Diatom reference communities in Québec (Canada) streams based on Kohonen self-organizing maps and multivariate analyses

Martine Grenier, Stéphane Campeau, Isabelle Lavoie, Young Seuk Park, and Sovan Lek

Abstract: The identification of biological reference conditions specific to each type of water body is essential for the development of sound biological indicators and criteria. The purpose of the present study was to establish the reference conditions of each stream type sampled in southern Québec (Canada) using benthic diatoms and environmental variables characterizing streams and watersheds. First, stream reaches were classified as a function of their natural watershed and habitat characteristics. Second, diatom communities were classified based solely on taxa abundance data. Resulting groups were graphically presented on ordinations to interpret, a posteriori, the environmental gradients associated with diatom groups and to identify the diatom communities representing the reference conditions of each of the stream reach groups. A final classification based solely on diatom reference communities found pH and conductivity to be the main discriminating factors, regardless of ecoregion and stream type. Although a specific diatom reference community may be identified for each stream group, our results suggest that many of these communities exhibit strong similarities. Only two reference communities may therefore be used, one for circumneutral conditions and one for alkaline conditions. These reference communities represent the baseline for biocriteria development.

Résumé : L'identification des conditions de référence biologiques spécifiques à chaque type de masse d'eau est essentielle pour le développement de bioindicateurs et de critères biologiques rigoureux. L'objectif de cette étude est d'identifier les conditions de référence spécifiques à chaque type de cours d'eau à partir des communautés de diatomées benthiques et des variables environnementales caractérisant les rivières et les bassins versants du Québec méridional (Canada). Dans un premier temps, les tronçons de rivières ont été classifiés en fonction des caractéristiques naturelles des bassins versants et des habitats lotiques. Dans un deuxième temps, les communautés de diatomées ont été classifiées d'après uniquement les données d'abondance, sans recours aux variables environnementales. Les groupes qui en résultent ont été représentés graphiquement sur des ordinations afin d'être en mesure d'interpréter, a posteriori, les gradients environnementaux influençant les diatomées et d'identifier les communautés qui représentent les conditions de référence de chacun des groupes de rivières. Finalement, une classification basée uniquement sur les communautés de référence a démontré que le pH et la conductivité sont les facteurs les plus discriminants, quels que soient l'écorégion et le type de cours d'eau. Bien qu'une communauté de référence puisse être identifiée pour chaque groupe de tronçons, nos résultats suggèrent que plusieurs de ces communautés sont similaires. Seulement deux communautés de référence ont par conséquent été établies, une pour les conditions circumneutres et l'autre pour les conditions alcalines. Ces communautés peuvent être considérées comme des références pour l'établissement de biocritères.

Introduction

The evaluation of aquatic ecosystems health should be realised by comparing the observed ecological conditions with the expected conditions, in the absence of human disturbance. These reference conditions can be derived from historical data (e.g., Nijboer et al. 2004), regional reference sites (e.g., Gosselain et al. 2005), prediction models (e.g., Wright et al. 1998), paleolimnological data (e.g., Simpson et al. 2005), or expert judgement. According to the US Environmental Protection Agency (Gibson et al. 1996) and the European Water Framework Directive (European Parliament

Received 14 October 2005. Accepted 18 May 2006. Published on the NRC Research Press Web site at http://cjfas.nrc.ca on 3 September 2006.
J18938

M. Grenier and S. Campeau.[1] Université du Québec à Trois-Rivières, 3351 Boul. des Forges, Trois-Rivières, QC G9A 5H7, Canada.
I. Lavoie. Trent University, 1600 West Bank Drive, Peterborough, ON K9J 7B8, Canada.
Y.S. Park[2] **and S. Lek.** Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France.

[1] Corresponding author (e-mail: stephane.campeau@uqtr.ca).
[2] Present address: Kyung Hee University, Dongdaemun-gu, Seoul 130-701, Republic of Korea.

Can. J. Fish. Aquat. Sci. 63: 2087–2106 (2006) doi:10.1139/F06-101

2000), regional reference sites provide the most realistic basis and are the most common approach used for the establishment of reference conditions. The use of a regional-based reference is preferred over individual site-specific references, because it allows for a broader application in water resource programs.

Ecoregions are usually the preferred classification scheme for establishing regional site classes. The ecoregion concept recognizes geographic patterns of similarity among ecosystems, grouped on the basis of environmental variables such as climate, soil type, physiography, and vegetation (Omernik 1987). However, the ecoregion scheme may be refined to define biological potential and identify reference conditions within ecoregions and subecoregions. A classification based on stream attributes may be useful to provide an ecological basis for identifying homogeneous areas from which reference conditions can be established.

Contrary to the regional reference site approach, prediction models usually make no a priori assumptions about the similarity of biological communities at different sites. The reference sites are classified using clustering methods based solely on the similarity of their species composition (Reynoldson et al. 1997). A test site is then matched to the appropriate reference group using, for example, a discriminant function analysis. Such models were used for the development of bioassessment tools such as RIVPACS (river invertebrate prediction and classification system; Wright et al. 1993), AusRivAS (Australian river assessment scheme; Parsons and Norris 1996), and BEAST (benthic assessment of sediment; Reynoldson et al. 1995).

The reference condition approach is being increasingly integrated into many bioassessment programs and is entrenched in the regulatory structure of environmental policies, such as the Clean Water Act (CWA) in the USA and the European Water Framework Directive (WFD). In USA, the need to identify reference conditions has led to a number of studies concerning periphyton communities (e.g., Pan et al. 2000), benthic invertebrate fauna (e.g., Barbour et al. 1999), and fish fauna (e.g., Baker et al. 2005). In Europe, the WFD has initiated a number of studies establishing the reference conditions for phytoplankton (e.g., Lepisto et al. 2004), macrophytes (e.g., Schaumburg et al. 2004*a*, 2004*b*; Meilinger et al. 2005), phytobenthos (e.g., Foerster et al. 2004; Gosselain et al. 2005; Tison et al. 2005), benthic invertebrate fauna (e.g., Rawer-Jost et al. 2004), and fish fauna (e.g., Oberdorff et al. 2001; Carrel 2002).

In Canada, benthic macroinvertebrates are the most widely used group of aquatic organisms in bioassessment. The reference condition approach has mainly been used in Ontario (e.g., Reynoldson et al. 1995; Linke et al. 1999; Winter et al. 2003), British Columbia (e.g., Reynoldson et al. 1997, 2001), and the Yukon (Bailey et al. 1998). The reference condition approach has also been used for fish communities, although to a lesser extent (e.g., Tonn et al. 2003). In the case of algal communities, extensive work has been conducted in Canada regarding the reference conditions of lake diatom assemblages derived from paleolimnological investigations (e.g., Smol et al. 1998). However, for lotic ecosystems, few studies have used diatoms and other periphytic algae as indicators of ecological integrity. Most of these studies were conducted in Ontario (e.g., Winter and Duthie 2000*a*, 2000*b*; Winter et al. 2003) and Québec (Wunsam et al. 2002; Campeau et al. 2005; Lavoie et al. 2006).

The purpose of the present study was to establish the regional reference conditions of each stream type sampled in southern Québec (Canada) using benthic diatoms and environmental variables characterizing streams and watersheds. First, stream reaches were classified as a function of their natural watershed and habitat characteristics using Kohonen self-organizing maps (SOM). Second, diatom communities were classified based solely on taxa abundance data. Resulting groups were graphically presented on ordinations to interpret, a posteriori, the environmental gradients associated with diatom groups and to identify the diatom communities representing the reference conditions of each of the stream reach groups. These reference communities represent the baseline for biocriteria development. They may also be considered as target communities in the context of stream and river restoration.

Material and methods

Study area

The study area is distributed within three ecoregions: the Canadian Shield, the St. Lawrence Lowlands, and the Appalachians (Fig. 1). These ecoregions were further subdivided into natural provinces by Li and Ducruc (2000). The southern part of the Canadian Shield is underlain by acidic igneous and metamorphic rocks (granite, gneiss, migmatite, etc.) covered by noncalcareous glacial tills low in clay-sized particles (Vincent 1989). The Canadian Shield contains an intricate hydrological network of lakes, peat bogs, marshes, beaver ponds, rivers, and streams. Its catchments are mostly covered by boreal forest with humo-ferric podzol soils (Clayton et al. 1978). The southern part of the Shield, however, overlaps the transition zone of mixed and boreal coniferous forests. The streams sampled in the Canadian Shield were low in nutrients, conductivity, and suspended solids and exhibited circumneutral pH (Table 1). These catchments are considered to be the least disturbed. However, some lake outlets may occasionally have higher nutrient levels.

The St. Lawrence Lowlands is a low-lying region with Paleozoic carbonate bedrock, overlain by glacial sediments and postglacial marine clays, and characterized by fertile soils. The lowlands encompass three natural provinces, which differ in geology, geomorphology, and land use (Fig. 1). The St. Lawrence Lowlands is the most heavily populated area of Québec and is characterized by intensive farmlands and large industrial centers. The streams sampled in this ecoregion were high in nutrients, conductivity, suspended solids, and pH (Table 1). These catchments exhibited a gradient from slightly impacted to very impacted streams with most of the latter being located in the Upper St. Lawrence Plain.

Located in southeastern Canada, the Appalachian Mountains are a geologically complex region with low and rounded relief. The rocks of this range are sedimentary, dating back to the Paleozoic era, and are covered by glacial tills. This region is also impacted by farming, but to a lesser extent. The streams sampled in this region had intermediate levels of nutrients, conductivity, and suspended solids. The Appalachian Complex of the Lower St. Lawrence is one of two natural provinces in the Appalachian region and exhibits

Fig. 1. Sampling locations in the St. Lawrence River basin (Québec, Canada). The Appalachians and the St. Lawrence Lowlands were subdivided into natural regions following Li and Ducruc (2000).

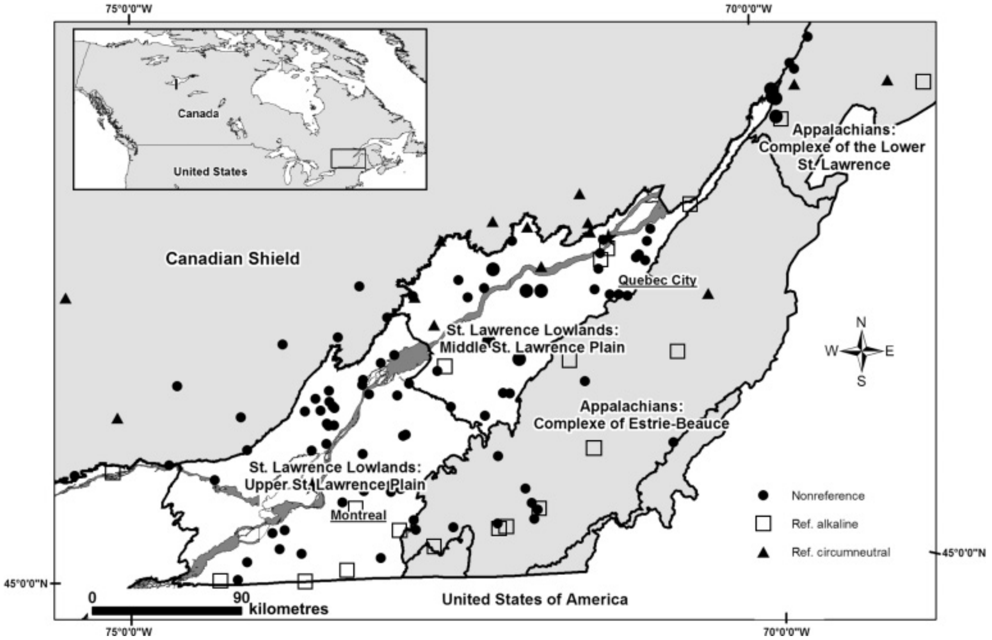


Table 1. Median values (M) and first and third quartiles (Q1 and Q3, respectively) for water chemistry variables arranged according to ecoregions of the St. Lawrence River basin (Québec, Canada).

			Canadian Shield			St. Lawrence Lowlands			Appalachians		
Code	Description		Q1	M	Q3	Q1	M	Q3	Q1	M	Q3
TP	Total phosphorus (mg·L^{-1} P)	Log	0.015	0.017	0.020	0.026	0.051	0.131	0.022	0.024	0.036
SRP	Soluble phosphorus (mg·L^{-1} P)	Log	0.010	0.010	0.011	0.011	0.023	0.057	0.010	0.010	0.012
TN	Total nitrogen (mg·L^{-1} N)	Log	0.182	0.210	0.258	0.330	0.633	1.353	0.275	0.410	0.560
NO_3	Nitrates-nitrites (mg·L^{-1} N)	Log	0.038	0.052	0.083	0.126	0.315	0.877	0.095	0.161	0.303
NH_3	Ammonia (mg·L^{-1} N)	Log	0.020	0.020	0.025	0.025	0.036	0.057	0.022	0.025	0.050
CHL	Chlorophyll *a* (mg·m^{-3})	Log	1.9	2.5	2.8	3.4	7.3	13.9	2.8	4.2	7.9
PH	pH	—	7.1	7.3	7.3	7.8	8.1	8.4	7.8	8.0	8.2
CON	Conductivity (μS·cm^{-1})	SQR2	29	38	70	160	273	393	136	163	233
T	Water temperature (°C)	—	19.2	20.2	20.9	19.7	21.5	22.5	19.8	21.2	22.2
O_2	Dissolved oxygen (mg·L^{-1})	—	8.7	9.1	9.7	8.2	9.1	10.0	8.7	9.2	9.9
TUR	Turbidity (NTU)	Log	0.8	1.3	2.0	2.5	5.6	10.7	1.4	2.7	4.1
SS	Suspended solids (mg·L^{-1})	Log	2.0	2.4	3.4	4.2	7.3	16.4	2.3	3.3	5.5
FC	Coliforms (UFC·100·mL^{-1})	Log	23	44	86	122	277	1024	56	123	215
DOC	Dissolved organic carbon (mg·L^{-1})	Log	4.0	4.6	5.2	4.7	6.0	7.6	3.9	5.7	6.7

Note: Log, log-transformed; SQR2, square root transformed; NTU, nephelometric turbidity units.

the highest concentrations of dissolved organic carbon due to the presence of wetlands in some of its catchments.

