Hydrol. Earth Syst. Sci. Discuss., 10, 411–447, 2013
www.hydrol-earth-syst-sci-discuss.net/10/411/2013/
doi:10.5194/hessd-10-411-2013

Hydrology and Earth System Sciences Discussions

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

Received: 21 December 2012 – Accepted: 3 January 2013 – Published: 14 January 2013

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

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

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 performances 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

HESSD

10, 411–447, 2013

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 characterized 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 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 et al., 1999). An alternative is the region of influence (ROI) approach (Burn, 1990a) 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; Griffis and Stedinger, 2007 for the case of floods and Gustard et al., 1992; Engeland and Hisdal, 2009 for the case of low flows). More recently, geostatistical methods have become popular 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. (2013) for the case of low flows). 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; Farquason 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

HESSD
10, 411–447, 2013

predictive performance of flood and low flow estimation methods in ungauged basins. In a second step we analyse a number of more detailed data sets, 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, 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 flows 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 a view to learning from their differences and similarities in performance in a general way.

HESSD
10, 411–447, 2013

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 characterize 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:

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

where $Q_{i,\mathrm{pred}}$: predicted specific discharge in cross-validation at gauge i, $Q_{i,\mathrm{obs}}$: observed specific discharge at gauge i, $\bar{Q}_{\mathrm{obs}}$: spatial mean of the observed specific discharge.

The target low flow index, on which this performance was reported, was 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

HESSD
10, 411–447, 2013

other low flow indicators (Table 1). 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 panel 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). Predictive performance on a catchment basis was given as the absolute normalised error (ANE), defined as:

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

The dataset for Level 2 assessment combines data from 1895 catchments. Three catchment characteristics are analysed: aridity index, mean elevation and catchment area. 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; and an indicator of the amount of rainfall data that is available for runoff estimation in ungauged basins, since larger catchments tend to contain a large number of rain gauges. The low flow regionalisation methods have been classified into the following groups:

– Process based methods: there is only a single cross-validation study we encountered in the literature (Engeland and Hisdal, 2009) of this type. The procedure

HESSD
10, 411–447, 2013

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: 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 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: 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: 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.

HESSD
10, 411–447, 2013
Part 2: Flood and low flow studies
J. L. Salinas et al.

– Short records: 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).

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:

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

The target flood index, on which this performance was mainly reported, was the 100-yr specific flood quantile, 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. Several studies compare different regionalisation approaches and/or subsets of data which results in a total of 49 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 panel) 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

HESSD
10, 411–447, 2013

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 low flow 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 panel) shows the Level 1 results of estimating low flows in ungauged basins. 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,

HESSD
10, 411–447, 2013

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 panel) show that the predictions in humid regions exhibit the largest errors and arid regions have the smallest errors. This means that the predictive performance clearly decreases with increasing aridity. There are a number of factors that may contribute to this dependence. The inter-annual variability (e.g. in terms of CV 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. This means that the floods are more difficult to estimate from short records. The stronger non-linearity also imply that the spatial hydrological variability in the flood producing processes will impact more strongly on 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. 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).

HESSD
10, 411–447, 2013

4.2 Which regionalisation method performs best?

The low flow regionalisation methods represented in the assessment included: one result from the process based methods group (continuous runoff models); four 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; ten global regression and seven regional regression results from the regression methods group; and five 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}, $Q_{7,10}$, $Q_{\mathrm{mon},5}$, all standardised by catchment area or mean flow, and the dimensionless (base flow index) 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 panel) 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. German region Baden-Württemberg) and studies in less seasonal climates (e.g. New South Wales and Victoria in Australia). The four results from geostatistical models give 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

HESSD
10, 411–447, 2013

Part 2: Flood and low flow studies

J. L. Salinas et al.

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 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 this 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 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

HESSD
10, 411–447, 2013

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 panel) 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. 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. These results are confirmed by studies that compared different approaches in the same region (grey lines in Fig. 3, right panel). It appears that it 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 of 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 that applying geostatistical methods which is because they have a longer tradition in hydrology. The first two columns of 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 panel) 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 higher stream gauge density in the larger studies but more detailed analyses 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 panel) 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 because of the higher stream gauge density in the larger studies 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 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 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 and 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 low land 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

HESSD
10, 411–447, 2013

catchments compared to low land 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 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 summarize 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.

