# A comparison between biotic indices and predictive models in stream water quality assessment based on benthic diatom communities

Maria J. Feio [a,*], Salomé F.P. Almeida [b], Sandra C. Craveiro [b], António J. Calado [b]

[a] Institute of Marine Research (IMAR), Department of Zoology, University of Coimbra, Largo Marquês de Pombal, P-3004-517 Coimbra, Portugal
[b] GeoBioTec – GeoBioSciences, GeoTechnologies and GeoEngineering (Research Unit), Department of Biology, University of Aveiro, P-3810-193 Aveiro, Portugal

## ARTICLE INFO

## ABSTRACT

Diatoms are widely used in stream bioassessment due to their broad distribution, extraordinary variability and the ability to integrate changes in water quality. The indices Specific Polluosensitivity Index (SPI), standardized Biological Diatom Index (BDI), European Economic Community Index (CEC) and Generic Diatom Index (GDI), originally developed in France, are often applied in Portugal to evaluate stream ecological quality based on diatom communities. Alternatively, predictive models resulting from the comparison between the communities of the study site and those of a set of reference sites representing undisturbed or the best available conditions of a given region have been proposed as valuable methods for evaluating the ecological status of streams. In the present study, we applied the four above-mentioned widely used diatom-based indices (SPI, BDI, CEC and GDI) and a predictive model (MoDi) to 54 sites located in central Portugal to assess the sensitivity of the five methods to a range of anthropogenic disturbances cumulatively affecting streams and represented by 27 variables (e.g., organic enrichment, changes in morphology of the channels, integrity of the riparian corridor, land use in the catchment). The results were analyzed comparatively through Spearman correlations, Boxplots and Stepwise Discriminant Analysis. This study confirmed the sensitivity of diatoms to organic and nutrient contamination (showed by the MoDi, BDI, CEC and SPI) and revealed the importance of suspended solids (through the MoDi, GDI, and SPI). The relevance of modifications in land use to diatoms was shown by all methods applied, except for the GDI. The MoDI also revealed the importance of changes in the structure and morphology of the reach and the channel, like the construction of artificial walls or embankments and connectivity; the BDI also related its assessments with the riparian zone integrity; and the SPI was not useful in detecting morphological pressures. The GDI produced the most divergent assessments and was less effective in revealing the anthropogenic disturbances. The use of the predictive model (MoDi) is therefore a good method for the assessment of streams in central Portugal because it expresses a great diversity of quantitative and qualitative changes in freshwater systems reflected in the structure (species richness and abundance) of diatom communities.

© 2008 Elsevier Ltd. All rights reserved.

* Corresponding author.
E-mail address: mjf@ci.uc.pt (M.J. Feio).

# 1. Introduction

Diatoms are the unicellular algal group most widely used as indicators of stream quality. They have several advantages as bioindicators: as a group they are ubiquitous and their variability spans most ecological conditions of the aquatic environment; benthic communities integrate variations of water quality in a particular location; sampling and preparation methods are relatively straightforward and preparations can be preserved indefinitely; and with some training, it is possible to identify the species through a number of taxonomic guides (Krammer and Lange-Bertalot, 1986, 1988, 1991a,b). As a result, monitoring programs all over the world have included diatoms and they are among the biological quality elements indicated in the European Water Framework directive (e.g., Directive 2000/60/EC, 2000; Ector and Rimet, 2005; Philibert et al., 2006).

For the quantification of river health based on diatom communities several indices have been developed regionally in many countries and widely used elsewhere, with or without adaptation. Currently used indices include the Specific Polluosensitivity Index (SPI – Coste, 1986), the standardized Biological Diatom Index (BDI – Lenoir and Coste, 1996), the Saprobity Index (SI – Sládeček, 1986), the European Economic Community Index (CEC – Descy and Coste, 1991), the Diatom Assemblage Index for organic pollution (DAI – Watanabe et al., 1986), the Trophic Diatom Index (TDI – Kelly and Whitton, 1995) and the Generic Diatom Index (GDI – Rumeau and Coste, 1988).

A more recent approach to water quality assessment are predictive models which measure the quality of a site, as the degree of alteration of its communities in relation to reference communities of sites of comparable environmental conditions (Reynoldson et al., 1997; Norris and Hawkins, 2000). This methodology, initially applied to macroinvertebrate communities, proved to be a powerful approach (Wright, 1995; Simpson and Norris, 2000; Reynoldson, 1995). Analogous models based on diatoms have lately been developed in several regions of the world (John, 2000; Mazor et al., 2006; Philibert et al., 2006; Feio et al., 2007).

With the ecological reference condition concept embedded in the current European legislation, water quality evaluation through the use of predictive models that measure the distance from a reference state seems quite appropriate. However, it is important to compare the performance of the model approach with the more widespread use of diatom-based indices. In particular, it is necessary to understand to what extent the environmental properties measured by the different methods are the same. In order to achieve this, this paper compares the assessments of the predictive model for diatoms (MoDi) to that of the common indices (SPI, BDI, CEC, and GDI) regarding their sensitivity to a range of anthropogenic disturbances often cumulatively found in streams (e.g., organic enrichment, changes in morphology of the channels, integrity of the riparian corridor, land use in the catchment).

# 2. Materials and methods

## 2.1. Study area

The study region, located in the centre of Portugal, includes three adjacent catchments of rivers Mondego, Vouga and Lis (Fig. 1), covering a total area of ≈11,200 km², and coincides
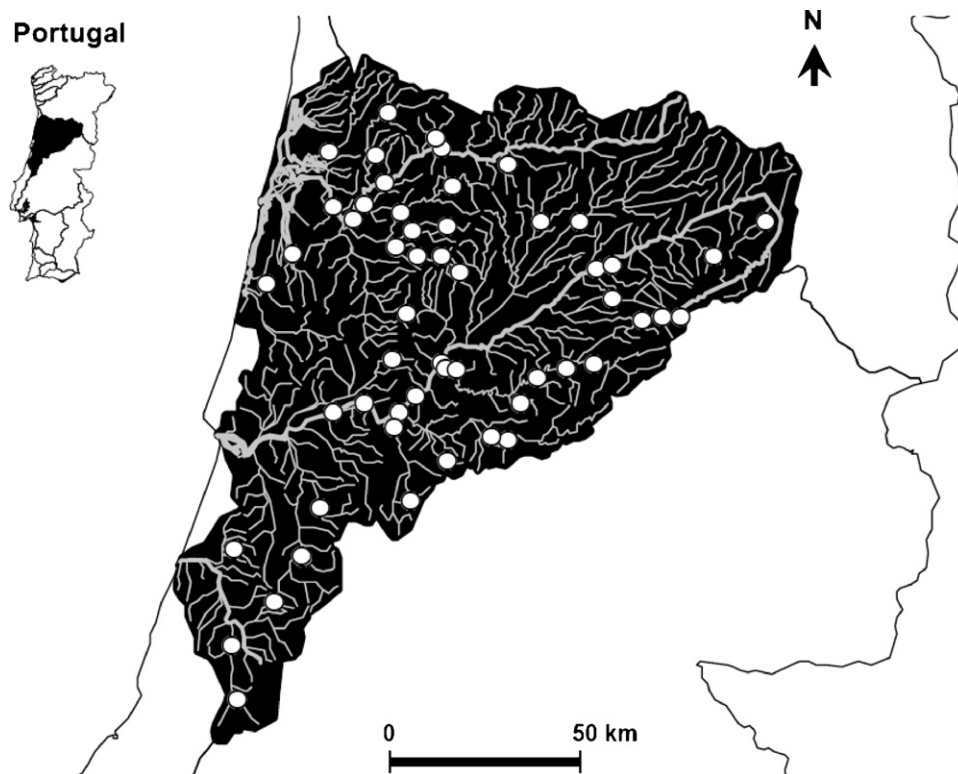


