the distribution of the parameters of a given model. In flood frequency analysis, this corresponds with combining the information provided by the locally observed flood data with additional information independent from those data, which in our case will be the historical floods. By condidering the parameters of the flood distribution as random variables themselves, the Bayesian framework considers the uncertainty inherent to the different sources of information used, and provides a computationally convenient way to estimate the uncertainty in parameters and quantiles metrics. These are given usually in terms of credibility bounds, the Bayesian counterpart of the confidence intervals in frequentist frameworks.

For a flood frequency distribution with parameters $\boldsymbol{\theta}$, the Bayesâ theorem states that

$$p(\boldsymbol{\theta}|\mathbf{D}) = \frac{l(\mathbf{D}|\boldsymbol{\theta})\cdot\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} l(\mathbf{D}|\boldsymbol{\theta})\cdot\pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}}$$

where $p(\boldsymbol{\theta}|\mathbf{D})$ is the posterior distribution of the parameters $\boldsymbol{\theta}$, after having observed the data $\mathbf{D}$; $l(\mathbf{D}|\boldsymbol{\theta})$ is the likelihood function; $\pi(\boldsymbol{\theta})$ is the prior distribution of the parameters. The integral in the denominator, computed on the whole parameter space $\boldsymbol{\Theta}$, serves as a normalization constant to obtain a unit area under the posterior probability density function $p(\boldsymbol{\theta}|\mathbf{D})$. For a more detailed exposition of the Bayesian approach to flood frequency hydrology, see e.g. *Viglione et al.* (2013).

Fuzzy Bayesian Inference (*Viertl*, 2008a,b) is a generalized framework for Bayesian inference when fuzzy samples $\mathbf{D}^*$ and/or fuzzy prior probability distributions $\pi^*(\boldsymbol{\theta})$ are present, which will be precisely the case in our treatment of non-precise historical floods. In this context, fuzzy probability distributions will be fuzzy-valued functions, as defined in the previous section, with some normalizing properties, analogous to the unit integral over the entire parameter for traditional probability distributions. Denoting the fuzzy posterior probabilty distribution as $p^*(\boldsymbol{\theta}|\mathbf{D}^*)$, the fuzzy-valued likelihood of the fuzzy sample $\mathbf{D}^*$ as $l^*(\mathbf{D}^*|\boldsymbol{\theta})$, and the fuzzy prior probability distribution as $\pi^*(\boldsymbol{\theta})$, the generalized Bayesâ theorem reads for their respective δ-level curves

$$\underline{p}_\delta(\boldsymbol{\theta}|\mathbf{D}^*) = \frac{\underline{l}_\delta(\mathbf{D}^*|\boldsymbol{\theta}) \cdot \underline{\pi}_\delta(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} \frac{1}{2}\left[\underline{l}_\delta(\mathbf{D}^*|\boldsymbol{\theta}) \cdot \underline{\pi}_\delta(\boldsymbol{\theta}) + \overline{l}_\delta(\mathbf{D}^*|\boldsymbol{\theta}) \cdot \overline{\pi}_\delta(\boldsymbol{\theta})\right] \mathrm{d}\boldsymbol{\theta}}$$

The expression for the upper δ-level curve $\overline{p}_\delta(\boldsymbol{\theta}|\mathbf{D}^*)$ is analogous, i.e. taking the upper δ-level curves in the numerator and keeping the same denominator. The normalizing contant must be equal for $\overline{p}_\delta(\boldsymbol{\theta}|\mathbf{D}^*)$ and $\underline{p}_\delta(\boldsymbol{\theta}|\mathbf{D}^*)$, in order to keep the sequential nature of the updating procedure in Bayesâ theorem (see e.g. *Viertl*, 2011b).

From the point of view of application, since the integral in the denominator cannot be generally expressed in closed form, simulation-based Monte Carlo sampling techniques such as the Markov chain Monte Carlo (MCMC) approaches are used. MCMC methods (including random walk Monte Carlo methods) are a class of algorithms for sampling directly from probability distributions by letting a Markov chain evolve, whose steady-state distribution is the desired one; here the interest lies on the posterior probability model (see e.g. *Robert and Casella*, 2004; *Gelman et al.*, 2004). Several MCMC algorithms have been used in flood frequency hydrology (e.g., *Kuczera*, 1999; *Reis and Stedinger*, 2005; *Ribatet et al.*, 2007). In this paper, we use the delayed rejection and adaptive Metropolis algorithm (*Haario et al.*, 2006; *Soetaert and Petzoldt*, 2010) to obtain the δ-level curves of the fuzzy posterior probability density function $p^*(\boldsymbol{\theta}|\mathbf{D}^*)$.

As a general formulation, the fuzzy likelihood function will be expressed as

$$l^*(\mathbf{D}^*|\boldsymbol{\theta}) = l_H^*(\mathbf{D}^*_{\mathbf{hist}}|\boldsymbol{\theta}) \cdot l_S(\mathbf{D}_{\mathbf{syst}}|\boldsymbol{\theta})$$

where $\mathbf{D}_{\mathbf{hist}}{}^*$ will correspond to the fuzzy sample containing the non precise historical discharges, and $\mathbf{D}_{\mathbf{syst}}$ stands for the systematic series, i.e. the discharges directly measured during the instrumental period. Note that the function $l_H^*(\cdot)$ is fuzzy-valued, while $l_S(\cdot)$ is not, as the fuzzyness of the systematic discharges are considered negligible if compared to the historical period.

Conceptualizing single historical events as fuzzy numbers representing the fuzzy peak discharges, and merging them into a fuzzy sample, the fuzzy likelihood function for historical discharges, simply a generalization of the non-fuzzy case as used in the literature (see e.g. *Stedinger and Cohn*, 1986; *Neppel et al.*, 2010; *Viglione et al.*, 2013), can be evaluated as

$$l_H^*(\mathbf{D}^*_{\mathbf{hist}}|\boldsymbol{\theta}) = \binom{h}{k} F_Q(Q_0|\boldsymbol{\theta})^{(h-k)} \prod_{j=1}^{k} f_Q^*(Q_j^*|\boldsymbol{\theta})$$

where h stands for the number of years corresponding to the historical period considered, k is the fuzzy sample size (i.e. number of fuzzy historical discharges), $F_Q(\cdot)$ is the cumulative distribution function for the peak discharges, Q_0 is a discharge threshold (only historical floods exceeding this thresholds are recorded), and $f_Q^*(\cdot)$ is the probability density function for the peak discharges, taken as a fuzzy function, as it is evaluated at each historical fuzzy peak discharge Q_j^*. The methodology on how to obtain the membership functions describing the fuzzy historical discharges represents the core of this study and is fully explained in the next subsection. In this paper, the Generalized Extreme Value (GEV) distribution is used as the statisical model for the frequency of peak annual discharges, and the cumulative distribution function has then the expression

$$F_Q(q|\boldsymbol{\theta}) = \exp\left[-\left(1-\theta_3\cdot\frac{q-\theta_1}{\theta_2}\right)^{1/\theta_3}\right]$$

when $\theta_3 \neq 0$, while the distribution converges to the Gumbel or EV1 model for $\theta_3 \to 0$. The GEV model has been chosen, as it has been recently reported in the literature that to better represent the average flood regime of European rivers (see e.g. *Salinas et al.*, 2014b,a). Nevertheless, any other 3-parameter extreme value distribution (e.g. Generalized Pareto, Generalized Logistic, Log-Pearson type 3, ...) could also be used, with local or regional justification, as they offer a similiar degree of flelixibility in the sense that they all possess a shape parameter to adjust the skewness of the flood frequency.

In case the flood perception threshold is considered to vary during the historical period analized, as it is usually the case, one can divide the historical period in q sub-periods with different perception thesholds, such that $h = h_1 + h_2 + \ldots + h_q$, apply the expression above for each sub-period with its own perception threshold, and multiply the values for all subperiods to obtain the expression for the fuzzy likelihood function for the entire period.