Many streams originate in the Canadian Shield or the Appalachians and flow downstream through the St. Lawrence Lowlands. As a result, the water chemistry of some streams flowing through the lowlands reflects the characteristics of upstream ecoregions. This is especially true for the large rivers of the northern shore flowing through the middle St. Lawrence Plain and the Ottawa Plain.

To account for interannual variability within diatom communities, sampling was conducted in the fall (September) of 2002 and 2003. These samples were collected at 126 sampling locations distributed along 32 rivers and streams in the St. Lawrence River basin. A total of 204 diatom samples were collected and analysed, 111 samples in 2002 (coded as "B") and 93 in 2003 (coded as "D"). The sampling locations are part of the water quality monitoring network of the Québec Ministry of the Environment. These sites were selected according to the availability of physico-chemical data and on the basis of land use information with the aim of sampling across a broad gradient of ecoregions and pollution levels. In 2003, the Québec Ministry of the Environment removed a number of sites from its monitoring network, and although new sites were added, the result was a lower number of sites sampled in 2003.

Water chemistry

Water analyses were performed by the Ministry of the Environment (Québec Government) as part of a water quality monitoring program that began in the 1970s. Most water samples were collected every 4 weeks. The following parameters were considered in this study: total phosphorus (TP), soluble reactive phosphorus (SRP), total nitrogen (TN), nitrate/nitrite nitrogen (NO_3), ammonia nitrogen (NH_3), chlorophyll *a* (CHL), pH, conductivity (CON), temperature (T), dissolved oxygen (O_2), turbidity (TUR), suspended solids (SS), faecal coliforms (FC), and dissolved organic carbon (DOC). Some water chemistry data were transformed to improve normality (Table 1). Because diatoms are known to integrate stream water chemistry through time, seasonal averages were used in the analyses. Averages of stream water chemistry were calculated from the six measurements taken in August and September over a 3-year period, including the year in which the diatoms were sampled. These 3-year seasonal averages explain more variance in diatom community structure than one-time chemistry measurements (Campeau et al. 2005).

Stream habitat and watershed characteristics

Stream reach embankment, width, and morphology, current velocity, water transparency, water level, substrate type, and riparian zone characteristics were evaluated at each site (Appendix A). A geographic information system (ArcGIS, version 8; ESRI, 1999) was used to determine watershed characteristics upstream of each sampling sites, such as watershed area, distance to source, geology, surficial deposits, land use, cropped area, animal units, population, and ecoregions (Appendix A). The following data were made available by several governmental agencies: digital maps were provided by the Québec Ministry of the Environment, geologic maps were provided by the Québec Ministry of Natural Resources, classified Landsat images were supplied by the Québec Ministry of Agriculture and Fisheries, and census data were contributed by Statistics Canada. Supplementary information from the United States Geological Survey (USGS) was used for the characterisation of watersheds located beyond the Canadian border.

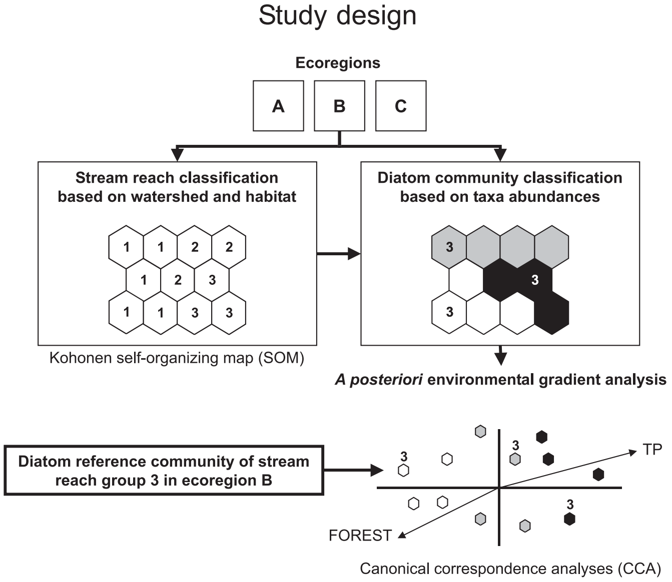
Diatom data

Benthic diatoms were scraped from the top surface of five rocks (composite sample) from riffles in unshaded areas where possible. Samples were collected within a ~5 m^2 area at depths varying from 20 cm to 50 cm, depending on water level and turbidity. The samples were preserved with Lugol's iodine and stored until they were processed. The samples were digested using hydrogen peroxide and mounted onto microscope slides using Naphrax. A minimum of 400 valves (Prygiel and Coste 1993) per slide were counted and identified at 1250× to the most precise possible taxonomic level under a Zeiss Axioskop II microscope with differential interference contrast imaging (DIC). Taxonomic identifications generally followed Krammer and Lange-Bertalot (1986, 1988, 1991*a*, 1991*b*), Reavie and Smol (1998), Fallu et al. (2000), Krammer (2000, 2002, 2003), and Lange-Bertalot (2001).

Stream reach classification and identification of reference diatom communities

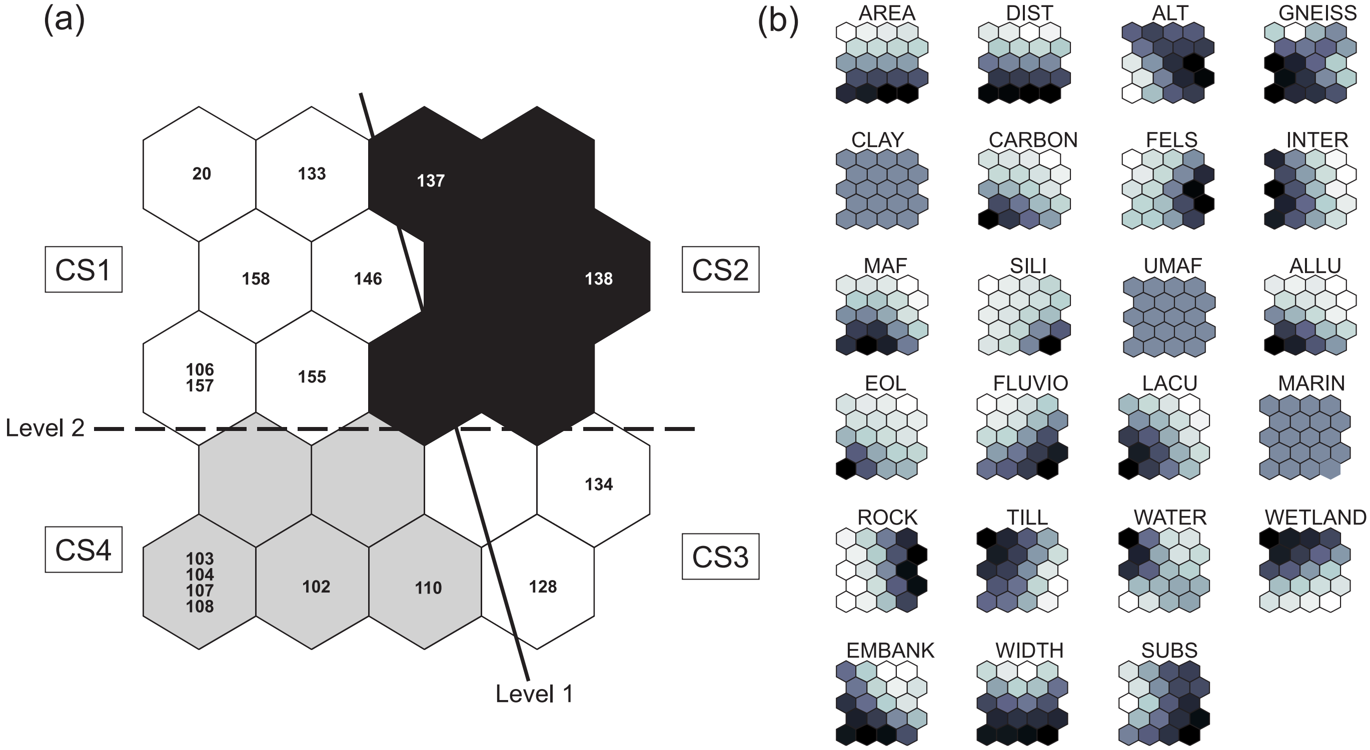
Diatom communities in natural environments are primarily influenced by water physico-chemistry (e.g., pH, CON, DOC) reflecting watershed attributes, such as geology, surficial deposits, and wetlands. Habitat characteristics, including substrate, shade, and current velocity, also play an important role in the structuring of diatom communities. Stream reaches in their natural conditions and with similar habitat and watershed characteristics should, therefore, have comparable benthic diatom communities. The first step in the determination of reference diatom communities was to classify stream reaches according to their habitat and watershed characteristics, excluding variables influenced by anthropogenic factors (Fig. 2). The second step consisted of the classification of diatom communities based solely on taxa abundance data. Finally, resulting groups were graphically presented on ordinations to interpret, a posteriori, the environmental gradients associated with each diatom group and to identify the diatom communities representing the reference conditions for each of the stream reach groups (Fig. 2). This procedure was performed for each ecoregion to discriminate, for example, between the reference diatom communities associated with circumneutral, acidic streams located on the Canadian Shield and communities representing the least-impacted conditions found in smaller, more alkaline agricultural streams of the St. Lawrence Lowlands.

Fig. 2. Schematic representation of the steps followed to determine the diatom-based reference conditions.



All classifications were performed using a Kohonen SOM combined with Ward's hierarchic cluster analyses. A Kohonen SOM is one of the most well-known neural networks with unsupervised learning rules; it performs a topology-preserving projection of the data space onto a regular two-dimensional space. As for the ordination methods, SOMs reduce the number of dimensions with a minimum loss of information. The data are projected onto a rectangular grid map containing multiple hexagonal cells. Sites with similar characteristics are mapped in the same vicinity. The distance between each hexagonal cell is then represented on a U matrix (Euclidean or Bray–Curtis) (Ultsch 1993). The U matrix calculates the distances between neighbouring units and visually presents them as grey-scaled clusters on the SOMs (Kohonen 2001). A Ward linkage method (Legendre

Fig. 3. (*a*) Kohonen self-organizing map (SOM) showing the four stream reach groups established for the Canadian Shield based on watershed and habitat characteristics and the two differentiation levels derived from the Ward's clustering method (the groups are described in Table 2). The numbers in the hexagonal cells represent the sampling site identification numbers (see Table 2). (*b*) SOM distribution map of environmental variables used to classify the stream reach groups. Dark cells represent high values, whereas light cells represent low values. The codes for the environmental variables are described in Appendix A.



and Legendre 1998), derived from the U matrix, was used to further classify the hexagonal cells in a reduced number of groups. A formula was used to calculate the number of cells needed to map the sites for each ecoregion (Y.-S. Park, personal communication, 2003). Kohonen SOMs were conducted using MATLAB software (MathWorks Inc. 2001) and the SOM toolbox developed by Sovan Lek's team at the Laboratoire Dynamique de la Biodiversité (Paul Sabatier University, Toulouse, France). Further information on SOM theory and its ecological applications can be found in Kohonen (2001) and Park et al. (2003). The SOM procedure has been used in a number of ecological studies and has been proven useful in the classification of aquatic communities (e.g., Giraudel and Lek 2001; Park et al. 2003, 2005) and the identification of reference conditions (Gosselain et al. 2005).

Diatom groups were graphically presented using canonical correspondence analyses (CCAs). The diatom community located at the lower extremity of the alteration gradient was considered to be the reference community. When several reference samples were available, the reference community was established by calculating the average relative abundance of each taxon. Preliminary CCAs were performed using all habitat- and watershed-related variables, excluding physico-chemical variables. Variables with a variance inflation factor (VIF) exceeding 10 were not included in the CCAs as they were highly correlated with other variables. Monte Carlo permutation tests were used to select variables explaining a significant portion of the variance ($p \leq 0.05$). A second series of CCAs was performed using only physico-chemical variables. The same procedure was repeated to eliminate multicollinearity and to select significant variables. A final CCA was conducted using selected physico-chemical variables and watershed-related variables. The selected watershed variables were included in the ordination as passive variables (added post hoc to the ordination by projection). All ordination analyses were performed using CANOCO 4.5 (ter Braak and Smilauer 2002).

Results

Stream reach classification for the Canadian Shield

The 17 sample sites from the Canadian Shield were classified into four groups sharing similar watershed and habitat characteristics (Fig. 3 and Table 2). Canadian Shield watersheds were found to be composed mainly of gneiss–paragneiss rocks covered by till. The stream widths were more than 5 m at all sites. The most important site separation (level 1 on the SOM) discriminated groups 1 and 4 from groups 2 and 3. The sites forming groups 1 and 4 had watersheds that consisted mainly of till and lacustrine deposits underlain by a higher proportion of intermediate rocks. The watersheds of groups 2 and 3 contained primarily fluvioglacial deposits and had a higher proportion of surficial bedrock (partly composed of felsic rocks). The sites forming groups 1 and 2 had smaller watersheds and narrower streambeds. Wetlands also occupied a significant portion of their watersheds. In the watersheds of group 4, alluvium and eolian deposits were underlain by a small proportion of carbonated rocks.

Table 2. Stream reaches of the Canadian Shield: stream reach group description, diatom reference samples (in bold), and most abundant diatom taxa in reference communities (>2%).

