Environmental Monitoring and Assessment (2005) **xxx:** 1–21

COMBINING LOGISTIC MODELS WITH MULTIVARIATE METHODS FOR THE RAPID BIOLOGICAL ASSESSMENT OF RIVERS USING MACROINVERTEBRATES

S. V. OLIVEIRA and R. M. V. CORTES*
Universidade de Trás-os-Montes e Alto Douro, Departamento Florestal, Apartment 1013, Vila Real, Portugal
*(*author for correspondence, e-mail: rcortes@utad.pt)*

(Received 4 March 2004; accepted 10 January 2005)

Abstract. This work represents an attempt to define a simple method to classify the relative degree of disturbance of sites in lotic systems on the basis of comparison of their faunistic composition with reference sites. Two ecotypes were selected in northern Portugal where benthic invertebrates were sampled in reaches with different levels of contamination. As a first stage, previous Geographic Information System information was used to define reference sites in each ecotype. Afterwards, multivariate techniques and non linear estimation models were combined to assess biological quality. This method allowed us to quantify sites according to increasing levels of contamination, after the probabilities of occurrence of taxa along a gradient of contamination taking into account the reference condition. The results suggest that this method is sensitive to organic pollution, easy to interpret, namely the species tolerance, and could be a good framework to establish regional rankings depending on the ecological impact of river sites.

Keywords: benthic communities, biomonitoring, ecotypes, logistic models, multivariate techniques, reference sites

1. Introduction

Different procedures have been developed to assess the quality of running waters based on the prediction of the composition of macroinvertebrate taxa using multivariate techniques, where observed taxa (O) is compared with the expected taxa (E) for each location. Methods like RIVPACS (Wright *et al.*, 1984; Wright, 2000), AUSRIVAS (Parsons and Norris, 1996) and BEAST (Reynoldson *et al.*, 1995, 1997) use these principles, since the expected fauna is derived from the environmental characteristics (geological, physical and chemical factors) of the area using a set of reference sites which are believed to be unpolluted or unstressed. Such multivariate methods require extensive sets of biotic and abiotic data, covering all types of ecological conditions and gathered seasonally. For instance, RIVPACS III currently draws actually on information from 614 reference sites belonging to 35 site groups within a wide range of running water sites across Great Britain, for which there is a historical record of invertebrate fauna (Wright, 2000). As is pointed out by Chessman (1999), when establishing an overall environmental difference to be used for prediction, it is necessary to rely on the parameters that are most closely associated with faunal differences. For instance, Sloane *et al.* (1997) attributed the reduction in O/E taxa at undisturbed test sites to drought conditions and not to contamination factors. The mentioned techniques are highly demanding since a wide range of environmental and biological data must be collected over a considerable period of time to establish accurate relationships and to distinguish between natural and human impacts. The reference condition approach is well-suited for large-scale biomonitoring programs (e.g. in UK, Australia and Canada), where the multivariate methods are extensively used, and can rely on the assemblages that characterize such sites, or on their geophysical attributes (rapid assessment protocols or multimetric systems, Barbour *et al.*, 1995).

Because often aquatic ecosystems can rapidly recover from most perturbations induced by humans (Karr *et al.*, 1996), a finding that biological integrity is being achieved not only reflects a current healthy condition but also means that the community has withstood and recovered from short-term stresses (Simon, 1999). Therefore metrics have to be carefully selected and validated in order to be sensitive to the environmental condition being monitored, a task that demands considerable work together with the definition of the reference condition. For instance, in France, Oberdorff *et al.* (2002) uses extensive data sets to develop a fish-based index employing classes of metrics derived from predictive models reflecting assemblage structure and function: 650 reference sites evenly distributed across the whole country were defined, covering a period of 13 years.

However, in countries like Portugal, the need to assess the degree of disturbance in specific reaches or catchments can rarely make use of previous biological inventories, since these are generally scattered in time and space and were established for specific purposes. The biomonitoring approach present in this work is then designed to assess the biological quality of river sites, and aims to compare the observed fauna with the expected or “target” fauna in geographical areas where continuous historical information is relatively scarce. By applying multivariate techniques and non linear estimation models this study intends to assess biological quality for local surveys using quality information based on the prediction of taxa, for a large geographical area that includes practically all northern Portugal. In contrast with RIVPACS and similar techniques a considerable simplification is introduced here. Instead of deriving indirectly the probability of occurrence for the benthic assemblages from a group of environmental parameters (like dissolved oxygen, pH, alkalinity, etc.) through discriminant functions, which require extensive surveys, probability is derived directly by logistic models from a gradient of contamination expressed by an ordination axis. For the definition of the reference situations we used a similar criteria to those indicated by Rosemberg *et al.* (2000) for whom local knowledge and expertise, published information or simple reconnaissance trips are possible tools. We followed this quicker and more practical procedure by using all the previous information related to the stream condition in northern Portugal (Cortes, 1992; Cortes *et al.*, 2002a,b) to define such “theoretical” reference sites. In the present case, the available environmental information was used to help in the definition of the ecotypes and the unstressed sites, in a similar way to the multimetric systems (Barbour *et al.*, 1995).

The development and refinement of bioassessment programs can become rather expensive before they become cost-effective. The development of regional methods, metrics and reference conditions requires an organized and well-thought-out design and the cost of obtaining information increases with its inherent accuracy, namely in the ability to quantify non-point source disturbances (Hughes, 1995; Barbour *et al.*, 1996). In conclusion, without losing accuracy, we tried to develop a cost efficient method designed for regional assessment purposes.

2. Materials and Methods

2.1. Sampling sites and invertebrate sampling

The surveyed area contained all the main rivers of northern Portugal (except the river Minho which constitutes the natural border with Spain).

The sites were grouped in two ecotypes, the partition entirely justified by a previous discriminant function analysis (DFA) from habitat variables adapted from the River Habitat Survey (Raven *et al.*, 1998) – Table I. This classification presents fewer cases misclassified in comparison with six other alternative geographical and typological variables. These ecotypes, which were then treated separately for biological assessment, included: (a) the most northwestern catchments of the rivers Lima, Cávado and Ave, smaller and with a more Atlantic influence; (b) the Portuguese part of the Douro catchment, clearly more extensive and with more continental characteristics, giving a more irregular flow throughout the year. This division agrees also with the classification of the river districts, which has lead to the development of two catchment plans that overlap such river types. Only stations

TABLE I

Results of discriminant analysis of different classification systems for habitat variables adapted from the River Habitat Survey (Raven *et al.*, 1998)

Classification model	Multivariate *p*-value	Discriminant misclassification rate (%)
Altitude (≤500 m/>500 m)	0.0202	13
Catchment area (<100 km^2/100–1000 km^2/≥1000 km^2)	0.0000	12
Mean catchment slope (≤0.5/>0.5)	0.2022	18
Latitude (<500000 m/≥500000 m)	0.0011	10
Stream order (1–3/4–6)	0.0005	6
Mean daily air temperature (≤12.5 °C/≥15 °C)	0.1053	24
Ecotype (Northwest/Douro)	0.0000	0

in the rhithron were considered to make possible the comparisons inside the same typological level (this study is a part of a more extensive work but the potamon sites were discarded). Thus, 31 sampling sites were selected in the NW streams and 45 in the Douro catchment. The characteristics of these streams and respective drainage basins have already been reported elsewhere (e.g. Cortes, 1992; Cortes *et al.*, 2002a).