The expression of $l_S(\mathbf{D_{syst}}|\boldsymbol{\theta})$ does not require any special treatment, as the systematic discharges are considered as precise (i.e. non-fuzzy), and would correspond to the product of the probability density function $f_Q(\cdot)$ evaluated at all measured peak discharges.

The choice of a prior probability distribution $\pi(\boldsymbol{\theta})$ is a crucial step in every Bayesian inference approach. If one has some kind of prior knowledge about the parameters of the flood frequency distribution coming from any other independent piece of evidence (like a regional distribution of flood frequency parameters, or results from a rainfall-runoff model; see e.g. *Viglione et al.*, 2013), an informative prior can and should be used in a real world application. In this paper, the focus is not on the selection of prior distributions for Bayesian inference frameworks, and therefore, if the information about historical floods is introduced in the likelihood function, a non-informative flat prior is taken, i.e. $\pi(\boldsymbol{\theta}) \propto 1$ for all values of $\boldsymbol{\theta}$.

Nevertheless, in the situation where historical floods are presented in indexed form (as described in sect. 2), introducing the information about the historical floods in form of a fuzzy prior probability distribution presents conceptual advantages, as will be justified in the following subsection, with the rationale on how and why the indexed time series where initially constructed. If one chooses a 3-parameter statistical model, as is the case in this paper with the GEV distribution, an indexed historical floods time series with 3 indices is needed, as a biyective mapping between indices and parameters will be stablished. From a practical point of view, this is not a problem, as historical time series consist almost always at least 3 indices, and in the case of more indices, two or more classes can be merged to obtain the required 3 indices. Then, given a 3 index-valued historical time series, a probability distribution containing information about the flood frequency could be evaluated as follows.

First, one can estimate the mean interarrival times between events inside each index. It can be demonstrated that, for a stationary stochastic process, consisting of a sequence of idependent identically distributed variables (in our case, the flood peak time series, modelled as a sequence of independent GEV realizations), the mean inter-arrival time of events larger than a given threshold has an Inverse-Gamma sampling distribution (see e.g. *Kottegoda and Rosso, 1997*). Given the estimated sample means of the inter-arrival times between events inside each index, we can get an expression for the three Inverse-Gamma sampling distributions. In this context, the sample mean inter-arrival time can be considered an estimator for the return period associated to the discharge value that conceptually in censoring the index-classes. Via e.g. copula modelling, we can construct a trivariate distribution with the obtained Inverse-Gamma's as marginals. This trivariate distribution, which can be interpreted as a trivariate sampling distribution for the three return periods associated to the three threshold discharges, can be uniquely transformed in a trivariate distribution for the parameters of the parent distribution. In this step, the Jacobian of the transformation from return periods $\mathbf{T} = (T_1, T_2, T_3)$ to parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$, needs to be evaluated as a function of the threshold discharges. If one chooses a set of 3 fuzzy discharge thresholds, as will be conceptually justified in the following subsection, the resulting probabilty distribution is a well defined fuzzy-valued probability distribution that can be used as a fuzzy prior in the generalized Bayesâ theorem, incorporating the information about the historical floods. Note that for this approach, the likelihood function is non-fuzzy, as it contains only information about the systematic discharges.

5.3.3 Linking historical records with fuzzy discharges

In the two previous sections we have briefly discussed the theoretical aspects of modelling non-precision of data with fuzzy numbers (sect. 3.1), and set the basis for a Bayesian inference framework with that non-precise information via the generalized Bayesâ theorem (sect. 3.2). The aim of this section is to link the exposed theoretical aspects of fuzzy statistical modelling with the inclusion of non-precise information about historical floods in a formal Bayesian flood frequency analysis, i.e. transforming the descriptions found in historical records into fuzzy peak discharges.

As stated in sect. 2, information about historical floods may present itself in different ways. It is usual, though, that some indication about the maximum water level during the flood is reported in a written way. In flood events affecting settlements, these indications could reference iconical buildings, bridges, squares, fountains, markets. In many cases, the issues reported correspond to some disturbance in a certain economic activity, e.g. a road or bridge damage limiting some sort of supply. Given that for a certain location, some kind of systematism in reporting flood extents and damages exists, one may set a series of locations where the flood extents are mentioned on a regular basis during the period considered.

This is the approach taken for the case study of the Rhine river at Basel (see sect. 4.1). For this particular location, the water level has been systematically reported to reach the locations listed in Fig. 3. The altitudes with respect to the river bed datum for these location can be considered constant, as there is enough evidence that they have not undergone any significant reconstruction (*Wetter et al.*, 2011). Information about the rating curve is given in *Wetter et al.* (2011), as pairs of discharge–water level values obtained with a one-dimensioanl hydraulic model, and are depicted in Fig. 3.

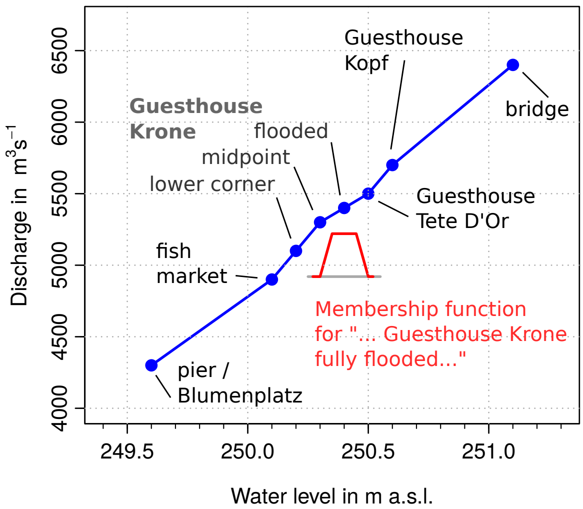


Figure 3. Rating curve (blue line) for the river Rhine at Basel, cross section at the Rheinsbrucke (taken from Wetter et al., 2011). Blue points stand for the locations mentioned systematically in the documentary evidence. An example of the construction of the membership function for the fuzzy water level corresponding to the description "...Guesthouse Krone fully flooded" is depicted in red.

The aim of this paper is not to assess the validity or uncertainties related with the hydraulic modelling itself for the Basel case study, therefore the estimated rating curve is taken as the real one. Therefore, we can focus on constructing fuzzy numbers for the water levels, and then obtain the memebership functions for the fuzzy peak discharges as a fuzzy function(as described in sect. 3.1.) given by the raing curve, with the fuzzy water level as argument. For a series of flood events, there is written evidence that the peak water level reached a certain location, and that the following one had not been flooded. As an example, the event of 1852, where the water level is reported to reach the Guesthouse "Tete D'Or", which allows us to infer that the building itself was not flooded, and therefore gives an indication of the possible maximum value of water level. As a general rule for this kind of information, we could build a trapezoidal characteristic function representing the fuzzy water level as follows. A memebership value of 0 is given to the water levels of the previous and following locations to the one described as flooded. Memebership values of 1 are given within half of the water level ranges between locations, modelling the non-precise written information that if one location has been flooded, e.g. a market square, the water level could range from the beginning until the end of the of the square, as seen from the advancing flood water front. Finally, the lower (and upper) δ-cuts of values 0 and 1 are linked by a linear increasing (and decreasing) membership function, as the modelled fuzzy water level is considered to have a trapezoidal membership function (see a schematic of such construction in Fig. 3). Previous to the year 1480, there is "vague written information", as described in *Wetter* (2012), about some of the peak water levels being reached during the flood events. These descriptions may include several locations, even contradictory in some instances, in their temporal or spatial evoulution. In such cases, a similar construction as the one described might be used, extending the membership function for the fuzzy water level to two, or even three locations, depending on the precision or confidence in the written evidence.