HESSD
10, 411–447, 2013

The Level 2 assessment for flood prediction studies, i.e. the assessment of the ANE error 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 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 likely that gauged sub-catchments 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 summarize the runoff prediction performance of different regionalisation approaches, stratified by the aridity index. Again, the right-top, right-middle and right- bottom panels show the performance for all catchment, 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

HESSD
10, 411–447, 2013

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 and middle panels). The median absolute normalised error is 1.0 indicating that typically the methods predict 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 of 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 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

HESSD
10, 411–447, 2013

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 inter-annual 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 assessment). 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

HESSD
10, 411–447, 2013

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 et al., 2005; 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.

Acknowledgement. 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), FWF Project P 23723-N21, and the Austrian Climate and Energy Fund (Project Nro. K10AC0K00003, CILFAD) for financial support.

References

Aschwanden, H. and Kan, C.: Die Abflussmenge Q347, Eine Standortbestimmung, Hydrologische Mitteilungen Communications hydrologiques, No. 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, Hydrol. Process., 9, 313–330, 1995.

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, 1990a.

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, 205 pp., 1993.

HESSD
10, 411–447, 2013

Part 2: Flood and low flow studies

J. L. Salinas et al.

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., 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.

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, Report 108, Institute of Hydrology, Wallingford, UK, 1992.

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

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, Rev. Obras Publ., 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.

HESSD

10, 411–447, 2013

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, Hydrol. Process., 20, 3851–3878, doi:10.1002/hyp.6161, 2006b.

Laaha, G. and Blöschl, 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 Blöschl, 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, 2013.

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., PHI Series, UNESCO, Paris, 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 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.

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.

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. Discuss., 10, 375–409, doi:10.5194/hessd-10-375-2013, 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, 166 pp., 2010.

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., 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., Blöschl, 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. Discuss., 10, 449–485, doi:10.5194/hessd-10-449-2013, 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. and Merz, R., Forum für Hydrologie und Wasserbewirtschaftung, Heft 30.11, 29–35, 2011.
Zhang, Z. and Kroll, C. N.: The base flow correlation method with multiple gauged sites, J. Hydrol., 347, 371–380, 2007b.
Zhang, Z. and Kroll, C. N.: A closer look at base flow correlation, J. Hydrol. Eng.-ASCE, 12, 190–196, 2007a.

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 ($q_{mon,5}$), 95–97 % specific runoff quantiles (q_{95}, q_{96}, q_{97}), normalized q_{95} specific runoff quantile (q_{95}/q_A) and baseflow index (BFI).

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	×
Castiglioni et al. (2001)	central Italy	Humid	51	G	q_{97}	0.89	
Plasse and Sauquet (2010)	France	Humid	1003	GR, RR, G, G	$q_{mon,5}$	0.43, 0.53–0.74, 0.61, 0.63–0.73	×
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	×
Laaha and Blöschl (2007)	Austria	Cold	325	RR	q_{95}	0.75	
Laaha et al. (2007)	Austria	Cold	298	G	q_{95}	0.75	×
Laaha and Blöschl (2006a,b)	Austria	Cold	325	GR, RR	q_{95}	0.57, 0.59–0.70	×
Laaha and Blöschl (2005)	Austria	Cold	325	SR	q_{95}	0.62, 0.93	×
Rees et al. (2002)	Himalayas, Nepal and India	Humid	40	GR	q_{95}/q_A	0.45, 0.53	
Aschwanden and Kan (1999)	Swizerland	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	

HESSD
10, 411–447, 2013

HESSD
10, 411–447, 2013

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 (49)	3023	5	1422

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 standarized by the mean annual flood (Q_{100}/Q_m).

