Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Hydrology 323 (2006) 193–214

Journal of Hydrology

www.elsevier.com/locate/jhydrol

A comparison of low flow regionalisation methods—catchment grouping

G. Laaha[a,*], G. Blöschl[b]

[a]Institut für Angewandte Statistik und EDV, Universität für Bodenkultur Wien, Gregor Mendel St. 33, A-1180 Wien, Austria
[b]Institut für Wasserbau und Ingenieurhydrologie, Technische Universität Wien, Karlsplatz 13, A-1040 Wien, Austria

Received 13 July 2004; revised 15 June 2005; accepted 2 September 2005

Abstract

Four catchment grouping methods are compared in terms of their performance in predicting specific low flow discharges q_{95}, i.e. the specific discharge that is exceeded on 95% of all days. The grouping methods are the residual pattern approach, weighted cluster analysis, regression trees and an approach based on seasonality regions. For each group, a regression model between catchment characteristics and q_{95} is fitted independently. Data from 325 sub-catchments in Austria ranging in catchment area from 7 to 963 km^2 are used in the analysis. The performance of the methods is assessed by leave-one-out cross-validation of the regressionestimates, which emulates the case of ungauged catchments. Results indicate that the grouping based on seasonality regions performs best and explains 70% of the spatial variance of q_{95}. The favourable performance of this grouping method is likely related to the striking differences in seasonal low flow processes in the study domain. Winter low flows are associated with the retention of solid precipitation in the seasonal snow pack while summer low flows are related to the relatively large moisture deficits in the lowland catchments during summer. The regression tree grouping performs second best (explained variance of 64%) and the performance of the residual pattern approach is similar (explained variance of 63%). The weighted cluster analysis only explains 59% of the spatial variance of q_{95}, which is only a minor improvement over the global regression model, i.e. without using any grouping (explained variance of 57%). An analysis of the sample characteristics of all methods suggests that, again, the grouping method based on the seasonality regions has the most favourable characteristics although all methods tend to underestimate specific low flow discharges in the very wet catchments.

Keywords: Low flows; Regionalisation; Regional regression; Classification; Cluster analysis; Regression tree; Classification tree; CART; Seasonality index

1. Introduction

Accurate estimates of low flow characteristics are needed for a range of purposes in water resources management and engineering including environmental flow requirements, water uses and discharges into

* Corresponding author. Tel.: +43 1 47654 5066; fax: +43 1 47654 5069.

E-mail addresses: laaha@edv1.boku.ac.at (G. Laaha), bloeschl@hydro.tuwien.ac.at (G. Blöschl).

0022-1694/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2005.09.001

streams, and hydropower operation (Smakhtin, 2001; Gustard et al., 2004). Low flow characteristics are best estimated from observed stream flow data but for sites where these data are unavailable hydrological regionalisation techniques can be used to infer them from other catchments where stream flow data have been collected (e.g. Demuth and Young, 2004).

The regionalisation of low flow characteristics is usually based on some sort of regression between the low flow characteristic of interest and catchment characteristics that are available for ungauged sites (e.g. Vogel and Kroll, 1992; Gustard et al., 1992; Schreiber and Demuth, 1997; Skop and Loaiciga, 1998). If the study domain is large or very heterogeneous in terms of the low flow processes a number of authors have suggested to split the domain into regions and apply a regression relationship to each of the regions independently. This is termed the regional regression approach. Gustard and Irving (1994), for example, tested a number of regression models between standardised Q_{95} low flows (the discharge that is exceeded on 95% of all days) and different soil group indices for 1530 catchments in Europe. Their global regression model of nine soil classes explained 29% of the spatial low flow variance while a regional regression model explained between 17 and 47% of the variance, depending on the region. In their study, the entire domain was subdivided into seven geographic regions. In a smaller scale study of 44 catchments in New Zealand, Clausen and Pearson (1995) showed that the variance of a drought index explained by catchment characteristics increased from 64% to between 68 and 91% if the domain is split into three geographically defined regions.

In some instances it is clear how to group a domain into regions of approximately uniform low flow behaviour but, more often, the choice is far from obvious. A number of methods of identifying homogeneous regions have, therefore, been proposed in the literature in the context of low flow regionalisation. All of these methods use low flow data and most of them use catchment characteristics as well. In the first technique, termed residual pattern approach, residuals from an initial, global regression model between flow characteristics and catchment characteristics are plotted, from which geographically contiguous regions are obtained by manual generalisation on a map (e.g. Hayes, 1992; Aschwanden and Kan, 1999). This is an obvious method of improving on a global regression model. A drawback of the residual pattern approach, however, is that the initial model may be far from correct as it extends over the entire domain of interest. The shapes of the regions so obtained may then be artefacts of an inadequate model and the regional regression model will have little physical significance. Once the regions have been identified, the ungauged site of interest needs to be allocated to one of the regions. As the regions in this approach are spatially contiguous, the ungauged site can be allocated by its geographical location. As a final step, the low flow value for the site of interest is estimated from multiple regressions between observed low flows and catchment characteristics fitted to each of these regions independently.

In the second technique, multivariate statistics such as cluster analyses are used to delineate regions. In the multivariate analyses, both low flow data and catchment characteristics are used. They are usually standardised and/or weighted to enhance the discriminatory power of the methods. The use of multivariate statistics in the context of low flow regionalisation has been explored in detail by Nathan and McMahon (1990). They tested a number of approaches based on a combination of different techniques of cluster analysis, multiple regression and principal component analysis. They used Andrews curves (Andrews, 1972) for visualising similarity in catchment characteristics which allowed them to fine-tune the catchment grouping. Based on data from 184 catchments in south-east Australia, Nathan and McMahon (1990) found that the relative performance of the methods depended on the low flow characteristic examined. Their overall assessment suggested that the weighted cluster analysis (Ward's method based on a Euclidean distance measure) using weights according to the coefficients of an initial stepwise regression model performed best. Since regions obtained by the cluster analysis approach are generally discontiguous in space, the allocation of ungauged sites to the most similar group requires decision criteria, which are usually based on catchment characteristics. Nathan and McMahon

(1990) assumed in their analysis that the catchment allocation was known and proposed to use Andrews curves for assigning ungauged catchments. Possible alternative methods are discriminant analyses and classification trees (Haines et al., 1988). As a final step, again, the low flow value for the site of interest is estimated from multiple regressions between observed low flows and catchment characteristics fitted to each of the regions independently.

A third technique are Classification And Regression Tree (CART) models (Breiman et al., 1984) which, to our knowledge, have not yet been used in low flow regionalisation. However, there do exist a number of interesting applications in hydrology, including the classification of satellite images of snow cover and the interpolation of ground snow measurement (e.g. Rosenthal and Dozier, 1996; Elder, 1995) and a first attempt of using the regression trees for regionalizing low flows is given in Laaha (2002). In the context of low flow regionalisation, the independent variables in the regressions trees are the catchment characteristics and the dependent variables are the low flows. Regression trees then divide a heterogeneous domain into a number of more homogeneous regions by maximising the homogeneity of low flows and catchment characteristics within each group simultaneously. Regression trees have a number of advantages over other models. Their structure is non-parametric, small trees are readily interpretable, there is no global sensitivity to outliers and they are able to handle non-linear relationships well. However, big trees are difficult to interpret, there is a lack of smoothness and there are potential problems with overfitting the data, so some method for pruning the tree is needed (Breiman et al., 1984). Once the regression tree is fitted to the data, it can be used to allocate ungauged sites to the regions obtained by the regression tree. Alternatively, classification trees can be used to allocate group names to catchment characteristics. Classification trees operate on categorical variables while regression trees operate on continuous variables. The final step of estimating low flows for the ungauged site of interest is a regional regression as in the other grouping methods.

In a fourth technique, the seasonality of low flows is used to delineate homogeneous regions. The rationale of this approach is that differences in the occurrence of low flows within a year are a reflection of differences in the hydrologic processes and are hence likely to be useful for finding homogeneous regions. Merz et al. (1999); Piock-Ellena et al. (2000) have illustrated that the seasonality approach is indeed useful in the context of flood frequency regionalisation in Austria. They used a cluster analysis based on circular statistics of flood occurrence within the year to identify homogeneous regions and plotted vector maps to visualise the spatial patterns of the seasonalities of floods and other hydrologic variables. In contrast, an application of a low flow seasonality index in the UK (Young et al., 2000) suggested there is little discriminatory power in this index because the spatial variability of low flow seasonality was rather weak. It is clear that the usefulness of this method hinges on the existence of clear spatial patterns in low flow seasonality. Laaha (2002) compared two seasonality measures in upper Austria and found that both measures were capable of classifying catchments into summer and winter low flow dominated sub-regions. An extension of this work is Laaha and Blöschl (2006) who visually delineated homogenous regions with respect to low flow seasonality from a number of seasonality measures. Their results indicated that, in a humid, mountainous country such as Austria, the spatial variations in low flow seasonality are indeed enormous. There is likely some potential in this approach. If the regions are spatially contiguous such as those of Laaha and Blöschl (2006), the ungauged site can be allocated by its geographical location. The final step of estimating low flows for the ungauged site of interest is an analogous regional regression to the other grouping methods.

While much work has been done in the literature on catchment grouping in the context of low flow regionalisation we are unaware of any comprehensive comparison of the grouping methods for the same data set to assess their relative merits. The aim of this paper, therefore, is to examine the relative performance of different grouping techniques to investigate what is the optimum grouping method for regionalising low flows. The comparison will be made on a comprehensive Austrian data set and the low flow characteristic chosen is the q_{95} specific low flow quantile, i.e. the specific discharge that is exceeded on 95% of all days.

2. Data

2.1. Study area

The study has been carried out in Austria, which is physiographically quite diverse. There are three main zones in terms of the geographical classification, high Alps in the west, lowlands in the east, and there is hilly terrain in the north (foothills of the Alps and Bohemian Massif). Elevations range from 117 to 3798 m a.s.l. Geological formations vary significantly, too. Austria has a varied climate with mean annual precipitation ranging from 500 mm in the eastern lowlands up to about 2800 mm in the western Alpine regions. Runoff depths range from less than 50 mm per year in the eastern part of the country to about 2000 mm per year in the Alps. Potential evapotranspiration ranges from about 730 mm per year in the lowlands to about 200 mm per year in the high alpine regions. This diversity is reflected in a variety of hydrologic regimes (Kresser, 1965) and low flows exhibit important regional differences in terms of their quantity and their seasonal occurrence (Laaha and Blöschl, 2003).

