Hydrobiologia (2009) 623:153–171
DOI 10.1007/s10750-008-9655-y

PRIMARY RESEARCH PAPER

The MEDiterranean Prediction And Classification System (MEDPACS): an implementation of the RIVPACS/AUSRIVAS predictive approach for assessing Mediterranean aquatic macroinvertebrate communities

José Manuel Poquet · Javier Alba-Tercedor · Tura Puntí · Maria del Mar Sánchez-Montoya · Santiago Robles · Maruxa Álvarez · Carmen Zamora-Muñoz · Carmen Elisa Sáinz-Cantero · Maria Rosario Vidal-Abarca · Maria Luisa Suárez · Manuel Toro · Ana Maria Pujante · Maria Rieradevall · Narcís Prat

Received: 13 May 2008 / Revised: 14 October 2008 / Accepted: 15 November 2008 / Published online: 12 December 2008

Abstract In Spain, a national project known as GUADALMED, focusing on Mediterranean streams, has been carried out from 1998 to 2005 to implement the European water framework directive (WFD) requirements. One of the main objectives of the second phase of the project (2002–2005) was to develop a predictive system for the Spanish Mediterranean aquatic macroinvertebrate communities. A combined-season (spring, summer, and autumn) predictive model was developed by using the latest improvements on the selection of best predictor variables. Overall model performance measures were used to select the best discriminant function (DF) models, and also to evaluate their biases and precision. The final predictive model was based on the best five DF models. Each one of these models involved eight environmental variables. Final observed (O), expected (E), and O/E values for the number of macroinvertebrate families (NFAM) and two biotic indices (IBMWP and IASPT) were calculated by averaging their values, previously weighted by the quality of each DF model. Regression analyses among the final O and E values for the calibration dataset showed a high proximity to the ideal theoretical model, where the final E values explained 73–84% of the variation present in the macroinvertebrate communities of the

Handling editor: Richard H. Norris

J. M. Poquet (✉) · J. Alba-Tercedor (✉) ·
C. Zamora-Muñoz · C. E. Sáinz-Cantero
Departamento de Biología Animal, Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain
e-mail: jmpoquet@ugr.es

J. Alba-Tercedor
e-mail: jalba@ugr.es

T. Puntí · M. Rieradevall · N. Prat
Departamento de Ecología, Universidad de Barcelona, 08028 Barcelona, Spain

M. del Mar Sánchez-Montoya · M. R. Vidal-Abarca ·
M. L. Suárez
Departamento de Ecología e Hidrología, Universidad de Murcia, 30100 Murcia, Spain

S. Robles
Cimera Estudios Aplicados SL, Parque Científico de Madrid, 28760 Madrid, Spain

M. Álvarez
Área de Ecología, Universidad de Vigo, 36200 Vigo, Spain

M. Toro
División de Ecología de los Sistemas Acuáticos Continentales, CEDEX, 28005 Madrid, Spain

A. M. Pujante
Red-Control SL, Parque Tecnológico de Valencia, 46980 Paterna, Spain

Spanish Mediterranean watercourses. The ANOVA performed among the reference (calibration and validation) and test datasets showed clear differences for the O/E values. Finally, the assessments carried out by the predictive model were sensitive to anthropogenic pressure present in the study area and allowed the definition of five ecological status classes according to the WFD requirements.

Keywords Predictive modelling · GUADALMED project · Bioassessment · Water framework directive · Ecological status

Introduction

Throughout the last century, biomonitoring of freshwater ecosystems has increased everywhere, and different methods have been developed (see Bonada et al., 2006b). Some of these methods have been integrated into regulatory laws (e.g. US Clean Water Act) and have been used by governments as tools to assess the health of freshwater ecosystems (Niemi & McDonald, 2004).

Since the early 1980s, two biomonitoring approaches have increased in popularity among aquatic biologists: the multimetric and the multivariate. The former utilizes measures that represent different characteristics of biological communities to summarize the overall ecological quality into one index value or score (De Pauw et al., 2006). The latter compares the observed and expected composition of biological communities. The expected values are derived from the relationship between a biological classification of reference sites and a set of environmental variables (Wright et al., 2000).

In Europe, the European water framework directive (WFD2000/60/EC; European Commission, 2000) introduces the obligation for its Member States to achieve and maintain good ecological status for all water bodies. Such ecological status must be assessed as a deviation from the reference condition, measuring the Ecological Quality Ratio (EQR = observed/expected) for different quality elements (macroinvertebrates, diatoms, macrophytes, and fish). While some European countries have adopted the multimetric approach in assessing biological communities (e.g., Hering et al., 2006), others have chosen to develop predictive systems based on the multivariate approach. These predictive systems, namely the SWEPAC in Sweden (Johnson & Sandin, 2001), the PERLA in the Czech Republic (Kokeš et al., 2006), or the Luxembourgian model (Ferréol et al., 2008), are based on the pioneer RIVPACS of Great Britain (Wright et al., 1984; Moss et al., 1987) and its Australian derivative AUSRIVAS (Simpson & Norris, 2000).

In the past, Armitage et al. (1990) applied the RIVPACS system to two rivers in a small area in northwest Spain. These authors recognize that the Iberian Peninsula represents a large zoogeographical block and that the development of its own predictive scheme would be better than applying other predictive systems. The Mediterranean streams are markedly influenced by the heterogeneity in temperature and rainfall regimes caused by the Mediterranean climate. Yearly variable discharge regimes of a maximum peak in winter and a minimum flow in summer create changes in freshwater communities over time (McElravy et al., 1989; Resh et al., 1990). Differences in faunal composition, as well as in environmental characteristics, would affect the assessment carried out using the British predictive system for any Iberian Mediterranean stream. These peculiarities highlight the need to develop a specific predictive system for this type of stream.

In 1998, a national project known as GUADALMED, focusing on Spanish Mediterranean streams, was promoted to satisfy the European water framework directive (WFD). One of the goals of the second phase of the project (2002–2005) was to develop a MEDiterranean Prediction And Classification System (MEDPACS) for aquatic macroinvertebrate communities. Despite the widespread development of predictive models around the world (e.g., Chessman et al., 1999; Hawkins et al., 2000; Joy & Death, 2002; Davis et al., 2006; Kennard et al., 2006; Hargett et al., 2007), to our knowledge, no predictive system has been put forward for Mediterranean streams. Recently, predictive models based on the Canadian predictive approach (BEAST; Reynoldson et al., 1995) were developed in Portugal (Feio et al., 2007a, b). They were confined to three watersheds located outside the Mediterranean area of the Iberian Peninsula (Köppen, 1923).

In this article, we report on the development of a predictive model for Mediterranean streams of a wide geographical area inside the Iberian Peninsula. The model, based on the RIVPACS/AUSRIVAS predictive approach, involves the use of the EQR for the number of macroinvertebrate families and for two previously developed biotic indices (IBMWP and IASPT, formerly BMWP' and ASPT', Alba-Tercedor & Sánchez-Ortega, 1988; Alba-Tercedor, 1996; Alba-Tercedor & Pujante, 2000; Alba-Tercedor et al., 2004). We applied the latest improvements on the selection of the best predictor variables as carried out by Van Sickle et al. (2006), as well as overall model performance measures to achieve the final predictive model. Furthermore, the sensitivity of the model was tested with an available dataset from the GUADALMED project, defining ecological status class boundaries as the new European legislative framework requires.

Materials and methods

Study area

The Iberian Peninsula is characterized by hot dry summers and cool wet winters. Annual temperatures range between 42°C and −2°C (mean value of 16–17°C), and annual precipitation varies from 280 to 1,000 mm (average about 600–650 mm). Important storm events often produce flooding during spring and autumn. High temperatures and low rainfall during the summer season lead to a natural water scarcity, generating drought events and in some cases the complete drying up of streams (Gasith & Resh, 1999).

After the Köppen (1923) climate criteria, the Iberian Mediterranean zone would include almost the entire southern two thirds of the peninsula since from a geographical point of view, the Mediterranean zone *sensu stricto* would be strictly composed by those watersheds that drain into the Mediterranean Sea (Fig. 1). The study area covered approximately 84,400 km^2, including large and very small watersheds (e.g., Júcar with 18,136 km^2 or Vicar with 12 km^2). Armengol et al. (2004) provide a detailed characterization of most of the watersheds. The second phase of the GUADALMED project involved 33 watersheds (135 watercourses) of the Iberian eastern coast and the Balearic Islands. Each sampling site was visited in spring, summer, and autumn of 2003, with the objective to maximize the information collected about the Iberian Mediterranean macroinvertebrate communities. Although some Mediterranean streams dry out in summer, to include this campaign into the sampling program allows emphasizing natural differences between permanent and temporal stream macroinvertebrate communities, obtaining a better biological characterization of sites. A dataset of 162

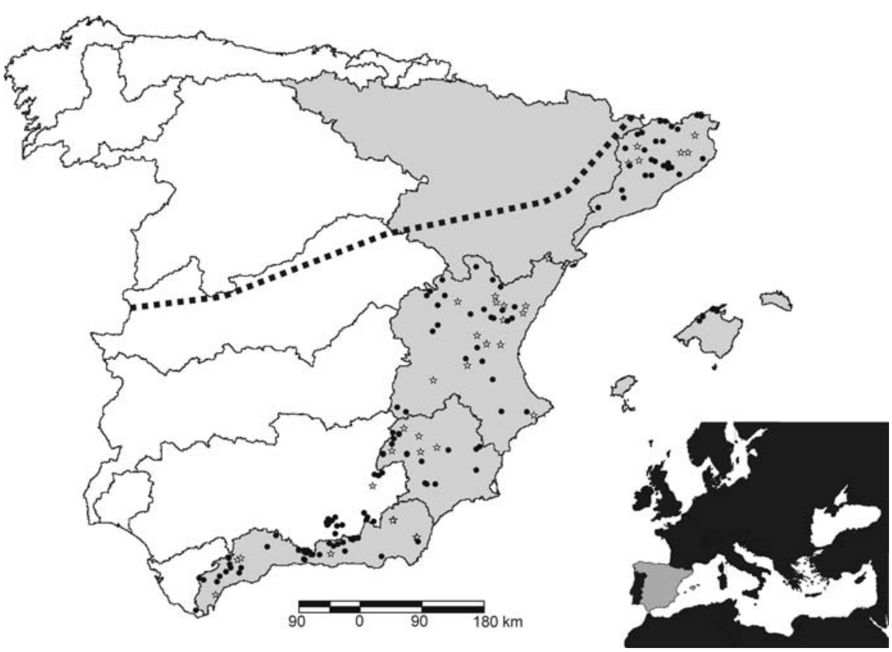

Fig. 1 Location of the calibration, validation, and test datasets (solid circles, open circles, and crosses, respectively) sampled during the GUADALMED project throughout the Spanish eastern area. The main watersheds of the country are shown, and Mediterranean watersheds *s. st.* are filled in grey. Dotted line represents the Köppen (1923) Mediterranean-climate boundary

potential reference sites was selected by expert judgment in order to cover as much natural environmental variability as possible, mostly inside of the Mediterranean zone *s. st.* and always inside Köppen's Mediterranean zone. Thus, sites ranged from very small streams at high altitudes to large streams in the lowlands, as well as streams with constant flow, fast flooding, and intermittent streams (e.g., *ramblas* characterized by Gómez et al., 2005).

