Hydrobiologia (2006) 566:247–260

M.T. Furse, D. Hering, K. Brabec, A. Buffagni, L. Sandin & P.F.M. Verdonschot (eds), The Ecological Status of European Rivers: Evaluation and Intercalibration of Assessment Methods
DOI 10.1007/s10750-006-0092-5

Uncertainty in diatom assessment: Sampling, identification and counting variation

Anna Besse-Lototskaya*, Piet F.M. Verdonschot & Jos A. Sinkeldam
*Freshwater Biology, Alterra, Wageningen University and Research, P.O. Box 47, 6700 AA Wageningen, The Netherlands (*Author for correspondence: E-mail: anna.besse@wur.nl)*

Key words: diatoms, multimetric index, standardisation, ecological assessment

Abstract

Despite the widespread application of periphytic diatoms to water quality assessment at a regional level, there is no standard European sampling protocol or associated assessment metrics. Furthermore, relatively little is known about the uncertainty in the results of such assessments. One of the objectives of the European project for the Standardisation of River Classifications (STAR) is to improve and standardise diatom assessment methods. An extensive diatom ring test, together with an audit of the project results, provided a better understanding and quantification of the uncertainty in quality assessment of running waters using diatoms. The variation in multimetric analysis shows that the choice of site and substrate for sampling, the inter-operator differences in diatom taxonomy and the counting techniques are the primary sources of uncertainty. To some extent, this variation also reveals the robustness of specific metrics in relation to the sources of uncertainty. Of the three most common substrate types tested (stone, macrophyte and sediment), macrophytes emerge as the most preferred substrate for diatom sampling when performing multimetric water quality assessment.

Introduction

Periphytic diatoms have been included in the assessment of river water quality since the early 1900s, are known to be reliable indicators of water conditions, and can be used successfully for assessing water quality in running waters (e.g., Coring, 1999; Ector & Rimet, 2005). It has been shown that diatoms react to changes in the intensity of eutrophication, acidity, saprobity, nitrogen, salinity and current velocity (e.g., Denys, 1991a, b; Battarbee et al., 1997; van Dam, 1997; Kelly, 1998; Coring, 1999). Many studies of running water diatom assemblages by different research schools present and compare the results from different regions, water types and microhabitats. However, there relatively little is still known about the comparability of these results and the degree of uncertainty in diatom assessment caused by differences in methodology for sampling, identification and counting techniques. Several studies show that, for example, the choice of substrate for sampling can play an important role in the assessment of the diatom community (Stevenson & Hashim, 1989; Snoeijs, 1991; Rolland et al., 1997; Rothfritz et al., 1997; Kelly et al., 1998). However, substrate type differs from stream type to stream type. For example, sampling stones in lowland streams is difficult or impossible because they are simply not present. The use of macrophytes as substrate induces problems caused by differences in the composition and abundance of diatom species colonising different parts of macrophytes, such as leaves, stem or root (Cazaubon, 1996). In contrast, Gomez & Licursi (2001) conclude that soft sediment provides the most appropriate diatom community for monitoring in lotic systems. Furthermore, factors such as the choice of

sampling site, and the methods for preparing and processing the sample and identifying the taxa can be crucial to the assessment results.

This paper attempts to identify and quantify the sources of uncertainty in the assessment of diatoms from running waters by comparing the results of an extensive diatom ring test that included simultaneous sampling from multiple sites and substrates, and replicate sampling and slide preparation, performed by different operators. The results of the ring test are also used to evaluate the reliability and precision of metrics derived from the diatom community. The objective of this paper was to test the diatom assessment methodology that was selected by the STAR project as a standard for studying European stream and river systems (Furse et al., this volume), and to audit the quality of diatom results from the project.

Materials

Diatom ring test

In order to identify and quantify the error in diatom assessment introduced by sampling, identification and counting, a ring test was performed during a training course on the Plaine River, France. A total of 116 samples were taken and analysed by 10 of the 14 partners that participated in the STAR diatom studies according to a standard diatom field and laboratory protocol (Furse et al., this volume). The samples were taken from 2 sites (PL0 and PL5), and from 3 substrate types (stones (H), macrophytes (M) and sediments (S)). The most common substrate type at both sampling sites was stone. Generally, each partner collected and analysed 3 samples from two out of three habitat types at each site (Table 1). Additionally, another test was performed where one participant (Alterra) prepared and analysed replicate slides from two of the ring test samples (4 in one sample, and 5 in another sample).

Audit of STAR diatom results

The main diatom assessment for the STAR project was performed by 14 partners according to the standardised diatom field and laboratory protocol (Furse et al., this volume). The samples were collected from various stream types, processed and analysed (i.e., identified and counted) by the partners in their respective laboratories.

The quality of the partners' analyses was audited by Alterra. Thirty eight percent of the STAR diatom samples were randomly selected from all the samples taken by each project partner. A total of 107 slides were analysed a second time by the auditor, where a new count and identification was undertaken according to the protocol.

Table 1. The distribution of the diatom samples per site (PL0 and PL5), substrate (H = stone, M = macrophyte, S = sediment) and partner

Partner	Code	Institute	PL0H	PL0M	PL0S	PL5H	PL5M	PL5S	Total
1	UK	Centre for Ecology and Hydrology, UK	3		3	3			9
2	D	University of Essen & Research & Institute Senkenberg, Germany		3	2		3	3	11
3	A	University of Agricultural Sciences, Vienna, Austria	3		3	3		3	12
5	S	Swedish University of Agricultural Sciences, Sweden	3	3		3	3		12
6	C	Masaryk University, Brno, Czech Republic	3		3	3		3	12
8	I1	Istituto di Recerca sulle Acque (IRSA-CNR), Italy	3		3	3		3	12
9	P	University of Evora, Portugal	3	3		3	3		12
10	DK	National Environmental Research Institute, Denmark		3	3		3	3	12
13	I2	Province of Bolzano (LABBIO), Italy	3	3		3	3		12
14	F	University of Metz, France		3	3		3	3	12
		Total	21	18	20	21	18	18	116

Methods

Taxonomic adjustment

In order to compare the diatom results between the partners, the nomenclatural differences between the partners first had to be resolved. Taxa were identified to the lowest achievable taxonomic level (species and/or variety/forma). By exchanging the results amongst the STAR partners through a round of comments, the level of identification was raised and the results improved. The taxonomic nomenclature used for all the results was then adjusted to the standardised STAR diatom taxa list that was agreed by all partners and experts.

