Ecological Indicators 90 (2018) 519–528

Contents lists available at ScienceDirect

Ecological Indicators

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

Original Articles

A predictive diatom-based model to assess the ecological status of streams and rivers of Northern Spain

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

[a] Departamento de Ecología y Biología Animal, Universidad de Vigo, 36310 Vigo, Spain
[b] Departamento de Zoología, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain
[c] Departamento de Informática, Universidad de Vigo, 36310 Vigo, Spain
[d] Acadia University Wolfville, Nova Scotia, Canada

ARTICLE INFO

Keywords:
Bioassessment
Diatoms
Predictive model
Reference conditions
Rivers
Stressors

ABSTRACT

In this study we developed a predictive diatom-based model to assess the ecological status of streams and rivers of Northern Spain. Diatom samples were collected with standard protocols over stones from 676 sites distributed along existing environmental conditions across Northern Spain, during seven years between 2002 and 2008 (n = 1056 samples). This dataset included a network of 91 reference sites selected by using criteria that confirmed the absence of relevant human pressures according WFD. A multinomial logistic regression using GAAC cluster-derived reference sites group as response variable was performed. The independent variables included obligatory typology factors (WFD System A typology descriptors), and other optional typology B descriptors were included in the model performed on a forward stepwise procedure. The Ecological Quality Ratios (EQRs) were obtained by dividing the observed similarity between the diatoms composition in each sample by the expected median similarity of each type reference diatom community. The model predictions (EQRs) responded significantly to eutrophication and intensive agriculture pressures, but were not related with sewages, hydromorphological alterations and extensive agriculture pressures. These results demonstrated the accuracy of the diatom model in predicting nutrient enrichment in Northern Spanish rivers and streams.

1. Introduction

Predictive models are relatively recent approaches to assess water quality, aiming to evaluate the quality of a stream or river site as the degree of alteration of its biota in relation to the same stream or river type biological reference communities (Reynoldson et al., 1997). There have been multiple approaches to develop predictive models using macroinvertebrate communities, proving to be powerful tools in the assessment of water quality (RIVPACS, Wright, 1995; BEAST, Reynoldson, et al., 1995; AUSRIVAS, Simpson and Norris, 1997; NORTI invertebrates, Pardo et al., 2014). Meanwhile, models based on diatoms have only lately been developed for different regions following similar methodologies as for invertebrates, and also showing significant response to several human pressures (Mazor et al., 2006; Philibert et al., 2006; Feio et al., 2009, 2012; Almeida and Feio, 2012). Diatoms are considered among the best indicators for eutrophication or nutrient enrichment conditions, even better than other organisms such as macrophytes and fish (Hering et al., 2006; Johnson et al., 2006).

During the last 20 years, benthic diatoms have gone from being poorly used to be one of the most used bioindicators together with benthic invertebrates for the assessment of river water quality in Europe (Poikane et al., 2016). Several European studies started to use diatoms as indicators of water quality (Kelly and Whitton, 1998; Kwandrans et al., 1998; Dell'Uomo et al., 1999) in parallel with advances in the field from other worldwide regions (Jüttner et al., 1996; Chessmann et al., 1999). However, significant advances in their use as bioindicators occurred since 2000 with the implementation of the Water Framework Directive (WFD, 2000/60/CE). WFD requirements led to the adjustment of existing assessment systems for water quality and encouraged the design of new classification systems for the ecological status of rivers. The new systems included new approaches for the concept of European reference conditions (Pardo et al., 2011, 2012), the establishment of national and European river typologies, the classification system for the ecological status and ecological classes boundaries, and the mandatory intercalibration exercise between European countries (van de Bund, 2009).

---

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

https://doi.org/10.1016/j.ecolind.2018.03.042
Received 24 January 2018; Received in revised form 16 March 2018; Accepted 17 March 2018

In Spain, assessment of the ecological status of rivers using diatoms is generally approached with the use of diatom metrics (Gomà et al., 2004; Blanco et al., 2008; Delgado et al., 2010; Martín et al., 2010; Delgado et al., 2012; Álvarez-Blanco et al., 2013), as integrators that summarize the complex ecology of stream diatom communities. In particular the Specific Polluosensitivity Index (IPS, Cemagref, 1982) has been used to assess the ecological status of Mediterranean rivers (Gomà et al., 2004; Blanco et al., 2008) and has enabled the comparison of diatom communities between different catchment areas (Martín et al., 2010). Moreover, the IPS diatom metric is used as the national classification system for all stream and rivers types in Spain (RD 817/2015), in spite of strong regional differences in the Spanish climate and geomorphological characteristics. These environmental differences determine distinctive attributes in geomorphology, biochemistry and hydrology causing variations in physical conditions and chemical composition and, consequently in the biogeography and composition of diatom communities (Stevenson, 2014; Delgado and Pardo, 2014).

A previous study in North Western Spain (Galicia), developed a multimetric diatom index (MDIAT) to classify the ecological status of the rivers of the region (Delgado et al., 2010), a system that was intercalibrated during the first European exercise (Kelly et al., 2009; van de Bund, 2009). Meanwhile, MDIAT only covered the more siliceous and mixed river types in NW Spain, there was a need to develop a more representative and comprehensive WFD classification system in order to cover all existing river types in the studied regions (e.g. from calcareous to siliceous).

The purpose of this study was to develop a predictive model based on diatom assemblages to assess the ecological status of rivers and streams of Northern Spain meeting the scientific requirements of WFD. We assembled biotic and abiotic data from minimally disturbed sites or reference sites where the absence of significant pressure criteria was verified (Pardo et al., 2012,) to develop the predictive diatom-based model NORTIdiat (NORThern Spain Indicator System for diatoms). The model response was tested on stress gradients including multiple stressors to assess the sensitivity of the predictions. Finally, the NORTIdiat was intercalibrated within the European intercalibration exercise based on its comparison with the ICM multimetric index, within the Central Baltic Geographical Intercalibration Group (CB GIG) to which Northern Spanish streams and rivers belong.

2. Material and methods

2.1. Study area

The study area is located in the Northern part of Spain, and covers the river basins managed by four River Basin Authorities (Galicia-Costa, Miño-Sil River basin, Cantabrian basins and Basque basins) (Fig. 1). The area covers 38,450 km^2, from Galicia in the Northwest to the Basque Country in the East. The Cantabrian and "Galicia-Costa" river basins are characterized by small, medium-high elevation gradient streams running from coastal mountains to the sea. Meanwhile, on the Western Atlantic coast the biggest and longest Miño-Sil Basin flows East-Westwards forming long valleys at low altitude. A more detailed description of the study area can be found in Pardo et al. (2014).

2.2. Sampling design

In this study 676 sampling sites distributed along the whole region of study and covering the wide existing environmental gradients in the area were sampled. Each sampling site was visited at least once during the period 2002–2008, generally in summer, except in 2006 and 2008, when several sites (115 and 30 sites, respectively) were visited also in spring. A total number of 1056 samples were collected. The sampling effort per year was not equally distributed along the study period, being samples scarce in 2002 and 2007 (28 and 17 samples per year, respectively) but in the rest of the years (≥146 per year). The maximum number of sites sampled in a year was 466 (in 2003).