Selection of reference sites

Final reference sites were selected according to ten objective criteria previously established during the first phase of the GUADALMED project (Bonada et al., 2004b), namely: (1) less than 10% of the watershed area is exposed to anthropogenic stressors (urban, industrial, and agricultural); (2) there is proper riparian vegetation according to the type of stream. In most cases, this condition refers to well-formed riparian vegetation, with 100% of tree cover. However, in some cases, such as high headwaters (>2,000 m a.s.l.), the natural situation could be represented by a riparian vegetation without trees. Similarly, the natural hydrological stress in temporal streams and *ramblas* could produce a riparian vegetation with less than 100% tree cover. (3) Riparian vegetation is formed by native species. (4) Natural riparian banks are without disturbances, such as human-built structures (factories, buildings, sports centers, etc.). (5) The fluvial channel is unmodified, i.e., the fluvial channel cannot present any modification, such as canalizations, breakwaters, or similar structures. (6) The streams are non-regulated, i.e., the site must not present any flow regulation, such as a dam upstream. (7) Sites have unaltered instream habitat, meaning the substrate should be that which is expected for the type of stream, e.g., with boulders in headwaters, gravel and pebbles in midlands, and sand or silt in lowlands. Finally, the last three reference criteria concern water quality, such as nutrient threshold concentrations: (8) <0.01 mg/l for nitrites, (9) <0.5 mg/l for ammonium, and (10) <0.05 mg/l for phosphates. Although nitrate concentration is used to denote anthropogenic impairments related to agriculture, they were excluded as reference criterion because some watersheds of the study area present high concentrations of this parameter in natural conditions (Gómez et al., 2005).

A site was considered as a reference when it fulfilled at least seven of ten criteria, except for modification of the natural fluvial channel or flow regulation. If a site failed to meet either of these criteria, it was not considered as a reference.

Sampling design and data collection

The sampling methodology was based on a rapid bioassessment protocol widely applied in Spain (Zamora-Muñoz et al., 1995; Zamora-Muñoz & Alba-Tercedor, 1996; Alba-Tercedor & Pujante, 2000; Alba-Tercedor et al., 2004). The design of the sampling methodology was carried out during the first phase of the GUADALMED project. This methodology has been used in several benthic studies (e.g., Bonada et al., 2005, 2006a, c; Sánchez-Montoya et al., 2007) and provides a standardized dataset as well as the ability to obtain the maximum information from the macroinvertebrate communities.

Among all groups involved in the project, a common protocol (Jáimez-Cuéllar et al., 2004) was established through an intercalibration exercise (Bonada et al., 2004a). In each site, a multi-habitat sample from all available habitats was collected with a kick net (250 µm-mesh size). Under high speed current, a 400-µm-mesh net was used. This change in mesh size does not affect the results obtained at the scale covered in this study, as Statzner et al. (2004) demonstrated.

We emptied collected material into trays and identified organisms to the family level (except for Hydracarina, Oligochaeta, and Ostracoda). This taxonomic level was chosen firstly because of the lack of knowledge about many of the species not yet described on the Iberian Peninsula and secondly by its easy identification in the field. In this way, Spanish scientists and water managers have reached a general agreement, using this taxonomic level for macroinvertebrates in freshwater biomonitoring programs. Furthermore, the good performance of predictive models previously developed by using the family level (e.g. Furse et al., 1984; Simpson & Norris, 2000) makes us confident about the reliability of the MEDPACS approach at this taxonomic level. A maximum of three individuals from each family was stored in a vial as field record (stage 1). Organisms not collected but seen in the field (e.g., Heteroptera) were also registered. The sampling finished when no new taxa were recorded. The remaining material was stored in 4% formalin and, once in the laboratory, was washed and placed in a partitioned tray. We sorted divisions with a computer-generated random number list until 200 individuals were separated (stage 2). Rank abundances were estimated for the whole sample: 1 = 1 to 3 individuals, 2 = 4 to 10 individuals, 3 = 11 to 100 individuals, and 4 > 100 individuals (stage 3). This three-stage protocol reduces the number of missing taxa in the field and provides a more comprehensive taxa list for the predictive model development.

For each season sampled, a presence-absence family list was obtained for each site, incorporating those families appearing at each stage of the whole process (field records, sub-sampling, and rank counting). In the predictive model development, seasonal presence-absence family lists (spring, summer, and autumn) were merged with the aim to get a unique family list for each site.

There are several criteria for defining rare taxa in benthic studies (Cao et al., 2001), but there is no agreement as to whether or not rare taxa should be removed from the dataset when using multivariate analysis (Cao et al., 1998; Cao & Williams, 1999; Marchant, 1999). We decided to exclude the rarest macroinvertebrate families from the dataset, considering rare taxa as those present at less than 2% of the sites.

Environmental variables

A set of 15 environmental variables was selected for model development from the GUADALMED database and measured at different scales: main watershed, reach watershed, reach, and site (Table 1). Environmental variables that could be influenced by human activities were not included (e.g., nutrient content). These variables are not suitable for this kind of model design to predict taxa expected at reference quality sites (Clarke et al., 2003). Furthermore, correlated environmental variables were also discarded ($R > 0.75$) to avoid redundant information. Environmental data were mainly available from the CEDEX (*Centro de Estudios Hidrográficos*, Spain) database. Variables were transformed to achieve normality when necessary for the model building process.

Multivariate analyses

The reference dataset was divided into calibration and validation datasets. The latter was set aside and used afterwards to validate the predictive model as an independent reference dataset. Calibration reference

Table 1 Environmental variables selected from the GUADALMED database to develop the discriminant function models

Scale	Name	Variable	Description
Watershed[a]	W_Area	Watershed area	Surface of the watershed (km^2)
	W_Car	% Carbonated watershed	Percentage of carbonated materials in the watershed
	W_Evp	% Evaporitic watershed	Percentage of evaporitic materials in the watershed
	W_Slp	Watershed slope	Mean slope of the main watershed (%)
	Alt_max	Maximum altitude	Maximum altitude for the main watershed (m)
Reach[a]	Latitude	Latitude	Latitude value measured in degrees
	Altitude	Altitude	Altitude of the site (m a.s.l.)
	St-Ord	Stream order	Stream order following the Sthraler system (scale 1:50,000)
	Slp	Slope	Mean slope of the reach watershed (%)
	Air-Temp	Air temperature	Annual mean temperature of the air (°C)
	Air-Temp-rng	Air temperature range	Annual temperature range of the air (°C)
	Evaporit	% Evaporitic	Percentage of evaporitic materials in the reach watershed
Site[b]	Dry	Temporality	Temporality of flow of the site along samplings (yes/no)
	Spr	Spring	Presence of a spring upstream the site (yes/no)
	sqrt_Alkal-rng	Alkalinity range	Square root-transformed water alkalinity range (meq/l)

[a] Variables measured using GIS

[b] Variables measured in field samplings

sites were classified into groups based on their macroinvertebrate composition (e.g., Wright et al., 1984; Parsons & Norris, 1996). A clustering technique (flexible UPGMA) recommended by Belbin and McDonald (1993) was run with a β value of -0.6 (Van Sickle et al., 2006), and based on the Bray-Curtis similarity measure (Reynoldson et al., 1995; Simpson & Norris, 2000). A non-metric multidimensional scaling (NMDS) was performed to aid in the decision-making for the resultant number of groups (Bailey et al., 2004). According to Wright et al. (1993), each group should have more than five reference sites; smaller groups will result in poor type-site representation and modeling errors (Simpson & Norris, 2000). However, the final decision about how many biological groups to use in the model development is subjective (Bailey et al., 2004), and classifications close to the selected one could be appropriate as well. In our case, the number of groups was determined by the researcher's judgment; however, alternative classifications with different numbers of groups and close to the selected one were also built to solve this problem.

Traditionally, the selection of the best predictor variables for macroinvertebrate classification groups has been done with stepwise methods (e.g., Parsons & Norris, 1996; Reynoldson et al., 1997; Hawkins et al., 2000; Hargett et al., 2007), generally the stepwise discriminant function analysis (stepwise-DFA). However, as Van Sickle et al. (2006) argue, stepwise DFA is vulnerable to the same problems as stepwise regression analysis. Stepwise procedures only explore a small subset of all possible models, where the order of entry (or deletion) of environmental variables, as well as their number, can affect the final model selected. Variable inclusion and removing rules of the model are based on partial *F*-tests of individual predictors, which can lead to biases in parameters, over-fitting, and incorrect significance tests (Whittingham et al., 2006). On the other hand, the stepwise selection relies on the selection of one best single model when other models may have a similarly good fit.