2.2. Discharge data

Discharge data used in this study are daily discharge series from 325 stream gauges. These data represent a complete set of gauges for which discharges have been continuously monitored from 1977 to 1996 and where hydrographs have not been seriously affected by abstractions and karst effects during low flow periods (Laaha and Blöschl, 2006). Catchments for which a significant part of the catchment area lies outside Austria have not been included as no full set of physiographic data was available for them. The catchments used here cover a total area of 49,404 km^2, which is about 60% of the national territory of Austria. Although a larger number of catchments are monitored in Austria, in this paper priority is given to a consistent observation period to make all records comparable in terms of climatic variability.

2.3. Disaggregation of nested catchments

Nested catchments were split into sub-catchments between subsequent stream gauges based on the hierarchical ordering of gauges presented in Laaha and Blöschl (2003). The advantage of using sub-catchments rather than complete catchments is that the application of regionalisation techniques to small ungauged catchments is more straightforward. Also, discharge characteristics of nested catchments are statistically not independent and disaggregation into sub-catchments between subsequent stream gauges makes them more independent. The disadvantage of the disaggregation is that errors may be somewhat larger, as the low flow characteristics are estimated from differences of the stream flow records at two gauges.

2.4. Low flow characteristic

Low flows were quantified by the Q_{95} flow quantile [$\Pr(Q > Q_{95}) = 0.95$], i.e. the discharge that is exceeded on 95% of all days of the measurement period. This low flow characteristic is widely used in Europe and was chosen because of its relevance for multiple topics of water resources management (e.g. Kresser et al., 1985; Gustard et al., 1992; Smakhtin, 2001). For gauged catchments without an upstream gauge the Q_{95} low flow quantile was calculated directly from the stream flow data. For sub-catchments Q_{95} was calculated from the differences of stream flows at the two gauges. Q_{95} was subsequently standardised by the catchment area to make the low flow characteristic more comparable across scales. The resulting specific low flow discharges q_{95} ($\mathrm{l\,s^{-1}\,km^{-2}}$) were considered to be representative of the characteristic unit runoff from the catchment area during sustained dry periods.

2.5. Catchment characteristics

We used 31 physiographic catchment characteristics in the low flow regionalisation in this paper. They relate to sub-catchment area (A [10^1 km^2]), topographic elevation (H), topographic slope (S), precipitation (P), geology (G), land use (L), and stream network density (D [10^2 m/km^2]). Topographic elevation is represented by the altitude of the streamgauge (H_0 [10^2 m], maximum altitude (H_+[10^2 m]), range of altitude (H_R [10^2 m]) and mean altitude (H_M [10^2 m]). Topographic slope (S) is represented by the mean slope (S_M [%]), and by area percentages of slight slope (S_{SL} [%]), moderate slope (S_{MO} [%]), steep slope (S_{ST} [%]). Precipitation (P) is represented by average annual precipitation (P [10^2 mm]), average summer precipitation (P_S [10^2 mm]), and average winter precipitation (P_w [102 mm]. Geology (G) is represented by the area percentages of Bohemian Massif (G_B [%]), Quaternary sediments (G_Q [%]), Tertiary sediments (G_T [%]), Flysch (G_F [%]), Limestone (G_L [%]), Crystalline rock (G_C [%]), shallow groundwater table (G_{GS} [%]), deep groundwater table (G_{GD} [%]), source region (G_{SO} [%]). Land use (L) is represented by the area percentages of urban (L_U [%]), agriculture (L_A [%]), permanent crop (L_C [%]), grassland (L_G [%]), forest (L_F [%]), wasteland-rocks (L_R [%]), wetland (L_{WE} [%]), water surfaces (L_{WA} [%]), glacier (L_{GL} [%]). All characteristics were first compiled on a regular grid and then combined with the sub-catchment boundaries of Laaha and Blöschl (2003); Behr (1989) to obtain the characteristics for each catchment. The catchment characteristics used in this paper are discussed in more detail in Laaha and Blöschl (2005).

3. Method

3.1. Classification of catchments

3.1.1. Residual pattern approach

The residual pattern approach to catchment grouping consisted of three steps:

(1) Perform stepwise regression to obtain a global regression model;
(2) Plot the residuals from the global regression model in geographic space;
(3) If residual patterns are apparent, delineate contiguous regions of similar sign and magnitude of residuals.

Stepwise regression may lead to over-fitted models where omission of a single catchment characteristic only slightly reduces the global model quality. When choosing the number of catchment characteristics in the global regression we therefore tended to use the more parsimonious model as it produced clearer residual patterns.

3.1.2. Weighted cluster analysis

Weighted cluster analysis has been recommended by Nathan and McMahon (1990) as the optimal technique to identify homogeneous regions and we used their method consisting of the following steps:

(1) Identify the catchment characteristics most relevant to the problem at hand by performing an overall stepwise regression analysis;
(2) Weight the selected catchment characteristics according to their relative importance, as determined by the magnitude of their β-coefficients which are the coefficients of the stepwise regression model based on standardised catchment characteristics;
(3) Perform a number of cluster analyses on the weighted catchment characteristics using different measures of similarity and linkage methods;
(4) Prepare plots of Andrews curves for each of the groupings derived in (3), and identify the set of clusters exhibiting the least within-group variation. This will give the optimal classification of catchments into homogeneous groups;
(5) Remove outliers in the optimum grouping based on the Andrews plots, if needed. Derive a set of mean catchment characteristics for each homogeneous group;
(6) Refine the optimum grouping derived by the cluster analysis by comparing the catchment characteristics of each catchment with the group mean and reclassify the catchment in case the catchment characteristics are too different.

In the spirit of Nathan and McMahon (1990), several cluster analysis techniques of the S-Plus statistics package were compared. These were two hierarchical cluster analysis methods, *hclust* (Hartigan, 1975) and *agnes* (Kaufman and Rousseeuw, 1990), which are similar to the algorithm used by Nathan and McMahon, as well as the *pam* partitioning method (Kaufman and Rousseeuw, 1990). Several combinations of linkage methods (single linkage, average linkage and complete linkage) and distance measures (Euclidean distance and Manhattan distance) were evaluated for different numbers of clusters. The most appropriate method was selected by a visual assessment of Andrews plots. In Andrews plots, a point in multi-dimensional space $x=[x_1, x_2,\ldots,x_n]$ is represented by a function of the form:

$$F(t) = x_1/\sqrt{2} + x_2\sin(t) + x_3\cos(t) + x_4\sin(2t) + x_5\cos(2t) + \cdots \quad (1)$$

plotted over the range of $-\pi \leq t \leq +\pi$. A set of multivariate observations can be displayed as a collection of these plots and those functions that remain close together for all values of t correspond to observations that are close to one another in terms of their Euclidean distance. This property implies that these plots can be used to both detect groups of similar observations and identify outliers in multivariate data.

Since the regions obtained by weighted cluster analysis, generally, are not contiguous, the prediction of low flow characteristics at ungauged sites requires a decision rule based on catchment characteristics in order to allocate the site of interest to the most appropriate region. Nathan and McMahon proposed a procedure similar to step (6), i.e. comparing the Andrews curve of an ungauged catchment with the mean curve of each cluster. Because of the subjectivity of a visual assessment, this method is not suitable for automatic cross-validation of the regional regression model. We therefore adopted an alternative approach and used classification trees for automatically allocating ungauged catchments to the most appropriate cluster. Similarly to the regression trees (see below), the classification tree was fitted based on 10-fold cross-validation to determine the optimum tree size for prediction.

3.1.3. Regression tree

In this paper, regression trees are proposed for obtaining homogeneous regions to be used in a regional regression approach. Regression trees are an exploratory technique for finding homogeneous regions among predictor variables (i.e. catchment characteristics) with respect to a target variable (i.e. q_{95} low flow). The regression tree is constructed by an algorithm known as *binary recursive partitioning* (Clark and Pregibon, 1991). By this algorithm, groups of catchments are subsequently subdivided by binary conditions (e.g. IF $P_s < 534$ mm THEN sub-group x ELSE sub-group y), starting from the most important catchment characteristics and proceeding to the less important ones. Each condition yields the optimal subdivision of a group, which minimises the sum of squared differences between observed values of q_{95} and the group mean, a measure that is termed the *deviance* of the node. The algorithm identifies the most important catchment characteristics, and potential interactions between catchment characteristics are handled implicitly (Venables and Ripley, 1999).

Tree construction can be carried out until each terminal node consists of one single catchment but this leads to a model with little significance for prediction or classification problems. To avoid such over-fitting, trees need to be pruned back, and the optimal number of nodes is best determined by an independent validation data set. If no such validation data set is available, one can split the data set into 10 (roughly) equally sized parts, subsequently use nine parts for calibration and one part for validation, and calculate the average prediction error (total deviance of a tree) for several tree sizes. This procedure, termed 10-fold cross-validation, is part of the S-Plus *tree*-package and was used in this study.

Regression trees have the convenient property of invariance against monotone transformation of predictor variables (i.e. catchment characteristics). However, the dependent variable (i.e. q_{95}) needs to be normally distributed for optimal tree construction. We examined the distribution of q_{95} in the data set of this paper and found that a square-root transformation of q_{95} yields a distribution that is close to normal. Since the regression tree is used for classification but not for prediction, no retransformation is needed which may be non-unique if the transformed variable changes sign.

The regression tree approach to catchment grouping consisted of the following steps:

(1) Perform transformation to normality;
(2) Fit an initial regression tree to the data;
(3) Determine the optimal tree size by 10-fold cross-validation;
(4) Prune the initial tree back to the tree size derived in (3).

While regression trees are suitable for allocating unobserved catchments to the most appropriate clusters, they are not suitable for cross-validation of the resulting regional regression model as the names of the clusters may change when models are refitted for subsets of the data. We therefore fitted a classification tree to the group names of the regression tree as (categorical) dependent variable, which exhibited an identical structure to the regression tree, but had the advantage of producing the same group names for various data subsets. This allowed us to assign each ungauged catchment to one of the clusters of the regression tree in the cross-validation.

3.1.4. Regions of similar low flow seasonality

Regions of similar low flow seasonality as defined by Laaha and Blöschl (2003) were used as the final scheme for catchment grouping. Laaha and Blöschl (2003) classified Austria into eight contiguous regions based on a visual assessment of two seasonality measures. The first seasonality measure was a seasonality index based on circular statistics (Young et al., 2000) that represented the mean and the variance of the days of low flow occurrence. The second seasonality measure were seasonality histograms (Laaha, 2002) which were used to refine the information from the seasonal statistics. Catchment elevation was used to assist in the delineation of the regions as, in Austria, topographic elevation is one of the main controls of hydrologic regimes. The method consisted of the following steps:

(1) Determine the Julian dates (i.e. days from 1 to 365) of days of low flow occurrence for each sub-catchment by selecting all days when daily discharge was below Q_{95};
(2) Calculate the seasonality index from the dates for each sub-catchment and plot the seasonality indices as a vector-map in geographical space;
(3) Delineate preliminary regions on the vector map;
(4) Plot monthly histograms of low flow occurrence for each sub-catchment and use them to refine the preliminary classification;
(5) Use topographic elevation to refine the exact position of the region boundaries.