2.3. Data collection and laboratory methods

2.3.1. Abiotic data

The same dataset of environmental variables as in Pardo et al. (2014) was used, as diatoms and invertebrates were collected simultaneously during sampling. Briefly, filtered water samples were collected in most sampling visits using standard methods and transported to the laboratory to quantify potential stressors (phosphates as $P\text{-}PO_4^{-3}$ mg/L; nitrates as $N\text{-}NO_3^-$ mg/L; nitrites as $N\text{-}NO_2^-$ mg/L; ammonium as $N\text{-}NH_4^+$ mg/L), and *in situ* Oxygen saturation was measured with a portable oximeter (YSI 556 MPS).

Sampling sites were also characterized according to the obligatory descriptors in the WFD typology System B (latitude [m], longitude [m], altitude [m], catchment area [km^2] and geology represented solely by the percentage of calcareous geology), together with other optional descriptors (catchment slope [%], mean channel slope [%], maximum and mean catchment altitude [m], annual precipitation [mm]) (Table 1). The network of reference sites identified by Pardo et al. (2014) of 108 reference sites was used in this study. However, diatom samples were available in only 91 of these reference sites. For detailed information about selection of reference sites see Pardo et al. (2014).

2.3.2. Diatom samples

In each site, five stones were randomly sampled in the middle of the stream from riffle areas. Diatom samples were pooled by scraping the stone surfaces with a toothbrush in order to remove all the periphyton (Kelly et al., 1998; EN 13946, 2003), and after that were fixed with formaldehyde (4%). Organic matter was eliminated using hydrogen peroxide (33%) and HCl (37%) was added to remove the calcium carbonate. Finally, after rinsing with distilled water, permanent slides were mounted using Naphrax®, a synthetic mounting medium with high refractive index (r.i. = 1.74) to identify diatom species. Diatoms were observed and identified at the lowest taxonomic level, using an optical microscope (Olympus BX41). In each sample a minimum of 400 diatom valves were identified and counted at 1000× magnification (NA 1.25). The identification and nomenclature were based on (Krammer and Lange-Bertalot, 1986, 1988, 1991), Lange-Bertalot (1993) as well as on recent bibliography.

2.4. Data analysis

2.4.1. Group assignment: Model construction and validation

We aimed to identify the natural groupings in diatom assemblages of reference samples. For that we applied Group-Average Agglomerative Clustering (GAAC) to the similarity Bray-Curtis matrix of log-transformed relative abundances of diatoms. GAAC measures the dissimilarity between two clusters by using the dissimilarities average between all two objects combinations (Quinn and Keough, 2002). Four major groups were differentiated at the 29% similarity level, although the smallest group was eliminated as it was only composed of two samples. The largest group was further subdivided in two groups at 36% similarity level. Group consistency was evaluated with an ANOSIM analysis (Clarke 1993) in order to ensure that each group was significantly different from the rest. Complementary, we ran a non-Metric Multidimensional Scaling (MDS) ordination on the same similarity matrix to visualize distinctiveness of groups and their separation on the ordination space. The software PRIMER v.6 (Clarke and Gorley, 2006) was used for the previous analyses.

Multinomial logistic regression using GAAC cluster-derived reference sites group as response variable was performed to predict the diatom groups from abiotic descriptor variables, using 10 environmental typology A and B potential predictors. Multinomial logistic regression can predict nominal variables (4 diatom assemblages) with a set of independent variables. The procedure for automatic forward

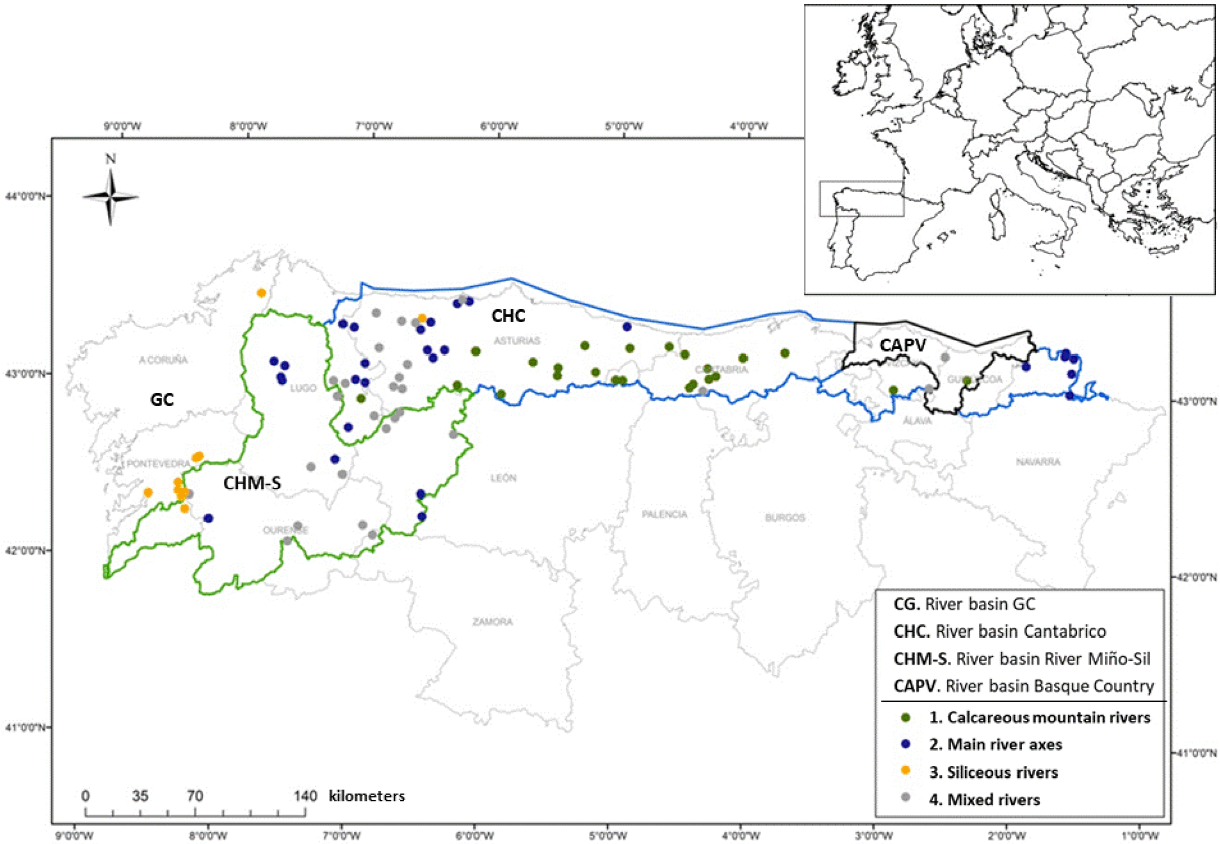

Fig. 1. Location of the study area in North Spain, and the network of reference sites used for the development of the NORTIdiat model. 1, Calcareous mountain rivers; 2, Main river axes; 3, Siliceous rivers; 4, Mixed rivers. CG: Water authorities Galicia-Costa; CHM-S: Water authorities Miño-Sil; CHC: Water authorities Cantábrico; CAPV: Water authorities País Vasco.

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

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

stepwise was selected, and a $p < 0.05$ threshold set for predictor inclusion, and a $p > 0.10$ for predictor exclusion. A cross-validation procedure "leave-one-out" was conducted to assess the predictive capacity of the model. To validate the resulting model groups a confusion matrix was generated to show the correspondence between the GAAC cluster-derived group and the multinomial regression model group's prediction, by removing one sample each time form the training set. Model equations were applied to predict the stream type (i.e. diatom group) of all non-reference sites used in this study. The multinomial logistic regression analyses were run with SPSS v. 15.