In some instances the indication about the maximum water level could come in terms of some narrative, as described in sect. 2, like "...the bridge looked like a raft on the river", or "...boats were boarded through the windows of the Guildhouse". These descriptions fit perfectly the nature of fuzzy modelling, in terms of being able to process non-precise linguistic expresions, and transform them in a fuzzy number determined by its membership function. Focusing on the example of the 1424 flood event in Basel, where the description "... people could wash their hands in the river", one could construct a fuzzy peak water level attending to a series of considerations. On the one hand, the peak water level almost surely did not exceed the level of the brigde, as it would have been probably reported otherwise. On the other hand, if people could wash their hands in the river while standing on the bridge, the water level could not have been lower than 20-40 cm below the level of the bridge itself. Given these assumptions, one could define a fuzzy water level, as the one depicted in Fig. 4 for that particular event. Other narratives could be transform into membership functions describing peak water levels in an analogous way, taking into account that most of these historical description depend strongly on human perception of the flood magnitude, and can be therefore subject to an extra level of unprecission in the fuzzification process.

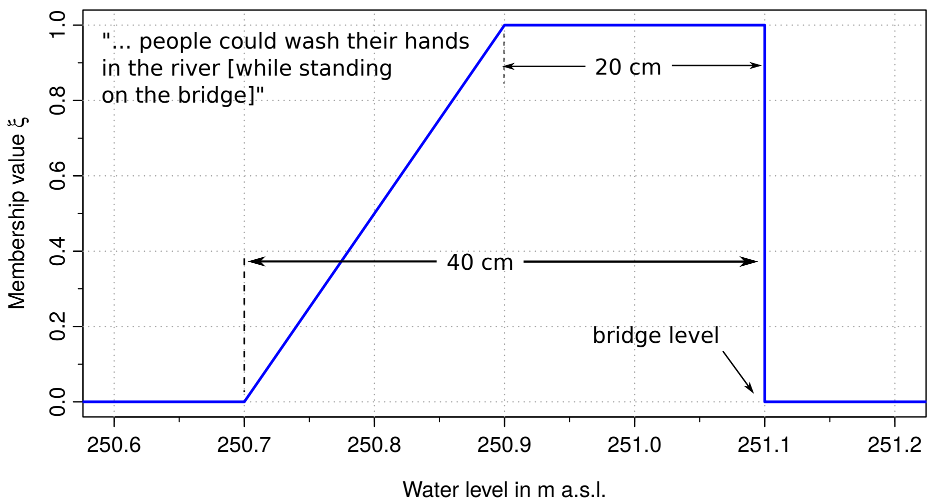


Figure 4. Depiction of the trapezoidal membership function (blue line), representing the fuzzy water level for the flood event of 1424 at Basel, described as "...people could wash their hands in the river [while standing on the bridge]" in the chronicles of Kaplan Hieronymus Brilinger.

As described in sect. 2, in occasions information about historical floods in presented in indexed form. These indexed are based on combination of economical/infrastructre damage, flood duration in time, extension of the flooded areas, etc (see Table 2). If an overlapping period exists, between the instrumental period (where measured peak discharges can be retreived) and the indexed historical series, one can establish a relationship between peak discharges and flood indices. This relationship must, and most probably will not be consistent, as the classification has been made in base of more indicators than peak discharges alone. It could happen therefore, that a certain peak discharge could belong to one flood class in some cases, and to another class in other occasions, depending on the rest of the indicators considered (i.e. damage, duration, or extension). Conceptually, we can model the nature of this imprecision in the discharge threshold between flood indices as fuzzy thresholds defined by membership functions. In this manner, and follwing the construction described in sect. 3.2, a trivariate fuzzy probabilty distribution for the parameters of the flood frequency curve can be obtained, which will be used as a prior in the fuzzy Bayesian inference framework.

The construction of the fuzzy discharge thresholds between flood indices can be performed in principle in many different ways. One simple methodology is described based on the case study of the Werra river at Meiningen (see sect. 4.2). Taking as subsample, the peak annual discharges from the overlapping period, we can assign to each value an index of 0, 1, 2 and 3 (an index of 0 represents flood peaks that have not been classified as 1, 2, or 3). Then, the discharge averages for each index are estimated. These values will correspond to membership values of 0 for the fuzzy thresholds. For example, the fuzzy threshold between indices 1 and 2, as can be seen in Fig. 5, is a trapezoidal fuzzy number ranging from 141 m^3s^{-1} (average discharge for index 1) and 170 m^3s^{-1} (average discharge for index 2). The fact of giving these two discharges a 0 value for the characteristic function of the fuzzy threshold between 1 and 2, can be conceptually understood as giving a possibility of 0 membership to the threshold, to discharge values lower (or larger) than the central values of each class. The rest of the trapezoidal fuzzy threshold, i.e. the δ-cut for δ equal to 1, is computed as one third (lower bound) and two thirds (upper bound) of the range defined by the index averages.

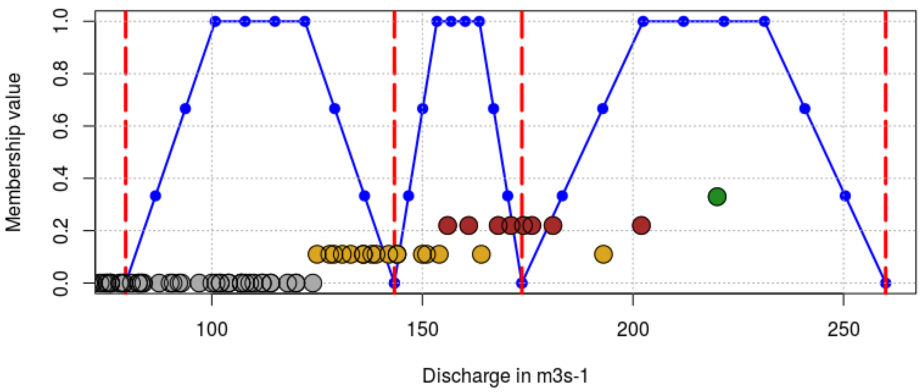


Figure 5. Trapezoidal membership functions (blue lines) for the three fuzzy discharge thresholds between flood indices. The data corresponds to the river Werra case study (see sect. 4.2). Grey, dark yellow, dark red, and green circles depict floods with index values 0, 1, 2, and 3, respectively. Red dashed vertical lines represent the average discharges for each flood index.

This approach for fuzzifying the discharge threshold has been taken for the sake of simplicity. More refined mathematical techniques could be applied; e.g. compute the kernel density estimation (*Rosenblatt*, 1956) for the discharges belonging to the different indexes, and construct the fuzzy thresholds taking into account the overlapping area of the densities. This particular tecnique has been applied to the data from the Werra case study, and the results in terms of fuzzy thesholds did not differ more than a 5%. Therefore, and given that for application purposes more complex mathematical methods could present more obstacles, the first approach described (i.e. the use of the averages in each flood index class) is recommended.

The following section presents the application of the described methodologies of (a) obtaining a fuzzy sample from historical descriptions, and (b) constructing a fuzzy prior probabilty distribution based on historical flood indices, to the case studies of the river Rhine at the swiss city of Basel and the river Werra at the german city of Meiningen, respectively.

5.4 Case Studies

5.4.1 Reconstruction based on maximum water levels descriptions since year 1256 – River Rhine at Basel (Switzerland)

The flood discharge (and index) series of the town of Basel, published by Wetter et al. (2011; also Wetter 2012) starting from the mid-13th century, is further applied for fuzzy analysis in the present study. This series forms an excellent basis of fuzzy investigations, because the discharge values were mainly calculated based on definite descriptive information related to the maximum level of flood event (e.g. people on the bridge could wash their hands in the water, or the water reached specific streets/squares or buildings in defined heights), and this information is available in published form and/or in the form of source references. Due to the fact that the river bed had no significant changes (i.e. not biased by sedimentation process), the height of the bridge was suggested to be around the same level throughout the study period, and the roughness of the built-up area in the town centre did not significantly change over time, basic environmental characteristics remained approximately the same over time. This is another characteristics that made this long historical time series suitable for further application in fuzzy Bayesian analysis.

