RIVER RESEARCH AND APPLICATIONS
River Res. Applic. **29**: 1147–1163 (2013)

CLASSIFICATION OF NATURAL FLOW REGIMES IN THE EBRO BASIN (SPAIN) BY USING A WIDE RANGE OF HYDROLOGIC PARAMETERS

M. ALBA SOLANS[a]* AND N. LEROY POFF[b]
[a] *Department of Environment and Soil Sciences, University of Lleida, Catalonia, Spain*
[b] *Biology Department, Colorado State University, Colorado, USA*

ABSTRACT

This paper presents a classification of different natural flow regimes found in Ebro basin, one of the largest in the Mediterranean region. Determination of flow regimes was based on multivariate analyses using long-term discharge series of unaltered flow data. Mean monthly discharges of the 30 'best' flow series and a total of 52 flow series containing unaltered flow data were selected to represent baseline flow conditions for tributaries throughout the basin. Metrics representing magnitude, duration and frequency components of flow were used to identify hydrologic differences across the basin. A total of six natural flow regimes were identified in the Ebro Basin, using a Ward cluster method. The flow patterns identified and their spatial distribution largely corresponded with climatic zones previously reported for the Ebro Basin, with regime types ranging from pluvio-oceanic in the western part of the basin to Mediterranean in the eastern region. Geologic characteristics of the catchment and altitude of headwaters were also found to play an important role in defining flow regime type. A 19-hydrologic variable subset was used to explain main hydrologic differences among groups (such as magnitude and frequency of extreme flow conditions or magnitude and variance of average flow conditions). However, stepwise discriminant analysis was not able to identify consistent subsets of hydrologic variables that adequately identified the six natural flow regime types in this basin. Canonical discriminant analysis was useful to understand class separation and for the interpretation of results.

KEY WORDS: classification; natural flow; hydrology; watershed management

Received 4 May 2011; Revised 11 June 2012; Accepted 20 June 2012

INTRODUCTION

Traditionally, human societies have sought to solve climatic uncertainty and precipitation scarcity by constructing reservoirs that guarantee both permanent and controlled access to fresh water resources. In more modern times, hydropower generation has played an important role in this direction. This policy acquired a special dimension in arid and semiarid regions. For example, Spain contains 2.5% of the world dams, ranking only behind China (46.2%), the USA (13.8%), India (9%) and Japan (5.6%) (World Commission on Dams, 2000). Most dams have been constructed during the second half of the 20th century, and still, an intensification of river regulation at a global scale is expected because of growing human populations and their demand for water, coupled with the uncertainties of anthropogenic climate change (Postel *et al.*, 1996; Postel, 1998; Vörösmarty *et al.*, 2000). However, dam construction alone has not solved water access problems in many countries or satisfied water demands in general; indeed, water conflicts seem to have increased globally (Gleick, 2003; Poff *et al.*, 2003). Dams typically alter the natural flow regime of rivers (e.g. Richter *et al.*, 1996; López-Moreno *et al.*, 2002; Batalla *et al.*, 2004; Poff *et al.*, 2007). Consequently, numerous dam-related impacts on freshwater native biodiversity and on riverine ecological processes have been reported (see Ligon *et al.*, 1995; Stanford *et al.*, 1996; Poff *et al.*, 1997; Kingsford, 2000; Bunn and Arthington, 2002; Poff and Hart, 2002; Bragg *et al.*, 2005, for examples). In addition, there is a growing recognition of goods and services that freshwater ecosystems offer to society (Palmer *et al.*, 2004) and of the high economic costs and questionable efficacy of river restoration projects (Bernhardt *et al.*, 2005).

Within this context, specialists from several disciplines and water managers agree that a sustainable management of water bodies and water resources is needed to maintain and sustain freshwater ecosystems and their services, which, in turn, support human well-being (e.g. Brisbane Declaration (unpublished); http://www.eflownet.org/download_documents/brisbane-declaration-english.pdf). Attempts to manage dams by integrating both ecosystem needs and human needs put some light into this direction (e.g. Palau, 2006; Suen and Eheart, 2006; Batalla and Vericat, 2008; Jacobson and Galat, 2008). Furthermore, some governments have started taking flow-dependent ecosystems into serious consideration when planning water resources management, for example, Australia (Arthington and Pusey, 2003), South Africa (King *et al.*, 2003)

---

*Correspondence to: M. Alba Solans, Department of Environment and Soil Sciences, University of Lleida, Catalonia, Spain.
E-mail: alba.solans@cedex.es

and the USA (NRC, 2007). Current water management frameworks that pursue the principle of sustainability (e.g. Brizga *et al*., 2002; Clarck, 2002; King *et al*., 2003; Richter *et al*., 2003; Arthington *et al*., 2006; Poff *et al*., 2010) rest upon two main premises: the key role of natural hydrologic variability to sustain functionality of freshwater systems, and the holistic view of the management process, where participation of a wide range of stakeholders is fundamental to define and achieve water policy goals.

In this direction, the EU Water Framework Directive (2000/60/CE, hereafter WFD) aims at the regulation of water uses within Europe to promote sustainable water management. The WFD encourages its members to take actions to prevent deterioration of groundwaters, freshwaters and associated ecosystems and to reach a good ecological potential of artificial and heavily modified water bodies and a good ecological status of the rest of surface water bodies by the year 2015. This context provides the framework for the elaboration of River Basin Water Management Plans, like the one that is being developed for Spain's River Ebro (hereafter EBWMP). The current plan (CHE, 2005) recognizes the need to update the 'ecological minimum flow' that is currently applied to rivers (10% of mean annual discharge), and it calls for setting new ecosystem flow standards based on scientific criteria.

The River Ebro has been the focus of several investigations because of an increasing social concern over the river's ecology and its capacity to supply water for human needs. In 1999, a first approach to identify regions physically similar was developed but without taking into account natural flow conditions (Munné and Prat, 1999). Oscoz *et al*. (2007) studied the ecosystem health across the Ebro Basin by using biological indexes; Vericat and Batalla (2006) analysed the river's sediment budget in relation to major reservoirs; Vericat *et al*. (2006) and Batalla *et al*. (2006) examined the morphological and sedimentary adjustments downstream from dams; and Palau *et al*. (2004) and Batalla and Vericat (2008) discussed the river's response to the flushing flows programme implemented in the lower Ebro River since 2002. In relation to freshwater ecosystem sustainability, Alcazar *et al*. (2008) proposed an ecological flow regime for the entire Ebro basin based on a minimum maintenance flow (Q_b parameter) obtained from a moving average forecasting model to increasing intervals of consecutive data (mean daily flows) (see Palau and Alcazar, 1996, for details). Alcazar and Palau (2010) explored the possibility of extrapolating the Q_b parameter across Ebro Basin by using hydrological and watershed features as explanatory variables. Bejarano *et al*. (2010) classified fluvial segments in Ebro basin by using mean monthly flow data from SIMPA model (CEDEX, Ministry of Environment and Public Works, Spain) as classification factors.

The ELOHA framework (i.e. Arthington *et al*., 2006; Poff *et al*., 2010) is the result of a recent consensus among international scientists to develop regional environmental flow standards. The first step of the ELOHA framework is the identification of regions under similar natural flow (baseline) conditions that can work as management units for ecological research and environmental flow guidelines design (Poff *et al*., 1997). Its implementation in the Ebro Basin (NE Iberian Peninsula) contextualizes the present work. Long-term mean monthly flow data of 52 gauges reflecting unaltered streamflows were identified to represent the range of flow variation in the basin. Of these, 30 sites contained particularly high-quality data. All flow data were provided by the Ebro Water Authorities (hereafter CHE). Fifty-four metrics representing magnitude, duration and frequency components of flow were used to identify hydrologic differences across the basin (e.g. Richter *et al*., 1996; Poff *et al*., 1997). Discriminant analysis served to visualize relationships between flow regime classes and to identify the most relevant metrics to separate and define the classes. In addition, variables characterizing watershed attributes and spatial climate variability were developed to examine influence of these factors on observed patterns of flow variability in the basin. A goal of this analysis was to develop a river classification for the Ebro Basin to help inform implementation of the new environmental flow standards aimed at resembling natural flow dynamics.

STUDY AREA

The Ebro River basin is the second largest in the Iberian Peninsula and one of the largest in the Mediterranean region. Draining approximately 85 400 km^2 along the southern-facing slopes of the Cantabrian Range and the Pyrenees, the northern-facing slopes of the Iberian Massif, and the western-facing slopes of the Catalan Ranges, the Ebro River empties into the Mediterranean Sea downstream from Tortosa, 180 km south of Barcelona (Figure 1).

Precipitation. The Ebro Basin receives both Atlantic and Mediterranean climate influences (CHE, 2005). The Atlantic climate is present in the northwestern corner of the basin, the western half of the Pyrenean Range and the northern part of the Iberian Range. The oceanic influence in Atlantic zone generates rain storms throughout the year, with a maximum in December–January and a minimum in July. There is a transition zone to the Mediterranean climate represented by the Western-Central Pyrenees and the central Iberian Range. The Mediterranean climate affects the Catalan Ranges (eastern boundary of the Ebro Depression) and the south-east corner of the Iberian range. Precipitation