2.4.2. EQR estimation

The EQRs were calculated by dividing the distance observed between any test sample and the median distance of the reference samples for its type (i.e. group predicted with the multinomial logit regression) in a 2-dimension NMDS plot based on Bray-Curtis similarity. In order to compare two single values, the set of reference samples was summarized in its centroid. The longer the distance between any sample and the centroid, the worse it's ecological status. This is a result of its community being largely different from the one expected if no anthropogenic activities were present. In order to comply with the normative definition of EQR, the observed distance was divided by the reference expected value under no anthropogenic impact. The reference value was estimated as the median of the distances calculated between all reference samples and the centroid of the NMDS ordination run with the reference samples included in each of the four diatom groups.

2.5. Stress gradient analysis

To check the EQRs response to stress gradients, a multiple regression model was conducted, using EQR values as dependent variables and the main stress gradients extracted in the environmental dataset as potential predictor factors. The same stress gradients as in Pardo et al. (2014, Table 5) were used, being the axes extracted from the multivariate analyses (Principal Component Analyses, PCA) of the main pressures and stressors affecting the whole dataset (reference and impacted sites). Variables included in the PCA were: land-use activities (artificial surfaces, intensive and low-intensity agriculture areas), and natural/semi-natural areas in the catchments estimated from CORINE land cover maps (Coordination of Information on the Environment, Land Cover 2000). The river authorities provided information on anthropogenic activities/interventions for each sampling site and at the catchment scale: sewage volume differentiating between domestic wastewater, urban wastewater and industrial wastewater; population density; the total number of dams upstream; upstream dam's cumulative height per catchment area; river protection (levees) and number of water transfers and diversions per catchment area, together with nutrient concentrations. Environmental gradients were extracted from PCA using varimax rotation and explained a total of 69.4% of the variance in anthropogenic stressors: (PCA #1: sewages input (highest loading of the following variables, domestic and urban wastewater and population density), 27.8%; PCA #2: eutrophication gradient (highest loadings of ammonium, nitrite and phosphate), 18.9%; PCA #3: hydromorphological alterations (highest loadings of dams number and height), 9.7%; PCA #4: intensive agriculture (highest loading of% intensive agriculture in the catchment), 6.9%; PCA #5: low intensity agriculture gradient opposed to an oxygenation gradient (highest loading of dissolved Oxygen in opposition to low intensive agriculture), 6.1%).

In the regression analysis when several samples exist for a given locality we used the mean EQR. The selection of variables was performed using the "best subset model approach" (see Burnham and Anderson, 2002) using the AIC criterion. Plausible models were those having the lowest AIC value and AIC value < 10.

3. Results

3.1. Group assignment: Model construction and validation

A total of 132 taxa were identified in the reference sites. *Achnanthidium minutissimum* (ADMI) and *Cocconeis euglypta* (CEUG) were the most frequent taxa, which appeared in more than 96% and 80% of the reference samples, respectively.

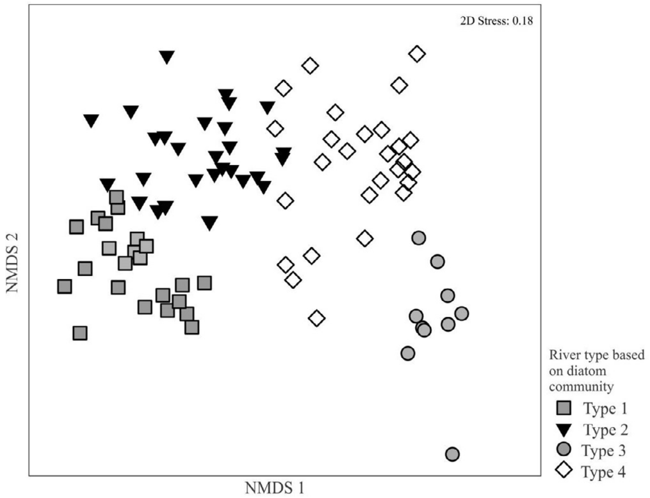

Fig. 3. Bray-Curtis based non-metric multidimensional scaling (NMDS) ordination of diatom reference samples. Symbols correspond to resulting river types extracted from GAAC clustering.

Based on GAAC clustering, four groups (i.e. types of diatoms assemblages) were identified in the reference samples network (Fig. 2). ANOSIM analyses evidenced that each group was significantly different from the rest (Global ANOSIM R = 0.794; $p < 0.01$; all pair-wise ANOSIM R > 0.571; $p < 0.001$). Group's separation was similarly evidenced in a 2-dimension NMDS ordination of samples (Fig. 3). The diatom assemblages corresponding to each of the 4 groups or the biological reference conditions for each river type are shown in Table 2. In the first group there were 5 dominant species: *Achnanthidium pyrenaicum, A. minutissimum, Gomphonema pumilum, Cocconeis lineata* and *C. euglypta.* Group 2 was characterized by the abundance of *A. minutissimum, C. euglypta* and *Achnanthidium subatomus.* Group 3 showed high abundances of *A. minutissimum,* followed by the less abundant, *Eunotia subarcuatoides* together with *E. intermedia* and *E. minor, Surirella roba, Navicula angusta* and *Peronia fibula.* Finally, the fourth group had a

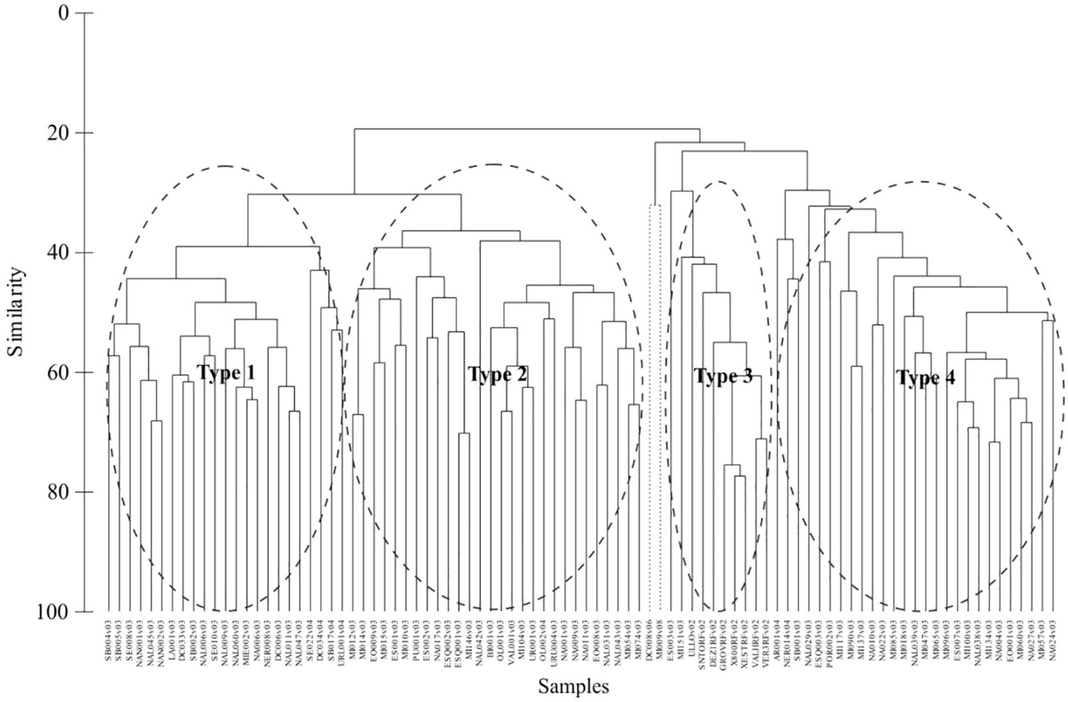

Fig. 2. Dendrogram showing the classification of reference sites according to their diatom composition. Identified stream types are shown. Dashed lines correspond to excluded samples.