River name	BQMA	Sampling sites	Latitude (°N)	Longitude (°W)	pH in reference conditions	Most abundant diatom taxa in reference community (mean abundance >2%)
CS1: Sites located at or near lake outlets						
Rivière des Envies (U)	5030113	20	46.84	–72.54	Circumneutral	1. *Achnanthidium minutissimum*
De la Petite Nation (U)	4040039	106	45.90	–75.09		2. *Tabellaria flocculosa*
Noire (U)	5040139	133	46.93	–72.12		3. *Brachysira microcephala*
L'Assomption (U)	5220017	146	46.29	–73.80		4. *Encyonopsis microcephala*
Maskinongé (U)	5260015	155	46.33	–73.36		5. *Fragilaria crotonensis*
Du Loup (U)	5280019	157	46.43	–72.97		6. *Brachysira procera*
Du Loup (M)	5280020	158	46.60	–73.19		7. *Fragilaria* sp. 1
CS2: Upstream reaches of the St. Charles River						
St. Charles (U)	5090003	137	46.86	–71.36	Circumneutral	1. *Achnanthidium minutissimum*
St. Charles (M)	5090016	138	46.91	–71.37		2. *Diatoma moniliformis*
						3. *Navicula notha*
						4. *Staurosira construens* var. *venter*
						5. *Encyonopsis microcephala*
						6. *Brachysira microcephala*
						7. *Achnanthes minutissima* var. *saprophila*
						8. *Tabellaria flocculosa*
						9. *Fragilaria capucina* form 5
						10. *Fragilaria capucina* form 6
						11. *Gomphonema sphaerophorum*
CS3: Upstream reaches of the St. Maurice River and Jacques-Cartier River						
St. Maurice (U)	5010386	128	47.56	–72.84	Circumneutral	1. *Achnanthidium minutissimum*
Jacques-Cartier (U)	5080004	134	47.07	–71.42		2. *Tabellaria flocculosa*
						3. *Brachysira microcephala*
						4. *Eunotia pectinalis*
						5. *Fragilaria capucina* form 5
						6. *Achnanthes minutissima* var. *saprophila*
						7. *Fragilaria capucina* form 3
CS4: Tributaries of the Ottawa River						
Du Nord (U)	4010010	102	45.91	–74.14	Circumneutral	1. *Achnanthidium minutissimum*
Rouge (D)	4020001	103	45.64	–74.69		2. *Tabellaria flocculosa*
Du Diable (U)	4020103	104	46.07	–74.63		3. *Brachysira microcephala*
Du Lièvre (U)	4060001	107	46.55	–75.50		4. *Fragilaria capucina* form 5
Du Lièvre (D)	4060004	108	45.59	–75.42		5. *Fragilaria capucina* form 6
Gatineau (U)	4080223	110	46.62	–75.92		6. *Fragilaria capucina* form 7

Note: BQMA, Banque de données sur la qualité du milieu aquatique. Positions: D, downstream; M, middle stream; U, upstream.

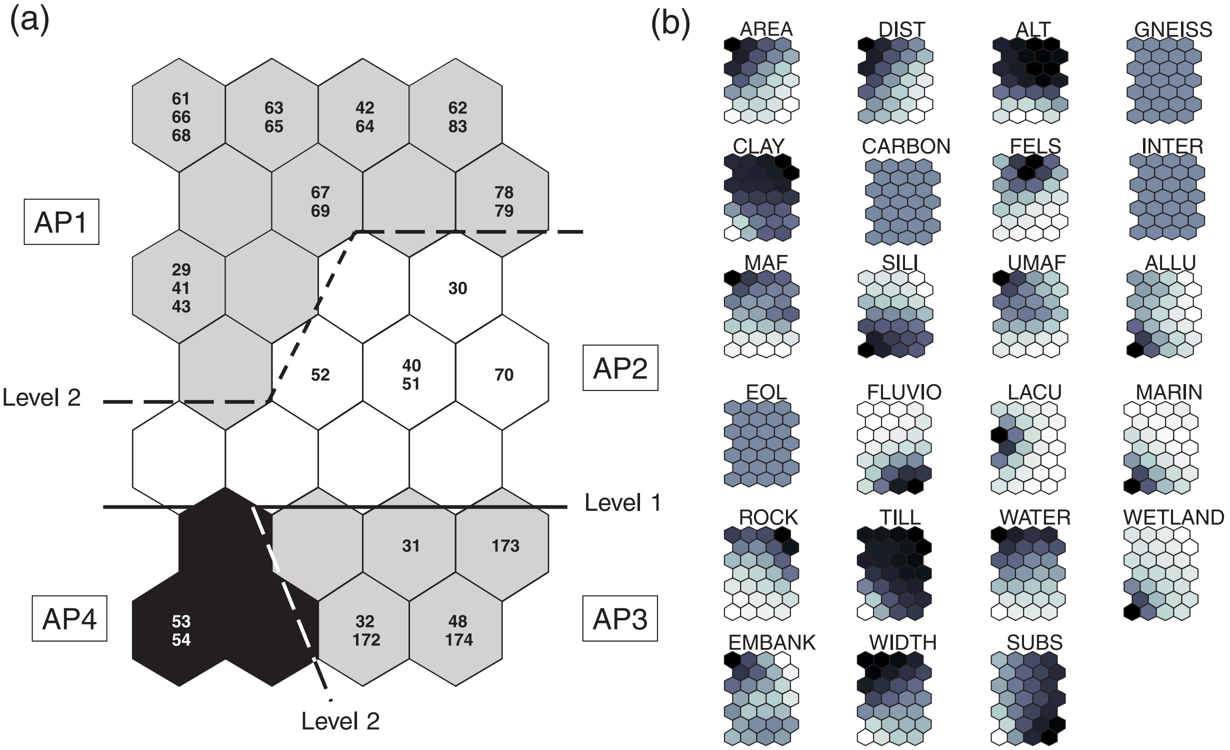
Stream reach classification for the Appalachians

The 29 sample sites located in the Appalachians were divided into four groups with similar watershed and habitat characteristics (Fig. 4 and Table 3). The surficial deposits of most sites located in Appalachian watersheds were dominated by till. The most significant site separation (level 1 on the SOM) discriminated groups 1 and 2 from groups 3 and 4. The sites belonging to groups 1 and 2 had larger watersheds (especially group 1) and wider stream beds and were underlain by a higher proportion of clay, mafic, and ultramafic rocks than those of groups 3 and 4. Lakes also occupied a large portion of the watersheds in groups 1 and 2 (e.g., Lake Memphremagog and Lake Brome). The watersheds of the sites located in the natural region of the Lower St. Lawrence Plain (groups 3 and 4; northeast of the Appalachians) contained mostly siliceous rocks. Sites upstream and downstream of the River Fouquette (group 4), located at the foot of the Appalachians, were distinct from all other groups because of their narrow stream beds (less than 2 m) and very small watershed size (50 km^2). Sites in this group had watersheds containing wetlands and consisted mainly of alluviums and marine deposits underlain by siliceous rocks.

Stream reach classification for the St. Lawrence Lowlands

The 80 sites sampled from the St. Lawrence Lowlands were classified into six groups with similar watershed and habitat characteristics (Fig. 5 and Table 4). Aside from

Fig. 4. (*a*) Kohonen self-organizing map (SOM) showing the four stream reach groups established for the Appalachians based on watershed and habitat characteristics and the two differentiation levels derived from the Ward's clustering method (the groups are described in Table 3). The numbers in the hexagonal cells represent the sampling site identification numbers (see Table 3). (*b*) SOM distribution map of environmental variables used to classify the stream reach groups. Dark cells represent high values, whereas light cells represent low values. The codes for the environmental variables are described in Appendix A.



group 2, most of the watersheds were dominated by marine deposits. The most important site separation (level 1 on the SOM) discriminated groups 1 and 2 from groups 3 to 6. Groups 1 and 2 represented wide rivers with large watersheds mostly located in an upstream ecoregion (Appalachians or Canadian Shield) containing a higher proportion of felsic rock. The watersheds of group 2 were dominated by gneiss–paragneiss rocks that are characteristic of the Canadian Shield. As a result, the water chemistry of some of the rivers that flow through the lowlands reflects the characteristics of an upstream ecoregion. The sites forming group 1 are all located on the Richelieu River. Groups 4 and 5 consisted of small streams of the St. Lawrence Plain watersheds covered mainly by marine deposits. The majority of the sites belonging to groups 3 and 4 were located in the natural region of the Upper St. Lawrence Plain, with watersheds consisting primarily of carbonated rocks. Sites belonging to group 5 were mainly located in the Middle St. Lawrence Plain and had watersheds consisting of siliceous and clay rocks covered by fluvioglacial and till deposits. These watersheds also had a higher proportion of wetlands. Watershed and stream sizes for groups 4 and 5 were generally smaller than those of groups 3 and 6.

Reference diatom communities of the Canadian Shield

The 29 diatom communities sampled on the Canadian Shield during the fall of 2002 and 2003 were classified into seven community types based solely on taxa relative abundances (Fig. 6). An a posteriori CCA was conducted using the seven groups to determine the direction of the environmental gradients influencing the structure of diatom communities on the Canadian Shield. Three outliers (altered communities) were removed from the ordination to improve the identification of reference communities. The physicochemical variables that explained a statistically significant amount of the variation in diatom community structures were CON and CHL. The first four ordination axes summarized 26.7% of the variation observed in diatom communities and explained 61% of the relationship between taxa and selected environmental variables. The eigenvalue for the first axis (λ_1) was 0.31 and for the second axis (λ_2) was 0.24. Forested area and the proportion of alluviums were statistically significant variables characterizing watersheds and habitats and were included in the ordination as passive variables. Based on physico-chemistry, samples representing reference conditions were positioned on the left side of the ordination. The communities representing the most altered sites were located at the foot of the Canadian Shield where there was a substantial proportion of alluvium present (correlated with population and agriculture) in the watershed. Samples from diatom groups 5 and 6 were positioned on the left of the CCA ordination and therefore were used to define the reference communities for each of the Canadian Shield stream reach groups (the diatom reference sites are listed in Table 2, along with each community's most abundant taxa). When several reference samples were available, the reference community was established by calculating the average relative abundance of each taxon. For example, samples from Des Envies River (D20), De La Petite Nation River (B106 and D106), Noire River (B133), and Assomption

Table 3. Stream reach groups of the Appalachians: stream reach group description, diatom reference samples (in bold), and most abundant diatom taxa in reference communities (>2%).

River name	BQMA	Sampling sites	Latitude (°N)	Longitude (°W)	pH in reference conditions	Most abundant diatom taxa in reference community (mean abundance > 2%)
AP1: Upstream reaches located in the Estrie–Beauce Complex						
Madawaska (U)	1170001	29	47.55	–68.64	Alkaline	1. *Achnanthidium minutissimum*
Chaudière (U)	2340004	41	46.18	–70.72		2. *Nitzschia sinuata* var. *tabellaria*
Chaudière (U)	2340006	42	45.69	–70.79		3. *Nitzschia fonticola*
Chaudière (M)	2340014	43	46.50	–71.07		4. *Nitzschia palea* var. *debilis*
St. François (U)	3020035	61	45.48	–71.94		5. *Encyonopsis microcephala*
St. François (U)	3020040	63	45.37	–71.85		6. *Fragilaria capucina* form 6
St. François (M)	3020081	66	45.66	–72.14		7. *A. minutissima* var. *saprophila*
Magog (U)	3020037	62	45.27	–72.10		8. *Fragilaria nanana*
Magog (M)	3020073	65	45.26	–72.16		9. *A.* cf. *latecephalum*
Magog (D)	3020176	68	45.40	–71.90		10. *Staurosira construens*
Au Saumon (U)	3020042	64	45.68	–71.40		11. *Fragilaria crotonensis*
Massawippi (D)	3020082	67	45.36	–71.86		
Coaticook (D)	3020177	69	45.31	–71.88		
Yamaska sud-est (U)	3030041	78	45.18	–72.66		
Yamaska (U)	3030094	79	45.28	–72.51		
Yamaska (U)	3030199	83	45.27	–72.80		
AP2: Upstream reaches with small watersheds located in the Estrie–Beauce Complex						
Cabano (D)	1170022	30	47.58	–68.92	Circumneutral	1. *Achnanthidium minutissimum*
Etchemin (U)	2330010	40	46.49	–70.45		2. *Fragilaria capucina* form 6
Bécancour (U)	2400005	51	46.05	–71.45		3. *Fragilaria capucina* form 5
Bécancour (U)	2400006	52	46.16	–71.56		4. *Encyonopsis microcephala*
Aux Cerises* (D)	3020187	70	45.29	–72.17		5. *Fragilaria capucina* form 7
						6. *A.* cf. *latecephalum*
						7. *Fragilaria capucina* form 4
						8. *Fragilaria capucina* form 3
AP3: Upstream reaches located in the Lower St. Lawrence Complex						
Du Loup (U)	2250002	31	47.58	–69.67	Circumneutral	1. *Achnanthidium minutissimum*
Du Loup (D)	2250005	32	47.84	–69.53		2. *Achnanthidium deflexum*
Dufour (U)	2260004	172	46.55	–71.88		3. *Fragilaria capucina* form 3
St. Denis (U)	2260005	173	46.54	–71.75		4. *Fragilaria capucina* form 6
Aux Perles (U)	2260006	174				3. *Staurosira construens* var. *venter*
Des Îles Brûlées (U)	2340086	48	46.51	–71.15		4. *Brachysira microcephala*
						5. *Fragilaria capucina* form 5
						6. *Fragilaria nanana*
						7. *Fragilaria capucina* form 7
						8. *Encyonopsis microcephala*
						9. *Cymbella delicatula*
AP4: Reaches of the Fouquette Stream						
Fouquette (D)	2E90001	53	47.71	–69.69	Alkaline	1. *Amphora pediculus*
Fouquette* (U)	2E90002	54	47.67	–69.66		2. *Navicula germainii*
						3. *Cocconeis placentula* var. *euglypta*
						4. *Navicula gregaria*
						5. *Achnanthidium minutissimum*
						6. *Rhoicosphenia abbreviata*
						7. *Meridion circulare*
						8. *Navicula cryptocephala*
						9. *Nitzschia sociabilis*
						10. *Navicula minima*

Note: BQMA, Banque de données sur la qualité du milieu aquatique. The asterisk (*) indicates that the reference community identified represents the least-disturbed conditions found and is not a true reference community. Positions: D, downstream; M, middle stream; U, upstream.