Study	Region	Climate	Number of catchments	Regiona-lisation method	Predicted variable	Error measure (RMSNE)	Used in Level 2
Jimenez et al. (2012)	Spain	Arid	217	R	q_{100}	0.54	×
Walther et al. (2011)	Germany (Saxony)	Cold	170	G, IM	q_{100}	0.46, 0.49	×
Kjeldsen and Jones (2010)	United Kingdom	Humid	602	IM	q_{100}	0.51, 0.50	×
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	×
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. (2006)	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	×
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	
Farquhason et al. (1992)	Arid and semi-arid basins worldwide	Arid	162	IM	q_{100}	0.73	

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. 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

HESSD
10, 411–447, 2013

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

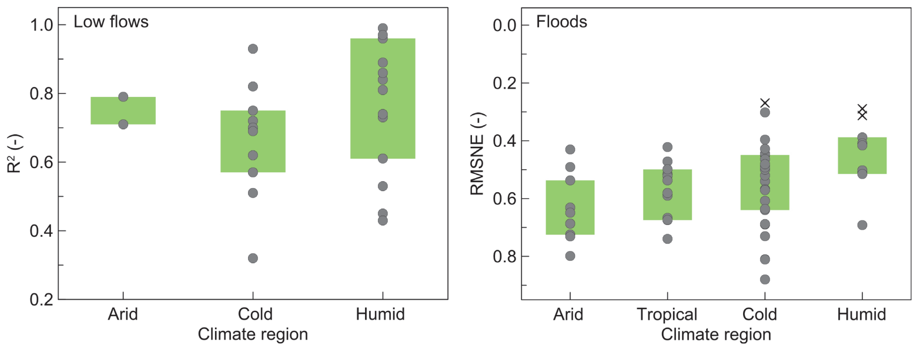

Fig. 2. Coefficient of determination (R^2) of predicting low flows in ungauged basins (left panel) and root mean squared normalised error (RMSNE) of predicting floods in ungauged basins (right panel), stratified by climate (Level 1 assessment). Each symbol refers to a result from the studies in Tables 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.

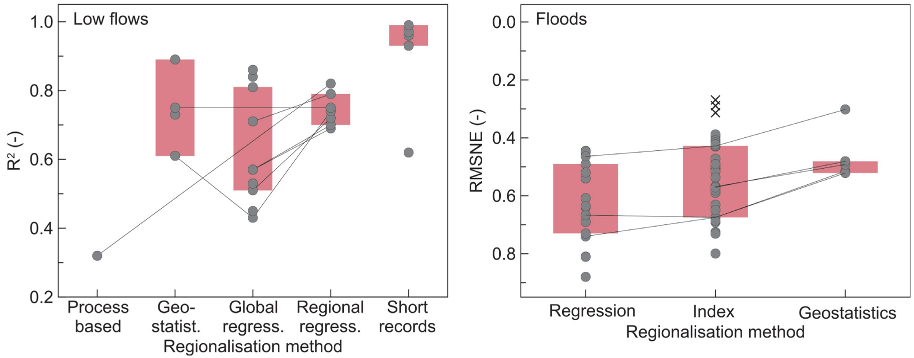

Fig. 3. Coefficient of determination (R^2) of predicting low flows in ungauged basins (left panel) and root mean squared normalised error (RMSNE) of predicting floods in ungauged basins (right panel), stratified by regionalisation method (Level 1 assessment). Each symbol refers to a result from the studies 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.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD
10, 411–447, 2013