From the methodological point of view, membership functions based on the records were given to single historical events, as described in sect 3.3. In this way, a fuzzy sample containing 36 historical floods is combined with a non-fuzzy sample containing the measured peak annual discharges from the available instrumental period (1869-2010). During the historical period (from 1256 to 1867), the flood perception threshold is considered to have varied four times, given the kind of evidence and sources presented in *Wetter et al.* (2011). Until the year 1500, the threshold discharge is taken as 5611 m^3s^{-1}, from 1501 to 1650 as 5261 m^3s^{-1}, from 1651 to 1780 as 4816 m^3s^{-1}, and from 1781 to 1868 as 4300 m^3s^{-1}. In all four cases, the perception threshold is taken as the lower bound for the smallest fuzzy discharge during each sub-period.

It is reported in *Wetter et al.* (2011) that two important river diversions reducing the magnitude of floods at Basel took place in the years 1714 and 1877. In the same study an average peak discharge reduction of 630 and 900 m^3s^{-1} was assessed via hydraulic modelling for only one and the two diversions simulataneously. In order to apply the fuzzy Bayesian inference framework described in sect. 3, the entire time series has been harmonized to present day conditions, simply by substracting the estimated peak discharge reduction in the corresponding period. The full mixed data sample (fuzzy and non-fuzzy) is depicted in Fig. 6, with the original time series in blue and the harmonized version in red.

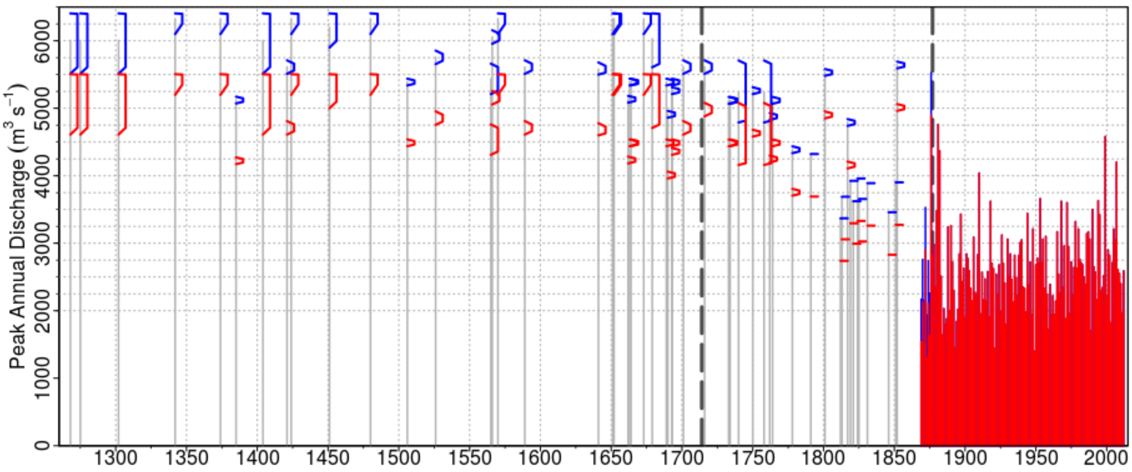


Figure 6. Fuzzy sample containing the historical floods (period 1256-1867; membership functions depicted vertically), together with the systematic peak discharges time series, river Rhine at Basel, cross section Rheinsbrucke. Timing of the 2 main hydraulic works (river diversions) shown as vertical dashed lines. Color blue stands for original reported/measured discharges, red includes the correction due to river diversions.

Figure 7 shows the fuzzy flood frequency curve (different δ-level curves represented by blue transparent polygons), with its fuzzy 5% and 95% Bayesian credibility bounds (cyan transparent polygons), resulting from the application of the fuzzy Bayesian inference process described in sect. 3.

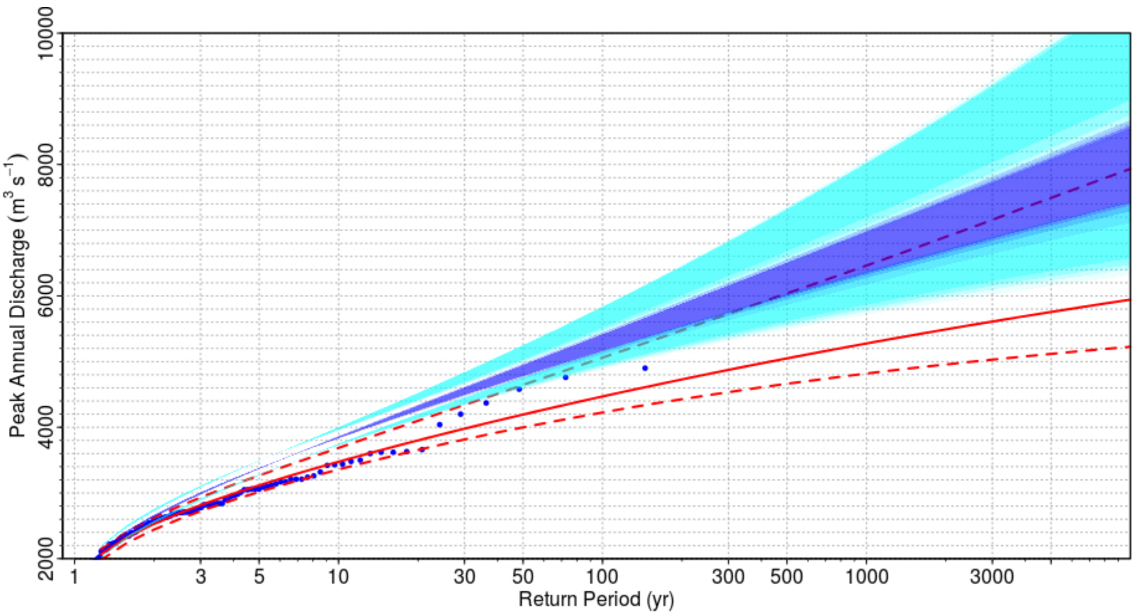


Figure 7. δ-level curves with $\delta = 0,\ 0.33,\ 0.66,$ and 1, for the fuzzy posterior flood frequency curve (blue polygons) and the 5% and 95% fuzzy Bayesian credibility bounds (cyan polygons) using fuzzy historical floods for the river Rhine at Basel. Non-fuzzy flood frequency curve obtained from using

only the systematic data (dark blue dots) is depicted as a red solid line, with 5% and 95% credibility bounds drawn as red dashed lines.

5.4.2 500-year series based on mixed source evidence – River Werra at Meiningen (Germany)

The exceptionally good quality historical flood series of the River Werra (in the eastern part of Germany), that formed the basis of the analyses carried out by *Mudelsee et al.* (2006), were developed by *Deutsch and Pörtge* (2003) (see also *Deutsch et al.*, 2004), based on various types of local sources (e.g. chronicles, annals, diaries, newpapers, pamphlets etc.). Based on the index classification methodology provided in Table 1, a 3-scaled flood magnitude classification was obtained. This series is available in publication in its original, indexed form (see in *Mudelsee et al.*, 2006).

The time series contains a total of 128 indexed floods from the year 1500 until 2003. Systematic discharge measurments were available for the period 1918-2012, therefore the overlapping period ranges from 1918 to 2003. For an indexed historical floods time series, as presented in sect. 3.2, one can include the information about the historical floods in the form of a fuzzy prior probability distribution. This has been the case for this example, where the fuzzy thresholds, obtained by the methodology described in sect. 3.3, were depicted in Fig. 5.