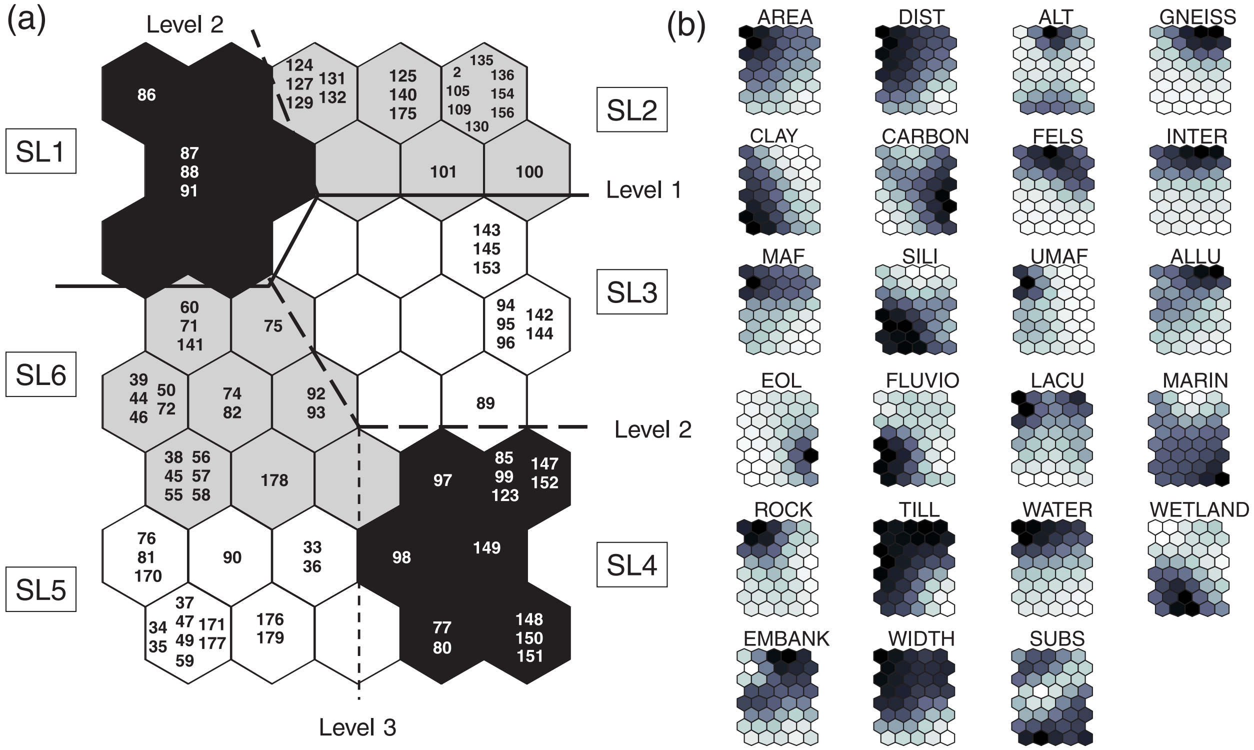


Fig. 5. (*a*) Kohonen self-organizing map (SOM) showing the six stream reach groups established for the St. Lawrence Lowlands based on watershed and habitat characteristics and the three differentiation levels derived from the Ward's clustering method (the groups are described in Table 4). The numbers in the hexagonal cells represent the sampling site identification numbers (see Table 4). (*b*) SOM distribution map of environmental variables used to classify the stream reach groups. Dark cells represent high values, whereas light cells represent low values. The codes for the environmental variables are described in Appendix A.

River (D146) all belong to stream reach group CS1 and were found on the left side of the CCA's first axis (Fig. 6). These samples represent the reference communities for other members of this group, such as the Maskinongé River (upstream) and Du Loup River (upstream). These rivers were located on the lower right portion of the ordination and are therefore more impacted than their reference sites. In the context of future restoration actions, one would expect that the diatom communities of these impacted rivers would change and resemble the reference community presented in Table 2.

Reference diatom communities of the Appalachians

The 50 diatom communities sampled in the Appalachians during the fall of 2002 and 2003 were classified into 10 type communities (Fig. 7). An a posteriori CCA was conducted using these 10 groups to determine the direction of the environmental gradients influencing the structure of diatom communities in the Appalachians. Four outliers (altered communities) were removed from the ordination to improve the identification of reference communities. The physico-chemical variables that explained a statistically significant amount of the variation in Appalachian diatom community structures were TP, CON, CHL, DOC, T, and O_2. The first four ordination axes summarized 16.6% of the variation observed in diatom communities and explained 79.3% of the relationship between taxa and the statistically significant environmental variables. The eigenvalues for the first and second axes were 0.31 and 0.2, respectively. Marine deposits, siliceous and ultramafic rocks, wetlands, and pastures were the watershed and habitat variables that explained a significant portion of the variation in diatom community structures and were included in the ordination as passive variables. The reference samples were positioned on the left portion of the ordination set, opposite to the pollution gradient. The communities representing the most altered sites were located at the foot of the Appalachians (marine deposits), where pasture areas are extensive. The samples from diatom groups 5 to 8 were positioned on the left portion of the CCA ordination and were used to define the reference communities for each stream reach group within the Appalachians (the diatom reference sites are listed in Table 3, along with each community's most abundant taxa).

Reference diatom communities of the St. Lawrence Lowlands

The 125 diatom communities sampled in the St. Lawrence Lowlands during the fall of 2002 and 2003 were classified into 10 type communities (Fig. 8). An a posteriori CCA was conducted using these 10 groups to determine the direction of the environmental gradients influencing the structure of diatom communities in the St. Lawrence Lowlands. Eleven outliers (altered communities) were removed from the ordination to improve the identification of reference communities. The physico-chemical variables that explained a statistically significant amount of the variation in diatom community were CON, pH, T, DOC, VEL, SS, and TUR. The first four ordination axes summarized 16.6% of the variation observed in diatom communities and explained 80.3% of the relationship between taxa and selected environmental variables. The eigenvalues for the first and second axes were

Table 4. Stream reach groups of the St. Lawrence Lowlands: stream reach group description, diatom reference samples (in bold), and most abundant diatom taxa in reference communities (>2%).

River name	BQMA	Sampling sites	Latitude (°N)	Longitude (°W)	pH in reference conditions	Most abundant diatom taxa in reference community (mean abundance > 2%)
SL1: Reaches of the Richelieu River						
Richelieu (D)	3040009	86	46.02	–73.13	Alkaline	1. *Achnanthidium minutissimum*
Richelieu (D)	3040010	87	45.40	–73.25		2. *Cocconeis placentula* var. *euglypta*
Richelieu* (U)	3040012	88	45.06	–73.33		3. *Fragilaria capucina* var. *vaucheriae*
Richelieu (M)	3040017	91	45.69	–73.19		4. *Staurosirella pinnata*
						5. *Nitzschia fonticola*
						6. *Nitzschia palea* var. *debilis*
						7. *Pseudostaurosira binodis*
						8. *Pseudostaurosira brevistriata*
						9. *Cocconeis pediculus*
SL2: Downstream reaches of the St. Lawrence Lowlands with most of their watershed located in the Canadian Shield						
Rivière des Envies (D)	5030114	2	46.62	–72.41	Circumneutral	1. *Achnanthidium minutissimum*
Du Nord (M)	4010002	100	45.56	–74.34		2. *Fragilaria capucina* form 7
Du Nord (D)	4010008	101	45.72	–74.09		3. *A. minutissima* var. *saprophila*
De la Petite Nation (D)	4040001	105	45.61	–75.13		4. *Tabellaria flocculosa*
Gatineau (D)	4080003	109	45.49	–75.75		4. *Fragilaria capucina* form 3
St. Maurice (D)	5010007	124	46.38	–72.61		5. *Navicula notha*
Shawinigan (D)	5010012	125	46.54	–72.77		6. *Fragilaria capucina* form 5
St. Maurice (M)	5010014	127	46.54	–72.75		7. *Brachysira microcephala*
Batiscan (D)	5030001	129	46.53	–72.34		8. *Fragilaria capucina* form 6
St. Anne (D)	5040007	130	46.57	–72.21		9. *Gomphonema manubrium*
St. Anne (U)	5040113	131	46.90	–71.85		10. *Nitzschia palea* var. *debilis*
St. Anne (M)	5040116	132	46.82	–71.97		11. *Adlafia* cf. *bryophila*
Jacques-Cartier (D)	5080006	135	46.68	–71.75		
St. Charles (D)	5090002	136	46.81	–71.26		
L'Assomption (D)	5220001	140	46.04	–73.44		
Maskinongé (D)	5260003	154	46.18	–73.03		
Du Loup (D)	5280001	156	46.24	–72.92		
Blanche (D)	5040006	175	46.03	–72.88		
SL3: Reaches of the Upper St. Lawrence Lowlands with most of their watershed overlying carbonated rocks						
L'Acadie (D)	3040013	89	45.43	–73.35	Alkaline	1. *Nitzschia sinuata* var. *tabellaria*
Chateauguay (M)	3090003	94	45.11	–74.09		2. *Achnanthidium minutissimum*
Chateauguay (U)	3090005	95	45.02	–74.17		3. *A.* cf. *latecephalum*
Trout River (U)	3090009	96	45.01	–74.30		4. *Navicula capitatoradiata*
L'Assomption (M)	5220004	142	45.94	–73.40		5. *Cymbella excisa* var. *procera*
De l'Achigan (D)	5220005	143	45.85	–73.45		6. *Gomphonema entolejum*
St. Esprit (D)	5220006	144	45.86	–73.46		7. *A. minutissima* var. *saprophila*
Ouareau (D)	5220012	145	45.95	–73.41		8. *Melosira varians*
Bayonne (D)	5240001	153	46.09	–73.17		9. *Geissleria decussis*
						10. *Staurosira construens* var. *venter*
SL4: Reaches of small streams of the Upper St. Lawrence Lowlands with part of their watershed overlying carbonated rocks						
Des Hurons (M)	3040007	85	45.49	–73.19	Alkaline	1. *Achnanthidium minutissimum*
Ruisseau Norton (M)	3090046	97	45.16	–73.68		2. *A. minutissima* var. *saprophila*
Des Anglais (U)	3090047	98	45.00	–73.65		3. *Nitzschia fonticola*
Ruisseau St. Louis (D)	3110003	99	45.27	–73.90		4. *Fragilaria capucina* var. *vaucheriae*
Mascouche (D)	4640003	123	45.72	–73.58		5. *Nitzschia palea* var. *debilis*
Ruisseau du Point-du-Jour (D)	5220063	147	45.85	–73.41		6. *Navicula minima*
Ruisseau Vacher (M)	5220239	148	45.93	–73.51		7. *Navicula sp. 10*
Ruisseau St. Pierre (M)	5220240	149	45.98	–73.44		
Ruisseau St. Esprit (M)	5220241	150	45.93	–73.62		

Table 4 (*concluded*).

River name	BQMA	Sampling sites	Latitude (°N)	Longitude (°W)	pH in reference conditions	Most abundant diatom taxa in reference community (mean abundance > 2%)
SL5: Reaches of small streams with most of their watershed located in the middle St. Lawrence Lowlands						
Boyer Sud (D)	2300002	34	46.72	–70.98	Alkaline	1. *Achnanthidium minutissimum*
Boyer Nord (D)	2300003	35	46.70	–71.00		2. *Cocconeis placentula* var. *euglypta*
Ruisseau du Portage (D)	2300004	36	46.79	–70.91		3. *Staurosirella pinnata*
Ruisseau Honfleur (D)	2300005	37	46.69	–70.93		4. *Nitzschia palea* var. *debilis*
Bras d'Henri (U)	2340051	47	46.54	–71.34		5. *Navicula capitatoradiata*
Bras d'Henri (D)	2340099	49	46.51	–71.22		6. *Nitzschia fonticola*
Des Pins (D)	3010038	59	46.00	–72.03		7. *Navicula germainii*
Yamaska Sud-Est* (D)	3030031	76	45.27	–72.92		8. *A. minutissima* var. *saprophila*
Yamaska Nord (D)	3030108	81	45.33	–72.81		9. *Encyonema silesiacum*
Aux Brochets (D)	3040015	90	45.12	–73.07		10. *Navicula cryptocephala*
La Chaloupe (D)	5230001	152	46.07	–73.18		11. *Planothidium lanceolatum*
Aux Perles (D)	2260002	170	46.29	–72.18		12. *Cyclotella meneghiniana*
Goudron (D)	2260003	171	46.18	–71.95		13. *Fragilaria capucina* var. *vaucheriae*
Gentilly* (M)	QC1	176	46.28	–72.18		
Rosaire (D)	QC2	177	46.18	–71.95		
Du Bois Clair* (U)	QC4	179	46.54	–71.75		
SL6: Reaches of large rivers located in the St. Lawrence Lowlands with most of their watershed in the Appalachians						
Chaudière* (D)	2340033	44	46.70	–71.28	Alkaline	1. *Achnanthidium minutissimum*
Beaurivage (D)	2340034	45	46.65	–71.30		2. *Nitzschia palea* var. *debilis*
Bécancour (D)	2400004	50	46.35	–72.44		3. *Nitzschia fonticola*
Nicolet (U)	3010007	55	46.00	–72.09		4. *Cocconeis pediculus*
Nicolet* (D)	3010008	56	46.15	–72.54		5. *A. minutissima* var. *saprophila*
Nicolet Sud-Ouest (D)	3010009	57	46.13	–72.60		6. *Cocconeis placentula* var. *euglypta*
Nicolet Sud-Ouest (U)	3010036	58	45.88	–72.23		7. *Nitzschia sinuata* var. *tabellaria*
St-François (D)	3020031	60	46.07	–72.82		8. *Navicula capitatoradiata*
St-François (D)	3020243	71	45.93	–72.50		9. *Nitzschia palea* var. *debilis*
Noire (M)	3030003	72	45.50	–72.90		
Yamaska (D)	3030023	74	46.00	–72.91		
Yamaska (D)	3030026	75	45.52	–72.98		
Yamaska (M)	3030123	82	45.78	–72.88		
Chateauguay (D)	3090001	92	45.29	–73.80		
Des Anglais (D)	3090002	93	45.18	–73.85		
L'Assomption (D)	5220003	141	45.75	–73.47		
Du Chêne (D)	QC3	178	46.55	–71.87		

Note: BQMA, Banque de données sur la qualité du milieu aquatique. The asterisk (*) indicates that the reference community identified represents the least-disturbed conditions found and is not a true reference community. Positions: D, downstream; M, middle stream; U, upstream.