Fig. 1 – Target area (Mondego, Vouga and Lis catchments) and study sites (white circles).

with the target area of the predictive model (Feio et al., 2007). This region has a temperate Atlantic climate and contains a wide diversity of landscapes, from mountain areas in the interior (up to 2000 m) to the coastal lowlands, and from granite and schist to limestone grounds. The main anthropogenic pressures are felt in the littoral, with higher population densities, industries (e.g., paper pulp) and agriculture (e.g., rice). In the interior the main impacts are the presence of big dams for power production and small weirs used to create water reservoirs to prevent the dislocation of introduced fishes but also some pastures, forestry and industry (e.g., cheese production) (Decretos regulamentares 9/2002, 15/2002, 23/2002).

## 2.2. Sampling

Fifty-four sites located in the target region, spread through different stream characteristics (size, geographic location, geology and altitude, and potential impacts) were used in this study (Fig. 1). During Spring, each site was visited for diatom sampling and to obtain data required by the predictive model to determine group membership: latitude, runoff (mm), water mineralization (low, medium and high, depending on the catchment's dominant geology), distance to source (km), slope (5), discharge ($m^3 s^{-1}$) and hardness ($mg l^{-1} CaCO_3$). Also, the potential anthropogenic disturbances affecting streams were measured through 27 variables grouped into the categories of land use, nutrients and organic contamination, water acidification and toxicity, suspended solids and morphology of the channel and banks (Table 1).

Water samples were collected for the laboratory measurement of nutrients, hardness and oxidability (methods indicated in Table 1). Other variables were obtained from cartographic documentation such as distance to source, slope and land use (1:25,000 digital military maps of Instituto Geográfico do Exército, Portugal; Atlas do Ambiente: Agência Portuguesa do Ambiente, 2007; Corine Land Cover, 2000).

The sampling, treatment and study of diatom communities was based on European standards (EN 13946 2003, EN 14407 2004, CEN TC230 N68 2003) and on Prygiel and Coste (2000). In a

---

**Table 1 – Description of the pressure variables considered by this study grouped by nature of the pressure, sources and methods and the transformations applied for discriminant analysis**

Land use
  Intensive agriculture (in the drainage area, %; Corine Land Cover, 2000)
  Extensive agriculture (in the drainage area, %; Corine Land Cover, 2000)
  Natural areas (in the drainage area, %; Corine Land Cover, 2000)
  Intensive agriculture 5 (in 5-km ratio around the site; %; Corine Land Cover, 2000)
  Extensive agriculture 5 (in 5-km ratio around the site; %; Corine Land Cover, 2000)
  Land use (degradation by agriculture and others; categories: 1, high quality to 5, bad quality; adapted from Pont et al., 2006; field and observation of data from Corine Land Cover, 2000)
  Urban area (impact of urbanization on the stream integrity; categories: 1, high quality to 5, bad quality; adapted from Pont et al., 2006; field and observation of data from Corine Land Cover, 2000)

Nutrients and organic contamination
  Nitrates ($NO_3^{2-} mg l^{-1}$; ion chromatographic method, A.P.H.A.); $log(x)$
  Nitrites ($NO_2^- mg l^{-1}$; ion chromatographic method, A.P.H.A.); $log(x)$
  Ammonium ($mg l^{-1}$; ion chromatographic method, A.P.H.A.); $log(x)$
  Phosphates ($mg l^{-1}$, ascorbic acid method, A.P.H.A.); $log(x)$
  Total N ($mg l^{-1}$; calculation, A.P.H.A.); $log(x)$
  Total P ($mg l^{-1}$; ascorbic acid method, A.P.H.A.); $log(x)$
  COD, chemical oxygen demand ($mg l^{-1}$; closed reflux, colorimetric method); $log(x + 1)$
  $BOD_5$, biological oxygen demand ($mg l^{-1}$; 5-day BOD test, A.P.H.A.); $log(x + 1)$
  Oxidability ($mg l^{-1}$; permanganate index, ISO 8467); $log(x)$
  $O_2$ ($mg l^{-1}$; Field measurement WTW OXI 92); $log(x)$
  Organic contamination and nutrient enrichment (deviation from reference values of nutrients and $BOD_5$ and COD parameters of the water; categories: 1, high quality to 5, bad quality; adapted from Pont et al., 2006; field and data observation)

Water acidification and toxicity
  pH (field measurement JENWAY 3310)
  Acidification and toxicity (loss of naturalness in the acidification and oxygenation level of the water; categories: 1, high quality to 5, bad quality; adapted from Pont et al., 2006; field and data observation)

Suspended solids
  TSS, total suspended solids ($mg l^{-1}$; A.P.H.A); $log(x)$
  Sediments discharge (loss of naturalness in the concentration of sediments transported by the stream water; categories: 1, high quality to 5, bad quality; adapted from Pont et al., 2006; field and data observation)

Morphology of the channel and banks
  Riparian zone (integrity of the riparian corridor; categories: 1, high quality to 5, bad quality; adapted from Pont et al., 2006; field observation)
  Morphological condition (loss of naturalness of the river channel and banks; categories: 1, high quality to 5, bad quality; adapted from Pont et al., 2006; field observation)
  Connectivity (categories: 1, high quality to 5, bad quality; adapted from Pont et al., 2006; field and cartographic information)
  HMS (Habitat Modification Score, calculated after field observations according to the River Habitat Survey, Environmental Agency, 2003)
  HQA (Habitat Quality Assessment, calculated after field observations according to the River Habitat Survey, Environmental Agency, 2003)

pre-defined reach of 50 m, diatoms were scraped from a total area of about 100 cm$^2$ from stones removed from a depth of 10–30 cm, preferably in unshaded spots and with a current velocity of 10–50 cm s$^{-1}$. The samples, preserved with Lugol's solution, were later oxidized with concentrated HNO$_3$ for 24 h. The diatoms were mounted in Naphrax and the identification of 400 valves was done to the lowest practicable taxonomic level, normally species or infra-specific rank, mainly with Krammer and Lange-Bertalot's flora (1986, 1988, 1991a,b). The counts were then converted into percentages of valves for each taxa.