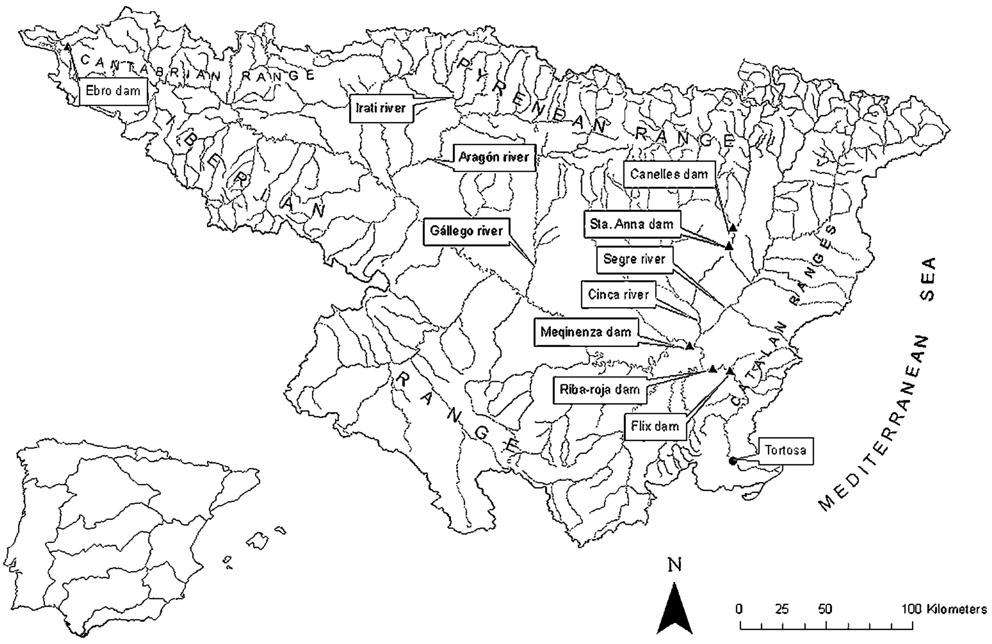

Figure 1. Location of Ebro basin in the Iberian Peninsula, with the main geographical features, tributaries and dams

in this region is characteristic of a typical Mediterranean climate, with maxima in autumn and spring and minima in winter and summer. Finally, the Eastern Pyrenees (Cinca and Segre basins) have a continental Mediterranean climate, with a more regular rainfall pattern, displaying a maximum in spring and a minimum in winter. Mean annual precipitation in the basin is 622 mm, ranging from over 2500 mm in the Pyrenees and the Cantabrian Range, to less than 300 mm in the inner Depression.

Hydrology and flow regulation. A previous, qualitative assessment of seasonal flow patterns at different sites in the Ebro Basin (CHE, 2005) showed differences in natural flow regime, mainly related to precipitation patterns. Streams flowing from the Cantabrian and the Pyrenean ranges (until the Irati basin, for location, see Figure 1) mainly showed a pluvio-oceanic regime, characterized by a perennial, nonseasonal flow with a maximum in winter and a severe minimum during summer. From Irati basin to the East, the flow regime is nivopluvial owing to the snow cover in the Pyrenees during winter and the subsequent snowmelt related runoff. It is characterized by a main flow peak in spring (May–June) and a smaller peak in autumn, with two baseflow periods in winter and summer. On the right side of Ebro valley, the streams located at the north-western part of the basin have a combined oceanic and snowmelt influence, shaping a pluvionival-oceanic regime. As the basin goes to south-east, the oceanic influence vanishes, snow retention disappears and river regimes acquire continental Mediterranean characteristics. The regime in this area is defined as pluvial-Mediterranean, and it is characterized by a marked unstable flow, with a maximum in autumn, the module of which barely go over 1.5, and a severe dry period in summer (July–August). In spring, there is a secondary maximum that normally has two peaks, first in February–March and second in May–June. Finally, EBWMP describes a strong equinoctial character in the most oriental streams.

Water yield follows the heterogeneous climatic and hydrological characteristics present in the basin. Owing to that, 23% of the annual runoff comes from rivers draining the Cantabrian Ranges, 31% from the Western Pyrenees, 41% from the Central-Eastern Pyrenees and 5% from the Iberian Massif and the Catalan Ranges (Batalla *et al.*, 2004). Mean annual discharge at Tortosa (the most downstream gauge in the River Ebro, see Figure 1) is 452 m^3/s.

Fluvial regimes of main tributaries (Segre, Cinca, Gállego and Aragón rivers) and the central and lower reaches of the Ebro mainstem are slightly to moderately altered (MMAa, 2000). The upper reach of the Ebro mainstem, downstream of the Ebro dam, and the Noguera Ribagorçana River downstream Canelles and Santa Ana dams, show the most altered fluvial regimes (see Figure 1). The largest complex of reservoirs is located in the lower reaches of the river and regulates 97% of the catchment area. The complex is formed by Mequinenza (constructed 1966, with a capacity of 1534 hm^3), Riba-roja (1969, 207 hm^3) and Flix dams (1948, 11 hm^3) (Figure 1). According to García and Moreno (2000) and the MMAa (2000) map of potential alteration, only the Ebro headwaters tributaries still have completely unimpaired flow regime.

Geology. Geologically, the Ebro basin has an endorheic origin, and as a consequence, it is formed by evaporitic rocks in most of the central and lower parts of the basin. Large-scale sedimentation during the secondary and the tertiary periods explains the presence of limestones and karstic formations in the Iberian, Catalan and Pyrenean ranges. Conglomerates are also present, mostly located in the eastern and western parts of the Depression. The higher elevations and summits of central and eastern Pyrenees are characterized by granitic batholiths reaching more than 3000 m a.s.l., which are interrupted by karstified carbonate lithologies and metamorphic rocks in the eastern sector (for more details, see http://www.chebro.es).

METHODS

Pre-treatment of data and site selection

Daily streamflow series from 278 gauging stations across the Ebro Basin were obtained from CHE (http://oph.chebro.es), and natural flow series were derived after removing data susceptible of receiving influences from hydraulic infrastructures (e.g. dams, channels, weirs or hydropower facilities). The location of sites with respect to hydraulic infrastructure activities was the main criterion used to identify impacts on flow regime. All cartographic data were obtained from the CHE website (http://oph.chebro.es/ContenidoCartografico.htm). Flow series affected by hydraulic infrastructure were shortened by retaining only the records prior to the start of these flow-altering activities. After that, each series was checked (when possible) with other flow time series from locations upstream or downstream of the same stream branch to check for flow consistency. A total of 89 gauges were found to have reliable flow data not altered by dams or other hydraulic activities.

Time series were transformed into monthly flow data, and a fill in process at monthly scale was carried out. Completed months at daily scale were transformed into average monthly flows, and any non-completed month at daily scale was considered a gap. Two infilling techniques were applied: an intra-site interpolation process up to 2-month gaps, and a month-to-month correlation between similar gauges up to 9-month gaps. The month-to-month correlation technique was applied only when sites were located so close geographically that climatic and watershed conditions were considered the same and a strong correlation between monthly streamflows existed. At each case, the series with more real flow data was retained in the analysis (sites 23 and 86), and its pair was removed to avoid data redundancy. At the end of the infilling process at the monthly scale, years that still had gaps were considered non-completed years and were removed from the dataset because completed years were required to obtain parameter estimates in the classification analysis (see *Selection of variables*)

We limited our analysis to gauges that had data in the period 1946–2002 because this period contained most part of the data. Moreover, some studies have put in doubt the reliability of data at some Ebro gauges before 1940 (MMAb, 2000). Only gauges with a minimum of 20 years of data were selected to dampen the effects of interannual climatic variability and to allow stable estimation of hydrologic statistics (Gan *et al.*, 1991; Richter *et al.*, 1997). Fifty five sites were found suitable in fulfilling both conditions. However, after a preliminary inspection at monthly flow patterns, three of them were removed because of unusual flow patterns: one of them had only data from wet periods (MMAb, 2000), and we could not find any reasonable explanation for the others, except that it was related with data quality. Therefore, as we decided they were unlikely to be part of the population we wished to make inferences about (outliers), we removed them from the analysis (Quinn and Keough, 2002). We assumed that the 52 sites remaining in the sample were representative of most flow variability in Ebro basin, and they were selected to carry out the classification of natural flow regimes. It is important to note that these sites represent high and middle altitude areas (250–1200 m a.s.l.), and therefore, inferences are limited to these zones.

Selection of variables

It is widely accepted that five critical components of flow regime control ecological processes in river systems: magnitude, frequency, duration, timing and rate of change (Poff and Ward, 1989; Richter *et al.*, 1996; Poff *et al.*, 1997). We developed 54 hydrologic variables (Table I) representing magnitude, frequency and duration components of flow, but rate of change and timing were not used because we based the analysis on monthly, not daily, flows. To compare hydrological properties from streams with different catchment sizes, many variables were obtained from modularized flow data by dividing each monthly flow by the grand mean flow of the corresponding gauge (Yevjevich, 1972; Poff and Ward, 1989). Because of the non-common zero-flow phenomenon at study gages and its low duration (normally less than one month), it was decided to study drought events from real data, at daily scale, to have a better description of them. The only variable affected by the change on the scale was yQ0 because the filling in process at monthly scale increased the record length of the series without affecting 0-flow events. Frequency of dry periods can be valued information to discriminate flow regimes in the Mediterranean area, and therefore, we included it, assuming a likely underestimation of yQ0 at some gages.

Table I. Variables used in classification of natural flow regimes (NFR) in Ebro basin. Variables with symbol (*) are the main variables responsible of separating the four types of NFR found with sites of Group 1. Variables with symbol (†) are the main variables responsible for separating the six types of NFR found with sites of Group 1+ Group 2

Variable code	Units	Hydrologic indicator	Definition
Flow magnitude			
Average flow conditions			
M*Y* * †	12	Mean monthly flow for month *Y*	Mean monthly flow for all months
V*Y* * †	12	Variability in monthly flows on month *Y*	Variance of monthly flows for all months
MyV	1	Mean annual volume	Mean of the sum of monthly flows in a hydrologic year
VyV	1	Variability in annual volumes	Variance of the sum of monthly flows in a hydrologic year
Low flow conditions			
MMIN * †	1	Mean minimum monthly flow	Mean of the minimum monthly flows, for the hydrologic year
VMIN	1	Variability in the minimum monthly flows	Variance of the minimum monthly flows
MSMIN	1	Specific minimum monthly flow	Mean of the minimum monthly flows divided by catchment area ($\mathrm{dm}^3/\mathrm{km}^2$)
MQ10	1	Baseflow index 1	Mean monthly flows that do not exceed the 10th percentile from the flow duration curve (flows that are exceeded, at least, the 90% of the time)
MQ25	1	Baseflow index 2	Mean monthly flows that do not exceed the 25th percentile from the flow duration curve (flows that are exceeded the 75%, or more, of the time)
High flow conditions			
MMAX *	1	Mean maximum monthly flow	Mean of the maximum monthly flows, for the hydrologic year
VMAX †	1	Variability in the maximum monthly flows	Variance of the maximum monthly flows
MSMAX †	1	Specific maximum monthly flow	Mean of the maximum monthly flows divided by catchment area ($\mathrm{dm}^3/\mathrm{km}^2$)
MQ75	1	High flow index 1	Mean monthly flows that exceed the 75th percentile from the flow duration curve (flows that are exceeded the 25%, or less, of the time)
MQ90 *	1	High flow index 2	Mean monthly flows that exceed the 90th percentile from the flow duration curve (flows that are exceeded the 10%, or less, of the time)
RMAX_MIN †	1	Variability index	Ratio between MMAX and MMIN
Flow frequency			
Low flow conditions			
FQ5	1	Frequency of low flow spells 1	Total number of low flow spells (low threshold equal to 5% of grand mean monthly flow) divided by the record length in years
FQ10	1	Frequency of low flow spells 2	Total number of low flow spells (low threshold equal to 10% of grand mean monthly flow) divided by the record length in years
yQ0	1	Percentage of years with dry periods	Number of years that occur any 0-flow event (daily scale) divided by the total record length
nQ0y †	1	Frequency of dry periods	Total number of 0-flow events (periods) divided by the number of years with 0-flow events
High flow conditions			
FQ3x	1	Flood frequency 1	Total number of high flow events (upper threshold equal to three times the grand