Invertebrate collections took place during the end of spring/summer 1999 using a 350 μm mesh net with a constant sampling time of 4 min (CPUE) and with an effort proportional to the size of each habitat. The contents of the hand-net were preserved in 4% formol and later sorted at the laboratory. Wherever possible, the organisms were identified to species level (except Diptera and Oligochaeta).

The decision about the reference sites was based on a classification coupling biological and physical features (Cortes *et al.*, 2002b) for the entire drainage network. That method required a previous classification of the geomorphic units (segments) using stream order, geology, topography and precipitation. The state of conservation of those units was then classified by overlapping seven primary data layers through Geographical Information Systems (GIS): biotic index, structure of the riparian corridors, number of exotic and native fish species, water quality and contamination load from urban and industrial sources. The reference condition considered the most pristine geomorphic units given by such a classification and was further statistically tested (see below).

2.2. Statistical analysis

All procedures used to assess the environmental disturbance of river sites are schematically illustrated in Figure 1 and will be detailed in the following paragraphs. The data sets from the different catchments were joined in two groups according to the referred ecotypes (Figure 1, step 1): one grouped data from the NW catchments, of smaller size, and the other included only the Douro catchment. Data was ordinated through detrended correspondence analysis (DCA), a variation of correspondence analysis (Figure 1, step 3), a metric technique using a "line of best-fit" approach from sample points. Underlying this method is the idea that species commonly exhibit bell-shaped response curves with respect to environmental gradients. This has already been described by Whittaker (1967) for terrestrial species of flora, but these unimodal curves are just as appropriate for describing the responses of benthic (Peeters and Gardeniers, 1998) or fish species (Oberdorff *et al.*, 2001). DCA axes represent indeed measures of species tolerance since these exhibit Gaussian response curves, expressed as standard deviation units, along those axes. For instance, sites that differ by four standard units in scores are expected to have no species in common (Jongman *et al.*, 1987). Data was previously log transformed ($\log x + 1$).

In order to select the reference stations based on environmental data the *k*-means classification was overlapped on the ordination (Figure 1, step 5). This is a

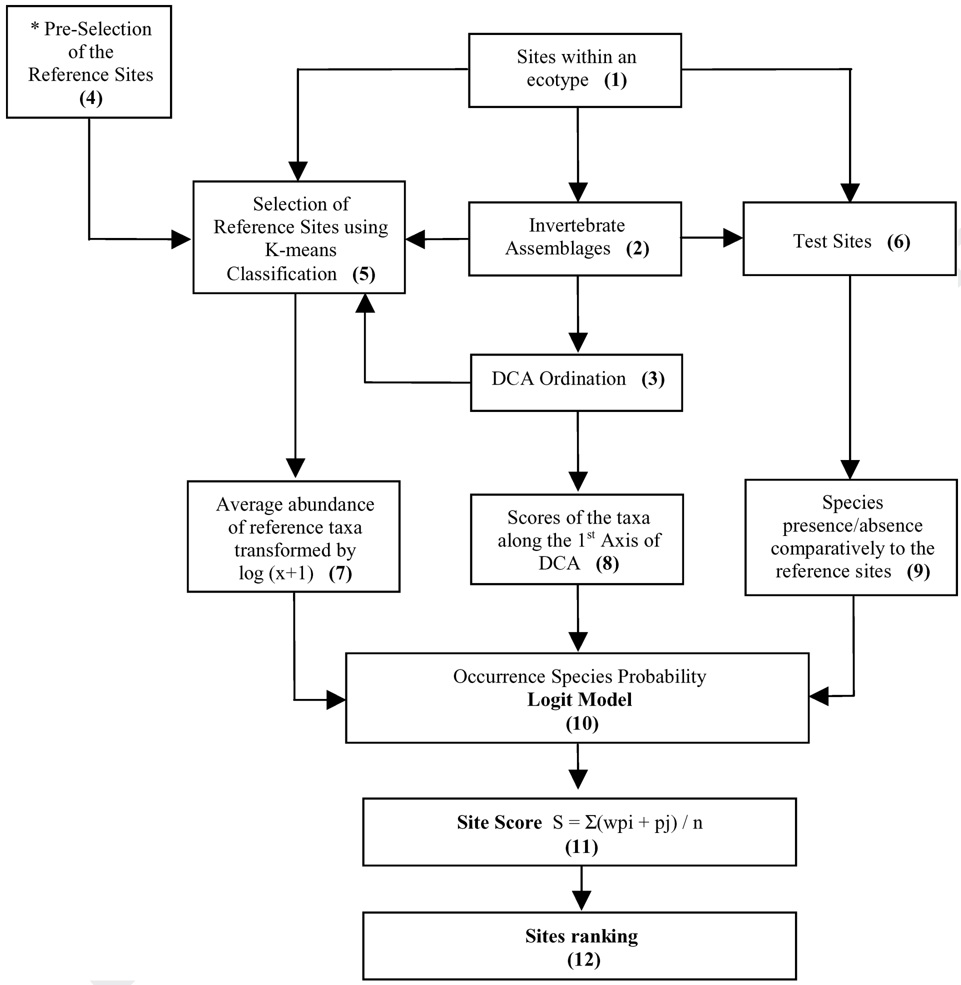

Figure 1. Main steps in the development of the model to assess the environmental disturbance of river sites. The numbers in brackets provide a guide to the order of the sequence process. *Pre-selection of the reference sites was based on biological and physical features.

non-hierarchical procedure with *a priori* specification of the number of classes (*k* clusters), where the objects are successively relocated in order to decrease the sum of the squared Euclidean distances of the different objects from the respective centroid (Podani, 2000). The initial partition was previously specified by predetermined objects representing our previous knowledge of the most disturbed and non-disturbed sites after the treatment of the information of all the drainage network by GIS (Figure 1, step 4), and data was previously standardized to the unit of variance. While DCA provided the representation of samples in a two-dimensional map, *k*-means clustering allowed aggregating those sites into distinct clusters, by splitting the ordination into groups representing different impact levels, which also allowed to group apart the reference sites.

Logistic regressions were used to predict the probability of occurrence of the different taxa, relating a binary response variable (presence or absence in the reference sites) and a quantitative explanatory variable representing the main environmental gradient. Since the data included a wide range of sites, from unpolluted reaches to eutrophic ones, the first axis of DCA theoretically should reflect a gradient of nutrient enrichment. This assumption was tested through the comparison of the sites position in the first axis with the previous information given by GIS and by a Pearson correlation with conductivity assuming that this variable expresses a gradient of contamination. Consequently, the explanatory variable was defined by this gradient, that is, by the scores of the taxa along this axis (instead of individual variables), since it could potentially summarize the global ecological conditions (Figure 1, step 8). The logistic model therefore estimates the probability of a certain species being present $p(x)$ in a site from a particular value of the environmental variable, in this case the gradient expressed by the first axis of the ordination (Figure 1, step 10). This probability is described by a curve, where the predicted values are between 0 and 1, so this is a restriction on the general logistic model and it is more correctly designated by logit model. We considered that the sigmoid curve displayed by this model was the most convenient representation, considering that the dispersion of sites along the DCA axes rarely exceeds two units of standard deviation in this type of stream (considering the centroids between polluted and reference sites). If the dispersion were larger (e.g. four units) the Gaussian logit curve would probably be more appropriate (ter Braak and Looman, 1996) allowing to estimate the optimum and tolerance of each species. Therefore, the model used is entirely in agreement with the idea of the unimodal distribution of species assumed by DCA, but it also takes into account the general length of the gradient.