## 2.3. Data analysis

The diatom indices BDI (Lenoir and Coste, 1996), SPI (Coste, 1986), CEC (Descy and Coste, 1991) and the GDI (Rumeau and Coste, 1988) were calculated with the software OMNIDIA developed by CLCI (Lecointe et al., 1993, 1999). The BDI, SPI and GDI use weighted average values of sensitivity, indicator value and relative abundance of each taxon. The SPI is based on Zelinka and Marvan's (1961) formula, derived from the Saprobic System, and comprises about 13,000 taxa which usually include all taxa observed (Prygiel et al., 1996). The BDI is feasible with lower taxonomic skills, and is based on the ecological profiles of 209 taxa whose presence probability for seven water quality classes is defined from 14 physical and chemical parameters, such as nutrients, organic content, pH and conductivity. The CEC is based on a double entry table where the taxa, according to their sensitivity to stress, are classified in different groups (low indicator value) and sub-groups (high indicator value). The crossing of the groups and sub-groups results in a value that indicates the quality class. The GDI uses only genera and admits therefore lower taxonomic skills (Prygiel et al., 1996).

All indices result in a value that, after linear adjustment, varies between 0 or 1 (bad quality) and 20 (high quality), corresponding to five water quality classes. Since the predictive model results in only four bands of quality, and to make the results comparable, the five classes of the indices were transformed into four classes: Classes 5 and 4 (very bad and bad) were considered equivalent to model Band 4, and the remaining Classes 1–3 were considered equivalent to the model Bands 1–3. For simplicity, both the four classes and the four bands of the model will herein be called "classes".

The predictive model (MoDi) applied in this study is based on the Benthic Assessment of Sediment (BEAST – Reynoldson, 1995; Reynoldson et al., 1997, 2001) and described in detail in Feio et al. (2007). The model allocates a site into one of four quality classes, with increasing distance to the group of reference sites in an ordination space, based on the biological communities. If a site is allocated to Class 1 it is considered equivalent to reference condition (high/good quality in our study); sites in Class 2 are potentially different (moderate quality), sites in Class 3 are different (poor quality) and sites in Class 4 are very different from the reference condition (bad quality). To use the model, biological data have to be transformed by square root as this is the transformation applied to reference data.

To determine the similarity of responses between four indices and the MoDi, Spearman correlations (SYSTAT 8.0) were obtained using the quality classes calculated for the 54 sites.

To understand the differences between the diatom indices and the model two methods were used: (1) the stepwise discriminant analysis (DA; forward, $p$ to enter and to remove = 0.15, tolerance = 0.001, SYSTAT 8.0) selected from a large group of disturbance variables (Table 1) those that best discriminate the test sites into their quality class, previously attributed by the indices or the model. Jackknife cross-validation was used to validate the results of the discriminant analysis by repeatedly eliminate a sample from the original data set and compute the discriminant function with the remaining observations (Daniel, 1989). For DA, some variables were transformed towards normality (Table 1), after a verification of normality through the Kolmogorov–Smirnoff test; (2) boxplots (SYSTAT 8.0) were applied to the variables selected in 1) to obtain information about the ranges of each disturbance variable within the quality classes attributed by the indices and model. In the graphs, the centre line marks the median of the sample. The length of each box shows the range within which the central 50% of the values fall, with the box edges (called hinges) at the first and third quartiles.

In order to categorize the patterns observed in the boxplots, four types of graphs were defined: Type 1 graphs show an important difference in the median value between Classes 3 and 4 while from Class 1 to Class 3 the values are similar; in Type 2 graphs there are evident changes between Classes 1 and 2, Classes 2 and 3 are similar and then there is another increase in the pressure values between Classes 3 and 4; Type 3 are those graphs showing a continuous pattern of increasing pressure over classes. In graphs of Type 4 there is a general tendency for the increase of median values over the classes
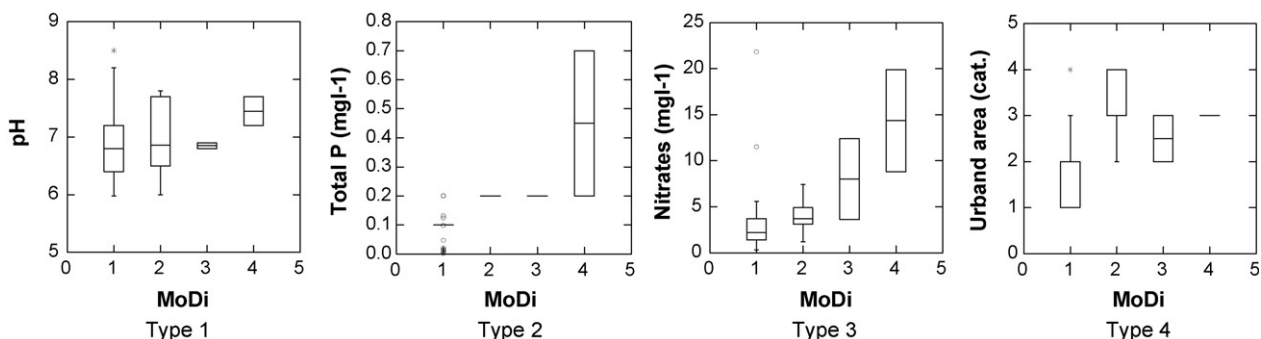