Diatom metrics

The comparison of the diatom results was based on a total of 17 metrics. The OMNIDIA program (Lecointe et al., 2003) was used to compute 14 different diatom metrics that are regularly used to assess several aspects of water quality, mainly in running waters (Table 2). Other parameters such as number of taxa, Shannon diversity (Zar, 1996) and evenness (Zar, 1996) were also used in the comparison. The values of the metrics IPS, SLAD, DSECY, L&M, SHE, WAT, TDI, EPI-D, ROTT, IDG, CEE, IBD and IDAP were transformed by the OMNIDIA program to a scale from 0 to 20; the scale of number of taxa and Shannon diversity is infinite; the evenness and %PT values range between 0–1 and 0–100, respectively.

Data analyses

The values of diatom metrics for all samples were compared using the average value and the standard deviation per respective group of samples.

The diatom ring test results were split into the following sets, for which the average value and the standard deviation were calculated:

- the whole database (1 set of 116 samples)
- per site (2 sets of 57 and 59 samples, respectively)
- per substrate (3 sets of 42, 36, and 38 samples, respectively)
- per partner (10 sets of 9–12 samples each)
- per replicate sample (39 sets of 2–3 samples each)

The average value and the standard deviation for the replicate slide dataset were plotted for each sample (2 sets of 4 and 5 samples).

The relative diatom counts were ordinated by redundancy analysis (RDA) with the program

Table 2. Metrics used for the comparison of the diatom results

Abbreviation	Full name	Reference
No. taxa	Number of taxa	
Diversity	Shannon diversity	(MVSP, 2001)
Evenness	Evenness	(MVSP, 2001)
IPS	Specific Pollution Sensitivity Metric	(Coste, 1987)
SLAD	Sládeček's pollution metric	(Sládeček, 1986)
DESCY	Descy's pollution metric	(Descy, 1979)
L&M	Leclercq & Maquet's pollution metric	(Leclercq & Maquet, 1987)
SHE	Steinberg & Schiefele trophic metric	(Steinberg & Schiefele, 1988)
WAT	Watanabe et al. pollution metric	(Lecointe et al., 2003)
TDI	Trophic Diatom metric	(Kelly & Whitton, 1995)
%PT	% pollution tolerant taxa	(Kelly & Whitton, 1995)
EPI-D	Pollution metric based on diatoms	(Dell'Uomo, 1996)
ROTT	Trophic metric	(Rott et al., 1999)
IDG	Generic Diatom Metric	(Lecointe et al., 2003)
CEE	Commission for Economical Community metric	(Descy & Coste, 1991)
IBD	Biological Diatom Metric	(Prygiel & Coste, 1999)
IDAP	Indice Diatomique Artois Picardie	(Lecointe et al., 2003)

CANOCO 4.5 (Ter Braak & Ŝmilauer, 2002). The data analysis is fully described by Verdonschot & Ter Braak (1994). RDA assumes a linear model for the relationship between the response of each taxon and the ordination axes and is used if the gradient length in the data is short (<4 units of standard deviation [SD]; Ter Braak, 1988). In our case the gradient length was smaller then 3 SD (axis 1: 2.1 and axis 2: 3.0) which implies that the data are quite homogeneous. RDA is the constraint form of PCA of taxon data, in which the components (axes) are constrained by linear combinations of environmental variables. The ordination results are presented as correlation biplots of sites and environmental variables (Verdonschot & Ter Braak, 1994). The eigenvalue of an ordination axis in RDA is the proportion of the total variance explained by that axis and indicates its relative importance. An unrestricted permutation test is used to test the validity of the total ordination. This technique is fully explained by Ter Braak & Ŝmilauer (2002) and Verdonschot & Ter Braak (1994). For this ordination the full diatom dataset was used, and the parameters partner, site, substrate and replicate were defined as nominal and included as environmental parameters.

For the audit database, the results from the partners were compared with the audit results. The average value and standard deviation were calculated for each partner versus audit sample (in total 107 sets each of 2 samples).

Furthermore, the average value of the SD was calculated for each of the above-mentioned datasets, in order to compare the methodological errors.

Diatom metrics from the ring test were also compared between partners and between replicates using an analysis of variance (ANOVA). The variance components were estimated by averages of restricted maximum likelihood (Patterson & Thompson, 1971). The hypothesis that there were no differences in the variance of metric values between replicate samples regardless of substrate type or sample site was tested with a chi-squared test. This test employs deviances differences as produced by restricted maximum likelihood. Analyses were performed with GenStat 8.11 (VSN International Ltd, 2002). If the probability (p) was less than 0.05, it was assumed that the hypothesis was not true, indicating that there was a relationship between the results and the sampling substrate or/and sampling site.

The correlation between each partner's and the auditor's metric was calculated, assuming that the correlation between the two is linear. The coefficient of determination (R^2) for each of the metrics is based on a dataset of 117 partner–audit samples. The R^2 above 0.5 is considered to indicate a relationship.

Results and discussion

The RDA ordination (total ordination is significant ($p < 0.002$)) of all samples resulted in a biplot (Fig. 1) that shows a clear separation of samples taken at site PL0 versus samples of site PL5. The two groups of samples are fully separated while at the same time each of the three replicate samples is plotted close to each other, while each of the groups of three replicate samples is clearly separated from the other groups of replicates.

Next, two ordination were run to establish the importance of either partner, or substrate, or

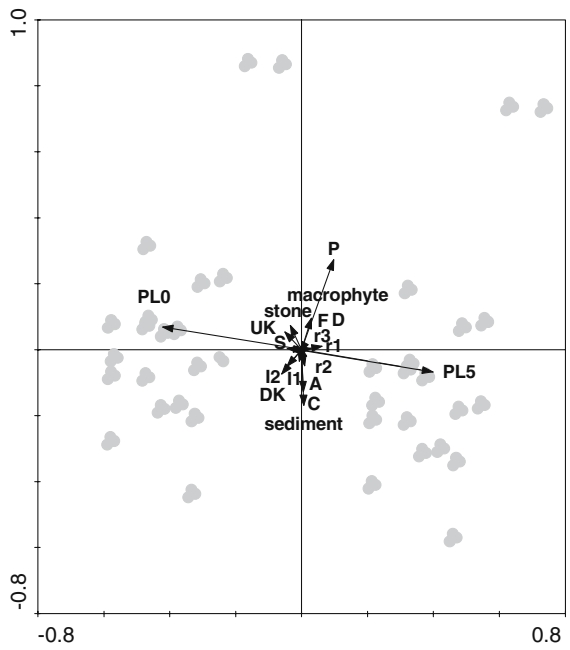


Figure 1. Ordination (RDA) diagram of the axis 1 and 2 (eigenvalues 0.18 and 0.11, respectively) showing the variation in the distribution of samples (grey dots) among environmental variables (arrows) in the two ring test sites (PL0 and PL5). Partner codes are given in Table 1. Replicates are coded r1, r2 and r3.

