National Water-Quality Assessment Program

Prepared in cooperation with The Academy of Natural Sciences,
Patrick Center for Environmental Research

Development and Application of Indices to Assess the Condition of Benthic Algal Communities in U.S. Streams and Rivers

By Marina Potapova and Daren M. Carlisle

Open File Report 2011–1126

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia: 2011

For additional information about this report write to:

Marina Potapova
The Academy of Natural Sciences
Patrick Center for Environmental Research
1900 Benjamin Franklin Parkway
Philadelphia, PA 19103

Daren Carlisle
U.S. Geological Survey, MS 413
National Water-Quality Assessment Program
12201 Sunrise Valley Drive
Reston, VA 20192

Suggested citation:
Potapova, Marina, and Carlisle, D.M., 2011, Development and application of indices to assess the condition of benthic algal communities in U.S. streams and rivers: U.S. Geological Survey Open File Report 2011–1126, 40 p.

Contents

Figures

Tables

Development and Application of Indices to Assess the Condition of Benthic Algal Communities in U.S. Streams and Rivers

By Marina Potapova[1] and Daren M. Carlisle[2]

Abstract

Multi-metric indices (MMIs) are a measure of a combination of characteristics of biological communities and are used as indicators of water quality and ecological health. Although MMIs for algal communities have been developed for specific regions of the United States, none of the indices have national applicability. The MMIs described in this report were developed by the National Water-Quality Assessment Program of the U.S. Geological Survey to assess the overall health of benthic algal communities in U.S. streams and rivers within five geographic regions that encompass the conterminous United States.

The traditional procedure for developing MMIs (also referred to as indices of biological integrity) is to select individual metrics that, separately, can distinguish between undisturbed sites (selected for this study as reference sites) and predetermined disturbed sites. The metrics are then combined into a single index. In addition to traditional approaches for selecting individual metrics, the current study used stepwise logistic regressions to select sets of metrics that best predicted whether sites were in an undisturbed or a disturbed condition. Multi-metric indices and logistic regression models were developed for five regions of the United States using calibration datasets and were evaluated using independent validation datasets. Applying the regional MMIs to validation sites, the percentage of sites correctly classified as “reference” or “disturbed” ranged from 66 to 92 percent. Most often, only two or three metrics, typically percentages of individual organisms belonging to diatom taxa indicative of reference or disturbed calibration sites, were needed to distinguish between reference and disturbed validation sites. The classification accuracy of these indices did not improve when non-diatom metrics based on the autecological characteristics of algal taxa were included. The autecological metrics that were used to differentiate reference and disturbed sites were based on the assignment of the same diatom taxa to various ecological categories and therefore were redundant. Individual autecological metrics that measure the effect of specific stressors are needed, however, to identify potential causes of impairment. The regional MMIs developed in this study were applied at National Water-Quality Assessment Program sampling sites to assess the overall biological health of algal

[1] The Academy of Natural Sciences, Patrick Center for Environmental Research.
[2] U.S. Geological Survey, National Water-Quality Assessment Program.

communities. The assessment of algal communities in urban and agricultural land-use settings indicated increased stress to ecological health when compared to communities in other land-use settings.

Introduction

Algal-based assessments of water quality commonly rely on various metrics based on the autecology and relative abundance of individual taxa comprising the community. Autecological metrics are usually based on a documented understanding of the associations among varying concentrations of individual water chemistry variables (such as dissolved organic matter or orthophosphate) as they relate to the distribution of algal species (Sládeček, 1986; Steinberg and Schiefele, 1988; Kelly and Whitton, 1995). In a few instances, autecological metrics have been developed on the basis of an understanding of the relationship of algal species distributions to multiple water-quality measures (Lange-Bertalot, 1979; Prygiel and others, 1996). In addition to the question of which taxon or community characteristics are the best indicators of specific chemical stressors, there is a need to assess overall biological condition, which distinguishes impaired communities from healthy biological communities. This assessment usually compares various characteristics of undisturbed communities to those of communities known to be impaired.

