Ecological Indicators 45 (2014) 51–62

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

An invertebrate predictive model (NORTI) for streams and rivers: Sensitivity of the model in detecting stress gradients

Isabel Pardo[a,*], Carola Gómez-Rodríguez[a], Rut Abraín[a], Emilio García-Roselló[b], Trefor B. Reynoldson[c]

[a] Department of Ecology and Animal Biology, University of Vigo, E-36330 Vigo, Spain
[b] Department of Informatics, University of Vigo, E-36330 Vigo, Spain
[c] Acadia University Wolfville, Nova Scotia, Canada

ARTICLE INFO

Article history:
Received 25 September 2013
Received in revised form 17 March 2014
Accepted 20 March 2014

Keywords:
Macroinvertebrates
Rivers
Bioassessment
Reference conditions
Pressure gradients

ABSTRACT

NORTI (NORThern Spain Indicators system) is a predictive model for assessing the ecological status of rivers of Northern Spain based on invertebrates. The system can be used to assign any test site to a type of river under minimal disturbed conditions. Macroinvertebrates were sampled with a multihabitat approach from 676 sites covering the variation in environmental conditions across Northern Spain, between 2000 and 2008 ($n = 1421$ samples), including a spatial network of 108 reference sites selected by the absence of significant pressures. A multinomial logistic regression was conducted using the GAAC cluster-derived groups of reference sites as response variable. Obligatory typology factors, following WFD System A, were included as forced entry terms in the model, other potential predictors were selected using a forward stepwise procedure. Ecological quality ratios (EQRs) were estimated from the observed similarity between the faunal composition of the sample of interest (test sample) and the expected median similarity for the reference community of each river type. The model predictions as EQRs responded significantly to the most important pressures: sewages inputs, eutrophication, hydromorphological alterations, and intensive and low intensity agriculture, demonstrating its accuracy in detecting impact in Northern Spanish streams and rivers.

1. Introduction

Predictive models have been used for the bioassessment of streams and rivers for nearly 30 years (Wright et al., 1984; Moss et al., 1987; Hawkins et al., 2000; Reynoldson et al., 1995, 2000; Simpson and Norris, 2000), and there is renewed interest in European countries (e.g. Kokeš et al., 2006; Feio et al., 2007) since the appearance of new water legislation such as the Water Framework Directive (WFD; 60/CE/2000). The technical basis underlying predictive modeling is substantially different from the “upstream/downstream” approach more traditionally used to evaluate the degree and magnitude of impact (Green, 1999). The approach compares impacted sites with an expected reference community. In essence, predictive modeling aims to predict the composition of biological communities based on stream environmental attributes that may influence the distribution and abundance of species. For a given site, the predicted community would represent the biotic conditions that would exist under no impact and, thus, can be used as a reference to infer the departure of such site from its natural status. It should be noted that the utility of predictive models strongly depends on the existence of a relationship between the selected environmental features and the species occurrence or/and abundance (Wright, 2000). In this sense, changes in the species habitat templet (*sensu* Southwood, 1977), influenced by natural (drought, floods) or human disturbances (climate change, organic or chemical pollution, hydromorphological alterations), or by both, can change patterns of species distribution, abundance, diversity and ecosystem functioning (Puccinelli, 2011).

Macroinvertebrates are commonly used as bioindicators of environmental stress in streams and rivers (Bennett et al., 2011), as they may respond to environmental change from a variety of disturbances ranging from nutrient enrichment to hydromorphological alteration (Johnson and Hering, 2009). This capability of macroinvertebrates should ideally be integrated in assessment systems in a way that may allow the detection of impact across multiple pressure gradients, as opposed to the selection of responses targeting a specific single stressor (i.e. organic pollution, such as

* Corresponding author. Tel.: +34 986812585.
E-mail address: ipardo@uvigo.es (I. Pardo).

http://dx.doi.org/10.1016/j.ecolind.2014.03.019
1470-160X/

the saprobic index (Zelinka and Marvan, 1961), or hydromorphological degradation (Lorenz *et al.*, 2004)), as it has been indicated for macrophytes (Kanninen et al., 2013), to account for synergies or antagonisms between stressors (Darling and Côte, 2008). Predictive models can detect multiple directional shifts in community composition as a response to single or multiple combined stressors, because they use the whole community in assessing disturbance, and mostly because the structure of the community is not constrained by a priori selection of biological responses to stressors causing degradation (i.e. construction of multimetrics combining biological metrics that respond individually to different stressors).

The WFD provides scientific and technical guidance documents for its implementation across Europe, establishing standardized ecological designs for biological assessment to ensure comparable environmental objectives among countries (e.g. Bennett et al., 2011; Kelly et al., 2012). A basic requirement is the division of natural aquatic ecosystems in water categories (rivers, lakes, wetlands, coastal and transitional waters) and, within each, in "types". The "type" comprises aquatic ecosystems of similar structure and function, conceptually they are characterized by type-specific chemical and hydromorphological conditions and inhabited by specific biotic communities. They need to be differentiated using abiotic descriptors (e.g. altitude, catchment area, geology), some being obligatory and some optional. It must be stressed that the biological communities cannot be used to infer typologies to avoid circularity in assessment, despite the fact that the expected unimpaired biological communities should be type specific (WFD: Directive, 2000/60/EC). Another important concept is the "ecological status", defined as the "expression of the quality of the structure and functioning of aquatic ecosystems associated with surface waters" (art. 2.21). The concept of reference condition (Reynoldson et al., 1997; Bailey et al., 1998; Reynoldson and Wright, 2000; Bailey et al., 2004) is at the core of the ecological status classification systems of the WFD used to gauge the effects of human activity (Karr and Chu, 1999). In brief, the European application of the concept used spatial networks of minimally disturbed streams and rivers *sensu* Stoddard et al. (2006) for each "type", compliant with a given set of pressure criteria thresholds. The check with pressure criteria should ensure the absence of significant impact at those sites (Pardo et al., 2012). Consequently, the type-specific derived biological reference conditions are the benchmark against which any test site belonging to the type is assessed. Finally, biological classification systems for the ecological status have to show a significant response to pressures, in line with the analytical framework for the assessment of pressures and impacts (Driving force-Pressure-State-Impact-Response; DPSIR), adopted by the European Environment Agency (1999). Moreover, they have to provide some understanding of the level of confidence and precision of the ecological status assessment, to estimate the confidence to which an individual water body can be assigned to an ecological status class (Clarke and Hering, 2006).

The aim of this study is to develop a predictive model for Northern Spain that meets the new scientific and technical requirements described in the new EU policy for water management. Within this framework, biotic and abiotic data were assembled from minimally disturbed sites or reference sites, where the absence of significant pressure criteria was verified. Here, we describe the development of the invertebrate NORTI (NORThern Spain Indicators system) predictive model for streams and rivers of Northern Spain, a fully WFD compliant method. We used stress gradients to test the invertebrate response to multiple sources of pollution and stream degradation, in order to assess the sensitivity of the predictive models. The classification system developed here is presently used by the Regional water Authorities in Northern Spain to assess the ecological status of streams and rivers.

2. Methods

2.1. Study area

The study area included most of the Northern coast of Spain, from the Western Atlantic corner (Galicia) to the beginning of the western Pyrenees (Navarra) in the East, covering an area of 38,450 km^2 (Fig. 1). The latitudinal range is small with the high mountains, where the rivers originate, very close to the Northern coast. The altitudinal range is high, from a maximum altitude of 2640 to 0 m.a.s.l. In the Western part, the longest river (River Miño) drains the largest catchment of 16,357 km^2. The dominant climate is oceanic, with abundant rainfall throughout the year (mean annual precipitation of 1500 mm [Coastal Galicia water district], mean annual precipitation of 1175 mm [Miño-Sil water district] and 1350 mm [Cantabrian water district]), and moderate variation in temperatures, with mild winters and cool summers. The geology in the area varies from mainly granitic and siliceous rocks in the West, to a dominance of carbonate rock in the Northeast. The highest human population density occurs near the coast and, in particular, in the Eastern and Western parts of the study area (total population in the study area is approximately 7 million). The main anthropogenic effects on streams and rivers are motivated by channel modification in urban areas, flow regulation for hydropower generation, point source organic and industrial pollution, and diffuse pollution from agriculture.

Rivers within the Northern catchments are short, rapid and plentiful, flowing South to North through steep valleys. The exceptions occur within the Miño-Sil Catchment, where some long rivers with multiple tributaries flow east-west through elongated and narrow valleys. Most streams and rivers are fast flowing high gradient streams, because of the proximity of their mountainous origin to the sea, their substrate is generally coarse as a consequence of large hydrological variation. Legislation requires the rivers in Northern Spain to maintain, at minimum, narrow forested riparian areas, thus, helping to maintain the hydromorphological integrity of streams and rivers, while supporting habitat and providing litterfall inputs that sustain stream functions in these oligotrophic streams (Pardo and Alvarez, 2006).

2.2. Data collection

In this study, a total of 676 sampling stations were identified spatially incorporating the wide range of existing environmental gradients, aiming to cover thoroughly the natural variability of stream types and reference conditions that exist in the studied area. Each sampling station was sampled at least once between 2000 and 2008. Most stations were sampled only once a year, in summer, except for 2006 and 2008, when 162 and 6 stations, respectively, were also visited in spring. The total number of samples collected was 1421. Sampling effort per year varied, the smallest number corresponding to 2000–2002 (<10 samples per year) and the highest in 2003, 2006 and 2008 ($\geq$290 per year).