replicate as explanatory variable. The ordination of all samples of site PL0 (total ordination is significant ($p < 0.002$)) shows a clear separation based on the substrate parameters (Fig. 2). The grouping of replicates remains. The ordination of all samples taken at site PL5 (total ordination is significant ($p < 0.002$)) shows a different pattern (Fig. 3). Here the differences between partners, especially Czech Republic, Portugal, Denmark, Italy (second partner) and Sweden are most explaining. Again, the grouping of replicates remains. The differences found between sites PL0 and PL5 are due to the differences in homogeneity, in the sense of variation in environmental conditions between the habitats present, between the two sites. At the more homogeneous site PL0 the samples collected by the different partners were more alike. At this site the differences between substrates prevailed. The more heterogeneous site PL5, where local variation between habitats was much larger, resulted in an in-between partner variation. The differences between partners, who collected their replicate samples from individual spots within the stream site, can possibly be due to the instream

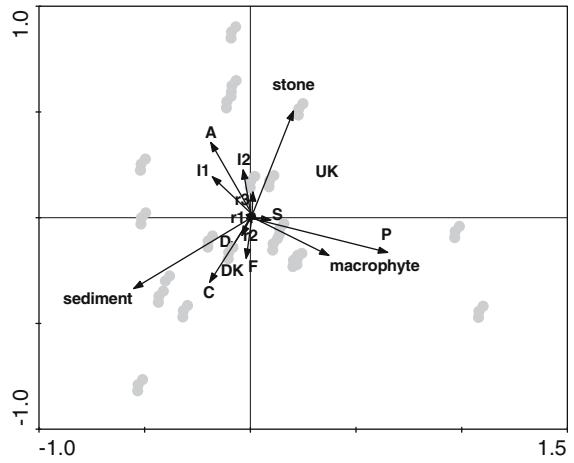


Figure 2. Ordination (RDA) diagram of the axis 1 and 2 (eigenvalues 0.24 and 0.18, respectively) showing the variation in the distribution of samples (grey dots) among environmental variables (arrows) at ring test site PL0. Partner codes are given in Table 1. Replicates are coded r1, r2 and r3.

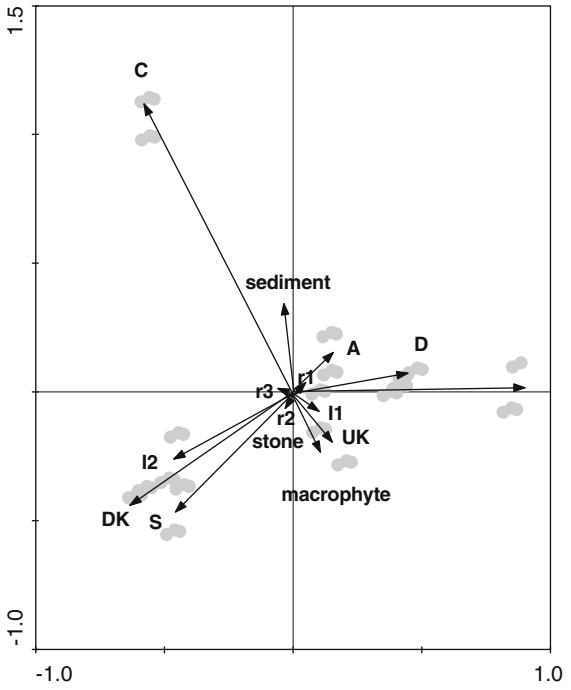
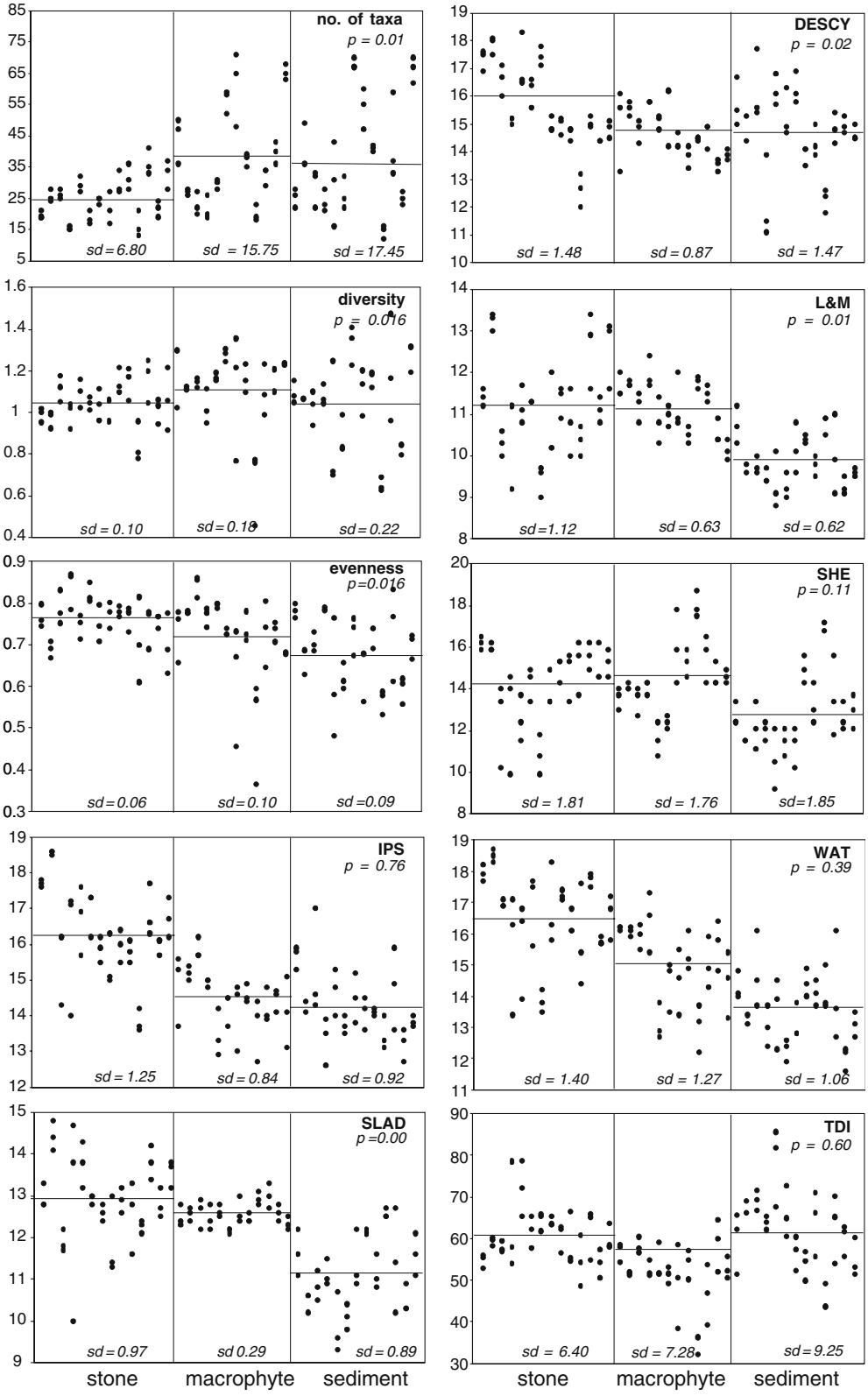


