The influence of methodological procedures on hydrological classification performance

F. J. Peñas[1], J. Barquín[1], T.H. Snelder[2], D.J. Booker[3] and C. Álvarez[1]

[1]Environmental Hydraulics Institute. "IH Cantabria". Universidad de Cantabria. Parque Científico y Tecnológico de Cantabria. C/Isabel Torres, 15, 39011, Santander, Spain.

[2]Aqualinc Research Ltd, PO Box 20–462, Bishopdale, Christchurch, 8543, New Zealand.

[3]National Institute of Water and Atmospheric Research, PoBox 8602, Christchurch, New Zealand

Correspondence to author: F.J. Peñas (penasfj@unican.es)

Keywords: Natural Flow Regime; Hydrological prediction; Flow series normalization; Classify First; Predict First; Water resource management

Abstract

Hydrological classification has emerged as a suitable procedure to disentangle the inherent hydrological complexity of river networks. This practice has contributed to determine key biophysical relations in fluvial ecosystems and the effects of flow modification. Thus, a plethora of classification approaches, which agreed in general concepts and methods but differed largely in specific procedures, have emerged in the last decades. However, few studies have compared the implication of applying contrasting approaches and specifications over the same hydrological data. In this work, using cluster analysis and modelling approaches, we classify the entire river network covering the northern third of the Iberian Peninsula. Specifically, we developed classifications of increasing level of detail, ranging from 2 to 20-Class levels, either based on raw and normalized daily flow series and using two contrasting approaches to determine class membership: Classify-Then-Predict (ClasF) and Predict-Then-Classify (PredF). Classifications were compared in terms of their statistical strength, the hydrological interpretation, the ability to reduce the bias associated to underrepresented parts of the hydrological space and their spatial correspondence. The results highlighted that both the data processing and the classification strategy largely influenced the classification outcomes and properties, although differences among procedures were not always statistically significant. The normalization of flow data removed the influence of flow magnitude and generated more complex classifications in which a wider range of hydrologic characteristics were considered. The application of the PredF strategy produced, in most of the cases, classifications with higher discrimination ability and presented greater ability to deal with the presence of distinctive gauges in the data set than using the ClasF strategy.

1 Introduction

Understanding the natural variability of hydrology at the regional scale has become crucial for river ecology and management because of three main reasons: (i) it is a primary factor influencing river geomorphology (Peñas et al., 2012;Richter et al., 1998;Benda et al., 2004), water (Álvarez-Cabria et al., 2010;Chinnayakanahalli et al., 2011) and biological characteristics (Poff and Zimmerman, 2010), (ii) its variability reflects climate (Morán-Tejeda et al., 2011) and catchment attributes (second order driver; Monk et al., 2007) and (iii) freshwater resources are essential to maintain many human activities (Naiman and Dudgeon, 2011).

Much progress has been made over the last 20 years in understanding hydrologic variability and how it promotes self sustaining ecosystems (Poff et al., 2006;Gurnell et al., 2000). However, the inherently complexity of flow regimes hinders both the quantification of direct responses of hydrology to catchment characteristics, and the identification of key hydrology and ecology relationships. The identification and characterization of relevant ecological aspects of the flow regime and the arrangement of similar rivers into a geographical context (Poff, 1996), through the definition of hydrological classifications, has emerged as a relevant procedure to structure analyses in hydroecological studies. Specifically, inductive hydrological classification approaches have been used to group river reaches into classes within similar attributes regarding the flow regime (Snelder et al., 2009) and ecological attributes (McManamay et al., 2012).

Many of the existing hydrological classifications following the inductive approach rely on the use of statistical procedures to minimize the redundancy of the hydrological information (Olden and Poff, 2003) and also, to reduce the intra-group and increase the inter-groups variability (Snelder and Booker, 2013). Nevertheless, many specific steps within the classification process may be influenced by a series of subjective decisions depending on the rationale, objectives and available data. For example, many hydrological classifications are based on normalized flow data (McManamay et al., 2012;Kennard et al., 2010;Reidy Liermann et al., 2012) while others used raw flow series (Zhang et al., 2012;Belmar et al., 2011;Alcázar and Palau, 2010). However, normalization can be viewed as a completely subjective choice that depends on the purpose of the classification (Olden et al., 2012). If the range of flow magnitude varies largely within a region, classification based on the raw flow series would be subjected uniquely to this attribute. In contrast, other flow attributes that present a lower degree of variability and that are not affect by the normalization of the series, would be masked in classifications. The main reason for normalization is to remove the scale dependence of flow magnitude indices to promote the classification of rivers according to a larger set of hydrological attributes. Therefore, the larger the number of hydrological aspects taken into account in the classification the larger its potential uses. For example, the normalization of flow series allows segregating rivers attending the intra-annual variability of flows magnitude, i.e. the shape of the hydrographs. Undoubtedly the shape of the hydrograph influences river reach ecology (Bunn and Arthington, 2002;Richter et al., 1998) and are key elements for understanding the relationship between climatic and streamflow patterns (Gámiz-Fortis et al., 2011). Nonetheless the size of a river reach and the absolute magnitude of flows also play a key role in ecological processes (Bunn and Arthington, 2002;Vannote et al., 1980) and it is a critical element to manage water resources.

In addition, the scientific and management utility of hydrologic classifications relies on the capacity to extrapolate the class membership to ungauged sites, providing a map of natural flow regimes at the regional scale (Snelder et al., 2009;Reidy Liermann et al., 2012). The Classify-then-Predict (ClasF) strategy has been the most common approach to fulfil this objective (e.g. Kennard et al., 2010;Reidy Liermann et al., 2012). ClasF predicts class membership to ungauged sites based on environmental data (climate, topography, geology or land-use). However, this method might pose some flaws when predicting onto an entire region, especially if the distribution of gauges is biased, i.e. specific kinds of rivers are under or overrepresented (Snelder and Booker, 2013). If this is the case, the cluster step would fail in accounting for those hydrological features underrepresented in the data set. This is a critical issue since the low representation in the gauged network does not imply a low representation in the entire river network. The way in which these underrepresented data or distinctive gauges (i.e. those ones presenting a large hydrologic dissimilarity to the other ones present in the data set) are classified may lead to the loss of their "rare" hydrologic character when classes are predicted to the whole river network. Due to this reason, some researchers have attempted other approaches such as the Predict-then-Classify (PredF) strategy (Ferrier and Guisan, 2006; Snelder and Booker, 2013). Using this approach, hydrological indices obtained from the flow series are predicted onto the entire river network based on climate and catchment characteristics. Then, classification of all river segments is performed as a final stage within the procedure.

The aim of this study was to investigate how the normalization of flow series data previous to the classification procedure and the use of ClasF and PredF influences (i) the classification performance, (ii) the hydrological interpretation of the classifications, (iii) their ability to reduce the bias associated to the underrepresented parts of the hydrological space and (iv) the degree of spatial correspondence between classifications. To achieve this aim we will develop hydrological classifications of natural conditions over an entire river network in the northern third of the Iberian Peninsula, covering catchments of contrasting climate and spatial configuration. We hypothesised that normalization of river flow data will tend to classify rivers according to their annual regime and not only to the size of the river and also increase the contribution of other hydrological variables not related to flow magnitude. In addition, we hypothesised that the application of the PredF classification procedure will reduce within class heterogeneity.

2 Methods

2.1 Study Area

The study area comprises the northern third of the Iberian Peninsula (Fig. 1) covering a total area greater than 124000 km^2. It represents heterogeneous environmental conditions and can be broadly segregated in three main zones. On one hand, the area draining into the Cantabric sea encompass several small basins with drainage areas ranging from 30 km^2 to 4907 km^2 covering a total area of 22000 km^2. Rivers are confined by the Cantabrian Cordillera, which reaches up to 2600 m.a.s.l. and runs parallel to the coast. Thus, they are characterized by high slopes and short main stream lengths. This region has a humid oceanic temperate climate (Rivas-Martínez et al., 2004). Precipitation is abundant throughout the year with mean of 1300 mm year^{-1}, with maximum rainfall~~s~~ in December (150 mm month^{-1}) and minimum in July (50 mm month^{-1}). However, the precipitation magnitude and distribution varies significantly according to local topography. Snow precipitation is frequent in winter above 1000 m.a.s.l. More than 50% of the surface is occupied by deciduous forest, scrubs and grasslands, while 10% is occupied by agriculture. The population in this area amounts to almost 3500000 inhabitants with a population density of 175 hab km^{-2} although it varies between regions. On the other hand, the Mediterranean area is mainly occupied by the Ebro basin along with a set of medium size basins in the eastern zone. The Ebro basin covers a total extension of 85530 km^2. It is enclosed by the Cantabrian Mountains and the Pyrenees (3400 m.a.s.l.) in the North, by the Catalan Coastal Chain (1712 m.a.s.l.) in the East and from the North-West to the South-East by the Iberian massif (2300 m.a.s.l.), which creates a dense river network in the catchment boundaries and an extended flat surface in the interior. The Ebro Basin receives both temperate and Mediterranean climate influences. The Pyrenean area (northwest) and the northern part of the Iberian massif present oceanic temperate climate that change gradually to a typical Mediterranean climate in the central Ebro depression. Annual precipitation is 656 mm, however it varies from 300 mm in the centre to the 1700 mm in the highest mountains (Bejarano et al., 2010) where snow is also common during the winter months. The precipitation regime in the Mediterranean region has its maxima in autumn and spring and minima in winter and summer. The temperature regime also oscillates through the year with temperatures over 30 ºC in summer and below 5 ºC during winter. Population density is below 35 hab km^{-2} which could be considered low, however more than 40% of the surface is occupied by agricultural land and, thus, the catchment is subjected to an intensive water resource control by more than 216 large dams and other water engineering systems. The eastern zone of the study area comprises several medium catchments ranging from 72 km^2 to 5000 km^2, occupying a total extension of 16500 km^2 that drain directly from the Pyrenees or the Catalan costal chain to the sea. This area is dominated by the Mediterranean oceanic climate in the coast and by a temperate climate in the mountains. Precipitation declines from an annual mean of 1200 mm $year^{-1}$ in the northern river heads to less than 500 mm $year^{-1}$ in the Southern catchments. Coniferous and broadleaf forest, scrubs and grasslands occupies more than 60% of the surface in the northern catchments which are progressively replaced by agriculture lands in the south. There are a total of 6600000 inhabitants in this area, mostly concentrated in the city of Barcelona and its metropolitan area. Therefore, most of the water resources are allocated to urban and industrial uses.