2.2.1. Benthic macroinvertebrates

At each station, macroinvertebrates were sampled from 20 subsampling units collected in a 100 m reach using a D-frame dip net (width = 0.25 m; mesh size = 0.5 mm). The subsampled units were selected following the multihabitat sampling approach (adapted from Barbour et al., 1999) in which five major habitats (representing > 5% of the total area) were sampled according to their proportional distribution in the reach. Each subsampling unit covered a distance of 0.5 m, thus yielding a sampled surface area of 0.125 m^2. All subsampling units were pooled into one sample, for a total sampled surface area per site of 2.5 m^2. Samples were preserved in 96% ethanol. In the laboratory, each sample was washed

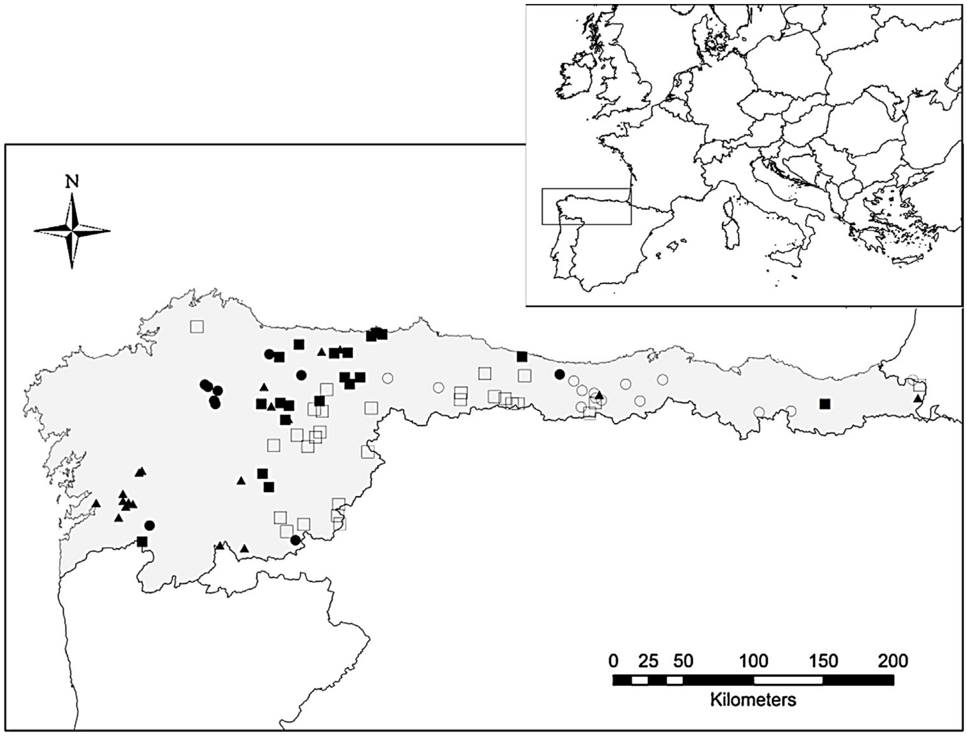

Fig. 1. Location of the studied area in North Spain (●, Main river axes; ○, Mixed calcareous rivers; ▲, Mixed siliceous rivers; ■, Mixed lowland rivers; □, Small mountain streams).

using three sieves (5 mm, 0.5 mm and 0.1 mm of mesh size) to increase sorting efficiency. When the number of individuals was too large, a further sub-sampling was carried out only in medium and fine fractions until a count of a minimum of 100 individuals was attained following Wrona et al. (1982), and multiplying the sub-sample counts for the fraction of the sample represented. The final abundance per sample was obtained by adding the counts from the 3 separated and identified fractions. Complementary *visu* inspection were produced on the median fraction to account for new and rare taxa of medium size. Individuals were identified to family level (except for Acari, Oligochaeta and Nematoda).

2.2.2. Water chemistry

Dissolved oxygen saturation was measured in situ with an oxygen meter (YSI 556 MPS). Additionally, in most sampling visits filtered water samples were collected to quantify nutrients, indicative of human-induced disturbance. Filtration was carried out in situ with a hand-pump through a glass fiber filter (Whatman GF/F with a pore size 0.045 μm). Samples were immediately frozen and transported to the laboratory. Phosphate [$P\text{-}PO_4^{-3}$ mg/L], nitrate [$N\text{-}NO_3^-$ mg/L], nitrite [$N\text{-}NO_2^-$ mg/L] and ammonium [$N\text{-}NH_4^+$ mg/L] were analyzed following the specific ISO standard methods for water samples (ISO 1996, 2003, 2005) by means of a continuous-flow analyzer (Auto-Analyzer 3, Bran + Luebbe, Germany).

2.2.3. Reach and catchment conditions

Besides the in-stream characterization, other abiotic variables were predetermined for each site, at both reach and catchment scales. The WFD recommends the use of particular abiotic factors for distinguishing stream types within a region. Stream types can be defined using two different WFD systems: System A is a standardized system with obligatory physical descriptors and boundaries, while System B is more flexible and allows the inclusion of both obligatory and optional variables and boundaries. Following System A, sampling stations were characterized by the obligatory variables of altitude [m], latitude [m], longitude [m], catchment area [km^2] and geology. In addition, several optional variables from System B (mean channel slope [%], annual precipitation [mm] and catchment slope [%]) were determined (Table 1).

2.2.4. Human activity gradients

Eleven variables were selected to quantify major stressors in the study area. Land-use activities (artificial surfaces, intensive and low-intensity agriculture areas), and natural/semi-natural areas in the catchments were estimated from CORINE land cover maps (Coordination of Information on the Environment, Land Cover 2000); see Pardo et al. (2012) for detailed information for corresponding CORINE land cover categories. The river catchment authorities provided information on anthropogenic activities/interventions for each sampling station and at the catchment scale: population density; sewage volume, differentiating between domestic wastewater, urban wastewater and industrial wastewater; the total number of dams upstream; upstream dams cumulative height per catchment area; river protection (levees) and the number of water transfers and diversions per catchment area.

2.3. Identification of reference sites

To identify reference sites, a screening was conducted based on the stressor criteria agreed by the Central/Baltic Geographic Intercalibration Group for the implementation of the WFD (Pardo et al., 2012). However, lower rejection thresholds were applied for land-use (artificial = 0%, total agriculture < 25%) in small streams (<100 km^2) given their higher sensitivity to human impact. Similarly, in medium and large rivers, lower protective thresholds were considered (artificial ≤ 0.6%, intensive agriculture < 25%) following recommendations by Pardo et al. (2011). Afterwards, expert judgment was applied to check if biota of those sites evidenced any human-induced disturbance that might not have been considered, but no site was rejected. Spearman correlations were performed

Table 1
Sources of information to derive the typology variables and their units.

	Descriptor	Unit	Source
System A	UTM X (m)	m	European Datum 1950 Zone 30 system
	UTM Y(m)	m	European Datum 1950 Zone 30 system
	Altitude (m)	m	Digital Terrain Model (DTM; Instituto Geográfico Nacional, 2004–2008)
	Catchment area (km^2) *log-transformed*	km^2	Digital Terrain Model (DTM; Instituto Geográfico Nacional, 2004–2008)
	Calcareous substrate (%)	%	Lithostratigraphic map (Instituto Geológico y Minero de España, 2006)
System B	Mean channel slope	%	Percentage from DTM
	Annual precipitation	mm	From precipitation map (Estrela y Quintas, 1996)
	Catchment slope	%	Percentage from DTM
	Maximun altitude	m	Digital Terrain Model (DTM; Instituto Geográfico Nacional, 2004–2008)
	Mean catchment altitude	m	Digital Terrain Model (DTM; Instituto Geográfico Nacional, 2004–2008)

between the pressure variables and ecological status values provided by the NORTI in reference sites, confirming the absence of significant relationships within the reference sites (all Spearman correlations $p > 0.01$). Only the first sample collected at a site was considered for the reference sites network building, yielding a network composed of 108 samples collected during summer.

2.4. Model construction and calculations of O/E

2.4.1. Model construction and validation

To identify biological groupings in the reference samples, a Group-Average Agglomerative Clustering (GAAC) was conducted on the Bray–Curtis similarity matrix of log-transformed relative abundances of 96 taxa. In GAAC, the similarity between two clusters is measured by the average of all the similarities between all combinations of two objects, one from each cluster (Quinn and Keough, 2002). Groups were set at different similarity levels (ranging from 52.5% to 59.8%) based on visual inspection of the dendrogram. Groups with less than 10 samples (1–5 sites groups) were removed from further analyses after examination of their invertebrate communities and geomorphological settings, to confirm that they were not really true representatives of any missing stream or river type but similar to other bigger taxa groups, and following Reynoldson and Wright (2000) recommendations for sufficient sites to suitably characterize the variation within a reference group. Differences among groups were evaluated with ANOSIM analysis (Clarke, 1993). Additionally, a non-Metric Multidimensional Scaling (MDS) ordination was performed on the Bray–Curtis similarity matrix to visually depict such differences among groups. All these analyses were run using PRIMER software v.6 (Clarke and Gorley, 2006).