Figure 3. Ordination (RDA) diagram of the axis 1 and 2 (eigenvalues 0.18 and 0.16, respectively) showing the variation in the distribution of samples (grey dots) among environmental variables (arrows) at ring test site PL5. Partner codes are given in Table 1. Replicates are coded r1, r2 and r3.

variation. The consequence is that to sample diatoms in a representative way one should collect subsamples spread over various spots of, for the eye, more or less the same substrates within a larger transect of the stream site.

The metric results of the ring test are presented in Figure 4. Ideally, all 116 samples of the ring test should show the same value for each metric if the sampling site or substrate type fully represents the composition of the diatom community in the river and when all partners use exactly the same sampling, sample processing, identification and counting techniques. However, our results show differences between samples taken from different sites and substrates, by different partners and between replicate samples. It is thus important to find out:

Figure 4. The diatom ring test results, per metric. The horizontal axe indicates the sample number and is sorted by substrate type. The data in each substrate zone are sorted by site and further by partner. Each sample includes multiple values from replicate samples. ''p'' is the metric of the chi-squared test examining the hypothesis of variance in values between replicates regardless the sampled substrate. ''sd'' is the average standard deviation of metric values depending on the substrate type. Dashed line indicates the average metric value, per substrate type.


no. of taxa
p = 0.01
sd = 6.80
sd = 15.75
sd = 17.45
DESCY
p = 0.02
sd = 1.48
sd = 0.87
sd = 1.47
diversity
p = 0.016
sd = 0.10
sd = 0.18
sd = 0.22
L&M
p = 0.01
sd = 1.12
sd = 0.63
sd = 0.62
evenness
p=0.016
sd = 0.06
sd = 0.10
sd = 0.09
SHE
p = 0.11
sd = 1.81
sd = 1.76
sd = 1.85
IPS
p = 0.76
sd = 1.25
sd = 0.84
sd = 0.92
WAT
p = 0.39
sd = 1.40
sd = 1.27
sd = 1.06
SLAD
p = 0.00
sd = 0.97
sd 0.29
sd = 0.89
TDI
p = 0.60
sd = 6.40
sd = 7.28
sd = 9.25
stone
macrophyte
sediment
stone
macrophyte
sediment

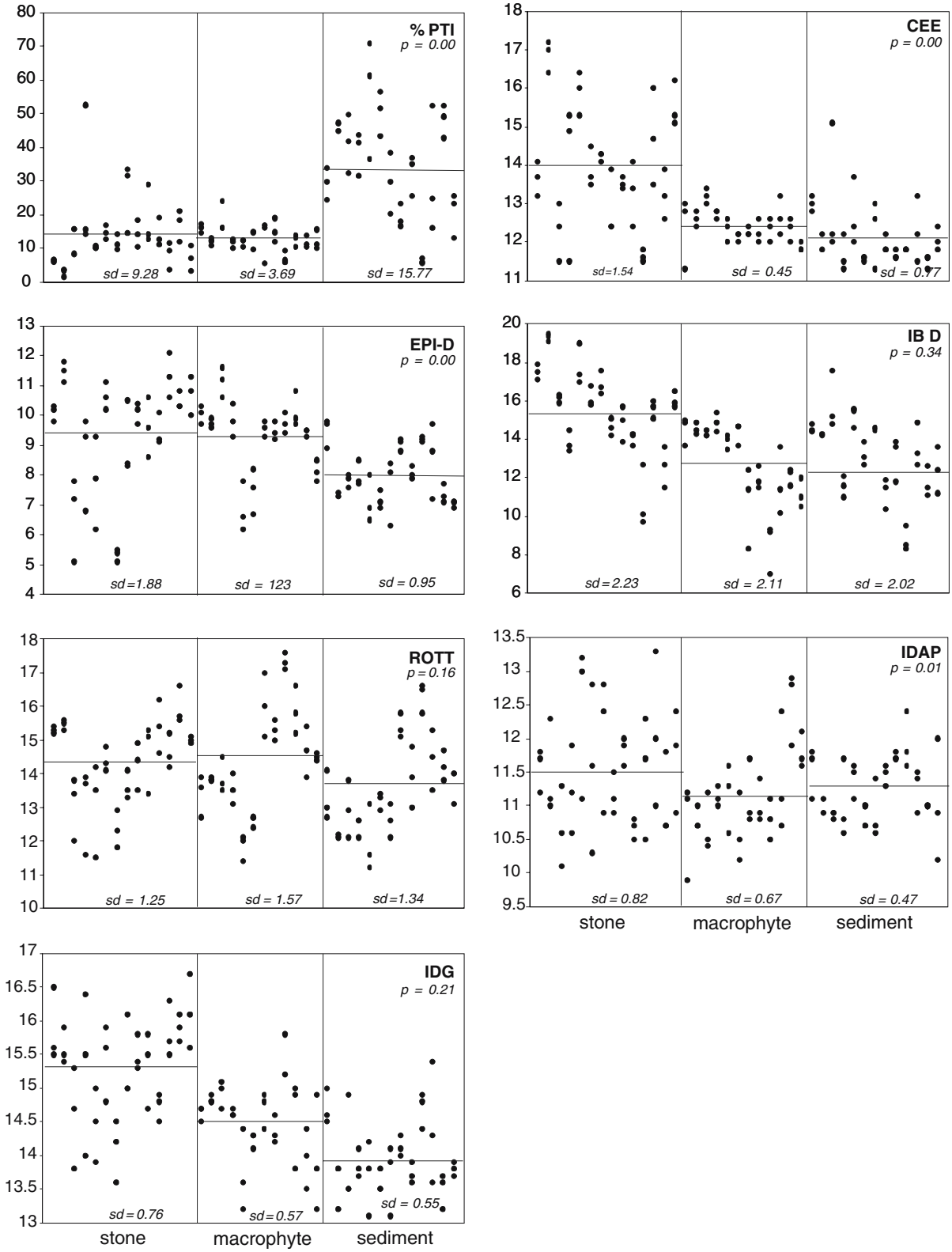


Figure 4. (Continued)

– what the sources of differences/uncertainty are;
– rank the sources of differences/uncertainty in order of importance and quantify the uncertainty;
– apply knowledge from the above to improve diatom assessments.