Van Sickle et al. (2006) have implemented the 'best-subset' approach as an alternative approach to the discriminant function variable selection. We applied this procedure to our reference dataset and to each one of the alternative biological classifications obtained previously. This approach, written as function scripts for the R language (Ihaka & Gentleman, 1996), explores all possible candidate discriminant function (DF) models for a given site classification. It ranks the set of all possible models by their Wilk's lambda separately within each model order. This is defined by the number of predictor variables included in the DF model (e.g., a DF model of the second model order would be a DF model with two environmental variables). The program enables retention of a number of best models of each order (models with the lowest Wilk's lambda) and calculates statistics for calibration and validation datasets.

As previous predictive models, the 'best-subset' program calculates for each site probabilities of belonging to each group (Mahalanobis distance between the site and the centroid of each group in the multidimensional space derived by the DFA). A site could vary within an environmental continuum, falling between the centers of two or more groups, and presenting with similar probabilities for belonging in each one of those groups (Clarke et al., 1996). The final probability of capture for one family in a site is calculated as the sum of the probabilities of belonging to each biological reference group, weighted by the frequency of occurrence of that family inside each group (Moss et al., 1987; Clarke et al., 1996). In this way, several site groups and a large proportion of reference sites actively contributed to the predictions of the macroinvertebrate families in a test site.

The 'best-subset' program calculates the observed (O) and expected (E) taxa for a given site, providing O/E values and the related statistics (mean and SD). The O value is the number of captured taxa recorded during the sampling process (stage 1–3), and the E value is the sum of probabilities of capture for each of the taxa at that site. These values can be obtained based on the full reference taxa list predicted to occur (Moss et al., 1987) or by the reference taxa list limited to different capture probabilities, excluding families at different thresholds ($P_t = 0.1, 0.2, 0.3, \ldots, 1.0$). Clarke and Murphy (2006) conclude that the statistical power needed to detect overall biological impacts of human activities based on the O/E for the number of taxa is similar for thresholds up to 0.7. In our case, we decided to use a threshold probability of $P_t \geq 0.5$, following the same criterion as the Australian approach (Simpson & Norris, 2000) and the predictive models developed in the USA (Hawkins et al., 2000; Hargett et al., 2007).

Overall DF model performance

The 'best-subset' program calculates the root mean squared error (RMSE) of O/E values to evaluate the overall model performance. The RMSE combines the bias and variability of prediction errors into a single measure of model performance (Van Sickle et al., 2006, p. 364, Eq. 1). Low RMSE values would denote an improvement in the overall model performance. The overall performance of any predictive model at calibration sites can be evaluated by comparison with an upper boundary for a model to be effective. Such a boundary is the RMSE (O/E) value of a null model. This null model is defined as a limit that would be achieved whenever a predictive model fails to explain any of the natural-gradient variation in assemblages (Van Sickle et al., 2005). In this case, the E value is a single fixed value for all reference sites. It is calculated as the sum of each probability of capture for any taxon, with each probability being estimated as the proportion of calibration sites at which the taxon was observed to occur (Van Sickle et al., 2005). In the calibration dataset, bias for the null and any predictive model is zero or nearly zero, but is not the case for the validation dataset. The possible bias in model predictions for validation sites and the variability present in O/E values between calibration and validation datasets provoke a direct comparison between RMSE (O/E) values for the validation dataset, and the null model could not give a reliable measure of a model's performance when applied to new sites. Thus, the 'best-subset' program calculates a separate baseline for assessing predictive model performance at validation sites, i.e., a RMSE (O/E) value for a null model where the E value is obtained from the validation dataset instead of the calibration dataset (Van Sickle et al., 2006). The overall performance of any predictive model assessing independent sites [RMSE (O/E) value for validation dataset] can be judged by how far it reduces below the upper baseline established by the null model for the validation dataset.

The 'best-subset' program also calculates the replicate sampling standard deviation (SD_R) for the calibration dataset. This value represents the minimum variation in O/E values because of replicate observed assemblages. No predictive model can be expected to show a SD for O/E at calibration sites less than SD_R. This value is based on true occurrence probabilities; however, Van Sickle et al. (2005) showed that it can be calculated based on model-predicted taxon occurrence probabilities. The derivation of SD_R assumes zero bias in O/E values at calibration sites, so that it serves as a lower baseline for RMSE (O/E) values.

MEDPACS assessments: the observed and expected IBMWP-IASPT

Once the expected number of families (NFAM) was calculated for each one of the best DF models selected, the expected values for IBMWP-IASPT biotic indices could be estimated within the framework of the MEDPACS. Like the original British BMWP, the IBMWP assigns a score within a range of 1–10 to each family of macroinvertebrates present in the Iberian Peninsula, according to their known tolerance to pollution (Alba-Tercedor & Pujante, 2000; Alba-Tercedor et al., 2004). The IBMWP for a site is the sum of the scores of the recorded families present in that site. The expected IBMWP for a site is given by the sum of probabilities of capture for all recorded families, each one multiplied by its IBMWP score. The IASPT was estimated as the IBMWP divided by the number of scoring families. The expected IASPT results from the expected IBMWP divided by the sum of probabilities of capture in a given site.

O and E values for the NFAM and IBMWP-IASPT biotic indices were calculated for calibration and validation datasets, and for each one of the best DF models selected. Unlike other predictive models, the final O and E values in MEDPACS were obtained by averaging values of the best DF models, weighted by the reduction percentage of the RMSE (O/E) for the validation dataset with respect to its upper limit. This value was used as a model's quality measure.

Each site was evaluated to determine whether it was within the range of the model (Moss et al. 1987; Clarke et al. 1996). The evaluation of a site placed outside of the environmental range of the predictive model could produce a bias in the assessment of ecological status. In the MEDPACS approach, a site was considered outside of the environmental range of the predictive model when it was identified as an outlier by three to five of the best DF models selected. When a site was considered as an outlier by less than three DF models, the O and E final values were calculated based on the remaining DF models.

Predictive model evaluation and sensitivity

We performed regression analyses between the final O and E values for the calibration dataset in order to check how well the E values were able to explain the corresponding O values. Similarly, these analyses were used to assess the nearness of the MEDPACS development to the ideal model. In a theoretical ideal predictive model, a linear relationship between O and E values would be expected. The O/E value for every reference site would be equal to one, regardless of the site's family richness. This would be translated as a slope of one and intercept zero in a regression equation for the O and E values (Linke et al., 2005), where the E values would explain the 100% of O value variation ($R^2 = 1$). We performed a one-way ANOVA, followed by Tukey multiple comparison tests, to look for significant differences among O/E mean values for the selected biological groups. Similarly, a set of 72 test sites available from the first phase of the GUADALMED project (1999–2000) was used to evaluate the sensitivity of the model. The family list of each test site was obtained as those used to develop the predictive model, merging the corresponding three seasonal family lists. These sites were within the range of the model (Fig. 1). In this case, analyses were performed among calibration, validation, and test datasets.

The ecological status class boundaries

The O/E gradient of each ecological indicator (NFAM and IBMWP-IASPT indices) must be categorized into one of five ecological status classes (high, good, moderate, poor, and bad) according to the WFD. Thus, calibration, validation, and test datasets were used together to create the corresponding O/E gradients. These gradients were correlated with a general stressor gradient, obtained by a standardized Principal Components Analysis (PCA), performed with the STATISTICA software v 7.1 (Statsoft, 2005). The PCA was based on land use variables (Corine Land Cover 2000), on dissolved oxygen content (mg/l), and on riparian forest quality, as measured by the QBR index (Munné et al., 2003). Previously, land use variables (percentage of watershed area earmarked for farming, grassland, roads, mines, urban area, and natural area) were transformed by arcsine [$\sqrt{(x/100)}$] to reduce their skewness (Legendre & Legendre, 1998). Linear and second order polynomial regressions were run among the general stressor gradient and each O/E gradient. We used the Bayesian Information Criterion (BIC; Schwarz, 1978) to identify which regression model (linear or polynomial) better describes the relationship between the O/E gradient for each ecological indicator and the general stressor gradient. Smaller BIC values indicate better, more parsimonious models (Quinn & Keough, 2002).

In those cases where a linear relationship was identified, ecological status class boundaries were defined according to the REFCOND Guidance (European Commission, 2003). The high-good class boundary was set up according to the 25th percentile of the reference distribution. The remaining O/E gradient was divided into equal divisions to define the lower ecological status classes. When a polynomial relationship was found, we followed the approach proposed by Munné and Prat (2006). We transferred the high-good class boundary from the O/E gradient to the general stressor gradient by means of the regression line. Afterwards, we set the lower ecological status classes, dividing the remaining stressor gradient into equal divisions, which we then transformed to their equivalents in the corresponding O/E gradient. In this sense, the ecological status class boundaries will be more realistic and representative of the anthropogenic pressure in the study area.

Results

Reference sites

The potential reference dataset was screened with GUADALMED reference criteria, and 128 sites were selected as reference sites. Each individual reference criterion was accomplished in more than 80% of the reference sites (with the exception of watershed land use and ammonium concentration criteria, whose percentages were lower). Thirty-four sites failed, either on more than three reference criteria or on the modification of the natural fluvial channel or flow regulation criteria, and were thus classified as non-reference (Table 2).

Table 2 Potential reference dataset split into reference and non-reference datasets after applying the GUADALMED reference criteria (Bonada et al., 2004b)

sReference criteria	128 reference sites	34 non-reference sites
<10% total land uses in the watershed	71	18
Natural riparian vegetation	95	88
Native riparian vegetation	97	79
Non-impaired riparian banks	95	79
Non-impaired fluvial channel	99	97
Non-regulated streams	100	38
Natural instream habitat	100	100
Ammonium (<0.5 mg/l)	34	9
Nitrites (<0.01 mg/l)	85	53
Phosphates (<0.05 mg/l)	81	44
Average of reference criteria	8.05	6.06
Standard error	0.09	0.22

Comparison in percentage of sites that complete each reference criterion applied to both subsets. Average and standard errors are shown for the number of criteria accomplished

Finally, 122 reference sites were available for the model development. The reference dataset was randomly split into calibration (80—98% reference sites) and validation (20—24% reference sites) datasets (Fig. 1). Twenty-one macroinvertebrate families (from 124 recorded) were considered as the rarest families and excluded from further analyses (see Methods).