3.2. Regional regression approach

For each group identified by the classification methods, a multiple regression model was fitted independently with specific low flow discharge q_{95} as the dependent variable and a set of catchment characteristics as the independent variables. Catchment characteristics are often subject to inter-correlations and multicollinearity. Rather than performing a selection of the most important variables prior to regionalisation, we used a stepwise regression approach. The stepwise regression procedure used *Mallow's* C_p (Weisberg, 1985, p. 216) as the criterion of optimality, which was calculated as:

$$C_p = \frac{\mathrm{RSS_p}}{\hat{\sigma}^2} + 2p - n \qquad (2)$$

The first term is the residual sum of squares of one considered model (RSS_p) with p coefficients divided by the residual error variance $\hat{\sigma}^2$ of the full model and corresponds to the relative optimality in terms of model error. Complexity of models is penalised by the second term, which adds the number of coefficients p minus the number of catchments n. C_p is therefore a penalised selection criterion which takes the gain of explained variance as well as the parsimony of models into account and yields models that are optimal in terms of prediction errors. Variable selection starts with one arbitrarily chosen catchment characteristic and subsequently adds variables that minimise the C_p criterion. After each step it is tested if replacing one of the variables by any remaining catchment characteristic will further decrease the criterion. The selection procedure continues until C_p reaches a minimum. The catchment characteristics obtained by the stepwise regression can hence be interpreted as important controls of low flows.

Fitting regression models in hydrology is often complicated by single extreme values or outliers. Eliminating outliers may improve the goodness-of-fit but this does not necessarily entail an increase in the predictive power of the model. On the other hand, extreme values may act as leverage points and force the fitted model close to them, particularly if the regression model is fitted by the least squares method, which increases the magnitude of the residuals of the remaining points. We therefore adopted an iterative robustified regression technique in this paper. Initial models were fitted by stepwise regression and then checked for leverage points using Cook's distance (e.g. Weisberg, 1985). These leverage points were removed from the sample and the regression model was refitted iteratively until no leverage points remained. The final model quality was assessed for all data including leverage points. q_{95} was used in all regional regressions without transformation, as exploratory analyses of the data suggested that transformations did not increase the predictive performance.

The regression models so obtained were checked for numerical stability of computation. Since numerical stability is sensitive to different scales of predictors, all catchment characteristics had been scaled by integer powers of 10 to give similar magnitudes in terms of their ranges (Section 2.5). Since linear regression is scale invariant (Weisberg, 1985, p. 185) the regression models, including their residual statistics, remain unaffected by the rescaling but the numerical stability is improved.

3.3. Analysis of predictive performance

3.3.1. Analysis of variance

In a first step, we were interested in how well the classification into homogeneous regions may explain the spatial variability of specific low flow discharges, q_{95}. A widely used measure of the explanatory power of groupings is the one-factorial analysis of variance (ANOVA) which was used here with q_{95} as the dependent variable and the classification number as the independent variable. The ANOVA may be interpreted as an assessment of a simple regionalisation model, where predicted q_{95} is simply the average low flow discharge in each group of a classification. The coefficient of determination (R^2) of this model, i.e. the ratio of the variance explained by the classification and the total variance of low flows, is a measure of the goodness-of-fit of this simple model. R^2 values close to 100% indicate an excellent goodness-of-fit while smaller values indicate a poorer goodness-of-fit.

3.3.2. Goodness-of-fit of component regressions

In a second step we examined how well the regression models in each of the regions fitted the data. We assessed the goodness-of-fit by the coefficient of determination of the regressions separately in each of the regions.

3.3.3. Cross-validation of regional regression

The value of the classification methods for the ultimate purpose of estimating low flow characteristics at ungauged sites cannot be fully assessed by goodness-of-fit statistics. A more appropriate measure of the prediction errors are the error statistics from leave-one-out cross-validation. In this paper, the cross-validation procedure consisted of the following steps:

(1) Remove catchment i from the data set;
(2) Update the catchment classification (if appropriate) for the remaining $n-1$ catchments;
(3) Assign catchment i to one of the regions obtained in (2);
(4) Estimate the coefficients of the regression equation for this region using all catchments in this region apart from catchment i;
(5) Apply the regression equation obtained in (4) to predict the low flow characteristic q_{95} at site i;
(6) Repeat steps (1)– (5) for all n catchments;
(7) Calculate the predictive error for each catchment i as q_{95} estimated in (5) minus observed q_{95} and analyse the error statistics.

In some of the classification methods the catchment classification was updated during the cross-validation procedure while in others it was not. In the weighted cluster analysis and the regression tree approaches the regions are discontiguous, and will hence significantly change if a single catchment is added. In these methods the classification was updated. In the residual pattern and the seasonality region approaches, however, the regions are contiguous and will therefore not change much if a single catchment is added. In these methods the classification was not updated.

From this prediction vector, the cross-validation prediction error V_{cv} was then estimated by:

$$V_{\mathrm{cv}} = \frac{1}{n}\sum_{i=1}^{n}(\hat{q}_{95i}^{(-i)} - q_{95i})^2 \tag{3}$$

where q_{95i} is the observed specific low flow discharge q_{95} for catchment i and $\hat{q}_{95i}^{(-i)}$ is the model prediction without using observed low flows from catchment i. The root mean squared error based on cross-validation is therefore

$$\mathrm{rmse}_{\mathrm{cv}} = \sqrt{V_{\mathrm{cv}}} \tag{4}$$

and the coefficient of determination based on cross-validation is:

$$R_{\mathrm{cv}}^2 = \frac{V_{\mathrm{q}} - V_{\mathrm{cv}}}{V_{\mathrm{q}}} \tag{5}$$

where V_q is the spatial variance of the observed specific low flow discharges q_{95}.

The advantage of cross-validation over other techniques of assessing predictive errors is its robustness and its general applicability to all regionalisation models. This is because cross-validation works well even if the regionalisation models are far from correct (Efron and Tibshirani, 1993). Cross-validation is hence a full emulation of the case of ungauged sites.

4. Results

4.1. Residual pattern approach

A preliminary global regression model was fitted to the data by stepwise regression. Since the primary purpose of the global model was to calculate a meaningful residual pattern, the residuals were carefully checked for the general assumptions underlying multiple regression, unbiasedness ($E[res_i]=0$) and homoscedasticity ($Var[res_i]=$constant), where res_i is the residual of catchment i. The analysis indicated slight heteroscedasticity which appeared to be a consequence of a significant skew of the distribution of q_{95}. We therefore transformed q_{95} by a square-root transformation which resulted in approximate normality. The global regression model was then fitted to the transformed data. The retransformation is non-unique if the variable changes sign but since all predictions were positive this was not a problem.

Stepwise regression resulted in seven catchment characteristics used as predictors. This equation was manually revised and the three predictors that contributed least to the model performance were removed to avoid overfitting. There was only a slightly loss in the goodness-of-fit when removing these predictors (R^2 decreased from 66 to 62%). The more parsimonious model indicated a clearer residual pattern than the full model based on seven predictors and hence seemed to be more suitable for detecting homogeneous regions. The residual map is presented in Fig. 1a. The residual pattern suggests that Austria can be classified into two main units. The first unit consists of flatlands and hilly terrain. In this unit, the magnitude of the residuals is generally low ($<1\,\mathrm{l\,s^{-1}\,km^{-2}}$ for most catchments, except for East-Tyrol) and the pattern of the residuals is random, so the global model seems to work well in this unit. The second unit consists of the Alpine catchments and the Molassezone in the North. In this unit, the magnitude of the residuals is larger although there are no clear patterns. We chose to subdivide the second unit into four regions based on the geology. This gave us a total of five regions as shown in the summary plot of Fig. 7(a). Region 0 relates to small residuals, region 1 relates to negative residuals, and the remaining regions 2–4 relate to positive residuals.

The coefficient of determination of this classification calculated by one-way ANOVA was $R^2=25\%$ (Fig. 6) which means that this classification explains 25% of the total spatial variance of the specific low flow discharges q_{95}. Although this is not much, the

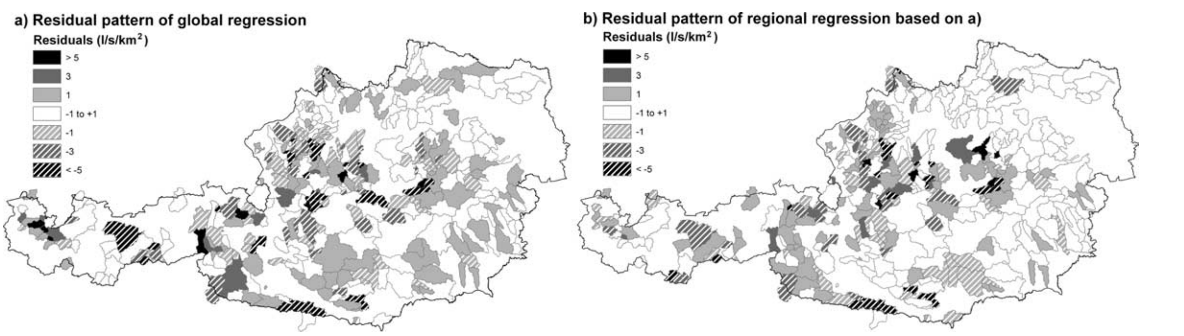

Fig. 1. Residual pattern of (a) the global regression model (goodness-of-fit residuals) and (b) the here from obtained regional regression model (cross-validated residuals). Positive residuals indicate an overestimation by the model.