(*Continues*)

Table I. (Continued)

Variable code	Units	Hydrologic indicator	Definition
			mean monthly flow) divided by the record length in years
FQ5x	1	Flood frequency 2	Total number of high flow events (upper threshold equal to five times the grand mean monthly flow) divided by the record length in years
Flow Duration			
Low flow conditions			
M2MIN	1	Mean annual 2-month minima discharge	Mean magnitude of 2-month-duration minimum annual flow
V2MIN	1	Variability in 2-month minima discharge	Variance of 2-month-duration minimum annual flow
M3MIN	1	Mean annual 3-month minima discharge	Mean magnitude of 3-month-duration minimum annual flow
V3MIN	1	Variability in 3-month minima discharge	Variance of 3-month duration minimum annual flow
belowQ50 †	1	Low flow duration	Maximum number of consecutive months during a year where the mean monthly flow remains below the grand mean monthly flow
MDQ0	1	Mean dry period duration	Mean duration of 0-flow periods (days)
MAXDQ0 *	1	Maximum dry period duration	Maximum duration of 0-flow periods (days)
High flow conditions			
M2MAX †	1	Mean annual 2-month maxima discharge	Mean magnitude of 2-month-duration maximum annual flow
V2MAX †	1	Variability in 2-month maxima discharge	Variance of 2-month-duration maximum annual flow
M3MAX †	1	Mean annual 3-month maxima discharge	Mean magnitude of 3-month-duration maximum annual flow
V3MAX	1	Variability in 3-month maxima discharge	Variance of 3-month-duration maximum annual flow

Cluster analysis

The normality of variables was examined, and some specific transformations were applied to meet assumptions of the analytical methods we employed (see Table II). Where normality could not be achieved, the variable was removed from the analysis. Two cluster analyses were developed, one with the 30 'best' sites, and another with all 52 sites. In the 30-site analysis, VMIN variable was not considered because of its difficulty to reach normality. Variables were standardized by subtracting the mean and dividing by the standard deviation. Afterwards, they were used as inputs in a hierarchical cluster analysis, Ward's method (Ward, 1963), to find groups of similar flow behaviour. A hierarchical tree diagram (or dendrogram) helped to decide the number of clusters. Cluster analysis was developed by MATLAB 7.0.1® software package.

Variable subset selection

We used a *stepwise discriminant analysis* to find a subspace of the 54-dimensional original space that optimizes separation among clusters previously identified and retains as much of the structure of the data as possible (Hand, 1981). Stepwise discriminant analysis is a widely used technique to select variables in classification analyses. It selects variables that maximize separation among groups by selecting step by step those variables with the largest partial F based on MANOVA tests. Variables are added one at a time, and at each step, the variables are re-examined to see if there is redundancy among the new set of variables (see Rencher, 2002, for details).

Stepwise discriminant analysis assumes multivariate normality and a common covariance matrix. With regard to the first condition, normality of any variable was checked by means of Q-Q plots, and also bivariate plots of each pair of variables were studied to detect curved trend and outliers. With regard to the second condition, a test of the null hypothesis of no difference in the multivariate dispersions of the six groups, using deviations from spatial medians and 9999 permutations of residuals suggested there was no convincing evidence for difference ($p = 0.15$). This test is a generalization of Levene's test applied in a multivariate context (Anderson, 2005). In contrast to traditional likelihood-based tests of homogeneity of variance–covariance matrices (e.g. Box's M test), the distance-based test for

Table II. Transformation functions applied to the variables, depending on the sample size

30 Sites		52 Sites	
Variable code	**Transformation function**	**Variable code**	**Transformation function**
MJAN	None	**MJAN**	None
MFEB	None	**MFEB**	None
MMAR	None	**MMAR**	x^2
MAPR	Ln (x)	**MAPR**	Ln (x)
MMAI	Ln (x)	**MMAI**	Ln (x)
MJUN	Ln (x)	**MJUN**	Ln (x)
MJUL	Ln (x)	**MJUL**	$x^{1/3}$
MAUG	Ln (x + 1)	**MAUG**	$x^{1/2}$
MSEP	None	**MSEP**	None
MOCT	None	**MOCT**	None
MNOV	Ln (x + 1)	**MNOV**	None
MDEC	Ln (x)	**MDEC**	Ln (x)
VJAN	Ln (x + 1)	**VJAN**	Ln (x + 1)
VFEB	Ln (x + 1)	**VFEB**	$x^{1/2}$
VMAR	Ln (x)	**VMAR**	$x^{1/2}$
VAPR	Ln (x)	**VAPR**	Ln (x)
VMAI	Ln (x)	**VMAI**	Ln (x)
VJUN	Ln (x)	**VJUN**	Ln (x)
VJUL	Ln (x)	**VJUL**	Ln (x)
VAUG	Ln (x)	**VAUG**	Ln (x)
VSEP	Ln (x)	**VSEP**	Ln (x)
VOCT	Ln (x)	**VOCT**	Ln (x)
VNOV	Ln (x)	**VNOV**	Ln (x)
VDEC	Ln (x)	**VDEC**	Ln (x)
MyV	Ln (x)	**MyV**	Ln (x)
VyV	Ln (x)	**VyV**	Ln (x)
MMIN	Ln (x + 1)	**MMIN**	$x^{1/2}$
VMIN	-	**VMIN**	Ln (x)
MSMIN	$x^{1/2}$	**MSMIN**	Ln (x)
MQ10	Ln (x)	**MQ10**	$x^{1/3}$
MQ25	$x^{1/2}$	**MQ25**	$x^{1/3}$
MMAX	Ln (x)	**MMAX**	Ln (x)
VMAX	Ln (x)	**VMAX**	Ln (x)
MSMAX	$x^{1/2}$	**MSMAX**	$x^{1/3}$
MQ75	Ln (x + 1)	**MQ75**	Ln (x)
MQ90	Ln (x + 1)	**MQ90**	Ln (x)
RMAX_MIN	Ln (x)	**RMAX_MIN**	Ln (x)
FQ5	$x^{1/2}$	**FQ5**	$x^{1/3}$
FQ10	$x^{1/2}$	**FQ10**	$x^{1/3}$
yQ0	$x^{1/2}$	**yQ0**	$x^{1/3}$
nQ0y	$x^{1/2}$	**nQ0y**	$x^{1/2}$
FQ3x	$x^{1/2}$	**FQ3x**	$x^{1/2}$
FQ5x	$x^{1/2}$	**FQ5x**	$x^{1/2}$
M2MIN	None	**M2MIN**	$x^{1/2}$
V2MIN	$x^{1/2}$	**V2MIN**	$x^{1/3}$
M3MIN	Ln (x)	**M3MIN**	$x^{1/3}$
V3MIN	$x^{1/2}$	**V3MIN**	$x^{1/3}$
belowQ50	Ln (x)	**belowQ50**	$x^{1/2}$
MDQ0	$x^{1/2}$	**MDQ0**	$x^{1/3}$
MAXDQ0	$x^{1/2}$	**MAXDQ0**	$x^{1/4}$
M2MAX	Ln (x + 1)	**M2MAX**	Ln (x)
V2MAX	Ln (x)	**V2MAX**	Ln (x)
M3MAX	Ln (x)	**M3MAX**	None
V3MAX	Ln (x)	**V3MAX**	Ln (x)

homogeneity of multivariate dispersions is robust to deviations from multivariate normality. However, this test is not sensitive to detect differences in correlation structure among groups. But, considering the regional scale of the study and the hydrological nature of all the variables, we did assume the common population correlation between groups. Although stepwise discriminant analysis is a popular technique used to reduce the complexity of a group of variables, it is advised to be taken by caution and carefully combined with cross-validation, like any other variable selection procedure (Manly, 1994). Stepwise discriminant analysis was carried out by means of SAS 9.1® statistical software package (i.e. SAS Function Code: PROC STEPDISC).

Validation

Cluster analysis was checked by cross-validation using the selected variable subset. Cross-validation consists of removing one observation at each step and using the remaining (n-1) observations to determine the discriminant (or classification) function. The estimated discriminant function is then used to classify the omitted observation. Under the assumption of a common population covariance matrix between groups, the discriminant functions are linear functions. Although normality is not required, when populations are normal with equal covariance matrices, discriminant analysis is asymptotically optimal, that is, the probability of misclassification is minimized and approaches optimality as the sample size increases (Rencher, 2002). This process is repeated (n-1) times so that each observation is left out once. The misclassification error rate obtained from this process is a nearly unbiased estimate of the expected error (Rencher, 2002). The validation process was also carried out by means of the SAS 9.1® statistical software package (i.e. SAS Function Code: PROC DISCRIM).