Table 2
Diatom assemblage composition (median relative abundance) of stream types identified by means of GAAC clustering. Only data from the network of reference samples is used.

		Stream type based on diatom community			
		1	2	3	4
ADAT	*Achnanthidium atomus*	1.0	0.0	0.0	0.0
ADMI	*Achnanthidium minutissimum*	20.0	55.0	297.5	85.0
ADPY	*Achnanthidium pyrenaicum*	177.0	0.0	0.0	0.0
ADSU	*Achnanthidium subatomus*	0.0	25.5	0.0	0.0
APED	*Amphora pediculus*	1.0	2.0	0.0	0.0
CPED	*Cocconeis pediculus*	1.0	0.0	0.0	0.0
CPLE	*Cocconeis euglypta*	10.0	45.0	0.0	2.0
CPLI	*Cocconeis lineata*	13.0	7.0	0.0	0.0
COPL	*Cocconeis pseudolineata*	1.0	5.0	0.0	0.0
DMES	*Diatoma mesodon*	0.0	0.0	0.0	1.0
ESLE	*Encyonema silesiacum*	1.0	0.0	0.0	0.0
EOMI	*Eolimna minima*	0.0	2.0	0.0	0.0
EUIN	*Eunotia intermedia*	0.0	0.0	6.5	0.0
EMIN	*Eunotia minor*	0.0	0.0	1.0	1.0
ESUB	*Eunotia subarcuatoides*	0.0	0.0	18.0	0.0
FRHO	*Frustulia rhomboides*	0.0	0.0	0.0	1.0
GEXL	*Gomphonema exilissimun*	0.0	0.0	0.0	2.0
GPUM	*Gomphonema pumilum*	14.0	0.0	0.0	0.0
GRHB	*Gomphonema rhombicum*	0.0	3.0	0.0	97.0
NAAN	*Navicula angusta*	0.0	0.0	7.0	0.0
NCTE	*Navicula cryptotenella*	1.0	1.5	0.0	0.0
NRCH	*Navicula reichardtiana* var. *reichardtiana*	1.0	0.0	0.0	0.0
NTPT	*Navicula tripunctata*	4.0	0.0	0.0	0.0
NDIS	*Nitzschia dissipata*	1.0	0.0	0.0	0.0
NFON	*Nitzschia fonticola*	2.0	0.0	0.0	0.0
PFIB	*Peronia fibula*	0.0	0.0	2.0	0.0
PLFR	*Planothidium frequentissimun*	0.0	1.0	0.0	0.0
POBG	*Psammothidium oblongellum*	0.0	0.5	0.0	7.0
RSIN	*Reimeria sinuata*	1.0	2.0	0.0	0.0
SRBA	*Surirella roba*	0.0	0.0	10.0	0.0

Table 3
Parameter estimates (B), standard error (SE) and significance (p) obtained with step-wise multinomial logit regression. The reference category is 4 (Mixed rivers).

Stream type		B	SE	p
1	Intercept	−920.372	394.172	0.020
	Maximum altitude (m)	0.006	0.002	0.014
	Precipitation	0.013	0.006	0.024
	Altitude (m)	0.006	0.004	0.137
	Catchment area (km^2) log-*transformed*	4.296	1.990	0.031
	Calcareous substrate (%)	0.158	0.055	0.004
	UTM Y (m)	0.000	0.000	0.021
2	Intercept	6.669	56.892	0.907
	Maximum altitude (m)	−0.001	0.001	0.194
	Precipitation	−0.001	0.001	0.347
	Altitude (m)	−0.001	0.002	0.470
	Catchment area (km^2) log-*transformed*	1.495	0.815	0.067
	Calcareous substrate (%)	0.007	0.013	0.586
	UTM Y (m)	0.000	0.000	0.935
3	Intercept	1048.429	762.048	0.169
	Maximum altitude (m)	−0.014	0.010	0.152
	Precipitation	−0.018	0.015	0.220
	Altitude (m)	−0.021	0.015	0.169
	Catchment area (km^2) log-*transformed*	−24.997	19.690	0.204
	Calcareous substrate (%)	−7.639	7.093	0.282
	UTM Y (m)	0.000	0.000	0.168

species assemblage dominated by *Gomphonema rhombicum, A. minutissimum, Cocconeis euglypta, Gomphonema exilissimun, Frustulia romboides, Eunotia minor* and *Diatoma mesodon.*

Six descriptors (UTM Y, catchment area, altitude, maximum altitude, precipitation and calcareous substrate) were selected in the step-wise multinomial logit model (Table 3) and explained 67% of the variability (Nagelkerke pseudo-R^2 = 0.87). Cross-validation results evidenced acceptable discrimination ability since model parameters correctly classify 65.9% of the cases (Table 4). The resulting river types were defined according to their most distinctive variables range of values and distinctiveness based on the environmental descriptors selected by the model (Table 3, Fig. 4), and where accordingly named as follows: Type 1: Calcareous mountain rivers; Type 2: Main river axes; Type 3: Siliceous rivers and Type 4: Mixed rivers (represented spatially in Fig. 1).

Table 4

Confidence matrix of cross-validation results (leave-one-out procedure). The reference site sample numbers are shown for each observed and predicted stream type, with correct predictions highlighted in bold. The percentage of correct assignment for each stream type and in global is shown.

		Predicted river type				Correct assignment
		Type 1	Type 2	Type 3	Type 4	
Observed river type	Type 1	**19**	3	0	0	86.40%
	Type 2	3	**16**	2	9	53.30%
	Type 3	0	3	**7**	2	58.30%
	Type 4	1	6	1	**16**	66.80%
						Global: 65.91%

3.2. Ecological quality ratios (EQR)

EQR values ranged from 0.120 to 1.570 and were obtained after dividing the observed distance for a particular test sample by the value observed in reference conditions for its type (i.e. expected value: Type 1 = 0.620; Type 2 = 0.592; Type 3 = 0.687; Type 4 = 0.594).