0.38 and 0.22, respectively. Population, forested area, marine deposits, carbonated rocks, gneiss–paragneiss rocks, felsic rocks, and intermediate rocks were the watershed and habitat variables that explained a significant amount of the variation in diatom community structures and were included in the ordination as passive variables. Based on physicochemistry, samples representing reference conditions were positioned on the right portion of the ordination. The samples from diatom groups 3 to 8 were positioned on the right portion of the CCA ordination and were used to define the reference communities for each stream reach group within the St. Lawrence Lowlands (the diatom reference sites and each community's most abundant taxa are listed in Table 4).

Combined analysis of all diatom reference communities

From the examination of Tables 2 to 4, we observed that reference diatom communities may exhibit strong similarities from one stream reach group to another. For example, all reference communities of the stream reach groups from the Canadian Shield are similarly dominated by *Achnanthidium minutissimum* and *Tabellaria flocculosa*. Moreover, reference communities may be similar from one ecoregion to another. For example, reference communities from the Appalachian stream groups AP2 and AP3 have a taxonomic composition similar to the reference communities of the stream group SL2 in the St. Lawrence Lowlands (*A. minutissimum* and different forms of *Fragilaria capucina*). To test the similarities between reference communities, we conducted a combined analysis of all reference diatom communities from the stream reach groups of the three ecoregions. The SOM classified all of the 14 diatom reference communities into only four communities (Fig. 9). An a posteriori CCA was conducted using all reference diatom communities to determine the direction of the environmental gradients influencing the structure of diatom reference communities (Fig. 10). The physico-chemical variables explaining a significant amount of the variation in diatom community structures were pH and CON. CON is correlated with FC, TN,

Fig. 6. (*a*) Kohonen self-organizing map (SOM) showing the seven diatom communities established for the Canadian Shield based solely on taxa relative abundances and the three differentiation levels derived from the Ward's clustering method. Median values for water chemistry are presented in boxes for each community. The numbers in the hexagonal cells represent the sampling site identification numbers (see Table 2). Sites that were sampled in 2002 are coded as "B" and those sampled in 2003 are coded as "D". (*b*) Canonical correspondence analysis (CCA) sites scores. The diatom SOM groups are represented by symbols. The numbers in parentheses indicate the corresponding stream reach groups (see Fig. 3). The codes for the environmental variables are described in Appendix A (broken arrows indicate passive variables).

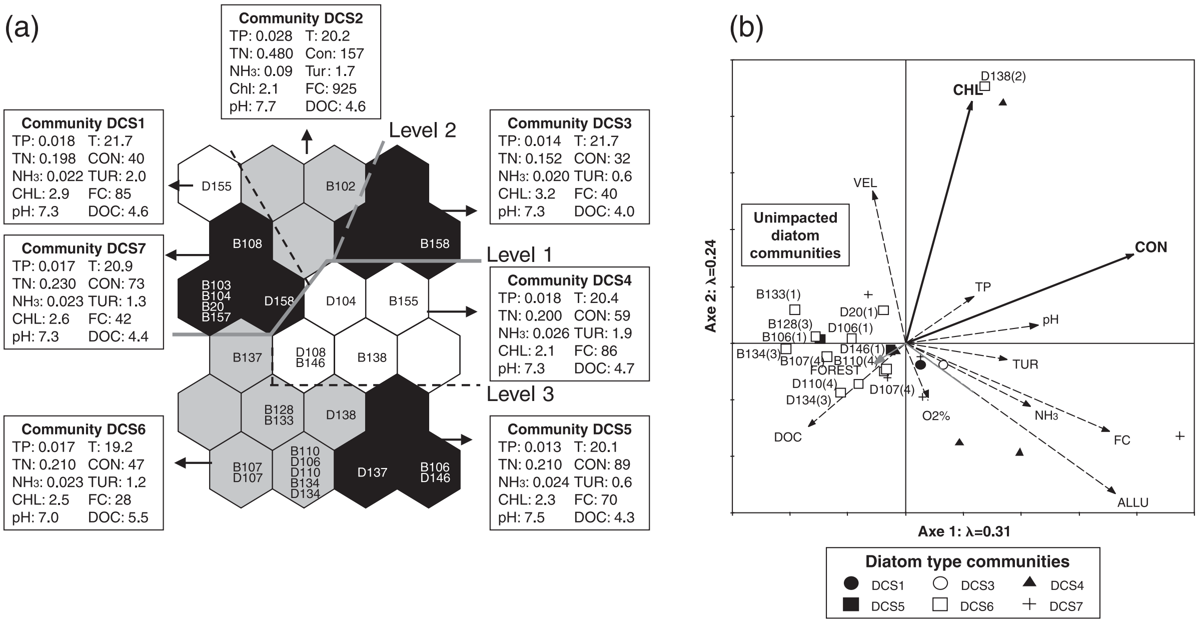


Fig. 7. (*a*) Kohonen self-organizing map (SOM) showing the 10 diatom communities established for the Appalachians based solely on taxa relative abundances and the four differentiation levels derived from the Ward's clustering. Median values for water chemistry are presented in boxes for each community. The numbers in the hexagonal cells represent the sampling site identification numbers (see Table 3). Sites that were sampled in 2002 are coded as "B" and those sampled in 2003 are coded as "D". (*b*) Canonical correspondence analysis (CCA) sites scores. The diatom SOM groups are represented by symbols. The numbers in parentheses indicate the corresponding stream reach groups (see Fig. 4). The codes for the environmental variables are described in Appendix A (broken arrows indicate passive variables).

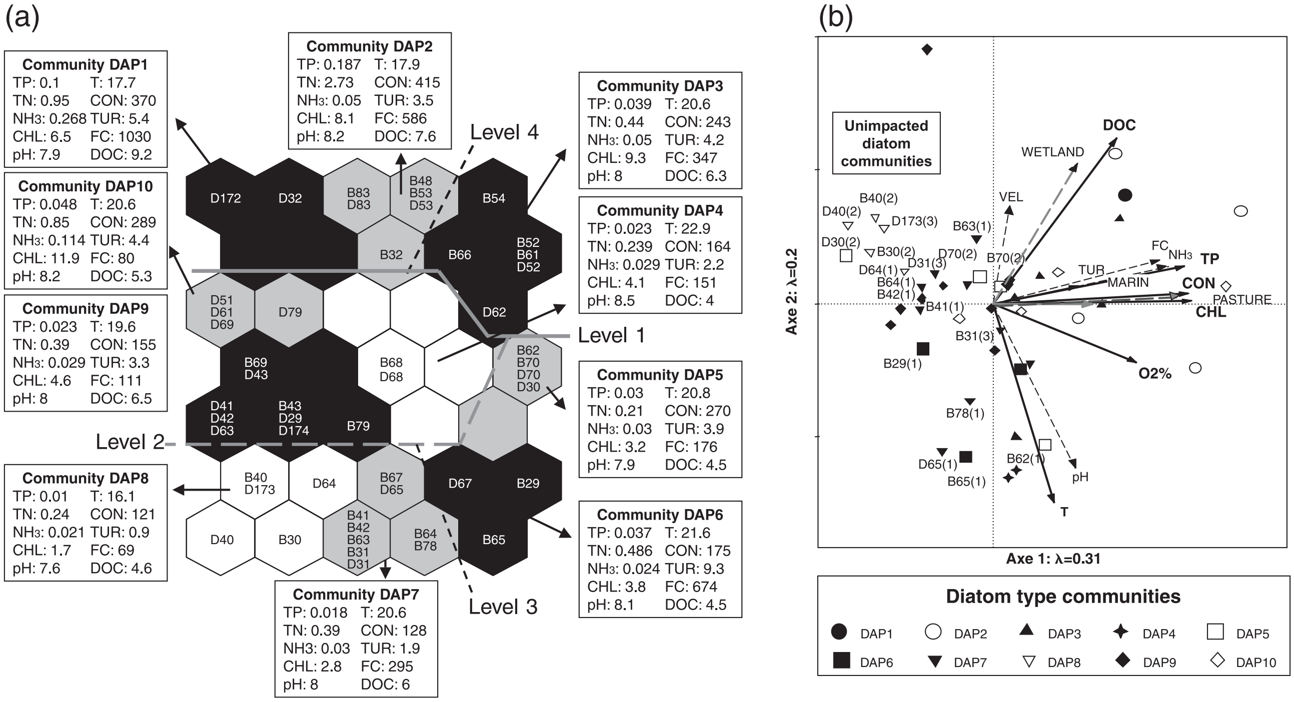


Fig. 8. (*a*) Kohonen self-organizing map (SOM) showing the 10 diatom communities established for the St. Lawrence Lowlands based solely on taxa relative abundances and the four differentiation levels derived from the Ward's clustering method. Median values for water chemistry are presented in boxes for each community. The numbers in the hexagonal cells represent the sampling site identification numbers (see Table 4). Sites that were sampled in 2002 are coded as "B" and those sampled in 2003 are coded as "D". (*b*) Canonical correspondence analysis (CCA) sites scores. The diatom SOM groups are represented by symbols. The numbers in parentheses indicate the corresponding stream reach groups (see Fig. 5). The codes for the environmental variables are described in Appendix A (broken arrows indicate passive variables).

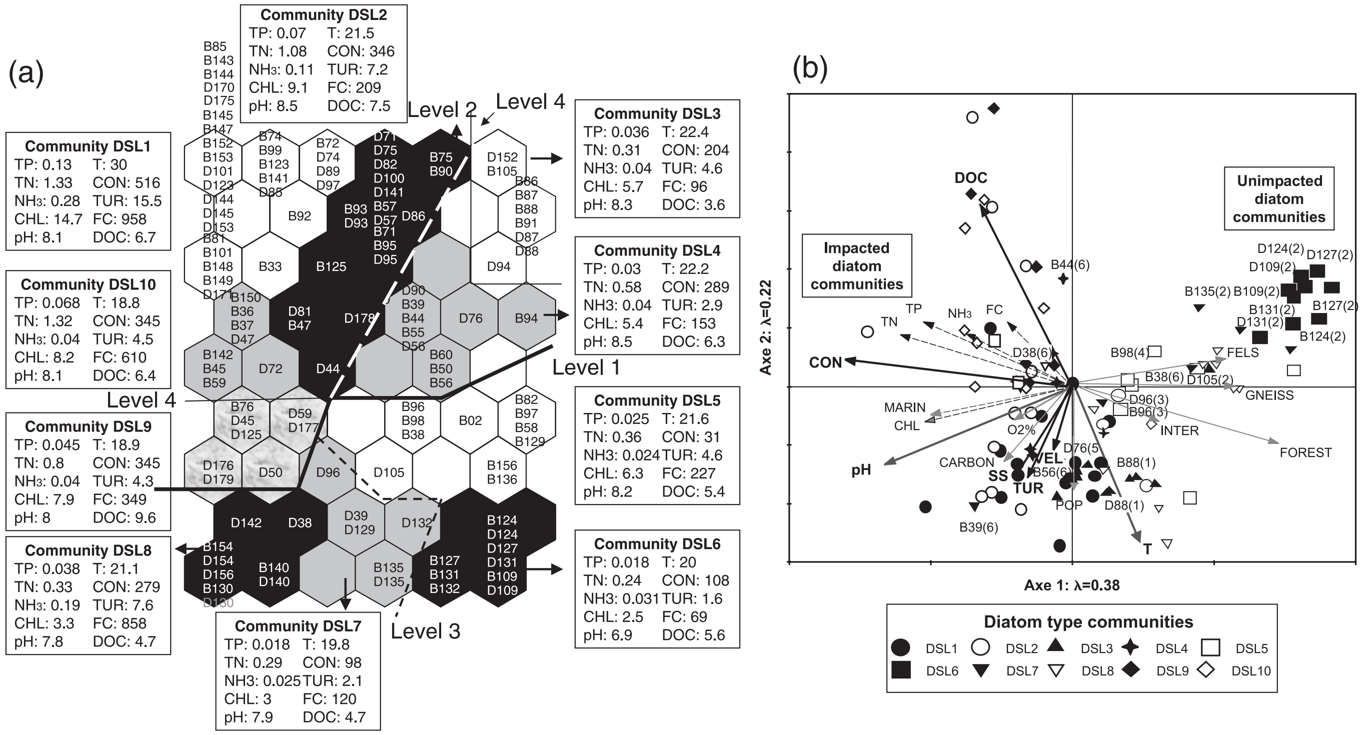


Fig. 9. Kohonen self-organizing map (SOM) showing the combined analysis of all diatom reference communities in the three ecoregions. The SOM classified all the diatom reference communities into four communities. The two differentiation levels derived from the Ward's clustering method are indicated. Median values for water chemistry are presented in boxes for each community. The underlined numbers represent the Canadian Shield sampling sites (Table 2), the numbers in italic represent the Appalachians sampling sites (Table 3) and the regular font represents the St. Lawrence sampling sites (Table 4). Sites that were sampled in 2002 are coded as "B" and those sampled in 2003 are coded as "D".