Canonical discriminant analysis

Canonical discriminant analysis (hereafter CDA), known as *descriptive discriminant analysis*, aims at identifying the relative contribution of the variables previously selected by stepwise discriminant analysis (see *Variable subset selection*) to separation of the groups defined by cluster analysis (see *Cluster analysis*) and finding the optimal plane on which the points can be projected to best illustrate the configuration of the groups (Rencher, 2002). CDA found linear combinations of the selected variables that best separate the k groups of multivariate observations; it is based on canonical correlations between the set of descriptor variables and the set of k grouping variables. These vectors are uncorrelated and have the highest multiple correlations with the groups. Canonical correlation squared is a measure of multivariate association and measures the association strength between the function and the different groups (Rencher, 2002). When variables are correlated, the best measure of variable importance is the correlation between each variable and the canonical discriminant function, known as *loadings* or *structure coefficients* (Quinn and Keough, 2002; Rencher, 2002). These correlations show the individual contribution of each variable to the group separation. CDA was undertaken by SAS 9.1® statistical software package (i.e. SAS Function Code: PROC CANDISC).

RESULTS

Site selection

The 52 sites were split into two groups: Group 1, 30 sites with high-quality data (i.e. data representing the whole study period of 1946–2002), and Group 2, 22 sites with medium-quality data (flow series with larger gaps and not fully representative of the whole study period) (Table III). These 52 sites represent high and middle elevations areas, most of them over 400 m a. s.l. because of the lack of unaltered flow data in the lowest basin zones (i.e. Ebro Depression). As said before, inferences from this study are limited to these areas.

Cluster analysis, variable subset selection and validation

The first goal was to test the hypothesis that different natural flow regimes do exist in Ebro basin. Only the best flow data (Group 1) were used in this preliminary analysis to attain the most robust classification. Cluster analysis with Group 1 flow series and 53 classificatory variables identified four main hydrologic regimes: A, B, C and D (see Figure 2). Stepwise discriminant analysis highlighted 15 variables responsible for maximizing separation among clusters. These variables, some of which had been transformed (see Table II), were as follows: flow means (*MFEB, MAPR, MMAI, MSEP* and *MDEC*), variability of flow means (*VJAN, VMAR, VAPR, VMAI, VSEP*, and *VDEC*), extreme flows (*MMIN, MMAX* and *MQ90*) and extreme drought conditions (*MAXDQ0*) (see Table I (*) for variable details). Cross-validation based on the 15-variable subset yielded highly robust clusters, with a misclassification rate of only 3.3% (1/30).

Once the existence of different natural flow regimes in Ebro basin was demonstrated, the second goal was to test

Table III. Ebro sites used in classification of NFR in Ebro basin, with the period and number of years of any flow time series

Group 1				**Group 2**			
Name	**ID gauge no.**	**No. of years**	**Period**	**Name**	**ID gauge no.**	**No. of years**	**Period**
Navarrete	41	54	47–02	Oña	93	40	60–02
Urroz de Villa	79	54	47–00	Nuévalos	8	37	65–01
Marañón	6	54	48–01	Muez	151	36	62–98
Aspurz	64	54	46–02	Pitarque	88	35	64–00
Liédena	65	53	46–00	Lecina de Barcabo	46	34	66–01
Yanguas	44	53	47–01	Javierregay	61	34	57–02
Sigues	63	52	51–02	Alins	135	34	66–02
Barasoain	86	52	48–00	Alcaine	127	34	64–02
Beceite	110	52	48–00	Horta San Juan	153	33	66–02
Puigcerdà	21	51	48–02	Tastavins	154	32	69–00
Jubera	58	51	49–00	Biota	155	31	69–00
Cuzcurrita	50	51	50–00	Batea	177	28	75–02
Embid de Ariza	57	50	46–00	San Miguel Pedroso	158	27	70–00
Binies	62	49	49–02	Azarrulla	157	27	66–01
Organyà	111	47	51–00	Tranquera	129	27	73–00
Boltaña	40	47	52–02	Jaraba	56	27	74–00
Olave	67	46	53–00	Torla	196	26	68–93
Santolea	30	46	52–01	Miranda de Ebro	165	26	77–02
Peralta de Alcofea	33	46	48–01	Orón	189	26	77–02
Berantevilla	75	46	50–00	Palazuelos	166	23	66–90
La Seu Urgell	23	45	49–02	Reinosa	178	22	73–94
San Pedro Manrique	43	44	57–00	Jaca	18	21	46–66
Las Cellas	91	44	52–00				
Zuriza	80	44	52–02				
Capella	47	43	56–00				
Los Fayos	90	43	52–96				
Lumbreras	142	42	52–94				
Barbastro	95	41	53–00				
Bergue	100	41	60–00				
Coll Nargó	148	37	58–00				

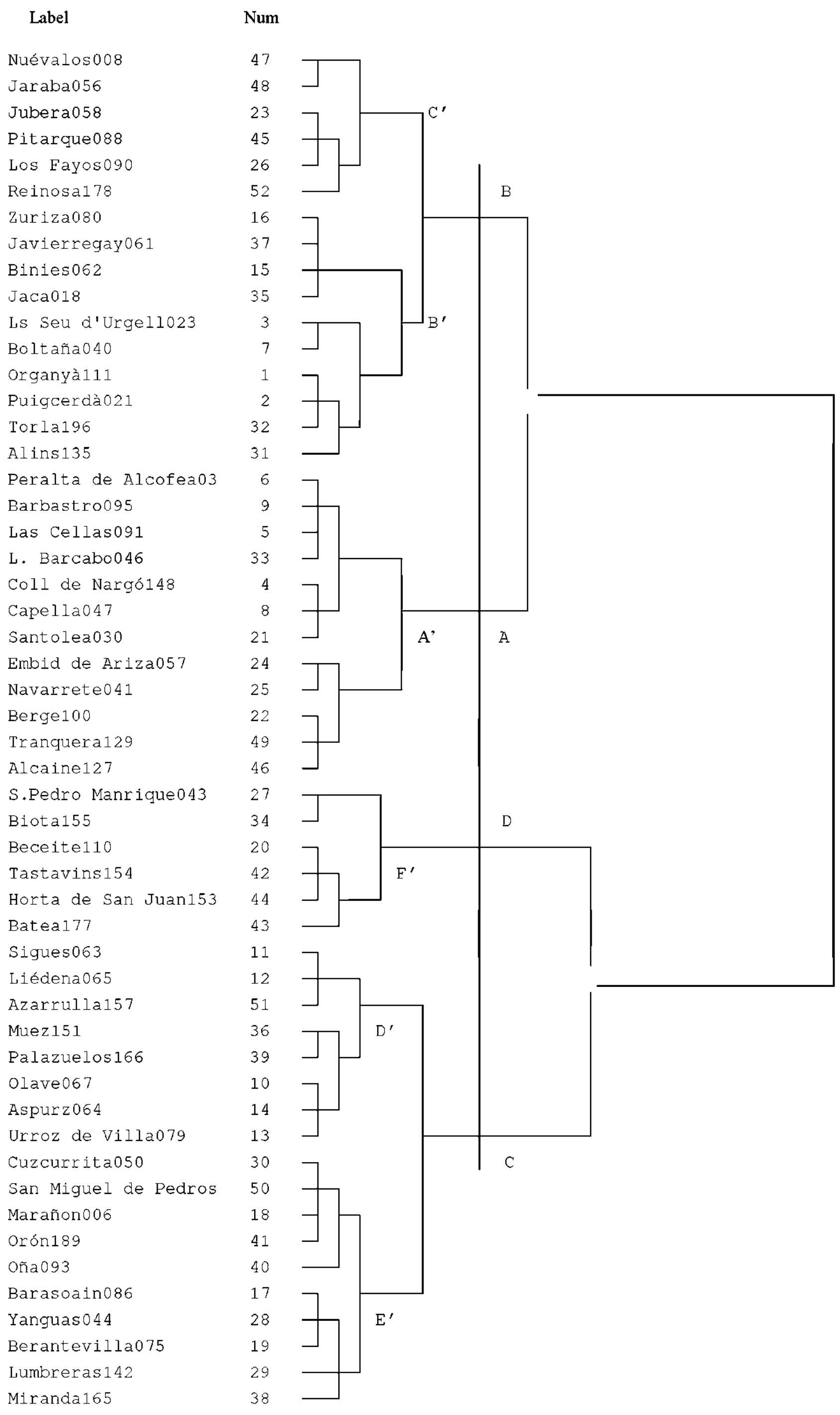

Figure 2. Dendrogram of cluster analysis with 52 gauges and 54 hydrologic variables. Six groups were identified: A', B', C', D', E' and F'. This classification was consistent with a classification obtained with the 30 'best' gauges and 53 hydrologic variables (A, B, C and D groups)

the consistency of flow groups by increasing the number of flow sites. A second classification using all available flow series (Group 1 + Group 2) was sought. The comparison of two classifications served, in addition, to test whether non-homogeneous data (from Group 2) would affect cluster composition. A cluster analysis with 52 observations and 54 hydrologic variables (here, including VMIN) yielded a similar cluster composition than the previous analysis. The difference between both classifications was the number of clusters indentified. The larger number of sites allowed a more precise classification with more specific subclusters: A', B', C', D', E' and F' (Figure 2). Differences between subclusters corresponded to different geographical settings, that is, to differences in climate and/or watershed characteristics, as well as to differences in intra-annual flow variability. We used the six-group classification to assess the distribution of natural flow regimes across the Ebro Basin. A combination of 19 variables was selected by means of stepwise discriminant analysis that maximized differences among the six groups. These transformed variables included average flow conditions, regarding magnitude (*MMAR, MMAI, MJUN, MJUL* and *MNOV*) and variability (*VMAR, VAPR, VMAI, VAUG* and *VOCT*) and extreme flow conditions, regarding magnitude (*MMIN, MSMAX* and *RMAX_MIN*), variance (*VMAX)*, frequency (*nQ0y*), duration (below Q50, *M2MAX* and *M3MAX*) and variability of duration (*V2MAX*). Cross-validation with the 19-variable subset showed an acceptable cluster stability with a misclassification rate of 7.7% (4/52). Cluster distribution is shown in Figure 3, and the annual flow patterns of 52 gauges clustered by groups are depicted in Figure 4. In general, there is good concordance among flow patterns within each cluster. Only cluster F' shows relatively high scatter, a fact that is related to sites 43 and 155, which are geographically distant from the rest of F' sites (see section *5.1*).