In order to predict biologically-relevant groups from abiotic variables, a multinomial logistic regression was conducted using the GAAC cluster-derived groups as response variable and the nine abiotic factors as potential predictors. This type of regression allows the prediction of multiple levels nominal variables with an array of independent variables, in a similar manner as discriminant analysis does (Hossain et al., 2002). Obligatory factors, following System A (WFD), were included, as permitted in the analysis, as forced entry terms in the model, being confirmed its inclusion in the model by its significant entrance in the model. The rest of the potential predictors were selected using an automatic forward stepwise procedure for variable selection based on their significance. A threshold of $p < 0.05$ was set for predictor inclusion and $p > 0.10$ for predictor removal. A "leave-one-out" cross-validation procedure was manually conducted in order to evaluate the predictive ability of the model. Validation results consist of a confusion matrix. The confusion matrix compares the GAAC cluster-derived group for each given site with the group predicted with the multinomial regression model (see Table 4 for illustration). Moreover, it should be noted that the prediction of the group was conducted under a cross-validation procedure and thus the site of interest had been previously removed from the training set to ensure independence in the comparison. The final model equations were used to predict the stream type (i.e. group) for all test sampled sites that did not fulfill the pressure criteria thresholds for reference conditions. Multinomial logistic regression analyses were performed with SPSS v. 16.

2.4.2. O/E calculation

The WFD states that monitoring systems for ecological quality assessments should be comparable across European countries. With this aim, the ecological status of a water body should be expressed as an ecological quality ratio (EQR) where the observed value of the classification parameter is divided by the expected value of the parameter in the absence of anthropogenic alteration. Such expected value was inferred from the network of reference sites corresponding to the same type of water body (i.e. same expected biological association).

In this study, EQR values were estimated from the observed similarity between each sample of interest (hereinafter test sample) and the expected median similarity for the reference samples for its type (i.e. group predicted with the multinomial logit regression). Using the following procedure: (1) we confirmed that the centroid (the ordination point central to the reference sites location), corresponded with a sample position resulting from a type reference community composed by the median composition of the taxa within the reference group of sites. (2) We calculated the BC similarity between each test site community and the median community of its corresponding type. The greater the similarity percentage between the test sample and the centroid (i.e. Bray–Curtis value close to 100%), the lowest the difference from reference. (3) To obtain the EQR, the observed similarity between the test site community and the median reference community divided by the expected similarity value. Such expected value calculated as the median value of the similarity between each reference sample and the centroid in an NMDS representing only the reference samples of the particular water body type (GAAC group).

2.5. Ecological status class assignment: class confidence

The ecological quality ratio (EQR) ranged from 0 to >1 because we used the median value from reference samples to calculate the EQR. The EQR scale was divided into five classes, as WFD requires, by considering the global median EQR from reference sites as the first High/Good boundary, and dividing the remaining scale between 4 to obtained the next 3 boundaries in a mathematical way. These class boundaries were intercalibrated at the European level in 2011 (Intercalibration technical report for rivers, data no published), being the final intercalibrated assessment bands: High (H) > 0.930; Good (G) = 0.930–0.700; Moderate (M) = 0.700–0.500; Poor (P) = 0.500–0.250; Bad (B) < 0.250. Additionally, the "probability of ecological class membership" or "confidence of class" was estimated for each single value of EQR that can be obtained from a unique sample collected at a particular site. The rationale behind lies in the fact that natural spatio-temporal variability in EQR values at a site may cause misclassification from its “true” class if a single sample is used for ecological classification. The “true” class would correspond to the one obtained given perfect information for that location and sampling period.

To compute the confidence of class, a two-step approach was conducted, following Kelly et al. (2009). First, the relationship between mean EQR values and their standard deviation was modeled with a second-order polynomial equation using data from all sampling locations surveyed at least twice. Second, assuming that the variability in the observed EQR values at a particular site is normally distributed, the confidence distribution associated to a given EQR value was assumed to follow a normal distribution in which the standard deviation is estimated from the polynomial equation obtained in step one. For each ecological class and EQR value, the confidence of class was computed using the cumulative distribution function in Excel (NORMDIST). This function provides the probability of such EQR being superior to the lower class boundary and hence the probability of being in the class of interest or in a better one. Thus, the probability value is computed using the EQR value, the ecological class lower boundary (introduced as the distribution mean) and the predicted standard deviation for such EQR value as $p_i = 1 - \text{NORMDIST}$ (EQR value; class boundary; predicted standard deviation). For any given EQR value, five equations are necessary to estimate its confidence of class, where p_i represents the probability of belonging to a status class worse than the class of interest (bad (B), poor (P), moderate (M), good (G) and high (H)). So, to compute the confidence of class for the bad status, the probability of being in a worse class than poor is used:

– Confidence of bad status class = $100p_{\mathrm{P}}$
  And the same rationale is applicable for the rest of the status classes:
– Confidence of poor status class = $100(p_{\mathrm{M}} - p_{\mathrm{P}})$
– Confidence of moderate status class = $100(p_{\mathrm{G}} - p_{\mathrm{M}})$
– Confidence of good status class = $100(p_{\mathrm{H}} - p_{\mathrm{G}})$
– Confidence of high status class = $100(1 - p_{\mathrm{H}})$

Graphically, this procedure equates to the plotting of the bell-shaped confidence distribution against the EQR range and class boundary limits, in order to calculate the area under the curve within each ecological class region.

2.6. EQR values and environmental conditions

We tested whether the EQRs responded to stress gradients or not, using multiple regression models, were the EQR values were the response variable and the main stress gradients the potential predictors. Stress gradients were extracted from a Principal Components Analysis with varimax rotation on anthropogenic stressors: water chemistry (ammonium, nitrate, nitrite, phosphate, oxygen saturation), land-use in the catchment (artificial, intensive agriculture, low intensity agriculture, natural or semi-natural areas), hydromorphological alterations in the catchment (length of levees used for the protection of riverside interventions, transfers and diversions, number and height of dams), population density and sewage volume (domestic, urban and industrial wastewater). Some variables were transformed in order to improve the linearity of relationships between variables (see Quinn and Keough, 2002). Five components with eigenvalues above 1 were extracted, explaining 69.4% of variance.

For the regression analysis, the mean EQR value was computed if several samples were collected for a given locality. Variable selection was based on a “best subset model approach” (see Burnham and Anderson, 2002) following the AIC criterion. The lower AIC value is, the better the model is. Plausible models were the model with the lowest AIC value as well as those with a difference in AIC value below 10 (see Burnham and Anderson, 2002).

3. Results

3.1. Model construction and validation

A total of 110 taxa were identified in the whole studied area, with Chironomidae, Baetidae and Elmidae being the most abundant, while Chironomidae, Baetidae and Hydrobiidae were the most frequent taxa.

Based on GAAC clustering, five groups (i.e. groups of invertebrate assemblages) were identified in the reference site network ($n = 108$) with similarity levels ranging from 52.5% to 59.8% (Fig. 2). Seventeen samples were excluded as they formed small groups (<10 samples) without clear relationship to other bigger ones. ANOSIM analyses demonstrated that each of the five groups differed from the rest (Global ANOSIM R = 0.635; $p < 0.01$; pair-wise ANOSIM R > 0.422; $p < 0.001$). Such group distinctiveness was also evidenced in a 2-dimension NMDS ordination of samples (Fig. 3). Groups also differed in their environmental characteristics (Table 2) and, could be classified as Main river axes (Type I), mixed-calcareous rivers (Type II), mixed-siliceous rivers (Type III), mixed-lowland rivers (Type IV) and small mountain rivers (Type V) (Table 2).

A total of 88 families were found at the 91 reference sites, with a total taxa richness per stream type ranging from 66 to 77 taxa, and mean taxa richness from 29.3 ± 1.0 to 39.2 ± 2.3 (Table 2). Mean abundance was minimal in type 2 (5299.2 ± 520.2) and maximum in type 4 (6756.9 ± 582.5). A SIMPER analysis between the stream types revealed that a high number of taxa contributed on a mean to the dissimilarity between the types (53.6 ± 0.6) being the mean dissimilarity among all pairwise types comparisons 42.2 ± 0.9. The taxa contributing most to the dissimilarity among river types (to the first 30% of dissimilarity contribution among types) either by its presence or by its abundance, corresponded to the families: *Ancylidae, Aphelocheiridae. Caenidae. Calopterygidae. Empididae, Ephemerellidae, Gammaridae, Goeridae, Gyrinidae, Heptageniidae, Hydraenidae, Hydrobiidae, Hydroptilidae, Lepidostomatidae, Leptoceridae, Leptophlebiidae, Nemouridae, Oligochaeta, Perlidae, Philopotamidae, Planariidae, Psychomyiidae, Scirtidae, Sericostomatidae*, and *Sphaeriidae*. The mean relative abundance and standard error of the taxa contributing up to 90% to the similarity within each group are presented in Table 2.

The final logistic model was composed of six environmental variables: geographical coordinates (UTM X and UTM Y), altitude, catchment area, calcareous substrate and catchment slope (Table 3) and explained 90% of the variability (Nagelkerke pseudo-$R^2 = 0.89$). Only one of the optional variables (catchment slope) was selected in the step-wise procedure of the multinomial logit model. The mean values and standard error of these variables in the reference samples network are presented in Table 2. Cross-validation results evidenced acceptable discrimination ability since model parameters correctly classified 64.8% of the cases. Model predictions for the river catchments sampled in this study ($n = 676$ sites) showed that the type 3 (mixed-siliceous rivers) was the most abundant stream type within the studied area.