Biological classification

The cluster analysis for the calibration dataset with the merged family presence-absence matrix (spring, summer, and autumn) led to five groups with a cut-off level of 76% of information remaining (Fig. 2). The ordination of these five groups in the first two axes of the NMDS analysis (20.21% stress level) shows little overlap. Groups 4 and 5 were the most scattered, while groups 1 and 2 were the most concentrated (Fig. 3).

There was a small difference with alternative classifications, a six-group classification (cut-off level

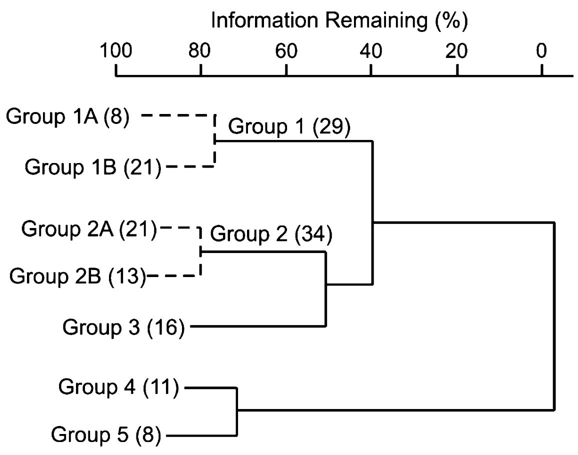

Fig. 2 Biological classification for the calibration dataset (98 reference sites), using cluster analysis. The number of sites in each group is shown in parentheses

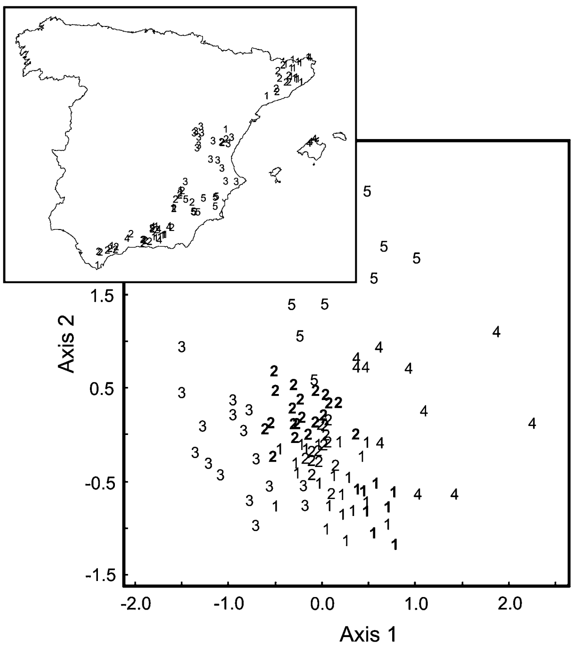

Fig. 3 NMDS two-dimensional ordination of reference communities obtained from the calibration dataset. The five-group classification is shown, as well as its geographical location (upper left-corner map). The splits of group 1 and 2 of the other two alternative classifications are also shown in the ordination space (groups 1A and 2A are highlighted in bold)

of 77%), and a seven-group classification (cut-off level of 80%). In the former, group 1 (29 sites) could be divided in two groups (Fig. 2), one composed of 8 sites (group 1A) and the other of 21 sites (group 1B). In the seven-group classification, group 2 (34 sites) could also be divided into two sub-groups, one composed of 21 sites (group 2A) and the other of 13 sites (group 2B).

The 'best-subset' approach: selection of the best DF models

The environmental variables (Table 1) and the three biological classifications described above (five, six, and seven groups) were used to run the 'best-subset' program. Fifteen model orders were defined, and the best five DF models (selected by their Wilk's lambda) were retained for each model order. Seventy-one best models were selected from 32,767 possible DF models for each biological classification. Figure 4 shows the overall model performance measures for each model selected. For each biological classification, the RMSE (O/E) values calculated for the calibration dataset (Fig. 4, solid squares) were located within the upper and lower boundaries for which the model was effective [null model RMSE (O/E) value and SD_R]. The RMSE (O/E) values decreased quickly in the first model orders. Nevertheless, once the seventh, the sixth, or the fifth model orders were reached for each biological classification (seven, six, and five groups), they became steady. Once it was checked that models were located between the upper and lower boundaries for the calibration dataset, the RMSE (O/E) values for the validation dataset were used to select the best model order for each biological classification. Thus, we elude selecting over-fitted DF models, i.e., models closely tailored to the calibration dataset that perform poorly when applied to independent sites, such as those belonging to the validation dataset. A similar pattern was found in the RMSE (O/E) for the validation dataset (Fig. 4, open squares). In this case, the stopping point for such decreasing tendency was given by the seventh and the eighth model orders. The five best combinations of eight environmental variables within all candidate variables (for five- and six-group classifications) and the five best combinations of seven environmental variables (for the seven-group classification) were selected as the model

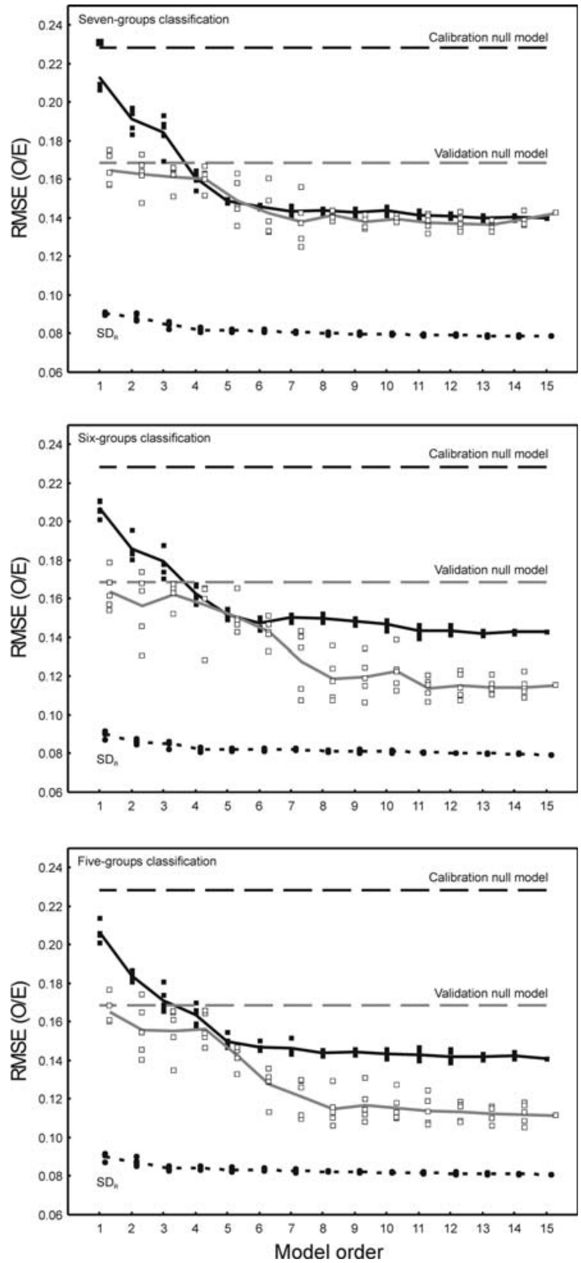

Fig. 4 Overall model performance measures calculated for the best 71 DF models. RMSE (O/E) values for both calibration (solid squares) and validation (open squares) datasets are shown for each biological classification studied, as well as the corresponding upper and lower baselines. Lines connect model order mean values

orders where the model performance exhibited no further improvements. In this case, models built with the seven-group classification presented RMSE (O/E) validation values higher than those of the other two classifications. These values were closer to the upper limit established by the validation null model than those values for models built with five- and six-group classifications. Therefore, models built with a seven-group classification were less effective in assessing independent reference datasets.

Differences among models for the other two biological classifications were minimal, with a small improvement in the five-group classification. Finally, the five best DF models developed for the five-group classification, with combinations of eight environmental variables, were used to calculate the final O, E, and O/E values for the number of families and the IBMWP-IASPT indices. All five DF models involved 12 environmental variables, leaving out the altitude of the site, the presence of a spring upstream, and the water alkalinity range.

Characterization of the reference site classification

The most common environmental variables among the five DF models selected (present in at least four of them) were used to characterize the assemblages of reference sites (Fig. 5). Thus, group 1 included reference sites from small streams with permanent flow, small watersheds, and low contents of evaporitic and carbonated materials. This group was mainly composed of headwaters from the Sierra Nevada and Pyrenees Mountains, located in the south and the north of the study area, respectively (Fig. 3). Sites of this group were mainly characterized by macroinvertebrate families from Trichoptera, Ephemeroptera, and Plecoptera orders, adapted to well-oxygenated watercourses (e.g., Sericostomatidae, Heptageniidae, Nemouridae, Perlidae, Limnephilidae, Chloroperlidae, or Ephemerellidae).