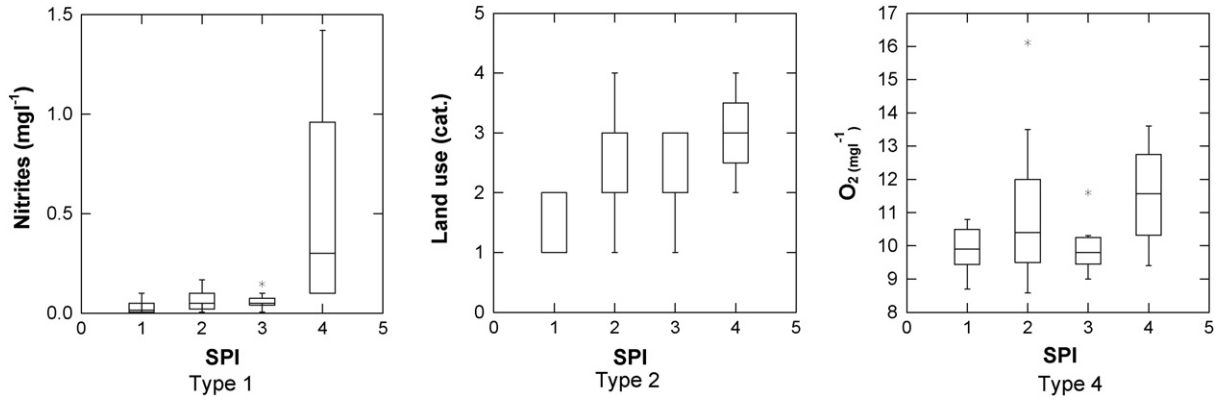Fig. 2 – Examples of box plots of each type for the predictive model.

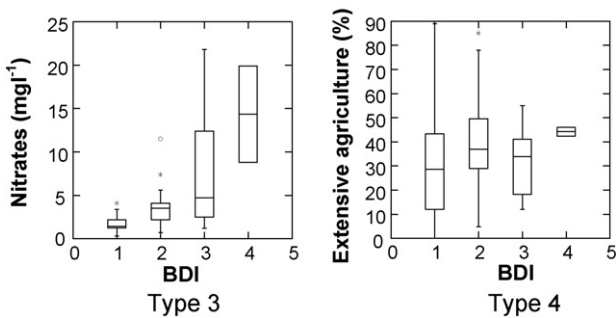**Fig. 3 – Examples of box plots of each type for the SPI.**



**Fig. 4 – Examples of box plots of each type for the BDI.**

but with some irregularities. Examples of these types of graphs are given in Figs. 2–6.

## 3. Results

The analysis of Spearman correlations between indices and model classes attributed to the test sites revealed that the most similar results are between the SPI and the CEC indices (0.858, $p < 0.001$) followed by those between the BDI and the SPI (0.699, $p < 0.001$) and BDI and CEC (0.664, $p < 0.001$). The predictive model showed to be significantly correlated with

three indices (BDI, CEC and SPI) with higher correlations obtained for the BDI (0.454, $p < 0.001$) (Table 2). The GDI was not correlated with the MoDi ($r = 0.150$, $p < 0.05$).

The stepwise discriminant analysis selected 16 variables that correctly discriminated the sites assessed by the MoDi into their previously attributed quality classes (classification error = 2%, Jackknifed error = 26%; Table 3). From those, nine variables belong to the categories of nutrients or organic contamination (e.g., nitrates, phosphates, $CBO_5$, conductivity), two variables from land use (urban area and extensive agriculture in the catchment), three variables belong to morphology of the channel and banks (connectivity and extensive agriculture, HMS) and pH was selected from the category of water acidification and toxicity.

For the SPI, the DA selected six variables in the categories of nutrients and organic contamination (nitrites and phosphates and conductivity), land use, suspended solids and acidification and toxicity (classification error = 35%, Jackknifed error = 39%; Table 3).

The BDI assessments were discriminated by 11 pressure variables in the categories of nutrients and organic contamination, morphology of the channel and banks (variable morphological condition and HQA) and land use (classification error = 22%, Jackknifed error = 37%; Table 3).

The DA selected eight variables for the CEC as discriminators of the quality classes. Among those are six variables related with nutrients and organic contamination (water
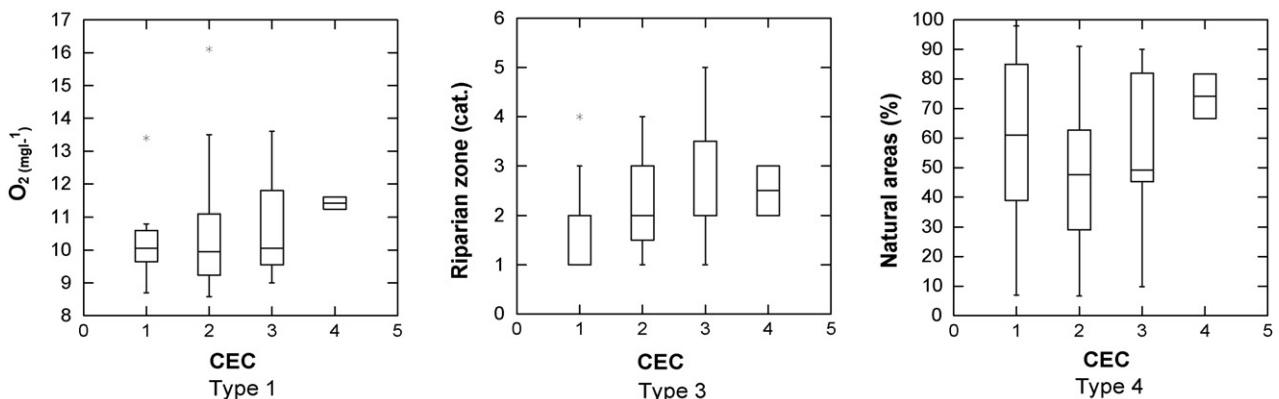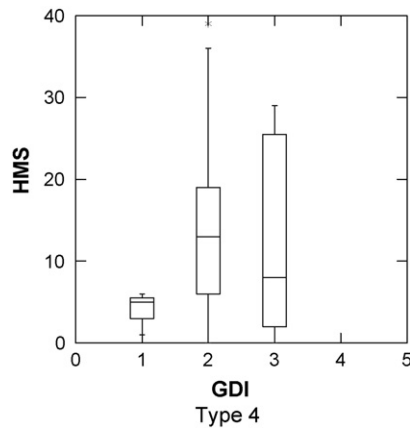


**Fig. 5 – Examples of box plots of each type for the CEC.**

**Fig. 6 – Examples of box plots of each type for the GDI.**

oxygenation, nitrates, total N, COD, oxidability), one in the category of suspended solids (sediments discharge), another from water acidification and toxicity and four variables in the category of land use (intensive and extensive agriculture, natural areas and land use) (classification error = 26%, Jack-knifed error = 35%; Table 3).

Finally, the DA selected for the GDI five variables: $O_2$, pH, conductivity, TSS and HMS covering four different categories: nutrients and organic contamination, water acidification and toxicity, suspended solids and morphology of the channel and banks (classification error = 31%, Jackknifed error = 37%; Table 3).