Using a multi-metric index (MMI) to combine information about several characteristics of a biological community (and potentially the ecosystem) is one way to assess biological condition. Multi-metric indices (also known as indices of biological integrity) have been developed for fish (Karr, 1981), macroinvertebrates (Kerans and Karr, 1994; Barbour and others, 1996; Fore and others, 1996), and algae (Hill and others, 2000; Fore, 2002; Fore and Grafe, 2002; Wang and others, 2005). The principal concept of the MMI is that individual metrics represent “measures of condition in individuals, populations, communities, ecosystems, and landscapes” (Karr and Chu, 1999) and therefore provide an integrated assessment when combined into a single index. The standard approach for developing MMIs involves evaluating different metrics representing the different levels of ecological organization; the ultimate goal is to select metrics that represent as many levels as feasible. For instance, the original MMI developed by Karr (1981) included metrics indicative of fish community structure and composition, trophic structure, reproductive function, population abundances, and individual condition (health). Attempts have been made to implement the same concept with algal MMIs, including indicators of richness of algal genera, community composition, measures of biomass, and alkaline phosphatase activity (Hill and others, 2000). Additional indicators that have been proposed for algal MMIs include tolerance/intolerance, autecological guilds, morphological guilds, and individual condition (Fore and Grafe, 2002) and indicators based on similarity measures (Wang and others, 2005). The ecological attributes represented by individual metrics within MMIs have varied considerably among these studies. However, most of the metrics selected for MMIs are limited to attributes of community composition (Fore and Grafe, 2002; Wang and others, 2005). This apparent inconsistency is partly due to the absence of some measures, such as the condition of individuals or community productivity, and to differences in study objectives. Most importantly, no MMIs have been developed for algal communities that are applicable across a large geographic scale, such as that which were represented by U.S. Geological Survey National Water-Quality Assessment (NAWQA) Program studies throughout the conterminous United States.

Purpose and Scope

This report presents the research and findings of a NAWQA Program study, the primary goal of which was to develop an MMI that could be used to assess the condition of algal communities across diverse land-use settings of the United States. Specifically, this study had two objectives. The first was to evaluate the efficacy of alternative metrics for use in MMIs that distinguish between stream benthic algal communities in undisturbed (reference) and disturbed watersheds in five regions of the conterminous United States. Several metrics based on the relative abundance of algal taxa associated with undisturbed and disturbed watersheds were developed specifically for this study (appendix 1). The second objective was to apply these MMIs to assess the condition of algal communities in streams and rivers sampled by the NAWQA Program across varying land-use settings and geographic regions of the United States (fig. 1).