The sites of group 2 were distributed along the whole study area. These sites are similar to those of group 1 with respect to their watershed areas and stream orders. They presented a higher content of carbonated and evaporitic materials, as well as high air temperatures (Fig. 5). In this group, macroinvertebrate families were more diverse than in group 1. In this case, together with some Trichoptera (Hydroptilidae, Leptoceridae, or Polycentropodidae), some Odonata are noticeable (e.g., Gomphidae, Aeshnidae, or Coenagrionidae), Coleoptera (e.g., Gyrinidae, Elmidae, Hydraenidae, or Dryopidae), Diptera (e.g., Stratyomidae, Athericidae, or Limonidae), and Mollusca (Sphaeridae, Hydrobiidae, or Planorbidae). Group 3 was similar to group 2 in relation to the

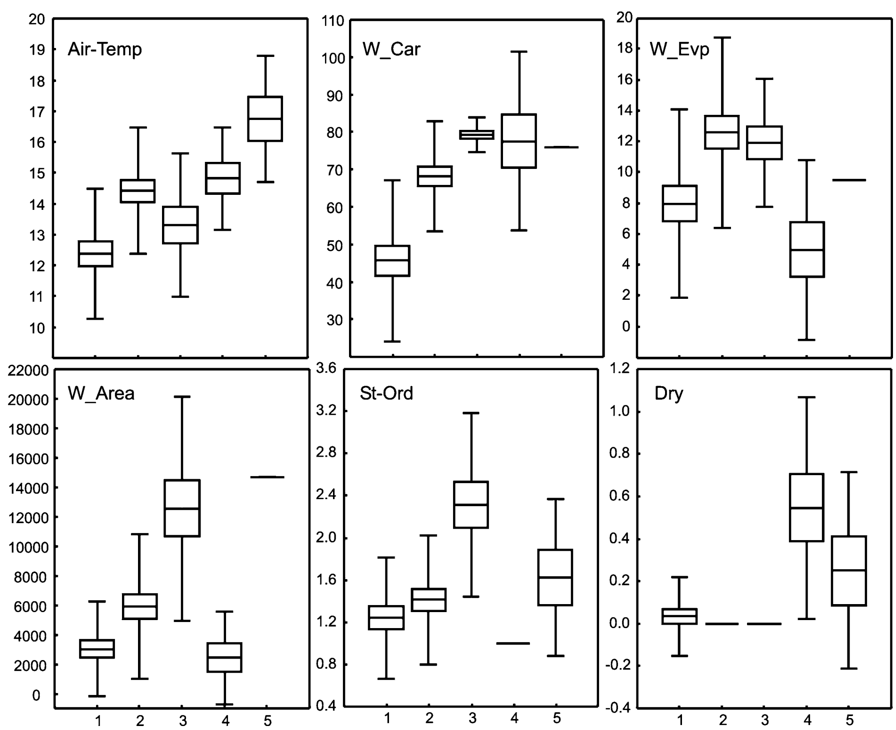

Fig. 5 Five-group classification box plots for the six most common environmental variables among the five best DF models selected. Mean, standard error, and standard deviation are represented by line, box, and whisker, respectively. See Table 1 for variable description

content of evaporitic materials present in their reach watersheds. However, these sites presented the highest watershed areas and stream orders. This group was composed of sites from midland permanent streams located in the Palancia, Júcar, and Turia River watersheds in the center of the study area (Fig. 3). This group showed characteristically families such as Hydrometridae, Gammaridae, Ecnomidae, or Thiaridae. The main characteristic of the last two groups (groups 4 and 5) was the temporal flow regime of their streams. However, both presented some differences, mainly in the content of evaporitic materials and watershed areas, as well as in their stream orders and annual air temperatures. Group 4 included temporal streams from the Balearic Islands and others of similar geomorphology from the south of the study area, whereas group 5 was composed of *ramblas* (fast flooding and intermittent streams) located in the Segura River watershed in the southeast of the study area (Fig. 3). Flow temporality of these groups cause macroinvertebrate assemblage in this kind of site to be reduced, and also dominated by Heteropterans (e.g., Naucoridae, Pleidae, and Nepidae). These organisms are able to breathe atmospheric oxygen and also are able to fly to another watercourse when the stream dries out.

Performance of the MEDPACS approach: final O, E, and O/E values

The predictive model identified five calibration sites and one validation site as outliers. These sites were discarded from further analyses. Thus, 93 calibration sites were used for regression analyses (between the final O and E values for the number of families and IBMWP-IASPT indices; Fig. 6). The observed number of macroinvertebrate families rose as the number of expected families increased. The E values obtained by the MEDPACS approach explained most of the variation in the observed number of families (NFAM $F_{1,91} = 238.45$, $P < 0.001$, $R^2 = 0.73$). Better results were obtained for IBMWP-IASPT indices (IBMWP $F_{1,91} = 306.10$, $P < 0.001$, $R^2 = 0.77$; IASPT $F_{1,91} = 488.82$, $P < 0.001$, $R^2 = 0.84$). On the other hand, the null hypothesis (intercept among O and E values equals zero) was not rejected for any of the three indices (NFAM $t_{91} = 0.914$, $P = 0.36$; IBMWP $t_{91} = 0.774$, $P = 0.44$; IASPT $t_{91} = 0.152$, $P = 0.88$).

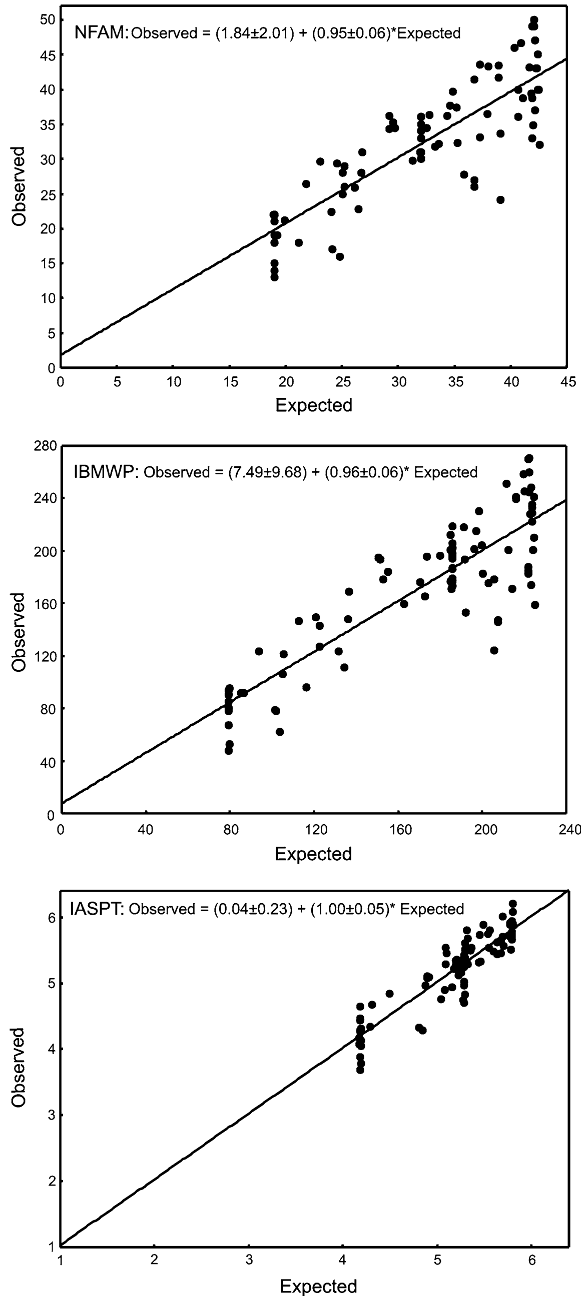

Fig. 6 Regression analysis scatter plots between calibration O and E values for the number of families (NFAM) and IBWMP-IASPT biotic indices. Regression equations show the corresponding standard error for intercepts and slopes

There was no evidence that calibration O/E values varied among classification groups (NFAM $F_{4,88} = 2.224$ $P = 0.072$; IBMWP $F_{4,88} = 1.543$ $P = 0.197$; IASPT $F_{4,88} = 0.89$ $P = 0.476$). Significant differences were found in the analysis of variance among

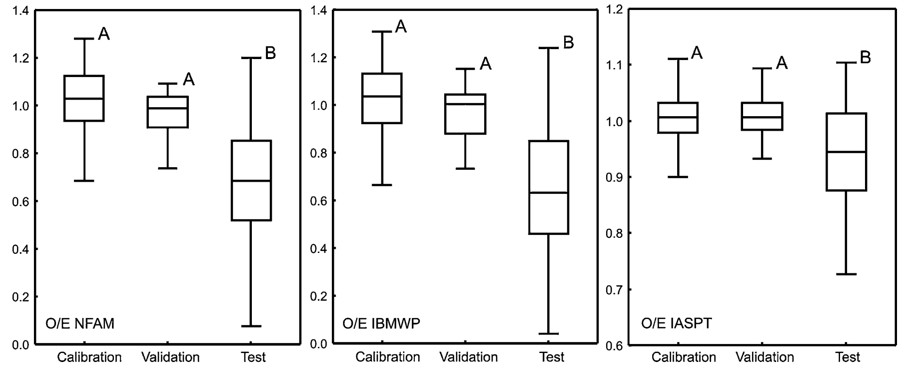

Fig. 7 Box plots of O/E values for calibration, validation, and test datasets for each ecological indicator studied. Median, 25–75%, and non-outlier range are represented by line, box and whisker, respectively. Capital letters identify homogeneous groups determined by the Tukey multiple comparison tests

the O/E values for the calibration, validation, and test datasets (NFAM $F_{1,185} = 49.83$, $P < 0.001$; IBMWP $F_{1,185} = 51.76$, $P < 0.001$; IASPT $F_{1,185} = 20.44$, $P < 0.001$). Tukey multiple comparison tests revealed that these differences were due to differences between the test and both reference datasets (Fig. 7).

The ecological status versus the general stressor gradient

One hundred and eighty-eight sites constituted the dataset that defined the general stressor gradient and the ecological status classes (93 calibration sites, 23 validation sites, and 72 test sites). We identified the first PCA axis as the general stressor gradient. This axis explained 36% of the variance present in the dataset (Fig. 8). The dissolved oxygen content, the quality of the riparian forest, and the watershed natural area were negatively correlated with this axis ($R = -0.42$, -0.77, and -0.83, respectively). On the other hand, the remaining land use variables were positively correlated with this axis, namely, watershed farming area ($R = 0.47$), watershed grassland area ($R = 0.28$), watershed urban area ($R = 0.79$), mines ($R = 0.47$), and roads ($R = 0.52$).