3.3. Stress gradient analysis

Two stress gradients, eutrophication (ammonium, nitrite and phosphate) and intensive agriculture (% intensive agriculture in the catchment), were included in the model selected as the best ($R^2 = 0.275$; $F_{2,396} = 75.13$ $p < 0.001$) and both evidenced a negative

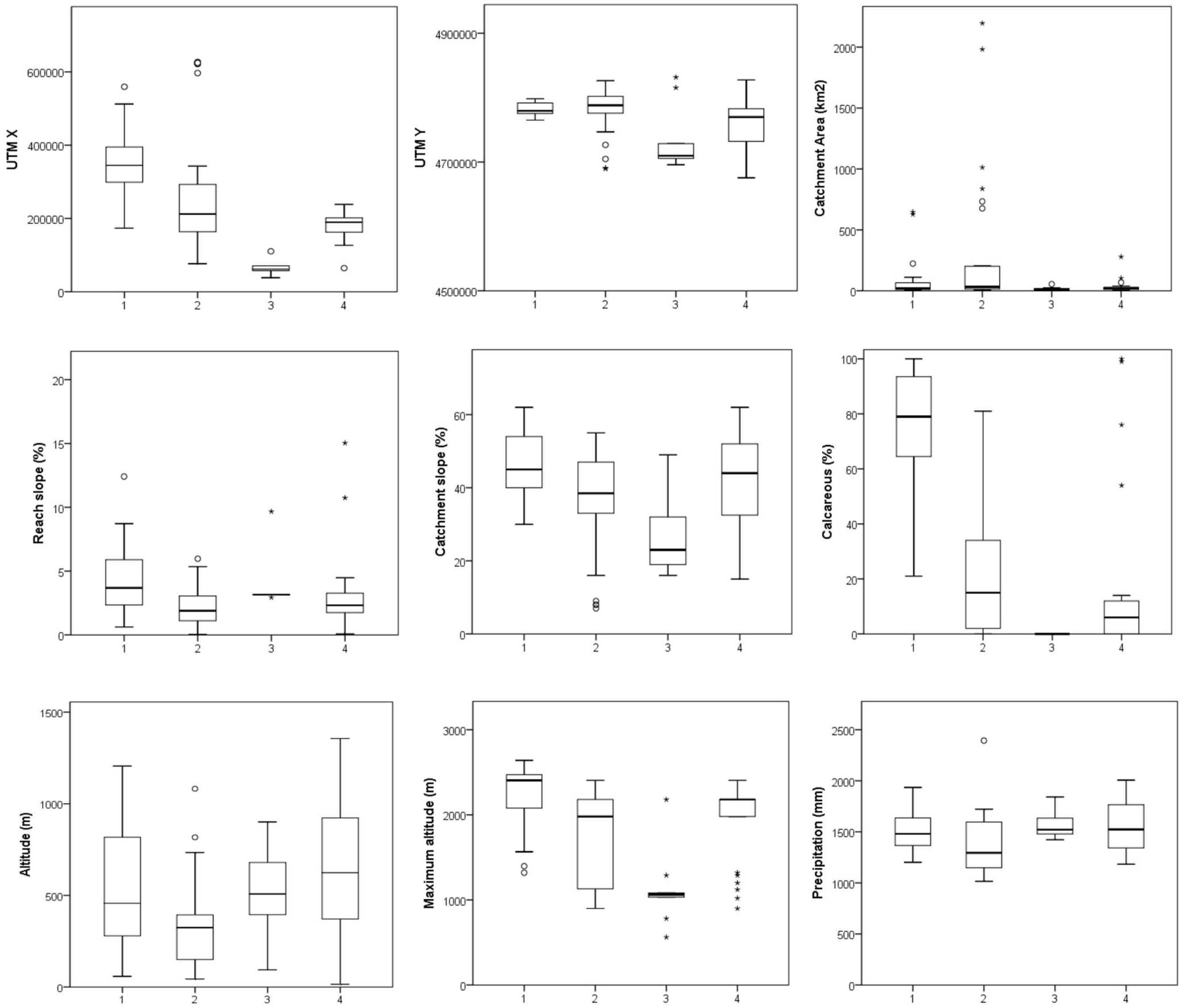

Fig. 4. Boxplots showing the values of the environmental variables characterizing the four river types in Northern Spain based on their diatoms assemblages. Data obtained from reference sites only.

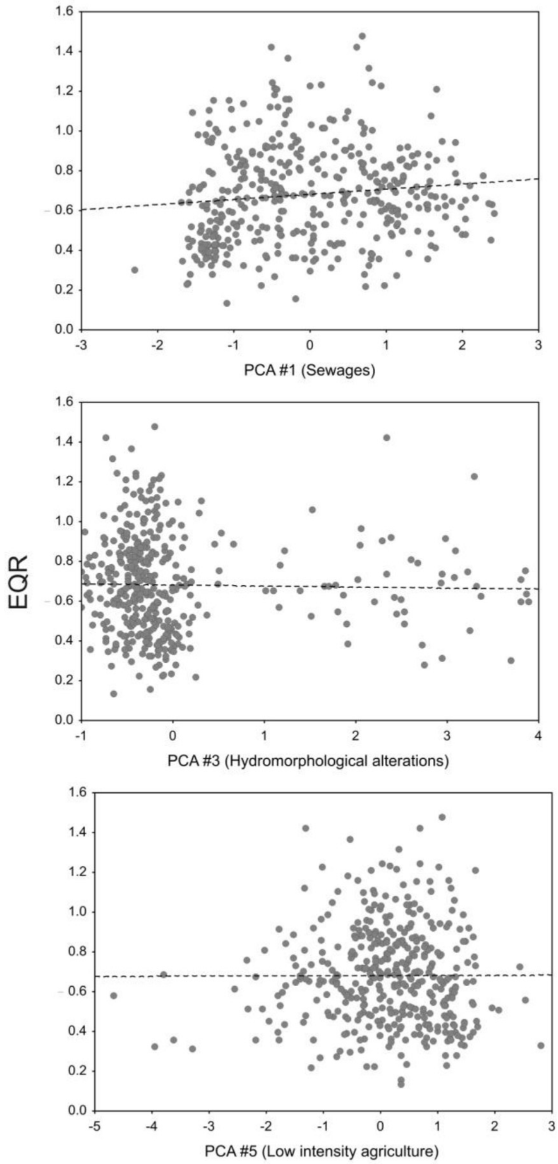

Fig. 5. Scatterplots showing the relationship between EQR_NORTIdiat values and the stress gradients identified in Pardo et al. (2014).

relationship with EQR values (Fig. 5). However, another seven models were also plausible and eutrophication and intensive agriculture were the only predictors included in all of them. In the model selected as the best, the standardized coefficient for eutrophication gradient (PCA #2) was -0.418 and the one for intensive agriculture (PCA #3) was -0.313, being both variables highly significant ($p < 0.001$).

4. Discussion

The NORTIdiat has proved to be a valid methodology in the assessment of the ecological status of streams and rivers in Northern Spain because it responded significantly to nutrient enrichment related pressures in the study area. The predictive diatom-based model (NORTIdiat) significantly characterized the four different diatom groups with obligatory and optional descriptors of the WFD typology that corresponded to a 4 river type's typology B that represented existing gradients in calcareous geology, catchment size, altitude, etc., characteristics of main river types existing in the studied area. Among the different existing approaches to predictive modelling using diatoms (Gevrey et al., 2004; Philibert et al., 2006, Carlisle et al., 2008, Feio et al., 2012; Almeida and Feio, 2012), the new approach uses the Bray-Curtis similarity between any test diatom community and the corresponding river type reference community, for the assessment of the ecological status. The NORTIdiat is a predictive model that fulfills all scientific aspects required to ecological status classification systems under the present WFD European legislation. It was developed using a spatial network of reference sites, minimally disturbed sites following Stoddard et al. (2006). Reference sites were identified and selected by the screening of a number of pressure criteria at various scales (catchment-riparian-stretch) developed by European Member States used in the European Rivers Intercalibration exercise (Pardo et al. 2012). The network of reference sites covered, from West to East, the North West and central North of Spain, being representative of most existing river geomorphologies and climates of the regions. For that the four biologically derived river types resulting in NORTIdiat represent the diatoms river typology validation inherent in the WFD, i.e. that the river types defined by geomorphological descriptors are inhabited by type specific biological reference communities.