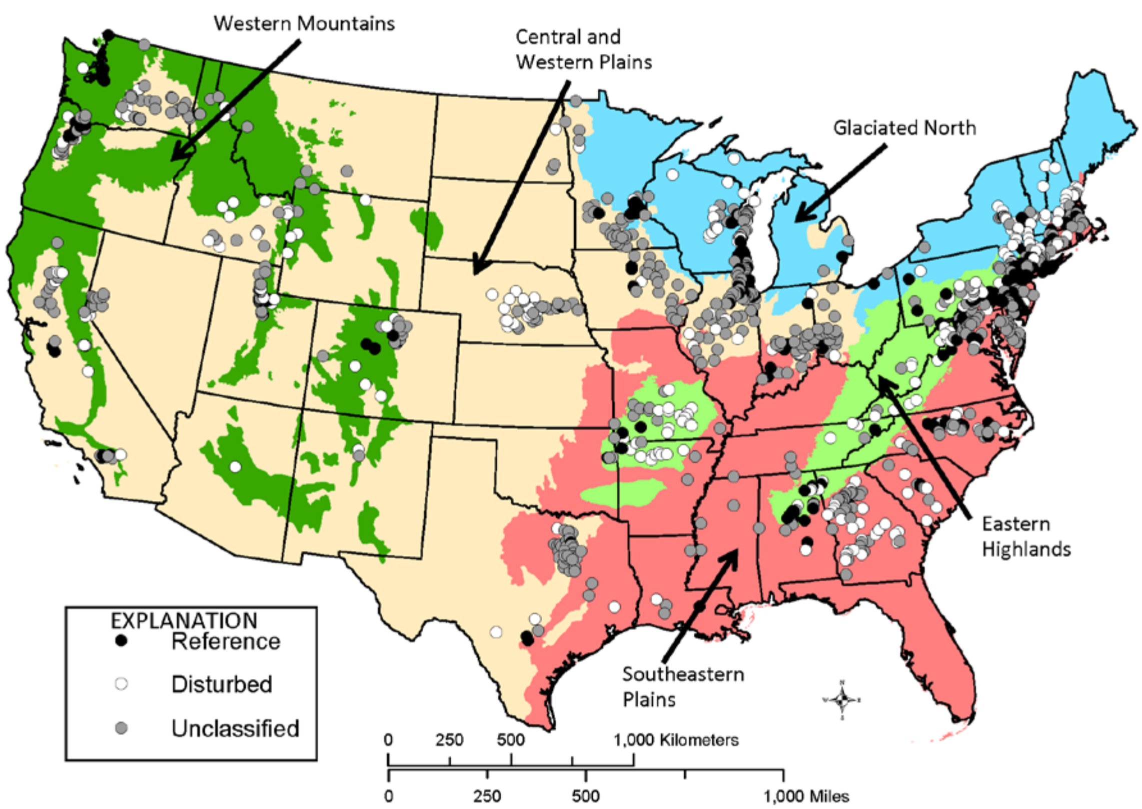

Figure 1. Map of the United States showing five regions and 1,071 sampling sites. White circles are reference sites, black circles are disturbed sites, and grey circles are unclassified sites.

General Approach

The general approach of this study was to apply several commonly used methods for assessing the condition of biological communities to benthic algal data collected by the NAWQA Program. Methods applied include traditional approaches as well as a statistical approach for selecting metrics that constitute the MMI. A distinct design feature of the NAWQA Program study was the targeting of selected sites in key land-use settings.

Methods of Index Development

Benthic algal data were collected by the NAWQA Program using nationally consistent methods. Undisturbed sites (the reference sites) and disturbed sites were identified among the sampling sites within each of five major regions of the conterminous United States. These sites were used to develop and evaluate MMIs to assess the condition of benthic algal communities.

Datasets

The datasets used in this study consisted of counts of benthic algae collected by the NAWQA Program at 1,071 river sites throughout the conterminous United States between 1993 and 2004 (fig. 1). Data were stratified into five aggregated regions of U.S. Environmental Protection Agency level-2 ecoregions, as described by Potapova and Charles (2007). The influence of spatial factors on algal community structure in U.S. streams and rivers and the need to develop algal indicators on a regional rather than national or continental scale were described previously by Potapova and Charles (2002). Algal samples used in this study were collected primarily from hard substrates using the methods of Porter and others (1993) and Moulton and others (2002). For sites where multiple samples in space or time had been collected, a single sample was randomly selected for this study. Diatoms were identified and enumerated from permanent slides at 1000X power and non-diatom algae were enumerated using Palmer-Maloney counting chambers at 400X power. Identifications and enumeration were done at the Patrick Center for Environmental Research (The Academy of Natural Sciences, Philadelphia, Pa.) using the methods described in Charles and others (2002).

The development of MMIs requires a set of sites classified a priori as reference quality and another set of sites classified a priori as disturbed, and that these classifications be based on information independent of biological data (Karr and Chu, 1999). For this study, sites were classified within each of the five regions into two contrasting groups: minimally disturbed (reference) and disturbed. The assignment of sites to the reference category was based on several criteria, including the professional judgment of local biologists, land cover data, and the evaluation of maps and imagery (Carlisle and Meador, 2007; Carlisle and others, 2008). Reference sites in some geographic areas (such as the midwestern United States in the Central and Western Plains) were considered “least disturbed” or “best attainable” (Stoddard and others, 2006). Disturbed sites were identified on the basis of an index of anthropogenic disturbance derived from 11 geospatial indicators of human activities (Falcone and others, 2009). These indicators included housing unit density (GeoLytics, 2001), percentage of urban land cover (Price and others, 2006), road density (GeoLytics, 2001), pesticide use (U.S. Geological Survey, 2007), nitrogen and phosphorus applications from fertilizers and manure (Ruddy and others, 2006), total dam storage (U.S. Army Corps of Engineers, 2006), percentage of streams that are in the category "canal/ditch/pipeline" (Horizon Systems Corporation, 2006), density of pollution dischargers (U.S. Environmental Protection Agency, 2006), density of mining operations (U.S. Geological Survey, 2006), and percentage of mining/transitional land cover (Price and others, 2006). Within each region, sites were ranked according to the disturbance index calculated from their contributing watersheds. Generally, sites within the highest quartile of disturbance within each region were classified as disturbed, although this threshold varied slightly among regions in order to achieve adequate numbers of reference and disturbed sites. In the Western Mountains region, all sites with a disturbance score greater than zero were classified as disturbed. In the Central and Western Plains region and in the Southeastern Plains region, sites with disturbance scores equal to or greater than 8 were classified as disturbed. Finally, sites in the Glaciated North and Eastern Highlands regions with disturbance scores greater than or equal to 4 were classified as disturbed. All other non-reference sites were not classified as disturbed (fig. 1), and therefore not used for the development of MMIs, but were later assessed for their degree of impairment. Less than 5 percent of the sites classified as reference sites, based on best professional judgment, had watershed disturbance index scores equal to or greater than the threshold value. These reference sites were removed from the reference category for the development of metrics and indices. All reference and disturbed sites within each of the five geographic regions were randomly subdivided into a calibration dataset (75 percent of the sites) or a validation dataset (25 percent of the sites). For each region, 5 to 10 MMIs with different combinations of metrics were developed with the calibration data and evaluated with validation data. The number of reference and disturbed sites in the calibration and validation datasets for each of the five regions is shown in table 1. The distribution of the disturbance index values for reference and disturbed sites in each region is shown in figure 2.