Part 2: Flood and low flow studies

J. L. Salinas et al.

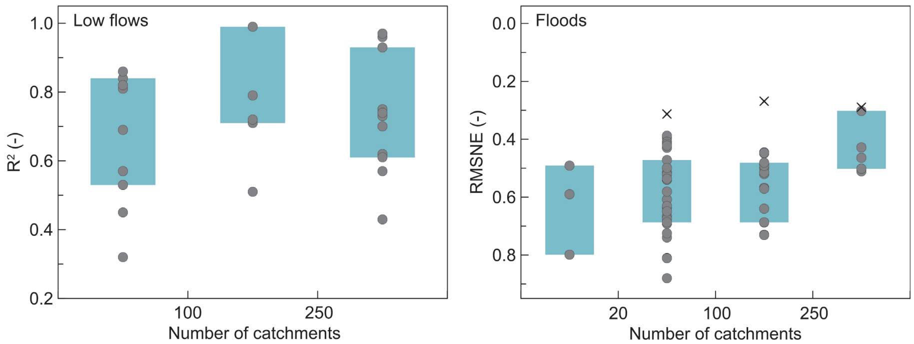

Fig. 4. Coefficient of determination (R^2) of predicting low flows in ungauged basins (left panel) and root mean squared normalised error (RMSNE) of predicting floods in ungauged basins (right panel), stratified by the number of catchments within each study (Level 1 assessment). Each symbol refers to 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.

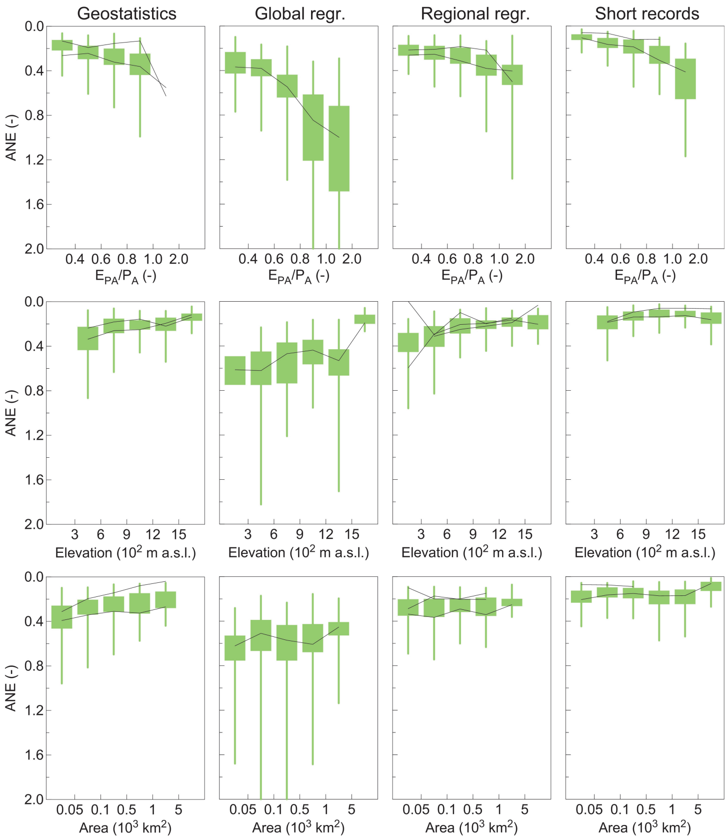

Fig. 5. Absolute normalized error (ANE) of predicting q95 low flows ($m^3\,s^{-1}\,km^{-2}$) in ungauged basins as a function of aridity (E_{PA}/P_A), mean elevation and catchment area for different regionalisation methods (Level 2 assessment). Lines connect median errors for the same study. Boxes are 40 %–60 % quantiles, whiskers are 20–80 % quantiles.