4.1. Relationship between diatom assemblages and typology descriptors

The NORTIdiat model is based on the different diatom assemblages inhabiting reference sites, and the logistic regression established which environmental variables were prevailing at sites where those four diatom assemblages were found. Both the cluster analyses and the MDS ordination evidenced the distinct species composition of the groups. A closer look at the 4 species group's composition evidenced the correspondence between the diatom assemblages across the N-S and W-E direction, catchment's geology (from 100% siliceous to 100% calcareous) and size (from 2 to 2196 km) and site altitude (from 15 to 1356 m).

The most ubiquitous species was *Achnanthidium minutissimum* being the first or second most abundant species in all the groups, a species quite common (Rimet et al., 2004) or dominant (Luis et al., 2009) in France and Portugal, respectively. *A. minutissimum* is the unique common taxa to all the groups identified, and in a similar way it has been reported to characterize most river types in non-impaired sites in Portugal (Feio et al., 2012). *A. minutissimum* is a very sensitive species usually found in non-disturbed sites, that seems to be absent at streams with P_PO_4 values > 0.3 mg/l (Delgado and Pardo, 2014). In high elevation calcareous mountain rivers of the study area (Type 1) *A. minutissimum* is substituted by the dominant *A. pyrenaicum*, where it coexists with other taxa showing similar relative abundances, *Gomphonema pumilum, Cocconeis lineata, C. euglypta, Nitzschia fonticola* and *Navicula tripunctata,* all of them species typically found in calcareous rivers (Delgado et al., 2013, Delgado and Pardo, 2014).

In the type 2 corresponding to main river axes, 3 codominant species appear: *A. minutissimum*, *C. euglypta* and *Achnanthidium subatomus,* the typical composition of oligotrophic, calcium-poor waters (Lange-Bertalot, 2001; Potapova and Charles, 2003). In contrast, the diatom assemblage of the rivers type 3, of siliceous granitic geology, is very different because of peculiar diatom species that do not appear in the other 3 groups such as *Eunotia subarcuartoides*, *E. intermedia*, *Surirella roba*, *Navicula angusta* and *Peronia fibula*; the assemblage being dominated by *A. minutissimum*. The presence and abundance of these species were previously defined as part of the reference diatom community of Galician rivers (Northwestern Spain) and used to develop the MDIAT multimetric index (Delgado et al., 2010) for siliceous rivers. Siliceous streams and rivers under reference conditions in Northern Spain are highly oligotrophic and this species assemblage tends to disappear with increasing levels of human disturbance (Delgado et al., 2010). The presence of different species of the genus *Eunotia* in type 3 relates this river type with acid and oligotrophic waters rich in oxygen and poor in organic nitrogen compounds (van Dam et al., 1994; De Nicola, 2000; Sala et al., 2002; Delgado and Pardo, 2014).

Finally, type 4 of mixed rivers is characterized by two co-dominant taxa *Gomphonema rhombicum* and *Achnanthidium minutissimum* in high abundances, coexisting with other species in the group such as *Psammothidium oblonguellum*, *Cocconeis euglypta* and *G. exilissimum*. *Gomphonema rhombicum* is a good indicator of well-oxygenated waters of low conductivity and low nutrients and organic matter content (Merino et al., 1994; Almeida et al., 2010).

4.2. Sensitivity to human pressures

Diatom communities have been used in the last decades to infer water quality alteration in streams and rivers, particularly in Europe (Whitton et al., 1991; Prygiel and Coste, 1993; Prygiel et al., 1999), and are currently used for the same purpose worldwide (Stevenson, 2014; Delgado et al., 2012; Teittinen et al. 2015; Holmes and Taylor, 2015; Tan et al., 2017) evidencing the sensitivity of the bioindication they provide. Rapid changes in diatom composition and abundance occur because of their short generation times able to generate a new biocoenosis in a few weeks (Schaumburg et al., 2007). For that they are considered fast response bioindicators to changes in water quality (Stevenson and Pan, 1999). In this study NORTIdiat showed that significant changes in diatom communities, implying a deviation from the type reference diatom community, occurred in response to gradients in both eutrophication and intensive agriculture. Other studies in Europe (Hering et al., 2006; Springe et al., 2006) and other parts of the world (Potapova and Charles, 2007; Smucker and Vis, 2009; Bellinger et al., 2013) are in agreement with these obtained relationships. Meanwhile few studies have assessed with predictive models the response of diatoms to human pressures. A Portuguese diatom predictive model responded to both nutrient/organic contamination and to changes in the structure and morphology of the reach and the channel (i.e. artificial walls or embankments and connectivity) (Feio et al., 2009). Meanwhile, NORTIdiat did not respond to hydromorphological alteration, represented in this study by variables at the catchment scale (the total number of dams upstream; upstream dams' cumulative height per catchment area; channelization and number of water transfers and diversions per catchment area), in agreement with other studies that demonstrate the absence of diatom response to hydromorphological alterations (Hering et al., 2006; Delgado et al., 2010). A previous predictive model based on invertebrates developed for the same river basins in Northern Spain responded to the five identified human disturbance gradients in the area, that included hydromorphological pressures (sewages, eutrophication, hydromorphological alterations, intensive agriculture and low intensity agriculture) (Pardo et al., 2014). While using the same modelling approach, the diatom predictive model only responded to eutrophication and intensive agriculture cover in the basin, both gradients evidencing nutrient enrichment and the influence of derived products from agriculture. We attributed to the relatively well preserved riparian areas of streams and rivers in the studied area the existence and protection of the highly heterogeneous in stream physical conditions able to favor diatom habitats conservation that can minimize catchment scale hydromorphological impacts.

4.3. The NORTIdiat intercalibration

The use of diatom metrics in national assessment methods required their intercalibration at the European level following WFD requirements (Kelly et al., 2009; Kelly et al., 2012). The NORTIdiat was intercalibrated as a non offical method in the second European intercalibration phase (final phytobenthos technical report drafts (in https://circabc.europa.eu/w/browse/2844e1cc-7776-48b1-b410-b4dcc8d10a27) where results from the first phase were checked according to the new guidance, and included other non intercalibrated national methods. The more alkaline river types 1 and 2 (1: Mountain calcareous rivers and 2: Main river axes) fulfilled all intercalibration requirements. Meanwhile, river types 3 and 4 could not be intercalibrated in spite of all 4 river types resulting in this study being derived in the same statistical manner. A pragmatic solution was the intercalibration alone of types 1 and 2, avoiding the problems found for the most siliceous rivers, but our results raise the question on whether the pICM can be used for the intercalibration of all European river types, as suggested previously by Almeida et al. (2014) because the use of very different diatom databases. Nevertheless, the MDIAT multimetric specific for siliceous acid rivers (Delgado et al., 2010) showed significant correlations with the intercalibration common metric (Kelly et al., 2009; EC, 2013), probably in this case because the MDIAT is a combination of metrics, among them the IPS and TI.

5. Conclusions

Here, we propose a predictive model based on diatoms NORTIdiat to assess the ecological status of streams and rivers of Northern Spain. We used the ecological framework defined by the WFD: deriving a river typology from diatom assemblages from reference sites, and estimating the Ecological Quality Ratios (EQRs) from the observed similarity between the diatom 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 eutrophication and intensive agriculture pressures demonstrating to be a valid system to assess the ecological status of streams and rivers submitted to organic and nutrient stressors in Northern Spanish Rivers. The relatively high hydromorphological heterogeneity of these Atlantic rivers and their riparian areas seemed to benefit diatom communities in the presence of other pollution pressures and hydromorphological alterations. The un-official intercalibration of the NORTIdiat river types revealed that in spite of being derived from the same methodological approach, siliceous river types could not be compared and intercalibrated using the phytobenthos intercalibration common metric (pICM), composed by two commonly used metrics the IPS and TI. We attribute this failure to the fact that those metrics do not account for the auto-ecological particular characteristics of very different diatom assemblages inhabiting siliceous rivers, results that deserve further research efforts and question the suitability of diatom metrics to assess and compare the wide range of European river types.