Canonical discriminant analysis

The separation of six clusters using the 19 variables was statistically significant ($p < 0.0001$) according to MANOVA tests (Wilk's Lambda, Pillai's Trace, Hotelling–Lawley and Roy). Squared canonical correlations of first and second components were 0.98 and 0.94, respectively, indicating strong associations among individual flow regimes within the clusters. Wilks' test indicated that two first canonical components captured more than 80% of explained group differences, indicating that mean vectors of the six clusters lay mostly in two dimensions (Figure 5). First canonical component accounts for 59.3% of explained flow group differences and highlights differences between D', E' and F' groups with respect to A', B' and C' groups. These differences are mainly explained by the highest correlated variables, which are MJUL (0.69), MMIN (0.71), RMAX_MIN (−0.76), M2MAX (−0.79) and M3MAX (−0.75). Box-plots applied to each significant variable based on transformed data showed that D', E' and F' groups have lower minimum flows (MJUL and MMIN), higher difference between maximum and minimum flows (RMAX_MIN) and higher long-duration maximum flows (M2MAX and M3MAX) than A', B' and C' groups (Figure 6). The second canonical component accounts for 21.6% of explained cluster differences and isolates F' group from other groups because of its higher flow variability in spring [VMAR (0.53), VMAI (0.66)] and autumn [VOCT (0.66)] and its higher high-flow variability [VMAX (0.47) and V2MAX (0.50)] (Figure 6). Box-plots applied to untransformed variables showed the same results than Figure 6. Therefore, it could be implied that

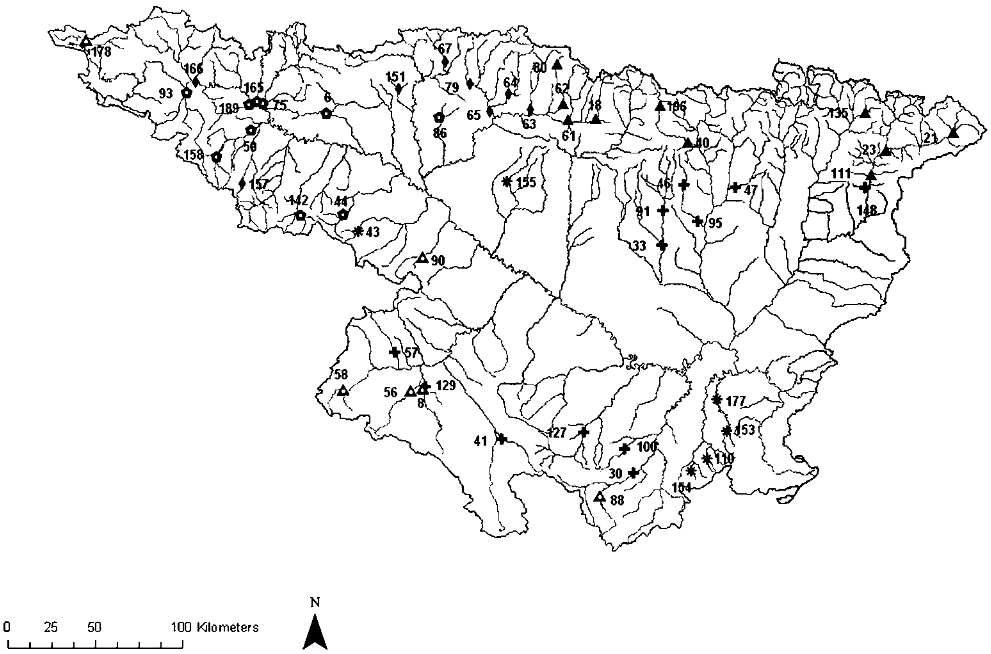

Figure 3. Clustering of the six flow regimes in Ebro basin based on the analysis of 52 gauges and 54 hydrologic variables. Flow regimes are related to the following climatic zones (according to EBWMP): A': Continental Mediterranean-pluvial (✚); B': Nivo-pluvial (▲); C': Continental Mediterranean-pluvial (with a groundwater-dominated flow pattern) (△); D': Pluvio-oceanic (◆); E': Pluvio-nival-oceanic (⬠); and F': Mediterranean (✱)

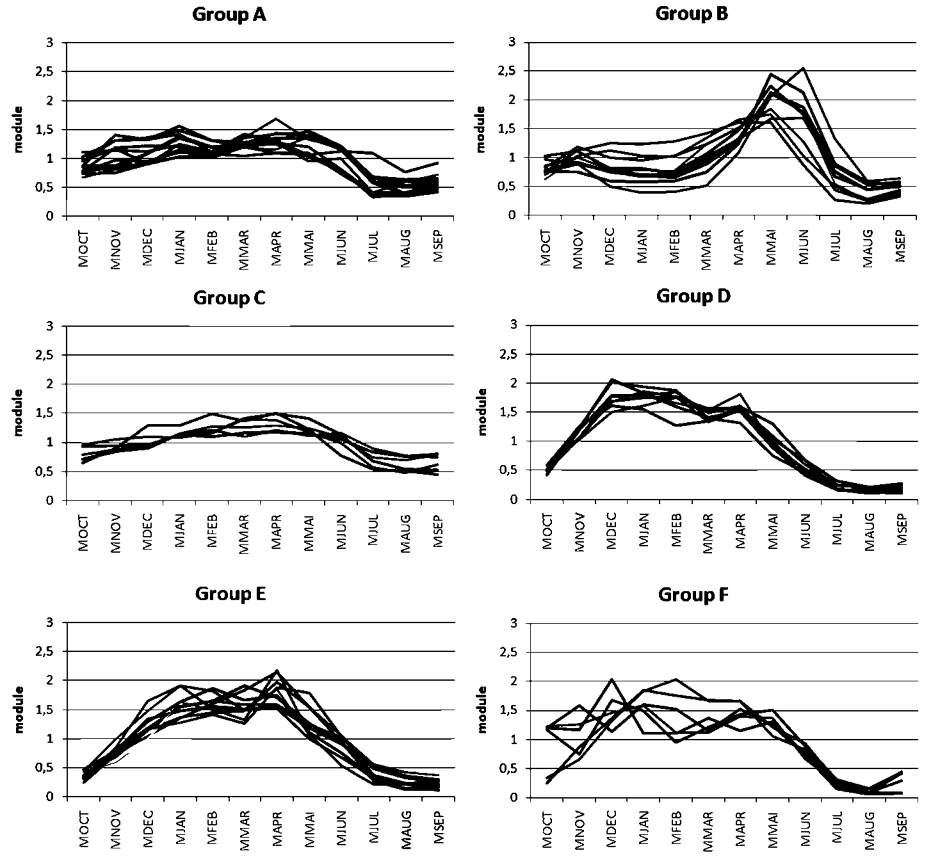

Figure 4. Annual flow patterns of the 52 sites used in classification analysis of NFR in Ebro basin, clustered by groups. The units in vertical axis are flow module ($Q_i/Q_{\text{grand mean}}$)

inferences from transformed variables may also apply to untransformed variables.

Plotting the first canonical components with six flow groups was useful to visualize relationships between groups. In addition, a series of canonical discriminant analyses applied in a stepwise manner helped to identify which aspects of the natural flow regime contributed most to group separation. At each step, only the group (or block of groups) with the major cluster distance with respect to the others was examined (see Figure 5). Once a subset of variables that maximized group differences was found, major loadings were assessed by box-plots, as described before. Cross-validation at different steps showed a high robustness of cluster differences (0% misclassification error for all cases), and canonical components shared a high (>80%) or very high (>95%) canonical correlation squared with the respective clusters in the different analyses.

Assessment of major loadings revealed several differences among the flow regime classes. The F' flow regime has the lowest low-flows (MMIN, MQ10 and M3MIN), the highest maximum flows (MMAX and M2MAX), the highest flashiness (RMAX_MIN) and the highest frequency of low flow spells (FQ5) and dry periods (y_Q0) in the basin. The D' and E' types in the western part of Ebro basin are characterized by lower flows in autumn (MOCT), higher flows in winter (MDEC) and lower minimum flows (MMIN) as well as lower variability of low flows (VJUL and VSEP) than central and eastern parts (A'-B'-C' types). Within the D'-E' flow block, the D' group has lower minimum flows (MAUG), lower variability of low flows (VJUN) and higher flows during autumn (MOCT) and winter (MNOV) than E'. The B' group has higher spring flows (MMAI) and lower variability of long-duration maximum flows (V2MAX) than A' and C' groups. Finally, the A' and C' groups are mainly distinguished primarily because C' has a lower V2MAX than does A'.

DISCUSSION

The Ebro Basin has a great variety of natural flow regime types because it receives precipitation from many climatically distinct sources: the southern slopes of the Pyrenean Range system, the northern slopes of the Iberian Range, the Atlantic Ocean and the Mediterranean Sea. Moreover, part of the Ebro

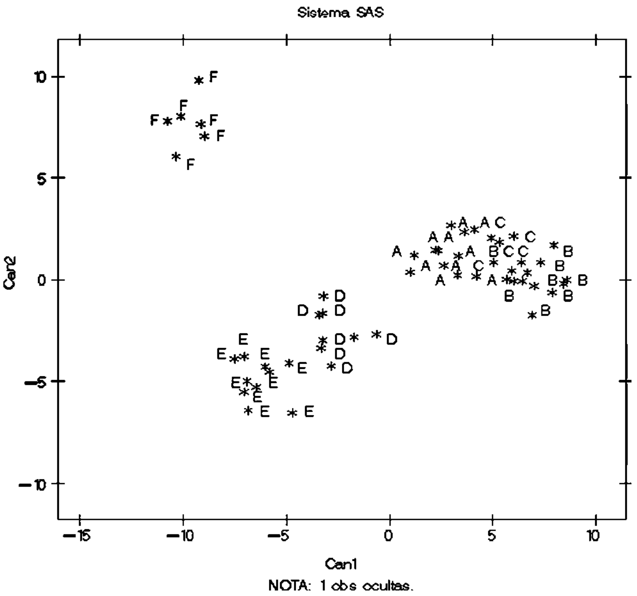

Figure 5. Plot of first and second canonical components of the CDA applied to the six flow groups, based on 19 hydrologic variables. First and second components capture 80.85% of the total explained group variability

Basin is a groundwater-dominated zone with extensive areas on calcareous rocks that control runoff, especially in central and southern sectors of the Iberian Range.