Despite small differences among BIC values (Table 3), we identified a linear relationship between the general stressor gradient and the NFAM and IBMWP O/E gradients. However, the IASPT O/E gradient showed a polynomial relationship with the general stressor gradient. We defined each ecological status class according to the procedures described in the Materials and Methods section. We set the high-good class boundary at the 25th percentile of the reference distribution (116 reference sites: calibration

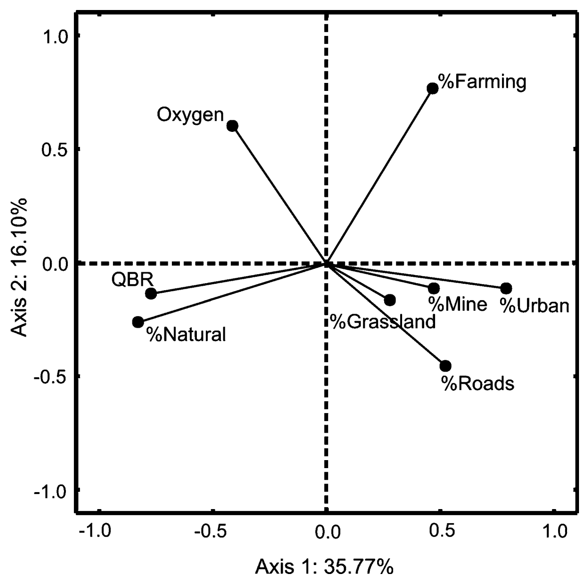

Fig. 8 Plot of the first two standardized PCA axes showing the position of the variables used to obtain the general stressor gradient

and validation datasets) for each ecological indicator (Fig. 9). Ranks for each ecological indicator studied are shown in Table 4, as well as the corresponding number of sites within each ecological status class.

Discussion

Predictive model development

In the process of building a predictive model, several decisions must be made. The number of biological groups, the set of environmental variables, and the probabilities of capture involved in the calculation of

Table 3 Bayesian Information Criterion (BIC) values calculated for linear and polynomial (second order) regressions

Ecological indicator	Linear regression	Polynomial regression (second order)
NFAM	−613.97	−613.36
IBMWP	−583.19	−581.19
IASPT	−999.02	−1,026.64

Analyses were performed between the general stressor gradient and the O/E gradient for each ecological indicator

O/E values are all issues that may produce different results for the same dataset. Alternative methods have been studied to achieve an optimal site classification (Moss et al., 1999), and some guidelines have been established to select the minimal number of sites for any biological group (Wright et al., 1993). However, the final decision about how many groups should be used in model development is somewhat subjective (Bailey et al., 2004) and is usually made using visual inspection by the researcher (e.g., Parsons & Norris, 1996). In the MEDPACS approach, alternative biological classifications close to the selected one were tested, with the aim of reducing such subjectivity (Figs. 2, 3, 4). Predictive models with a high number of groups have a higher probability of classifying a site into the incorrect group. This could affect the results obtained by predictive models where predictions are made only on the basis of the most probable biological group, as in the BEAST approach (Reynoldson et al., 1995). However, in RIVPACS/AUSRIVAS-type models, such an effect is lessened because the final probabilities of capture are calculated with probabilities of belonging to every group (Moss et al., 1987; Clarke et al., 1996). Small differences in the number of groups should not have a big effect on the model's overall performance. In this study, the three biological classifications tested were different in no more than two biological groups, and clear differences for seven- and five-group classifications were found in the overall model performance for an independent reference dataset (Fig. 4). Thus, these results pointed out the need to test alternative classifications close to the selected one during the model-building process, with the aim of ensuring selection of the best biological classification.

The development of a combined model as a first step in the MEDPACS approach arose from the

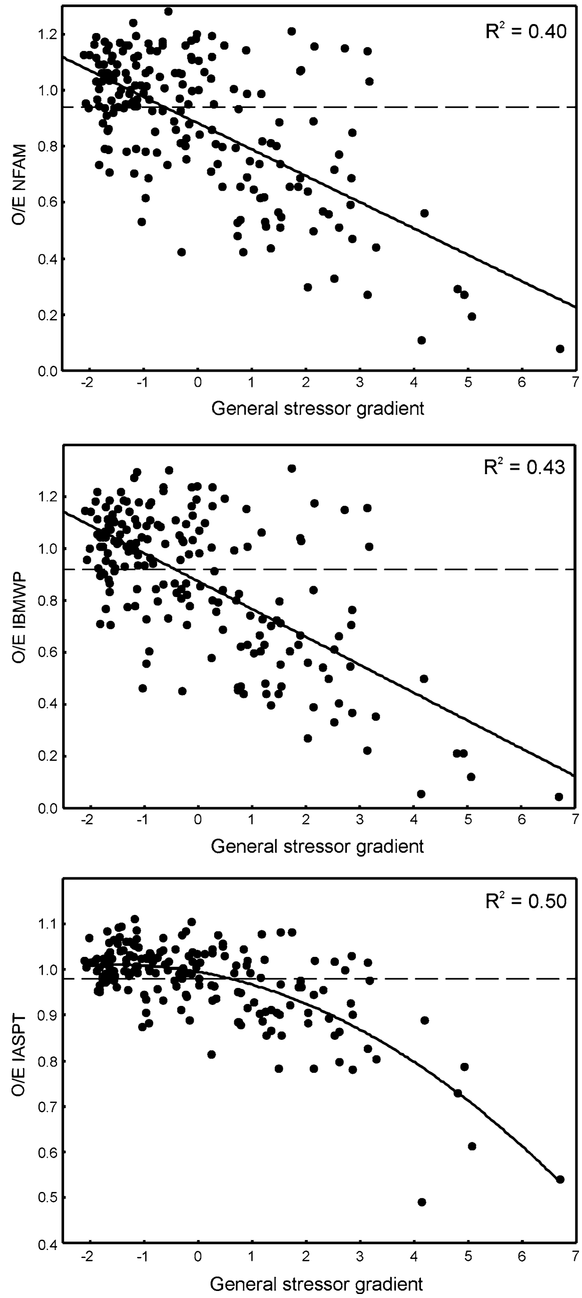

Fig. 9 Ecological indicator O/E gradients versus the general stressor gradient defined; dashed line represents the 25th percentile reference distribution value

ability of this kind of model to detect impairments along any season involved (in our case, spring, summer, and autumn). Combined models provide more precise results because they use more information about macroinvertebrate communities (Furse

Table 4 Ecological status class boundaries for each ecological indicator, as well as the corresponding number of sites catalogued in each class

Ecological status	NFAM	Number of sites	IBMWP	Number of sites	IASPT	Number of sites
High	≥0.93	103	≥0.92	102	≥0.98	113
Good	0.92–0.71	44	0.91–0.69	42	0.97–0.92	38
Moderate	0.70–0.47	29	0.68–0.46	26	0.91–0.82	26
Poor	0.46–0.24	9	0.45–0.23	12	0.81–0.69	8
Bad	0.23–0.00	3	0.22–0.00	6	0.68–0.00	3

Ranks are set up according to the REFCOND Guidance (NFAM and IBMWP) or the Munné and Prat (2006) approach (IASPT). See Materials and Methods section for details

et al., 1984; our unpublished data). One-season models are able to assess only impairments inside of the season studied. Thus, impairments produced outside the model's season would be undetected. Combined models reduce the variation among sites as the number of shared taxa increase. Moreover, they need quality control programs that involve several seasons to be applied. For all of this, the trade-off between the precision desired and the cost in the assessments of the ecological status should be taken into account. The high versatility of the predictive modeling allows the development of both combined (e.g., two, three, or four seasons) and one-season models. Thus, after new model development, the MEDPACS approach is conceived as a whole platform to assess the ecological status of Spanish Mediterranean streams, adapted to any assessment program designed by water managers.

The most common environmental variables selected by the best DF models in the model developed were large-scale variables. These types of variables are becoming widely used as tools to explain patterns in macroinvertebrate communities in biological assessments (e.g., Bis et al., 2000; Hargett et al., 2007). Usually, they are related to watershed features and geographical and climatic processes, and are also of great importance for macroinvertebrate communities (including those in Mediterranean watercourses; Clarke et al., 2003). For this reason, these types of variables are considered good choices for the development of predictive models (e.g., Hawkins et al., 2000; Kokeš et al., 2006). However, the temporal flow regime has an important role in determining macroinvertebrate communities (Caruso, 2002; Acuña et al., 2005; Bonada et al., 2006c, 2007; Dewson et al., 2007) and thereafter for the predictive modeling of such communities, as the predictive model showed for the Spanish Mediterranean watercourses.

Small-scale (site) variables are also important, such as stream chemistry (e.g., conductivity), or local features, such as a stream's morphology (width, depth). These types of variables were selected in other predictive models (e.g., Wright et al., 1993; Simpson & Norris, 2000). However, to be able to use these environmental variables, one must to be sure that no human activity can influence them (at least in the area of the predictive model implementation) because they can be very sensitive to human stressors (Clarke, 2000; Clarke et al., 2003).

Evaluation and sensitivity of the MEDPACS approach

Different sources of uncertainty related to variation in the observed fauna, as well as errors associated with the expected fauna, exist in the model building process (Clarke, 2000). Those sources could produce differences between the final O and E values. However, they could be minimized with meticulous work at each stage of the model-building process. Another source of variation arises from the reference sites themselves. According to Stoddard et al. (2006), reference sites represent the best known physical, chemical, and biological habitat conditions given today's state of the landscape. This concept implies that not all of them present the same level of non-perturbation. Thus, the O/E average for reference sites will be close to one, but roughly half of the reference sites will be better than the average, and will tend to have O/E values greater than one, and roughly half will have values less than one (Clarke, 2000). Predictive models try to explain this variation through environmental characteristics, but a residual variation remains. The less residual variation among reference sites, the more sensitive an assessment of a test site will be, and we will be able to detect the smallest deviation of the test site's biota from the expected reference condition (Bailey et al., 2004).

Predictive models with O and E regression slopes lower than one for the calibration dataset would be models that, on average, would expect more families of macroinvertebrates than observed. This would mean that O/E values would be displaced towards zero, assessing sites as non-reference when they should actually be assessed as reference (type I error). However, slopes higher than unity would mean O/E values are biased above one, assessing sites as reference when they should not be (type II error). These biases have important implications related to restoration issues. The objective of predictive models is to assess the ecological status of streams and to facilitate the establishment of the corresponding restoration programs where they are needed.