We directly compare the European system A river typology of the Central Baltic area, where the Northern Spanish rivers have been intercalibrated (Bennett et al., 2011), with the resulting NORTI types, observing a broad agreement with some of the system A types (Table 6). We also run a direct comparison between the performance of the O/E EQR resulting from the NORTI types and the O/E EQRs derived with the same method from the European A types, evaluating the precision of the estimates (i.e. standard deviation) for the reference population as suggested by Aroviita et al. (2008). The comparison indicated a higher performance of the O/E NORTI

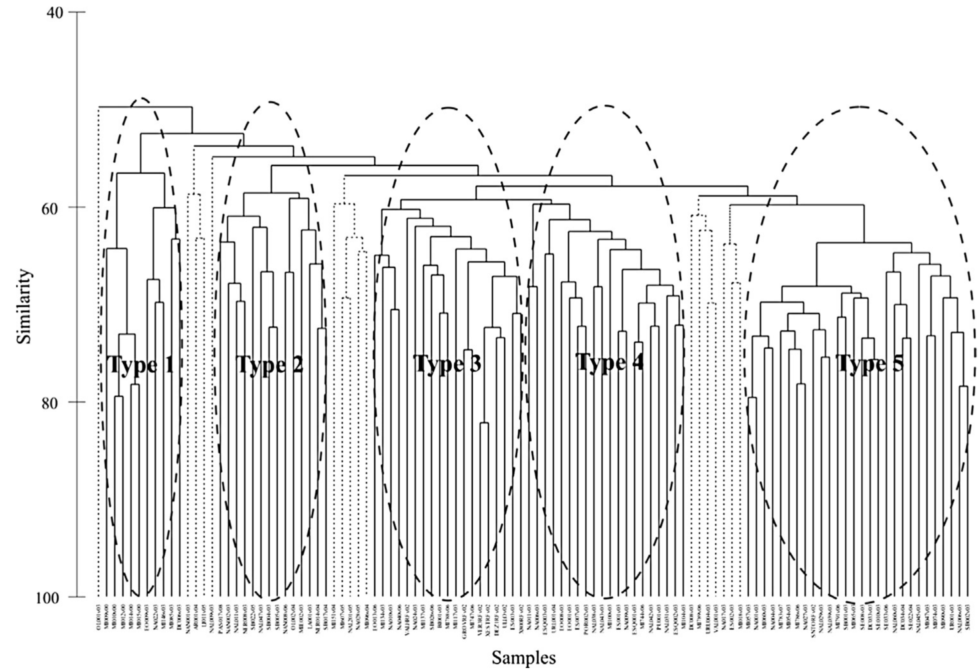

Fig. 2. Dendrogram showing the classification of reference sites according to their macroinvertebrate communities. Identified stream types are shown. Dashed lines correspond to excluded samples.

EQRs (0.950 ± 0.116 mean ± SD) than the O/E EQRs of the typology A (0.978 ± 0.228), indicative of less variation in the reference sites estimation.

3.2. *EQR and ecological status class assignment: confidence of class*

EQR values ranged from 0.063 to 1.307 and were obtained after dividing the observed similarity for a particular sample by the median similarity value calculated from all pairwise comparisons of reference sites in each type (i.e. expected median similarity value: Type 1 = 72.26%; Type 2 = 70.57%; Type 3 = 71.64%; Type 4 = 74.51%; Type 5 = 76.31%). The assignment to ecological classes following standard boundaries evidenced an over representation of samples in the best (good and high) ecological classes (B = 26 samples; P = 89 samples; M = 242 samples; G = 781 samples and H = 283).

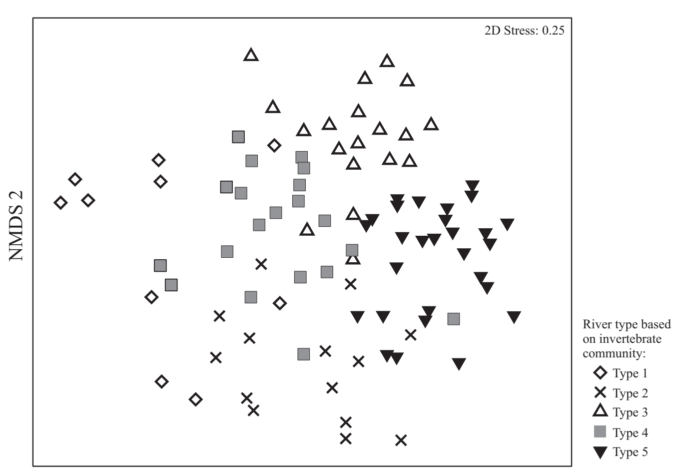

Fig. 3. Bray–Curtis based non-metric multidimensional scaling (NMDS) ordination of macroinvertebrate reference samples. Symbols correspond to observed stream type, as extracted from GAAC clustering.

Variability in EQR values measured at a given sampling station at different times ranged from almost negligible (minimum standard deviation in a sampling station = 0.0004) to large (maximum standard deviation in a sampling station = 0.313, with only 6.3% of the stations with a standard deviation above 0.20). In general, largely different values of standard deviation of EQRs were observed along the EQR range and no clear pattern was depicted. Thereby, standard deviation of EQRs showed no relationship with mean EQR (Fig. 4).

The probability of ecological class membership evidenced that very high confidence (> 85%) can be only attained for very small (<0.128) and very large (>1.031) EQR values. For the rest of the EQR values, maximum confidence of classes (between 75% and 80%) is attained in the center of the ecological class region while minimum confidence is observed in the class boundaries, as expected. Notably, the lowest maximum confidence corresponds to the moderate class (Fig. 4). As an example, for the NORTI EQR of 0.75 (0.70 good boundary) on ≥2 samples on a site and the sampling uncertainty SD of 0.088, the observed status class is good, with a

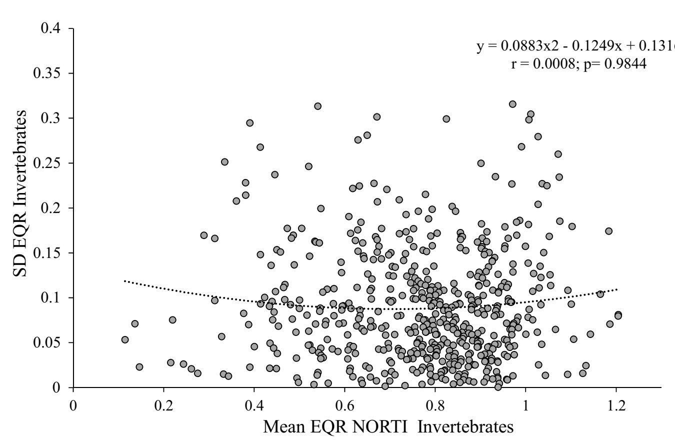

Fig. 4. Within-site variability in EQR values for sites with ≥2 samples. The line is fitted to a polynomial function.

Table 2
Abiotic catchment characteristics (mean ± SE) and description of the invertebrate assemblages composition (in bold, taxa contributing up to 90% to the similarity of the group) of the 5 stream types identified by GAAC clustering (data represent mean relative abundance ± SE). Only data from the network of 91 reference sites is used.