Choice of substrate for sampling

One of the most interesting and important questions that was posed for our diatom ring test was the choice of substrate for sampling and its influence on the diatom assessment results. Up to now, there is no standard method for sampling periphytic diatoms in running waters, and previous diatom studies have used various substrate types for sampling (among others Descy & Coste, 1991; Gomez, 1998; Rott et al., 1998). The results of studies testing the relationship between the diatom community and the substrate type were often conflicting. While several (Rothfritz et al., 1997; Winter & Duthie, 2000) showed that there is no consistent difference in the results of water quality monitoring using diatom community structure from different substrates, Snoeijs (1991) found that different types of substrates were colonised differently. Studies by Stevenson and Hashim (1989), Rolland et al. (1997) and Rothfritz et al. (1997) revealed significant differences in diatom diversity between different substrate types. Recent studies suggest hard natural or artificial substrates are the most suitable and reliable for ecological studies of periphytic diatoms (e.g., Kelly et al., 1998). Our diatom ring test included sampling from stone, macrophyte and sediment substrates and these are the most common substrates in running waters. Two sites were sampled by 10 partners in 2–3 replicates. When sorted according to substrate type, the variation and values of the resulting metric values between replicate samples and partners were compared by substrate type. This was done in order to relate the choice of substrate to the reliability of the results (Fig. 4).

Variation

In order to establish a possible relationship between the type of the substrate sampled and the results of the ring test, we compared the variation in each metric (Fig. 4, Table 3). The chi-squared test revealed that 8 metrics varied with habitat, namely number of taxa, SLAD, DESCY, L&M, % PT, EPI-D, CEE and IDAP ($p < 0.05$). The variation between replicate samples and between the partners, expressed as SDs, is also considerably lower for one (or several) substrate types than the other(s) (Table 3). The number of taxa varied least in the samples taken from the stone substrate; SLAD, DESCY, % PT and CEE, varied least in the macrophyte substrate; EPI-D and IDAP varied least in the sediment substrate; and L&M varied least in macrophyte and sediment samples. There was no relationship between the type of substrate and the variation of the nine other metrics (Table 3).

The relationship established between type of substrate and the variability in the results of 8 of the 17 water quality assessment metrics confirms the importance of the substrate choice when sampling diatoms. When using one of the diatom metrics, one should consider sampling the substrate providing the least variability. The macrophyte substrate is thus preferred for sampling diatoms when water quality is assessed using SLAD, DESCY, % PT and CEE. Sediment offers the most reliable substrate when the assessment is based on EPI-D and IDAP. Either macrophyte, or sediment substrate is preferred to stone when using L&M. The relatively high inter-partner variability in number of taxa for all substrates (especially for the macrophytes and the sediments) can be explained by the diatom counting technique. When we decided the sampling protocol for STAR (Furse et al., this volume), we agreed that 300 valves of diatoms (where possible) should be identified and counted from each sample. However, in a number of cases the protocol was not followed strictly; more valves were counted and/or the slide was surveyed for additional (rare) taxa after the count was completed. These technical discrepancies probably led to significant differences in the number of taxa found between the partners. The stone substrate, however, shows here the least variability of all substrates, and is thus the most representative substrate for sampling when using the number of taxa.

In the case of diversity, evenness, IPS, SHE, WAT, TDI, IDG and IBD, the choice of substrate for sampling does not play a significant role.

The macrophyte substrate gives the least variability (Table 3) and therefore appears to be the

Table 3. Variation (inter-partner and replicate) by substrate type, results of the chi-squared test of variance in replicate samples, and preferred substrate type, for the 17 diatom water quality assessment metrics

Metric	Standard deviation (sd)			*p*	Preferred substrate
	Stone	Macrophyte	Sediment		
Number of taxa	6.80	15.75	17.45	0.01	Stone
Diversity	1.05	1.11	1.05	0.16	None
Evenness	0.06	0.10	0.09	0.16	None
IPS	1.25	0.84	0.92	0.76	None
SLAD	0.97	0.29	0.89	0.00	Macrophytes
DESCY	1.48	0.87	1.47	0.02	Macrophytes
L&M	1.12	0.63	0.62	0.01	Macrophytes/sediment
SHE	1.81	1.76	1.85	0.11	None
WAT	1.40	1.27	1.06	0.39	None
TDI	6.40	7.28	9.25	0.60	None
%PT	9.28	3.69	15.77	0.00	Macrophytes
EPI-D	1.88	1.23	0.95	0.00	Sediment
ROTT	1.25	1.57	1.34	0.16	None
IDG	0.76	0.57	0.55	0.21	None
CEE	1.54	0.45	0.77	0.00	Macrophytes
IBD	2.23	2.11	2.02	0.34	None
IDAP	0.82	0.67	0.47	0.01	Sediment

most appropriate for diatom sampling in 5 out of 8 cases. In the remaining 3 cases the results do not differ between any of the substrate types. Thus, we conclude that if one uses a multiple metric approach, the macrophyte substrate is preferred more than the stone or the sediment as the substrate to be sampled.

Average value

The differences in the metric average values of each metric in each substrate type can also indicate the reliability of the substrate type and should be taken into account when choosing a substrate for sampling. To compare the discrepancies between the diatom metrics directly, the metrics should be given a common scale, as each metric is based on a different range of scores. Scaling is difficult because some metrics can score infinitely, so we only perform a qualitative comparison.

The average values of 11 of the 17 metrics vary significantly between substrates (number of taxa, IPS, SLAD, L&M, WAT, TDI, %PT, EPI-D, IDG, CEE and IBD; Fig. 4). This sequence of metrics is not related to the degree inter-substrate variation.

In the case of the number of taxa, the mean is considerably lower for the stone substrate compared to the macrophyte or the sediment substrate. The samples collected from stone never contained more than 40 taxa, whereas the samples from macrophytes and sediment included up to 68–70 taxa. The difference is due to the rare taxa collected by a number of partners on the macrophyte and the sediment substrates. In general, the partners collected less taxa on the stone substrate compared to the other substrates, although one would assume that the stone substrate supports a diatom community of multiple seasons, and is therefore richer in diversity. Our data oppose this hypothesis and suggest that if one strives to collect a high variety in diatom taxa, one should not focus on stone. However, stone is still the most reliable substrate (least variance) when comparing the number of taxa between the partners, independent of the counting technique used. All other diatom water quality assessment metrics are based on the relative abundance of different taxa and should therefore either be less or unaffected affected by the presence or absence of rare taxa in the samples.