2.2 Hydrologic Data

The initial data set consisted in series of mean daily flow recorded at 428 gauging stations operated by different Spanish water agencies and regional governments. Only gauges unaffected by impoundments (defined as large engineering structures) or large upstream abstractions were selected for analyses. In addition, we selected those gauges with available data for the period 1976-2010 and analysed the quality of the series. First, an analysis of the flow series was carried out to eliminate those years without desirable data quality, which could be due to the presence of (i) periods of consecutive repeated values, (ii) non-natural extreme low flows for short time periods, (iii) periods of zero flow values in non- intermittent rivers, (iv) non-natural flow magnitude rises and falls or (v) large differences between two periods, probably due to changes of flow recorder method. Years with more than 30 days of missing data were removed from the analysis. In the last step, we discarded the gauges that accounted with less than 8 years. After applying these restrictions, 156 gauges were selected with an average length of 17 years of data (Table 1).

In this study we developed four sorts of classifications (Fig. 2). Two of them were obtained from normalized flow series and the other two from non-normalized (raw) series. Normalization is used to eliminate the influence of flow magnitude (Snelder et al., 2009). Flow series were normalized by dividing all daily flow values by the mean annual flow (Poff et al., 2006)

A set of 103 and 101 hydrologic indices, which represent a wide range of ecologically meaningful aspects of the flow regime (Olden and Poff, 2003), were calculated for the raw and normalized flow series, respectively (Appendix A). These indices characterize the central tendency and dispersion of: (i) magnitude of annual and monthly flows conditions, (ii) magnitude of severe high and low flow conditions, (iii) timing of flows, (iv) frequency and duration of high flow pulses and (v) rate of change of flow (Richter et al., 1996;Olden and Poff, 2003). It must be pointed out that among the indices representing flow magnitude, l1 and lcv, were excluded from the set of indices extracted from the normalized flow series. After dividing each daily flow data by the mean annual flow, l1 became equal to 1 in all the gauges. In addition, lcv became equal to lca (as lcv = lca/l1).

Given the strong correlation between several indices, the initial set of indices was reduced to a set of non-correlated synthetic indices using the procedure outlined in Olden and Poff (2003) and followed by many others (Chinnayakanahalli et al., 2011;Zhang et al., 2012;Belmar et al., 2011). According to Olden and Poff (2003), a principal components analysis (PCA) was used to determine the patterns of correlation between the hydrological indices. It allow identifying the subsets of synthetic indices, that describe the major sources of variation while minimize redundancy. The broken stick method (Jackson, 1993) was applied to obtain and define the optimal set of PCs to be retained. Each of the selected PC was used as a hydrologic synthetic index in subsequent analysis. Two PCAs were carried out independently, one for the hydrologic indices calculated from the raw flow series and another for the hydrologic indices calculated from the normalized flow series. Each PC was standardized before conducting further analysis to give them equal weights. Snelder and Booker (2013) demonstrated that this additional step increased classification performance.

2.3 Environmental Data

A Synthetic River Network (SRN) was delineated using a 25-m digital elevation model (DEM) using the NestStream software (Miller, 2003). The SRN comprises 667406 segments with lengths ranging from 16 to 800 m and was used as a spatial network to integrate the hydrological and environmental information.

Climate, topography, land cover and geology are hypothesised to be important discriminators of the hydrologic regime regardless of geographic location. Thus, environmental variables were used to explain the hydrological character of the recorded flow series and predict this character onto the whole river network. Predictor variables describing several environmental attributes including climate, topography, land cover and geology were extracted from existing databases provided by several national and regional organizations. The variables for each segment represented the mean value of the variables in the upstream catchment. An initial set of 25 environmental variables with potential influence on the hydrological regimes were selected. Pearson's correlation coefficient between each pair of variables was calculated and variables with correlation higher than 0.7 were discarded. A final set of 16 variables were selected (Table 2):

i) Climate (n=3): Precipitation, precipitation range and evapotranspiration were derived from monthly climate variables calculated in a 1 km x 1 km grid map. This map was obtained by means of an interpolation procedure based on data recorded in more than 5000 weather stations of the Spanish network. These data were originally developed to be implemented into the Integrated System for Rainfall-Runoff modelling (in Spanish SIMPA model) by the Centre for Hydrographic Studies (CEDEX, Ministry of Public works and Ministry of Agriculture and Environment, Spain).

ii) Topography (n=5): Catchment area, slope, elevation, confluence density and drainage density were derived from the 25 m DEM.

iii) Land cover (n=6): The percentage surface occupied by broadleaf forest, coniferous forest, pasture, agricultural land, denuded areas and urban areas was derived from the Soil Occupancy Information System (in Spanish SIOSE) developed by the National Geographic

Institute of the Spanish Government. SIOSE presents a scale of 1:25000 and integrates satellite and aerial images from several sources of information.

iv) Geology (n=2): The average rock hardness and the terrain permeability were derived from the litostatigraphic and permeability map at scale 1:200,000 developed by the Spanish Geologic and Miner Institute of the Spanish Government. The base of the calculation of these variables was the percentage of area occupied by the original classes of rocks included in the data layer. These classes were then reclassified into broader ones and then, we assigned them a numerical value based on geological hardness and soil permeability (see Snelder et al, 2008 for details).

2.4 Classification procedures

In this study, we derived classifications with increasing numbers of levels using the synthetic hydrologic indices extracted from the raw or the normalized flow series and using two contrasting strategies (sensu Snelder and Booker, 2013): (i) the classify-then-predict (rawClasF and norClasF) and the (ii) predict-then-classify (rawPredF and norPredF). The prefix raw and nor indicated whether classification was based on the hydrological indices extracted from the raw or normalized flow series respectively.

Given the high number of gauges removed due to the presence of impoundments or abstraction upstream, it is probable that selected gauges do not represent the whole spectrum of natural hydrologic conditions in the study area. In addition, the SRN developed for this study presented many rivers of first and second order which are underrepresented in the gauge data base. The prediction of the class membership (ClasF) or the hydrological synthetic indices (PredF) beyond the hydrological space represented in the selected gauges could lead to misleading results. Therefore, the prediction stage of the ClasF and PredF approaches was not based on the whole SRN (667406 segments) but in a reduced SRN. All the segments of the SRN that presented values of the predictor variables out of the range (maximum/minimum) defined by these predictors in the selected gauges were discarded. The reduced SRN kept 178297 segments.

2.4.1 Classify-Then-Predict classification (ClasF)

Partitioning Around Medoids (PAM; Kauffman and Rousseeuw, 1990) algorithm based on the synthetic indices was used to cluster gauges (Fig. 2). This technique allowed to specify the number of clusters. We produced classifications with numbers of classes ranging from 2 to 20. We then used Random Forest (RF; Breiman, 2001) to develop predictive models that relate class memberships and the environmental variables (Fig. 2). We fitted one specific RF for each classification level (2 to 20-Class level) and then, these models were used to establish the most probable class of each segments of the SRN for each classification, i.e. 19 sets of predictions.

2.4.2 Predict-Then-Classify classification (PredF)

For the PredF strategy, empirical models were first fitted to each of the standardized synthetic indices as a function of environmental variables using RFs (Fig. 2). Then predictions of the synthetic indices are made for each segment of the SRN. Finally, classifications were produced by clustering all the modelled segments using the PAM algorithm varying again between 2 and 20-Class levels.

As stated before, ClasF and PredF strategies are based in the use of RF (Breiman, 2001). RF fits many classification and regression trees (CART; Breiman et al., 1984), each of them grown with a randomized subset of sites and predictor variables from the initial data. Each CART is then used to predict the sites initially excluded from the data set, named the out-of-bag (OOB) samples. These predictions are used to calculate the predictive accuracy of the model and the importance of each predictor variable (Snelder et al., 2011).

2.5 Comparison of classification performance

The performance of the classifications was measured using the classification strength (CS; Van Sickle, 1997) and ANOVA.

CS estimate the degree of dissimilarity between gauges explained by the classifications (Snelder and Booker, 2013). This analysis was performed on the hydrological indices with the highest loading on each of the retained PCs. Briefly, CS resulted from the difference between the mean dissimilarity of the gauges in the same class (D_{within}) and the mean dissimilarity of gauges in the other classes ($D_{between}$). Higher values of CS indicated a greater uniformity within classes and greater differences between classes (Van Sickle, 1997). We calculated CS for each classification (rawClasF, rawPredF, norClasF and norPredF each with 2-20-Class levels). We applied the restriction that classes comprising a minimum of five gauges to reduce the influence in the analysis of classes represented by a very low number of gauges.

In addition, we performed an ANOVA on all the hydrological indices (103 and 101 for raw and normalized series, respectively) with the class membership as the explanatory variable. ANOVA allowed analysing the potential of classifications to discriminate each of the hydrological indices. The coefficient of determination (r^2) was calculated for each level (2 to 20-Class level) of the 4 classifications. The restriction of the five gauges per class was also applied.

Following the procedure outlined in Snelder and Booker (2013) and Snelder et al. (2012), both the CS and ANOVA analysis were performed on gauges not used in the fitted models by means of a five-fold cross validation procedure (Hastie et al., 2001). This allowed us focusing on the “predictive performance” of the classifications. Each cross validation procedure was repeated 5 times in order to “smooth out” the variability inherent to each subset. Therefore, results of 25 estimates of predictive CS and r^2 statistics for each hierarchical level of classifications were obtained. Based on the “one standard error rule”, two classifications were assumed significantly different if standard errors of the statistics did not intersect.

2.6 Hydrological interpretation of classifications

We selected the five hydrological indices included in the initial set (103 and 101 indices for the raw and normalized series, respectively) with the highest values in each retained PCs to interpret the hydrological meaning of the new synthetic indices. The retained PCs accounted with the greatest part of the hydrological variability so, they are the major determinants of the classification patterns. In addition, we used the ANOVA results to interpret each classification by looking at the different coefficients of determination for specific indices. We assumed that the higher the coefficient of determination the higher the importance of that index to discriminate among classes.

2.7 Analysis of distinctive gauges

We also analysed how each classification strategy resolved the problem associated with the presence of distinctive gauges (DGs). DG can be defined as those that showed the most distinctive regimes (i.e. gauges presenting the largest hydrologic dissimilarity relative to the other ones present in the data set). The way the classification procedure deals with the DGs is very important. For instance, DGs can be grouped to other ones that are completely dissimilar or in very exclusive classes with lower dissimilarity between gauges but a very restricted number. In both cases, the hydrologic character represented by the DGs may be underrepresented when classes are predicted to the whole river network. We quantified how the different strategies deal with the presence of DGs in the data set.