The box plots analysis showed that the global classification attributed by the model and indices is not only consequence of a continuous increase of the pressures selected by the discriminant analysis, which may not vary in the same way between classes, but have mostly a cumulative effect over diatom communities. Table 4 resumes the patterns found for all indices and model concerning their responses to different pressures.

For the MoDi it is possible to integrate the most evident patterns in four types of graphs/responses. Conductivity and pH showed an important difference in the median value between Classes 3 and 4 and similar values between Class 1 and Class 3, which seems to indicate that only strong changes

in these variables affect diatom community structure (Type 1 graphs, example in Fig. 2). A different situation occurs for phosphates, nitrites, organic contamination and total P where there are evident changes between Classes 1 and 2, Classes 2 and 3 are similar and then there is another increase in the pressure values between Classes 3 and 4 (Type 2 graphs, example in Fig. 2). This may indicate that a slight change in nutrient concentration in clean waters is reflected in the communities and that the effect is only felt again for much higher nutrient concentrations. Nitrates and TSS showed a continuous pattern of increasing pressure over the model classes (Type 3 graphs, example in Fig. 2). Finally the variables urban area, extensive agriculture in the catchment, morphological condition, connectivity, and HMS showed a tendency for the increase of median values (and therefore decrease in quality) over model classes, although with some irregularities (graphs Type 4, example in Fig. 2). For ammonium, COD and $CBO_5$ the patterns in the boxplots were not clear.

For the SPI only three types of graphs appeared with clear patterns: sediments discharge and nitrites showed a Type 1 graph, land use Type 2, and water oxygenation Type 4 (Fig. 3). Any pressure was reflected in the index classes in a continuous way. Phosphates and acidification and toxicity had no clear patterns reflected in the graphs.

With the BDI the nutrients (phosphates, nitrites, nitrates, total N, and total P) increased continuously from Class 1 to Class 4 (graph Type 3, Fig. 4) whereas morphological condition, % of extensive agriculture in the catchment and organic contamination resulted in a more or less clear increase in the disturbances level with the BDI classes (graph Type 4, Fig. 4). Ammonium, conductivity and HQA did not result in clear patterns.

The observation of the boxplots for the CEC revealed a Type 1 pattern for water oxygenation ($O_2$ mg l$^{-1}$), a Type 3 graph for the categorical variables urban area and degradation of riparian zone, with a continuous increase over the CEC quality classes, and a Type 4 graph for oxidability, nitrates, total N and % intensive agriculture in the catchment. For extensive agriculture, natural areas, acidification and toxicity and sediments discharge the patterns were not clear.

Since the GDI index only classified the sites into three classes, the patterns are less clear and only for TSS and HMS there is a more obvious tendency for the increase in disturbance level with the increase of class, but it is still a Type 4 graph.

## 4. Discussion

Discriminant analysis was used in this study to determine the pressure variables that best explained the results of four diatom indices and a predictive model. Although we used for DA variables scoring from 1 to 5 (such as the "riparian zone" or "morphological condition") we used them as interval variables since equal intervals were assumed between classes. Moreover, Pohar et al. (2004) demonstrated that five classes is a sufficiently high number for letting the estimated mean and variance to be close to the population values of the continuous explanatory variables. Therefore we consider that these variables did not have a negative effect over the discriminant analysis performance.

| Table 2 – Spearman correlation coefficients ($r$) between indices and predictive model (MoDi) and respective significance levels ($p$) | | | |
|---|---|---|---|
| | BDI | CEC | GDI | MoDi |
| SPI | $r = 0.699$ $p < 0.001$ | $r = 0.858$ $p < 0.001$ | $r = 0.475$ $p < 0.001$ | $r = 0.295$ $p < 0.05$ |
| BDI | | $r = 0.664$ $p < 0.001$ | $r = 0.380$ $p < 0.01$ | $r = 0.454$ $p < 0.001$ |
| CEC | | | $r = 0.462$ $p < 0.001$ | $r = 0.314$ $p < 0.05$ |
| GDI | | | | $r = 0.150$ $p > 0.05$ |
| The number of samples used in all correlations was 54. | | | |

| Table 3 – Stepwise results for the five assessment methods applied with selected variables (F to remove and tolerance) and % of correct classifications | |
|---|---|
| Variables (F-to-remove, tolerance) | % correct classifications, % after Jackknifed cross-validation |
| SPI | |
| Sediments discharge (21.90, 0.69) | 65%, 61% |
| Land use (6.14, 0.70) | |
| Nitrites (6.09, 0.10) | |
| $O_2$ (3.96, 0.69) | |
| Phosphates (3.78, 0.11) | |
| Acidification and toxicity (2.72, 0.837) | |
| BDI | |
| Conductivity (560.79, 0.03) | 78%, 63% |
| Nitrites (11.87, 0.05) | |
| Phosphates (11.63, 0.08) | |
| Ammonium (6.82, 0.11) | |
| Morphological condition (4.55, 0.42) | |
| Organic contamination (4.51, 0.45) | |
| Nitrates (3.42, 0.05) | |
| Total P (3.20, 0.07) | |
| HQA (2.78, 0.52) | |
| Total N (2.66, 0.05) | |
| Extensive agriculture (2.61, 0.75) | |
| CEC | |
| Oxidability (8.39, 0.65) | 74%, 65% |
| Extensive agriculture (5.78, 0.08) | |
| Natural areas (5.04, 0.06) | |
| Sediments discharge (4.92, 0.49) | |
| Total N (3.90, 0.67) | |
| COD (3.43, 0.76) | |
| Acidification and toxicity (2.44, 0.77) | |
| $O_2$ (2.41, 0.79) | |
| GDI | |
| TSS (9.78, 0.20) | 69%, 63% |
| $O_2$ (6.30, 0.79) | |
| pH (3.18, 0.88) | |
| Conductivity (3.02, 0.20) | |
| HMS (2.64, 0.80) | |
| MoDi | |
| Conductivity (336.95 0.01) | 98%, 74% |
| Nitrites (18.12, 0.03) | |
| Phosphates (11.30, 0.07) | |
| $BOD_5$ (11.02, 0.001) | |
| COD (10.12, 0.001) | |
| Urban area (9.94, 0.31) | |
| Ammonium (6.68, 0.10) | |
| Nitrates (6.44, 0.21) | |
| TSS (5.04, 0.21) | |
| Extensive agriculture (4.83, 0.51) | |
| HMS (4.33, 0.53) | |
| Connectivity (3.97, 0.38) | |
| Organic contamination (3.66, 0.35) | |
| pH (3.01, 0.55) | |
| Total P (2.37, 0.06) | |
| Morphological condition (2.24, 0.39) | |

For all methods we obtained low cross-validation errors between 26 and 37%. These values are acceptable in ecological studies where systems are complex, and indicate that there is only a small percentage of the classification not explained by the variables considered in this study. The MoDi and the BDI showed a lower stability of their discriminant results, with larger differences between the classification matrix and the jackknife cross-validation matrix (Daniel, 1989). However the final classification after the cross-validation is still good, especially for the MoDi (74%).