Relating natural flow regimes and climatic and watershed factors

A close link between geographical site location and characteristics of natural flow regime was found (see Figure 3). Flow classes were consistent with flow regime description in EBWMP, and both cluster distribution and annual flow patterns were consistent with climatically defined regions reported in EBWMP (see *Study area*).

According to the present work, clusters D’ and E’ belong to the northwestern sector of the basin that includes streams from the Cantabric and Pyrenean Ranges (until Irati basin) and the northern part of Iberian Range (see Figure 3). Not surprisingly, this distribution fits well with the pluvio-oceanic and the pluvio-nival-oceanic flow regimes defined in CHE (2005). Figure 4 shows D’ and E’ flow regimes with a regular flow from December to April that differ for the peak flow timing, with D’ having a December peak and E’ an April peak. D’ flow regime is mainly located in Western Pyrenees (i.e. west from Irati basin), whereas E’ flow regime is present in the lower parts of Cantabrian Range and the northern sector of the Iberian Range. E’ flow regime likely experiences the effects of snow cover in winter and the corresponding melt in spring that would produce a delay on winter peak. It would correspond to a pluvio-nival-oceanic regime. Moreover, D’ and E’ flow regimes extend over the same area than oceanic climate regime with flow annual patterns emulating the annual precipitation pattern, that is, maximum values in winter and minimum in summer. Therefore, a close link between D’ and E’ flow regimes and oceanic climate regime appears to exist.

B’ flow regime extends through the Pyrenean Range, from Irati basin to the East, and it fits very well with nivo-pluvial streamflow distribution reported in CHE (2005). This regime is characterized by a low winter flow and a high peak on spring because of snow melting and spring precipitations occurring simultaneously. Therefore, these results point to altitude as another key factor responsible for flow variability in Ebro basin. Sites at lower altitudes than B’, with the same hydro-climatic conditions but with no snowfall during winter, have flow peaks directly related to precipitation maxima. These locations correspond to A’ flow regime, and it is also found in central and southern sectors of the Iberian Range. These areas would correspond to the transition zone between Atlantic (i.e. oceanic) and Mediterranean climate regimes, with a continental character in the central parts of the basin (CHE, 2005). The EBWMP defines the flow in these areas as continental Mediterranean pluvial, which is characterized with a consistent baseflow with two maxima, in autumn and in spring, and a severe dry period in summer. However, the A’ flow pattern does not show severe droughts in summer. The presence of soft carbonate rocks in the centre of the Ebro Depression and especially in the central and southern sectors of Iberian Range could provide a plausible physical explanation for such permanent summer flow. Headwaters of the central and southern Iberian Range often have karstic compositions that exert an important control on runoff (CHE, 2005). This phenomenon might explain the presence of a C’ flow regime in these areas, whose low variable groundwater-dominated flow pattern reinforces the idea of the lithological control on river’s hydrology. These hypotheses should be tested by local studies and basin-wide fieldwork.

Finally, the F’ flow pattern is found in the southeastern part of the basin and, agreeing with EBWMP, would correspond to rivers that typically experience a Mediterranean regime with a strong equinoctial character. The F’ annual pattern (Figure 4) has two main peaks, one in early winter and another in spring, and it seems to be linked directly to the character of equinoctial precipitation, which, in this area, has high variability because of alternating droughts and intense rains. This variability on precipitations would explain the extreme flow conditions found in the F’ group (see *Canonical discriminant analysis*. Sites 43 and 155, however, which are located in the centre of Ebro basin, are not affected by Mediterranean climate regime. The fact that both sites are included in the F’ group could be explained, first, because of

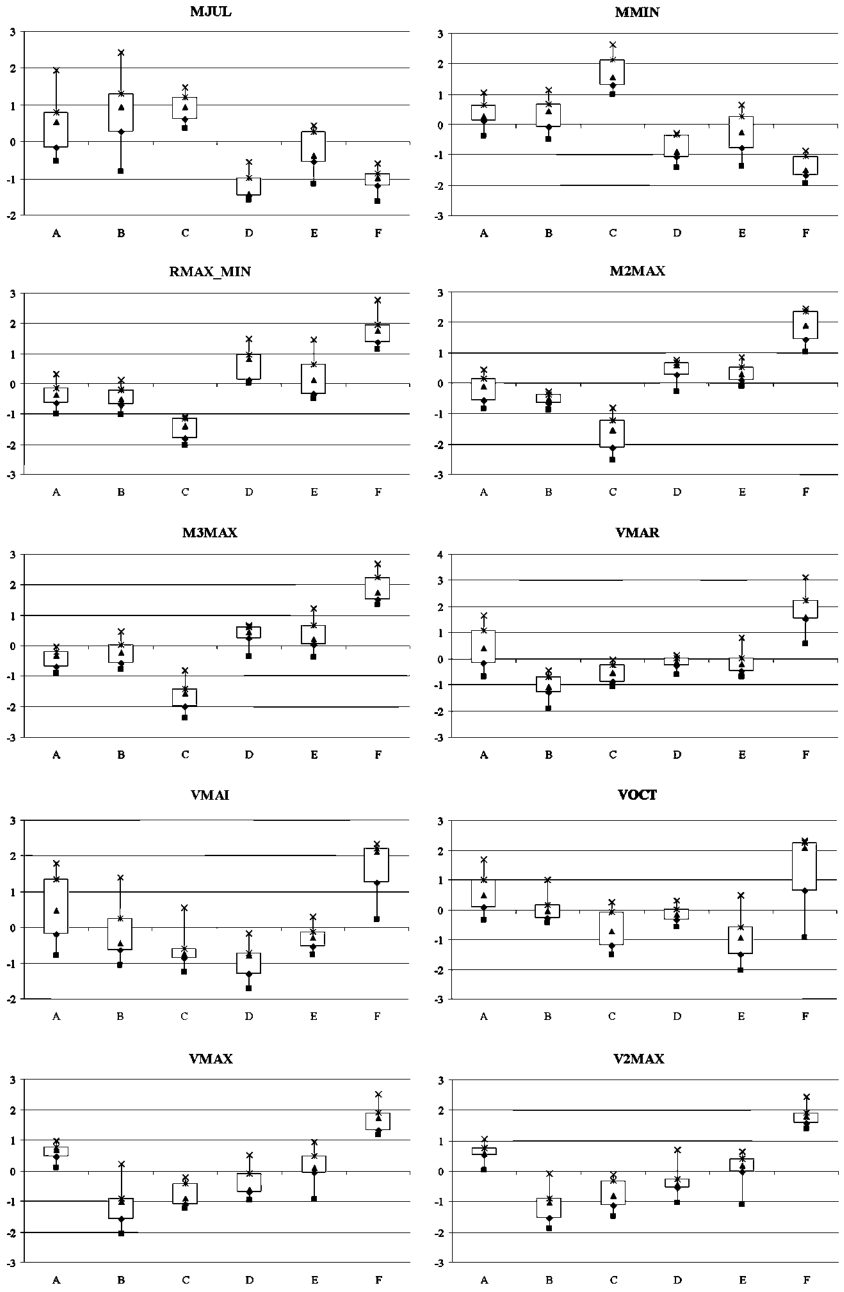

Figure 6. Box-plots of the major loadings of first and second canonical components of the CDA applied to the six flow regime types identified in Ebro basin. The variables are transformed and standirdized. This figure is available in colour online at wileyonlinelibrary.com/journal/rra.

the arid conditions existing in the inner Depression zone, where they are located (see Figure 3). The scarcity of precipitation in this region would produce extreme low flow conditions and recurrent droughts. And second, the small catchment size coupled with an impervious geology could force intense responses of flow after rainstorms, creating high peak flows with short durations.

The consistency of the 52-gauge classification with respect to physical features in the basin indicates that the 54-variable set succeeded in characterizing main flow differences in the Ebro Basin and that, despite the intense pretreatment of data, multivariate analyses demonstrated to be very effective in combining information and identifying differences among flow groups. Its efficiency in processing a large amount of data is the main reason why multivariate analyses have been widely used in regional hydrologic assessments (e.g. Poff, 1996; Chiang *et al.*, 2002; Thoms and Parsons, 2003; Baeza and García de Jalón, 2005; Poff *et al.*, 2006; Sanborn and Bledsoe, 2006; Moliere *et al.*, 2009).

Historical flow data

Cluster analysis of 52 long-term mean monthly series based on 54 hydrologic variables representing magnitude, duration and frequency components of flow allowed us to identify six distinct flow regimes that, overall, are consistent with hydroclimatic and physical differences (i.e. topography, geology) in the basin. This classification was also consistent with a more restricted classification obtained with the 30 best flow series and 53 hydrologic variables.

As these results were obtained using monthly flow data, it means that coarse time step data (average monthly flows) may be sufficient to explain flow variability when this variability is strongly dependent on physical factors (e.g. climate), as occurs in the Ebro Basin. In support of this idea, Poff (1996) assessed the correspondence between daily and monthly flow data for the characterization of flow spates at 12 flow regime types in the continental USA and found a median over 60% of correspondence between both statistics in almost all types. At two flow types, the median correspondence was over 80%.

Furthermore, the fact that consistent results were obtained with non-continuous time series allowed more sites to be incorporated into the analysis. A larger sample permitted a better characterization of central tendencies in the flow regimes. As a consequence, a more accurate classification was obtained. Snelder *et al.* (2009) also used this kind of data to make historical flow assessments in France, a country with similar gauging history to Spain. Long-term continuous, unaltered flow data are, in general, rare, and it has been assumed as a critical problem to assess reference flow conditions (Özçelik and Baykan, 2009).

Finally, the length of time series effectively seems to be an important factor to be taken into account. Analyses not shown here indicated that the inclusion of a third group of gauges with less than 20 years of data did not yield consistent cluster results. This finding agrees with other works that claim a relative homogeneity and length of database to characterize the main attributes of flow regimes (Gan *et al.*, 1991; Richter *et al.*, 1997; Poff *et al.*, 2006).