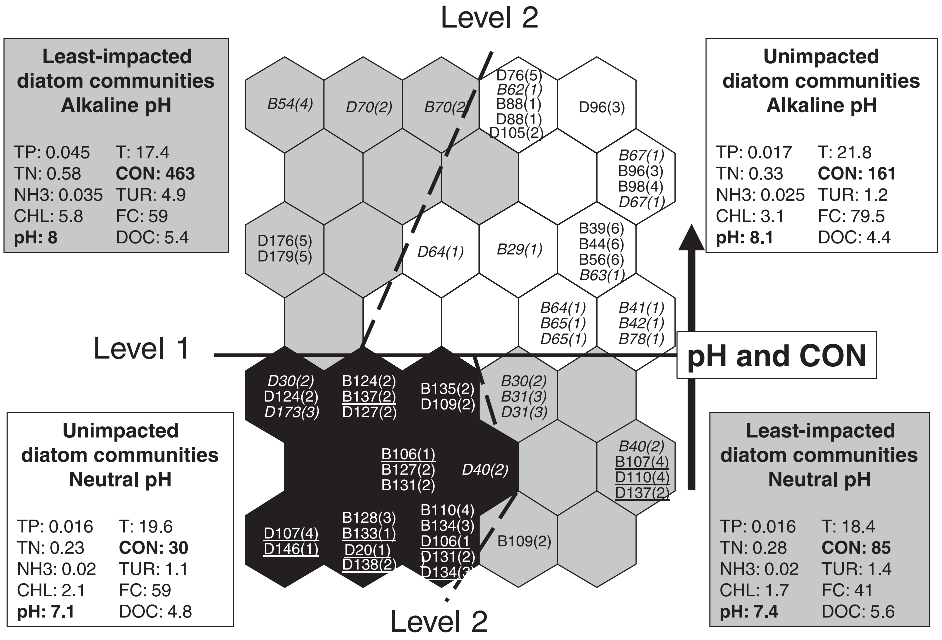
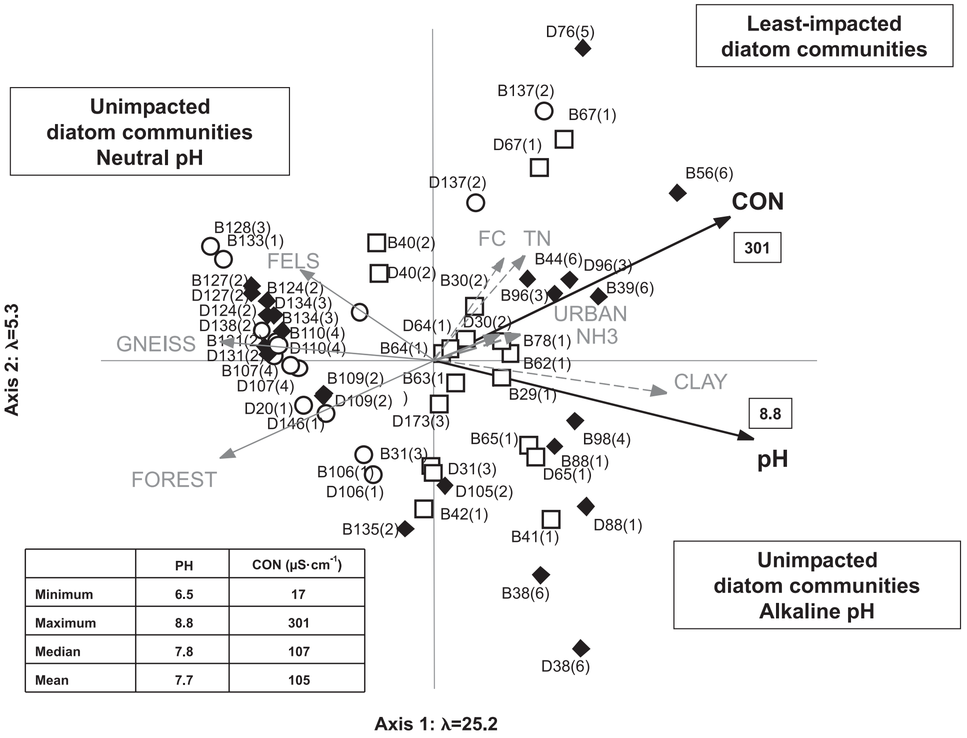


Fig. 10. Canonical correspondence analysis (CCA) sites scores showing the combined analysis of all diatom reference communities in the three ecoregions. The numbers represent the sampling site identification numbers (Tables 2–4). Sites that were sampled in 2002 are coded as "B" and those sampled in 2003 are coded as "D". The numbers in parentheses indicate the corresponding stream reach groups (Tables 2–4). The symbols represent the three ecoregions (circles, Canadian Shield; diamonds, St. Lawrence Lowlands; squares, Appalachians). The codes for the environmental variables are described in Appendix A (broken arrows indicate passive variables).



	PH	CON (μS·cm^{-1})
Minimum	6.5	17
Maximum	8.8	301
Median	7.8	107
Mean	7.7	105

and NH_3, which indicates that conductivity is related to a degradation gradient. These correlated variables were included in the ordination as passive variables. The first four ordination axes summarized 22.6% of the variation observed in diatom communities. The eigenvalue for the first axis (λ_1) was 0.26 and for the second axis (λ_2) was 0.05. Forested area, urban area, and the presence of gneiss–paragneiss and felsic rocks were the significant variables characterizing watersheds and habitats and were included in the ordination as passive variables. The presence of gneiss–paragneiss and felsic rocks is inversely correlated with the presence of clay rocks. This relationship may indicate that natural alkaline pH could be explained by the high proportion of clay rocks in these watersheds. The most important site separation (level 1 on the SOM and axis 1 on the CCA) discriminated reference communities based on pH and conductivity. Samples on the left side of the ordination represent reference communities for the streams with naturally neutral pH and lower conductivity, whereas samples on the right side of the ordination represent reference communities for the streams with naturally alkaline pH and higher conductivity. A pH of 7.65 represents the separation line between both reference communities. Samples on the upper part of the ordination were slightly altered or had a higher conductivity in their natural state, but represent the least-impacted conditions for some stream reach groups of the Appalachians and the St. Lawrence Lowlands. Although a specific diatom reference community may be identified for each stream reach group, these results suggest that many of these communities exhibit strong similarities. As a result, only two reference communities may be used, one for circumneutral conditions and the other for alkaline conditions.

Discussion

Establishing reference conditions: abiotic and biotic approaches

In the reference-condition approach, a test site is compared with an appropriate set of reference sites characterizing the biological condition of a region. Two major analytical approaches for the comparison of test sites with reference conditions have been used so far: abiotic and biotic methods. Abiotic methods classify reference sites based on geographic and physical attributes. Some authors consider watershed variables as being the most useful spatial framework for aquatic ecosystem management (e.g., US Water Environment Federation 1992; Maxwell et al. 1995). However, because climate, geology, soil, and vegetation type are variables that are not specific to a single watershed, most authors consider ecoregions to constitute a superior spatial framework for the determination of water quality standards and restoration goals. Reference sites are chosen from streams with catchments belonging to specific ecoregions or subecoregions (e.g., Barbour et al. 1995, 1996). These regions are predefined using geomorphological characteristics such as climate, physiography, geology, soils, and vegetation (Omernik 1987). Some authors consider the combination of both the ecoregion and the watershed as necessary for the development of a regional reference site network of watersheds with similar reference communities within the same ecoregion (e.g., Hughes 1995; Omernik and Bailey 1997; Rogers and Wasson 1997). The ecoregion scheme has been widely used in multimetric methods to study macroinvertebrate communities. Several studies concluded that significant biotic variation among sites was related to ecoregion, especially where there were marked differences in topography between ecoregions (e.g., Gerritsen et al. 2000). However, although significant, the amount of variation related to landscape features is usually quite low (Hawkins et al. 2000). Local habitat features appear to account for more biotic variation than larger-scale environmental features. In the case of diatom communities, the study of Pan et al. (2000) showed that diatom community structure in reference streams did not vary with either ecoregion or catchment. As pointed out by Hawkins et al. (2000), classification based on both stream reach-level and larger-scale landscape features may provide a better tool for the prediction of aquatic community composition. In France, for example, 22 hydro-ecoregions were visually separated according to natural discontinuities in stream typology (Wasson et al. 2002). The final typology was, however, realized by evaluating biological reference conditions. Their results showed a good correspondence between the 22 hydro-ecoregions, the invertebrate communities, and the diatom communities. Different European countries are currently dividing their territory into “subecoregions” or aquatic “landscape units” (e.g., Austria: Fink et al. 2000). According to Hering et al. (2003), these classifications led to the establishment of about 100 stream types in Europe. Although the method and the variables used may differ from the above studies, our work shares the common approach of determining stream typology before the identification of the reference conditions specific to each typology.

Contrary to abiotic methods, biotic approaches make no a priori assumptions about the similarity of biological communities at different sites. Rather, the reference sites are classified using clustering methods based on the similarity of their species composition (Reynoldson et al. 1997). A method is then required to match a test site to the appropriate reference group. These reference sites can then be used to predict the community structure expected at the test site following restoration. The predictive model may be based, for example, on a discriminant function. Such biotic approaches were used for the development of bioassessment tools such RIVPACS (Wright et al. 1993), AusRivAS (Parsons and Norris 1996), and BEAST (Reynoldson et al. 1995).

Recent studies carried out in Europe (Coste et al. 2004; Descy et al. 2005; Gosselain et al. 2005) also investigated the identification of stream reference conditions based on biota. Reference sites for the French territory were first selected according to the floristic composition of diatom communities. Only the sites that did not experience environmental pressures were considered. Diatom-based index values were then calculated using the specific polluosensitivity index (IPS: Indice de Polluo-sensibilité Spécifique; Coste 1982) and the biological diatom index (IBD: Indice Biologique Diatomées; Lenoir and Coste 1996) for each of the selected reference sites. The index values for the selected reference sites were considered to be in a “good ecological state”. In our opinion, the procedure of reference site selection based on diatom indices (IPS and IBD) is biased because of the development of indices that do not consider ecoregions and stream reach group variables. Reference communities for a specific type of environment may, therefore, not be adequately represented in the diatom index. Furthermore, the IPS and IBD were developed on the basis of the relationships between diatom communities and physicochemistry data, which imply that the definition of the reference communities is dependent on the physico-chemical conditions of the stream. This circular argument disregards the use of nonredundant information in the characterisation of diatom community structure when attempting to supply additional information on the ecosystem’s status. Finally, the identification of environmental pressures on each ecoregion was based on land use analysis. Although this method is certainly valid, there is a possibility that certain sources of degradation may not be detected and that the gradient discriminating between reference, intermediate, and degraded conditions may not be adequately identified.

The method used in our study combines abiotic and biotic classifications. The two-step procedure presented in this study was first used to classify stream reaches as a function of their natural watershed and habitat characteristics, both of which are known to influence diatom communities. This classification was conducted for each of the three ecoregions and identified four stream reach groups on the Canadian Shield, four in the Appalachians, and six in the St. Lawrence Lowlands. In parallel, diatom communities were classified based solely on taxa abundance data. This classification was also conducted for each ecoregion and identified seven diatom communities typical of the Canadian Shield, 10 diatom communities typical of the Appalachians, and 10 diatom communities typical of the St. Lawrence Lowlands. Resulting groups were graphically presented on ordinations to interpret, a posteriori, the environmental gradients associated with the diatom groups and to identify the diatom communities representing the reference conditions of each of the stream reach groups. The reference community for each stream reach group was found at the lower end of the alteration gradient, indicating nonimpacted or “least-disturbed” conditions. As a result, the selection of reference samples was based solely on community structure. Water physico-chemistry and land use characteristics were used only to interpret the position of each diatom community along the pollution gradient.

Finally, a classification based solely on diatom reference communities found pH and conductivity to be the main discriminating factors, regardless of ecoregion and stream type. Although a specific diatom reference community may be identified for each stream group, our results suggest that many of these communities exhibit strong similarities. Therefore, only two reference communities may be used, one for circumneutral conditions and one for alkaline conditions. These results suggest that pH and conductivity, which partially depend on geology and the presence of wetlands, have a major influence on the composition of reference diatom communities. Similar results were obtained in Europe and the USA (e.g., Potapova and Charles 2002; Gosselain et al. 2005).

Although only two reference communities are sufficient, it is still useful to compare a test site with its specified stream reach group reference community, especially where there is a lack of pristine conditions. For example, in the case of alkaline conditions, most of the reference sites are from large rivers. These reference communities may not adequately represent the expected diatom community in small, unaltered streams. It is therefore useful to go back to the reference community identified for the stream reach group representing small agricultural streams. This reference community represents least-disturbed conditions rather than pristine conditions. It defines an intermediate restoration goal, which may be more realistic than the pristine conditions, especially for the heavily impacted streams of the St. Lawrence Lowlands.

Low pH reference diatom communities (pH below 7.65)

In Quebec, the presence of gneiss–paragneiss rocks and felsic rocks, with low buffering capacity, seems to be responsible for pH variations in the reference streams. This applied primarily to Canadian Shield reference sites and St. Lawrence Lowlands reference sites, which have a large portion of their watershed located on the Canadian Shield. Low pH also seems to be partly explained by the presence of ultramafic and siliceous rocks, although these relationships need to be confirmed by other analyses based solely on Appalachian reference sites. Finally, the presence of wetlands appears to explain the high concentrations of DOC in some watersheds, which contribute to lower the pH level, especially in the Appalachians.

As a result, reference communities identified for the streams with low buffering capacity in the Appalachians and St. Lawrence Lowlands have similar communities to those found on the Canadian Shield. Low pH reference diatom communities are dominated by taxa indicators of oligotrophic conditions, e.g., *A. minutissimum* (Leland 1994) and *F. capucina*, and more acidic conditions, e.g., *T. flocculosa* (Van Dam et al. 1994) and *Brachysira microcephala* (Reavie and Smol 2001). According to their taxonomic composition, almost all reference communities identified in Canadian Shield and Appalachian watersheds approach undisturbed biotypes.