The considered curve is represented by the expression:

$$p = \frac{\exp(b_0 + b_1 x)}{1 + \exp(b_0 + b_1 x)}$$

where b_0 and b_1 are the curve parameters and the part $(b_0 + b_1 x)$ is termed the linear predictor and x represents the independent variable. Two logit models were then used to predict taxa separately for Douro and the NW catchments. The ordinary least-squares regression cannot be used to estimate the parameters because the errors are not distributed in a normal way and have no constant variance (Jongman *et al.*, 1987); thus, the maximum likelihood criterion was applied to estimate the parameters of this regression. The term logit stems from the fact that one can easily linearize this model via the logit transformation: one can transform the probability p as: $p' = \log_e[p/(1-p)]$, therefore this term is also called logit link. This model then falls within the general framework of generalized linear models. We should mention however that, in this case, the independent variable actually consists of two variables: one which contains the binary codes indicating species presence/absence relative to the considered reference sites (Figure 1, step 9), and a variable that contains the counts (Figure 1, step 7), that is, the average abundance of reference taxa (transformed by log $(x + 1)$). These are defined as the species that appear at least in one of the reference sites and the calculation of their averages refers only to these reference sites. In this way, the data file set up may be considered as a cross tabulation table of the scores of species on the first axis of the ordination by their presence (absence) in the undisturbed sites, where an additional variable contains the averages of the taxa in these sites. As it can be seen, this procedure means that even very large inventories can be summarized in a relatively small file.

We must emphasize that the use of logit models corrects and standardizes the species values along DCA axes (representing an ecological gradients) and allows to achieve an easy interpretation of the tolerance status of the different taxa. In fact, these values range between 0 and 1 (species present or absent in the reference sites were respectively classed as 0 or 1), getting closer to this number as the disturbance increases. Probability values for each species must not be confounded with the prediction ability of the model, which can be assessed by the regression residuals (observed minus predicted values).

In order to make the selection of the intolerant species less dependent on rare occurrences and thereby to select bio-indicators less influenced by local factors, the predicted values for the species in the reference sites were divided by the number of sites where they occur, which represents a weighting procedure (taxa with more clear preference for these sites acquires then lower values). Finally, after fitting the logit model, the scores of the sites (S) can be calculated (Figure 1, step 11):

$$S = \sum \frac{wp_i + p_j}{n}$$

where wp_i represents the weighted predicted value (probability) for species present in reference sites, p_j the predicted value for remaining species, and n the total number of species in each site.

This procedure makes it possible to rank and to attribute values to the sites according to the predicted values of the species. Therefore, as far as the test sites differ in the S value relative to the reference sites, the more polluted they are (as a larger proportion of the taxa is absent from the reference sites). The S values are grouped by five quality classes (Figure 1, step 12) with the class 1 corresponding to wpthe optimal condition and 5 to the most impaired site (Table II). To validate this classification a Pearson correlation was made between selected environmental variables that are related to human disturbance and all the site scores (S) of both basins. These variables were adapted from Raven *et al.* (1998) with the exception of QBR (an index that characterizes the structure of the riparian vegetation), which was adapted from Munné *et al.* (1998).

TABLE II
Five-point scale for indicator measurements. Quality classes with their scoring amplitudes and the attributes of those classes

Quality class of site	Scores (*S*)	Attributes
I	<0.2	Highest quality, comparable to the best situations without human disturbance
II	0.2–0.4	Good health
III	0.4–0.6	Fair
IV	0.6–0.8	Poor. High levels of contamination
V	>0.8	Unhealthy. Highest levels of contamination

Data was analysed through STATISTICA 6.0 (StatSoft, 2001), for the classification (*k*-means) and logistic regression and CANOCO (ter Braak and Smilauer, 1998) for ordination (DCA).

3. Results

The first axes of both ordinations (eigenvalues: 0.6160 for the NW catchments and 0.3887 for the Douro catchment) display clear gradients of disturbance: disturbed and non-disturbed sites are displayed apart on the edges of these axes (Figures 2 and 3). Thus, the disturbed sites were located in basins draining urban and industrial areas or with severe alteration of the physical habitat. The main exception is Cávado 2, whose separation is dictated by an assemblage that integrates a few rare species. Correlation analysis showed that first axis of DCA and conductivity was strongly correlated in both the ecotypes ($p < 0.05$). This supports the assumption that such axis represents the main gradient of eutrophication, since conductivity is generally an adequate water quality variable expressing nutrient concentration (Webb and Walling, 1996; Ladson *et al.*, 1999; Griffith *et al.*, 2002). Therefore, we can apply the proposed methodology that fits the species scores along this axis according to the logit model.

Classification departure for the *k*-means clustering was subjectively created from a pre-defined number of 11 clusters for the NW catchments and 12 for the Douro basin. These groups arose from the GIS analyses of the biological and physical characteristics of the watersheds. Results from both techniques (multivariate analysis and GIS) were also compared in order to extract reference sites that exhibited a consistent character. Thus, the reference sites established were:

- NW catchments: Ázere 1, Homem 1, Adrão 1, Vez 1, Estorãos 1, Carcerelha 1, Tamente 1.
- Douro catchments: Tinhela 1, Tuela 2, Bessa 1, Baceiro 1, Tuela 1, Pinhão 1, Balsemão 1.

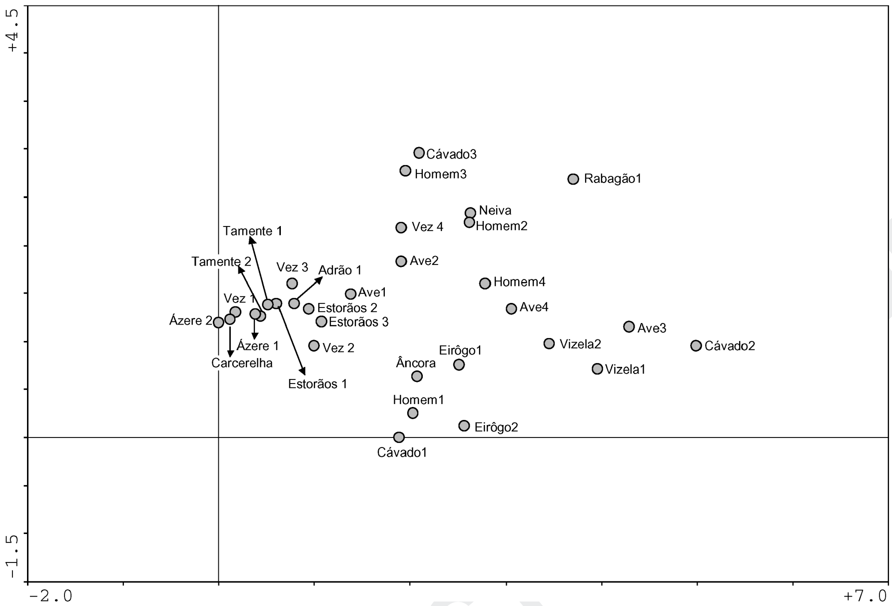

Figure 2. DCA ordination of sites data of the NW catchments (notice that arrows are only represented for a better perception of sites labeling).