Variable subset stability

A 19-variable subset succeeded in capturing differences among six flow regimes in the Ebro basin, producing a misclassification error of 7.7%, and a 15-variable subset distinguished broad differences between flow groups in the basin, with a misclassification error of 3.3% (Figure 2). Comparing both variable subsets, only five variables performed consistently, that is, were common to both subsets: MMAI, VMAR, VAPR, VMAI and MMIN. The main difference between the two analyses was in regard to the sample size, with the former comprising 52 sites, and the latter 30 sites. The inconsistency in variable subsets demonstrates some instability in the variable selection process that may be related to multicollinearity (Quinn and Keough, 2002). Hence, the 19-variable subset might not be taken as a reference variable subset to allocate new gauges into Ebro flow classes. The instability of the 19-variable subset was also demonstrated when the subset failed in performing the clusters obtained with the original 54-variable set. In the same direction, some preliminary multicollinearity analyses applied to the 54-variable set and not included here showed that any removal of variable correlation from $|0.65|$ to $|0.8|$ could not consistently reproduce the original clusters either. These results confirm that as long as the correlation between two variables is less than 1, it is not possible to reduce the number of variables without some loss of information.

However, the low misclassification error (7.7%) obtained with the 19-variable subset justified its use in posterior multivariate analyses because a reduction in variable redundancy and multicollinearity meant an improvement in the robustness of MANOVA test statistics and parameter estimates of linear discriminant functions (Johnson and Field, 1993; Quinn and Keough, 2002).

Stepwise selection method has been criticized for its tendency to exclude some significant variables and to include some variables that are weak discriminators. However, it gives good guidance in the absence of other alternatives and should be always combined with a cross-validation process (Manly, 1994). These results confirm that variable selection may be useful to develop statistical analyses when multicorrelation may be a problem to obtain robust statistical results.

Annual flow patterns

Results found in the present work demonstrate that a visual assessment of annual flow patterns alone is not sufficient to develop a flow regime classification, but it can complement multivariate analyses, for example, by helping validate gauge association and making initial interpretations and formulating hypotheses.

For example, it was found that, although A' and C' groups have similar annual flow patterns, an adequate selection of flow metrics with ecological meaning allowed meaningful divergences between them to be related to different geographical location and geologic setting. Conversely, annual flow patterns in the F' group were not consistent, but they showed similar low/high flows and flashiness properties.

Assessment of annual flow patterns allowed us to make a visual validation of flow matching among groups that helped interpret the reliability of results. For example, from the original 55-site sample, it was useful to detect unexpected flow responses with respect to sample records. This allowed us to remove three gauges from the sample that otherwise could have added noise to the analyses and led to a misinterpretation of results.

Coupling mean annual flow pattern and geographical location allowed deriving preliminary relationships between flow regime characteristics and climate, altitude and geologic features. For example, a close link between E'-D' flow regimes and oceanic climate regime was found. In addition, it seemed evident that the F' flow regime was a direct response to Mediterranean climate conditions and that A', B' and C' groups were located in climate transition zones. As the mentioned groups (F', E'-D' and A'-B'-C') represent major variability in the basin (> 80%), it was concluded that climate is the major factor responsible of flow variability in the Ebro Basin. Other flow differences, accounting approximately for the 20% of group variance, seem to be related with site altitude and geology factors. Mesoscale studies at the subbasin level and fieldwork are, however, needed to confirm these hypotheses. A similar conclusion was arrived at by Bejarano *et al.* (2010) using simulated flow data from the Ebro Basin. They found that natural intra-annual flow variability in the basin responded first to ecoregion and rainfall patterns, second to geological nature of catchments and size, and third to elevation and slope. Also, other studies made in other contexts around the world report similar findings. For example, Jowet and Duncan (1990) found that climate accounted for a broad regional distribution of flow regime types in New Zealand. Chiang *et al.* (2002) showed that similar hydrologic responses were identified by watershed variables and precipitation, and Snelder *et al.* (2009) mapped a natural flow regime classification in France based on the proposition that watersheds having similar topography, superficial geology and climate would result in similar hydrologic regimes, regardless of geographical location.

Recently, some authors have advocated for the classification of natural flow regimes as an essential step previous to any regional water resource plan or/and restoration program with the purpose to further transform stream classification into ecologically meaningful groups (Hughes *et al.*, 1986; Poff and Ward, 1989; Biggs *et al.*, 1990; Bonada *et al.*, 2002; Thoms and Parsons, 2003; Baeza and García de Jalón, 2005; Arthington *et al.*, 2006; González del Tánago and García de Jalón, 2006; Sanborn and Bledsoe, 2006; Moliere *et al.*, 2009; Snelder *et al.*, 2009; Bejarano *et al.*, 2010; Poff *et al.*, 2010). Many examples of stream classification can be found in the literature, applied at different spatial scales (from ecoregion to channel reach) and using a wide range of environmental variables (e.g. topographic, climatic, hydrologic, geologic, vegetative or chemical). Here, we offer an example of flow classification in a middle size basin from the Mediterranean region using natural streamflow attributes as classification variables. Flow data, when available, are the best guarantee to achieve a reliable characterization of historical flow regime. Stream classification would help in the identification and characterization of management units, within which environmental flow guidelines could be implemented. This approach has been accepted as the first step to set environmental flow standards at regional scale (Arthington *et al.*, 2006; Poff *et al.*, 2010). Providing plausible explanations for flow variation across the basin in terms of climate, altitude and geology, our work adds even more consistency to an assessment based on robust statistical tools. The hydroclimatic relation to the classification can be helpful to the Ebro River Basin Management Plan that is currently being elaborated.

ACKNOWLEDGEMENTS

The first author wants to thank Ebro's Authorities and especially Miguel Ángel García Vera for his support and sound advice during the development of the study. Also, thanks to Prof. José Salas of Colorado State University for his advice about statistical infilling techniques and stochastic modelling. Thanks to Andrés Mellado Díaz and Jorge Rubén Sánchez González from CEDEX for their advice in multivariate statistics. Also, thanks to Ramon Batalla from the University of Lleida and two anonymous referees for improving the quality of the paper with their comments to earlier versions. Part of this research was carried out while the first author was a PhD student granted by the Spanish Ministry of Education and Sciences under the framework of the Research Project REN2001-0840-C02-01/HID (2002–2005).

REFERENCES

Alcazar J, Palau A. 2010. Establishing environmental flow regimes in a Mediterranean watershed based on a regional classification. *Journal of Hydrology* **388**: 41–51.

Alcazar J, Palau A, Vega-García C. 2008. A neural net model for environmental flow estimation at the Ebro River basin, Spain. *Journal of Hydrology* **349**: 44–55.

Anderson MJ. 2005. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* **62**: 245–253.

Arthington AH, Pusey BJ. 2003. Flow restoration and protection in Australian rivers. *River Research and Applications* **19**: 377–395.

Arthington AH, Bunn SE, Poff NL, Naiman RJ. 2006. The challenge of providing environmental flow rules to sustain river ecosystems. *Ecological Applications* **16**(4): 1311–1318.

Baeza D, García de Jalón D. 2005. Characterization of streamflow regimes in central Spain, based on relevant hydrobiological parameters. *Journal of Hydrology* **310**: 266–279.

Batalla RJ, Vericat D. 2008. Hydrological and sediment transport dynamics of flushing flows: Implications for management in large mediterranean rivers. *River Research and Applications* **25**(3): 297–314.

Batalla RJ, Gómez CM, Kondolf GM. 2004. Reservoir-induced hydrological changes in the Ebro River basin (NE Spain). *Journal of Hydrology* **290**: 117–136.

Batalla RJ, Vericat D, Palau A. 2006. Sediment transport during a flushing flow in the Lower Ebro River. In *Sediment dynamics and the hydromorphology of fluvial system*, Rowan JS, Duck RW, Werritty A (eds). IAHS Publications: Wallingford; 37–44.

Bejarano MD, Marchamalo M, García de Jalón D, González del Tánago M. 2010. Flow regime patterns and their controlling factors in the Ebro basin (Spain). *Journal of Hydrology* **385**: 323–335.

Bernhardt ES, Palmer MA, Allan JD, Alexander G, Barnas K, Brooks S, Carr J, Clayton S, Dahm C, Follstad-Shah J, Galat D, Gloss S, Goodwin P, Hart D, Hassett B, Jenkinson R, Katz S, Kondolf GM, Lake PS, Lave R, Meyer JL, O'Donnell TK, Pagano L, Powell B, Sudduth E. 2005. Synthesizing U.S. river restoration efforts. *Science* **308**: 636–637.

Biggs BJF, Duncan MJ, Jowett IG, Quinn JM, Hickey CW, Davies-Colley RJ, Close ME. 1990. Ecological characterisation, classification, and modelling of New Zealand rivers: an introduction and synthesis. *New Zealand Journal of Marine and Freshwater Research* **24**: 277–304.

Bonada N, Prat N, Munné A, Rieradevall M, Alba-Tercedor J, Àlvarez M, Avilés J, Casas J, Jáimez-Cuellar P, Mellado A, Moyà G, Pardo I, Robles S, Ramon G, Suárez ML, Toro M, Vidal-Abarca MR, Vivas S, Zamora-Muñoz C. 2002. Ensayo de una tipología de las cuencas mediterráneas del proyecto GUADALMED siguiendo las directrices de la directiva marco del agua. *Limnetica* **21**(3–4): 77–99.

Bragg OM, Black AR, Duck RW, Rowan JS. 2005. Approaching the physical-biological interface in rivers: a review of methods for ecological evaluation of flow regimes. *Progress in Physical Geography* **29**(4): 506–531.

Brizga SO, Arthington AH, Choy SJ, Kennard MJ, Mackay SJ, Pusey BJ, Werren GL. 2002. Benchmarking, a "top-down" methodology for assessing environmental flows in Australian rivers. Environental Flows for River Systems Working Conference and 4th International Ecohydraulics Symposium. Cape Town, South Africa.

Bunn SE, Arthington AH. 2002. Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. *Environmental Management* **30**(4): 492–507.

Chiang SM, Tsay TK, Nix SJ. 2002. Hydrologic regionalization of watersheds. I: Methodology development. *Journal of Water Resources Planning and Manangement* **128**(1): 3–11.