In the MEDPACS approach, the results obtained in the regression analyses performed for the calibration dataset (Fig. 6) showed a clear relationship among the final O and E values. The E values obtained explained a high percentage (73–84%) of the variability present in the Spanish Mediterranean macroinvertebrate communities. The slope values obtained were very close to one, and although the intercepts were not equal to zero, the null hypothesis tested (intercepts equals zero) was not rejected for any of the three ecological indicators. In this way, it can be considered that the MEDPACS approach is close to the ideal model. Furthermore, no differences were found among the assessments (O/E) carried out for each biological reference group; the model was thus able to assess the ecological status for any reference site with the same reliability. At the same time, the analysis of variance between the O/E values for calibration, validation, and test datasets proved the good performance of the MEDPACS approach in the assessment of independent reference and non-reference datasets (Fig. 7).

Ecological status class boundaries

The assessments carried out by the MEDPACS approach have shown a clear negative relationship with the anthropogenic pressure present in the study area (Fig. 9). This approach could be used as a reliable evaluation tool for the Spanish Mediterranean watercourses; nevertheless, some weaknesses of the technique should be improved. A high number of sites were catalogued in the upper ecological status classes for each ecological indicator (Table 4), reflecting the high proportion of reference sites used to build the corresponding O/E gradients. This fact denotes the need to include in the analysis a more extensive test dataset with a severe anthropogenic impact in order to achieve more comprehensive O/E gradients. However, the variation explained by the general stressor gradient was no more than 50% (NFAM $R^2 = 0.40$; IBMWP $R^2 = 0.43$; IASPT $R^2 = 0.50$). A refinement of the general stressor gradient should include more stressor variables in the PCA analysis. As the MEDPACS approach will be used in new test sites, and more information will be accessible about pressures on the areas, these issues will be evaluated and integrated into the predictive system, achieving more adjusted and realistic ecological status class boundaries.

Previously Alba-Tercedor and Pujante (2000) considered that the idea of developing a RIVPACS-type system for the whole of Spain was completely realistic. We have shown that the use of multivariate methodologies in the assessment of the ecological status of the Spanish Mediterranean watercourses is acceptably reliable. In addition, the combined model developed is the first step in the complete development of the MEDPACS predictive system. Further research should be addressed towards increasing the application area of the MEDPACS predictive system inside the Iberian Peninsula, as well as achieving a national predictive system to contribute to a comprehensive assessment of Spanish watercourses.

Acknowledgments The authors would especially like to acknowledge the help and advice of M.T. Furse, R.T. Clarke, R.H. Norris, S. Nichols, S. Linke, and R.C. Bailey, including the warm welcomes extended to the first author during his stays in their laboratories. We thank N. Bonada for help and comments on previous drafts of this manuscript, as well as C.P. Hawkins and J. Van Sickle for supplying the 'best-subsets' scripts, and two anonymous reviewers for their comments that contributed to the improvement of this paper. This research was supported by the GUADALMED-2 project (REN2001-3438-C07), by the Eurolimpacs project (GOCE-CT-2003-505540), as well as by pre-doctoral grants to J.M. Poquet,

T. Puntí, and M.M. Sánchez-Montoya from the Spanish Ministry of Science and Technology.

References

Acuña, V., I. Muñoz, A. Giorgi, M. Omella, F. Sabater & S. Sabater, 2005. Drought and postdrought recovery cycles in an intermittent Mediterranean stream: structural and functional aspects. Journal of the North American Benthological Society 24: 919–933.

Alba-Tercedor, J., 1996. Macroinvertebrados acuáticos y calidad de las aguas de los ríos. IV Simposio del Agua en Andalucía (SIAGA) vol II. Instituto Tecnológico Geo-Minero de España, Madrid: 203–213.

Alba-Tercedor, J. & A. M. Pujante, 2000. Running-water biomonitoring in Spain: opportunities for a predictive approach. In Wright, J. F., D. W. Sutcliffe & M. T. Furse (eds), Assessing the Biological Quality of Fresh waters: RIVPACS and Other Techniques. Freshwater Biological Association, Ambleside, Cumbria: 207–216.

Alba-Tercedor, J. & A. Sánchez-Ortega, 1988. Un método rápido y simple para evaluar la calidad biológica de las aguas corrientes basado en el de Hellawell. Limnetica 4: 51–56.

Alba-Tercedor, J., P. Jáimez-Cuellar, M. Álvarez, J. Avilés, N. Bonada, J. Casas, A. Mellado, M. Ortega, I. Pardo, N. Prat, M. Rieradevall, S. Robles, C. E. Sáinz-Cantero, A. Sánchez-Ortega, M. L. Suárez, M. Toro, M. R. Vidal-Abarca, S. Vivas & C. Zamora-Muñoz, 2004. Caracterización del estado ecológico de ríos mediterráneos ibéricos mediante el índice IBMWP (antes BMWP'). Limnetica 21(2002): 175–185.

Armengol, J., N. Prat & N. Bonada (eds), 2004. Resultados del proyecto GUADALMED. Limnetica 21 (2002). Publicacions i Edicions Universitat de Barcelona. Barcelona.

Armitage, P. D., I. Pardo, M. T. Furse & J. F. Wright, 1990. Assessment and prediction of biological quality. A demonstration of a British macroinvertebrate based method in two Spanish rivers. Limnetica 6: 147–156.

Bailey, R. C., R. H. Norris & T. B. Reynoldson, 2004. Bioassessment of Freshwater Ecosystems: Using the Reference Condition Approach. Kluwer Academic Publishers, Dordrecht.

Belbin, L. & C. McDonald, 1993. Comparing three classification strategies for use in ecology. Journal of Vegetation Science 4: 341–348.

Bis, B., A. Zdanowicz & M. Zalewski, 2000. Effects of catchment properties on hydrochemistry, habitat complexity and invertebrate community structure in a lowland river. Hydrobiologia 422/423: 369–387.

Bonada, N., N. Prat, A. Munné, M. Plans, C. Solà, M. Álvarez, I. Pardo, G. Moyá, G. Ramón, M. Toro, S. Robles, J. Avilés, M. L. Suárez, M. R. Vidal-Abarca, A. Mellado, J. L. Moreno, C. Guerrero, S. Vivas, M. Ortega, J. Casas, A. Sánchez-Ortega, P. Jáimez-Cuellar & J. Alba-Tercedor, 2004a. Intercalibración de la metodología GUADALMED Selección de un protocolo de muestreo para la determinación del estado ecológico de los ríos mediterráneos. Limnetica 21(2002): 13–33.

Bonada, N., N. Prat, A. Munné, M. Rieradevall, J. Alba-Tercedor, M. Álvarez, J. Avilés, J. Casas, P. Jáimez-Cuellar, A. Mellado, G. Moyá, I. Pardo, S. Robles, G. Ramón, M. L. Suárez, M. Toro, M. R. Vidal-Abarca, S. Vivas & C. Zamora-Muñoz, 2004b. Criterios para la selección de condiciones de referencia en los ríos mediterráneos. Resultados del proyecto GUADALMED. Limnetica 21(2002): 99–114.

Bonada, N., C. Zamora-Muñoz, M. Rieradevall & N. Prat, 2005. Ecological and historical filters constraining spatial caddisfly distribution in Mediterranean rivers. Freshwater Biology 50: 781–797.

Bonada, N., H. Dallas, M. Rieradevall & N. Prat, 2006a. A comparison of rapid bioassessment protocols used in 2 regions with Mediterranean climates, the Iberian Peninsula and South Africa. Journal of North American Benthological Society 25: 487–500.

Bonada, N., N. Prat, V. H. Resh & B. Statzner, 2006b. Developments in aquatic insect biomonitoring: a comparative analysis of recent approaches. Annual Review of Entomology 51: 495–523.

Bonada, N., M. Rieradevall, N. Prat & V. H. Resh, 2006c. Benthic macroinvertebrate assemblages and macrohabitat connectivity in Mediterranean-climate streams of northern California. Journal of North American Benthological Society 25: 32–43.

Cao, Y. & D. D. Williams, 1999. Rare species are important in bioassessment (Reply to the comment by Marchant). Limnology and Oceanography 44: 1841–1842.

Cao, Y., D. D. Williams & N. E. Williams, 1998. How important are rare species in aquatic community ecology and bioassessment? Limnology and Oceanography 43: 1403–1409.

Cao, Y., D. P. Larsen & R. S. J. Thorne, 2001. Rare species in multivariate analysis for bioassessment: some considerations. Journal of North American Benthological Society 20: 144–153.

Caruso, B. S., 2002. Temporal and spatial patterns of extreme low flows and effects on stream ecosystems in Otago, New Zealand. Journal of Hydrology 257: 115–133.

Chessman, B. C., I. Growns, J. Curreys & N. Plunkett-Cole, 1999. Predicting diatom communities at the genus level for the rapid biological assessment of rivers. Freshwater Biology 41: 317–331.

Clarke, R. T., 2000. Uncertainty in estimates of biological quality based on RIVPACS. In Wright, J. F., D. W. Sutcliffe & M. T. Furse (eds), Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques. Freshwater Biological Association, Ambleside, Cumbria: 39–54.

Clarke, R. T. & J. F. Murphy, 2006. Effects of locally rare taxa on the precision and sensitivity of RIVPACS bioassessment of freshwaters. Freshwater Biology 51: 1924–1940.

Clarke, R. T., M. T. Furse, J. F. Wright & D. Moss, 1996. Derivation of a biological quality index for river sites: comparison of the observed with the expected fauna. Journal of Applied Statistics 23: 311–332.

Clarke, R. T., J. F. Wright & M. T. Furse, 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. Ecological Modelling 160: 219–233.

Davis, J., P. Horwitz, R. H. Norris, B. C. Chessman, M. McGuire & B. Sommer, 2006. Are river bioassessment methods using macroinvertebrates applicable to wetlands? Hydrobiologia 572: 115–128.

De Pauw, N., W. Gabriels & P. L. M. Goethals, 2006. River monitoring and assessment methods based on macroinvertebrates. In Ziglio, G., M. Siligardi & G. Flaim (eds), Biological Monitoring of Rivers: Applications and Perspectives. Wiley, Chichester: 113–134.