Table 1
Components of the regional regression model based on the residual pattern approach

Group	Region	R^2 (%)	Model
0	N, E, SE of Austria, E-Tyrol, W-Tyrol	87	$\hat{q}_{95} = -3.46 + 0.67P - 0.19L_{GL} - 0.03G_F + 0.10S_M$
1	Central-Alps and Pre-Alps	60	$\hat{q}_{95} = -0.81 + 0.69P + 0.41H_R - 0.52H_M + 0.08S_M$
2	Part of the Northern Calcerous Alps	15	$\hat{q}_{95} = 7.66 + 0.12G_Q$
3	Carinthia	82	$\hat{q}_{95} = 1.51 + 1.02D - 0.08S_{MO}$
4	Bregenzerwald (Vorarlberg)	32	$\hat{q}_{95} = 14.49 - 0.12S_{MO}$

R^2 denotes the goodness-of-fit coefficient of determination. Symbols see Sections 2.4 and 2.5.

delineated regions were used as a basis for a regional regression model. The model consisted of five independent regionally restricted models. A statistical summary of these component models is presented in Table 1. Three out of the five regions are well represented by the regional models (regions 0, 1, and 3). However, the regression models for region 2 (Northern Calcerous Alps) and region 4 (Bregenzerwald) indicate very poor model performance, which suggests that there may be significant heterogeneity of low flow processes within these regions. Note that R^2 represents the model goodness-of-fit coefficient of determination and hence does not fully capture the predictive performance for ungauged sites.

The predictive performance of the complete regional regression model was finally checked by cross-validation. Ungauged catchments were assigned based on the regions in Fig. 7(a). The overall predictive performance was found as $R^2_{cv} = 63\%$. This is significantly better than the coefficient of determination of the classification (goodness-of-fit $R^2 = 25\%$). This improvement is also apparent when comparing the residual pattern of the global regression model (Fig. 1(a)) with that of the regional regression model (Fig. 1(b)). The latter pattern is more random and the magnitudes of the residuals are significantly smaller. This means that there is a lot of value in using regionally restricted regression models over one single, global regression model.

4.2. Weighted cluster analysis

For the weighted cluster analysis, all catchment characteristics were standardised to zero mean and unit variance. A stepwise regression was then conducted between q_{95} and the standardised catchment characteristics in order to identify the most relevant catchment characteristics. The catchment characteristics so obtained and the respective β-coefficients of the regression are presented in Table 2. These β-coefficients were checked for plausibility and subsequently used as weights in the weighted cluster analysis.

A number of cluster analyses were carried out, combining different distance measures and linkage methods for a range of numbers of clusters. In each case, the homogeneity of the groups was assessed by a visual examination of Andrews plots. This comparison suggested that the hierarchical cluster analysis (*agnes*) that combines Ward's method and a Euclidean distance metric (using 10 clusters) was preferable to other methods and slightly preferable to the *pam* partitioning method (10 clusters, Euclidean metric). Fig. 2 shows the Andrews curves for the optimum classification method (*agnes*, 10 clusters). Each panel represents a cluster and each line corresponds to one catchment. The x_i of Eq. (1) are the catchment characteristics in Table 2 from left to right, standardised to zero mean and unit variance, and weighted by the β-coefficients. The Andrews curves were subsequently examined for

Table 2
Catchment characteristics and associated weights obtained by a preliminary stepwise regression

Catchment characteristic	H_R	L_R	G_F	P_W	G_{GD}	S_M	G_Q
Weight (β-coefficient)	0.22	−0.27	−0.12	0.42	0.13	0.33	0.11

Symbols see Section 2.5.

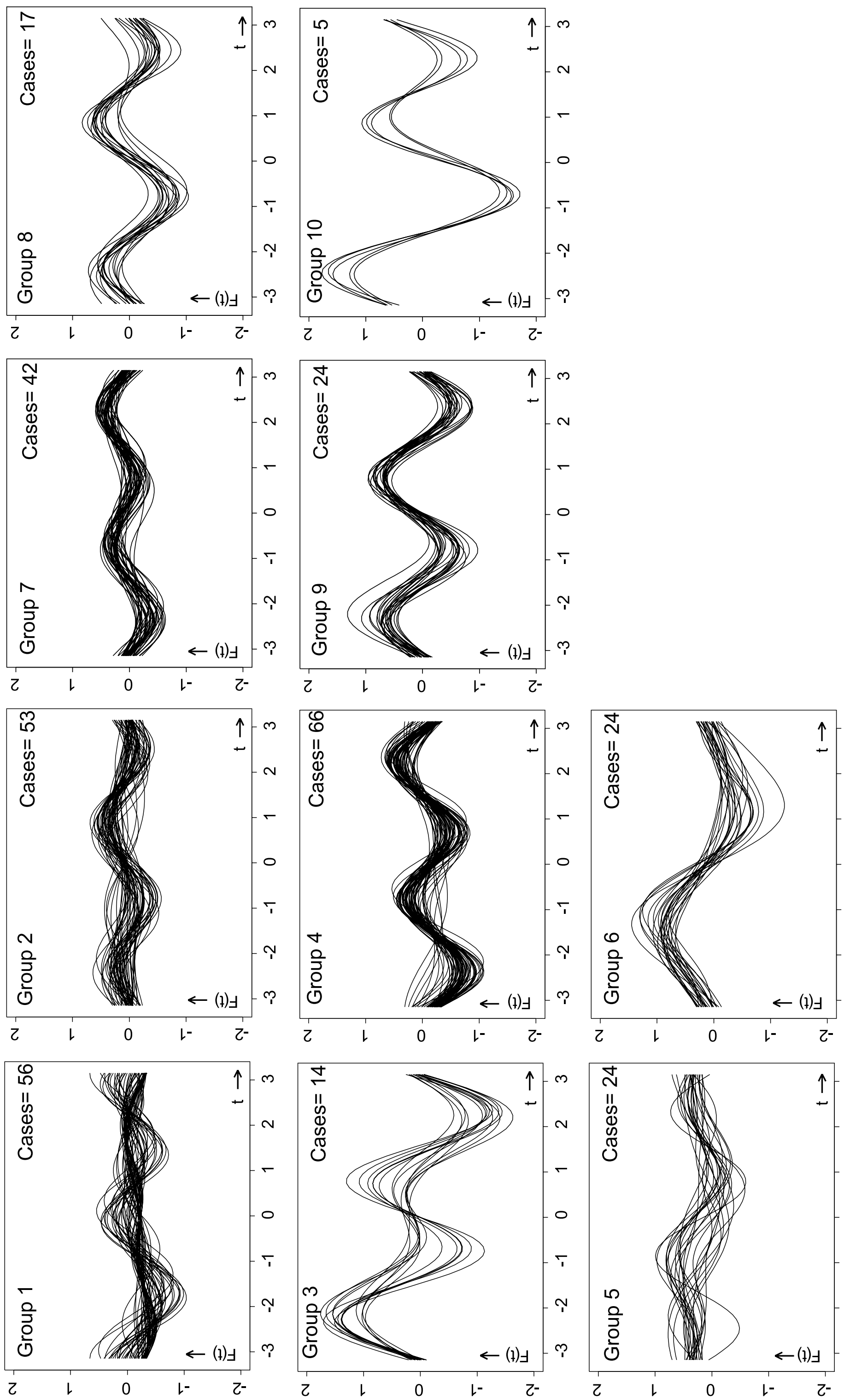

Fig. 2. Andrews curves of weighted catchment characteristics.

homogeneity. Overall, the between-group variability is much larger than the within-group variability, although in groups 4 and 5 individual catchments appear to be different from the rest. However, given that a robustified regression technique was used which gives little weight to single outliers, we deemed the groups sufficiently homogeneous for the further analysis. We were hence able to avoid any subjective steps of manual re-classsification of outliers. The coefficient of determination of the classification by the weighted cluster analysis alone (i.e. without regional regressions) was $R^2 = 56\%$ (Fig. 6) which means that this classification explains 56% of the total spatial variance of the specific low flow discharges q_{95}. This is significantly more than that of the residual patterns approach.

In a next step, the clusters were plotted on a map (see summary plot in Fig. 7(b)). Even though the cluster analysis did not use any information on the geographical location of catchments, most of the clusters are contiguous and there are only some of the Alpine catchments that are scattered in terms of their location. This result gives additional credence to the weighted cluster analysis approach. The spatial contiguity of the regions is apparently related to the spatial dependence of the weighted catchment characteristics.

Regression models were then fitted to each of the regions independently. They are shown in Table 3. Most regions are represented rather poorly by the respective multiple regression models. For some regions (regions 8, 6, 4, 3), however, the model performance is very good. These differences may be related to the weights of the catchment characteristics. Constant weights have been used across the entire study domain, which may be more appropriate in some parts of the domain than in others, as local deviations from the average behaviour may exist. The catchment characteristics used in the context of a weighted cluster analysis are hence not able to fully represent regional anomalies in the low flow patterns.

Even though most of the clusters in Fig. 7(b) were coherent we did not judge them to be sufficiently contiguous for allocating ungauged catchments to regions in a unique way. The grouping of Fig. 7(b) was therefore approximated by a classification tree (Fig. 3). The quality of approximation was assessed by the misclassification error, which is the ratio of misclassified catchments and all classified catchments. The overall misclassification error is 21 out of 325 catchments (i.e. $21/325 = 0.06$) which represents an excellent approximation to the grouping from the weighted cluster analysis. Fig. 3 shows in detail what catchment characteristics are most significant in representing the clusters. This result is similar to the weights found by the regressions using standardised catchment characteristics in Table 2. Note that region 10 does not appear in the classification tree as the number of catchments is very small in this region. Also note that some of the catchment groups appear in two nodes (e.g. group 4) which means that this group consists of both terminal nodes in the classification tree.

The predictive performance of the complete regional regression model was finally examined by cross-validation, using the classification tree of Fig. 3 for assigning ungauged catchments to the regions. The cross-validation gave a predictive performance of $R^2_{\mathrm{cv}} = 59\%$. Although the variance explained by the

Table 3
Components of the regional regression model based on the weighted cluster analysis

Group	Region	R^2 (%)	Model
1	Upper Austria	35	$\hat{q}_{95} = 8.30 + 5.45H_0 + 2.01A - 1.08L_F + 1.37P_S$
2	Central Alps	32	$\hat{q}_{95} = 8.20 + 2.07G_Q + 3.62P_W + 0.91A$
3	Northern Calcerous Alps I	66	$\hat{q}_{95} = 9.36 - 2.10S_{MO} + 2.60G_F$
4	Flatland and hilly terrain (N, E of Austria)	67	$\hat{q}_{95} = 4.66 + 2.45P - 0.30G_F$
5	High Alps I (Tyrol, Carinthia)	44	$\hat{q}_{95} = 7.75 + 3.26P_S$
6	High Alps II (Tyrol, Carinthia)	70	$\hat{q}_{95} = -1.67 + 4.24S_M$
7	Low Alps (Styria and Carinthia)	41	$\hat{q}_{95} = 5.89 + 1.69H_+ - 0.87S_{MO}$
8	Flyschzone (Upper- and Lower Austria)	75	$\hat{q}_{95} = 17.35 - 1.98G_F + 11.04A$
9	Northern Calcerous Alps II	32	$\hat{q}_{95} = 10.65 - 1.87D + 3.55G_Q$
10	Pre-alps (Bregenzerwald)	0	$\hat{q}_{95} = 8.45$

R^2 denotes the goodness-of-fit coefficient of determination. Symbols see Section 2.4 and 2.5.