The DA results confirmed the sensitivity of diatoms to organic pollution, as indicated in many other studies (e.g., Van Dam et al., 1994; Rimet et al., 2005; Dere et al., 2006; Blanco et al., 2007) and also to nutrient load (e.g., Weckström and Juggins, 2005; Dela-Cruz et al., 2006; Potapova and Charles, 2007; Blanco et al., 2007) with all variables in the category of nutrients and organic contamination being related to the indices and model outputs at least once. Among those, nitrates, nitrites, phosphates, water oxygenation and conductivity were the variables more often chosen by the DA. The MoDi assessments were those discriminated by a high number of variables from nutrients and organic contamination (9 among 12) followed by the BDI (8), CEC (5), SPI and GDI (2).

Other studies (Pan et al., 1999; Griffith et al., 2002; Tison et al., 2005) indicate that coarse-scale factors, such as land cover/use in watersheds, geology and relief, explain diatom community variation. Our study confirmed that land use is an important disturbance affecting diatoms since the variables % of extensive agriculture and land use were selected by the DA for two and three indices or model, respectively. On the other hand, urban area and % of natural areas were selected only once each which shows that the most relevant anthropogenic activity affecting water quality in our region is agriculture. However, the variables % intensive agriculture in 5 km ratio and % of extensive agriculture in 5 km ratio were not related to any of the methods tested, even though their correspondent variables at the catchment scale (e.g., % agriculture) were shown to be important. This apparent contradiction indicates that the land use at the catchment scale has a greater influence over the water quality and diatoms communities than the most immediate activities. Finally the DA shows that the CEC (4 among 7) is apparently the most effective method in detecting this kind of pressures, followed by the MoDi (2) and the BDI, SPI (1). The GDI was unable to detect any of the land use variables.

However, the DA results show that other types of disturbance variables, not often associated with changes in diatoms structure (but see Potova, 1996), can produce an effect over the assessments of some methods. In the present study morphological changes such as the construction of artificial walls or embankments in the stream (reflected in the categorical variable morphological condition), alterations in the natural connectivity with lower and upper reaches (by small weirs, or dams) or modifications in the riparian corridor were reflected in the assessments of one or two indices/model. For these type of pressures the MoDi was the most efficient method, detecting changes in three parameters (among 5), followed by the BDI (2), and the CEC and GDI (1). This is not a surprising result because the main aim of the diatom indices was to evaluate water quality changes (mainly organic and

| Pressures | SPI | BDI | CEC | GDI | MoDi |
|---|---|---|---|---|---|
| **Table 4 – Type of boxplots (1–4) observed for each pressure variable selected by the discriminant analysis (DA)** | | | | | |
| Land use | | | | | |
|   Intensive agriculture | – | – | 4 | – | – |
|   Extensive agriculture | – | 4 | * | – | 4 |
|   Natural areas | – | – | * | – | – |
|   Intensive agriculture 5 | – | – | – | – | – |
|   Extensive agriculture 5 | – | – | – | – | – |
|   Land use | 2 | – | 3 | – | – |
|   Urban area | – | – | – | – | 4 |
| Nutrients and organic contamination | | | | | |
|   Nitrates | – | 3 | 4 | – | 3 |
|   Nitrites | 1 | 3 | – | – | 2 |
|   Ammonium | – | * | – | – | * |
|   Phosphates | * | 3 | – | – | 2 |
|   Total N | – | 3 | 4 | – | – |
|   Total P | – | 3 | – | – | 2 |
|   COD | – | – | * | – | * |
|   $BOD_5$ | – | – | – | – | * |
|   Oxidability | – | – | 4 | – | – |
|   $O_2$ | 4 | – | 1 | * | – |
|   Conductivity | – | * | – | * | 1 |
|   Organic contamination and nutrient enrichment | – | 4 | – | – | 2 |
| Water acidification and toxicity | | | | | |
|   pH | – | – | – | * | 1 |
|   Acidification and toxicity | * | – | * | – | – |
| Suspended solids | | | | | |
|   TSS | – | – | – | 4 | 3 |
|   Sediments discharge | 1 | – | * | – | – |
| Morphology of the channel and banks | | | | | |
|   Riparian zone | – | – | 3 | – | – |
|   Morphological condition | – | 4 | – | – | 4 |
|   Connectivity | – | – | – | – | 4 |
|   HMS | – | – | – | 4 | 4 |
|   HQA | – | * | – | – | – |

'–' indicates the variables not selected by the DA; '*' signs the variables selected by the DA that did not show a clear pattern in the boxplots.

nutrient contamination, pH, conductivity) rather than changes in the structure and morphology of the stream and channel (these latter variables were not considered in the databases when the indices were constructed). In contrast, the construction of the predictive model included a much larger set of variables, not only water chemical and physical parameters but also those related to channel morphology and structure. Potapova (1996) also found that differences in algal community structure of continental and coastal rivers were influenced by the discharge pattern. The discharge pattern is related to flow regimes and current velocity and can thus be modified by the construction of artificial walls or dams (parameters selected in our study).

The category of suspended solids can also be indirectly related to the previous one since the presence of transversal structures leads to changes in the hydrological regime thus promoting alterations in the sediments transport and deposition. So, it is not surprising that a variable from this category was selected by the DA for all assessment methods, except the BDI.

The pattern (graph type) observed in the boxplots was not coherent across the five methods for the same variable, which indicates that even when two methods react to the same

pressure they have differential responses over the classification system. Yet, for the BDI and also the MoDi most of the graphs showed an approximately continuous pattern (type 3 or 2) for nutrients and organic contamination, indicating that a small and gradual increase of this type of disturbances is reflected in the methods applied. On the other hand, other kinds of pressures, such as land use or morphology, show more irregular patterns, which imply that they influence the overall assessment but there is not a direct relation between the amount of pressure and the effect at the community level.

The category of water acidification and toxicity shows the most unclear patterns. However a number of studies refer the sensitivity of diatoms to pH and acidification (e.g., Yangdong et al., 1996; Orendt, 1998; Tipping et al., 2001). Therefore, our results were probably limited by the low range of values obtained for these variables since the Mondego, Vouga and Lis catchments are not specially affected by acidification.