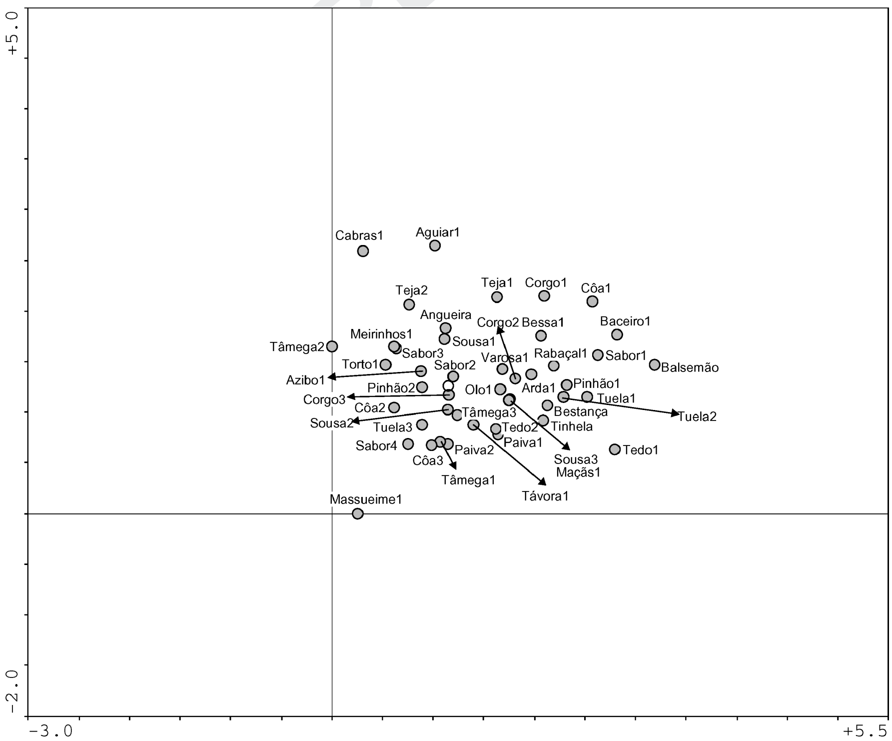

Figure 3. DCA ordination of sites data of the Douro catchment (arrows are linked to site labeling).

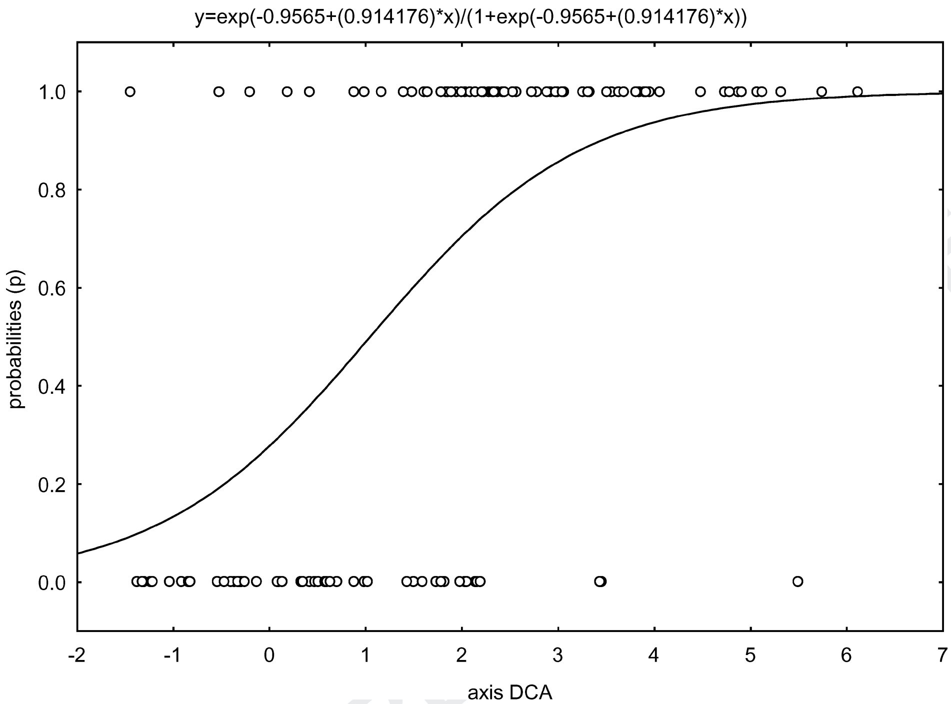

Figure 4. Application of logit model for the NW catchments. The dots represent the observed taxa (with values 0 or 1 according to the absence or presence in the reference sites) and the curve the fitted model that sets the predictions for each taxon.

Fitting the logit model allowed us to estimate the regression parameters, which were all significant for a $p < 0.05$ level (Figures 4 and 5). Consequently, the models were:

$$\text{NW catchment}: \; p = \frac{\exp(-0.9565 + 0.9142x)}{1 + \exp(-0.9565 + 0.9142x)}$$

$$\text{Douro catchment}: \; p = \frac{\exp(0.6880 - 0.8651x)}{1 + \exp(0.6880 - 0.8651x)}$$

Q1

The final loss was, respectively, 92.1760 and 141.1058, considering that the regression coefficients were estimated in order to maximize the likelihood function.

Table III displays the 10 most tolerant and intolerant taxa for the two groups of catchments considered and their predicted values. Both ecotypes have in common *Protonemura* sp. and *Aphelocheirus occidentalis*, displaying a high sensitivity to pollution. Trichoptera constitutes the dominant group for this type of taxa. There are no tolerant taxa that are shared by those groups, and the most representative