	Main river axes 1	Mixed-calcareous rivers 2	Mixed-siliceous rivers 3	Mixed lowland rivers 4	Small mountain rivers 5
UTM X	158.0 ± 81.4	416.2 ± 95.1	151.6 ± 143.5	216.2 ± 96.2	259.8 ± 104.6
UTM Y	4769.5 ± 46.1	4784.1 ± 8.5	4736.0 ± 45.1	4784.8 ± 39.4	4757.5 ± 38.4
Altitude (m)	367.2 ± 376.8	327.6 ± 175.7	480.9 ± 247.6	296.3 ± 183.5	801.6 ± 302.8
Catchment slope (%)	21.5 ± 16.6	38.6 ± 8.1	32.7 ± 11.8	43.7 ± 9.1	48.5 ± 11.3
Catchment area (km^2)	885.7 ± 694.6	65.7 ± 70.8	20.6 ± 13.8	88.6 ± 140.8	21.8 ± 20.7
Calcareous substrate (%)	21.0 ± 22.7	79.9 ± 26.0	10.3 ± 25.5	29.0 ± 31.5	27.1 ± 31.6
n° sites	10	14	19	20	28
Similarity intra-group	61.1	60.5	63.1	63.3	66.3
Total taxa richness	75	73	76	77	66
Mean taxa richness ± SE	39.2 ± 2.3	32.6 ± 1.6	38.2 ± 1.1	37.3 ± 1.0	29.3 ± 1.0
Mean abundance ± SE	6351.8 ± 451.2	5299.2 ± 520.2	5567.7 ± 900.6	6756.9 ± 582.5	6233.2 ± 923.0
Taxa					
Ancylidae	**2.3 ± 1.1**	**0.7 ± 0.3**	**0.4 ± 0.1**	**2.8 ± 0.5**	**0.5 ± 0.1**
Aphelocheiridae	**0.3 ± 0.1**	0.0 ± 0	0.0 ± 0	0.0 ± 0	0.0 ± 0
Athericidae	**0.8 ± 0.2**	**0.5 ± 0.1**	**1.8 ± 0.6**	**0.7 ± 0.1**	**0.4 ± 0.1**
Baetidae	**10.6 ± 2.5**	**18.8 ± 2.9**	**5.0 ± 1.1**	**6.6 ± 1.2**	**12.3 ± 2.1**
Brachycentridae	**0.4 ± 0.1**	**1.0 ± 0.5**	**1.8 ± 0.4**	**1.4 ± 0.4**	**1.7 ± 0.5**
Caenidae	**4.7 ± 1.8**	**1.1 ± 0.6**	0.1 ± 0.1	0.3 ± 0.1	0.1 ± 0.1
Calopterygidae	0.1 ± 0.1	0.1 ± 0.1	**0.5 ± 0.2**	**0.3 ± 0.1**	0.0 ± 0
Chironomidae	**20.0 ± 3.6**	**16.1 ± 2.2**	**21.0 ± 1.8**	**25.5 ± 3**	**19.9 ± 2.3**
Elmidae	**10.9 ± 1.7**	**13.3 ± 1.2**	**13.4 ± 2**	**9.1 ± 1.3**	**6.1 ± 1**
Empididae	**0.2 ± 0.1**	0.1 ± 0.1	**0.7 ± 0.1**	**0.7 ± 0.2**	0.6 ± 0.3
Ephemerellidae	**3.3 ± 1.4**	**0.9 ± 0.3**	**0.4 ± 0.1**	**3.5 ± 0.9**	0.5 ± 0.2
Gammaridae	0.4 ± 0.2	**4.8 ± 1.7**	0.3 ± 0.2	0.7 ± 0.5	0.4 ± 0.3
Goeridae	0.4 ± 0.4	**0.5 ± 0.3**	0.2 ± 0.1	**2.3 ± 0.8**	1.3 ± 0.6
Gyrinidae	**0.1 ± 0**	0.1 ± 0.1	**0.8 ± 0.2**	**0.4 ± 0.2**	0.1 ± 0
Heptageniidae	**5.1 ± 1.8**	**6.0 ± 1.8**	**5.9 ± 1**	**6.6 ± 1.3**	**12.5 ± 1.4**
Hydraenidae	0.1 ± 0.1	**0.5 ± 0.1**	**2.6 ± 0.7**	0.3 ± 0.1	**1.5 ± 0.4**
Hydrobiidae	**1.9 ± 1.3**	**1.6 ± 0.5**	0.0 ± 0	**3.6 ± 1.5**	0.0 ± 0
Hydropsychidae	**6.5 ± 2.8**	**5.4 ± 1.3**	**8.3 ± 1.3**	**9.9 ± 1.2**	**7.8 ± 1.4**
Hydroptilidae	0.1 ± 0.1	**1.1 ± 0.6**	1.3 ± 0.7	0.3 ± 0.1	0.1 ± 0
Lepidostomatidae	**0.2 ± 0.1**	0.0 ± 0	**0.6 ± 0.2**	**1.2 ± 0.4**	0.2 ± 0.1
Leptoceridae	**1.0 ± 0.3**	0.2 ± 0.1	0.3 ± 0.2	**0.4 ± 0.1**	0.1 ± 0
Leptophlebiidae	**2.7 ± 2.3**	**1.6 ± 0.6**	**6.1 ± 1.1**	**1.3 ± 0.5**	**3.0 ± 0.7**
Leuctridae	**4.4 ± 0.8**	**7.9 ± 2.1**	**5.3 ± 0.9**	**5.3 ± 1**	**6.8 ± 0.7**
Limoniidae	0.3 ± 0.3	**0.4 ± 0.2**	0.2 ± 0.1	0.1 ± 0	**0.5 ± 0.1**
Nemouridae	1.6 ± 1.6	0.1 ± 0	**2.9 ± 0.7**	**1.9 ± 0.4**	**5.9 ± 0.9**
Oligochaeta	**6.8 ± 2.4**	**0.5 ± 0.1**	**4.8 ± 1.6**	**1.1 ± 0.4**	**0.5 ± 0.1**
Perlidae	0.2 ± 0.2	0.2 ± 0.1	**0.6 ± 0.2**	0.2 ± 0.1	**2.3 ± 0.6**
Philopotamidae	0.3 ± 0.3	0.3 ± 0.2	**1.6 ± 0.6**	**1.1 ± 0.3**	**1.3 ± 0.4**
Planariidae	0.0 ± 0	0.1 ± 0.1	**0.8 ± 0.2**	0.3 ± 0.2	0.1 ± 0
Polycentropodidae	**0.7 ± 0.3**	**0.4 ± 0.2**	**0.5 ± 0.2**	**0.5 ± 0.1**	0.1 ± 0.1
Psychomyiidae	**0.6 ± 0.2**	0.0 ± 0	0.3 ± 0.2	**0.9 ± 0.4**	0.2 ± 0.1
Rhyacophilidae	**0.8 ± 0.6**	**0.6 ± 0.1**	**0.6 ± 0.1**	**0.7 ± 0.1**	**1.1 ± 0.2**
Sericostomatidae	**0.4 ± 0.2**	**0.9 ± 0.4**	**1.5 ± 0.4**	**1.8 ± 0.6**	**3.4 ± 0.6**
Simuliidae	**2.1 ± 0.7**	**10.5 ± 3.5**	**1.6 ± 0.3**	**3.7 ± 1.5**	**5.1 ± 1.2**
Sphaeriidae	**3.5 ± 1.3**	0.1 ± 0.1	0.0 ± 0	**0.2 ± 0.1**	0.0 ± 0

Taxa with abundance lower than 10 individuals/2.5 m^2 in all types are suppressed.
Type 1: Aeshnidae; Calamoceratidae; Cordulegastridae; Corixidae; Dryopidae; Dytiscidae; Erpobdellidae; Gerridae; Glossiphoniidae; Gomphidae; Limnephilidae; Lymnaeidae; Planorbidae; Sialidae; ***Type 2***: Aeshnidae; Blephariceridae; Ceratopogonidae; Dixidae; Ephemeridae; Gerridae; Limnephilidae; ***Type 3***: Aeshnidae; Cordulegastridae; Dixidae; Dytiscidae; Erpobdellidae; Gerridae; Limnephilidae; Scirtidae; Sialidae; ***Type 4***: Aeshnidae; Ceratopogonidae; Cordulegastridae; Ephemeridae; Gerridae; Limnephilidae; Uenoidae; ***Type 5***: Ceratopogonidae; Dixidae; Dytiscidae; Erpobdellidae; Limnephilidae; Tipulidae.

probability of 69.60%, but there is also an estimated probability (28.19%) of being moderate (Fig. 5).

3.3. Stress gradient analysis

The extraction of five principal components explained 69.4% of the variance in anthropogenic stressors (Table 5). These main pressure gradients corresponded to sewages inputs (PCA #1; 27.8%); eutrophication (PCA #2; 18.9%); hydromorphological alterations (PCA #3; 9.7%), intensive agriculture (PCA #4; 6.9%) and a low intensity agriculture gradient opposed to an oxygenation gradient (PCA #5; 6.1%). All extracted gradients were included in the model selected as best (AIC = −1832.95; R^2 = 0.323; $F_{5,476}$ = 45.5, p < 0.001) and showed a negative relationship with EQR values (Fig. 6; note that the relationship between PCA #5 and low intensity agriculture is negative (Table 5)). However, a model excluding the intensive agriculture gradient (PCA #4) was also plausible (AIC = −1832.87; R^2 = 0.320; $F_{5,477}$ = 56.2, p < 0.001) and this variable was only marginally significant in the model selected as best. In the model selected as best, the eutrophication gradient (PCA #2) and the sewages input (PCA #1) were the most relevant predictors as they showed the largest standardized coefficients (Standardized coefficients and significance: PCA #1 = −0.277***; PCA #2 = −0.410***; PCA #3 = −0.162***; PCA #4 = −0.063 (n.s.); PCA #5= 0.189***).

4. Discussion

The NORTI classification system based on macroinvertebrates is a new predictive model designed to fulfill all scientific aspects required for classification systems under the present European

Table 3
Parameter estimates (B), standard error (SE) and significance (p) obtained with step-wise multinomial logit regression. The reference category is 5 ("small mountain rivers").

Stream type		Parameter	*B*	SE	*p*
Code	Name				
1	Major river axes	Intercept	−138.567	161.446	0.391
		UTM X (m)	0.000	0.000	0.981
		UTM Y (m)	0.000	0.000	0.417
		Altitude (m)	0.003	0.005	0.591
		Catchment slope (%)	−0.290	0.135	0.032
		Catchment area (km^2) *log-transformed*	8.710	2.957	0.003
		Calcareous substrate (%)	0.041	0.047	0.380
2	Mixed-calcareous rivers	Intercept	−440.262	273.591	0.108
		UTM X (m)	0.000	0.000	0.078
		UTM Y (m)	0.000	0.000	0.106
		Altitude (m)	0.004	0.005	0.433
		Catchment slope (%)	−0.454	0.158	0.004
		log Catchment area (km^2)*log-transformed*	4.671	2.466	0.058
		Calcareous substrate (%)	0.140	0.054	0.010
3	Mixed-siliceous rivers	Intercept	184.374	82.756	0.026
		UTM X (m)	0.000	0.000	0.256
		UTM Y (m)	0.000	0.000	0.031
		Altitude (m)	−0.009	0.003	0.004
		Catchment slope (%)	−0.066	0.046	0.156
		Catchment area (km^2) *log-transformed*	−0.826	1.472	0.575
		Calcareous substrate (%)	−0.006	0.019	0.766
4	Mixed lowland rivers	Intercept	106.122	87.897	0.227
		UTM X (m)	0.000	0.000	0.079
		UTM Y (m)	0.000	0.000	0.247
		Altitude (m)	−0.011	0.003	0.001
		Catchment slope (%)	−0.003	0.051	0.953
		Catchment area (km^2) *log-transformed*	1.066	1.425	0.455
		Calcareous substrate (%)	0.011	0.018	0.540

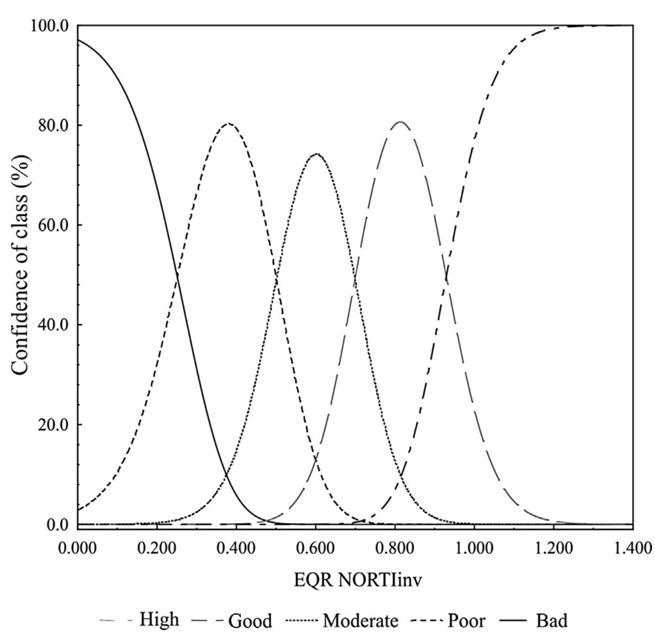

Fig. 5. Probability of ecological class membership (confidence of class) for any observed EQR value in the study area.