Table 1. The number of reference and disturbed sites in National Water-Quality Assessment Program datasets used for this study, by region and nationally.

[Subsets of the sites (in parentheses) are calibration (first number) and validation (second number) sites]

Region	Number of reference sites	Number of disturbed sites	Total number of sites
Glaciated North	55 (42:13)	47 (35:12)	146
Eastern Highlands	64 (48:16)	54 (41:13)	156
Southeastern Plains	72 (54:18)	61 (46:15)	311
Central and Western Plains	54 (41:13)	51 (38:13)	400
Western Mountains	31 (24:7)	13 (9:4)	58
National total	276 (209:67)	226 (169:57)	1,071

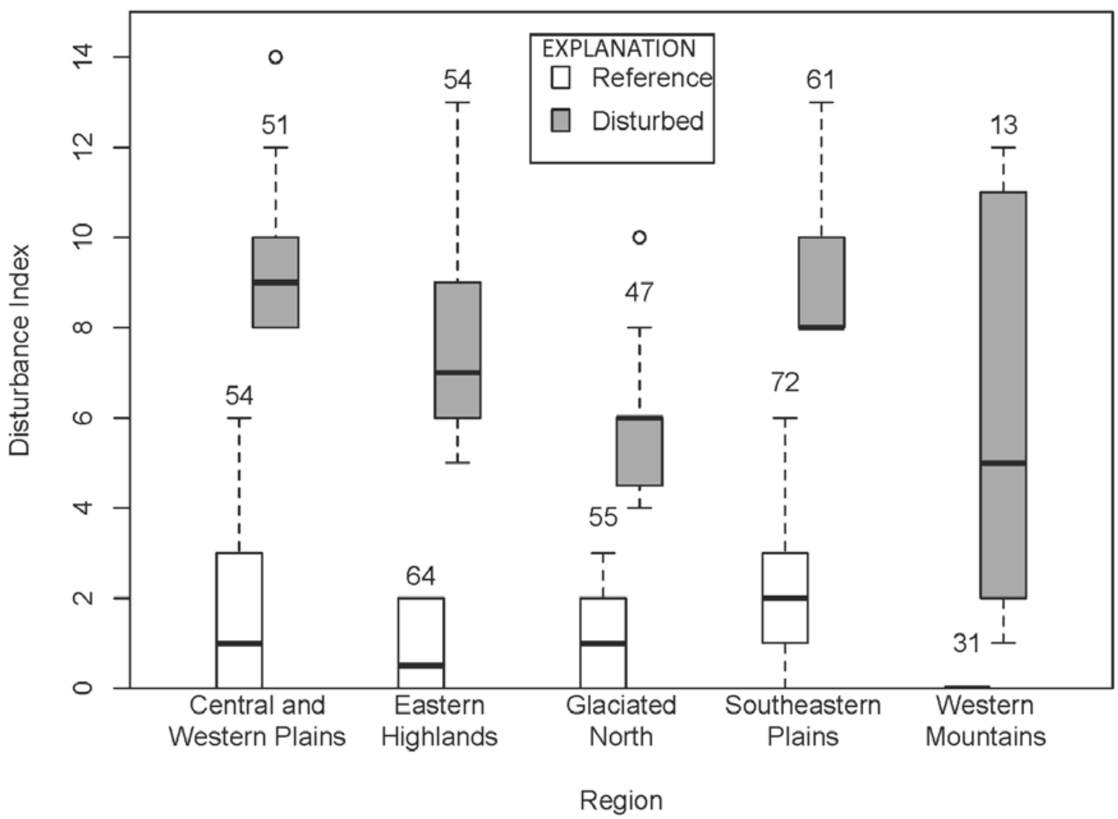

Figure 2. Values of the disturbance index in sampling sites grouped by regions and by the category “reference” or “disturbed.” The box represents interquartile range, which contains 50 percent of the observations; the whiskers are lines that extend to the highest and lowest numbers excluding outliers, which are values outside 1.5 times the box length. The horizontal line across a box is the median. The number of streams in each category is provided above each box.