Independent analyses were made for classifications based on raw and normalized flow series. Firstly, we calculated, based on the synthetic indices scores, the dissimilarity between each pair of gauges and then, the corresponding mean dissimilarity for each gauge. This value allowed selecting the gauges with the most distinctive hydrological regime, i.e. the DGs. We ordered the gauges from the most to the less dissimilar gauge and analysed how the dissimilarity values decayed. We selected 4 DGs for each type of series (raw or normalized), corresponding to the first important inflexion point in the decay trend of the dissimilarity. It is important to stress than dissimilarity values decreased from DG1 to DG4. Finally, we recorded the classes where the DGs belong after classifying the SRN.

For each DG two analyses were performed. Firstly, we calculated the distance between the DG and the medoid of the classes. This value was weighted by the mean distance between the medoid and all the other gauges belonging to the class. This distance indicated how much different is the DG relative to the other gauges included in the class. Secondly, we analysed the proportion of the classification domain assigned to the classes where the distinctive gauges were included. Low frequency of a class in the observed space (i.e. in the gauge network) does not implied low frequency in the SRN. Therefore, we expected higher frequencies of the class in the SRN than those observed in the gauge network. Low frequency of these classes indicated the inability of the procedure to predict properly the hydrological characteristics represented by the DGs.

2.8 Correspondence between classifications

The spatial agreement between each pair of classifications was evaluated by means of the Adjusted Rand Index (ARI; Hubert and Arabie, 1985). ARI analyses the relationship of each pair of gauges and how they differ between two cluster solutions. It ranges between 0 (indicating that agreement between two clustering solutions is not better than chance) and 1 (indicating perfect agreement). Given the large number of segments in the SRN, we randomly selected a subset of 1000 segments and computed ARI for all pairs of the four classifications. This process was repeated 10 times to avoid the effect of the variability in the selected data set.

Bespoke functions written in R were used to analyse flow series and calculate hydrological indices (Snelder and Booker 2013).

3 Results

3.1 PCA and Predictive mapping

The broken stick method selected the first five PCs of the PCA performed on the raw series. They explained 91% of the variance, accounting the PC1 alone for the 68 % (Table 3). The OOB misclassification rate of the RF models in the rawClassF ranged from 0.13 for the 2 classes level to 0.77 for the 20-Classes level (Fig. 3). The most important predictor variables of the RF were catchment area, precipitation, agriculture, pasture and elevation. For the rawPredF classification, the mean OOB r^2 for the RF models of the 5 synthetic indices was 0.4 decreasing from 0.65 for PC1 to 0.18 for the PC5. Predictors varied according to the modelled PC, but most of them included topography (catchment area, slope), climate (precipitation) and land cover (agriculture, coniferous and broadleaf forest) variables.

Parallel, the first six PCs of the PCA performed on the normalized flow series were retained. They explained 83.3 % of the variance (Table 3). The OOB misclassification rate of the RF models in the norClasF strategy ranged from 0.22 to 0.66 (Fig. 3). The most important variables differed between classifications comprising different class levels but in general precipitation, elevation, gradient and broadleaf forest were present in most models. For the norPredF strategy the mean OOB r^2 was 0.31 for the 6 PCs decreasing from 0.63 for PC2 to 0.08 for the PC6. The most important variables were not consistent between RF models although precipitation, elevation, pasture and broadleaf forest were present in most of them.

3.2 Comparison of classification performance

CS statistics for the classifications based on the raw flow series (rawClasF and rawPredF) showed similar patterns. CS increased from 2 to 6-Class level and in general, the analysis did not reveal significant differences (i.e. overlapped among standard error bars) beyond the 6-Class level (Fig. 4A). RawPredF showed generally higher CS values than rawClasF, although differences were not always significant.

The discrimination power of classifications for each of hydrological indices (ANOVA) got higher with increasing number of classes (Fig. 5 and Supplementary material, Table S1).

However, in most cases there were not significant differences between classifications comprising a number of classes ranging from 6 to 20 classes. Moreover, rawPredF outperformed rawClasF, especially for those indices representing flow magnitude and duration (Fig. 5 and Supplementary material, Table S1).

NorPredF presented a progressive increment of CS from 2 to 10-Class level where it reached the maximum value, suffering then only slight variations (Fig. 4B). NorClasF presented a more unstable CS pattern than norPredF. Except for specific class levels (2 and 4-Class levels), norPredF reached higher CS than norClasF.

The discrimination ability of norClasF and norPredF on individual indices showed similar patterns to those found for classifications based on raw series. An increase in r^2 with increasing number of classes and the presence of an inflexion located between 6 and 10-Class levels were observed (Fig. 6 and Supplementary material, Table S2). In addition, although norPredF performed better than norClasF, differences were not significant in several cases.

In general, the classifications based on the raw flow series (rawClasF and rawPredF) provided slightly higher CS (Fig. 4) and r^2 values (Figs. 5 and 6) than those based on normalized series (norClasF and norPredF).

3.3 Hydrological Interpretation of classifications

According to the hydrological indices with the highest values on each axis in the PCA performed on the raw flow series, PC1 represented the magnitude of the mean annual and high flows, while PC2 represented the frequency of high flow events and the magnitude of low flows. PC3 was also related to the frequency of high flow events while PC4 and PC5 represented the interannual variability of different hydrological characteristics (Table 3). The hydrological interpretation of the PCs became more difficult as explained variance decreased. In addition, ANOVA analysis revealed higher r^2 values of indices related to flow magnitude and frequency than those representing other aspects of the flow regime (Fig. 5 and Supplementary material, Table S1).

The PCA performed on the normalized flow series showed that PC1 represented the variability of the annual mean flow and the magnitude and duration of extreme low flows and

PC2 represented the variability of the magnitude and duration of high flow events. PC3 to PC6 are mainly related with indices representing flow magnitude in different months. Thus, they represented the shape and variability of the hydrograph across the year. In regard to the ANOVA, the highest r^2 values were obtained for the indices representing mean monthly flows. The maxima reached by the indices representing mean and duration of extreme flows was 0,3 (Fig. 6 and Supplementary material, Table S2). In addition, both norClasF and norPredF showed high discrimination ability on indices representing the frequency of high flow events, despite these indices not identified as important in any PCs.

3.4 Analysis of distinctive gauges

Three of the four DGs selected from the raw flow series were situated in the Ebro catchment and one in the Cantabric region. The distance between each distinctive gauge and its respective class medoid in the rawClasF classifications was lower than the distance in the rawPredF classification more than two thirds of the times. However, the relative differences were generally below 20 % (Table 4). In addition, for the rawClasF it was observed that the proportion of the classification domain assigned to the classes in which the distinctive gauges were included presented very low frequencies. This was especially visible beyond the 6-Class level where this proportion was below 1 % for the four distinctive gauges (Fig. 7A). Regarding the rawPredF the proportions of the classes containing the distinctive gauges were higher than those for the rawClasF (Fig. 7B).

The classifications based on the normalized flow series presented two distinctive gauges situated in the Ebro catchment and the other two in two Catalan catchments. NorPredF showed smaller distances between the distinctive gauges and their respective class medoids than norClasF 95 % of the times. In addition, more than half of the times differences were over 40 % (Table 4). The comparison of the frequency of the classes containing the distinctive gauges did not reveal important differences between norClasF and norPredF (Fig. 7C and 7D).

3.5 Correspondence between classifications

The ARIs for each pair of classifications were in the range 0.12-0.4 for the 6-Class level and in the range 0.14-0.34 for the 11, 16-Class level and the mean of all classification levels (Table 5). The highest ARI was obtained between rawPredF and norPredF (≥ 0.4). Contrary, rawClasF and norClasF showed the lowest correspondence (≤ 0.15).

4 Discussion

As expected the different data specification and classification procedures analysed in this study exerted a significant influence in the classifications outcomes. The normalization of flow data generated hydrological classifications that were not completely subjected to the flow magnitude and the size of the river as if data were not normalised. Consequently, classifications based on normalized series were more difficult to interpret and predict. In addition, classifications based on PredF outperformed those obtained with ClasF and presented a greater ability than ClasF to deal with the underrepresented parts of the hydrological space.

4.1 Comparison of classification performance

Similar classification performance measured through CS and ANOVA was observed in relation to the results obtained by Snelder and Booker (2013) in New Zealand rivers. The specific classification characteristics depend upon the selected gauged network and the hydrological behaviour of the rivers in the target study zone. However, the similarity of the results with those obtained by Snelder and Booker (2013) highlights the possibility to discern more clearly the benefits and drawbacks of the different classification strategies and data specification.

Our analysis demonstrated that the PredF performed better than ClasF and significant differences in the ability to discriminate hydrological characters were found for several class levels. The higher performance of PredF classifications is supported by the conceptual basis of this approach and can be explained by two main reasons: the equalization of data and loss of information produced in ClasF strategy coupled with the effective data processing of the PredF strategy. Firstly, ClasF imposes sharp barriers to the observed hydrological space, i.e. the gauged network, and not over the whole hydrologic space of the fluvial network. The creation of classes produced a equalization of hydrologic data within classes. Given that the subsequent prediction step enforces congruence of all the river segments of the SRN with those previously created classes, the equalization of data could be linked to a loss of information when classes are predicted. Moreover, the real extent to which such discrete groupings exist is uncertain (Kennard et al., 2010). In contrast, the aim of PredF is to account for the whole hydrological variability in the SRN before conducting the classification. This process generated a more complete distribution of the hydrologic variables which is in accordance with the actual hydrology of the SRN. This avoided the bias associated to gauge location. Moreover, PredF does not assume any interactions between the various dependent variables for each RF, which is true as the PCA created orthogonal and independent variables.

In addition, it must be pointed out that although prediction of classes and synthetic indices is not entirely comparable, results indicated that similar prediction performance can be assumed for both strategies. Hence, the prediction of classes or synthetic indices in not a major determinant in the better results obtained by PredF.

In general, the specification of the initial hydrological data has also significant consequences in the classification performance. Classifications based on raw flow series had higher discrimination ability for individual indices than those based on normalized flow series (Figs. 5 and 6). As discussed below, classifications based on raw series discriminated rivers based almost exclusively on flow magnitude, which greatly depends on river size. In contrast, classifications based on normalized flow series considered a greater range of hydrological aspects. Obviously, the variability of river size shows a clear pattern within river networks and thus, it is a straightforward approach to segregate river reaches. In contrast, the consideration of a higher spectrum of hydrologic aspects hampered the creation of so evident classes and thus classifications achieved lower discrimination ability.