The average values of all the other metrics (diversity, evenness, DESCY, SHE, ROTT, IDAP) (Fig. 4) did not vary significantly between substrates.

For the majority of the metrics where average values varied significantly between substrates, the average values from samples collected from macrophytes seldom exceed or are less than those of the other substrates. This is interpreted as a favourable feature of sampling from macrophytes.

Summarising the results of our ring test, the samples from the macrophyte substrate generally reveal the lowest inter-partner and inter-replicate variability, and show average values in comparison to the samples from the stone and sediment, and therefore should be used as preferred substrate for diatom sampling when performing multimetric water quality assessments.

Sources of uncertainty in diatom assessment

In order to assess and reduce the impact of error and uncertainty on diatom metrics, it is crucial to understand the sources of uncertainty, to quantify the error for each step of diatom sampling, and to identify measures that can decrease this uncertainty.

Uncertainty is introduced during all steps of sampling and can be quantified (Table 4). Our diatom ring test, together with the audit of analytical quality, provided a dataset that we used to quantify the sources of uncertainty and to set them in the order of significance. The variations (standard deviation and average standard deviation) in the diatom metric results in total, by site, substrate type, partner, replicate sample and by replicate slide are listed in Table 5.

The variation between samples from different sites (ring test) represents the uncertainty caused by the choice of sampling site; the variation between samples from different substrates (ring test) shows the contribution of the substrate choice during sampling to the uncertainty; the variation between replicate samples (ring test) adds to the uncertainty during sample collection. The variation between replicate slides in the ring test (the comparison of diatom composition from different slides taken from the same sample and identified and counted by the same operator) is similar to the variation between the partner and the auditor (diatom compositions taken from the same slide and identified and counted by different operators) because the random fields chosen for diatom identification and counting are different between both. The variation between partners (ring test) represents the uncertainty over all steps of sampling.

The total variation, expressed in the average SD between all samples of the ring test, comprises the sum of errors introduced at all steps of sampling and is highest in 12 out of 17 diatom metrics. The variation between the samples from different sites is generally the second highest (8 out of 17 cases) and is always lower than or equal to the total variation. The variation due to partner and substrate is generally ranked third or fourth. Partner is ranked third 9 times and fourth 5 times and substrate is ranked third 4 times and fourth 8 times. For the remaining audit is ranked fifth 7 times, replicate samples sixth 10 times and replicate slides seventh 16 times. Thus the order

Table 4. Linking diatom sampling steps to measures of uncertainty

sampling steps	**quantification of uncertainty**		
choice of sampling site	variation between samples from different sites		variation between partners
choice of substrate for sampling	variation between samples from different substrates		
sample collection	variation in replicate samples		
slide preparation	variation in replicate slides	Variation partner-auditor	

All sampling steps include uncertainty due to taxonomic identification and counting.

Table 5. Average standard deviation values (sd) for the ring test and the project audit results (after correction)

Average sd	Total	Site	Partner	Substrate	Audit	Repl. samples	Repl. slides
Number of taxa	15.11 (1)	14.81 (2)	6.15 (5)	13.33 (3)	11.60 (4)	4.05 (6)	1.97 (7)
Diversity	0.17 (2)	0.17 (3)	0.13 (5)	0.17 (4)	0.34 (1)	0.08 (7)	0.10 (6)
Evenness	0.09 (1)	0.08 (2)	0.07 (4)	0.08 (3)	0.05 (5)	0.04 (6)	0.03 (7)
IPS	1.39 (1)	1.34 (2)	1.12 (3)	1.00 (4)	0.64 (5)	0.57 (6)	0.32 (7)
SLAD	1.11 (1)	1.08 (2)	0.95 (3)	0.72 (4)	0.39 (6)	0.41 (5)	0.19 (7)
DESCY	1.38 (1)	1.10 (4)	1.13 (3)	1.27 (2)	0.50 (5)	0.49 (6)	0.34 (7)
L&M	1.04 (2)	1.04 (1)	0.94 (3)	0.79 (4)	0.41 (5)	0.39 (6)	0.16 (7)
SHE	1.94 (1)	1.63 (4)	1.84 (3)	1.81 (2)	0.52 (6)	0.83 (5)	0.36 (7)
WAT	1.76 (2)	1.76 (1)	1.54 (3)	1.24 (4)	0.79 (5)	0.69 (6)	0.45 (7)
TDI	8.78 (1)	8.37 (2)	7.24 (4)	7.64 (3)	5.91 (5)	4.35 (6)	3.50 (7)
% PT	14.57 (1)	14.15 (2)	12.36 (3)	9.58 (4)	2.14 (6)	5.12 (5)	1.97 (7)
EPI-D	1.55 (1)	1.48 (2)	1.29 (4)	1.35 (3)	0.48 (6)	0.55 (5)	0.30 (7)
ROTT	1.41 (1)	1.10 (4)	1.31 (3)	1.38 (2)	0.38 (6)	0.55 (5)	0.11 (7)
IDG	0.85 (1)	0.84 (2)	0.71 (3)	0.63 (4)	0.45 (5)	0.41 (6)	0.31 (7)
CEE	1.34 (2)	1.30 (3)	1.02 (4)	0.92 (5)	2.35 (1)	0.49 (6)	0.31 (7)
IBD	2.45 (1)	2.01 (3)	1.98 (4)	2.12 (2)	0.55 (6)	0.71 (5)	0.47 (7)
IDAP	0.69 (2)	0.68 (3)	0.59 (5)	0.65 (4)	1.21 (1)	0.49 (6)	0.38 (7)

of magnitude of variation is (from highest to smallest): total variation > sampling site variation > partner variation > substrate type variation > audit variation > sample collection variation (replicate sample) > slide preparation variation (replicate slide).

The variation between partners is larger than between substrates as it also includes replicate sample and slide variation. The audit variation is greater than the replicate sample and slide variation. Audit variation = slide variation + variation in analytical quality between partners. Some partners' analysts may consistently over or under count. Besides, as the audit comprised a variation in sites and water qualities, the identification uncertainty increased in comparison to the assessment of homogenous two ring test sites, despite the fact that identification uncertainty was reduced as much as possible by performing taxonomic adjustments to the standardised list of taxa.