The correlation analysis between the five indices/model showed that the assessments provided highly significant correlations between the SPI, BDI, CEC and GDI, even though the r values for the correlation of the GDI with the others are lower. These results can be explained by the common sensitivity of BDI, CEC and SPI to nutrients and

organic contamination (Descy and Coste, 1991; Gomà et al., 2005; Blanco et al., 2007).

The lower correlations between the GDI (Rumeau and Coste, 1988) and the others maybe due to the lower level of taxonomic discrimination (genus) used in this index when all other methods (including the model) used species or sub-species. Prygiel et al. (1996) underlined the importance of discriminating the species of some genera, such as *Navicula* and *Nitzschia*, due to their widely differing ecologies. The ongoing taxonomic rearrangement of diatoms that leads to the splitting of large genera into a growing number of smaller ones results in smaller diversity of ecological preferences within each genus. Incorporating the modern generic ranges into the GDI is therefore likely to improve the performance of the index because the new sets of co-generic species will tend to be more coherent in their ecological characteristics. And although identification at the generic level becomes more complex with the increase in the number of genera, it is still much simpler than the accurate identification of species or infraspecific taxa. A reformulated GDI may then become an interesting tool for water quality evaluation requiring relatively little taxonomic preparation.

Finally the high correlations found between the MoDi and the BDI are supported by the results of the DA since these two discriminated a higher number of variables in nutrients and organic contamination category (also the one including more variables), especially conductivity (first variable selected) and had also as common discriminators the variables extensive agriculture and morphological condition.

## 5. Conclusions

This study confirmed that diatoms are an important bioindicator for assessing streams ecological quality since the global assessments express their high sensitivity to changes in water quality due to organic and nutrient contamination but also to changes in the natural morphology of the channel and banks or land use.

However we showed that different methods for attributing quality classes to streams based on diatom communities may produce different assessments due to differences in their sensitivity to various types of pressures. The predictive model approach was found to be the better approach to detect changes in nutrients and organic contamination, morphology of the channel and banks, water acidification and toxicity, suspended solids and was the second best approach for land use; the CEC could be used to detect changes in land use and nutrients but performed poorly with the other disturbances; the BDI could detect some changes in morphological conditions but is especially useful for nutrients and organic contamination; the SPI could also be used to assess nutrients and organic contamination but was not as effective as the model the BDI or the CEC. Finally the GDI was found to be the least useful tool.

In conclusion the predictive model (MoDi) is appropriate for the assessment of streams in central Portugal because it expresses quantitative and qualitative changes in diatom communities and reflects a wider variety of impacts. Considering the underlying philosophy of the predictive model

("reference condition approach", Reynoldson et al., 1997) it is also the most adequate method to determine the ecological status of streams required by the WFD (Directive 2000/60/EC, 2000).

## REFERENCES

Agência Portuguesa do Ambiente, 2007. Atlas do Ambiente Digital, http://www.iambiente.pt/atlas/est/index.jsp.

Blanco, S., Bécares, E., Cauchie, H.-M., Hoffmann, L., Ector, L., 2007. Comparison of biotic indices for water quality diagnosis in the Duero Basin (Spain). Arch. Hydrobiol. Suppl. 161, 267–286.

CEN TC230 N68, 2003. Water Quality: Guidance for Routine Sampling of Benthic Algae in Shallow Swift Running Waters. Comité Europeén de Normalisation.

CORINE Land Cover, 2000. Instituto do Ambiente, Portugal.

Coste, M., 1986. Les méthodes microfloristiques d'évaluation de la qualité des eaux. Cemagref, Bordeaux, 15 pp + 46 annexes.

Daniel, L.G., 1989. Use of the Jackknife statistic to establish the external validity of discriminant analysis results. In: Proceedings of the Annual Meeting of the Southwest Educational Research Association (Houston).

Decreto Regulamentar 9/2002, 2002a. Plano de Bacia Hidrográfica do Mondego. Diário da República, I SÉRIE-B 51, 1695–1745.

Decreto Regulamentar 15/2002, 2002b. Plano de Bacia Hidrográfica do Vouga, volume II, capítulo III, Lisboa. Diário da República, I SÉRIE-B 78, 2354–2392.

Decreto Regulamentar 23/2002, 2002c. Plano de Bacia Hidrográfica do Lis. Diário da República, I SÉRIE-B 78, 2967–3015.

Dela-Cruz, J., Pritchard, T., Gordon, G., Ajani, P., 2006. The use of periphytic diatoms as a means of assessing impacts of point source inorganic nutrient pollution in south-eastern Australia. Freshwater Biol. 51, 951–972.

Dere, Ş., Dalkiran, N., Karacaoğlu, D., Elmaci, A., Dülger, B., Şentürk, E., 2006. Relationships among epipelic diatom taxa, bacterial abundances and water quality in a highly polluted stream catchment, Bursa, Turkey. Environ. Monit. Assess. 112, 1–22.

Descy, J.-P., Coste, M., 1991. A test of methods for assessing water quality based on diatoms. Verh. Internat. Verein. Limnol. 24, 2112–2116.

Directive 2000/60/EC, 2000. Water Framework Directive of the European Parliament and the Council, of 23 October 2000, Establishing a Framework for Community Action in the Field of Water Policy. Off. J. Eur. Commun., L327, 1–72.

Ector, L., Rimet, F., 2005. Using bioindicators to assess rivers in Europe: an overview. In: Lek, S., Scardi, M., Verdonschot, P.F.M., Descy, J.-P., Park, Y.-S. (Eds.), Modelling Community

Structure in Freshwater Ecosystems. Springer, Berlin, pp. 7–19.

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

EN 14407, 2004: Water Quality: Guidance Standard for the Identification, Enumeration and Interpretation of Benthic Diatom Samples from Running Waters.

Environmental Agency, 2003. River Habitat Survey in Britain and Ireland. Field Survey Guidance Manual: 2003 version. EA, SEPA, Environment and Heritage Service.

Feio, M.J., Almeida, S.F.P., Craveiro, S.C., Calado, A.J., 2007. Diatoms and macroinvertebrates provide consistent and complementary information on environmental quality. Fundament. Appl. Limnol. Arch. Hydrobiol. 168, 247–258.

Gomà, J., Rimet, F., Cambra, J., Hoffmann, L., Ector, L., 2005. Diatom communities and water quality assessment in Mountain Rivers of the upper Segre basin (La Cerdanya, Oriental Pyrenees). Hydrobiologia 551, 209–225.