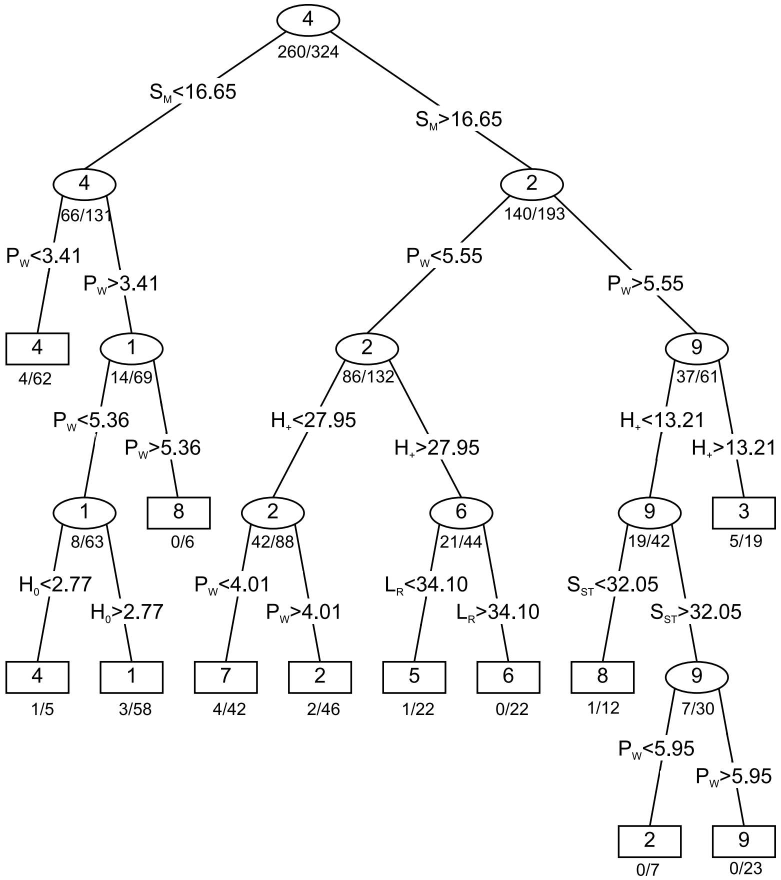

Fig. 3. Approximation of the classification based on the weighted cluster analysis by the classification tree. Ellipses indicate interior nodes, rectangles indicate terminal nodes (groups of catchments). Numbers within nodes represent group numbers (Table 3), numbers below nodes represent misclassification error rate (misclassified catchments/classified catchments).

grouping alone was relatively large, the weighted cluster analysis does not appear to be as useful for delineating regions for the regional regressions.

4.3. Regression tree

In the regression tree approach, the target variable was the specific low flow discharge q_{95} transformed by a square-root transformation. As descriptive variables, the complete set of non-standardised catchment characteristics was used. From an initial regression tree that was completely fitted to data, the optimal tree size was determined by 10-fold cross-validation. Fig. 4 shows the cross-validated total deviance of trees of different sizes. Since the cross-validated deviance is a measure of the prediction error

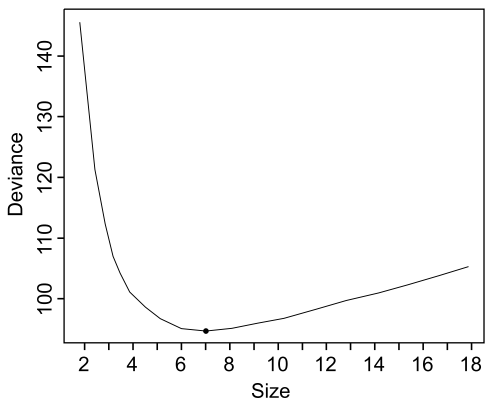

Fig. 4. Cross-validated deviance of regression trees as a function of number of splits. The minimum prediction error (cross-validated deviance) is obtained by a tree size of seven terminal nodes.

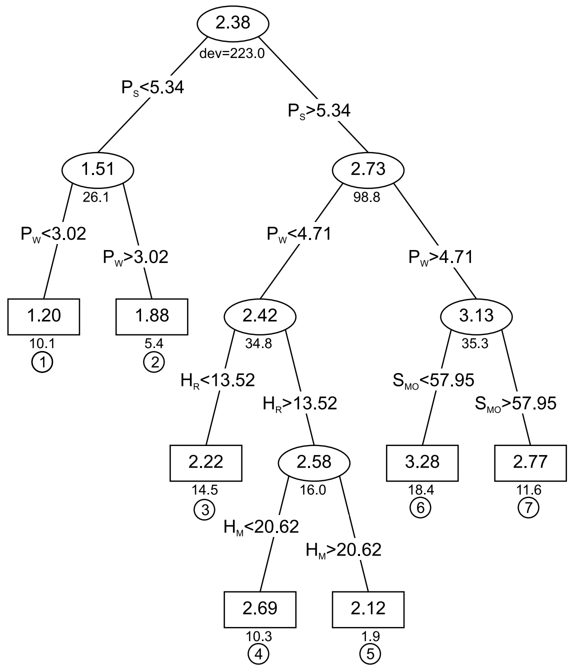

Fig. 5. Regression tree model. Ellipses indicate interior nodes, rectangles indicate terminal nodes (groups of catchments), circles represent group numbers. Numbers within nodes represent node means of square root-transformed specific discharge q_{95}, numbers below nodes represent node deviances in terms of square root-transformed specific discharge.

of the model, the optimum size of the regression tree is where the prediction error is at a minimum. Fig. 4 indicates that the optimum size is seven nodes. The initial regression tree was then pruned back to seven nodes using cost-complexity pruning (Clark and Pregibon, 1991).

The regression tree so obtained is shown in Fig. 5 and divides Austria into seven regions. The structure of the regression tree indicates that the resulting classification partitions the landscape into regions of similar relief and similar seasonal precipitation. The variance explained by the grouping, calculated by one-way ANOVA, is 62% (Fig. 6). This is the largest value of all classification approaches. This means that the regression tree is an excellent classification method if one is interested in finding groups that are most distinct in terms of both catchment characteristics and catchment response.

Regression equations were now fitted to each region independently (Table 4). Two regions (regions 1 and 5) are well represented by the regression models, three regions (regions 2, 4, 7) exhibit a moderate model fit, and two regions (regions 3, 6) are poorly represented by the models. In the main, the goodness-of-fit of the regional regression model is similar to that of the weighted cluster analysis (Table 3). Overall, the regions so obtained are consistent with both the geographical classification of Austria and the main geological units (Fig. 7(c)). As the regions are not sufficiently contiguous to permit a unique allocation of ungauged catchments we allocated them by a classification tree. The cross-validation of regional regression estimates based on the regression tree approach was found as $R^2_{cv} = 64\%$. This is significantly better than the estimates from the weighted cluster analysis where the performance was only $R^2_{cv} = 59\%$. The main difference in terms of predictive performance of the two methods seems to be related to the allocation of ungauged catchments. The classification tree for the grouping in the weighted cluster analysis method exhibited a significantly larger misclassification rate than the classifications in the regression tree approach. It appears that one advantage of the regression tree method is a very efficient classification and allocation of ungauged catchments.

Table 4
Components of the regional regression model based on the regression tree

Group	Region	R^2 (%)	Model
1	Flatland and hilly terrain (N, E of Austria)	70	$\hat{q}_{95} = -2.28 + 0.33P + 0.04G_{GS} + 0.25H_M + 0.40S_{ST}$
2	Mühlviertel and Pre-alps (Lower Austria)	51	$\hat{q}_{95} = 2.25 - 0.60D - 0.08L_{GL} + 1.91P_W$
3	Foothills of Alps	25	$\hat{q}_{95} = -0.19 + 0.57D + 0.03G_{GD} - 0.10G_{GS}$
4	Central Alps	54	$\hat{q}_{95} = -1.99 + 0.90P - 0.20G_T + 0.11G_Q$
5	High Alps (Tyrol)	67	$\hat{q}_{95} = -9.57 + 0.30S_M$
6	Calcerous Alps I ($S_{MG} < 57.95\%$)	13	$\hat{q}_{95} = 14.68 + 0.19L_A - 0.56D$
7	Calcerous Alps II ($S_{MG} < 57.95\%$)	47	$\hat{q}_{95} = 10.51 + 0.05G_L - 1.47P_W + 0.15L_G$

R^2 denotes the goodness-of-fit coefficient of determination. Symbols see Section 2.4 and 2.5.

4.4. Regions of similar low flow seasonality

The last approach to catchment grouping considered in this study is based on types of low flow seasonality as defined by Laaha and Blöschl (2003). Most regions of the grouping of Laaha and Blöschl (2003) are contiguous with the exception of three subtypes of winter low flows (types A, B, C), which are scattered within the winter low flow dominated Alpine region. Since this approach was focused on contiguous regions, these three types were merged into one single type of winter low flows. The resulting classification consists of eight regions of approximately homogeneous seasonality (Fig. 7(d)). Since all regions are contiguous, the allocation of ungauged sites is well defined by their location and no re-classification was needed in the cross-validation procedure. Examples of the seasonal distribution of low flows for each of the regions are given in Fig. 6. From Fig. 6 it is quite clear that the seasonality of low flows shows major differences in the study domain, so one would expect seasonality to possess significant predictive power for delineating regions of similar low flow processes.

Regional regressions were now fitted independently to each of the regions. The results are summarised in Table 5. In most regions, the models fit well, with coefficients of determination ranging from 60 to 70%. The regression models for the Pre-Alps of Styria and Lower Carinthia (regions 3 and 4) exhibit even better coefficients of determinations of 89 and 83%, respectively. The exception is the Alpine, winter low flow dominated region (A–C), where the goodness-of-fit is only $R^2 = 51\%$. This low coefficient is not surprising as three types of seasonality have been lumped into a single region.