Algal Metrics

One hundred twenty-four published and commonly used algal metrics were evaluated for inclusion in MMIs. One hundred metrics from existing literature sources were calculated and can be grouped into eight general categories (appendix 1), including total algal abundance, broad taxa groups, diversity measures, growth habit of diatoms, nutrient status indicators, organic pollution indicators, water ionic content indicators, and diatoms indicating general impairment. In addition, 24 metrics of general disturbance were developed in this study; these represented the relative abundance of diatoms and the relative cell density, cell density, and biovolume of non-diatom algae associated with reference and disturbed sites using the methods described below.

The 24 metrics developed in this study were based on the Indicator Species Analysis (ISA) method described by Dufrêne and Legendre (1997). The ISA method was used to identify species associated with reference and disturbed sites based on prior classification of watershed disturbance. Separate ISAs contrasting reference and disturbed sites were conducted with calibration data for each region; an additional ISA contrasted reference and disturbed calibration sites nationwide because study researchers expected that indicator species may be scale dependent. The nonclassified sites were not used in this analysis. Indicator values were calculated as products of the average relative frequency and relative abundance of each species in a group (that is, reference versus disturbed). Taxon indicator abilities identified by ISA can be biased for common taxa, so an adjustment was made by carrying out a series of Canonical Correspondence Analyses (CCAs) with a single constraining environmental variable "reference "or "disturbed" site for each of the six datasets (five regional and one national). If only one constraining variable is used, the CCA species scores along the first axis are proportional to weighted averages of the species in relation to that variable. In this analysis, the constraining variable was binary (reference versus disturbed sites), so the species at the extremes of the axis are those found predominantly in reference or disturbed sites. The taxa with unusual (in the lower or upper quartile) scores along the first CCA axis and found in at least three sites in regional datasets and in six sites in the national dataset were added to the lists of indicator taxa. Both ISA and CCA were performed using the PC-ORD/4 (MjM Software, Gleneden Beach, Oreg.). All metrics expressed as percentages of individual organisms were based either on a percentage of valves (for diatoms) or cells (for non-diatom algae) in the laboratory count.

Traditional Development of Multi-Metric Indices

The procedure for the development of MMIs is described in detail in Karr (1981), Barbour and others (1996), and Stoddard and others (2006). Approaches specific to algae are described by Fore and Grafe (2002) and Wang and others (2005). The methods used by Wang and others (2005) for selecting metrics for the MMI and scoring the resulting index were followed in this study. The ability of each metric to differentiate between reference and disturbed sites was evaluated by the Mann-Whitney nonparametric test, and only those metrics that showed significant differences ($P<0.05$) between reference and disturbed sites were retained. Those retained metrics with median values of zero at reference or disturbed sites were excluded. The remaining metrics were then assigned a score based on the distinctness of the distributions of values at reference compared to disturbed sites (table 2) (Barbour and others, 1996; Wang and others, 2005). The highest score, a "3," was assigned to metrics that had nonoverlapping distributions between reference and disturbed sites. A score of "2" was assigned to metrics that had overlapping interquartile ranges, but both medians were outside of this range. A score of "1" was assigned when one of the two medians was inside the overlapping portion of the interquartile range. The lowest score of "0" was given when both medians were inside the overlapping portion of the range. Metrics that decreased with disturbance (for example, the median of disturbed sites is lower than the median of reference sites) were scaled by dividing each value by the 90th percentile of the reference sites. Metrics that increased with disturbance were scaled by dividing the metric value by the 90th percentile of the disturbed sites and subtracting this number from one (Wang and others, 2005). The MMI was calculated as the sum of all scaled selected metrics multiplied by 10 and divided by the number of component metrics.

Table 2. Metrics ranked by their ability to differentiate between reference and disturbed sites based on an examination of boxplots.

[A score of 3 indicates no overlap of the interquartile ranges; a score of 2 indicates interquartile ranges overlapped, but both medians were outside of the overlap area. The metric's names are given in appendix 1. Underlined metrics were selected for inclusion in indices; --, indicates metric did not differentiate between reference and disturbed sites in that region.