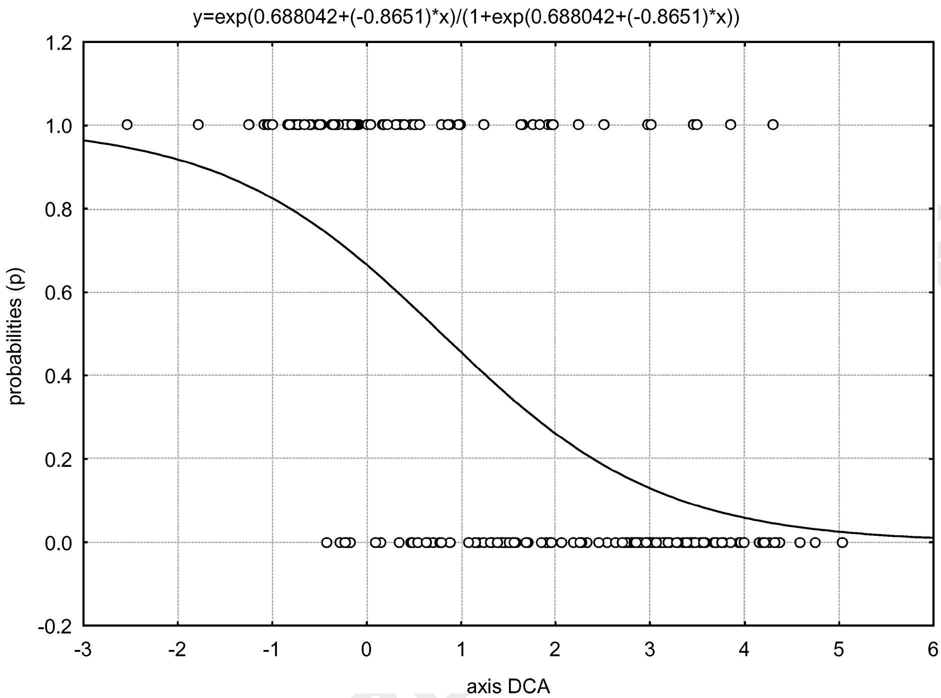

Figure 5. Application of logit model for the Douro catchment. The dots represent the observed taxa (with values 0 or 1 according to the absence or presence in the reference sites) and the curve the fitted model that sets the predictions for each taxon.

members are included in the orders of Hirudinea, Gastropoda (NW catchments) and Diptera (Douro catchment).

Figure 6 presents, graphically, the values computed for each site on the basis of the predicted values for all the taxa, ranked according to progressive levels of contamination. Contamination does not necessarily increase continuously along the longitudinal gradient for the same streams (see also Table IV). However, headstreams occupy the higher ranks (better condition), whereas in the lower ranks are generally positioned the higher order streams. The exceptions are Rabagão 1 and Cávado 2, but this is a distortion probably related to the low number of invertebrates collected in these stations.

By comparing the results for both the areas we can verify that the NW catchments have a larger dispersion of their sites, showing clearly, distinct degrees of perturbation from the best to the worst conditions (Vizela 1 and Ave 3). This scattering was not observed in the Douro catchment because a higher homogeneity between stations reproduces a general lower degree of human impacts.

Finally, Table V shows that there exists a relationship between site scores (S) and the environmental variables that are sensitive to the various stressing agents such as the conductivity (cond), percentage of dominant substrate (% ripC), percentage

TABLE III
The 10 most tolerant and intolerant taxa for the NW and Douro catchments and their predicted values obtained by logit model

NW catchments		Douro catchments	
Intolerant taxa			
Habrophlebia fusca	0.0358	Silo graellsi	0.0151
Anacaena sp.	0.0502	Protonemura meyeri	0.0155
Athripsodes sp.	0.0512	Larcasia partita	0.0159
Protonemura sp.	0.0518	Sericostoma sp.	0.0159
Micrasema longulum	0.0556	Polycentropus sp.	0.0160
Micrasema moestum	0.0559	*Aphelocheirus occidentalis*	0.0161
Oulimnius sp.	*0.0571*	Elmis sp.	0.0167
Thremma tellae	0.0647	Atherix sp.	0.0173
Micronecta sp.	0.0701	Centroptilum luteolum	0.0174
Aphelocheirus occidentalis	0.0709	Callyarcis humilis	0.0186
Tolerant taxa			
Glossiphonia heteroclita	0.9904	Ephydridae	0.9467
Batracobdella paludosa	0.9904	Dolichopodidae	0.9467
Pisidum casertanum	0.9904	Hydrovatus sp.	0.9033
Corixa panzeri	0.9904	Sciomyzidae	0.9026
Helobdella stagnalis	0.9865	Coelostoma sp.	0.8534
Erpobdella monostriata	0.9831	Dugesia tigrina	0.8363
Gyraulus albus	0.9802	Cercion lindeni	0.8306
Ancylus fluviatilis	0.9764	Meladena coriacea	0.8285
Pristina sp.	0.9754	Lymnaea truncatula	0.8248
Hydroporus sp.	0.9714	Limoniidae	0.8248

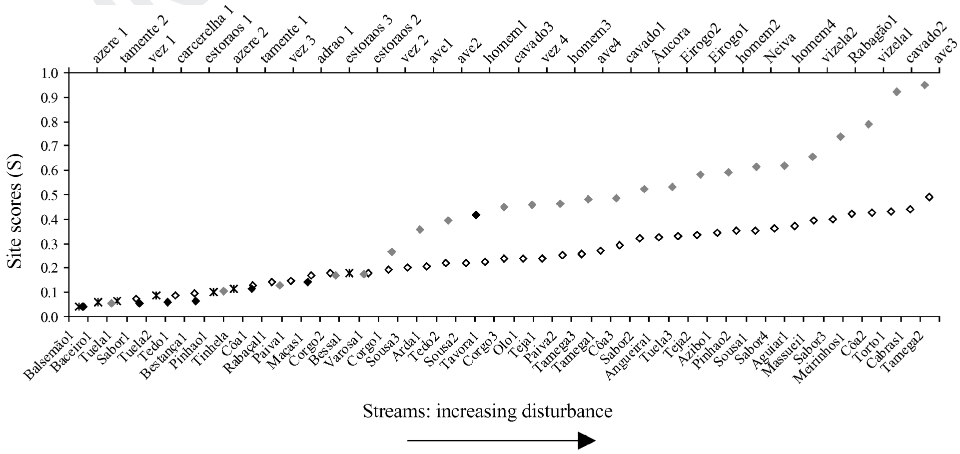

Figure 6. Diagram of all sites of the NW (upper rank) and Douro (lower rank) basins. The values were calculated for each site on the basis of the predicted values for all the taxa, ranked according to increasing levels of contamination. The black marks (♦ and ∗) indicate the reference sites pre-selected.

TABLE IV
The index values for each site of NW and Douro basins

	NW catchments		Douro catchment	
Rank	Sites	Index values	Sites	Index values
1	Ázere 1	I	Balsemão1	I
2	Tamente 2	I	Baceiro1	I
3	Vez 1	I	Tuela1	I
4	Carcerelha 1	I	Sabor1	I
5	Estorãos 1	I	Tuela2	I
6	Ázere 2	I	Tedo1	I
7	Tamente 1	I	Bestança1	I
8	Vez 3	I	Pinhão 1	I
9	Adrão 1	I	Tinhela 1	I
10	Estorãos 3	II	Côa 1	I
11	Estorãos 2	II	Rabaçal 1	I
12	Vez 2	II	Paiva 1	I
13	Ave 1	II	Maças 1	I
14	Ave 2	III	Corgo 2	I
15	Homem 1	III	Bessa 1	I
16	Cávado 3	III	Varosa 1	I
17	Vez 4	III	Corgo 1	I
18	Homem 3	III	Sousa 3	II
19	Ave 4	III	Arda 1	II
20	Cavado 1	III	Tedo 2	II
21	Âncora 1	III	Sousa 2	II
22	Eirôgo 2	III	Távora 1	II
23	Eirôgo 1	IV	Corgo 3	II
24	Homem 2	IV	Olo 1	II
25	Neiva 1	IV	Teja 1	II
26	Homem 4	IV	Paiva 2	II
27	Vizela 2	IV	Tâmega 3	II
28	Rabagão 1	IV	Tâmega 1	II
29	Vizela 1	V	Côa 3	II
30	Cavado 2	V	Sabor 2	II
31	Ave 3	V	Angueira 1	II
32			Tuela 3	II
33			Teja 2	II
34			Azibo 1	II
35			Pinhão 2	II