HESSD
10, 411–447, 2013

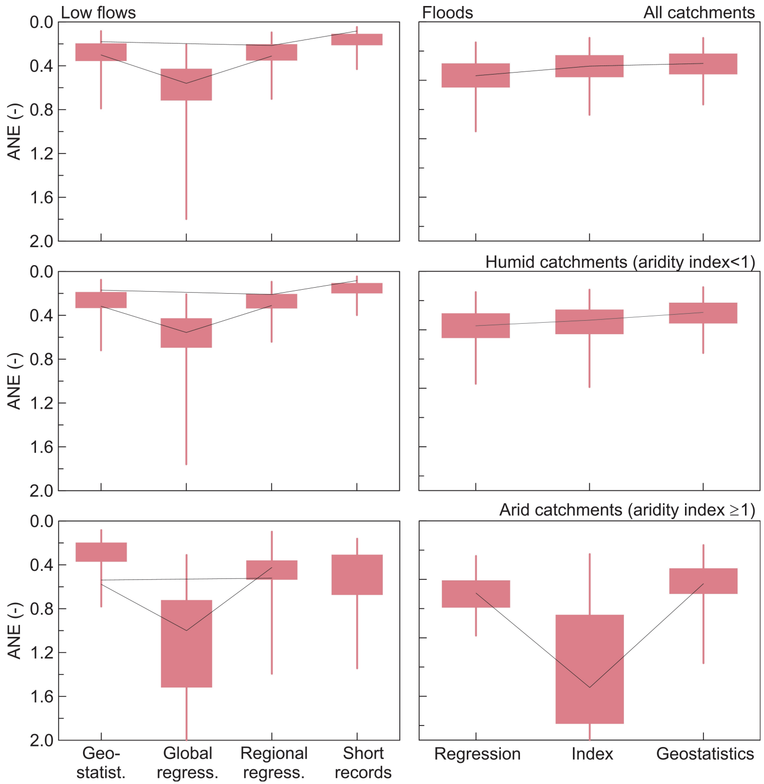

Fig. 6. Absolute normalized error (ANE) of predicting q_{95} low flows ($m^3\,s^{-1}\,km^{-2}$), left panels, and q_{100} floods ($m^3\,s^{-1}\,km^{-2}$), right panels, in ungauged basins for different regionalisation methods, stratified by aridity (Level 2 assessment). Top: all catchments. Centre: humid 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.

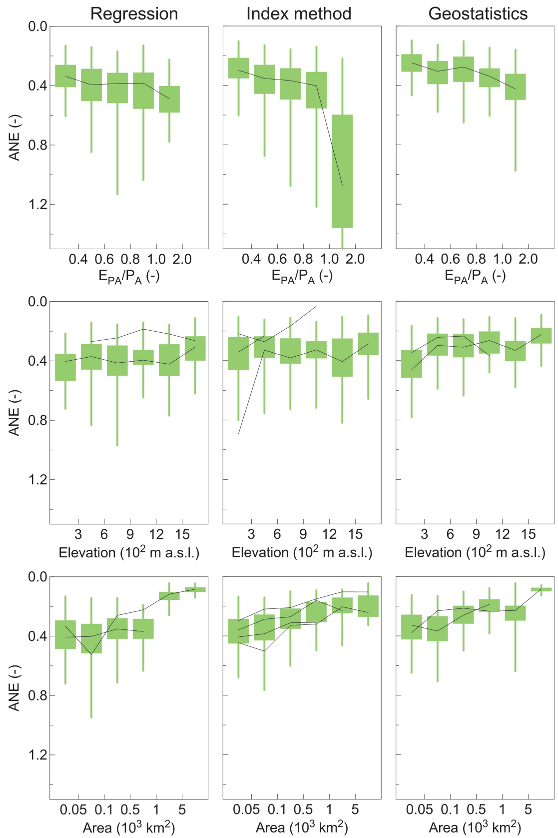

Fig. 7. Absolute normalized error (ANE) of predicting q_{100} floods ($m^3 s^{-1} km^{-2}$) in ungauged basins as a function of aridity (E_{PA}/P_A), mean elevation and catchment area for different regionalisation methods (Level 2 assessment). Lines connect median errors for the same study. Boxes are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

HESSD
10, 411–447, 2013