4.2 Hydrological interpretation of classifications

To our knowledge this is the first study that has compared the consequences of classifying river networks attending to the initial data specification: the use of raw or normalized flow series. The PCA performed on the raw series showed that the first PC explained more than two thirds of the hydrological variability in the study region. This PC was mainly related to the magnitude of mean annual and high flows. Thereby, the magnitude of flow was the major determinant to segregate rivers, as expected. In addition, indices accounting with the frequency of high flow events were also represented in other PCs and therefore, this flow attribute also showed a relatively important contribution in the classifications (Table 6).

Moreover, the ANOVA analysis also showed that all the indices related to flow magnitude, even those not included as the most important ones in any PC presented important differences between classes. This is not surprising given the high correlation between all the flow magnitude indices. However, although these classifications segregated river reaches according to flow magnitude, they were unable to incorporate the severity of droughts, i.e. the magnitude that these episodes represent in relation to the mean flow condition. The pattern of droughts in the study area is an essential element that should be considered in the classifications given the Mediterranean character of the study zone. The fact that the high differences in flow magnitude between large and small rivers have accounted with the largest percentage of variability, have probably masked the effects of low flow attributes. Contrary to our results, Belmar et al (2011) and Chinnayakanahalli et al (2011) working in areas influenced by the Mediterranean climate found that several hydrologic characteristics related to drought were considered in the synthetic hydrologic indices, even if the series were not normalized by the mean annual flow. We expected that the characteristic intermittency of many Mediterranean streams had been represented in the synthetic indices. However, the lack of this attribute in our classifications may be attributed to the scarcity of gauges situated in intermittent streams.

On the other hand, the interpretation of the classifications based on normalized flow series differed completely to those derived from raw flow series (Table 6). The main differences can be summarized in two essential aspects. First, the proportion of variance explained by the different PCs was more evenly distributed in the normalized than in the raw flow series. Therefore these classifications were not uniquely subjected to just one hydrologic attribute. Second, it was observed that the indices with the highest loading in each PC and hence, their interpretation, varied considerably depending on the data processing and specification (Table 6). Given the higher number of flow attributes influencing the classifications based on normalized flow series, their interpretation was more difficult than those based on raw flow series. In this regard, magnitude and duration of low flow conditions were represented in PC1. Hence, the Mediterranean character of the rivers was one of the main attributes for classification. In addition, PC3 to PC6 were related to the magnitude of flows in different months and periods through the year, therefore classification accounted with the shape of the hydrograph as it has been observed in other works (Bejarano et al., 2010;Solans and Poff, 2013;Snelder et al., 2009). Contrary to expected, other indices not related to flow magnitude, such as the frequency of high flow events were not included as important indices in any PC. Nonetheless, the ANOVA analysis showed the high ability of classifications based on normalized flow data to discriminate the indices representing frequency (Fig. 6). Therefore it was assured that such an important hydrological aspect played an important role to define the classification patterns. In addition, classifications and hydrologic attributes based on normalized flow series were also more difficult to predict. For instance, while flow magnitude indices depended almost uniquely in the catchment area and climate patterns, other hydrologic attributes such as duration or frequency of different flow events were related with other environmental variables that were more difficult to characterize.

The interpretation and meaning of classifications are essential to determine their further use. As stated before, classifications based on raw flow series segregated rivers according almost exclusively to flow magnitude. This provides an important loss of hydrologic information which limits its use to water resource and flooding management issues. However, results demonstrated that these classification did not take into account the drought patterns in the study area. Hence, even within the water resources field, these classification would not result effective in dealing with low flow issues, such as the environmental flows or reliability of water supply in drought situations. In contrast, segregating rivers according to a larger spectrum of hydrological attributes widens its potential applications. For instance, many other hydrological attributes different from magnitude may be potentially altered by human perturbations. Hence, following the principles established in the ELOHA framework (Poff et al. 2010), classifications based on normalized flow series may be more valuable in evaluating the hydrologic alteration caused by human perturbations or the response of freshwater ecosystems to these flow alteration.

Finally, it must be pointed out that any of the classifications, whether they were based on raw or normalized data, failed to represent some other important hydrologic aspects such as timing of extreme flow events and rate of change (Table 6). These attributes presented a modest spatial variability within the study area which ultimately resulted in a small contribution to the hydrologic classifications.

4.3 Analysis of distinctive gauges

The analyses demonstrated that the PredF approach presented greater capability than ClasF to deal with the underrepresented parts of the hydrological space in the data set. If data were not normalized, rawClasF approach generated classes that were comprised by the distinctive gauge plus a very limited number of gauges, in most of the cases less than four. In these cases, the distance between the DG and the medoid of the class was similar to the mean distance calculated for the other gauges included in the class. Therefore, it can be assumed that these classes were relatively homogeneous in regard to its hydrologic characteristics. However, when classes where predicted to the SRN, their frequencies were normally lower than 1 %. This means that the hydrological characteristics accounted in these classes where almost lost after the prediction step in of the rawClasF. Moreover, their frequencies were probably well below the actual frequencies of those river classes.

On the other hand, the normalization of the flow series smoothed the differences between gauges due to the reduction of the influence of low magnitude, which implied that DGs in the norClasF classifications were not isolated into such exclusive classes as those found in rawClasF. This greatly reduced the problem associated with the low frequency of these classes when they were predicted to the SRN. However, the distance between the DGs and the medoid was normally over two times the mean distance of the other gauges included in the class. This indicated that DGs were grouped to other gauges that are not hydrologically similar. Hence, it is assumed that the hydrologic characteristics accounted by the DG were not represented at all in any of the classes.

By contrast, when the PredF approach was applied, these rare hydrologic characteristics are predicted to a larger number of segments before classifying the SRN. Consequently, the proportion of segments accounting with these rare characteristics increased. In the subsequent step of classification, these segments accounting with the rare hydrological characteristics were grouped in specific classes and hence, the frequencies of these classes were more adjusted to the actual distribution of river types in the study area.

4.4 Correspondence between classifications

The ARI analysis showed that the correspondence between rawClasF and rawPredF and between norClasF and norPredF presented a similar pattern. The ARI values in these two cases were around 0.2 which implies important differences in the spatial distribution of classes. This indicated that the strategy used to predict class membership to the SRN (ClasF vs. PredF) is a critical specification in the classification procedure. In contrast to the expected outcome, ARI analyses also showed that classifications obtained through the PredF approach, regardless of the initial data processing (i.e. rawPredF or norPredF), presented the highest spatial correspondence. This result highlights that the prediction of the hydrological characteristics to the SRN before classifying is probably generating classifications more adjusted to the actual spatial distribution of river types, even if classifications presented different interpretation.

5 Conclusion

In conclusion, this study showed that the methodological specifications used throughout the classification process greatly influences classification outcomes and performance. Although the comparison between ClasF and PredF did not reveal significant differences for several classification levels, the classifications based on PredF produced, in general, higher classification performance, greater ability to deal with the presence of distinctive gauges in the data set and a spatial distribution of classes more adjusted to the actual river types. PredF produced classes that presented higher intra-class homogeneity and higher inter-class heterogeneity than ClasF. In general, the segregation of gauges before the prediction step in the ClasF produced a loss of information due to the presence of under and overrepresented hydrologic characteristics. In contrast, the prediction of the hydrologic characteristics previous to the classification step avoided these bias associated to gauge location. These features are very valuable when applying these classifications with different objectives. For instance, classifications developed through PredF represents the best strategy to further detect not only the hydrological alteration caused by human perturbations but the ecological impact associated to this alteration. Given all these strengths, we recommend the application of the PredF strategy to develop hydrological classifications at the regional scale. Finally, the specification of flow data influenced the interpretation of the hydrological classes. The normalization of flow data removed the effect of flow magnitude and generated classifications in which a larger spectrum of hydrologic characteristics was considered. This widens the potential range of management and ecological applications of the classification as classifications would not be subjected to a unique hydrological attribute. In all the cases, the selection of the most suitable number of classes is difficult to be accomplished from completely objective criteria, as many times, classifications with different level of detail presented similar statistical performance.

APPENDIX A: Hydrological indices used in the classification

Table A1: Hydrological indices used in the classification. Overall mean and standard deviation (referred in the manuscript by the prefix sd) of annual values for each index except for I1, I2, lca, lcv, ikur, X5, X25, X75 and X95. I1 was not calculated for Normalized flow series.

Group	Name	Description
1) Magnitude of annual and monthly flows	l1	Linear moment that represents the mean of the calculated flow duration curve
	l2	Linear moment that represents the variance of the calculated flow duration curve
	lca	Linear moment that represents the skewness of the calculated flow duration curve
	lcv	Linear moment that represents the coefficient of variation of the calculated flow duration curve
	ikur	Linear moment that represents the kurtosis of the calculated flow duration curve
	M1-M12	Mean monthly flow. Standard deviation for each index was calculated.
	MxM1-MxM12	Maximum monthly flow

	MnM1- MnM12	Minimum monthly flow
2) Magnitude and duration of annual extremes	1LF	Magnitude of minimum annual flow of 1 day duration. sd was also calculated
	7LF	Magnitude of minimum annual flow of 7 day duration.
	30LF	Magnitude of minimum annual flow of 30 day duration.
	90LF	Magnitude of minimum annual flow of 90 day duration.
	X75	Mean magnitude of flow exceeded 75% of the time
	X95	Mean magnitude of flow exceeded 95% of the time
	1HF	Magnitude of maxima annual flow of 1 day duration
	7HF	Magnitude of maxima annual flow of 7 day duration
	30HF	Magnitude of maxima annual flow of 30 day duration
	90HF	Magnitude of maxima annual flow of 90 day duration
	X25	Magnitude of the flows exceeded 25 %

		of the time. High flow pulses
	X5	Magnitude of the flows exceeded 5 % of the time.
	ZFD	Number of zero flow days
	BFI	Seven-day minimum flow divided by mean annual daily flows
3) Timing of extreme flow events	JMin	Julian day of minimum flow
	JMax	Julian day of annual maximum flow
	Pred	Predictability (sensu Colwell, 1974)
4) Frequency and duration of high pulses	FRE1	Number of high flow events per year using an upper threshold of 1 time median flow over all years
	FRE3	Number of high flow events per year using an upper threshold of 3 time median flow over all years
	FRE7	Number of high flow events per year using an upper threshold of 7 time median flow over all years
	nPHigh	Number of high pulses within each year
	dPHigh	Duration of high pulses within each

		year
4) Rate and frequency of flow changes	Pos	Mean of all positive differences between days
	nPos	Number of days with increasing flow
	Neg	Mean of all negative differences between days
	nNeg	Number of days with decreasing flow
	Rev	Number of hydrologic reversals

Acknowledgements

This study was partly funded by the Spanish Ministry of Economy and Competitiveness as part of the project MARCE (Ref: CTM-2009-07447; http://marce.ihcantabria.es/) and the Ministry of Agriculture, Food and Environment as part of the project RECORAM (Ref: 132/2010). José Barquín is supported by a Ramon y Cajal grant (Ref: RYC-2011-08313) of the Ministry of Economy and Competitiveness. We would also like to thanks Confederación Hidrográfica del Cantábrico, Confederación Hidrográfica del Ebro, Agencia Vasca del Agua, Agencia Catalana del Agua and Gobierno de Navarra for providing flow series data. Authors also thank two anonymous reviewers and Erwin Zehe for their comments which improved the manuscript.