Metric abbreviation*	Glaciated North	Eastern Highlands	Southeastern Plains	Central and Western Plains	Western Mountains
	Score per region				
DiaR	3	3	2	3	3
DiaD	3	3	3	3	3
Dia	3	3	3	3	3
SAR	3	2	--	1	3
TPh_r	3	--	--	--	--
TPl_r	3	--	--	--	--
TP_r	3	2	--	--	--
TPh	3	--	--	2	--
TPl	3	2	--	--	2
TP	3	2	--	--	2
TNh_r	3	--	--	--	--
TNl_r	3	3	--	--	--
TN_r	3	--	2	--	--
TNh	3	--	--	2	--
TNl	3	2	--	--	2
TN	3	--	--	--	--
CHLl	3	3	2	--	--
CHLh	3	3	2	2	--
SULFl	3	2	--	--	--
SULFh	3	--	2	--	--
CONDh	3	--	2	--	--
EUTR	3	2	--	--	--
EUTRandH	3	2	--	2	--
BF	3	2	2	2	2
O_h	3	--	2	--	2
MOTILE	3	2	--	--	--
DiaR_n	--	3	2	2	3
DiaD_n	--	3	2	2	--
Dia_n	--	3	2	2	--
O_l	--	2	--	--	--
FB	--	2	--	--	--
A_MESO	--	2	--	2	2
Bt	--	2	--	--	--
AP	--	2	--	--	--
SAD	--	2	--	--	--
Nhf	--	--	2	2	--
SP-OL	--	--	2	--	--
MESOTR	--	--	--	2	--
SAD_n	--	--	--	1	--
CD	--	--	--	--	3
CP	--	--	--	--	3
O_vl	--	--	--	--	2

*Abbreviations are given for metrics included in the models presented in table 3.

The threshold for classifying validation sites as "reference" or "disturbed" was the average of the 75th percentile of the disturbed calibration sites and the 25th percentile of the reference calibration sites for each region/regression model, which is a threshold commonly used in assessments with MMIs (Wang and others, 2005). The ability of the MMIs to distinguish between reference and disturbed conditions was estimated by Correct Classification Rate (CCR), which is the percentage of correctly classified sites, and also by sensitivity (percentage of correctly classified disturbed sites, or true positive rate, CCRd) and specificity (percentage of correctly classified reference sites, CCRr). Correct classification rates were calculated for both calibration and validation datasets.

For each region, 5 to 10 MMIs with different combinations of metrics were developed with the calibration data and evaluated with validation data. The MMIs included the single best (highest score) metric from each of the eight categories discussed previously, the best metrics from one or more categories, or other combinations of metrics from one or more categories (table 3).

Table 3. Multi-metric indices and logistic regression model summary.

[CCR, overall correct classification rate; CCRr, correct classification rate of reference sites; CCRd, correct classification rate of disturbed sites. The formulas are given to calculate the values of multi-metric indices (see Methods section, "Traditional MMI Development"). The values below threshold indicate impairment of algal communities. %, percent]

Region/ Index or model	Metrics and formula of the index/model	Threshold	Calibration set			Validation set		
			CCR	CCRr	CCRd	CCR	CCRr	CCRd
Glaciated North								
NAWQA diatom metrics, single category	(1 + DiaR_GN/92 - DiaD_GN/90)*5	4.03	70/77 (91%)	39/42	31/35	23/25 (92%)	12/13	11/12
NAWQA diatom metrics, various categories	(3 + DiaR_GN)/92 - DiaD_GN/90 + TP/0.93 + TN/0.91 - CHLh/73 - SULFh/80)*1.67	4.89	69/77 (90%)	40/42	29/35	22/25 (88%)	12/13	10/12
NAWQA diatoms and non-diatom metrics	(1 + DiaR_GN/92 - DiaD_GN/90 + SAR_GN/99.7)*3 .33	4.23	71/77 (92%)	39/42	32/35	23/25 (92%)	12/13	11/12
NAWQA and non-NAWQA metrics	(3 + DiaR_GN/92 - DiaD_GN/90 + SAR_GN/99.7 + TP/0.93 + TN/0.91 - CHLh/73 - SULFh/80)*1.43	4.89	69/77 (90%)	39/42	30/35	23/25 (92%)	12/13	11/12