Figure 8 shows the results of the fuzzy Bayesian inference process of combining the systematic flood sample and a fuzzy prior probability distribution containig the information of the historical flood indices. These results are again in terms of a series of δ-level curves for the fuzzy posterior flood frequency curve and the fuzzy 5% and 95% credibility bounds (color code as in Fig.7)

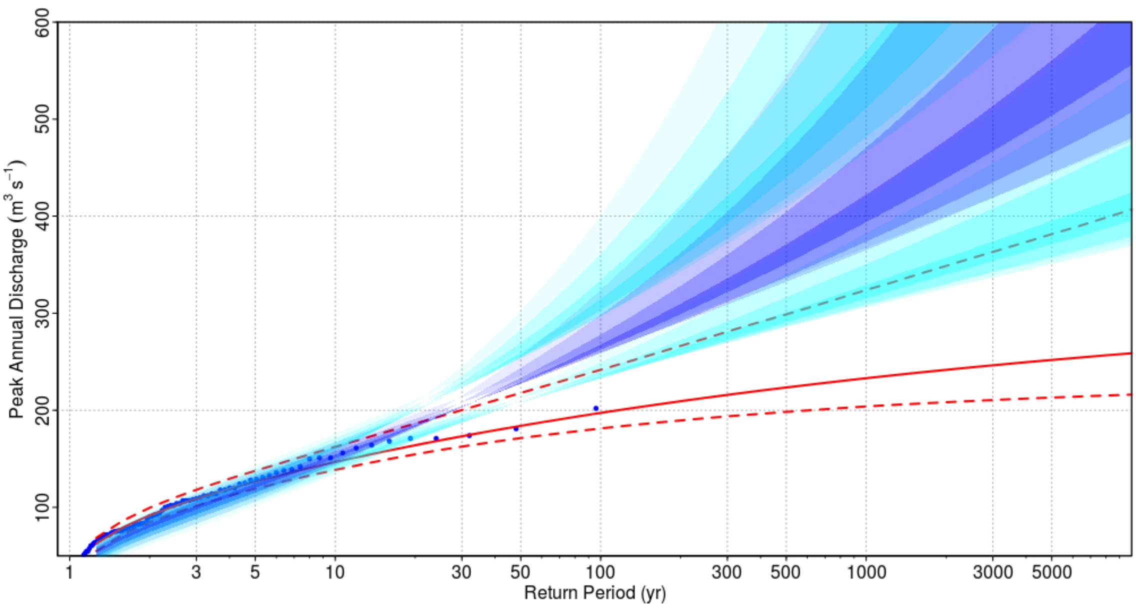


Figure 8. δ-level curves with $\delta = 0,\ 0.33,\ 0.66,$ and 1, for the fuzzy posterior flood frequency curve (blue polygons) and the 5% and 95% fuzzy Bayesian credibility bounds (cyan polygons) using fuzzy historical floods for the river Werra at Meiningen. Non-fuzzy flood frequency curve obtained from

using only the systematic data (dark blue dots) is depicted as a red solid line, with 5% and 95% credibility bounds drawn as red dashed lines.

5.5 Discussion and Conclusions

The methodology proposed for tranforming historical records into fuzzy numbers representing peak discharges during historical flood events, has shown to flexible in terms of being able to adapt to various types of linguistic evidence. In one of the case studies analized, fuzzy numbers for peak discharges could be obtained from historical records describing floodings reaching certain locations. The particular kind of descriptions describing the evidences of floodings need to be carefully studied for each case study before transforming them into fuzzy information, as they can vary strongly between regions, or even for similar locations different types of historical sources can differ in the way they report the flood events. The obtention of all the documentary and archival information usually requires a considerable amount of specialised work, therefore it is recommended to use historical series already published, assuming that their validity has been checked. Trying to find a general rule to construct fuzzy numbers from historical information, represents itself a challenge from a scientific point of view, as the degree of fuzziness present in the records changes from case to case. In this sense, this paper pretends to show the viability of the particular methodology applied in the first case study to obtain a fuzzy sample of historical floods, which fullfill the mathematical requirements to be used in a fuzzy Bayesian inference framework, while or other cases of interest, the approach might and possibly should be changed and adapted in order to succesfully retreive a well defined time series of fuzzy historical floods.

For the case study representing the indexed flood time series, a simple averaging approach has been defined to the fuzzy discharge thresholds between flood classes, but has been done so for the sake of applicability. The only general requirement of this type of historical flood time series, is an overlapping period between the historical (i.e. indexed) and the instrumental period (i.e. with measured peak discharges). In this sense, the approach used for constructing the fuzzy prior probability distribution with the historical information is more straightforward to generalize to analogous indexed time series.

The final results for both case studies are presented in terms of a fuzzy estimate of the flood frequency curve, plus the fuzzy 5% and 95% Bayesian credibility bounds. Analogous studies from the literature, which combine systematic with historical discharges in a non-fuzzy Bayesian framework, also report the results in the same terms, but in their non-fuzzy counterparts (see e.g. *Gaume et al.*, 2010; *Neppel et al.*, 2010; *Viglione et al.*, 2013).

The fuzzy Bayesian inference framework applied for flood frequency estimation in presence of non-precise historical floods presents two main novelties, if compared with the existent flood hydrology literature. On the one hand, it allows the inclusion of information about historical floods events defined by an arbitrary expression for their imprecision, described by memebership functions of peak discharges. Note that in this study, mainly trapezoidal characteristic functions have been used for simplicity, but any other analytic expresion giving normalised values between 0 and 1 can be used. So far in the literature, only precisely known historical discharges, or information about a lower and upper bound for historical discharges, was included in formal methods for flood frequency estimation. Fuzzy modelling offers therefore a higher flexibility in representing all the information that can be retreived from the historical records via membership functions, than just a lower and upper bound. This will help to better exploit the information on historical flood events to be included in the Bayesian framework.

On the other hand, the representation of non-precise historical discharges with fuzzy number, and their incorporation in the fuzzy Bayesian inference, depicts in an explicit way the propagation of imprecision from the fuzzy imput (non-precise historical floods) to the fuzzy output (δ-level curves for the fuzzy posterior flood frequency curve and the fuzzy 5% and 95% credibility bounds), which has not been reported before in the flood frequency literature. This propagation of imprecision or fuzziness represents the entire range of possible values that the output (e.g. a fuzzy 100-yr flood quantile) can take, with different membership values, as opposed to the traditional probabilistic framework, in which the uncertainty in the inputs (expressed in terms of error probability distributions) propagates into the outputs. Heuristically, the propagation of fuzziness could be understood as the propagation of a perfectly correlated error structures, where the input membership functions would play the role of the input uncertainty distributions. While the overall ranges obtained with the fuzzy Bayesian inference procedure could not be appropiate for flood design purposes, where a precise value is required, they give a valuable insight on the effect that the incorporation of non-precise historical information can have in the flood frequency regime.

Future lines of research related with this study, could include e.g. developing some kind of formal set of rules or mapping between non-precise descriptions of historical flood events and their fuzzy numbers, in a similar fashion as performed in the present paper, but in more general grounds, so they can be applied in a wider variety historical records. These kind of analysis would require a closer collaboration between hydrologists and environmental historians, exploiting the potential synergies and feedbacks between the two disciplines.

5.6 References

Bárdossy, A. (1996), The use of fuzzy rules for the description of elements of the hydrological cycle, *Ecological Modelling*, *85*(1), 59–65, doi:10.1016/0304-3800(95)00011-9, fuzzy Logic in Ecological Modelling.

Bárdossy, A. (2008), Fuzzy Rule-Based Flood Forecasting, in *Practical Hydroinformatics*, *Water Science and Technology Library*, vol. 68, edited by R. J. Abrahart, L. M. See, and D. P. Solomatine, chap. 13, pp. 177–187, doi:10.1007/978-3-540-79881-1_13.