References

Alcázar, J., and Palau, A.: Establishing environmental flow regimes in a Mediterranean watershed based on a regional classification, J. Hydrol., 388, 41-51, doi: 10.1016/j.jhydrol.2010.04.026, 2010.

Álvarez-Cabria, M., Barquín, J., and Juanes, J. A.: Spatial and seasonal variability of macroinvertebrate metrics. Do macroinvertebrate assemblages track river health?, Ecol. Indic., 10, doi: 370-379, 10.1016/j.ecolind.2009.06.018, 2010.

Bejarano, M. D., Marchamalo, M., García de Jalón, D., and González del Tánago, M.: Flow regime patterns and their controlling factors in the Ebro basin (Spain), J. Hydrol., 385, 323-335, doi: 10.1016/j.jhydrol.2010.03.001, 2010.

Belmar, O., Velasco, J., and Martínez-Capel, F.: Hydrological Classification of Natural Flow Regimes to Support Environmental Flow Assessments in Intensively Regulated Mediterranean Rivers, Segura River Basin (Spain), Environ. Manage., 47, 992-1004, doi: 10.1007/s00267-011-9661-0, 2011.

Benda, L., Poff, N. L., Miller, D., Dunne, T., Reeves, G., Pess, G., and Pollock, M.: The Network Dynamics Hypothesis: How Channel Networks Structure Riverine Habitats, Bioscience, 54, 413-427, doi: http://dx.doi.org/10.1641/0006-3568(2004)054[0413:TNDHHC]2.0.CO;2, 2004.

Breiman, L.: Random Forest, Mach. Learn., 45, 5-32, doi: 10.1023/A:1010933404324, 2001.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J.: Classification and regression trees, Wadsworth, Inc., Monterey, Calif., U.S.A., 1984.

Bunn, S. E., and Arthington, A. H.: Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity, Environ. Manage, 30, 492-507, doi: 10.1007/s00267-002-2737-0, 2002.

Colwell, R. K.: Predictability, Constancy, and Contingency of Periodic Phenomena, Ecology, 55, 1148-1153, 1974.

Chinnayakanahalli, K. J., Hawkins, C. P., Tarboton, D. G., and Hill, R. A.: Natural flow regime, temperature and the composition and richness of invertebrate assemblages in streams of the western United States, Freshwater Biol., 56, 1248-1265, doi: 10.1111/j.1365-2427.2010.02560.x, 2011.

Gámiz-Fortis, S. R., Hidalgo-Muñoz, J. M., Argüeso, D., Esteban-Parra, M. J., and Castro-Díez, Y.: Spatio-temporal variability in Ebro river basin (NE Spain): Global SST as potential source of predictability on decadal time scales, J. Hydrol., 409, 759-775, doi: 10.1016/j.jhydrol.2011.09.014, 2011.

Gurnell, A. M., Hupp, C. R., and Gregory, S. V.: Linking hydrology and ecology, Hydrol. Process., 14, 2813-2815, doi: 10.1002/1099-1085(200011/12)14:16/17<2813::AID-HYP120>3.0.CO;2-Q, 2000.

Hastie, T., Tibshirani, R., and Friedman, J. H.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer-Verlag, New York, 2001.

Hubert, L., and Arabie, P.: Comparing Partitions, J. Classif., 2, 193-218, 1985.

Jackson, D. A.: Stopping Rules in Principal Components-Analysis - a Comparison of Heuristic and Statistical Approaches, Ecology, 74, 2204-2214, doi: 10.2307/1939574, 1993.

Kauffman, L., and Rousseeuw, P. J.: Finding groups in data. An introduction to cluster analysis, Wiley and Sons, New-York, 1990.

Kennard, M. J., Pusey, B. J., Olden, J. D., MacKay, S. J., Stein, J. L., and Marsh, N.: Classification of natural flow regimes in Australia to support environmental flow management, Freshwater Biol., 55, 171-193, doi: 10.1111/j.1365-2427.2009.02307.x , 2010.

McManamay, R. A., Orth, D. J., Dolloff, C. A., and Frimpong, E. A.: A regional classification of unregulated stream flows: Spatial resolution and hierarchical frameworks, River Res. Appl., doi: 10.1002/rra.1493, 2012.

Miller, D.: Programs for DEM Analysis., Earth System Institute, 38 pp, 2003.

Monk, A. W., Wood, P. J., and Hannah, D. M.: Examining the influence of flow regime variability on instream ecology, in: Hydroecology and Ecohydrology: past, present and future, edited by: Wood, P. J., Hannah, D. M., and Sadler, J. P., John Wiley & Sons, Ltd., Chichester, 2007.

Morán-Tejeda, E., López-Moreno, J. I., Ceballos-Barbancho, A., and Vicente-Serrano, S. M.: River regimes and recent hydrological changes in the Duero basin (Spain), J. Hydrol., 404, doi: 241-258, doi: 10.1016/j.jhydrol.2011.04.034, 2011.

Naiman, R. J., and Dudgeon, D.: Global alteration of freshwaters: influences on human and environmental well-being, Ecol. Res., 26, 865-873, doi: 10.1007/s11284-010-0693-3, 2011.

Olden, J. D., and Poff, N. L.: Redundancy and the choice of hydrologic indices for characterizing streamflow regimes, River Res. Appl., 19, 101-121, doi: 10.1002/rra.700, 2003.

Olden, J. D., Kennard, M. J., and Pusey, B. J.: A framework for hydrologic classification with a review of methodologies and applications in ecohydrology, Ecohydrology, 5, 503–518, doi: 10.1002/eco.251, 2012.

Peñas, F. J., Barquín, J., Snelder, T., Booker, D., Fernandez, D., and Álvarez-Cabria, M.: Do rivers reaches differ in habitat-flow relationships according to hydrologic classification and river size?, 9th International Symposium on Ecohydraulics Proceedings, Vienna, Austria, 17-21 September 2012, 14956, 2012

Poff, N. L.: A hydrogeography of unregulated streams in the United States and an examination of scale-dependence in some hydrological descriptors, Freshwater Biol., 36, 71-91, doi: 10.1046/j.1365-2427.1996.00073.x, 1996.

Poff, N. L., Olden, J. D., Pepin, D. M., and Bledsoe, B. P.: Placing global stream flow variability in geographic and geomorphic contexts, River Res. Appl., 22, 149-166, doi: 10.1002/rra.902, 2006.

Poff, N. L., and Zimmerman, J. K. H.: Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows, Freshwater Biol., 55, 194-205, doi: 10.1111/j.1365-2427.2009.02272.x, 2010.

Reidy Liermann, C. A., Olden, J. D., Beechie, T. J., Kennard, M. J., Skidmore, P. B., Konrad, C. P., and Imaki, H.: Hydrogeomorphic classification of Washington state rivers to support emerging environmental flow management strategies, River Res. Appl., 28, 1340-1358, doi: 10.1002/rra.1541, 2012.

Richter, B. D., Baumgartner, J. V., Powell, J., and Braun, D. P.: A method for assessing hydrologic alteration within ecosystems, Conserv. Biol., 10, 1163-1174, doi: 10.1046/j.1523-1739.1996.10041163.x, 1996.

Richter, B. D., Baumgartner, J. V., Braun, P. D., and Powell, J.: A spatial assessment of hydrologic alteration within a river network, Regul. Rivers, 14, doi: 0.1002/(SICI)1099-1646(199807/08)14:4<329::AID-RRR505>3.0.CO;2-E,1998.

Rivas-Martínez, S., Penas, A., and Díaz, T. E.: Bioclimatic Map of Europe, Bioclimates, Cartographic Service. University of León, León, Spain, 2004.

Snelder, T. H., Pella, H., Wasson, J. G., and Lamouroux, N.: Definition Procedures Have Little Effect on Performance of Environmental Classifications of Streams and Rivers, Environmental Management, 42, 771-788, 10.1007/s00267-008-9188-1 2008.

Snelder, T. H., Lamouroux, N., Leathwick, J. R., Pella, H., Sauquet, E., and Shankar, U.: Predictive mapping of the natural flow regimes of France, Journal of Hydrology, 373, 57-67, 10.1016/j.jhydrol.2009.04.011, 2009.

Snelder, T. H., Lamouroux, N., and Pella, H.: Empirical modelling of large scale patterns in river bed surface grain size, Geomorphology, 127, 189-197, 10.1016/j.geomorph.2010.12.015, 2011.

Snelder, T. H., Barquin Ortiz, J., Booker, D. J., Lamouroux, N., Pella, H., and Shankar, U.: Can bottom-up procedures improve the performance of stream classifications?, Aquatic Sciences, 74, 45-59, 10.1007/s00027-011-0194-7, 2012.

Snelder, T. H., and Booker, D.: Natural flow regime classifications are sensitive to definition procedures, River Research and Applications, 7, 822-838, 10.1029/2009WR008839, 2013.

Solans, M. A., and Poff, N. L.: Classification of Natural Flow Regimes in the Ebro Basin (Spain) by using a Wide Range of Hydrologic Parameters, River Res. Appl., doi: 10.1002/rra.2598, 2012.

Van Sickle, J.: Using mean similarity dendogram to evaluate classifications, J. Agr. Biol. Envir. St., 2, 370-388, 1997.

Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R., and Cushing, C. E.: The river continuum concept, Can. J. Fish. Aquat. Sci., 37, 130-137, 1980.

Zhang, Y., Arthington, A. H., Bunn, S. E., Mackay, S., Xia, J., and Kennard, M.: Classification of flow regimes for environmental flow assessment in regulated rivers: The Huai River Basin, China, River Res. Appl., 28, 989-1005, doi: 10.1002/rra.1483, 2012.

Table 1. Number of retained years for flow time-series used in the analysis.