(*Continued on next page*)

TABLE IV
(*Continued*)

	NW catchments		Douro catchment	
Rank	Sites	Index values	Sites	Index values
36			Sousa 1	II
37			Sabor 4	II
38			Aguiar 1	II
39			Massueime1	II
40			Sabor 3	II
41			Meirinhos 1	II
42			Côa 2	III
43			Torto 1	III
44			Cabras 1	III
45			Tâmega 2	III

of herbaceous covering (% herbC), anthropogenic perturbation (anthP), bank alterations (Bank alt), riparian habitat quality (QBR), land use segment (LUS), urbanization segment (US) and sediment load segment (SLS). In this way, the sectors less impacted physically display lower scores, after all defining the reference conditions. These correlations also show that this technique is relatively independent from stream typology.

4. Discussion

The method presented here intends to be directed to a rapid but precise assessment of water-quality impairments, in a way similar to that of Chessman *et al.* (1999). In the present approach we described the basis for classifying the degree of disturbance for a high ecological range of sites. The probability of occurrence for the expected fauna (E) is obtained after fitting a logit model to the species distribution along a DCA axis expressing the main gradient of environmental stress. E scores can then be used to set the classification of any new site inside the same ecotype, after the identification of taxa and its tolerance classification based on the determined probability values. This technique can also be extended to other geographical areas. To achieve that purpose it is necessary to select "*a priori*" reference sites and to compute a new DCA followed by the convenient non-linear model (e.g. logit) to fit the taxa along the axis expressing the main anthropogenic impacts, ending in the sites classification.

When compared to RIVPACS this method has the advantage of not demanding a large set of biological and environmental data. Unlike RIVPACS, or similar models, the identification level does not need to be standardized (e.g. genus or family

TABLE V
Pearson correlations between final site scores (S) and various environmental variables ($N = 76$)

Environmental variables	r	p-value
pH	−0.0171	0.884
O_2 (dissolved oxygen)	−0.0060	0.959
v (velocity)	−0.1551	0.181
vmax (maximum water velocity)	−0.4059*	0.000
w/h (stream width:depth ratio)	−0.0230	0.843
cond (conductivity)	0.4124*	0.000
% domS (percentage of dominant substrate)	0.3218*	0.005
% ripC (percentage of riparian canopy cover)	−0.2916*	0.011
% aquC (percentage of aquatic vegetation covering)	0.1842	0.111
% herbC (percentage of herbaceous covering)	0.2757*	0.016
anthP (anthropogenic perturbation)	0.4954*	0.000
MLU (marginal land use)	−0.1727	0.136
SAV (structure of arboreal vegetation)	−0.1969	0.088
IRS (influence of riparian vegetation structure)	−.0441	0.705
Bank alt (bank alterations)	0.4035*	0.000
QBR (Riparian Habitat Quality)	−0.3498*	0.002
HMS (Habitat Modification Score)	0.1573	0.175
LUS (land use segment)	0.3115*	0.006
US (urbanization segment)	0.3908*	0.000
SLS (sediment load segment)	0.5841*	0.000
distS (distance from source)	0.1215	0.296

Note: *Significance levels (p) are shown and marked correlations are significant at $p < 0.05$.

level), since site classification was set on relative comparisons, which reconciles this procedure to the different degrees of expertise (or available identification keys for the different taxa). Besides, in contrast to presence/absence data files, we used quantitative information on the taxa present to assess test sites and to measure the difference from the reference condition, which goes in the direction of the BEAST procedure.

However, our technique does not allow taxa prediction from habitat or water quality parameters like the typical predictive models. The present approach also incorporates aspects included in the multimetric methods since there is *a priori* definition of the reference sites from each sub-ecoregion or ecotype. Such regions were defined here by using characteristics linked to the geomorphologic aspects, such as climate, physiography and geology, but also direct or indirect information related to human activities in a design that contemplate the principles expressed in Barbour *et al.* (1995). The laborious definition of such sites, like it was adopted in our framework, could be replaced by a general field survey, or by up-dating the available information sources. Nevertheless, they must be confirmed by a consistent multivariate technique, such as cluster analysis, even if it is recognized that invertebrate taxa vary continuously along environmental gradients, and sites, therefore, show a weak tendency to cluster into discrete groups (Hawkins and Vinson, 2000). Barbour *et al.* (2000) also describe a flexible strategy for the establishment of the reference condition, by assuming that heavy agricultural, industrial or urbanized areas reference conditions can be established based on diverse procedures, such as historical records, simulation models or expert judgments. Obviously, when historical information is scarce and a quick answer is needed, we cannot follow the advice of Reynoldson *et al.* (1995) to do multimetric and multivariate analyses side by side and to base the ultimate decision of site disturbance on the interpretation of both approaches.

As previously mentioned, our designed ecological evaluation of human impacts preferred the community composition to the multimetric alternative. We think that the development of such metrics requires more laborious work, since they must be sensitive to human disturbances and must have well-understood unidirectional responses. But, the fluctuation of the flow regime of the Iberian river systems and environmental harshness is responsible for poorly predictable macroinvertebrate assemblages in this region (Gasith and Resh, 1999; Aguiar *et al.*, 2002). Such a high temporal variability in community structure has also the potential to limit the sensitivity of other biomonitoring approaches, such as AUSRIVAS and RIVPACS systems, as is recognized by Bunn and Davies (2000), besides the mentioned huge effort to build those systems. For these authors, low persistence in benthic community structure makes it extremely difficult to construct robust predictive models, since temporal changes in community composition may be more stochastic and unrelated to inter-annual variation in environmental parameters.

Logistic regression set the predictions of each taxon along the defined ecological gradient, which allowed us to classify the stations according to the differences between the observed composition and the reference sites. This is a reliable analysis since it considers the theoretical distribution of species along a gradient, which tends to a unimodal curve if the gradient length is considerable. We assumed that in the same ecotype there are no relevant differences in the environmental conditions that impose distinct faunistic expectations. Logistic regression has already been used by Moss *et al.* (1999) to compare different techniques of prediction, namely in combination with multiple discriminant analysis, also using the axis scores. However, these authors did not classify sites prior to the development of the prediction, as in the present work (disturbed *versus* reference sites). Moreover, the logistic regressions used were derived for each taxon separately (see also Oberdorff *et al.*, 2001) to estimate the probability of occurrence of each test site, necessitating a much higher demanding computational load. Also, the axis scores were based on selected environmental variables for each site, whereas this study used only the scores of the species. However, that design allowed a validation of the method by comparing, for the considered river segments, the observed to the expected rates, which was not possible here. On the contrary, this procedure is far simpler and can be tested by establishing correlations between the final site scores and the available physical and chemical parameters describing their environmental condition. It should though be mentioned that the described technique depends on the ability of the first ordination axis to exhibit a gradient of contamination. Naturally, if that situation is not displayed by this axis it is more advisable to use the second or third axis, if these ones offer a better separation based on the anthropogenic impacts.