Bárdossy, A., and M. Disse (1993), Fuzzy rule-based models for infiltration, *Water Resources Research*, *29*(2), 373–382, doi:10.1029/92WR02330.

Bárdossy, A., and L. Samaniego (2002), Fuzzy rule-based classification of remotely sensed imagery, *IEEE Transactions on Geoscience and Remote Sensing*, *40*(2), 362–374, doi:10.1109/36.992798.

Bárdossy, A., I. Bogardi, and W. Kelly (1988), Imprecise (fuzzy) information in geostatistics, *Mathematical Geology*, *20*(4), 287–311, doi:10.1007/BF00892981.

Bárdossy, A., I. Bogardi, and W. Kelly (1989), Geostatistics utilizing imprecise (fuzzy) information, *Fuzzy Sets and Systems*, *31*(3), 311–328, doi:10.1016/0165-0114(89)90203-0.

Bárdossy, A., I. Bogardi, and L. Duckstein (1990a), Fuzzy regression in hydrology, *Water Resources Research*, *26*(7), 1497–1508, doi:10.1029/WR026i007p01497.

Bárdossy, A., I. Bogardi, and W. Kelly (1990b), Kriging with imprecise (fuzzy) variograms. II: Application, *Mathematical Geology*, *22*(1), 81–94, doi:10.1007/BF00890298.

Bárdossy, A., I. Bogardi, and W. E. Kelly (1990c), Kriging with imprecise (fuzzy) variograms. I: Theory, *Mathematical Geology*, *22*(1), 63–79, doi:10.1007/BF00890297.

Bárdossy, A., L. Duckstein, and I. Bogardi (1995), Fuzzy rule-based classification of atmospheric circulation patterns, *International Journal of Climatology*, *15*(10), 1087–1097, doi:10.1002/joc.3370151003.

Bender, M. J., and S. P. Simonovic (2000), A fuzzy compromise approach to water resource systems planning under uncertainty, *Fuzzy Sets and Systems*, *115*(1), 35–44, doi:10.1016/S0165-0114(99)00025-1.

Benito, G., and V. R. Thorndycraft (2005), Palaeoflood hydrology and its role in applied hydrological sciences, *Journal of Hydrology*, *313*(1–2), 3–15, doi:10.1016/j.jhydrol.2005.02.002.

Benito, G., A. Díez-Herrero, and M. Fernández de Villalta (2003), Magnitude and Frequency of Flooding in the Tagus Basin (Central Spain) over the Last Millennium, *Climatic Change*, *58*(1-2), 171–192, doi:10.1023/A:1023417102053.

Bogardi, I., A. Bardossy, L. Duckstein, and R. Pongracz (2003), Fuzzy logic in hydrology and water resources, in *Fuzzy Logic In Geology*, edited by R. V. Demicco and G. J. Klir, chap. 6, pp. 153–190, Elsevier, Academic Press.

Brázdil, R., Z. W. Kundzewicz, and G. Benito (2006), Historical hydrology for studying flood risk in Europe, *Hydrological Sciences Journal*, *51*(5), 739–764, doi:10.1623/hysj.51.5.739.

Brázdil, R., Z. W. Kundzewicz, G. Benito, G. Demarée, N. Macdonald, and L. A. Roald (2012), Historical Floods in Europe in the Past Millennium, in *Changes in Flood Risk in Europe*, edited by Z. W. Kundzewicz, IAHS Special Publication 10, IAHS Press, Wallingford.

Brázdil, R., et al. (2005), *Historické a současné povodně v České republice (Historical and recent Floods in the Czech Republic)*, Masarykova univerzita v Brně, Český hydrometeorologický ústav v Praze, Brno - Praha.

Chachi, J., S. Taheri, and N. Arghami (2014), A hybrid fuzzy regression model and its application in hydrology engineering, *Applied Soft Computing*, *25*(0), 149–158, doi:10.1016/j.asoc.2014.09.010.

Cohn, T. A., W. L. Lane, and W. G. Baier (1997), An algorithm for computing moments-based flood quantile estimates when historical flood information is available, *Water Resources Research*, *33*(9), 2089–2096, doi:10.1029/97WR01640.

Deutsch, M., and K.-H. Pörtge (2003), *Hochwasserereignisse in Thüringen (Flood events in Thuringia)*, Schriftenreihe 63, Thüringer Landesanstalt für Umwelt und Geologie, Jena.

Deutsch, M., J. Böhner, K.-H. Pörtge, and K. T. Rost (2004), Untersuchungen historischer Hochwasserereignisse in Thüringen - dargestellt am Beispiel der Werra. (Investigation of historical floods in Thuringia - depicted by the case study of the Werra River), *Zeitschrift f. Geomorphologie, N.F., Suppl.-Bd.*, *135*, 11–32.

Elleder, L., J. Herget, T. Roggenkamp, and A. Nießen (2013), Historic floods in the city of Prague – a reconstruction of peak discharges for 1481–1825 based on documentary sources, *Hydrology Research*, *44*(2), 202–214, doi:10.2166/nh.2012.161.

England, J. F., J. D. Salas, and R. D. Jarrett (2003), Comparisons of two moments-based estimators that utilize historical and paleoflood data for the log Pearson type III distribution, *Water Resources Research*, *39*(9), 1243, doi:10.1029/2002WR001791.

Esogbue, A. O., M. Theologidu, and K. Guo (1992), On the application of fuzzy sets theory to the optimal flood control problem arising in water resources systems, *Fuzzy Sets and Systems*, *48*(2), 155–172, doi:10.1016/0165-0114(92)90330-7.

Gaume, E., L. Gaál, A. Viglione, J. Szolgay, S. Kohnová, and G. Blöschl (2010), Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites, *J. Hydrol.*, *394*, 101–117, doi:10.1016/j.jhydrol.2010.01.008.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004), *Bayesian Data Analysis*, Texts in Statistical Science, 2nd ed., Chapman and Hall/CRC, New York.

Glaser, R., and H. Stangl (2004), Climate and floods in central Europe since AD 1000: Data, methods results and consequences, *Surv. Geophys.*, *25*, 485–510.

Glaser, R., et al. (2010), The variability of European floods since AD 1500, *Climatic Change*, *101*(1-2), 235–256, doi:10.1007/s10584-010-9816-7.

Haario, H., M. Laine, A. Mira, and E. Saksman (2006), DRAM: Efficient adaptive MCMC, *Statistics and Computing*, *16*(4), 339–354, doi:10.1007/s11222-006-9438-0.

Hall, J., et al. (2014), Understanding flood regime changes in Europe: a state-of-the-art assessment, *Hydrology and Earth System Sciences*, *18*(7), 2735–2772, doi:10.5194/hess-18-2735-2014.

Herget, J., and H. Meurs (2010), Reconstructing peak discharges for historic flood levels in the city of Cologne, Germany, *Global and Planetary Change*, *70*(1–4), 108–116, doi: 10.1016/j.gloplacha.2009.11.011.

Herget, J., T. Roggenkamp, and M. Krell (2014), Estimation of peak discharges of historical floods, *Hydrology and Earth System Sciences*, *18*(10), 4029–4037, doi:10.5194/hess-18-4029-2014.

Kiss, A., and J. Laszlovszky (2013), 14th-16th-Century Danube Floods and Long-Term Water-Level Changes in Archaeological and Sedimentary Evidence in The Western and Central Carpathian Basin: an Overview with Documentary Comparison, *Journal of Environmental Geography*, *6*(3-4), 1–11, doi:10.2478/jengeo-2013-0001.

Kjeldsen, T., et al. (2014), Documentary evidence of past floods in Europe and their utility in flood frequency estimation, *Journal of Hydrology*, *517*, 963–973.