In a final step, the predictive performance for the case of ungauged catchments was assessed by cross-validation. The cross-validated coefficient of

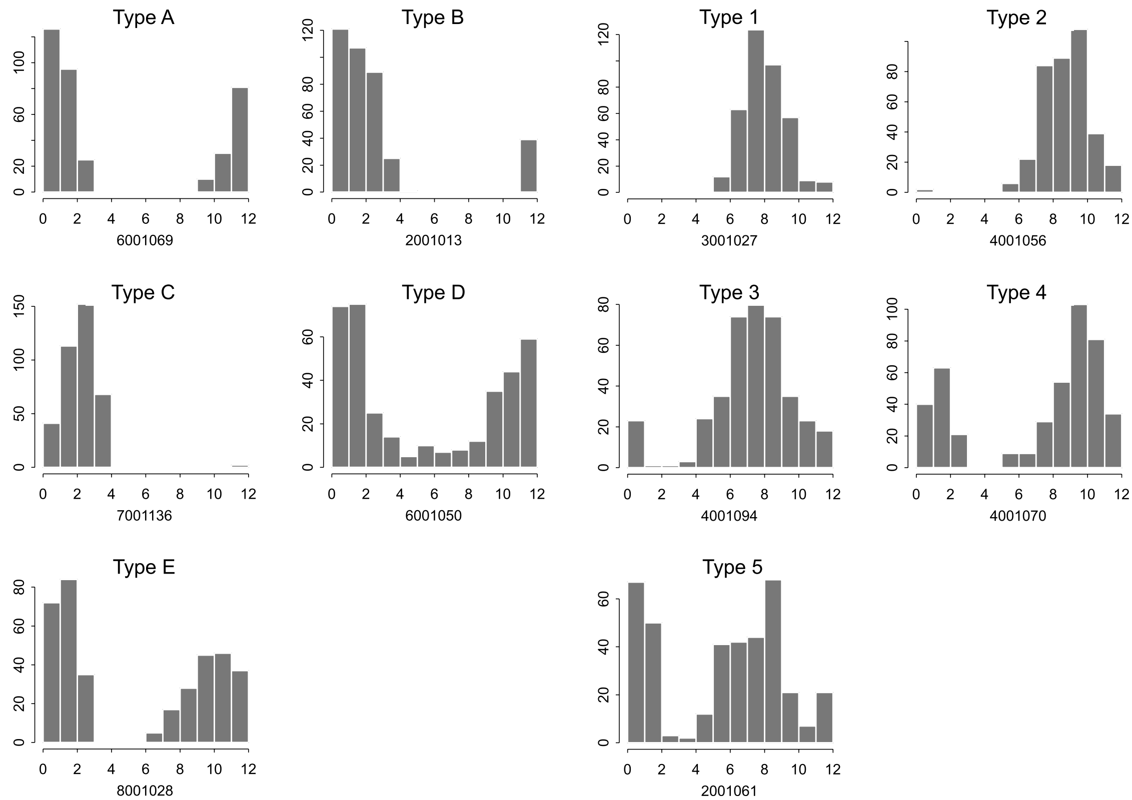

Fig. 6. Seasonality types of low flows in Austria illustrated by the non-exceedance frequencies of Q_{95} for each month for a typical catchment in each region. Letters relate to winter low flows, numbers relate to summer low flows (Fig. 7(d)).

Table 5
Components of the regional regression model based on regions of similar low flow seasonality

Group	Region	R^2 (%)	Model
A–C	Alps	51	$\hat{q}_{95} = 0.67 + 0.40P + 0.17G_Q - 0.01G_C + 6.43L_{WE} + 0.14S_M - 0.04L_R - 0.20H_0$
1	Flatland & hilly terrain (N, E of Austria)	71	$\hat{q}_{95} = -0.12 + 0.11S_M + 0.05G_{GS} + 0.02G_C$
2	Bohemian Massif	64	$\hat{q}_{95} = -3.31 + 1.96P_W$
3	Foothills of Alps (Upper Austria)	68	$\hat{q}_{95} = -10.04 - 0.76D + 3.27P - 2.22H_o$
4	Flyschzone	63	$\hat{q}_{95} = -6.17 + 0.06G_L + 2.07P_S - 0.06L_F$
5	Lower Carinthia	83	$\hat{q}_{95} = -17.48 + 3.56D + 20.06L_{WE}$
D	Pre-Alps (Styria)	89	$\hat{q}_{95} = -7.99 + 1.08P + 0.04L_F$
E	Pre-Alps (Vorarlberg)	60	$\hat{q}_{95} = 18.20 - 0.18S_{MO}$

R^2 denotes the goodness-of-fit coefficient of determination. Symbols see Sections 2.4 and 2.5.

determination for the approach based on seasonality regions was $R^2_{cv} = 70\%$. This is a better predictive performance than the other grouping methods. It appears that the stream flow characteristics as illustrated in Fig. 6 contain a lot of information highly relevant to low flow regionalisation.

5. Discussion

5.1. Variance explained by grouping alone (ANOVA)

In a first step of comparing the methods of catchment grouping we examined the part of the variance (R^2) of specific low flows q_{95} that can be explained by the grouping alone without using regressions. The R^2 values are large if the variability between the estimated group means of q_{95} (SS_G) are large relative to the variability of the residuals (observed q_{95} minus group mean) within each group (SS_R). R^2 is a goodness-of-fit measure.

The regression tree approach performs best (Fig. 6). Out of the total sum of squared specific low flow discharges of 5246 $l^2\,s^{-2}\,km^{-4}$ the regressions tree explains 3244 $l^2\,s^{-2}\,km^{-4}$, i.e. the variance explained by the grouping, calculated by one-way ANOVA, is 62%. This means that the regression tree is an excellent classification method if one is interested in finding groups that are most distinct in terms of both catchment characteristics and low flow catchment response. We believe that the reason for the good performance is that the splitting algorithm simultaneously maximises group homogeneity in terms of catchment characteristics and low flows. The regression tree is flexible in that it can choose the locally most relevant catchment characteristics, as each group can be subdivided by different decision criteria. This means that there is no need to select global similarity measures. This is an advantage for low flow regionalisation where global similarity measures may not exist. Application of the regression tree is straightforward and it provides an objective and robust classification. The most relevant catchment characteristics are apparent in the structure of the fitted regression tree. In contrast to the weighted cluster analysis, the regression tree is suitable for non-linear relationships between low flows and catchment characteristics which is an additional advantage. Using regression trees prior to linear regressions is therefore an attractive approach of combining the merits of non-linear and linear models.

The weighted cluster analysis approach performs second best and explains 56% of the variance of q_{95}. The weighting of the catchment characteristics by the coefficients of a regression model transfers information on low flow discharges to the distance measures used in the cluster analysis, which seems to be a rather efficient approach. However, it should be noted that the weighted cluster analysis consists of 10 groups so one would expect a better goodness-of-fit than for the other methods. The seasonality regions and residual pattern approaches yield low R^2 values of 34 and 25%, respectively. It is clear that these two methods give little weight to finding regions that are most homogeneous in terms of low flows. It is also interesting that even though there are large differences in the goodness-of-fit between the groupings, they are all significant at the 95% level (Table 6). For comparison Fig. 7 presents the catchment groupings

Table 6
Variance explained by the groupings alone without using regressions

Classification method	Number of Groups	SS_G	SS_R	SS_T	R^2(%)	p-value
Residual pattern approach	5	1319.1	3927.1	5246.3	25	<0.001
Weighted cluster analysis	10	2911.8	2334.4	5246.2	56	<0.001
Regression tree	7	3244.4	2001.9	5246.3	62	<0.001
Seasonality regions	8	1787.0	3459.3	5246.2	34	<0.001

SS_G is the sum of squares of the mean group specific low flows q_{95}, SS_R is the sum of squares of the residuals of group mean minus observed q_{95} and SS_T is the total sum of squares of the observed q_{95}. Units of SS are $l^2\ s^{-2}\ km^{-4}$. R^2 is the coefficient of determination of the group mean and the p-values are the empirical significance levels of F-tests of the group means.

obtained by the four classification methods. Group numbers are as of Tables 1, 3, 4 and 5. There are some similarities between the classifications, which reflect the main topographical units of Austria.

5.2. Goodness-of-fit of regression models

In a second step we compared the goodness-of-fit of the regressions models for each of the groups identified by the various grouping methods. We also compared these goodness-of-fit values to the global regression model.

The global regression model uses four catchment characteristics as predictors. These are H_R (range of altitude), L_R (fraction of wasteland or rocks), G_F (fraction of Flysch) and P_W (average winter precipitation). The global model explains 62% of the variance in q_{95}. This is the same value as the best grouping method without regressions. It is interesting to compare this result to studies in the literature that used a similar number of catchments as in this paper (325 catchments) and examined specific discharges as in this paper, rather than discharges. Gustard et al. (1992) obtained $R^2=57\%$ between Q_{95} standardised by the mean flow and portion of hydrologically defined soil classes for 694 catchments in the UK. Schreiber and Demuth (1997) obtained $R^2=56\%$ between specific mean annual 10-day minimum

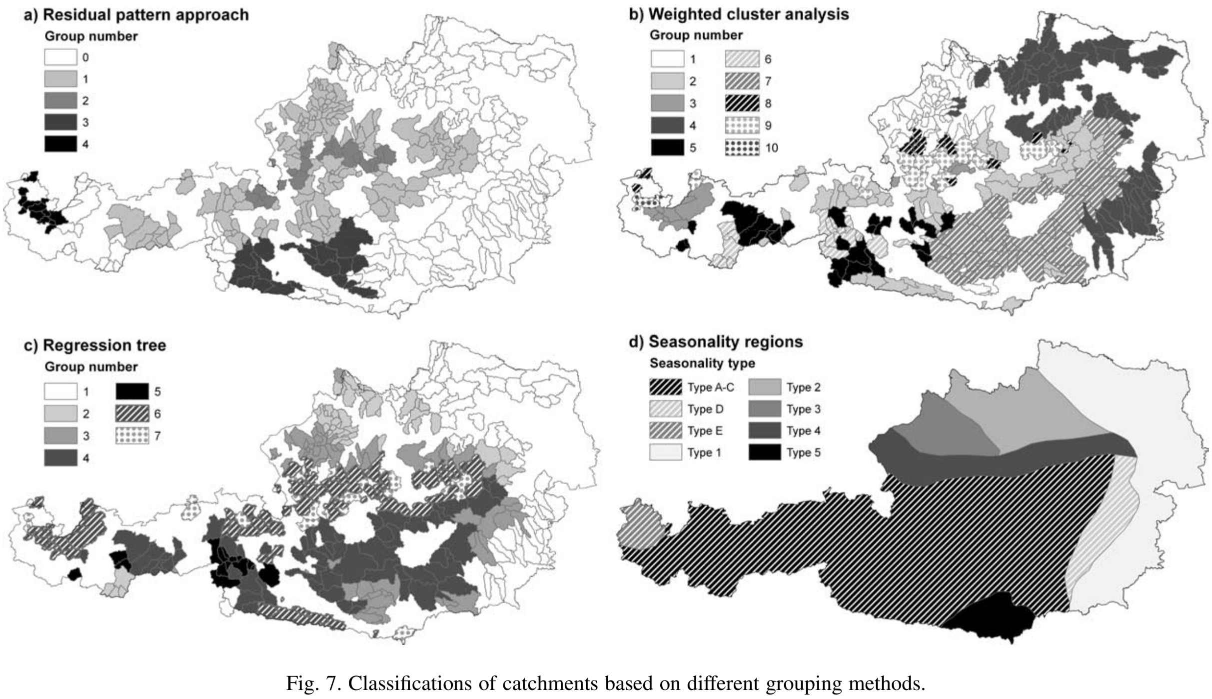

Fig. 7. Classifications of catchments based on different grouping methods.

discharge MAM(10) and a number of catchment characteristics for 169 catchments in south-west Germany, and Aschwanden and Kan (1999) obtained $R^2=51\%$ between specific discharge (q_{95}) and a number of catchment characteristics for 143 head-water catchments in Switzerland. The R^2 obtained in this study are hence somewhat larger than those from the literature. It is likely that the difference is related to the hydrological heterogeneity of Austria with clear regional differences in low flows. The better goodness-of-fit in this study may also be related to using sub-catchments rather than complete catchments which may make the catchment characteristics more relevant to low flow regionalisation.