Acknowledgements

This study was funded by the water Authorities of North Spain (Confederaciones Hidrográficas del Cantábrico y Miño-Sil), 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

Almeida, S.F.P., Craveiro, S.C., Calado, A.J., 2010. On the identity and distribution in Northern Portugal of three *Gomphonema* species currently misidentified as *Gomphonema clevei*. Diatom Res. 25, 13–27.

Almeida, S.F.P., Elias, C., Ferreira, J., Tornés, E., Puccinelli, C., Delmas, F., Dörflinger, G., Urbanič, G., Marcheggiani, S., Rosebery, J., Mancini, L., Sabater, S., 2014. Water quality assessment of rivers using diatom metrics across Mediterranean Europe: a methods intercalibration exercise. Sci. Total Environ. 476, 768–776.

Almeida, S.F.P., Feio, M.J., 2012. DIATMOD: diatom predictive model for quality assessment of Portuguese running waters. Hydrobiologia 695, 185–197.

Álvarez-Blanco, I., Blanco, S., Cejudo-Figueiras, C., Becares, E., 2013. The Duero Diatom Index (DDI) for river water quality assessment in NW Spain: design and validation. Environ. Monit. Assess. 185, 969–981.

Bellinger, B.J., Angradi, T.R., Bolgrien, D.W., Jicha1, T.M., Hill1, B.H., Reavie, E.D., 2013. Longitudinal variation and response to anthropogenic stress in diatom assemblages of the Lower Mississippi River. USA. River Syst. 21, 29–54.

Blanco, S., Ector, L., Huck, V., Monnier, O., Cauchie, H.M., Hoffmann, L., Bécares, E., 2008. Diatom assemblages and water quality assessment in the Duero Basin (NW Spain). Belg. J. Bot. 141, 39–50.

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

Carlisle, D.M., Hawkins, C.P., Meador, M.R., Potapova, M., Falcone, J.A., 2008. Biological assessments of appalachian streams based on predictive models for fish, macroinvertebrate, and diatom assemblages. J. North Am. Benthol. Soc. 27, 16–37.

Cemagref, 1982. Etude des Méthodes Biologiques Quantitatives d'appréciation de la qualité des eaux - Rapport Division Qualité des Eaux Lyon. Agence Financiére de Bassin Rhône - Méditerranée - Corse, Pierre-Bénite, pp. 218.

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

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.

De Nicola, D.M., 2000. A review of diatoms found in highly acidic environments. Hydrobiologia 433, 111–122.

Delgado, C., Pardo, I., García, L., 2010. A multimetric diatom index to assess the ecological status of coastal Galician rivers (NW Spain). Hydrobiologia 644, 371–384.

Delgado, C., Pardo, I., García, L., 2012. Diatom communities as indicators of ecological status in temporary Mediterranean streams (Balearic Islands, Spain). Ecol. Ind. 15, 131–139.

Delgado, C., Ector, L., Novais, M.H., Blanco, S., Hoffmann, L., Pardo, I., 2013. Epilithic diatoms of springs and spring-fed streams in Majorca Island (Spain) with the description of a new diatom species Cymbopleura margalefii sp. nov. Fottea 13, 87–104.

Delgado, C., Pardo, I., 2014. Comparison of benthic diatoms from Mediterranean and Atlantic Spanish streams: community changes in relation to environmental factors. Aquat. Bot. 120, 304–314.

Dell'Uomo, A., Pensieri, A., Corradetti, D., 1999. Diatomes Epilithiques du fleuve Esino (Italie centrale) et leur utilisation pour 1'evaluation de la qualite biologique de l'eau, Cryptogamie. Algal 20, 253–269.

EN 13946, 2003. Water Quality: Guidance Standard for the Routine Sampling and Pretreatment of Benthic Diatoms for Rivers.

EC, 2013. Commission decision of 20 September 2013 establishing, pursuant to directive 2000/60/EC of the European parliament and of the Council, the values of the member state monitoring system classifications as a result of the intercalibration exercise and repealing decision 2008/915/EC. Off. J. Eur. Union L 266/1-47.

Feio, M.J., Almeida, S.F.P., Craveiro, S.C., Calado, A.J., 2009. A comparison between biotic indices and predictive models in stream water quality assessment based on benthic diatom communities. Ecol. Ind. 9, 497–507.

Feio, M.J., Aguiar, F.C., Almeida, S.F.P., Ferreira, M.T., 2012. AQUAFLORA: a predictive model based on diatoms and macrophytes for streams water quality assessment. Ecol. Ind. 18, 596–598.

Gevrey, M., Rimet, F., Park, Y.S., Giraudel, J.L., Ector, L., Lek, S., 2004. Water quality assessment using diatom assemblages and advanced modelling techniques. Freshwater Biol. 49, 208–220.

Gomà, J., Ortiz, R., Cambra, J., Ector, L., 2004. Water evaluation in Catalonian Mediterranean rivers using epilithic diatoms as bioindicators. Vie Milieu 54, 81–90.

Hering, D., Johnson, R.K., Kramm, S., Schmutz, S., Szosz-kiewicz, K., Verdonschot, P.F.M., 2006. Assessment of European streams with diatoms, macrophytes, macroinvertebrates and fish: a comparative metric-based analysis of organism response to stress. Freshwater Biol. 51, 1757–1785.

Holmes, M., Taylor, J.C., 2015. Diatoms as water quality indicators in the upper reaches of the Great Fish River, Eastern Cape, South Africa. Afr. J. Aquat. Sci. 40, 321–337.

Johnson, R.K., Hering, D., Furse, M.T., Clarke, R.T., 2006. Detection of ecological change using multiple organism groups: metrics and uncertainty. Hydrobiologia 566, 115–137.

Jüttner, I., Rothfritz, H., Ormerod, S.J., 1996. Diatoms as indicators of river quality in the Nepalese Middle Hills with consideration of the effects of habitat-specific sampling. Freshwater Biol. 36, 475–486.

Kelly, M.G., Whitton, B.A., 1998. Biological monitoring of eutrophication in rivers. Hydrobiologia 384, 55–67.

Kelly, M.G., Cazaubon, A., Coring, E., Dell'Uomo, A., Ector, L., Godsmith, B., Guasch, H., Hürlimann, J., Jarlman, A., Kawecka, B., Kwandrans, J., Laugaste, R., Lindstrom, E.A., Leitao, M., Marvan, P., Padisák, J., Pipp, E., Prygiel, J., Rott, E., Sabater, S., Van Dam, H., Vizinet, J., 1998. Recommendations for the routine sampling of diatoms for water quality assessments in Europe. J. Appl. Phycol. 10, 215–224.

Kelly, M., Bennett, C., Coste, M., Delgado, C., Delmas, F., Denys, L., Ector, L., Fauville, C., Ferreol, M., Golub, M., Jarlman, A., Kahlert, A., Lucey, J., Ni Chathain, B., Pardo, I., Pfister, P., Picinska-Faltynowicz, J., Schranz, C., Schaumburg, J., Tison, J., van Dam, H., Vilbaste, S., 2009. A comparison of national approaches to setting ecological status boundaries in phytobenthos assessment for the European Water Framework Directive: results of an intercalibration exercise. Hydrobiologia 621, 169–182.