N. of years	N. of gauges	Frequency	Freq. acum.
>19	52	33.3	33.3
19	3	1.9	35.3
18	7	4.5	39.7
17	6	3.8	43.6
16	16	10.3	53.8
15	7	4.5	58.3
14	8	5.1	63.5
13	8	5.1	68.6
12	11	7.1	75.6
11	9	5.8	81.4
10	9	5.8	87.2
9	9	5.8	92.9
8	11	7.1	100.0

Table 2. Environmental variables used to predict classes or the synthetic hydrologic indices onto the ungauged segments of the river network (TG: Topography; CL: Climatic LC: Land Cover; GL: Geology)

Variable	Type	Units	Description	Source
Precipitation	CL	Mm	Annual catchment precipitation	SIMPA
Precipitation range	CL	mm	Range between maximum and minimum	SIMPA
Evapotranspiration	CL	Mm	Annual catchment evapotranspiration	SIMPA
Catchment area	TG	Km^2	Total catchment area	DEM
Slope	TG	%	Average catchment gradient	DEM
Elevation	TG	m	Average catchment elevation	DEM
Confluence density	TG	-	Number of rivers confluences by	DEM
Drainage density	TG	-	Number of segments divided by the	DEM
Broadleaf forest	LC	%	Surface occupied by broadleaf forest	SIOSE
Coniferous forest	LC	%	Surface occupied by coniferous	SIOSE
Pasture	LC	%	Surface occupied by pasture	SIOSE
Agriculture	LC	%	Surface occupied by agricultural land	SIOSE
Denuded	LC	%	Surface occupied by denuded areas	SIOSE
Urban	LC	%	Surface occupied by urban areas	SIOSE
Permeability	GL	-	Terrain permeability	IGM
Hardness	GL	-	Rock hardness	IGM

Table 3. The 5 hydrologic indices with the highest loadings in each PC and variation explained by the retained PCs using the raw (above) and the normalized flow series (below). A minus sign indicates negative relation with the PC.

Axe	Hydrologic variables with the highest values in the PCs	Variation Explained (%)
PC1	-l1, -X25, -90HF, - 30HF, -M11	68
PC2	-FRE7, -FRE3, -lcv, BFI, sdBFI	10.6
PC3	-FRE1, -nPH, -FRE3, dPH, sdZFD	5.9
PC4	sdnPos, sdnNeg, ikur, lca	3.6
PC5	-sdnPH, sdJMax, -sdRev, -sdFRE3, -sdJmin	3.5
PC1	- l2, X75, 90LF, 30LF, 7LF	38.6
PC2	sd30HF, sd7HF, sd3HF, sd90HF, sdM5	20.4
PC3	-M10, -sdM10, -MXM10, -FRE1, sdM9	11.6
PC4	ikur, X25, MnM9, MnM2, MnM11	7.1
PC5	-M1, M5, sdZFD, -sdM1, -MxM1,	6.1
PC6	SdM8, MXM8, sdnPH, -MxM11, -sdM11	4.5

Table 4. Euclidean distance between the distinctive gauges (DG) and the medoid of the classes in which they were included for the 4, 6, 8, 10, 12, 16 and 20-Class levels classification. Distances were weighted by the mean difference of all the gauges included in the same class as the DG. Empty cells indicated that the gauge is the unique gauge in the class. Bold letters indicate the procedure that showed the lowest distance.

Raw series								
	MG 1		MG 2		MG 3		MG 4	
	rawClasF	rawPredF	rawClasF	rawPredF	rawClasF	rawPredF	rawClasF	rawPredF
4	2.95	**2.92**	**2.52**	2.97	**1.30**	1.46	**1.46**	1.65
6	6.45	**4.05**	**2.15**	3.07	**1.40**	1.53	1.67	**1.60**
8		4.30	**2.06**	2.28	**1.20**	1.60	1.83	**1.45**
10		3.64	**2.06**	2.91	**1.20**	1.50	1.58	**1.51**
12			3.63	**3.51**	**1.35**	1.51	1.88	**1.70**
16		3.15		2.18	**1.05**	1.47	1.47	**1.71**
20		2.71		2.39	**1.05**	1.26	**1.26**	1.66

Normalized series								
	MG 1		MG 2		MG 3		MG 4	
	norClasF	norPredF	norClasF	norPredF	norClasF	norPredF	norClasF	norPredF
4	3.46	**1.67**	1.96	**1.55**	1.85	**1.60**	1.98	**1.69**
6	2.16	**1.42**	1.93	**1.50**	2.88	**1.72**	1.96	**1.34**
8	2.22	**1.69**	1.94	**1.30**	1.87	**1.40**	1.44	**1.39**
10	1.59	**1.71**	2.10	**1.44**	1.89	**1.46**	1.46	**1.25**
12	1.66	**1.32**	2.14	**1.33**	1.88	**1.73**	1.45	**1.19**
16		1.34	0.94	**0.75**	1.83	**1.22**	1.82	**1.20**
20		1.45	1	**1.50**	1.83	**0.92**	1.82	**0.91**

Table 5. Adjusted Rand Index (ARI) for the 6, 11 and 16-Class level and the mean of all class levels classifications following the four approaches.

Level	Classification	Classification		
		rawClasF	rawPredF	norClasF
6	rawPredF	0.22		
	norClasF	0.12	0.16	
	norPredF	0.19	0.39	0.19
11	rawPredF	0.23		
	norClasF	0.14	0.23	
	norPredF	0.19	0.32	0.23
16	rawPredF	0.20		
	norClasF	0.17	0.17	
	norPredF	0.17	0.34	0.21
Mean of all levels	rawPredF	0.22		
	norClasF	0.16	0.18	
	norPredF	0.18	0.32	0.21

Table 6. Relative representativeness of each flow regime aspect according to the data processing previous to classification procedure. (–None; *Limited; ** Moderate; *** High)

Flow Aspect		Raw	Normalized
Magnitude of annual flows	Mean	***	
	Variability	*	***
Magnitude of monthly flows	Mean	-	***
(shape of the hydrograph)	Variability	-	**
Magnitude and duration of low flows	Mean	-	***
	Variability	-	-
Magnitude and duration of high flows	Mean	***	-
	Variability	-	***
Timing of extreme flow events	Mean	-	-
	Variability	*	-
Frequency and duration of high pulses	Mean	**	**
	Variability	-	-
Rate and frequency of flow change	Mean	-	-
	Variability	*	-

Figure caption

Fig. 1. Map of unregulated gauges (●; n=156) in the study area. Black lines divide the Cantabric, the Ebro and the Catalan catchments. (CS: Cantabric sea; MS: Mediterranean sea).

Fig. 2. Schematic diagram summarising the 4 classifications strategies.

Fig. 3. Out-of-Bag misclassification rate of the random forest models developed for the 2 to 20-Class level classifications using ClasF strategy based on the synthetic indices derived from the raw (△; rawClasF) and the normalized flow series (◇: norClasF).

Fig. 4. Performance of the classifications based on the Classification Strength statistic A) classifications based on raw flow series (●: rawPredF; △: rawClasF); B) classifications based on normalized series (■: norPredF; ◇: norClasF).

Fig. 5. Performance of the classifications derived from the raw flow series based on ANOVA analysis on individual indices analysis (●: rawPredF; △: rawClasF). We selected one index representing each aspect of the natural flow regime to illustrate the results (the values obtained for the 103 indices are included in Supplementary material, Table S1).

Fig. 6. Performance of the classifications derived from the normalized flow series based on individual indices analysis(■norPredF; ◇ norClasF). We selected one index representing each aspect of the natural flow regime to illustrate results (the values obtained for the 101 indices are included in Supplementary material, Table S2).

Fig. 7. Frequency (%) of the segments of the classification domain assigned to the classes where the distinctive gauges were included. (A: rawClasF ;B: rawPredF; C: norClasF; D: norPredF).

Figure 1

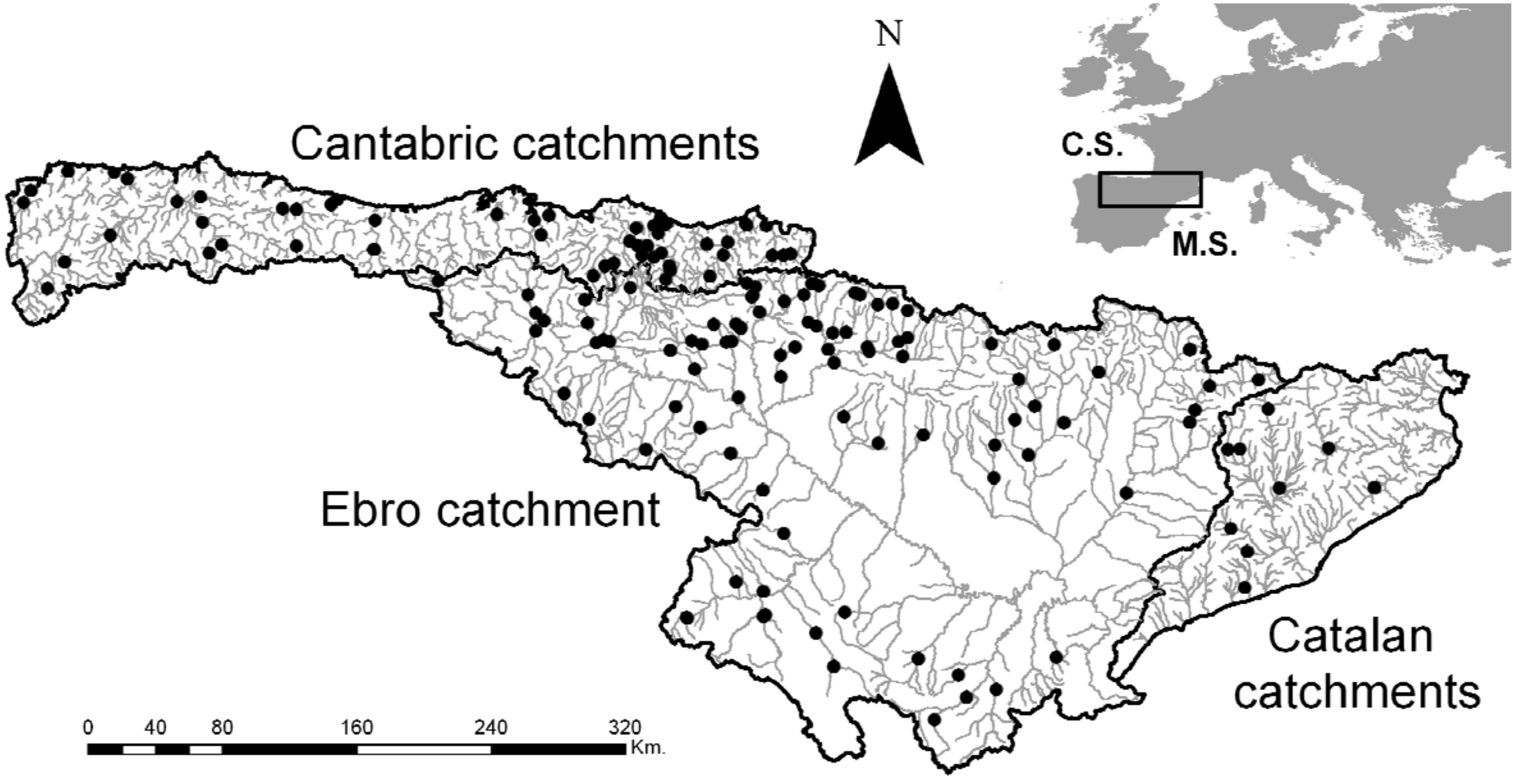

N
C.S.
M.S.
Cantabric catchments
Ebro catchment
Catalan catchments
0
40
80
160
240
320
Km.