High pH reference diatom communities (pH above 7.65)

Conversely, alkaline streams of the Appalachians have reference communities similar to those found in alkaline streams of the St. Lawrence Lowlands. Their high pH seems to be explained by the presence of carbonated and (or) clay rocks. These reference communities are dominated by such species as *A. minutissimum*, *Nitzschia sinuata* var. *tabellaria*, *Nitzschia fonticola*, and *Nitzschia palea* var. *debilis*. However, in the Appalachians, there is an absence of undisturbed reference samples for group 4 represented by the sites of the River Fouquette. The similarities between watershed characteristics of the River Fouquette and those of the small rivers located in the middle St. Lawrence Lowlands (group 5) suggests that they share the same restoration goals. However, wetlands occupy a significant portion of the River Fouquette's watershed (4.8%), potentially contributing to the acidification of the water and suggesting that the reference conditions may instead be acidic. More reference sites are needed to clarify this assumption. Moreover, the insufficient number or the lack of reference samples for certain stream categories, particularly in the St. Lawrence Lowlands (small agricultural streams), made it difficult to identify real restoration goals for these environments. Most of the taxa found in reference communities of small agricultural streams are also indicators of impacted conditions, e.g., *Cocconeis placentula* var. *euglypta* and *Planothidium lanceolatum*. An increase in the number of reference sites is necessary to clarify the structure of these reference communities.

Development of a bioassessment tool

In Canada, benthic macroinvertebrates are the group of aquatic organisms most widely used in bioassessment (e.g., Linke et al. 1999; Reynoldson et al. 2001; Winter et al. 2002). Several recent studies carried out in Canada and United States show the potential of diatom communities as indicators of water quality (e.g., Leland and Porter 2000; Winter and Duthie 2000*b*; Potapova and Charles 2002). Despite the fact that diatoms have been proven to be good indicators of environmental conditions, there are no diatom indices currently being used in Canadian biomonitoring programs. European countries have a long history of developing diatom-based indices for biological assessment and biocriteria. A variety of indices have been developed, with the most popular indices being the IPS (Coste 1982), the IBD (Lenoir and Coste 1996; Prygiel and Coste 2000), the trophic diatom index (Kelly and Whitton 1995), and the Sladecek index (SLA) (Sladecek 1973). All of the above indices were developed based on a weighted average equation (Zelinka and Marvan 1961) in which the optima and tolerance values for each taxon were determined regardless of the reference conditions specific to each ecoregion or stream type. Species optima and tolerances are also generally derived from physico-chemical data, which implies that biomonitoring depends on environmental variables and leads to a circular argument. The main reason for using biota is precisely the complementary and nonredundant information provided on the ecosystem status. The most logical approach, in term of ecological integrity, is to derive bioindication information directly from community structure.

Based on the results of the present study, Lavoie et al. (2006) developed a diatom-based index that integrates different types of stream alterations and provides information related to the "distance" from the less-impacted state. The eastern Canadian diatom index (IDEC) used correspondence analysis (CA) to develop a "chemistry-free" index in which the position of the sites along the gradient of maximum variance (first axis) is strictly determined by diatom community structure and is therefore independent of measured environmental variables. The index value indicates the distance of each diatom community from its specific reference community. A high index value represents a non- or less-impacted site, whereas a low index value represents a more heavily impacted site. Two subindices were developed based on two sets of reference communities. The IDEC circumneutral subindex includes the sites that have reference communities characteristic of slightly acidic or neutral environments, whereas the IDEC alkaline subindex includes the sites that have reference communities characteristic of environments in which pH values are naturally higher than 7.65. The distinction between the two subindices is fundamental to ensuring that each stream has the potential to reach a high IDEC value following complete restoration of its ecosystem.

Development of a predictive model

The present work represents the preliminary phase in the elaboration of a model to predict the diatom community composition expected at a given site following restoration. To predict diatom community composition, predictive modelling assumes that a site is in its reference state. As stated by Reynoldson et al. (1997), if a test site can be associated with a group of reference sites representing the reference condition, then those reference sites can be used to predict community composition expected at the test site in the absence of disturbance. Stream and watershed attributes for groups of reference sites may be compared to identify a subset of variables used in the prediction of group membership. The RIVPACS predicts macroinvertebrate fauna at a given site from a small number of environmental parameters. By comparing the observed fauna with the predicted or “target” fauna, a measure of site integrity can be obtained (Wright et al. 2000). The AusRivAS is based on the RIVPACS model with the exception that major habitats are sampled and modelled separately. The BEAST (Reynoldson et al. 1995, 1997) is similar to the AusRivAS and RIVPACS approaches but uses abundances of macroinvertebrates instead of their presence or absence. Similar approaches were recently employed in Europe to predict the community structure of aquatic communities using advanced modelling techniques such as artificial neural networks, bayesian models, and genetic algorithms (Lek et al. 2005). From our results, it appears that a predictive model could be developed using, as predictors, the variables responsible for pH variations in reference streams, such as the proportion of gneiss–paragneiss, felsic, or clay rocks. These variables could be used to predict the group (circumneutral or alkaline community) to which impacted streams in southern Québec belong.

Acknowledgements

We thank Muriel Gevrey, Mylène Vallée, Martin Matteau, and Amélie Turcot for their field and laboratory assistance. Special thanks are due to Paul Hamilton (Canadian Museum of Nature, Ottawa) for providing detailed comments on diatom taxonomy and to Kristin Mueller for editing this manuscript. This project was supported by grants from the Fonds d’Action Québécois pour le Développement Durable (FAQDD), the Programme d’Aide à la Recherche et au Développement en Environnement (Ministry of the Environment, Québec), the Fonds québécois de la recherche sur la nature et les technologies (FQRNT), and the Natural Sciences and Engineering Research Council of Canada (NSERC) awarded to Stephane Campeau. Martine Grenier and Isabelle Lavoie were supported by scholarships from NSERC, Ontario Graduate Studies (OGS), and Fondation de la Faune du Québec. We thank the Québec Ministry of the Environment for providing water chemistry data and digital maps. This work is a contribution to the Groupe de Recherche Interuniversitaire en Limnologie (GRIL).

References

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

Baker, E.A., Wehrly, K.E., Seelbach, P.W., Wang, L., Wiley, M.J., and Simon, T. 2005. A multimetric assessment of stream condition in the northern lakes and forests ecoregion using spatially explicit statistical modelling and regional normalization. Trans. Am. Fish. Soc. **134**(3): 697–710.

Barbour, M.T., Stribling, J.B., and Karr, J.R. 1995. Multimetric approach for establishing biocriteria and measuring biological condition. *In* Biological assessment and criteria. Tools for water resource planning and decision making. *Edited by* W.S. Davis and T.P. Simon. Lewis Publishers, Boca Raton, Florida. pp. 63–77.

Barbour, M.T., Gerritsen, J., Griffith, G.E., Frydenborg, R., McCarron, E., White, J.S., and Bastian, M.L. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. J. North Am. Benthol. Soc. **15**: 185–211.

Barbour, M.T., Gerritsen, J., Snyder, B.D., and Stribling, J.B. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish. 2nd ed. US Environmental Protection Agency, Office of Water, Washington, D.C. EPA 841-B-99-002.

Campeau, S., Lavoie, I., Grenier, M., and Dillon, P. 2005. Développement d’un indice d’eutrophisation pour le suivi de la pollution agricole au Québec. Rapport déposé au ministère de l’Environnement du Québec et au Fonds d’action québécois pour le développement durable. Université du Québec à Trois-Rivières, Québec, Canada.

Carrel, G. 2002. Prospecting for historical fish data from the Rhone River basin: a contribution to the assessment of reference conditions. Arch. Hydrobiol. **155**(2): 273–290.

Clayton, J.S., Ehrlich, W.A., Cann, D.B., Day, J.H., and Marshall, I.B. 1978. Soils of Canada. Research Branch, Department of Agriculture, Ottawa, Ontario.

Coste, M. 1982. Étude des méthodes biologiques quantitatives d’appréciation de la qualité des eaux. Cemagref, Lyon, France.

Coste, M., Tison, J., and Delmas, F. 2004. Flores diatomiques des cours d’eau : proposition de valeurs limites du « Bon État » pour l’IPS et l’IBD. Document de travail – Unité de Recherche Qualité des Eaux – Cemagref Bordeaux, France.

Descy, J.P., Gerard, P., Darchambeau, F., Demol, T., Fauville, C., Gosselain, V., Lepieur, F., and Vanden Bossche, J.P. 2005. Définition des conditions de référence biologiques des rivières en Wallonie. Rapport final du Programme Intégré de Recherche Environnement–Eau–DGRNE, Belgique.

ESRI. 1999. ArcGIS [computer software]. Version 8.0. ESRI, Redlands, California.

European Parliament and the Council of the European Union. 2000. Directive of the European Parliament and the Council 2000/60/EC establishing a framework for community action in the field of water policy. ENV 221, PE-CONS 3639/1/00.

Fallu, M.A., Allaire, N., and Pienitz, R. 2000. Freshwater diatoms from northern Québec and Labrador (Canada): species–environment relations in lakes of boreal forest, forest tundra and tundra regions. *In* Bibliotheca Diatomologica. Vol. 45. J. Cramer, Berlin, Germany.

Fink, M.H., Moog, O., and Wimmer, R. 2000. Fließgewässer-aturräume Ősterreichs. Monographien des Umweltbundesamtes, Wien.

Foerster, J., Gutowski, A., and Shaumburg, J. 2004. Defining types of running waters in Germany using benthic algae: a prerequisite for monitoring according to the Water Framework Directive. J. Appl. Phycol. **16**(5): 407–418.

Gerritsen, J., Barbour, M.T., and King, K. 2000. Apples, oranges, and ecoregions: on determining pattern in aquatic assemblages. J. North Am. Benthol. Soc. **19**(3): 487–496.

Gibson, G.R., Barbour, M.T., Stribling, J.B., Gerritsen, J., and Karr, J.R. 1996. Biological criteria: technical guidance for streams and small rivers. Revised ed. US Environmental Protection Agency, Office of Water, Washington, D.C. EPA 822-B-96-001.

Giraudel, J.L., and Lek, S. 2001. A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. Ecol. Model. **146**: 329–339.

Gosselain, V., Campeau, S., Gevrey, M., Coste, M., Ector, L., Rimet, F., Tison, J., Delmas, F., Park, Y.-S., Lek, S., and Descy, J.P. 2005. Diatom typology of low-impacted conditions at a multi-regional scale: combined results of multivariate analyses and SOM. *In* Modelling community structure in freshwater ecosystems. *Edited by* S. Lek, M. Scardi, P.F.M. Verdonschot, J.P. Descy, and Y.-S. Park. Springer, Berlin, Germany. pp. 317–342.

Hawkins, C.P., Norris, R.H., Gerritsen, J., Hughes, R.M., Jackson, S.K., Johnson, R.K., and Stevenson, R.J. 2000. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. J. North Am. Benthol. Soc. **19**: 541–556.

Hering, D., Buffagni, A., Moog, O., Sandin, L., Sommerhäuser, M., Stubauer, I., Feld, C., Johnson, R.K., Pinto, P., Soulikidis, N., Verdonshot, P.F.M., and Zahrádková, S. 2003. The development of a system to assess the ecological quality of streams based on macroinvertabrates — design of the sampling program within the AQEM project. Int. Rev. Hydrobiol. **88**: 345–361.

Hughes, R.M. 1995. Defining acceptable biological status by comparing with reference conditions. *In* Biological assessment and criteria: tools for water resource planning and decision making. *Edited by* W. Davis and T. Simon. Lewis Publishers, Boca Raton, Florida. pp. 31–47.

Kelly, M.G., and Whitton, B.A. 1995. The trophic diatom index: a new index for monitoring eutrophisation in rivers. J. Appl. Phycol. **7**: 433–444.

Kohonen, T. 2001. Self-organizing maps. 3rd ed. Springer, Berlin, Germany.

Krammer, K. 2000. Diatoms of Europe. Diatoms of the European inland waters and comparable habitats. Vol. 1. The genus *Pinnularia*. A.R.G. Gantner Verlag K.G., Ruggell, Liechtenstein.

Krammer, K. 2002. Diatoms of Europe. Diatoms of the European inland waters and comparable habitats. Vol. 3. *Cymbella*. A.R.G. Gantner Verlag K.G., Ruggell, Liechtenstein.

Krammer, K. 2003. Diatoms of Europe. Diatoms of the European inland waters and comparable habitats. Vol. 4. *Cymbopleura*, *Delicata*, *Navicymbula*, *Gomphocymbellopsis*, *Afrocymbella*. A.R.G. Gantner Verlag K.G., Ruggell, Liechtenstein.

Krammer, K., and Lange-Bertalot, H. 1986. *Bacillariophyceae*. 1. Teil: *Naviculaceae*. *Edited by* H. Ettl, J. Gerloff, H. Heynig, and D. Mollenhauer. Süβwasserflora von Mittleuropa, Band 2/1. Gustav Fischer Verlag, Stuttgart and New York.

Krammer, K., and Lange-Bertalot, H. 1988. *Bacillariophyceae*. 2. Teil: *Bacillariaceae*, *Epithemiaceae*, *Surirellaceae*. *Edited by* H. Ettl, J. Gerloff, H. Heynig, and D. Mollenhauer. Süβwasserflora von Mittleuropa, Band 2/2. Gustav Fischer Verlag, Stuttgart and New York.

Krammer, K., and Lange-Bertalot, H. 1991*a*. *Bacillariophyceae*. 3. Teil: *Centrales*, *Fragilariaceae*, *Eunotiaceae*. *Edited by* H. Ettl, J. Gerloff, H. Heyinig, and D. Mollenhauer. Süβwasserflora von Mittleuropa, Band 2/3. Gustav Fischer Verlag, Stuttgart and Jena.

Krammer, K., and Lange-Bertalot, H. 1991*b*. *Bacillariophyceae*. 4. Teil: *Achnanthaceae* Kritische Ergänzungen zu *Navicula* (*Lineolatae*) und *Gomphonema*. *Edited by* H. Ettl, G. Gärtner, J. Gerloff, H. Heynig, and D. Mollenhauer. Süβwasserflora von Mittleuropa, Band 2/4. Gustav Fischer Verlag, Stuttgart and New York.