The R^2 values of the component models vary vastly depending on the grouping method and the region (Tables 1, 3, 4, and 5). For the residual pattern approach, the R^2 values vary from 15 to 87%, for the weighted cluster analysis they vary from 0 to 75%, for the regression tree they vary from 13 to 70% and for the seasonality regions they vary from 51 to 89%. Overall the seasonality regions provide the best goodness-of-fit of the component regression models.

Aschwanden and Kan (1999) obtained R^2 values between 59 and 84% using the residual pattern approach and regional regressions of q_{95} in a very similar analysis to this paper. This R^2 range is a similar order of magnitude found for the residual patterns approach in this study. The low goodness-of-fit for one of the regions of 15% in this study (region 2, Table 1) may be related to karstic effects as this is a limestone area of the Pre-alps. It is possible that the specific discharges derived from the observations are inaccurate as the hydrologic catchment areas in these regions may differ from the topographic catchment areas but are not well known. Most other studies in the literature used discharge rather than specific discharge and so are not directly comparable to the results in this paper. As catchment size usually explains around 80–90% of the variability of low flow discharges (e.g. Dingman and Lawlor, 1995; Vogel and Kroll, 1992) it is clear that the R^2 values for discharges will be much larger than the R^2 values for specific discharges, particularly if there are significant variations in catchment size within the sample. Dingman and Lawlor (1995); Vogel and Kroll (1992), for example, reported R^2 values of more than 90%.

Table 7
Predictive performance of regional regression models based on different grouping methods

Catchment grouping	Allocation of ungauged site via	R^2_{cv} (%)
Residual pattern approach	Geographic location	63
Weighted cluster analysis	Classification tree	59
Regression tree	Classification tree	64
Seasonality regions	Geographic location	70
No grouping	–	57

R^2_{cv} is the coefficient of determination of cross-validated estimates.

5.3. Predictive performance of regional regressions for various grouping methods

The global regression model, i.e. without using any grouping, gives an $R^2_{cv}=57\%$ in the cross-validation mode (Table 7). This is a significantly lower value than the goodness-of-fit R^2 of the global model ($R^2=62\%$). Part of the difference may be related to an overfitting of the global regression model although this is unlikely to explain the full difference as only four catchment characteristics have been used as predictors. A more important reason for the difference may be heteroscedasticity of the sample and the existence of outliers, which contribute significantly to the estimation error. This issue is discussed later in this paper.

In the regional regression models, the grouping based on seasonality regions performs best (Table 7). The explained variance, in a cross-validation mode, is $R^2_{cv}=70\%$. This is significantly more than for the global model ($R^2_{cv}=57\%$). It appears that delineating regions based on the seasonality of low flows provides information on the hydrological regimes not captured by the catchment characteristics and the low flow discharges. Note that all four grouping methods use information on the low flow discharge q_{95}, albeit in different ways, and all grouping methods, with the exception of the seasonality regions approach, use catchment characteristics as well.

It is interesting that this performance is significantly better than that of an alternative model proposed by Laaha and Blöschl (2006) which gave $R^2_{cv}=58\%$ for the same data set. The model of Laaha and Blöschl (2006) is a global regression model that uses a region index as a predictor variable in addition to the catchment characteristics. This index value differs by the region and has been calibrated. It appears that the seasonality types are not mainly related to the magnitude of the low flows, so they are not very efficient as a predictor variable. However, the relationship between catchment characteristics and low flows appears to be significantly different for different seasonality regions. Various processes may combine in different ways in different seasonality regions, as a result of differences in the hydrologic and climatic regime. The seasonality grouping is hence very efficient in the context of the regional regression approach of using separate regressions in each of the groups. The favourable performance of the grouping method based on seasonality regions may be related to the striking differences in low flow seasonalities in the study domain (Fig. 6). These differences are clearly related to different processes. Winter low flows are a result of the retention of solid precipitation in the seasonal snow pack of the catchment and of freezing processes in the soils. In contrast, summer low flows are related to the relatively large moisture deficits in the lowland regions of Austria during summer. It appears that grouping the domain according to low flow seasonalities does capture some of the effects of these processes.

The regression tree grouping performs second best ($R^2_{\mathrm{cv}} = 64\%$) and the performance of the residual pattern approach is similar ($R^2_{\mathrm{cv}} = 63\%$). As compared to the global regression model ($R^2_{\mathrm{cv}} = 57\%$) there is some improvement in the performance although it is not large. The weighted cluster analysis, only yields a minor improvement ($R^2_{\mathrm{cv}} = 59\%$) over the global model. The improvement of the regional regression models (including grouping) over the global model (without grouping) is related to the degree of non-linearity that can be captured by the grouping method. In the weighted cluster analysis method, the performance is similar to the fully linear global model, so does poorly in representing any non-linearity. The other two methods do capture some of the non-linearity.

It is interesting that the relative performance of the grouping methods combined with regional regressions differs from the relative goodness-of-fit of the grouping methods alone. While for the grouping methods alone the regression tree approach performed best, it is the grouping based on seasonality regions that performs best when combining the grouping with regional regressions. It is clear that in the latter case, the important feature the catchment groupings need to capture is the way the catchment characteristics are related to low flows rather than the low flows themselves. Within group homogeneity and between group heterogeneity in terms of low flow discharges are hence not a good indicator for the predictive performance of low flow regional regressions. Cross-validation of the regression estimates is certainly a preferable way of measuring the performance of regionalisation methods.

It should be noted that in the residual pattern and the seasonality region approaches the regions were not updated in the cross-validation procedure. This was because the regions were deemed sufficiently contiguous not to change much if a single catchment is added. It is possible that the cross-validation performance of these two methods may very slightly decrease if the regions were updated but given the relative magnitude of the cross-validated coefficients of determinations it is unlikely that this will change the ranking of the predictive performance of the methods.

5.4. Heteroscedasticity, outliers and bias

As a final step of assessing the methods of catchment grouping we examined scatter plots of predicted vs. observed specific low flow discharges q_{95} (Fig. 8). The scatter plots allow a detailed examination of the performance of individual catchments including the existence of outliers and a potential heteroscedasticity of the observations and the predictions. Overall the relative scatter of the methods (Fig. 8) corresponds well with the cross-validated coefficients of determination in Table 7 and it is clear that the seasonality regions approach performs best and the weighted cluster analysis approach performs poorest. The weighted cluster analysis approach overestimates low flows significantly for three catchments and the magnitude of the estimation error is relatively large for a number of catchments. The outliers tend to increase with q_{95}, which suggests that the predictions are heteroscedastic. One would usually apply a variance-stabilising transformation in this case, such as taking the logarithms of q_{95}. However, since preliminary analyses indicated little effect on the model

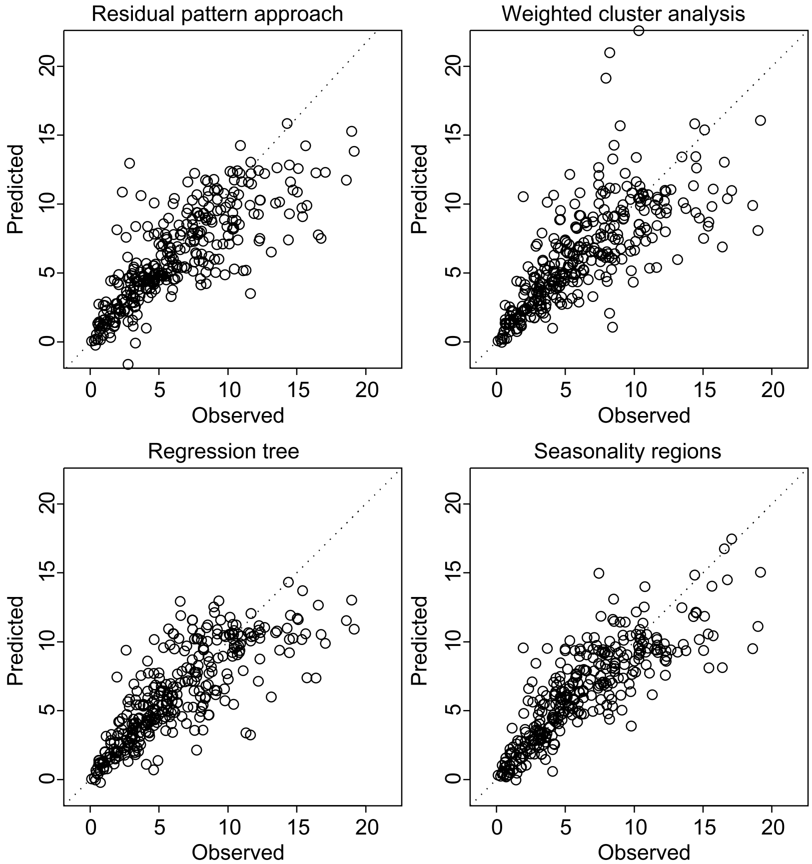

Fig. 8. Scatter plots of predicted vs. observed specific low flow discharges q_{95} ($\mathrm{l\,s^{-1}\,km^{-2}}$) in the cross-validation mode. Each panel corresponds to one regional regression model and each point corresponds to one catchment.

parameters, the level of heteroscedasticity was considered acceptable in the context of this paper as the main focus was on evaluating the potential of catchment grouping on low flow regionalisation. The residual pattern approach generally performs quite well although it gives negative predictions of q_{95} for two catchments and a few outliers. The regression tree approach performs equally well for the bulk of the catchments, but appears slightly superior to the residual pattern approach as far as outliers are concerned. The approach based on seasonality regions performs best. The points are scattered around the 1:1 line indicating low prediction errors for a broad range of discharges. The scatter is almost homoscedastic and there are only a few minor outliers.