Table 4
Confidence matrix of cross-validation results (leave-one-out procedure). The number of reference sample sites is shown for each observed and predicted stream type, being correct predictions highlighted in bold. Percentage of correct assignment for each stream type and in global is also shown.

		Predicted stream type					Correct assignment
		Type 1	Type 2	Type 3	Type 4	Type 5	
Observed stream type	Type 1	**5**	0	0	2	1	**62.5%**
	Type 2	0	**11**	2	2	0	**73.3%**
	Type 3	0	1	**10**	4	5	**50.0%**
	Type 4	4	1	4	**11**	0	**55.0%**
	Type 5	1	1	3	1	**22**	**78.6%**
							Global: 64.8%

legislation requirements from WFD (i.e. typology; reference conditions, normative definitions). It responded to the main human pressures impairing stream and river ecosystems in Northern Spain supporting the accuracy of the classification system in detecting biological impact. Thereby, and consistent with previous assessment provided by predictive models (Wright et al., 1984; Moss et al., 1987; Parsons and Norris, 1996; Marchant et al., 1997; Hawkins et al., 2000), it demonstrated to be a valid system to assess the ecological status of streams and rivers. NORTI was developed with a spatial network of minimally disturbed sites (reference sites) following Stoddard et al. (2006), in agreement with pressure criteria used within the European Geographical Intercalibration Groups of the Common implementation strategy described in Pardo et al. (2012), but with more preventive thresholds for artificial and agricultural land uses in the catchment, as indicated by previous studies (Pardo et al., 2011). Moreover, the reference pressure thresholds tested in this study did not impacted the invertebrate communities of Northern Spain, supporting the designation of the reference streams and rivers in the studied area. On the other hand, the inclusion of WFD obligatory and optional variables in the river's typology was supported by the use of statistical criteria for acceptance, concluding in a reliable classification of stream types. Watershed scale (Bioclimatic region, geology, altitude) and stream-segment variables (mean channel and valley slopes), following Frissell et al. (1986), predicted the invertebrate assemblages. Since climate is similar in the area, geographic, geological and topographic descriptors captured the existing environmental variability in the studied catchments, and as in other comprehensive studies the large scale

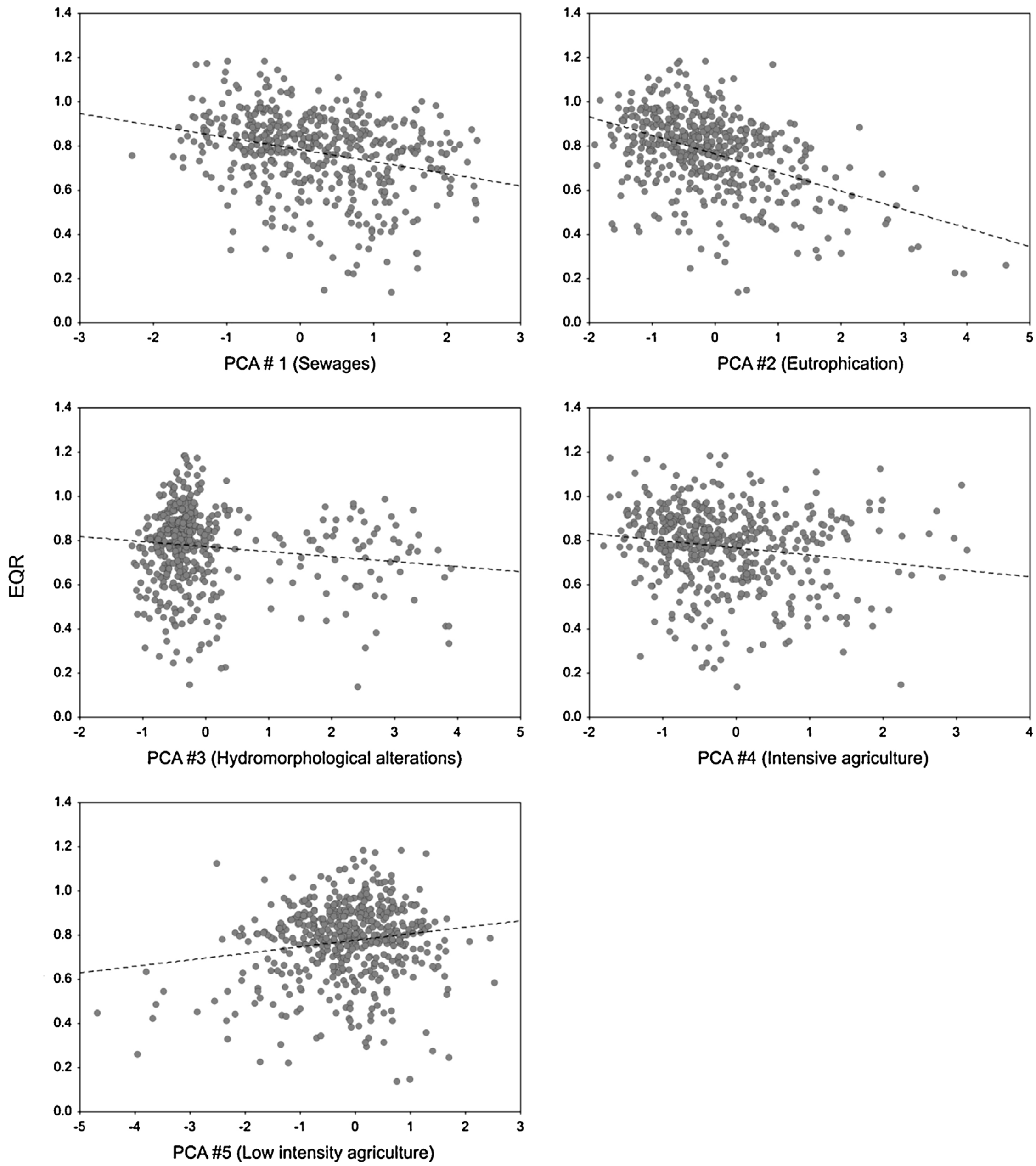

Fig. 6. Scatterplots showing the relationship between EQR values and PCA components.

factors explained most of the variation in invertebrate composition in European rivers (Verdonschot and Nijboer, 2004). Notwithstanding, it is noteworthy that high heterogeneity in environmental conditions exists in Northern Spain, in spite of its reduced area, and, thus, the need to identify different river types.

There was a broad coherence between NORTI and System A typologies that can be attributed to the fact that invertebrate communities are adapted to and specific of a variety of riverine local environmental conditions (Pardo and Armitage, 1997), as the ones covered in this study, across a wide range of spatial scales (Frissell et al., 1986). The NORTI performance in accounting for the natural variation in macroinvertebrate assemblages was more precise than the system A river typology, contrary to results from Aroviita et al. (2008) reporting a similar performance between the system A typology-based approach and a RIVPACS-type predictive model. The system A typology only registers the natural variability in geomorphological environmental characteristics, while the invertebrate river groups predicted by NORTI represent meaningful ecological river types (system B), both in terms of invertebrate communities and environmental characteristics.

The partitioning of the high natural environmental heterogeneity in the area into river types, allowed NORTI to respond significantly to the dominant disturbance gradients, more strongly to sewages and eutrophication, in spite of the family level used. Hawkins et al. (2000) found weak effects of land use variables on stream biota using RIVPACS predictive models, and attributed to the family level or to a low degree of impairment the result, suggesting that family assessments may be relatively inaccurate in environmentally heterogeneous regions. Meanwhile the North of Spain, comprising very old regions of the Hesperic massif, maintains a generally good conservation state, supporting diverse invertebrate assemblages (Pardo, 2000; Pardo and Alvarez, 2006), not preventing the family level to be sensitive to dominant pressures. In other studies, the application of Bray–Curtis to (RIVPACS)-type models to

Table 5
Variable loadings in Principal Components Analysis (PCA). Note that variable transformation are not specified. Highest variables loadings are indicated in bold for each axis.