Another advantage of the method we have designed is the possibility of extracting a classification for the tolerance level of the identified taxa, from the most stressor-intolerant species to the most tolerant ones, for the specific stressing agent (in this case, organic contamination). Also, our classification is based on visible and clear standards, ranging from 0 to 1. Such a possibility is relevant, not only for the site's classification, but also for the ecological characterization of the taxa for the considered catchment or region. The intolerant species were weighted according to the presence in the reference sites, otherwise there could have been an excessive contribution of rare species. For instance, we realized that without that weighting about 50% of the taxa classified as the intolerant species appeared only in one site (however, with more than one individual). Such a fact shows that ordination is significantly affected by outliers (Belbin and McDonald, 1993). Standardization and the option of removing such rare species may also avoid them making a high contribution to the final classification of sites. This is a question that has been subjected to discussion by many authors who use multivariate analysis for bioassessment (e.g. Cao *et al.*, 1998, 2001). According to these authors, removing rare species can statistically improve the precision of species prediction, but can substantially shorten the list of species to be compared, which could lead to an underestimation of the difference between undisturbed and impacted sites. These authors refer to two types of errors in bioassessment: (a) to detect an impact, when there is none, or to overestimate a potential disturbance; (b) to fail to detect an impact or to underestimate one. The exclusion of rare species is related to error b, which is the most common one. In the present case, or in similar cases where bioassessment takes place in a defined geographical area or ecotype, the objective is to establish a precise comparison between sites, therefore we decided to include the rare species. In typical predictive systems, like AUSRIVAS and BEAST, because they are designed for large areas, it is fully justified to remove large numbers of rare taxa as in those conditions abundant species can differentiate the most important gradients instead of local ones, simplifying sorting, identification and data treatment. Faith and Norris (1989) and Cao *et al.* (2001), also consider that only for very large spatial scales does species abundance give additional information. A different perspective appears in Dufréne and Légendre (1997) and Dohet *et al.* (2002), for whom the identification of bioindicator species should be based in concepts like "fidelity" (species abundance in the reference sites) and "high specificity" (species occurrence in those sites). We believe that our approach, by incorporating these two concepts, where intolerant taxa were weighted according to their specificity and their average abundance in the reference sites was included as a variable in the logit model, was a consistent option to obtain a more precise classification of the ecological disturbance.

The present methodology was applied to a restricted geographical area, excluding the sites corresponding to the Potamon to avoid the inclusion of other river ecotypes. A further improvement could be the inclusion of a wider ecological spectrum. In fact, the bio-assessment methods can represent either a one-time investment (Rosemberg *et al.*, 2000) or they can be continually improved (Wright, 1995, 2000). It is also essential to ensure that the procedures are robust considering temporal changes or even inter-operator variability. Clarke *et al.* (2002) assessing sampling variation in BMWP scores and RIVPACS variability consider that river monitoring systems based on a comparison of O/E fauna or reference condition requires estimates of the uncertainty and errors in site quality.

In conclusion, we believe that this technique can be widely expanded since it may encompass a multitude of different scale objectives, without relying on a static model of pre-defined indicators or metrics. Such flexibility (that also includes different taxonomic expertise) makes this technique potentially very useful when there is a lack of historical information and it is necessary to perform screening assessments to identify impairment. This work introduces also a simple scoring method reflecting pollution tolerance of individual taxa (pollution-intolerant species have lower scores) and further challenges or applications are to derive accurate biotic indices for specific bioregions, since the pristine reaches are correctly determined to represent the reference situation.

The European Water Framework Directive (2000/60/EC) requires the definition of distinct levels of degradation for surface waters following the identification of reference condition for each type of water body. We believe that, after the identification of the pristine sites, this method may be very useful in order to classify the test sites according to the disturbance level.

References

Aguiar, F. C., Ferreira, M. T. and Pinto, B.: 2002, 'Relative influence of environmental variables on macroinvertebrate assemblages from an Iberian basin', *J. N. Am. Benthol. Soc.* **21**, 43–53.

Barbour, M. T., Stribling, J. B. and Karr, J. R.: 1995, 'The multimetric approach for establishing biocriteria and measuring biological condition', in: W. S. Davis and T. P. Simon (eds), *Biological Assessment and Criteria. Tools for Water Resource Planning and Decision Making*, Lewis Publishers, Boca Raton, Florida, pp. 63–77.

Barbour, M. T., Diamond, J. M. and Yoder, C. O.: 1996, 'Biological assessment strategies: Applications and limitations', in: K. L. Dickson, D. R. Groethe and D. K. Reed Jundkins (eds), *Whole Effluent*

Toxicity Testing: An Evaluation of Methods and Prediction of Receiving Systems Impacts, SETAC Press, Pensacola, Florida, pp. 245–270.

Barbour, M. T., Swietlik, F., Jackson, S. K., Courtemanch, D. L., Davies, S. P. and Yoder, C. O.: 2000, 'Measuring the attainment of biological integrity in the USA: A critical element of ecological integrity', *Hydrobiologia* **422/423**, 453–464.

Belbin, L. and McDonald, C.: 1993, 'Comparing three classification strategies for use in ecology', *J. Veg. Sci.* **4**, 341–348.

Bunn, S. E. and Davies, P. M.: 2000, 'Biological processes in running waters and their implications for the assessment of ecological integrity', *Hydrobiologia* **422/423**, 61–70.

Cao, Y., Larsen, D. P. and Thorne, R. St-J.: 2001, 'Rare species in multivariate analysis for bioassessment: Some considerations', *J. N. Am. Benthol. Soc.* **20**, 144–153.

Cao, Y., Williams, D. D. and Williams, E.: 1998, 'How important are rare species in aquatic community ecology and bioassessment?' *Limnol. Oceanogr.* **43**, 1403–1409.

Chessman, B. C.: 1999, 'Predicting the macroinvertebrate faunas of rivers by multiple regression of biological and environmental differences', *Freshw. Biol.* **41**, 747–757.

Chessman, B., Growns, I., Cureys, J. and Plunkett Cole, N.: 1999, 'Predicting diatom communities at the genus level for the rapid biological assessment of rivers', *Freshw. Biol.* **41**, 317–331.

Cortes, R. M. V.: 1992, 'Seasonal pattern of benthic communities along the longitudinal axis of river systems and the influence of abiotic factors on the spatial structure of those communities', *Arch. Hydrobiol.* **126**, 85–103.

Cortes, R. M. V., Ferreira, M. T., Oliveira, S. V. and Oliveira, D.: 2002a, 'Macroinvertebrate community structure in a regulated river segment with different flow conditions', *River Res. Appl.* **18**, 367–382.

Cortes, R. M. V., Oliveira, S. V., Cabral, D. A., Santos, S. and Ferreira, T.: 2002b, 'Different scales of analysis in classifying streams: From a multimetric towards an integrate system approach', *Large Rivers Suppl. Arch. Hydrobiol.* **13**, 209–224.