Lange-Bertalot, H. 2001. Diatoms of Europe. Diatoms of the European inland waters and comparable habitats. Vol. 2. *Navicula* sensu stricto. 10 genera separated from *Navicula* sensu lato. *Frustulia*. A.R.G. Gantner Verlag K.G., Ruggell, Liechtenstein.

Lavoie, I., Campeau, S., Grenier, M., and Dillon, P. 2006. A diatom-based index for the biological assessment of eastern Canadian rivers: an application of correspondence analysis (CA). Can. J. Fish. Aquat. Sci. **63**: 1793–1811.

Legendre, L., and Legendre, P. 1998. Numerical ecology. 2nd ed. Elsevier Science Publishers, Amsterdam, the Netherlands.

Lek, S., Scardi, M., Descy, J.-P., and Park, Y.-S. 2005. Modelling community structure in freshwater ecosystems. Springer Verlag, Berlin, Germany.

Leland, H.V. 1994. Distribution of benthic algae in Yakima River Basin, Washington, in relation to geology, land use, and other environmental factors. *In* Proceedings Abstracts American Water Resources Association's Symposium on the National Water-Quality Assessment Program, 7–9 November 1994, Chicago, Illinois. *Edited by* S.K. Sorenson. US Geological Survey Open-File Report 94–397.

Leland, H.V., and Porter, S.D. 2000. Distribution of benthic algae in upper Illinois River basin in relation to geology and land use. Freshw. Biol. **44**: 279–301.

Lenoir, A., and Coste, M. 1996. Development of a practical diatom index of overall water quality applicable to the French National Water Board network. *In* Use of algae for monitoring rivers II. *Edited by* B.A. Whitton and E. Rott. Studia Student, GmbH, Innsbruck, Austria. pp. 29–43.

Lepisto, L., Holopainen, L.L., and Vuoristo, H. 2004. Type-specific and indicator taxa of phytoplankton as a quality criterion for assessing the ecological status of Finnish boreal lakes. Limnologica, **34**(3): 236–248.

Li, T., and Ducruc, J.P. 2000. Les Provinces naturelles du Québec: niveau I du cadre écologique du Québec. Les Publications du Québec, Québec, Canada.

Linke, S., Bailey, R.C., and Schwindt, J. 1999. Temporal variability of stream bioassessments using benthic macroinvertebrates. Freshw. Biol. **42**: 575–584.

MathWorks Inc. 2001. MATLAB [computer software]. Version 6.0. The MathWorks Inc., Natick, Massachusetts.

Maxwell, J.R., Edwards, C.J., Jensen, M.E., Paustian. S.J., Parrot, H., and Hill, D.M. 1995. A hierarchical framework of aquatic ecological units in North America (Nearctic zone). US Department of Agriculture, Forest Service, North Control Forest Experiment Station, St. Paul, Minnesota. General Technical Report No. NC-176.

Meilinger, P., Schneider, S., and Melzer, A. 2005. The reference index method for the macrophyte-based assessment of rivers — a contribution to the implementation of the European Water Framework Directive in Germany. Int. Rev. Hydrobiol. **90**(3): 322–342.

Nijboer, R.C., Johnson, R.K., Verdonschot, P.F.M., Sommerhauser, M., and Buffagni, A. 2004. Establishing reference conditions for European streams. Hydrobiology, **516**(1): 91–105.

Oberdorff, T., Pont, D., Hugueny, B., and Chessel, D. 2001. A probabilistic model characterizing fish assemblages of French rivers: a framework for environmental assessment. Freshw. Biol. **46**: 399–415.

Omernik, J.M. 1987. Ecoregions of the conterminous United States. Ann. Assoc. Am. Geogr. **77**: 118–125.

Omernik, J.M., and Bailey, R.G. 1997. Distinguishing between watersheds and ecoregions. J. Am. Water Works Assoc., **33**(5): 935–949.

Pan, Y., Stevenson, R.J., Hill, B.H., and Herlihy, A. 2000. Ecoregion and benthic diatom assemblages in Mid-Atlantic Highlands streams, USA. J. North Am. Benthol. Soc. **19**: 518–540.

Park, Y.-S., Verdonschot, P.F.M., Chon, T.S., and Lek, S. 2003. Patterning and predicting aquatic macroinvertebrate diversities using artificial neural network. Water Res. **37**: 1749–1758.

Park, Y.-S., Oberdorff, T., and Lek, S. 2005. Patterning riverine fish assemblages using an unsupervised neural network. Modelling community structure in freshwater ecosystems. Springer Verlag, Berlin, Germany.

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

Potapova, M.G., and Charles, D.F. 2002. Benthic diatoms in USA rivers: distributions along spatial and environmental gradients. J. Biogeogr. **29**: 167–187.

Prygiel, J., and Coste, M. 1993. The assessment of water quality in the Artois–Picardie water basin (France) by the use of diatom indices. Hydrobiology, **269–279**: 343–349.

Prygiel, J., and Coste, M. 2000. Guide méthodologique pour la mise en oeuvre de l'Indice Biologique Diatomées. Agences de l'eau – Cemagref, Douai, France. NFT 90–354.

Rawer-Jost, C., Zenker, A., and Bohmer, J. 2004. Reference conditions of German stream types analysed and revised with macroinvertabrate. Limnologica, **34**(4): 390–397.

Reavie, E.D., and Smol, J.P. 1998. Freshwater diatoms from the St. Lawrence River. Bibliotheca Diatomologica Band 41. J. Cramer, Berlin, Germany.

Reavie, E.D., and Smol, J.P. 2001. Diatom–environmental relationships in 64 alkaline southeastern Ontario (Canada) lakes: a diatom-based model for water quality reconstructions. J. Paleolimnol. **25**: 25–42.

Reynoldson, T.B., Bailey, R.C., Day, K.E., and Norris, R.H. 1995. Biological guidelines for freshwater sediment based on BEnthic Assessment of SedimenT (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., and Rosenberg, D.M. 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. J. North Am. Benthol. Soc. **16**(4): 833–852.

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

Rogers, C.F., and Wasson, J.G 1997. Cadrage géographique de la structure et de la dynamique physique des hydrosystèmes du bassin rhodanien français: l'approche Hydro-écorégions. Communication présentée au Séminaire G.I.P. Hydrosystèmes « Journée du Rhône » le 27/03/97 à la direction régionale du CNRS Rhône-Alpes, 60100 Villeurbanne, France.

Schaumburg, J., Schranz, C., Foester, J., Gutowski, A., Hofmann, G., Meilinger, P., Schneider, S., and Schmedtje, U. 2004*a*. Ecological classification of macrophytes and phytobenthos for rivers in Germany according to the Water Framework Directive. Limnologica, **34**(4): 283–301.

Schaumburg, J., Schranz, C., Hofmann, G., Stelzer, D., Schneider, S., and Schmedtje, U. 2004*b*. Macrophytes and phytobenthos as indicator of ecological status in German lakes — a contribution to the implementation of the Water Framework Directive. Limnologica, **34**(4): 302–314.

Simpson, G.L., Shilland, E.M., Winterbottom, J.M., and Keay, J. 2005. Defining reference conditions for acidified waters using a modern analogue approach. Environ Pollut. **137**(1): 119–133.

Sladecek, V. 1973. System of water quality from the biological point of view. Ergebnisse der limnologie, Beih. Arch. Hydrobiol. **7**: 1–218.

Smol, J.P., Cumming, B.F., Dixit, A.S., and Dixit, S.S. 1998. Tracking recovery patterns in acidified lakes: a paleolimnological perspective. Restor. Ecol. **6**: 318–326.

ter Braak, C.J.F., and Smilauer, P. 2002. Reference manual and user's guide to CANOCO for Windows. Version 4.5. Center for Biometry, Wageningen.

Tison, J., Park, Y.-S., Coste, M., Wasson, J.G., Ector, L., Rimet, F., and Delmas, F. 2005. Typology of diatom communities and the influence of hydro-ecoregions: a study on the French hydrosystem scale. Water Res. **39**(14): 3177–3188.

Tonn, W.M., Paszkowski, C.A., Scrimgeous, G.J., Aku, P.K.M., Lange, M., Prepas, E.E., and Westcott, K. 2003. Effects of forest harvesting and fire on fish assemblages in boreal plains lakes: a reference condition approach. Trans. Am. Fish. Soc. **132**: 514–523.

Ultsch, A. 1993. Self-organizing neural networks for visualizing and classification. *In* Information and classification. *Edited by* O. Opitz, B. Lausen, and R. Klar. Springler-Verlag, Berlin, Germany. pp. 307–313.

US Water Environment Federation. 1992. Water quality 2000: a national water agenda for the 21st century. Water Environment Federation, Alexandria, Virginia.

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

Vincent, J.S. 1989. Le Quaternaire du sud-est du Bouclier canadien. *In* Le Quaternaire du Canada et du Groenland. *Edited by* R.J. Fulton. Ch. 3. Vol. 1. Geological Survey of Canada, Ottawa, Ontario.

Wasson, J.G., Chandesris, A., and Pella, H. 2002. Définition des hydro-écorégions de France métropolitaine. Approche régionale de typologie des eaux courantes et éléments pour la définition des peuplements de référence d'invertébrés. Rapport Cemagref Lyon BEA/LHQ et MATE/DE, France.

Winter, J.G., and Duthie, H.C. 2000*a*. Epilithic diatoms as indicators of stream total N and total P concentration. J. North Am. Benthol. Soc. **19**(1): 32–49.

Winter, J.G., and Duthie, H.C. 2000*b*. Stream biomonitoring at an agricultural test site using benthic algae. Can J. Bot. **78**(10): 1319–1325.

Winter, J.G., Somers, K.M., Dillon, P.J., Paterson, C., and Reid, R.A. 2002. Impacts of golf courses on macroinvertabrate community structure in Precambrian Shield streams. J. Environ. Qual. **31**(6): 2015–2025.

Winter, J.G., Dillon, P.J., Paterson, C., Reid, R.A., and Somers, K.M. 2003. Impact of golf course construction and operation on headwater streams: bioassessment using benthic algae. Can. J. Bot. **81**: 848–858.

Wright, J.F., Furse, M.T., and Armitage, P.D. 1993. RIVPACS: a technique for evaluating the biological quality of rivers in the UK. Eur. Water Pollut. Control. **3**(4): 15–25.

Wright, J.F., Furse, M.T., and Moss, D. 1998. River classification using invertebrates: RIVPACS applications. Aquat. Conserv. **8**: 617–631.

Wright, J.F., Sutcliffe, D.W., and Furse, M.T. 2000. Assessing the biological quality of freshwaters: RIVPACS and similar techniques. Freshwater Biological Association, Ambleside, UK.

Wunsam, S., Cattaneo. A., and Bourassa, N. 2002. Comparing diatom species, genera and size in biomonitoring: a case study from streams in the Laurentians (Quebec, Canada). Freshw. Biol. **47**: 325–340.

Zelinka, M., and Marvan, P. 1961. Zur Präzisierung der biologischen Klassification der Reinheit fliessender Gewässer. Arch. Hydrobiol. **57**: 389–407.

Appendix A

Table A1. Description of the environmental variables.

Variables	Description	Units
AP	Appalachians	
CS	Canadian Shield	
SL	St. Lawrence Lowlands	
CHL	Chlorophyll *a*	$mg\cdot cm^{-3}$
CON	Conductivity	$\mu S\cdot cm^{-1}$
DOC	Dissolved organic carbon	$mg\cdot L^{-1}$ C
FC	Faecal coliforms	UFC 100 mL^{-1}
NH_3	Ammonia ($N\text{-}NH_3$)	$mg\cdot L^{-1}$ N
NO_3	Nitrates–nitrites	$mg\cdot L^{-1}$ N
O_2	Dissolved oxygen	$mg\cdot L^{-1}$
pH	pH	pH
T	Temperature	°C
TN	Total nitrogen	$mg\cdot L^{-1}$ N
TP	Total phophorus	$mg\cdot L^{-1}$ P
TUR	Turbidity	NTU (nephelometric turbidity units)
SRP	Soluble phosphorus	$mg\cdot L^{-1}$ P
SS	Suspended solids	$mg\cdot L^{-1}$
VEL	Current velocity	$m\cdot s^{-1}$
ALT	Altitude	m
AREA	Watershed area	km^2
DIST	Distance to source	km
EMBANK	Embankment	m
SUBS	Dominant substrate	Ordinal variable
WIDTH	Stream width	Ordinal variable
FOREST	Forested area	% of watershed
URBAN	Urban area	% of watershed
WATER	Water surface area	% of watershed
WETLAND	Wetlands + bogs	% of watershed
ALLU	Alluvium deposits	% of watershed
EOL	Eolian deposits	% of watershed
FLUVIO	Fluvioglacial deposits	% of watershed
LACU	Lacustrine deposits	% of watershed
MARIN	Marine deposits	% of watershed
ROCK	Surficial bedrock	% of watershed
TILL	Till deposits	% of watershed
GNEISS	Gneiss and paragneiss	% of watershed
CLAY	Clay rocks (e.g., mudrock and schist)	% of watershed
CARBON	Carbonated rocks (e.g., limestone, marble, and dolomite)	% of watershed
FELS	Felsic rocks (e.g., granite and tonalite)	% of watershed
INTER	Intermediate rocks (e.g., syenite)	% of watershed
MAF	Mafic rocks (e.g., basalt, diorite, and anorthosite)	% of watershed
SILI	Siliceous rocks (e.g., sandstone, arkose, and quartzite)	% of watershed
UMAF	Ultramafic rocks	% of watershed
ANIMAL	Animal population	Animal units
HERBIC	Area with herbicide use	% of watershed
MANURE	Area with manure use	% of watershed
PASTURE	Pasture area	% of watershed
POP	Human population	Number