	PCA #1 Sewages	PCA #2 Eutrophication	PCA #3 Hydromorphological alterations	PCA #4 Intensive agriculture	PCA #5 Low intensity agriculture
Explained variance	27.8%	18.9%	9.7%	6.9%	6.1%
N-NH_4	0.12	**0.77**	−0.05	0.11	−0.15
N-NO_2	0.24	**0.78**	0.02	0.21	−0.01
N-NO_3	0.05	0.35	−0.09	0.64	0.34
O_2	0.15	−0.46	0.00	−0.12	**0.63**
P-PO_4	0.06	**0.74**	0.05	0.21	−0.02
Artificial land use	0.41	0.46	0.06	0.17	0.17
Intensive agriculture	−0.23	0.09	0.03	**0.85**	0.07
Low intensity agriculture	0.63	−0.03	−0.12	0.17	**−0.51**
Natural areas	−0.14	−0.16	0.13	−0.89	0.25
Population density	**0.72**	−0.02	0.11	−0.04	−0.39
Urban wastewater	**0.82**	0.14	0.20	−0.07	0.11
Industrial wastewater	0.70	0.21	0.32	0.08	0.07
Domestic wastewater	**0.87**	0.15	0.21	−0.05	0.08
Dams number	0.20	0.01	**0.92**	−0.02	0.02
Dams height	0.06	−0.01	**0.90**	−0.08	−0.06
Riverside protection	0.46	0.23	0.33	−0.06	0.22
Transfer/diversions	0.36	−0.03	0.68	−0.07	0.19

observed and predicted assemblages directly, improved the relationships between O/E and stressors (Van Sickle, 2008), indicating that the compositional nature of BC, as used in this study, improves the indication of the biota to stressors.

The composition and abundance of invertebrate communities in reference sites was the basis for defining invertebrate assemblages. The types were characterized by a high taxa richness of 70 taxa per type. From the 35 dominant taxa (Table 2) characterizing the reference conditions, 30 corresponded to insects, and 19 corresponded to Ephemeroptera, Plecoptera and Trichoptera (EPT). EPT families showed high abundances (mostly *Leuctridae, Heptageniidae, Hydropsychidae* and *Baetidae)* as expected for fast flowing temperate rivers. Crustaceans (*Gammaridae*) were well represented in calcareous and mixed rivers, also appearing in main river axes, while being absent from siliceous or mixed streams. River type preferences were also manifested by the *Aphelocheiridae* bugs (mainly *Aphelocheirus occidentalis*) from main river axes, with bivalves *Sphaeridae* and mayflies *Caenidae* also for lowland rivers associated with less coarse sediments (gravel-silt). Interestingly, the *Hydrobiidae* (mostly *Potamopyrgus antipodarum*) was absent from mountainous and from siliceous rivers, indicative of the limitation of its dispersal in these river types. Our results highlight a dominance of insects of reophilous character in Northern Spanish streams and rivers, in agreement with other extensive studies where alkalinity, as a surrogate for geology, had a strong influence in taxa composition from a dominance of Ephemeroptera, Plecoptera and Trichoptera in less alkaline stream systems toward Mollusca and Crustacea (Death and Joy, 2004), but with the exception of the Diptera *Chironomidae*, that showed similar percentages (from 16% to 25%) under the minimally disturbed conditions of river types in this study.

The new statistical approach described here basically consists of two steps: first, abiotic variables are used to predict river types of biological significance. In other words, river types are originally defined by their invertebrate assemblages but, once the typology is defined, the physiographic catchment features are the ones to infer the type of new locations. Second, a new statistical approach was used to assess ecological status based on the biological community of the sample and the O/E relationship. The novelty lies in the fact that the Bray–Curtis similarity is calculated for such test site community and the type' reference community (median similarity value between each reference site and the median community for the types). This similarity informs about the deviation of the whole community from the expected one in reference conditions. The

Table 6
Direct correspondence in percentages of river sites between the system A river typology (WFD European intercalibration) and the 5 NORTI stream and river types (common intercalibration river types: RC2, small lowland siliceous – rock; RC3, small mid-altitude siliceous; RC4, medium lowland mixed; RC5, large lowland mixed; RC6, small, lowland, calcareous). In bold stronger agreements between river types.

NORTI stream and river types	System A (EU Intercalibration) typology				
	RC2	RC3	RC4	RC5	RC6
1. Main river axes		9.0	**42.1**	100.0	
2. Mixed calcareous rivers	4.1	3.6	8.8		**67.8**
3. Mixed siliceous rivers	**65.9**	**50.9**	9.7		2.3
4. Mixed lowland rivers	**30.1**	16.1	**37.0**		**21.6**
5. Small mountain streams		**20.5**	2.3		8.2

Bray–Curtis similarity value is a traditional "taxon base" β diversity measure (Magurran, 2004) that we used to standardize the similarity between any test site and the expected reference community for the site type (O/E), by dividing the observed similarity by the median of the expected similarity within the reference group. The Bray–Curtis has also shown to perform better than other diversity or community indices to detect impairment at the community level (Field et al., 1982; Perkins, 1983), being comparable its response to the population level response (Pontasch et al., 1989). Bailey et al. (2004) used MCDist in a similar way as in this study, to indicate how the reference condition varied in several study cases. They calculate the Bray–Curtis distance of the community from the average reference community (named MCDist) as single descriptor of the biota. In this study, we used the median value of the Bray–Curtis as the expected value for the reference state, and calculate with it the O/E value between any test site and the reference community.

The original predictive models relied on classification of reference stream communities and then used discriminant function analysis (DFA), a same purpose technique as the multinomial logistic model, to select a suite of habitat attributes that best matched the biological classification (Wright et al., 1984). We applied multinomial regression, another widely used methodology to classify qualitative variables, for its greater flexibility in statistical requirements and the reduction in computational time when using the obtained equations into software development. Multinomial regression does not require normal distribution for the residuals neither homoscedasticity (Guisande et al., 2006), and tends to provide higher percentages of classification than discriminant analysis (Guisande et al., 2011). For comparison purposes we performed a discriminant analysis with the six variables used in NORTI (some did not fulfill normality assumptions, neither transformed), and run the analysis for separate groups because the hypothesis of equal variance was not met (M box $p < 0.05$). The initial global percentage of correct classification (prior to cross-validation) of the multinomial regression was 78%, slightly higher than the 74.7% of the discriminant analysis that provided 4 significant functions.

The RIVPACS assessment was site specific and was done by comparing the list of expected taxa predicted by the model for the site with the observed taxa, also used in Australia (Simpson and Norris, 2000) and in the USA (Hawkins et al., 2000). This approach was adopted in Canada to the assessment of the Great Lakes (Reynoldson et al., 2000) and of the Fraser River in British Columbia (Reynoldson et al., 2001). The NORTI approach is similar to the Canadian BEAST by using ordination to conduct the assessment on species composition and abundance, and in the use of taxa composition in any test site to assess the deviation from the reference sites. However, the NORTI differs from it in the O/E calculation, as it uses the Bray–Curtis similarity in taxa composition to the reference, instead of using bands of biological quality (ellipses of probabilities) to assess how similar a test site is to the reference sites (Rosenberg et al., 1999). The O/E NORTI value is classified in classes from High to Bad, using the European intercalibrated boundaries. Moreover, the probability of assess confidence of WFD status class based on

EQR class limits (probability of belonging to a class or another) is given per each O/E individual value, calculated as indicated in this study.

NORTI assessment is recommended for summer (from June to September), as summer samples were the ones used for reference conditions establishment, and to prevent higher uncertainty due to samples collected at different times of the year. Seasonal changes in community composition might be responsible for a model lower performance when used in other seasons, as evidenced for other community derived predictive models (Reece et al., 2001). However, it should also be noted that the response of the NORTI model to pressures is good even when including spring samples and that the uncertainty level for the EQR of a site is similar to previous studies (Clarke, 2013). Shifts in community composition in the studied area are maximum between autumn-winter and summer months, being spring assemblages more similar to summer (Pardo, 2000). The recommendation of using summer samples is to ensure minimum disturbances from autumn-winter flow on the biota (Pardo, 2000), while low discharge levels influence maximum effluents impairment. The NORTI predictive-model approach satisfies the main requirements for any WFD biological monitoring, being presently used by the water authorities of Northern Spain to assess the ecological status of rivers.

Acknowledgements

This study was funded by the Water Autorities of North Spain (Confederaciones Hidrográficas del Cantábrico y Miño-Sil, grants CO-140-04 and CO-048-08), and previous projects FD97-1481-C02-02 (AMB). We will like to thank Agustín Monteoliva and the Ecohydros S.L. team for their hard work and efforts in field and laboratory work during many years and to Fernando Sandoval for his assistance with maps.

References

Aroviita, J., Koskenniemi, E., Kotanen, J., Hämälaïnen, H., 2008. A priori typology-based prediction of benthic macroinvertebrate fauna for ecological classification of rivers. Environ. Manage. 42, 894–906.

Bailey, R.C., Kennedy, M.G., Dervish, M.Z., Taylor, R.M., 1998. Biological assessment of freshwater ecosystems using a reference condition approach: comparing predicted and actual benthic invertebrate communities in Yukon streams. Freshw. Biol. 39, 765–774.

Bailey, R.C., Norris, R.H., Reynoldson, T.B., 2004. Bioassessment of Freshwater Ecosystems: Using the Reference Condition Approach. Springer, New York, USA.

Barbour, M.T., Gerritsen, J., Snyder, B.D., Stribling, J.B., 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish. EPA 841-B-99-002, 2nd ed. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Bennett, C., Owen, R., Birk, S., Buffagni, A., Erba, S., Mengin, N., Murray Bligh, J., Ofenböck, G., Pardo, I., van de Bund, W., Wagner, F., Wasson, J.G., 2011. Bringing European river quality into line: an exercise to intercalibrate macro-invertebrate classification methods. Hydrobiologia 667, 31–48.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multi-Model Inference. A Practical Information-Theoretic Approach, 2nd ed. Springer, Springer Verlag, New York, USA, pp. 488.

Clarke, K.R., 1993. Non-parametric multivariate analyses of changes in community structure. Aust. J. Ecol. 18, 117–143.