Kuczera, G. (1999), Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference, *Water Resources Research*, *35*(5), 1551–1557, doi:10.1029/1999WR900012.

Leese, M. N. (1973), Use of censored data in the estimation of Gumbel distribution parameters for annual maximum flood series, *Water Resources Research*, *9*(6), 1534–1542, doi: 10.1029/WR009i006p01534.

Menger, K. (1951), Ensembles flous et fonctions aleatoires, *C. R. Acad. Sci.*, *232*, 2001–2003.

Merz, B., and A. H. Thieken (2005), Separating natural and epistemic uncertainty in flood frequency analysis, *Journal of Hydrology*, *309*(1–4), 114–132, doi:10.1016/j.jhydrol.2004.11.015.

Merz, R., and G. Blöschl (2008a), Flood frequency hydrology: 1. temporal, spatial, and causal expansion of information, *Water Resour. Res.*, *44*, W08,432, doi:10.1029/2007WR006744.

Merz, R., and G. Blöschl (2008b), Flood frequency hydrology: 2. combining data evidence, *Water Resour. Res.*, *44*, W08,433, doi:10.1029/2007WR006745.

Mudelsee, M., M. Deutsch, M. Börngen, and G. Tetzlaff (2006), Trends in flood risk of the River Werra (Germany) over the past 500 years, *Hydrological Sciences Journal*, *51*(5), 818–833, doi: 10.1623/hysj.51.5.818.

Nasseri, M., A. Ansari, and B. Zahraie (2014), Uncertainty assessment of hydrological models with fuzzy extension principle: Evaluation of a new arithmetic operator, *Water Resources Research*, *50*(2), 1095–1111, doi:10.1002/2012WR013382.

Neppel, L., et al. (2010), Flood frequency analysis using historical data: accounting for random and systematic errors, *Hydrological Sciences Journal*, *55*(2), 192–208, doi: 10.1080/02626660903546092.

O'Connell, D. R. H., D. A. Ostenaa, D. R. Levish, and R. E. Klinger (2002), Bayesian flood frequency analysis with paleohydrologic bound data, *Water Resources Research*, *38*(5), 1058, doi: 10.1029/2000WR000028.

Özelkan, E. C., and L. Duckstein (2001), Fuzzy conceptual rainfall–runoff models, *Journal of Hydrology*, *253*(1–4), 41–68, doi:10.1016/S0022-1694(01)00430-9.

Parent, E., and J. Bernier (2003), Bayesian POT modeling for historical data, *Journal of Hydrology*, *274*(1–4), 95–108, doi:10.1016/S0022-1694(02)00396-7.

Payrastre, O., E. Gaume, and H. Andrieu (2011), Usefulness of historical information for flood frequency analyses: Developments based on a case study, *Water Resources Research*, *47*(8), W08,511, doi:10.1029/2010WR009812.

Plate, E. J. (2002), Flood risk and flood management, *Journal of Hydrology*, *267*(1–2), 2–11, doi: 10.1016/S0022-1694(02)00135-X, advances in Flood Research.

Prodanovic, P., and S. P. Simonovic (2002), Comparison of fuzzy set ranking methods for implementation in water resources decision-making, *Canadian Journal of Civil Engineering*, *29*(5), 692–701, doi:10.1139/l02-063.

Reis, D. S. J., and J. R. Stedinger (2005), Bayesian MCMC flood frequency analysis with historical information, *Journal of Hydrology*, *313*(1–2), 97–116, doi:10.1016/j.jhydrol.2005.02.028.

Retsö, D. (2014), Documentary evidence of historical floods and extreme rainfall events in Sweden 1400–1800, *Hydrology and Earth System Sciences Discussions*, *11*(9), 10,085–10,116, doi: 10.5194/hessd-11-10085-2014.

Ribatet, M., E. Sauquet, J. M. Grésillon, and T. B. M. J. Ouarda (2007), Usefulness of the reversible jump Markov chain Monte Carlo model in regional flood frequency analysis, *Water Resources Research*, *43*(8), W08,403, doi:10.1029/2006WR005525.

Robert, C., and G. Casella (2004), *Monte Carlo statistical methods*, Springer Texts in Statistics, Springer New York, doi:10.1007/978-1-4757-4145-2.

Rohr, C. (2004), Überschwemmungen an der Traun zwischen Alltag und Katastrophe. Die Welser Traunbrücke im Spiegel der Bruckamtsrechnungen des 15. und 16. Jahrhunderts, *Jahrbuch des Musealvereines Wels*, *33*, 281–328.

Rohr, C. (2006), Measuring the frequency and intensity of floods of the Traun River (Upper Austria), 1441–1574, *Hydrological Sciences Journal*, *51*(5), 834–847, doi:10.1623/hysj.51.5.834.

Rosenblatt, M. (1956), Remarks on some nonparametric estimates of a density function, *The Annals of Mathematical Statistics*, *27*(3), 832, doi:10.1214/aoms/1177728190.

Russell, S., and P. Campbell (1996), Reservoir Operating Rules with Fuzzy Programming, *Journal of Water Resources Planning and Management*, *122*(3), 165–170, doi:10.1061/(ASCE)0733-9496(1996)122:3(165).

Salinas, J. L., A. Castellarin, S. Kohnová, and T. R. Kjeldsen (2014a), Regional parent flood frequency distributions in Europe – Part 2: Climate and scale controls, *Hydrology and Earth System Sciences*, *18*(11), 4391–4401, doi:10.5194/hess-18-4391-2014.

Salinas, J. L., A. Castellarin, A. Viglione, S. Kohnová, and T. R. Kjeldsen (2014b), Regional parent flood frequency distributions in Europe – Part 1: Is the GEV model suitable as a pan-European parent?, *Hydrology and Earth System Sciences*, *18*(11), 4381–4389, doi:10.5194/hess-18-4381-2014.

Schulz, K., and B. Huwe (1997), Water flow modeling in the unsaturated zone with imprecise parameters using a fuzzy approach, *Journal of Hydrology*, *201*(1–4), 211–229, doi:10.1016/S0022-1694(97)00038-3.

Schumann, A. H. (2011), Introduction – Hydrological Aspects of Risk Management, in *Flood Risk Assessment and Management*, edited by A. H. Schumann, pp. 1–10, Springer Netherlands, doi:10.1007/978-90-481-9917-4_1.

Schumann, A. H., and D. Nijssen (2011), Application of Scenarios and Multi-Criteria Decision Making Tools in Flood Polder Planning, in *Flood Risk Assessment and Management*, edited by A. H. Schumann, pp. 249–275, Springer Netherlands, doi:10.1007/978-90-481-9917-4_12.

Shrestha, B., L. Duckstein, and E. Stakhiv (1996), Fuzzy Rule-Based Modeling of Reservoir Operation, *Journal of Water Resources Planning and Management*, *122*(4), 262–269, doi:10.1061/(ASCE)0733-9496(1996)122:4(262).

Simonovic, S. P., and R. Verma (2008), A new methodology for water resources multicriteria decision making under uncertainty, *Physics and Chemistry of the Earth, Parts A/B/C*, *33*(5), 322–329, doi:10.1016/j.pce.2008.02.006, integrated Water Resources Management in a Changing World.

Soetaert, K., and T. Petzoldt (2010), Inverse modelling, sensitivity and Monte Carlo analysis in R using package FME, *Journal of Statistical Software*, *33*, 1–28.

Stedinger, J. R., and T. A. Cohn (1986), Flood Frequency Analysis With Historical and Paleoflood Information, *Water Resources Research*, *22*(5), 785–793, doi:10.1029/WR022i005p00785.

Sturm, K., R. Glaser, J. Jacobeit, M. Deutsch, R. Brázdil, C. Pfister, J. Luterbacher, and H. Wanner (2001), Floods in Central Europe since AD 1500 and their Relation to the Atmospheric Circulation, *Petermanns Geographische Mitteilungen*, *148*(6), 18–27.