Table 3. Multi-metric indices and logistic regression model summary.—Continued

[CCR, overall correct classification rate; CCRr, correct classification rate of reference sites; CCRd, correct classification rate of disturbed sites. The formulas are given to calculate the values of multi-metric indices (see Methods section, "Traditional MMI Development"). The values below threshold indicate impairment of algal communities. %, percent]

Region/ Index or model	Metrics and formula of the index/model	Threshold	Calibration set			Validation set		
			CCR	CCRr	CCRd	CCR	CCRr	CCRd
Glaciated North—continued								
Non-NAWQA diatom metrics	(3 - EUTRandH/85 - BF/42 + O_h/66 - MOTILE/46)*2.5	5.48	68/77 (88%)	37/42	31/35	19/25 (76%)	10/13	9/12
Logistic regression	Dia_GN		74/77 (96%)	40/42	34/35	23/25 (92%)	12/13	11/12
Eastern Highlands								
NAWQA diatom metrics, single category	(1 + DiaR_EH/86 - DiaD_EH/76)*5	3.90	79/89 (89%)	42/48	37/41	19/29 (66%)	10/16	9/13
NAWQA diatom metrics, various categories	(3 + DiaR_EH/86 - DiaD_EH/76 + CHLl/97 - CHLh/51 - SULFh/65 + TN_EH/0.88 + TP_EH/0.90)*1.43	4.96	69/89 (78%)	38/48	31/41	17/29 (59%)	10/16	7/13
NAWQA diatoms and non-diatom metrics	(2+ DiaR_EH/86 - DiaD_EH/76 + SAR_EH/94.1 - SAD_EH/71)*2.5	4.43	75/89 (84%)	40/48	35/41	23/29 (79%)	12/16	11/13
NAWQA and non-NAWQA metrics	(7 + DiaR_EH/86 - DiaD_EH/76 - MOTILE/54 - O_l/34 - Bt/33 - FB/93 - AP/34 - EUTRandH/83)*1.25	5.84	69/89 (78%)	37/48	32/41	20/29 (69%)	11/16	9/13
Non-NAWQA metrics only	(6 - MOTILE/54 - O_l/34 - Bt/33 - FB/93 - AP/34 - EUTRandH/83)*1.67	6.66	67/89 (75%)	36/48	31/41	20/29 (69%)	11/16	9/13
Logistic regression	DiaD_EH		74/89 (83%)	41/48	33/41	18/29 (62%)	10/16	8/13

Table 3. Multi-metric indices and logistic regression model summary.—Continued

[CCR, overall correct classification rate; CCRr, correct classification rate of reference sites; CCRd, correct classification rate of disturbed sites. The formulas are given to calculate the values of multi-metric indices (see Methods section, "Traditional MMI Development"). The values below threshold indicate impairment of algal communities. %, percent]

Region/ Index or model	Metrics and formula of the index/model	Threshold	Calibration set			Validation set		
			CCR	CCRr	CCRd	CCR	CCRr	CCRd
Southeastern Plains								
NAWQA diatom metrics, single category	(1 + DiaR_SP)/98 - DiaD_SP/79)*5	4.45	81/100 (81%)	44/54	37/46	27/33 (82%)	14/18	13/15
NAWQA diatom metrics, various categories	(2 + DiaR_SP/98 - DiaD_SP/79 - CHL2/58 + TN_SP/0.98)*2.5	7.20	77/100 (77%)	42/54	35/46	25/33 (76%)	14/18	11/15
NAWQA diatom and non-diatom metrics	No non-diatom metric had score above 0							
NAWQA and non-NAWQA metrics	(4 + DiaR_SP/98 - DiaD_SP/79 - Nhf/51 - BF/30 - EUTRandH/79 + O_h/77 + SP_OL/50)*1.43	5.40	77/100 (77%)	42/54	35/46	25/33 (76%)	13/18	12/15
Non-NAWQA metrics only	(3 - Nhf/51 - BF/30 - EUTRandH/79 + O_h/77 + SP_OL/50)*2		72/100 (72%)	40/54	32/46	23/22 (70%)	13/18	11/15
Logistic regression	Dia_SP		79/100 (79%)	40/54	39/46	24/33 (73%)	11/18	13/15
Central and Western Plains								
NAWQA diatom metrics, single category	(1 + DiaR_WCP/79 - DiaD_WCP/84)*5	3.58	70/79 (89%)	37/41	33/38	22/26 (85%)	10/13	12/13