Clarck MJ. 2002. Dealing with uncertainty: adaptive approaches to sustainable river management. *Aquatic Conservation: Marine and Freshwater Ecosystems* **12**: 347–363.

Confederación Hidrográfica del Ebro (CHE). 2005. Caracterización de la demarcación y registro de zonas protegidas. Ministerio de Medio Ambiente y Medio Rural y Marino.

Gan KG, McMahon TA, Finlayson BL. 1991. Analysis of periodicity in streamflow and rainfall data by Colwell's indices. *Journal of Hydrology* **123**: 105–118.

García MA, Moreno MC. 2000. Los aprovechamientos en la Cuenca del Ebro: Afección en el régimen hidrológico fluvial. Internal Technical Report n° 2000-PH-24.1, Confederaciónn Hidrográfica del Ebro, Zaragoza.

Gleick PH. 2003. Global freshwater resources: soft-path solutions for the 21st century. *Science* **302**: 1524–1528.

González del Tánago M, García de Jalón D. 2006. Caracterización jerárquica de los ríos españoles. Propuesta de tipología de tramos fluviales para su clasificación atendiendo a la directiva marco del agua. *Limnetica* **25**(3–4): 81–98.

Hand DJ. 1981. *Discrimination and Classification*. John Wiley & Sons: USA.

Hughes RM, Larsen DP, Omernik, JM. 1986. Regional reference sites: a method for assessing stream potentials. *Environmental Management* **10**(5): 629–635.

Jacobson RB, Galat DL. 2008. Design of a naturalized flow regime-an example from the Lower Missouri River, USA. *Ecohydrology* **1**: 81–104.

Johnson CR, Field CA. 1993. Using fixed-effects model multivariate analysis of variance in marine biology and ecology. *Oceanography and Marine Biology Annual Review* **31**: 177–221.

Jowet IG, Duncan MJ. 1990. Flow variability in New Zealand rivers and its relationship to in-stream habitat and biota. *New Zealand Journal of Marine and Freshwater Research* **24**: 305–317.

King J, Brown C, Sabet H. 2003. A scenario-based holistic approach to environmental flow assessments for rivers. *River Research and Applications* **19**: 619–639.

Kingsford RT. 2000. Ecological impacts of dams, water diversions and river management on floodplain wetlands in Australia. *Austral Ecology* **25**: 109–127.

Ligon FK, Dietrich WE, Trush WJ. 1995. Downstream ecological effects of dams. A geomorphic perspective. *Bioscience* **45**(3): 183–192.

López-Moreno JI, Beguería S, García-Ruiz JM. 2002. Influence of the Yesa reservoir on floods of the Aragón River, central Spanish Pyrenees. *Hydrology and Earth System Sciences* **6**(4): 753–762.

Manly BFJ. 1994. *Multivariate Statistical Methods*, 2ndedn . Chapman & Hall: London, UK.

Ministerio de Medio Ambiente (MMAa). 2000. La situación actual y los problemas existentes y previsibles. In *Libro Blanco del Agua en España*. Ministerio de Medio Ambiente: Madrid. Chapter 3: 73–505.

Ministerio de Medio Ambiente (MMAb). 2000. Análisis de los sistemas hidráulicos. *Plan Hidrológico Nacional* **3**.

Moliere DR, Lowry JBC, Humphrey CL. 2009. Classifying the flow regime of data-limited streams in the wet-dry tropical region of Australia. *Journal of Hydrology* **367**: 1–13.

Munné A, Prat N. 1999. *Regionalización de la cuenca del Ebro para el establecimiento de los objetivos del estado ecológico de sus ríos*. Informe para la Confederación Hidrográfica del Ebro (Oficina de Planificación Hidrológica): Zaragoza.

National Research Council (NRC). 2007. *River Science at the U.S. Geologic Survey*. The National Academies Press: Washington, DC.

Oscoz J, Gomà J, Ector L, Cambra J, Pardos M, Durán C. 2007. Estudio comparativo del estado ecológico de los ríos de la cuenca del Ebro mediante macroinvertebrados y diatomeas. *Limnetica* **26**(1): 143–158.

Özçelik C, Baykan NO. 2009. An improved time series model for monthly stream flows. *Stochatic Environmental Research and Risk Assessment* **23**: 587–601.

Palau A. 2006. Integrated environment management of current reservoirs and regulated rivers. *Limnetica* **25**(1–2): 287–302.

Palau A, Alcazar J. 1996. The basic flow: an alternative approach to calculate minimum environmental instream flows. Ecohydraulics 2000. Proceedings 2nd International Symposium on Habitat Hydraulics, v A. Quebec, Canada; 547–558.

Palau A, Batalla RJ, Rosico E, Meseguer A, Vericat D. 2004. Management of water level and design of flushing floods for environmental river maintenance downstream of the Riba-Roja Reservoir (lower Ebro River, NE Spain). Proceedings of the International Conference HYDRO 2004: A new era for hydropower. Porto, Portugal.

Palmer M, Bernhardt ES, Chornesky E, Collins S, Dobson A, Duke C, Gold B, Jacobson R, Kingsland S, Kranz R, Mappin M, Martinez ML, Micheli F, Morse J, Pace M, Pascual M, Palumbi S, Reichman OJ, Simons A, Townsend A, Turner M. 2004. Ecology for a crowded planet. *Science* **304**: 1251–1252.

Poff NL. 1996. A hydrogeography of unregulated streams in the United States and an examination of scale-dependence in some hydrological descriptors. *Freshwater Biology* **36**: 71–91.

Poff NL, Hart D. 2002. How dams vary and why it matters for the emerging science of dam removal. *Bioscience* **52**(8): 659–668.

Poff NL, Ward JV. 1989. Implications of streamflow variability and predictability for lotic community structure: a regional analysis of streamflow patterns. *Canadian Journal of Fisheries and Aquatic Sciences* **46**(10): 1805–1818.

Poff NL, Allan JD, Bain MB, Karr JR, Prestegaard KL, Richter BD, Sparks RE, Stromberg JC. 1997. The natural flow regime. A paradigm for river conservation and restoration. *Bioscience* **47**(11): 769–784.

Poff NL, Allan JD, Palmer MA, Hart DD, Richter BD, Arthington AH, Rogers KH, Meyer JL, Stanford JA. 2003. River flows and water wars: emerging science for environmental decision making. *Frontiers in Ecology and the Environment* **1**(6): 298–306.

Poff NL, Olden JD, Merritt DM, Pepin DM. 2007. Homogenization of regional river dynamics by dams and global biodiversity implications. In *National Academy of Sciences of the USA*, Mooney HA (ed.). Stanford University: Stanford; 104[14], 5732–5737.

Poff NL, Olden JD, Pepin DM, Bledsoe BP. 2006. Placing global stream flow variability in geographic and geomorphic contexts. *River Research and Applications* **22**: 149–166.

Poff NL, Richter BD, Arthington AH, Bunn SE, Naiman RJ, Kendy E, Acreman M, Apse C, Bledsoe BP, Freeman M, Henriksen J, Jacobson RB, Kennen J, Merritt DM, O'Keeffe J, Olden JD, Rogers K, Tharme RE, Warner A. 2010. The Ecological Limits of Hydrologic Alteration (ELOHA): a new framework for developing regional environmental flow standards. *Freshwater Biology* **55**: 147–170.

Postel SL. 1998. Water for food production will there be enough in 2025?. *Bioscience* **48**(8): 629–636.

Postel SL, Daily GC, Ehrlich PR. 1996. Human appropriation of renewable fresh water. *Science* **271**: 785–788.

Quinn GP, Keough MJ. 2002. *Experimental design and data analysis for biologists*. Cambridge University Press: United Kingdom.

Rencher AC. 2002. *Methods of multivariate analysis*. 2ndedn . Wiley-Interscience: Hoboken. USA.

Richter BD, Baumgartner JV, Powell J, Braun DP. 1996. A method for assessing hydrologic alteration within ecosystems. *Conservation Biology* **10**(4): 1163–1174.

Richter BD, Baumgartner JV, Wigington R, Braun DP. 1997. How much water does a river need? *Freshwater Biology* **37**: 231–249.

Richter BD, Mathews R, Harrison DL, Wigington R. 2003. Ecologically sustainable water management: managing river flows for ecological integrity. *Ecological Applications* **13**(1): 206–224.

Sanborn SC, Bledsoe BP. 2006. Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon. *Journal of Hydrology* **325**: 241–261.

Snelder TH, Lamouroux N, Leathwick JR, Pella H, Sauquet E, Shankar U. 2009. Predictive mapping of the natural flow regimes of France. *Journal of Hydrology* **373**: 57–67.

Stanford JA, Ward JV, Liss WJ, Frissell CA, Williams RN, Lichatowich JA, Coutant ChC. 1996. A general protocol for restoration of regulated rivers. *Regulated Rivers: Research & Management* **12**: 391–413.

Suen JP, Eheart JW. 2006. Reservoir management to balance ecosystem and human needs: Incorporating the paradigm of the ecological flow regime. *Water Resources Research* **42**: W03417. DOI:10.1029/2005WR004314

Thoms MC, Parsons M. 2003. Identifying spatial and temporal patterns in the hydrological character of the Condamine-Balonne River, Australia, using multivariate statistics. *River Research and Applications* **19**: 443–457.

Vericat D, Batalla RJ. 2006. Sediment transport in a large impounded river: The Lower Ebro, NE Iberian Peninsula. *Geomorphology* **79**: 72–92.

Vericat D, Batalla RJ, Garcia C. 2006. Breakup and reestablishment of the armour layer in a highly regulated large gravel-bed river: the lower Ebro. *Geomorphology* **76**: 122–136.

Vörösmarty CJ, Green P, Salisbury J, Lammers RB. 2000. Global water resources: Vulnerability from climate change and population growth. *Science* **289**: 284–288.

Ward JH. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* **58**: 236–244.

World Commission on Dams. 2000. Dams and development. A new framework for decision-making. *The report of the World Commission on Dams*. London, Earthscan.

Yevjevich V. 1972. *Probability and statistics in hydrology*. Water Resources Publications: Littleton, USA.