Griffith, M.B., Hill, B.H., Herlihy, A.T., Kaufmann, P.R., 2002. Multivariate analysis of periphyton assemblages in relation to environmental gradients in Colorado Rocky Mountain Streams. J. Phycology 38, 83–95.

ISO 8467, 1986. Water Quality, Determination of Permanganate Index. ISO International Standard (ISO), 8467, Geneva, Switzerland.

John, J., 2000. A Guide to Diatoms as Indicators of Urban Stream Health. National River Health Program. Urban Sub Program, Report no. 7, Land and Water Resources Research and Development Cooperation LWRRDC Occasional Paper 14/99.

Kelly, M.G., Whitton, B.A., 1995. The trophic diatom index: a new index for monitoring eutrophication in rivers. J. Appl. Phycol. 7, 333–433.

Krammer, K., Lange-Bertalot, H. (Eds.), 1986. Süßwasserflora von Mitteleuropa. Bacillariophyceae 1 Teil: Naviculaceae, Band 2/1. 1986. Gustav Fischer Verlag, Stuttgart, p. 876.

Krammer, K., Lange-Bertalot, H. (Eds.), 1988. Süßwasserflora von Mitteleuropa. Bacillariophyceae 2 Teil: Bacillariaceae, Epithemiaceae, Surirellaceae, Band 2/2. 1988. Gustav Fischer Verlag, Stuttgart, p. 596.

Krammer, K., Lange-Bertalot, H. (Eds.), 1991a. Süßwasserflora von Mitteleuropa. Bacillariophyceae 3 Teil: Centrales, Fragilariaceae, Eunotiaceae, Band 2/3. 1991. Gustav Fischer Verlag, Stuttgart, p. 600.

Krammer, K., Lange-Bertalot, H. (Eds.), 1991b. Süßwasserflora von Mitteleuropa. Bacillariophyceae 4 Teil: Achnanthaceae, Kritische Ergänzungen zu Navicula (Lineolatae) und Gomphonema. Band 2/4. 1991. Gustav Fischer Verlag, Stuttgart, p. 437.

Lecointe, C., Coste, M., Prygiel, J., 1993. ''Omnidia'': software for taxonomy, calculation of diatom indices and inventories management. Hydrobiologia 269/270, 509–513.

Lecointe, C., Coste, M., Prygiel, J., Ector, L., 1999. Le logiciel OMNIDIA version 2, une puissante base de données pour les inventaires de diatomées et pour le calcul des indices diatomiques européens. Cryptogr. Algol. 20, 132–134.

Lenoir, A., Coste, M., 1996. Development of a practical diatom index of overall water quality applicable to the French national water board network. In: Whitton, B.A., Rott, E. (Eds.), Use of Algae for Monitoring Rivers II. Studia Student G.m.b.H, Innsbruck, Austria.

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.

Norris, R.H., Hawkins, C.P. (Eds.), 2000. Monitoring river health. Hydrobiologia 435, 5–17.

Orendt, C., 1998. Macroinvertebrates and diatoms as indicators of acidification in forest spring brooks in a region of eastern Germany (Leipzig–Halle–Bitterfeld) highly impacted by industrial activities. Arch. Hydrobiol. 143, 435–467.

Pan, Y., Stevenson, R.J., Hill, B.H., Kaufmann, P.R., Herlihy, A.T., 1999. Spatial patterns and ecological determinants of benthic algal assemblages in mid-Atlantic streams, USA. J. Phycol. 35, 460–468.

Philibert, A., Gell, P., Newall, P., 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.

Pohar, M., Blas, M., Turk, S., 2004. Comparison of logistic regression and linear discriminant analysis: a simulation study. Metodološki zvezki 1, 143–161.

Pont, D., Hugueny, B., Beier, U., Goffaux, D., Melcher, A., Noble, R., Rogers, C., Roset, N., Schmutz, S., 2006. Assessing river biotic condition at a continental scale: a European approach using functional metrics and fi sh assemblages. J. Appl. Ecol. 43, 70–80.

Potapova, M., 1996. Epilithic algal communities in rivers of the Kolyma Mountains, NE Siberia, Russia. Nova Hedwigia 63, 309–334.

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

Prygiel, J., Coste, M. (Eds.), 2000. Guide Méthodologique pour la mise en oeuvre de l'Indice Biologique Diatomées. NFT90-354, France, 134 pp.

Prygiel, J., Lévêque, L., Iserentant, R., 1996. A new practical diatom index for the assessment of water quality in monitoring networks. Rev. Sci. Eau. 9, 97–113.

Reynoldson, T.B., 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 to assess water—quality impairment using benthic macroinvertebrates. J. North. Am. Benthol. 16, 833–852.

Reynoldson, T.B., Rosenberg, D.M., Resh, V.H., 2001. Comparisons of models predicting invertebrate assemblages for biomonitoring in the Fraser River catchment, British Columbia. Can. J. Fish. Aquat. Sci. 58, 1395–1410.

Rimet, F., Cauchie, H.M., Hoffmann, L., Ector, L., 2005. Response of diatom indices to simulated water quality improvements in a river. J. Appl. Phycol. 17, 119–128.

Rumeau, A., Coste, M., 1988. Initiation à la systématique des diatomées déau douce. Bull. Fr. Pêche Piscic. 309, 1–69.

Simpson, J.C., Norris, R.H., 2000. Biological assessment of river quality: development of AUSRIVAS modes 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, U.K., pp. 125–142.

Sládeček, V., 1986. Diatoms as indicators of organic pollution. Arch. Hydrochim. Hydrobiol. 14, 555–566.

Tison, J., Park, Y.-S., Coste, M., Wasson, J.G., Ector, L., Rimet, F., Delmas, F., 2005. Typology of diatom communities and the influence of hydro-ecoregions: a study on the French hydrosystem scale. Water Res. 39, 3177–3188.

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

Watanabe, T., Asai, K., Houki, A., 1986. Numerical estimation to organic pollution of flowing water by using the epilithic

diatom assemblage—diatom assemblage index (DAIpo). Sci. Total Environ. 55, 209–218.

Weckström, K., Juggins, S., 2005. Coastal diatom-environment relationships from the Gulf of Finland, Baltic Sea. J. Phycol. 42, 21–35.

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

Yangdong, P., Jan Stevenson, R., Hill, B.H., Herlihy, A.T., Collins, G.B., 1996. Using diatoms as indicators of ecological conditions in lotic systems: a regional assessment. J. North Am. Benthol. Soc. 15, 481–495.

Zelinka Von, M., Marvan, P., 1961. Zur Prazisierung der biologischen Klassifikation der Reinheit fliessender Gewässer. Arch. Hydrobiol. 57, 389–407.