Figure 2

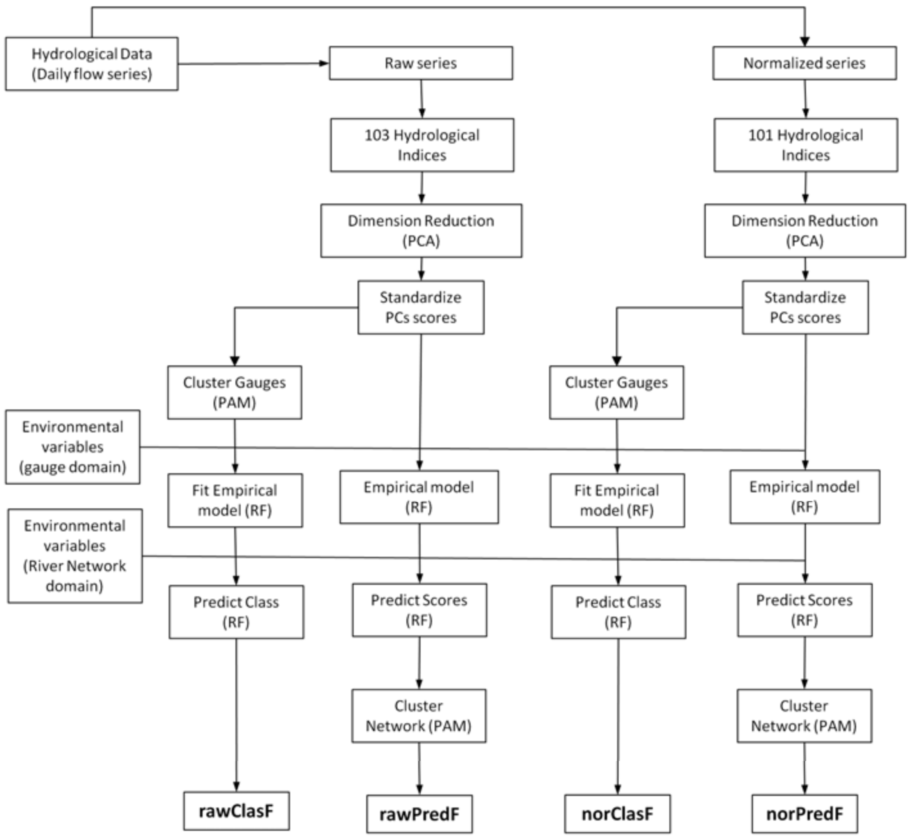
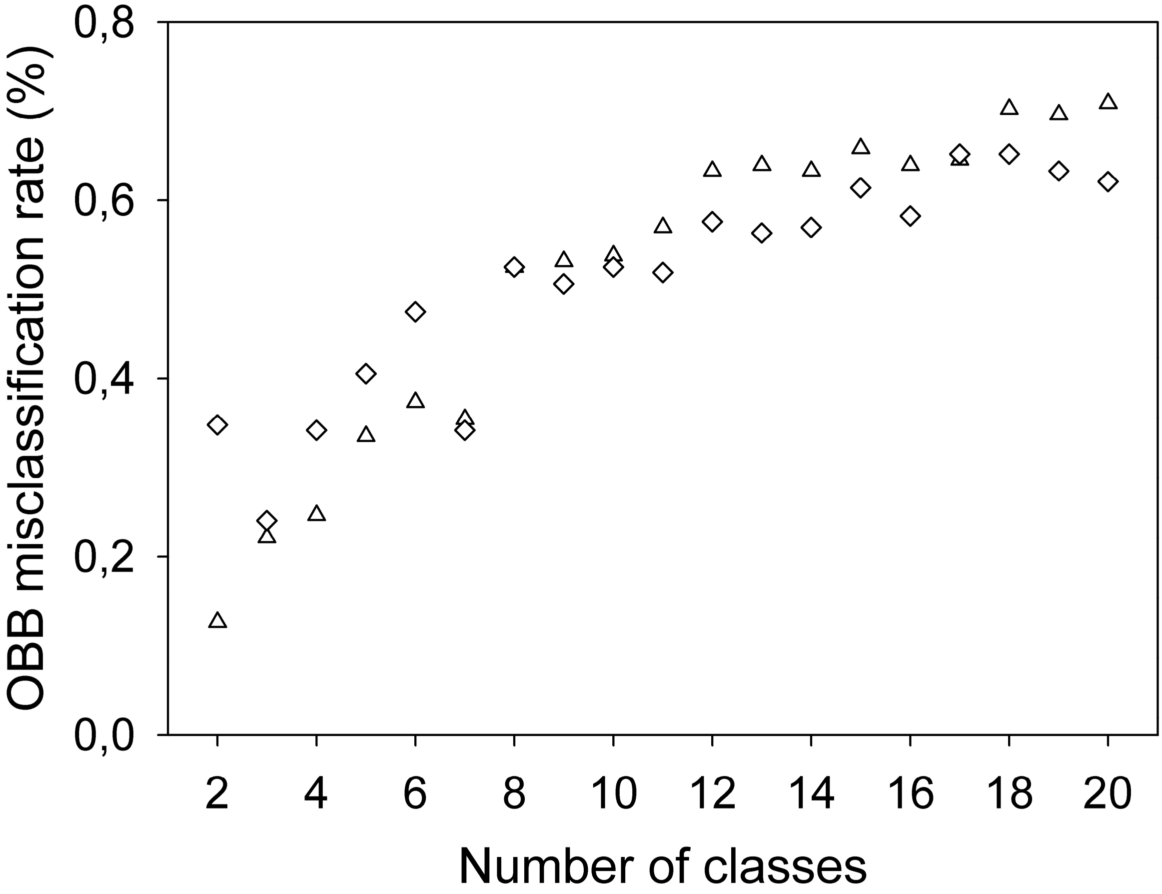

Hydrological Data (Daily flow series)
Raw series
Normalized series
103 Hydrological Indices
101 Hydrological Indices
Dimension Reduction (PCA)
Dimension Reduction (PCA)
Standardize PCs scores
Standardize PCs scores
Cluster Gauges (PAM)
Cluster Gauges (PAM)
Environmental variables (gauge domain)
Fit Empirical model (RF)
Empirical model (RF)
Fit Empirical model (RF)
Empirical model (RF)
Environmental variables (River Network domain)
Predict Class (RF)
Predict Scores (RF)
Predict Class (RF)
Predict Scores (RF)
Cluster Network (PAM)
Cluster Network (PAM)
rawClasF
rawPredF
norClasF
norPredF

Figure 3



0,8
0,6
0,4
0,2
0,0
OBB misclassification rate (%)
2
4
6
8
10
12
14
16
18
20
Number of classes

Figure 4

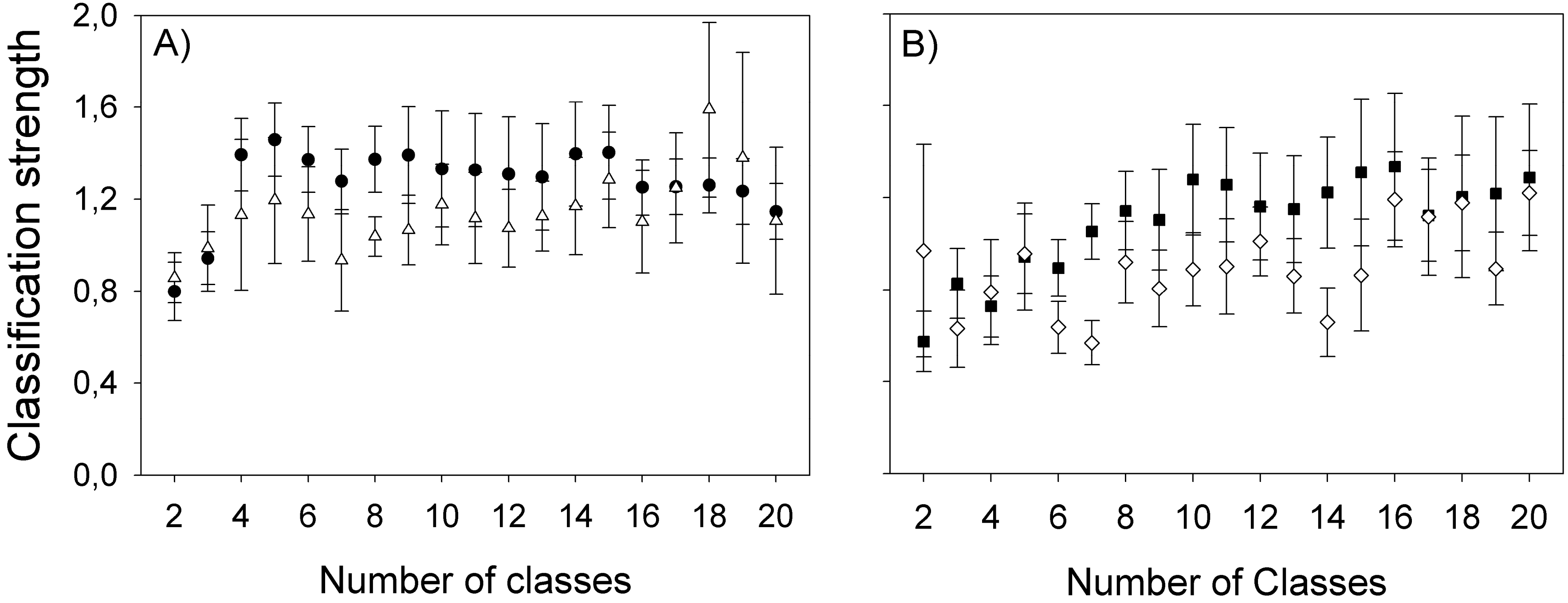

A)
B)
Classification strength
2,0
1,6
1,2
0,8
0,4
0,0
2 4 6 8 10 12 14 16 18 20
Number of classes
Number of Classes