In our experiment, we tested all steps of the sample processing procedure for diatom assessment. One must realise that, in the approach chosen, site variation includes partner, substrate, replicate sample and replicate slide variation, and substrate variation includes replicate sample and replicate slide variation. The variation is cumulative, in the order indicated. This is large in sampling site, partner variation, and substrate type, as all three approach the value for the total variation. The variation is small for slide and somewhat larger but still small for sample variation. The audit variation exceeds sample variation slightly. In general, uncertainty increases when inter-partner variation is introduced. This variation is related to the differences in environmental circumstances at the sampled sites, and also includes differences, though small, in sample treatment and preparation of slides between laboratories. All variation also includes uncertainty due to variation in identification (number of taxa and number of rare species detected with or without an extra survey for rare taxa) despite the adjustment to standard taxonomic list, and due to the number of valves counted.

Audit of the project results

The audit was performed to test whether partners kept to the protocol and whether the identification was correct. Ideally, project and audit results should be highly correlated because the auditor uses the same slide as the project partner did, and thus, only replicate slide variation is still present. The correlation between the partner and audit diatom metric results is generally high ($R^2 > 0.5$) (Table 6). However, the audit results show that the

Table 6. Correlation of diatom metrics between the STAR results and the audit

Metric	R^2 coefficient
Number of taxa	0.575
Diversity	0.557
Evenness	0.512
IPS	0.699
SLAD	0.751
DESCY	0.570
L&M	0.696
SHE	0.732
WAT	0.623
TDI	0.528
%PT	0.896
EPI-D	0.620
ROTT	0.748
IDG	0.818
CEE	0.288
IBD	0.771
IDAP	0.004

average variation between the audit and the project partner is always higher than the replicate slide variation but is generally lower than the total variation of the diatom ring test (Table 5). Unfortunately, during the STAR project the diatom protocol was not always strictly followed. Sometimes, different number of valves was counted and different substrates were sampled. Together with the different stream types surveyed these differences caused the observed variation.

The coefficient of determination R^2 is indicative for the susceptibility of a diatom metric to the determination differences between operators. Thus, the metric % PT appears to be the most robust ($R^2 = 0.90$), whereas metrics CEE ($R^2 = 0.29$) and IDAP ($R^2 = 0.00$) are probably sensitive to the identification and counting techniques (Table 6).

Conclusions and suggestions

This study based on an extensive diatom ring test and an audit of analytical quality, showed that sampling protocol plays a crucial role in the assessment of water quality using diatoms. The choice of sampling site and substrate type, and the taxonomic identification contribute the most to the uncertainty in the resulting water quality metrics. There is much controversy in the literature about the relationship between the substrate type and the composition of diatom assemblages, and its influence on water quality assessments. Of the three most common substrate types (stone, macrophyte and sediment), macrophyte generally gave the most consistent results, and thus should be used as preferred substrate for diatom sampling (if used with care) when performing multimetric water quality assessment. However, some diatom metrics perform better with samples from other substrates. In order to standardise the substrate choice in the sampling protocol, a further evaluation is needed in which the multimetric results are tested in relation to the substrate type only, while excluding all other variables such as site and partner.

Besides the choice of the substrate type, taxonomic identification proves to be another important contributor to the uncertainty in diatom assessment. Some partners are more experienced and more careful than others, and some have more skill. Some metrics are affected more by the degree of skill of the analysts, whilst others are more robust and do not require such highly skilled staff. This skill includes the ability to identify diatoms correctly and to differentiate between them, and to could accurately evaluate the sample without under- or over-estimation. These skills are often undervalued and the time that they take to develop is often not realised. If staff are not given sufficient time, the results will suffer. Good operational (c.f. research) laboratory management is a balance between analytical quality and number of samples analysed.

Furthermore, the ordination showed that if one would like to get a full and representative picture of the diatoms present at a stream stretch, one should collect subsamples. In order to overcome the within stream variation, the subsamples should be spread over various spots of, for the eye, more less the same substrates within a larger stretch of the stream site.

The taxonomic adjustment to a standardised list, incorporated in the protocol (Furse et al., this volume) is an excellent tool to reduce the uncertainty. However, the audit of the project results demonstrated that continuous active co-operation between diatom taxonomists is necessary in order to further improve the standard taxonomic list and

thus, reduce the uncertainty in the assessment results. Our experiment also proved that the standardisation of the counting technique of diatom valves could significantly reduce the uncertainty in metric results. It is essential to follow strictly the protocol of counting 300 valves and afterwards searching the slide for rare taxa in order to assess the total diversity.

Diatom metrics that are currently used for assessing water quality usually have regional character, mainly because of the limited geographical variation in the data used. Comparing results between the different ecoregions is therefore problematic, but can be improved by using a multimetric analysis. Our study shows that some metrics are more sensitive to application by different operators and at different sampling sites in different geographical regions and substrate types than the others. The computation and evaluation existing data from different European stream sites in multimetric context could possibly lead to the establishment of one or several metrics that could be used to assess water quality conditions at a European scale.

Acknowledgements

This paper is a result of the EU-funded project STAR (6th Framework Programme; contract number: EVK1-CT-2001–00089). We thank J. van der Molen for the initiation of the experiment and co-ordinating the collecting of the data. We thank all our STAR partners for their data collection, identification and input. We are grateful to our colleagues at Wageningen University and Research M. van den Hoorn for technical assistance in processing the data, and to P. Goedhart for the statistical analyses. We also thank John Murray-Bligh and an anonymous reviewer for their useful comments.

References

Battarbee, R. W., R. J. Flower, S. Juggins, S. T. Patrick & A. C. Stevenson, 1997. The relationship between diatoms and surface water quality in the Hoylandet area of Nord-Trondelag, Norway. Hydrobiologia 348: 69–80.

Cazoubon, A., 1996. Algal epiphytes, a methodological problem in river monitoring. In Whitton, B. A. & E. Rott (eds), Use of Algae for Monitoring Rivers II. Innsbruck, Institut für Botanik, Universität Innsbruck, 47–50.

Coring, E., 1999. Situation and developments of algal (diatom)-based techniques for monitoring rivers in Germany. In Prygiel, J., B.A. Whitton & J. Bukowska (eds), Use of Algae for Monitoring Rivers III. Agence de l'Eau Artois-Picardie, Douai, 122–127.

Coste, M., 1987. Etude des méthods biologique quantitatives d'appréciation de la qualité des eaux. Rapport Division Qualité des Eaux Lyon. Agence de l'Eau Rhône, 28.