Dewson, Z. S., A. B. W. James & R. G. Death, 2007. A review of the consequences of decreased flow for instream habitat and macroinvertebrates. Journal of the North American Benthological Society 26: 401–415.

European Commission, 2000. Directive 2000/60/EC of the European Parliament and the Council of 23rd October 2000 establishing a framework for community action in the field of water policy. Official Journal of the European Communities L327: 1–72.

European Commission, 2003. Common implementation strategy for the water framework directive (2000/60/EC). Guidance Document No. 10, Rivers and lakes—typology, reference conditions and classification systems. Office for official publications of the European Communities, Luxembourg.

Feio, M. J., S. F. P. Almeida, S. C. Craveiro & A. J. Calado, 2007a. Diatoms and macroinvertebrates provide consistent and complementary information on environmental quality. Fundamental and Applied Limnology 168: 247–258.

Feio, M. J., T. B. Reynoldson, V. Ferreira & M. A. S. Graça, 2007b. A predictive model for freshwater bioassessment (Mondego River, Portugal). Hydrobiologia 589: 55–68.

Ferréol, M., A. Dohet, H.-M. Cauchie & L. Hoffmann, 2008. An environmental typology of freshwater sites in Luxembourg as a tool for predicting macroinvertebrate fauna under non-polluted conditions. Ecological Modelling 212: 99–108.

Furse, M. T., D. Moss, J. F. Wright & P. D. Armitage, 1984. The influence of seasonal and taxonomic factors on the ordination and classification of running-water sites in Great Britain and on the prediction of their macro-invertebrate communities. Freshwater Biology 14: 257–280.

Gasith, A. & V. H. Resh, 1999. Streams in Mediterranean climate regions: abiotic influences and biotic responses to predictable seasonal events. Annual Review of Ecology and Systematics 30: 51–81.

Gómez, R., I. Hurtado, M. L. Suárez & M. R. Vidal-Abarca, 2005. Ramblas in south-east Spain: threatened and valuable ecosystems. Aquatic Conservation: Marine and Freshwater Ecosystems 15: 387–402.

Hargett, E. G., J. R. ZumBerge, C. P. Hawkins & J. R. Olson, 2007. Development of a RIVPACS-type predictive model for bioassessment of wadeable streams in Wyoming. Ecological Indicators 7: 807–826.

Hawkins, C. P., R. H. Norris, J. N. Hogue & J. W. Feminella, 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecological Applications 10: 1456–1477.

Hering, D., C. K. Feld, O. Moog & T. Ofenböck, 2006. Cook book for the development of a Multimetric Index for biological condition of aquatic ecosystems: experiences from the European AQEM and STAR projects and related initiatives. Hydrobiologia 566: 311–324.

Ihaka, R. & R. Gentleman, 1996. R: a language for data analysis and graphics. Journal of Computational and Graphical Statistics 5: 239–314.

Jáimez-Cuellar, P., S. Vivas, N. Bonada, S. Robles, A. Mellado, M. Álvarez, J. Avilés, J. Casas, M. Ortega, I. Pardo, N. Prat, M. Rieradevall, C. E. Sáinz-Cantero, A. Sánchez-Ortega, M. L. Suárez, M. Toro, M. R. Vidal-Abarca, C. Zamora-Muñoz & J. Alba-Tercedor, 2004. Protocolo GUADALMED (PRECE). Limnetica 21(2002): 187–204.

Johnson, R. K. & L. Sandin, 2001. Development of a prediction and classification system for lake (Littoral, $SWEPAC_{LLI}$) and Stream (Riffle, $SWEPAC_{SRI}$) macroinvertebrate communities. Stencil, Department of Environmental Assessment, SLU, Uppsala.

Joy, M. K. & R. G. Death, 2002. Predictive modelling of freshwater fish as a biomonitoring tool in New Zealand. Freshwater Biology 47: 2261–2275.

Kennard, M. J., B. J. Pusey, A. H. Arthington, B. D. Harch & S. J. Mackay, 2006. Development and application of a predictive model of freshwater fish assemblage composition to evaluate river health in eastern Australia. Hydrobiologia 572: 33–57.

Kokeš, J., S. Zahrádková, D. Němejcová, J. Hodovský, J. Jarkowský & T. Soldán, 2006. The PERLA system in the Czech Republic: a multivariate approach for assessing the ecological status of running waters. Hydrobiologia 566: 343–354.

Köppen, W., 1923. De klimate der Erde. Bornträger, Berlin.

Legendre, P. & L. Legendre, 1998. Numerical ecology, 2nd ed. Elsevier, Amsterdam.

Linke, S., R. H. Norris, D. P. Faith & D. Stockwell, 2005. ANNA: a new prediction method for bioassessment programs. Freshwater Biology 50: 147–158.

Marchant, R., 1999. How important are rare species in aquatic community ecology and bioassessment? A comment on the conclusions of Cao et al. Limnology and Oceanography 44: 1840–1841.

McElravy, E. P., G. A. Lamberti & V. H. Resh, 1989. Year-to-year variation in the aquatic macroinvertebrate fauna of northern California stream. Journal of North American Benthological Society 8: 51–63.

Moss, D., M. T. Furse, J. F. Wright & P. D. Armitage, 1987. The prediction of the macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. Freshwater Biology 17: 41–52.

Moss, D., J. F. Wright, M. T. Furse & R. T. Clarke, 1999. A comparison of alternative techniques for prediction of the fauna of running-water sites in Great Britain. Freshwater Biology 41: 167–181.

Munné, A. & N. Prat, 2006. Comparing quantitative and qualitative metrics based on macroinvertebrates to measure biological quality and define reference conditions in mediterranean rivers types. In Libro de resumenes del XIII Congreso de la Asociación Española de Limnología—V Congreso Ibérico de Limnología. Asociación Española de Limnología. Barcelona: 61.

Munné, A., N. Prat, C. Solà, N. Bonada & M. Rieradevall, 2003. A simple field method for assessing the ecological quality of riparian habitat in rivers and streams: QBR

index. Aquatic Conservation: Marine and Freshwater Ecosystems 13: 147–163.
Niemi, G. J. & M. E. McDonald, 2004. Application of ecological indicators. Annual Review of Ecology, Evolution and Systematics 35: 89–111.
Parsons, M. & R. H. Norris, 1996. The effect of habitat specific sampling on biological assessment of water quality using a predictive model. Freshwater Biology 36: 419–434.
Quinn, G. P. & M. J. Keough, 2002. Experimental Design and Data Analysis for Biologists. Cambridge University Press, Cambridge.
Resh, V. H., J. F. Jackson & E. P. McElravy, 1990. Disturbance, annual variability, and lotic benthos: examples from a California stream influenced by a Mediterranean climate. Memorie dell'Istituto Italiano di Idrobiologia 47: 309–329.
Reynoldson, T. B., R. C. Bailey, K. E. Day & R. H. Norris, 1995. Biological guidelines for freshwater sediment based on Benthic Assessment of Sediment (the BEAST) using a multivariate approach for predicting biological state. Australian Journal of Ecology 20: 198–219.
Reynoldson, T. B., R. H. Norris, V. H. Resh, K. E. Day & D. M. Rosenberg, 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. Journal of North American Benthological Society 16: 833–852.
Sánchez-Montoya, M. M., T. Puntí, M. L. Suárez, M. R. Vidal-Abarca, M. Rieradevall, J. M. Poquet, C. Zamora-Muñoz, S. Robles, M. Álvarez, J. Alba-Tercedor, M. Toro, A. M. Pujante, A. Munné & N. Prat, 2007. Concordance between ecotypes and macroinvertebrate assemblages in Mediterranean streams. Freshwater Biology 52: 2240–2255.
Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6: 461–464.
Simpson, J. & R. H. Norris, 2000. Biological assessment of water quality: development of AUSRIVAS models and outputs. In Wright, J. F., D. W. Sutcliffe & M. T. Furse (eds), Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques. Freshwater Biological Association, Ambleside, Cumbria: 125–142.
StatSoft, 2005. STATISTICA (data analysis software system), version 7.1. www.statsoft.com.
Statzner, B., S. Dolédec & B. Hugueny, 2004. Biological trait composition of European stream invertebrate communities: assessing the effects of various trait filter types. Ecography 27: 470–488.
Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson & R. H. Norris, 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. Ecological Applications 16: 1267–1276.
Van Sickle, J., C. P. Hawkins, D. P. Larsen & A. H. Herlihy, 2005. A null model for the expected macroinvertebrate assemblage in streams. Journal of the North American Benthological Society 24: 178–191.
Van Sickle, J., D. D. Huff & C. P. Hawkins, 2006. Selecting discriminant function models for predicting the expected richness of aquatic macroinvertebrates. Freshwater Biology 51: 359–372.
Whittingham, M. J., P. A. Stephens, R. B. Bradbury & R. P. Freckleton, 2006. Why do we still use stepwise modelling in ecology and behaviour? Journal of Animal Ecology 75: 1182–1189.
Wright, J. F., P. D. Armitage, M. T. Furse & D. Moss, 1984. The classification of sites on British rivers using macroinvertebrates. Verhandlungen International Verein Limnology 22: 1939–1943.
Wright, J. F., M. T. Furse & P. D. Armitage, 1993. RIVPACS—a technique for evaluating the biological quality of rivers in the UK. European Water Pollution Control 3: 15–25.
Wright, J. F., D. W. Sutcliffe & M. T. Furse (eds), 2000. Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques. Freshwater Biological Association, Ambleside.
Zamora-Muñoz, C. & J. Alba-Tercedor, 1996. Bioassessment of organically polluted Spanish rivers, using a biotic index and multivariate methods. Journal of the North American Benthological Society 15: 332–352.
Zamora-Muñoz, C., C. E. Sáinz-Cantero, A. Sánchez-Ortega & J. Alba-Tercedor, 1995. Are biological indices BMWP' and ASPT' and their significance regarding water quality seasonally dependent? Factors explaining their variations. Water Research 29: 285–290.