Viertl, R. (2008a), Fuzzy Bayesian Inference, in *Soft Methods for Handling Variability and Imprecision*, *Advances in Soft Computing*, vol. 48, edited by D. Dubois et al., pp. 10–15, Springer Berlin Heidelberg, doi:10.1007/978-3-540-85027-4_2.

Viertl, R. (2008b), Foundations of Fuzzy Bayesian Inference, *Journal of Uncertain Systems*, *2*, 187–191.

Viertl, R. (2011a), *Statistical Methods for Fuzzy Data*, John Wiley and Sons, West Sussex.

Viertl, R. (2011b), *Bayes' theorem and fuzzy information*, pp. 91–96, John Wiley & Sons, Ltd, doi:10.1002/9780470974414.ch14.

Viglione, A., R. Merz, J. L. Salinas, and G. Blöschl (2013), Flood frequency hydrology: 3. a Bayesian analysis, *Water Resour. Res.*, *49*, 1–18, doi:10.1029/2011WR010782.

Wetter, O. (2012), Hochwasser-"¡Katastrophen"¿ in Basel vom 13. bis 21. Jahrhundert. Rekonstruktion, Deutung und Lerneffekte, in *Krisen Ursachen, Deutungen und Folgen*, edited by D. David, J. Mathieu, J. M. Schaufelbuehl, and T. Straumann, pp. 47–64, Chronos Verlag, Zurich.

Wetter, O., C. Pfister, R. Weingartner, J. Luterbacher, T. Reist, and J. Trösch (2011), The largest floods in the High Rhine basin since 1268 assessed from documentary and instrumental evidence, *Hydrological Sciences Journal*, *56*(5), 733–758, doi:10.1080/02626667.2011.583613.

Zadeh, L. A. (1965), Fuzzy sets, *Information and Control*, *8*(3), 338–353, doi:10.1016/S0019-9958(65)90241-X.

Chapter 6

Summary of results and overall conclusions

The findings presented in this thesis help better characterize and quantify different facets of uncertainty involved in the flood frequency estimation process. While these different facets of uncertainty are usually lumped together, the present work aims at throwing light at four well defined sources of these uncertainties, by analysing the model and data related aspects that constrain them, and by defining the characteristic spatial and temporal scales under which they operate.

The uncertainties computed and analysed in Chapter 2 of this thesis, will be only applicable in the case of absence of local discharge data, while the spatial scale at which it has been assessed is the entire globe. In this sense, the patterns in these uncertainties need to be taken as avergaged behaviours. The results showed that predictive performance of flood tends to decrease with increasing aridity and tends to increase with increasing catchment area. In particular, index methods are significantly biased for arid catchments and tend to overestimate the 100 yr floods in the catchments analysed. Arid regions would therefore need more gauges to capture the temporal and spatial variability in order to reduce the uncertainties in the estimates, but achieving this is unrealistic in many arid parts of the world where (due to economic reasons) data density is typically lower than in humid regions. Methods that are able to exploit the specifics of the region would be needed here. Use of readily available landscape information, such as erosional patterns, based on the idea of reading the landscape, may assist in improving the predictions of runoff extremes.

The uncertainites induced by statistical model selection have been assessed on a regional scale, covering the entire European continent for Chapter 3 and three countries in Chapter 4. In this context, the quantification of the uncertainty has been expressed in terms of sample variabilities of L-moments, and has a regional validity, as at-site statistical model choice is subject ot other sources of uncertainties, given mostly by lack of knowledge in local flood generation processes. The results for these chapters showed that that the GEV alone can not be considered as a single candidate for a pan-European flood frequency distribution, being not able to reproduce the entire variety of hydrological processes leading to the different shapes of flood frequency. The inclusion of information on the underlying hydrological processes is therefore of high importance in reducing the uncertainty in the model choice. curves. Furthermore, the results presented Chapter 4 could be seen as a first attempt at defining a set of âprocess-drivenâ regional parent flood frequency distributions in a European context. By defining hydro-climatical regional with some prevalence of a given statistical model, the facet of uncertainty associated with model selection could be very much reduced.

In Chapter 5, a different facet of unertainty is presented, in term of imprecision of historical data. Given the new framework presented for extracting that imprecision and propagating it into the flood estimate, the modeller perception of the non-precise historical records may have a significant effect in the quantification of this particular kind of uncertainty. So far in the literature, only precisely known historical discharges, or information about a lower and upper bound for historical discharges, was included in formal methods for flood frequency estimation. Fuzzy modelling offers therefore a higher flexibility in representing all the information that can be retreived from the historical records via membership functions, than just a lower and upper bound. This will help to better exploit the information on historical flood events to be included in the Bayesian framework. The propagation of the modelled imprecision from the historical records to the flood frequency estimate is then computed in an explicit way, allowing to the modeller to assess the effects of incorporating different levels of non-precise historical information in the flood frequency regime estimation.

Table 6.1 reports a summary of the general framework developed in the thesis, for identifying the nature of the diferent facets of uncertainties described, and quantifying them.

Chapter	**Nature** of uncertainty	Method to **quantify** uncertainty	Spatial **scale**
2	· Ungauged basins (lack of data) · Uncertainty related with climate, model and data	· Leave one out cross validation · Regionalisation error as surrogate for predictive uncertainty	· Global
3	· Sampling uncertainty · Limited sample sizes · Related with the (3rd – 4th moment relationship)	· Resampling · Monte Carlo simulations	· Europe
4	· Model Choice (statistical distribution) · Related with the (3rd – 4th moment relationship)	· Regional control of climate and scale · 3rd – 4th moment relationship = f(*A*, *MAP*)	· Regional
5	· Sampling uncertainty · Imprecision of historical discharges	· Flood Frequency Hydrology · Fuzzy-Bayesian Inference	· Local

Table 6.1. Framework developed in this PhD thesis for characterising and quantifying uncertainty components in flood frequency estimation.

The results presentd in this thesis have implications for both hydrological science base knowledge, and applied engineering hydrology. On the one hand, linking uncertainties in flood frequency estimation with hydrological and climatological indicators helps identify regions where an improved hydrological process understanding in needed. For example, for the case of arid regions, both the uncertainties in prediction of floods in ungauged catchments and the regional uncertainties caused by model selection, resulted significantly higher than in humid cathments.

On the other hand, an improved quantification of the uncertainties helps in obtaining more robust and reliable design decisions and flood risk zones, tasks that belong to the very essence of applied flood hydrology. For example, by including the information about the historical floods in the flood frequency estimation, long term features of the flood regime could be captured and expressed, even in a non-precise way in terms of the fuzzy estimates for flood quantiles.

Acknowledgements

In Chapters 3 and 4, the L-moment ratios dataset used was compiled in the frame of the COST Action ES0901 (European procedures for flood frequency estimation – FloodFreq), contributions made by members of Working Group 2 of the COST Action are greatly appreciated. For Chapter 5, the systematic data for the Rhine river at Basel (gauging station Rheinsbrücke) was provided by the Wasser- und Schifffahrtsverwaltung des Bundes (Deutschland), and the data for the Werra river at Meiningen was provided by Thüringer Landesanstalt für Umwelt und Geologie Jena. Both contributions are acknowledged.

Financial support for this thesis was provided by the Austria Science Funds (FWF) as part of the Vienna Doctoral Programme on Water Resource Systems (DK-plus W1219-N22), the FWF Project P23723-N21, the Innovative Ideas program of the TU Wien, the Austrian Climate and Energy Fund (Project K10AC0K00003, CILFAD), the COST Office grant ES0901 (European procedures for flood frequency estimation – FloodFreq), and the ERC FloodChange project (ERC Advanced Grant FP7-IDEAS-ERC-AG-PE10 291152).