One apparent deficiency of all models is the large scatter and clear bias for very wet catchments. In catchments where observed specific low flow discharges are more than about $12\,\mathrm{l\,s^{-1}\,km^{-2}}$ the low flows are consistently underestimated, and the random prediction error is also rather large. It appears that none of the models can cope very well with these large discharges. Part of the errors may be related to biases in the observed values. A specific discharge of $12\,\mathrm{l\,s^{-1}\,km^{-2}}$ corresponds to 378 mm of low flow depth per year which is a relatively large value for Austrian conditions. In all catchments in the $q_{95} > 12\,\mathrm{l\,s^{-1}\,km^{-2}}$ range, with the exception of two catchments, limestone is the main geologic formation (75% of the catchment area on average) so karst effects are likely to occur. It is possible that the specific discharges derived from the observations are inaccurate as the hydrologic catchment areas in these regions may differ from the topographic catchment areas. A more detailed analysis is needed to ascertain the extent to which the low flow observations in these catchments are actually biased. It should also be noted that it is not uncommon for regionalisation models to have a tendency for underestimating large values. For example, the flood regionalisation analysis of Merz and Blöschl (2005) showed that flood quantiles in the same study area were consistently underestimated by their method for catchments with above-average specific flood discharges.

6. Conclusion

We compared four catchment grouping methods in terms of their performance in predicting specific low flow discharges q_{95}. These methods are the residual pattern approach, weighted cluster analysis, regression trees and an approach based on seasonality regions. The grouping based on seasonality regions performs best and explains 70% of the variance in a cross-validation mode. The favourable performance of this grouping method is likely related to the striking differences in seasonal low flow processes in the study domain. Winter low flows are a result of the retention of solid precipitation in the seasonal snow pack of the catchments and of freezing processes in the soils while summer low flows are related to the relatively large moisture deficits in the lowland regions of Austria during summer. The regression tree grouping performs second best (explained variance of 64%) and the performance of the residual pattern approach is similar (explained variance of 63%). The weighted cluster analysis only explains 59% of the spatial variance of q_{95}, which is only a minor improvement over the global regression model, i.e. without using any grouping, in a cross-validation mode (explained variance of 57%).

We further examined the part of the variance (R^2) of specific low flows q_{95} that can be explained by the grouping alone without using regressions. In this comparison, the regression tree approach performs best and explains 62% of the spatial variance. This means that the regression tree is an excellent classification method if one is interested in finding groups that are most distinct in terms of both catchment characteristics and low flow catchment response. The weighted cluster analysis approach performs second best (explained variance of 56%). The seasonality regions and residual pattern approaches yield low R^2 values of 34 and 25%, respectively. It is clear that these two methods give little weight to finding regions that are most homogeneous in terms of low flows.

An analysis of the sample characteristics of all methods suggests that, again, the grouping method based on the seasonality regions has the most favourable characteristics although all methods tend to underestimate specific low flow discharges in the very wet catchments. The favourable performance of the seasonality regions approach was further reflected in an analysis of the goodness-of-fit of the regressions between catchment characteristics and q_{95} for each of the groups identified by the various grouping methods. Here, the seasonality regions approach explained between 51 and 89% of the spatial variance of q_{95}, depending on the region. A global regression model that uses range of altitude, fraction of rock, fraction of Flysch, and average winter precipitation as the predictor variables explains 57% of the variance in q_{95}.

This study has examined a single low flow characteristic (q_{95}) and it would be interesting to see whether the relative performance of the grouping methods remains the same if different characteristics are examined. There is also some potential in using short discharge series in the low flow regionalisation as short series and, perhaps, snapshot discharge measurements may be available in a much larger number of catchments (Laaha and Blöschl, 2005). We are currently assessing techniques that combine both sources of information, estimates from regionalisation and estimates from short records, and the results will be reported in the near future.

Acknowledgements

Analyses are based on data provided by the Geological Survey of Austria, by the Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management, and by the Austrian Hydrographic Service. We are very grateful to H. Strelec and two anonymous reviewers for many valuable comments on the manuscript.

References

Andrews, D.F., 1972. Plots of high dimensional data. Biometrics 28, 125–136.

Aschwanden, H., Kan, C., 1999. Le débit d'étiage Q347—Etat de la question. Communications hydrologiques, vol. 27. Service hydrol. et géol. national, Berne.

Behr, O., 1989. Digitales Modell des Oberflächenentwässerungssystems von Österreich. Institut für Hydraulik Gewässerkunde und Wasserwirtschaft, Technische Universität Wien (Forschungsbericht 11, Wien).

Breiman, L., Friedman, J.H., Olshen, R., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth International Group, Belmont, California.

Clark, L.A., Pregibon, D., 1991. Tree-based models. In: Chambers, J.M., Hastie, T.J. (Eds.), Statistical Models. Chapman & Hall/CRC, New York.

Clausen, B., Pearson, C.P., 1995. Regional frequency analysis of annual maximum streamflow drought. J. Hydrol. 173, 111–130.

Demuth, S., Young, A.E., 2004. Regionalisation procedures. In: Tallaksen, L.M., van Lanen, H.A.J. (Eds.), Hydrological Drought: Processes and Estimation Methods for Streamflow and Groundwater Developments in Water Science, vol. 48. Elsevier.

Dingman, S.L., Lawlor, S.C., 1995. Estimating low flow quantiles from drainage-basin characteristics in New Hampshire and Vermont. Water Resour. Bull. 31 (2), 243–256.

Efron, B., Tibshirani, R.J., 1993. An introduction to the bootstrap. Monographs on Statistics and Applied Probability, vol. 57. Chapman & Hall, New York.

Elder, K., 1995. Snow distribution in alpine watersheds. PhD dissertation, University of California, Santa Barbara, 309 pp.

Gustard, A., Irving, K.M., 1994. Classification of the low flow response of European soils. FRIEND: Flow Regimes from International Experimental and Network Data. IAHS Publication No. 221, pp. 113–117.

Gustard, A., Bullock, A., Dixon, J.M., 1992. Low flow estimation in the United Kingdom. Institute of Hydrology, Report No. 108, 88 pp., append.

Gustard, A., Young, A.E., Rees, G., Holmes, M., 2004. Low flows, drought and operational hydrology. In: Tallaksen, L.M., van Lanen, H.A.J. (Eds.), Hydrological Drought: Processes and Estimation Methods for Streamflow and Groundwater Developments in Water Science, vol. 48. Elsevier, NL.

Haines, A.T., Finlayson, B.L., McMahon, T.A., 1988. A global classification of river regimes. Appl. Geog. 8, 255–272.

Hartigan, J.A., 1975. Clustering Algorithms. Wiley, New York.

Hayes, D.C. 1992. Low flow characteristics of streams in Virginia. US Geological Survey, Water Supply Paper 2374.

Kaufman, L., Rousseeuw, P.J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York.

Kresser, W., 1965. Österreichs Wasserbilanz. Österr. Wasserwirtschaft 17, 213–221.

Kresser, W., Kirnbauer, R., Nobilis, F., 1985. Überlegungen zur Ermittlung von Niederwasserkenngrößen Mitteilungsblatt des Hydrographischen Dienstes, vol. 54. Bundesministerium für Land- und Forstwirtschaft, Wien.

Laaha, G., 2002. Modelling summer and winter droughts as a basis for estimating river low flows. In: FRIEND 2002 — Regional Hydrology: Bridging the Gap between Research and Practice. Proceedings of the Fourth International Conference held in Cape Town, South Africa, March 2002. IAHS Publ. 274, pp. 289–295.

Laaha, G., Blöschl, G., 2003. Saisonalität von Niederwasserspenden in Österreich. Mitteilungsblatt des Hydrographischen Dienstes in Österreich, vol. 82. Bundesministerium für Land- und Forstwirtschaft, Wien.

Laaha, G., Blöschl, G., 2005. Low flow estimates from short stream flow records—a comparison of methods. J. Hydrol. 306, 264–286.

Laaha, G., Blöschl, G., 2006. The value of seasonality indices for regionalizing low flows. J. Hydrol. Process in press.

Merz, R., Blöschl, G., 2005. Flood frequency regionalisation—spatial proximity vs. catchment attributes. J. Hydrol. 302, 283–306.

Merz, R., Piock-Ellena, U., Blöschl, G., Gutknecht, D., 1999. Seasonality of flood processes in Austria. IAHS Publ. 255, 273–278.

Nathan, R.J., McMahon, T.A., 1990. Identification of homogeneous regions for the purpose of regionalization. J. Hydrol. 121, 217–238.

Piock-Ellena, U., Pfaundler, M., Blöschl, G., Burlando, P, Merz, R. (2000) Saisonalitätsanalyse als Basis für die Regionalisierung von Hochwässern. Wasser, Energie, Luft, 92. Jahrgang, Heft 1/2, 13-21, Baden.

Rosenthal, W., Dozier, J., 1996. Automated mapping of montane snow cover at subpixel resolution from the Landsat Thematic Mapper. Water Resour. Res. 32, 115–130.

Schreiber, P., Demuth, S., 1997. Regionalisation of low flows in southwest Germany. Hydrol. Sci. J. 42 (6), 845–858.

Skop, E., Loaiciga, H.A., 1998. Investigating catchment hydrology and low-flow characteristics using GIS. Nordic Hydrol. 29 (2), 105–128.

Smakhtin, V.U., 2001. Low flow hydrology: a review. J. Hydrol. 240, 147–186.

Veanbles, W.N., Ripley, B.D., 1999. Modern Applied Statistics with S-Plus, third ed. Springer, New York.

Vogel, R.M., Kroll, C.N., 1992. Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics. Water Resour. Res. 28 (9), 2451–2458.

Weisberg, S., 1985. Applied Linear Regression, second ed. Wiley, New York.

Young, A.R., Round, C.E., Gustard, A., 2000. Gustard Spatial and temporal variations in the occurrence of low flow events in the UK. Hydrol. Earth Syst. Sci. 4 (I), 35–45.