Dell'Uomo, A., 1996. Assessment of water quality of an Appanine river as a pilot study for diatom-based monitoring of Italian water courses. In Whitton, B. A. & E. Rott (eds), Use of Algae for Monitoring Rivers II. Innsbruck, Institut für Botanik, Universität Innsbruck, 64–72.

Denys, L., 1991a. A check-list of the diatoms in the holocene deposits of the Western Belgian coastal plain with a survey of their apparent ecological requirements. I. Introduction, ecological code and complete list. Ministère des Affaires Economiques – Service Géologique de Belgique.

Denys, L., 1991b. A check-list of the diatoms in the holocene deposits of the Western Belgian coastal plain with a survey of their apparent ecological requirements. II. Centrales. Ministère des Affaires Economiques – Service Géologique de Belgique.

Descy, J.-P., 1979. A new approach to water quality estimation using diatoms. Nova Hedwigia 64: 305–323.

Descy, J.-P. & M. Coste, 1991. A test of methods for assessing water quality based on diatoms. Verhandlung Internationale Vereingung de Limnologie 24: 2112–2116.

Ector, L. & F. Rimet, 2005. Using bioindicators to assess rivers in Europe: an overview. In Lek, S., M. Scardi, P. F. M. Verdonschot, J. -P. Descy & Y. -S. Park (eds), Modelling Community Structure in Freshwater Ecosystems. Springer: 7–19.

Furse, M., D. Hering, O. Moog, P. Verdonschot, R. K. Johnson, K. Barbec, K. Gritzalis, A. Buffagni, P. Pinto, N. Friberg, J. Murray-Bligh, J. Kokes, R. Alber, P. Usseglio-Polatera, P. Haase, R. Sweeting, B. Bis, K. Szoszkiewicz, H. Soszka, G. Springe, F. Sporka & I. Krno, 2006. The STAR project: context, objectives and approaches. Hydrobiologia 566: 3–29.

Gomez, N., 1998. Use of epipelic diatoms for evaluation of water quality in the Matanza-Riachuelo (Argentina), a pampean plain river. Water Research 32: 2029–2034.

Gomez, N. & M. Licursi, 2001. The Pampean Diatom Index (IDP) for assessment of rivers and streams in Argentina. Aquatic Ecology 35: 173–181.

Kelly, M. G., 1998. Use of the trophic diatom index to monitor eutrophication in rivers. Water Research 32: 236–242.

Kelly, M. G., A. Cazaubon, E. Coring, A. Dell'Uomo, L. Ector, B. Goldsmith, H. Guasch, J. Hürlimann, A. Jarlman, B. Kawecka, J. Kwandrans, R. Laugaste, E.-A. Lindstrøm, M. Leitao, P. Zarvan, J. Padisák, E. Pipp, J. Prygiel, E. Rott, S. Sabater, H. van Dam & J. Vizinet, 1998. Recommendations for the routine sampling of diatoms for water quality assessments in Europe. Journal of Applied Phycology 10: 215–224.

Kelly, M. G. & B. A. Whitton, 1995. The Trophic Diatom Index: a new index for monitoring eutrophication in rivers. Journal of Applied Phycology 7: 433–444.

Leclercq, L. & B. Maquet, 1987. Deux noveaux metrics diatomique et the qualite chimique des eaux courantes. Comparison avec differents metrics existants. Cahiers de Biologie Marine 28: 303–310.

Lecointe, C., M. Coste & J. Prygiel, 2003. Omnidia 3.2. Diatom Index Software Including Diatom Database with Taxonomic Names, References and Codes of 11645 Diatom Taxa.

MVSP: Multi-Variate Statistical Package, 1986–2001. Kovach Computing Services, Pentraeth, Wales, UK.

Patterson, H. D. & R. Thompson, 1971. Recovery of interblock information when block sizes are unequal. Biometrika 58: 545–554.

Prygiel, J. & M. Coste, 1999. Progress in the use of diatoms for monitoring rivers in France. In Prygiel, J., B. A. Whitton & J. Bukowska (eds), Use of Algae for Monitoring Rivers III. Douai, Agence de l'Eau Artois-Picardie: 165–179.

Rolland, T., S. Fayolle, A. Cazaubon & S. Pagnetti, 1997. Methodical approach to distribution of epilithic and drifting algae communities in a French subalpine river: inferences on water quality assessment. Aquatic Sciences 59: 57–73.

Rott, E., H. C. Duthie & E. Pipp, 1998. Monitoring organic pollution and eutrophication in the Grand River, Ontario, by means of diatoms. Canadian Journal of Fisheries and Aquatic Sciences 55: 1443–1453.

Rott, E., P. Pfister, H. Van Dam, E. Pipp, K. Pall, N. Binder & K. Ortler, 1999. Indikationslisten für aufwuchsalgen. Wien. Bundesministerium für Land - und Forstwirtschaft 248.

Sládeček, V., 1986. Diatoms as indicators of organic pollution. Acta Hydrochimica et Hydrobiologica 14(5): 555–566.

Snoeijs, P. J. M., 1991. Monitoring pollution effects by diatom community composition: a comparison of sampling methods. Archiv für Hydrobiologie 121: 497–510.

Steinberg, C. & S. Schiefele, 1988. Biological indication of trophy and pollution of running waters. Zeitschrift für Wasser- und Abwasser-Forschung 21: 227–234.

Stevenson, R. J. & S. Hashim, 1989. Variation in diatom community structure among habitats in sandy streams. Journal of Phycology 25: 678–686.

Ter Braak, C. J. F., 1988. CANOCO – A FORTRAN Program for canonical community ordination by [partial] [detrended] [canonical] correspondence analysis, principal component analysis and redundancy analysis (version 2.1). Report LWA-88-02. Agricultural Mathematics Group, Wageningen.

Ter Braak, C. J. F. & P. Ŝmilauer, 2002. CANOCO Reference Manual and Users Guide to Canoco for Windows. Software for Canonical Community Ordination (version 4.5). Centre for Biometry, Wageningen, the Netherlands.

Van Dam, H., 1997. Partial recovery of moorland pools from acidification: indications by chemistry and diatoms. Netherlands Journal of Aquatic Ecology 30: 203–218.

Verdonschot, P. F. M. & C. J. F. Ter Braak, 1994. An experimental manipulation of oligochaete communities in mesocosms treated with chlorpyrifos or nutrient additions: multivariate analyses with Monte Carlo permutation tests. Hydrobiologia 278: 251–266.

Winter, J. G. & H. C. Duthie, 2000. Stream epilithic, epipelic and epiphytic diatoms: habitat fidelity and use in biomonitoring. Aquatic Ecology 34: 345–353.