Kelly, M.G., Gómez-Rodríguez, C., Kahlert, M., Almeida, S.F.P., Bennett, C., Bottin, M., Delmas, F., Descy, J.-P., Dörflinger, G., Kennedy, B., Marvan, P., Opatrilova, L., Pardo, I., Pfister, P., Rosebery, J., Schneider, S., Vilbaste, S., 2012. Establishing expectations for Pan-European diatom based ecological status assessments. Ecol. Ind. 20, 177–186.

Krammer, K., Lange-Bertalot, H., 1986-1991. Bacillariophyceae, 1-5. VEB Gustav Fisher Verlag, Jena, Germany.

Kwandrans, J., Eloranta, P., Kawecka, B., Wojtan, K., 1998. Use of benthic diatom communities to evaluate water quality in rivers of southern Poland. J. Appl. Phycol. 10, 193–201.

Lange-Bertalot, H., 1993. 85 Neue Taxa und über 100 weitere neu definierte Taxa ergänzend zur Süsswasserflora von Mitteleuropa vol. 2/1–4. Bibliotheca Diatomologica 27, 1–164.

Lange-Bertalot, H., 2001. Navicula sensu stricto: 10 genera separated from Navicula sensu lato, Frustulia. In: Lange-Bertalot, H., org. Diatoms of the Europe. Ruggell: A.R.G. Gantner Verlag Kommanditgesellschaft. 526.

Luis, A.T., Teixeira, P., Almeida, S.F.P., Ector, L., Matos, J.X., Ferreira da Silva, E.A., 2009. Impact of acid mine drainage (AMD) on water quality, stream sediments and periphytic diatom communities in the surrounding streams of Aljustrel mining area (Portugal). Water Air Soil Pollut. 200, 147–167.

Martín, G., Toja, J., Estela, S., Fernández, M.R., Reyes, I., Casco, M.A., 2010. Application of diatom biotic indices in the Guadalquivir River Basin, a Mediterranean basin. Which one is the most appropriated? Environ. Monit. Assess. 170, 519–534.

Mazor, R.D., Reynoldson, T.B., Rosenberg, D.M., Resh, V.H., 2006. Effects of biotic assemblage, classification, and assessment method on bioassessment performance. Can. J. Fish. Aquat. Sci. 63, 394–411.

Merino, V., García, J., Hernàndez-Mariné, M., 1994. Use of diatoms for pollution monitoring in the Valira Basin (Andorra). In: Proceedings of the 13th International Diatom Symposium. Koeltz Scientific Publishers, Koenigstein, 107–119.

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., Bennett, C., Birk, S., Buffagni, A., Erba, S., Mengin, N., Murray-Bligh, J., Ofenböeck, G., 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., Gómez-Rodríguez, C., Abraín, R., García-Roselló, E., Reynoldson, T.B., 2014. An invertebrate predictive model (NORTI) for streams and rivers: sensitivity of the model in detecting stress gradients. Ecol. Ind. 45, 51–62.

Philibert, A., Gell, P., Newall, P.R., Chessman, B., Bate, N., 2006. Development of diatom-based tools for assessing stream water quality in south-eastern Australia: assessment of environmental transfer functions. Hydrobiologia 572, 103–114.

Poikane, S., Kelly, M., Cantonati, M., 2016. Benthic algal assessment of ecological status in European lakes and rivers: challenges and opportunities. Sci. Total Environ. 568, 603–613.

Potapova, M., Charles, D.F., 2003. Distribution of benthic diatoms in US rivers in relation to conductivity and ionic composition. Freshwater Biol. 48, 1311–1328.

Potapova, M., Charles, D.F., 2007. Diatom metrics for monitoring eutrophication in rivers of the United States. Ecol. Ind. 7, 48–70.

Prygiel, J., Coste, M., 1993. The assessment of water quality in the Artois-Picardie water basin (France) by the use of diatom indices. Hydrobiologia 269 (270), 343–349.

Prygiel, J., Whitton, B.A., Bukowska, J., 1999. Use of Algae for Monitoring Rivers III. Agence de l'Eau Artois-Picardie, Douai.

Quinn, G.P., Keough, M.J., 2002. Experimental Design and Data Analysis for Biologists. Cambridge University Press, UK.

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.

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 toassess water-quality impairment using benthic macroinvertebrates. J. North Am. Benthol. Soc. 16, 833–852.

Rimet, F., Ector, L., Cauchie, H.M., Hoffmann, L., 2004. Regional distribution of diatom assemblages in the headwater streams of Luxembourg. Hydrobiologia 520, 105–117.

Sala, S.E., Duque, S.R., Nunez-Avellaneda, M., Lamaro, A., 2002. Diatoms from the Colombian Amazon. Cryptogamie Algol. 23, 75–99.

Schaumburg, J., Schranz, C., Stelzer, D., Hofmann, G., 2007. Handlungsanweisung für die ökologisc ewertung von Seen zur Umsetzung der EU-Wasserrahmenrichtlinie: Makrophyten und Phytobenthos: 75 pp. - Bayerisches Landesamt für Wasserwirtschaft im Auftrag der Länderarbeitsgemeinschaft Wasser, February 007. - München.

Simpson, J.C., Norris, R.H., 1997. Assessing the biological quality of fresh waters: RIVPACS and other techniques. In: Wright, J.F., Sutcliffe, D.W., Furse, M.T. (Eds.), Proceedings of an International Workshop held in Oxford, UK, 125–142.

Smucker, N.J., Vis, M.L., 2009. Use of diatoms to assess agricultural and coal mining impacts on streams and a multiassemblage case study. J. North Am. Benthol. Soc. 28, 659–675.

Springe, G., Sandin, L., Briede, A., Skuja, A., 2006. Biological quality metrics: their variability and appropriate scale for assessing streams. Hydrobiologia 566, 153–172.

Stevenson, R.J., 2014. Ecological assessments with algae: a review and synthesis. J. Phycol. 50, 437–461.

Stevenson, R.J., Pan, Y., 1999. Assessing Environmental Conditions in Rivers and Streams with Diatoms. In: Stoermer, E.F., Smol, J.P. (Eds.), The Diatoms: Application for the Environmental and Earth Sciences. Cambridge University Press, Cambridge, pp. 11–40.

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.

Tan, X., Zhang, Q., Burford, M.A., Sheldon, F., Bunn, S.E., 2017. Benthic diatom based indices for water quality assessment in two subtropical streams. Front. Microbiol. 8, 601. http://dx.doi.org/10.3389/fmicb.2017.00601.

Teittinen, A., Taka, M., Ruth, O., Soininen, J., 2015. Variation in stream diatom communities in relation to water quality and catchment variables in a boreal, urbanized region. Sci. Total Environ. 531, 279–289.

Van Dam, H., Mertens, A., Sinkeldam, J., 1994. A coded checklist and ecological indicator values of freshwater diatoms from The Netherlands. Neth. J. Aquat. Ecol. 28, 117–133.

van de Bund, WJ (ed), 2009. Water Framework Directive intercalibration technical report. Part 1: Rivers. JRC Scientific and Technical Reports, Ispra, Italia.

Whitton, B.A., Rott, E., Friedrich, G., 1991. Use of Algae for Monitoring Rivers. Institut fürn Botanik, Universität Innsbruck, Innsbruck.

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