Dohet, A., Dolisy, D., Hoffmann, L. and Dufrêne, M.: 2002, 'Identification of bioindicator species among Ephemeroptera and Trichoptera in a survey of streams belonging to the rhithral classification in the Grand Duchy of Luxembourg', *Verh. Int. Ver. Limnol.* **28**, 381–386.

Dufrêne, M. and Legendre, P.: 1997, 'Species assemblages and indicator species: the need for a flexible asymmetrical approach', *Ecol. Monogr.* **67**, 345–366.

Faith, D. P. and Norris, R. H.: 1989, 'Correlation of environmental variables with patterns of distribution and abundance of common and rare freshwater macroinvertebrates', *Biol. Conserv.* **50**, 77–98.

Gasith, A. and Resh, V. H.: 1999, 'Stream in Mediterranean climate regions: Abiotic influences and biotic responses to predictable seasonal events', *Ann. Rev. Ecol. Syst.* **30**, 51–81.

Griffith, J. A., Martinko, E. A., Whistler, J. L. and Price, K. P.: 2002, 'Preliminary comparison of landscape pattern normalized difference vegetation index (NDVI) relationships to central plains stream conditions', *J. Environ. Qual.* **31**, 846–859.

Hawkins, C. P. and Vinson, M. R.: 2000, 'Weak correspondence between landscape classifications and stream invertebrate assemblages: Implications for bioassessment', *J. N. Am. Benthol. Soc.* **19**, 501–517.

Hughes, R. M.: 1995, 'Defining acceptable biological status by comparing with reference conditions', in: W. S. Davis and T. P. Simon (eds), *Biological Assessment and Criteria. Tools for Water Resource Planning and Decision Making*, Lewis Publishers, Boca Raton, Florida, pp. 31–47.

Jongman, R. H. G., Ter Braak, C. J. F. and Van Tongeren, O. F. R.: 1987, 'Data analysis in community and landscape ecology', Pudoc, Wageningen.

Karr, J. R., Fausch, K. D., Angermeir, P. L., Yant, P. R. and Schlosser, I. J.: 1996, 'Assessing biological integrity in running waters: A method and its rationale', *Ill. Nat His. Surv. Spec. Publ.* 5. **Q3**

Ladson, A. R., White, L. J., Doolan, J. A., Finlaysons, B. L., Hart, B. T., Lake, P. S. and Tilleard, J. W.: 1999, 'Development and testing of an index of stream condition for waterway management in Australia', *Freshw. Biol.* **41**, 453–468.

Moss, D., Wright, J. F., Furse, M. T. and Clarke, R. T.: 1999, 'Comparison of alternative techniques for prediction of the fauna of running water sites in Great Britain', *Freshw. Biol.* **41**, 167–181.

Munné, A., Solà, C. and Prat, N.: 1998, 'QBR: Un índice rápido para la evaluación de la calidad de los ecosistemas de ribera', *Tecnologia del Agua, Barcelona* **175**, 20–37.

Oberdorff, T., Pont, D., Hugueny, B. and Chessel, D.: 2001, 'A probabilistic model characterizing fish assemblages of French rivers: A framework for environmental assessment', *Freshw. Biol.* **46**, 399–415.

Oberdorff, T., Pont, D., Hugueny, B. and Porchers, J. P.: 2002, 'Development and validation of a fish-based index for the assessment of "river-health" in France', *Freshw. Biol.* **47**, 1720–1734.

Parsons, M. and Norris, R. H.: 1996, 'The effect of habitat-specific sampling on biological assessment of water quality using a predictive model', *Freshw. Biol.* **36**, 419–434.

Peeters, E. T. H. M. and Gardeniers, J.: 1998, 'Logistic regression as a tool for defining habitat requirements of two common gammarids', *Freshw. Biol.* **39**, 605–615.

Podani, J.: 2000, *Introduction to the Exploration of Multivariate Biological Data*, Backhuys Publishers, Leiden, The Netherlands.

Raven, P. J., Holmes, N. T. H., Dawson, F. H., Fox, P. J. A., Everard, M., Fozzard, I. R. and Rouen, K. J.: 1998, *River Habitat Quality*, Environmental Agency, Bristol, U.K.

Reynoldson, T. B., Bailey, R. C., Day, K. E. and Norris, R. H.: 1995, 'Biological guidelines for freshwater sediment based on benthic assessment of sediment (the BEAST) using a multivariate approach for predicting biological state', *Aust. J. Ecol.* **20**, 198–219.

Reynolsdon, T. B., Norris, R. H., Resh, V. H., Day, K. E. and Rosenberg, D. M.: 1997, 'The reference condition: A comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates', *J. N. Am. Benthol. Soc.* **16**, 833–852.

Rosemberg, D. M., Reynoldson, T. B. and Resh, V. H.: 2000, 'Establishing reference conditions in the Fraser River catchment, British Columbia, Canada, using the BEAST (benthic assessment of sediment) predictive model', in: J. F. Wright, D. W. Stutcliffe and M. T. Furse (eds), *Assessing the Biological Quality of Fresh Waters*, Freshwater Biological Association, Ambleside, Cumbria, pp. 181–194.

Simon, T. P.: 1999, 'Introduction: Biological integrity and use of ecological health concepts for application to water resource characterization', in: T. P. Simon (ed), *Assessing the Sustainability and Biological Integrity of Water Resources Using Fish Communities*, CRC Press, Boca Raton, Florida, pp. 3–16.

Sloane, P., Simpson, J. and Norris, R.: 1997, *ACT Component of the Monitoring River Health Initiative*, Cooperative Research Center for Freshwater Ecology, Canberra, Australia.

StatSoft: 2001, *STATISTICA for Windows (Computer Program Manual)*, StatSoft, Inc., Tulsa, Oklahoma.

ter Braak, C. F. J. and Looman, W. N.: 1996, 'Weighted averaging logistic regression and Gaussian response model', *Vegetatio* **65**, 3–11.

ter Braak, C. F. J. and Smilauer, P.: 1998, *CANOCO Reference Manual and Users Guide to Canoco for Windows*, Centre for Biometry, Wageningen.

Webb, B. M. and Walling, D. E.: 1996, 'Water quality. II. Chemical characteristics', in: G. Petts and P. Calow (eds), *River Flows and Channel Forms*, Blackwell, Oxford, U.K., pp. 102–129.

Whittaker, R. H.: 1967, 'Gradient analysis of vegetation', *Biol. Rev.* **49**, 207–264.

Wright, J. F.: 1995, 'Development and use of a system for predicting the macroinvertebrate fauna in flowing waters', *Aust. J. Ecol.* **20**, 181–197.

Wright, J. F.: 2000, 'An introduction to RIVPACS', in: J. F. Wright, D. W. Stutcliffe and M. T. Furse (eds), *Assessing the Biological Quality of Fresh Waters*, Freshwater Biological Association, Ambleside, Cumbria, pp. 1–24.

Wright, J. F., Moss, D., Armitage, P. D. and Furse, M. T.: 1984, 'A preliminary classification of running-water sites in Great Britain based on macro-invertebrate species and the prediction of community type using environmental data', *Freshw. Biol.* **14**, 221–256.

Queries

Q1. Au: Please check the value −0.9565 in the denominator.

Q2. Au: This reference is not given in the reference list. Please check.

Q3. Au: Please check the year in this reference.