Clarke, K.R., Gorley, R.N., 2006. PRIMER v6: User Manual/Tutorial. PRIMER-E, Plymouth.

Clarke, R.T., 2013. Estimating confidence of European WFD ecological status class and WISER Bioassessment Uncertainty Guidance Software (WISERBUGS). Hydrobiologia 704, 39–56.

Clarke, R.T., Hering, D., 2006. Errors and uncertainty in bioassessment methods – major results and conclusions from the STAR project and their application using STARBUGS. Hydrobiologia 566, 433–439.

Darling, E.S., Côte, I.M., 2008. Quantifying the evidence for ecological synergies. Ecol. Lett. 11, 1278–1286.

Death, R.G., Joy, M., 2004. Invertebrate community structure in streams of the Manawatu–Wanganui region, New Zealand: the roles of catchment versus reach scale influences. Freshw. Biol. 49, 982–997.

Directive, 2000/60/EC of the European Parliament and of the Council establishing a framework for the Community action in the field of water policy.

European Environment Agency, 1999. Environmental Indicators: Typology and Overview. European Environment Agency, Technical Report No. 25, Copenhagen.

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

Field, J.G., Clarke, K.R., Warwick, R.M., 1982. A practical strategy for analysing multispecies distribution patterns. Mar. Ecol. Prog. Ser. 8, 37–52.

Frissell, C.A., Liss, W.J., Warren, C.E., Hurley, M.D., 1986. A hierarchical framework for stream habitat classification: viewing streams in a watershed context. Environ. Manage. 10, 199–214.

Green, R.H., 1999. Review of potentially applicable approaches to benthic invertebrate data analysis and interpretation. Aquatic Effects Technology Evaluation (AETE) Program, Canada Centre for Mineral and Energy Technology, Natural Resources Canada, Ottawa.

Guisande, C., Barreiro, A., Maneiro, I., Riveriro, I., Vergara, A., Vaamonde, A., 2006. Tratamiento de datos. Diaz de Santos, España.

Guisande, C., Vaamonde, A., Barreiro, A., 2011. Tratamiento de datos con R, Statistica y SPSS. Diaz de Santos, España.

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

Hossain, M., Wright, S., Petersen, L.A., 2002. Comparing performance of multinomial logistic regression and discriminant analysis for monitoring access to care for acute myocardial infarction. J. Clin. Epidemiol. 55, 400–406.

Johnson, R.K., Hering, D., 2009. Response of taxonomic groups in streams to gradients in resource and habitat characteristics. J. Appl. Ecol. 46, 175–186.

Kanninen, A., Hellsten, S., Hämäläinen, H., 2013. Comparing stressor-specific indices and general measures of taxonomic composition for assessing the status of boreal lacustrine macrophyte communities. Ecol. Indic. 27, 29–43.

Karr, J.R., Chu, E.W., 1999. Restoring Life in Running Waters: Better Biological Monitoring. Island Press, Washington, DC, USA.

Kelly, M., Gómez-Rodríguez, C., Kahlert, M., Almeida, S.F.P., Bennet, C., Bottin, M., Delmas, F., Descyg, J.P., Dörflingerh, G., Kennedyi, B., Marvanj, P., Opatrilovak, L., Pardo, I., Pfisterl, P., Roseberyf, J., Schneiderm, S., Vilbastenet, S., 2012. Establishing expectations for pan-European diatom based ecological status assessments. Ecol. Indic. 20, 177–186.

Kelly, M., Bennion, H., Burgess, A., Ellis, J., Juggins, S., Guthrie, R., 2009. Uncertainty in ecological status assessments of lakes and rivers using diatoms. Hydrobiologia 633, 5–15.

Kokeš, J., Zahradkova, S., Nemejcova, D., Hodovsky, J., Jarkovsky, J., Soldan, T., 2006. The PERLA system in the Czech Republic: a multivariate approach for assessing the ecological status of running waters. Hydrobiologia 566, 343–354.

Lorenz, A., Hering, D., Feld, C.K., Rolauffs, P., 2004. A new method for assessing the impact of hydromorphological degradation on the macroinvertebrate fauna of five German stream types. Hydrobiologia 516, 107–127.

Magurran, A.E., 2004. Measuring Biological Diversity. Blackwell Publishing, Malden, USA.

Marchant, R., Hirst, A., Norris, R.H., Butcher, R., Metzeling, L., Tiller, D., 1997. Classification and prediction of macroinvertebrate communities from running waters in Victoria, Australia. J. North Am. Benthol. Soc. 16, 664–681.

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

Pardo, I., 2000. Patterns of community assembly in a fourth order stream. Arch. Hydrobiol. 148, 301–320.
Pardo, I., Poikane, S., Bonne, W., 2011. Revision of the Consistency in Reference Criteria Application in the Phase I of the European Intercalibration Exercise. JRC Scientific and Technical Reports. Office for Official publications of the European Communities, Luxembourg.
Pardo, I., Gómez-Rodríguez, C., Wasson, J.G., Owen, R., van de Bund, W., Kelly, M., 2012. The European reference condition concept: a scientific and technical approach to identify minimally impacted river ecosystems. Sci. Total. Environ. 420, 33–42.
Pardo, I., Alvarez, M., 2006. Comparison of resource and consumer dynamics in Atlantic and Mediterranean streams. Limnetica 25, 271–286.
Pardo, I., Armitage, P.D., 1997. Species assemblages as descriptors of mesohabitats. Hydrobiologia 344, 111–128.
Parsons, M., Norris, R.H., 1996. The effect of habitat specific sampling on biological assessment of water quality using a predictive model. Freshw. Biol. 36, 19–434.
Perkins, J.L., 1983. Bioassay evaluation of diversity and community comparison indexes. J. Water Pollut. Control Fed. 55, 522–530.
Pontasch, K.W., Smith, E.P., Cairns, J., 1989. Diversity indices, community comparison indices and canonical discriminant analysis: interpreting the results of multispecies toxicity tests. Water Resour. 23, 1229–1238.
Puccinelli, E., 2011. How can multiple stressors combine to influence ecosystems and why it is important to address this question. Integr. Environ. Assess. Manage. 8, 201–202.
Quinn, G.P., Keough, M.J., 2002. Experimental Design and Data Analysis for Biologists. Cambridge University Press, UK.
Reece, P.F., Reynoldson, T.B., Richardson, J.S., Rosenberg, D.M., 2001. Implications of seasonal variation for biomonitoring with predictive models in the Fraser River catchment, British Columbia. Can. J. Fish. Aquat. Sci. 58, 1411–1418.
Reynoldson, T.B., Day, K.E., Pascoe, T., 2000. The development of the BEAST: a predictive approach for assessing sediment quality in the North American Great Lakes. In: Wright, J.F., Sutcliffe, D.W., Furse, M.T. (Eds.), Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques. Freshwater Biological Association, Ambleside, Cumbria, UK, pp. 165–180.
Reynoldson, T.B., Rosenberg, D.M., Resh, V.H., 2001. Comparison of models predicting invertebrate assemblages for biomonitoring in the Fraser River catchment, British Columbia. Can. J. Fish. Aquat. Sci. 58, 1395–1410.
Reynoldson, T.B., Wright, J.F., 2000. The reference condition: problems and solutions. In: Wright, J.F., Sutcliffe, D.W., Furse, M.T. (Eds.), Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques. Freshwater Biological Association, Ambleside, Cumbria, UK, pp. 293–303.
Reynoldson, T.B., Norris, R.H., Resh, V.H., Day, K.E., Rosenberg, D.M., 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. J. North Am. Benthol. Soc. 16, 833–852.
Reynoldson, T.B., Bailey, R.C., Day, K.E., 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.
Rosenberg, D.M., Reynoldson, T.B., Resh, V.H., 1999. Establishing Reference Conditions for Benthic Biomonitoring in the Fraser River Catchment, British Columbia, Canada. Fraser River Action Plan, Environment Canada, Vancouver, B.C. DOE FRAP 1998-32.
Simpson, J., Norris, R.H., 2000. Biological assessment of water quality: development of AusRivAS models and outputs. In: Wright, J.F., Sutcliffe, D.W., Furse, M.T. (Eds.), Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques. Freshwater Biological Association, Ambleside, Cumbria, UK, pp. 125–142.
Southwood, T.R.E., 1977. Habitat, the templet for ecological strategies? J. Anim. Ecol. 46, 337–365.
Stoddard, J.L., Larsen, D.P., Hawkins, C.P., Johnson, R.K., Norris, R.H., 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. Ecol. Appl. 16, 1267–1276.
Van Sickle, J., 2008. An index of compositional dissimilarity between observed and expected assemblages. J. North Am. Benthol. Soc. 27, 227–235.
Verdonschot, P.F.M., Nijboer, R.C., 2004. Testing the European stream typology of the Water Framework Directive for macro-invertebrates. Hydrobiologia 516, 37–55.
Wright, J.F., 2000. An introduction to RIVPACS. In: Wright, J.F., Sutcliffe, D.W., Furse, M.T. (Eds.), Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques. Freshwater Biological Association, Ambleside, Cumbria, UK, pp. 1–24.
Wright, J.F., Moss, D., Armitage, P.D., 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.
Wrona, F.J., Culp, J.M., Davies, R.W., 1982. Macroinvertebrate subsampling: a simplified apparatus and approach. Can. J. Fish. Aquat. Sci. 39, 1051–1054.
Zelinka, M., Marvan, P., 1961. Zur präzisierung der biologischen klassifikation der reinheit fließender gewässer. Arch. Hydrobiol. 57, 389–407.