Figure 5

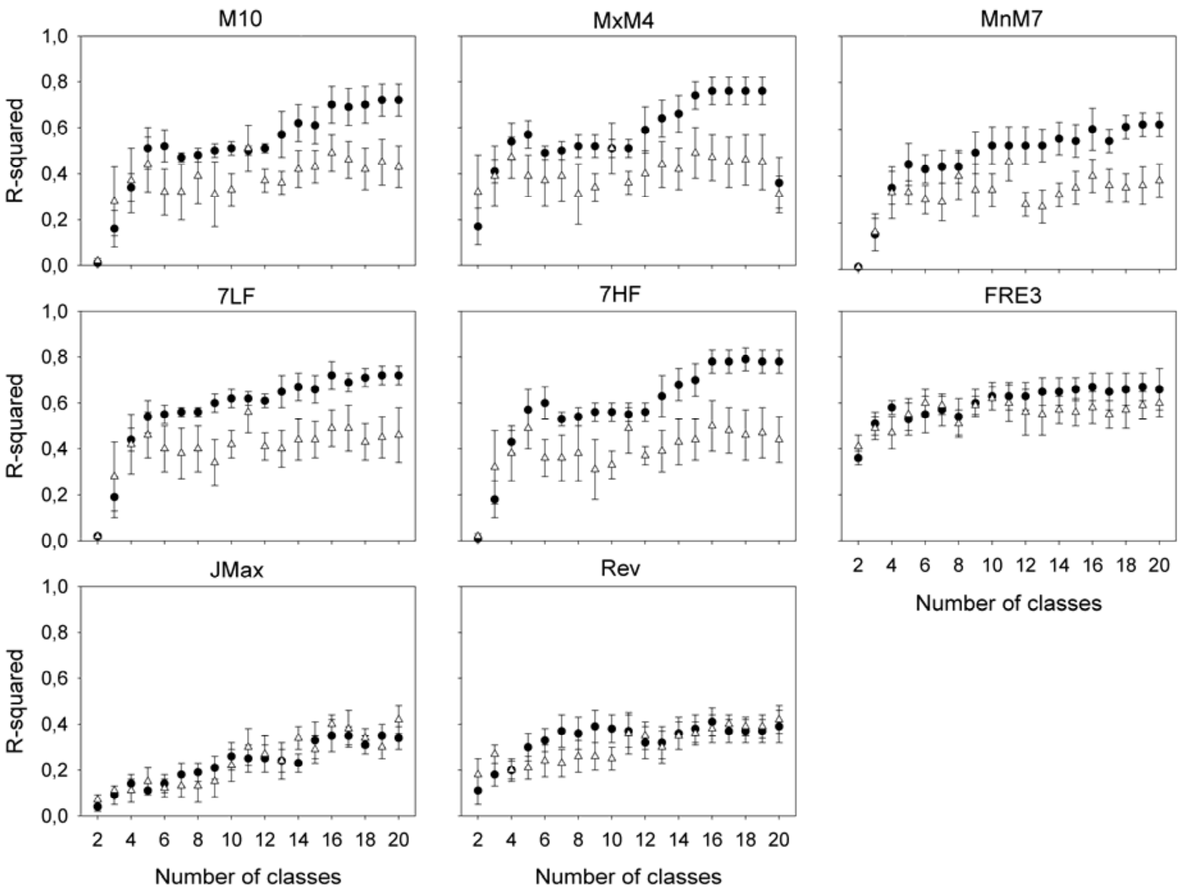

M10
MxM4
MnM7
7LF
7HF
FRE3
JMax
Rev
R-squared
Number of classes
0,0
0,2
0,4
0,6
0,8
1,0
2
4
6
8
10
12
14
16
18
20

Figure 6

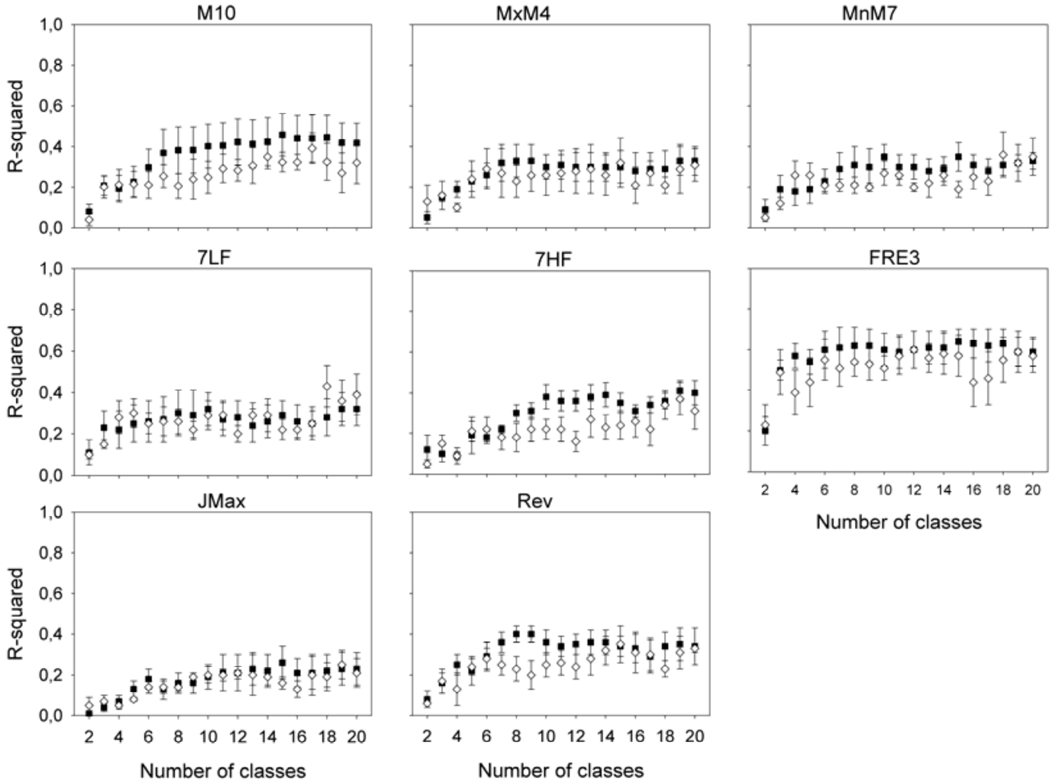
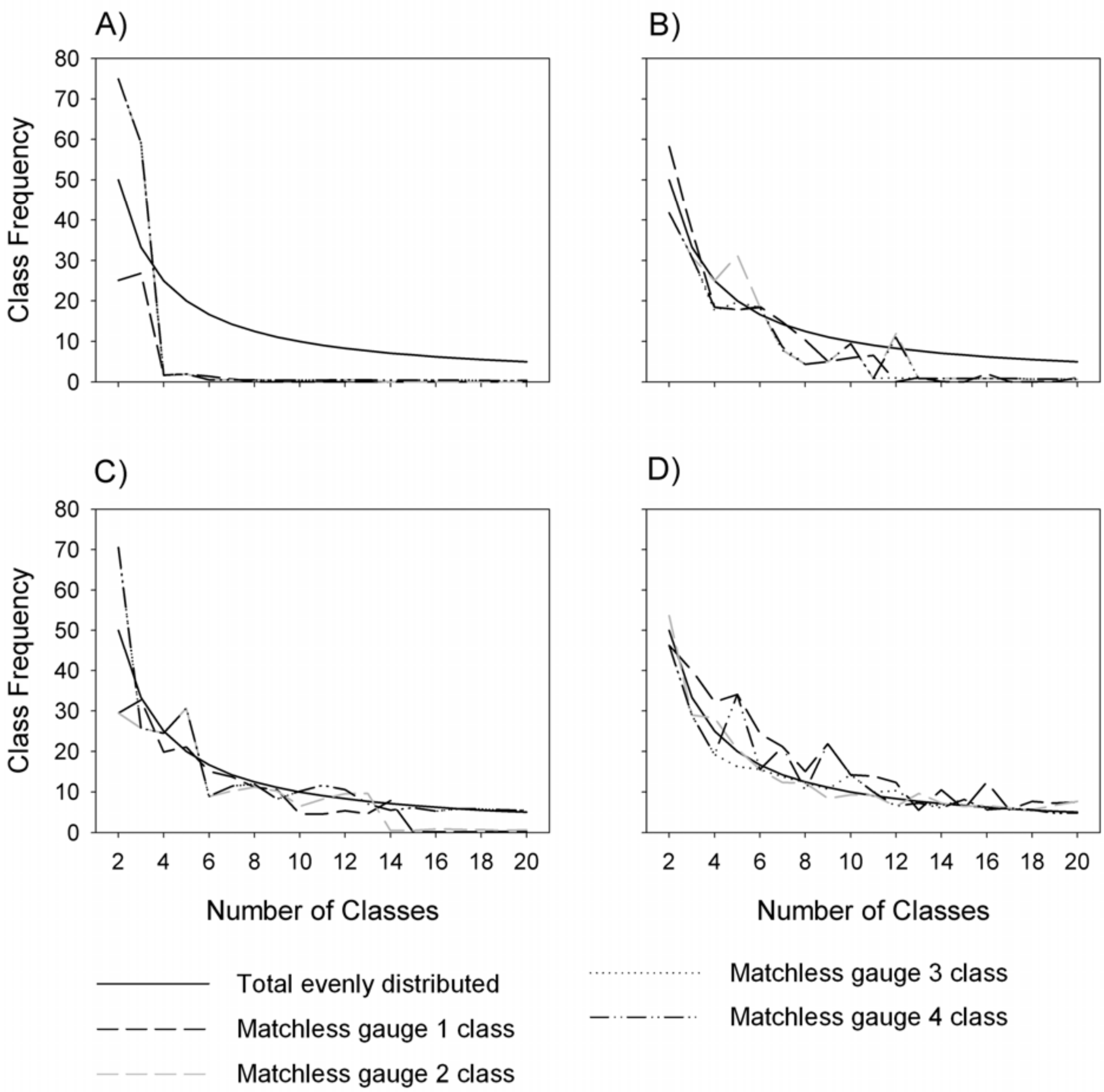

M10
MxM4
MnM7
7LF
7HF
FRE3
JMax
Rev
R-squared
Number of classes
1,0
0,8
0,6
0,4
0,2
0,0
2 4 6 8 10 12 14 16 18 20

Figure 7


A)
B)
C)
D)
Class Frequency
Number of Classes
Total evenly distributed
Matchless gauge 1 class
Matchless gauge 2 class
Matchless gauge 3 class
Matchless gauge 4 class