Table 3. Multi-metric indices and logistic regression model summary.—Continued

[CCR, overall correct classification rate; CCRr, correct classification rate of reference sites; CCRd, correct classification rate of disturbed sites. The formulas are given to calculate the values of multi-metric indices (see Methods section, "Traditional MMI Development"). The values below threshold indicate impairment of algal communities. %, percent]

Region/ Index or model	Metrics and formula of the index/model	Threshold	Calibration set			Validation set		
			CCR	CCRr	CCRd	CCR	CCRr	CCRd
Central and Western Plains—continued								
NAWQA diatom metrics, various categories	(4 + DiaR_WCP/79 - DiaD_WCP/84 - CHLh/83 - TNh/90 - TPh/91)*2	3.80	64/79 (81%)	34/41	30/38	21/26 (81%)	10/13	11/13
NAWQA diatom and non-diatom metrics	(2 + DiaR_WCP/79 - DiaD_WCP/84 + SAR_WCP/94.3 - SAD/53.2)*2.5	4.43	70/79 (89%)	37/41	33/38	21/26 (81%)	12/13	9/13
NAWQA and non-NAWQA metrics	(5 + DiaR_WCP/79 - DiaD_WCP/84 - Nhf/54 - BF/61 - A_MESO /52 - EUTRandH /91 + MESOTR/15)*1.43	4.66	62/79 (78%)	32/41	30/38	22/26 (85%)	10/13	12/13
Non-NAWQA metrics only	(4 - Nhf/54 - BF/61 – A_MESO/52 - EUTRandH/91 + MESOTR/15)*2	5.12	55/79 (70%)	28/41	27/38	20/26 (77%)	10/13	10/13
Logistic regression	Dia_WCP		69/79 (87%)	35/41	34/38	23/26 (88%)	11/13	12/13
Western Mountains								
NAWQA diatom metrics only	(1+DiaR_WM/86 - DiaD_WM/71)*5	5.25	30/33 (91%)	23/24	7/9	5/11 (45%)	3/7	2/4
NAWQA diatom metrics,various categories	(1 + DiaR_WM/86 - DiaD_WM/71 + TPl/93)*3.33	4.46	25/33 (76%)	18/24	7/9	6/11 (55%)	4/7	2/4
NAWQA metrics	(2 + DiaR_WM/86 - DiaD_WM/71 - SAD_WM/71)*3.33	5.99	29/35 (83%)	22/24	7/9	7/11 (64%)	4/7	3/4

Table 3. Multi-metric indices and logistic regression model summary.—Continued

[CCR, overall correct classification rate; CCRr, correct classification rate of reference sites; CCRd, correct classification rate of disturbed sites. The formulas are given to calculate the values of multi-metric indices (see Methods section, "Traditional MMI Development"). The values below threshold indicate impairment of algal communities. %, percent]

Region/ Index or model	Metrics and formula of the index/model	Threshold	Calibration set			Validation set		
			CCR	CCRr	CCRd	CCR	CCRr	CCRd
Western Mountains								
NAWQA and non-NAWQA metrics	(4 + DiaR_WM/86 - DiaD_WM/71 + TPl/93 - O_vl/10 - A_MESO/68 - CP/0.22)*1.67	6.70	29/33 (89%)	21/24	8/9	8/11 (73%)	6/7	2/4
Non-NAWQA metrics only	(4 - O_vl/10 - A_MESO/68 - CP/0.22 - BF/18)*2.5	9.05	23/33 (70%)	17/24	6/9	7/11 (64%)	5/7	2/4
Logistic regression	DiaD_WM		31/33 (94%)	24/24	7/9	7/11 (64%)	6/7	1/4

Alternative Processes for Selecting Metrics

In addition to developing a traditional MMI, other approaches for selecting indicators were explored. Logistic regression was used as an exploratory tool for determining combinations of metrics that could distinguish between reference and disturbed sites in each of the five regions. This technique is convenient for identifying combinations of independent variables that provide the best prediction of a nominal dependent variable (Zuur and others, 2007). The dependent variable in this study was the classification of a site as reference or disturbed, and all 124 candidate metrics were considered as possible independent variables. Unlike in the MMI development, no metrics were excluded from consideration based on preliminary tests. The rationale for using a logistic regression was that any metric might be included in the model if it increased the model's predictive power, even if the metric did not distinguish between reference and disturbed sites when considered individually. The number of metrics was thus comparable to the number of observations in regional datasets, which can lead to spurious correlations and overfitting of the models. Thus, in the case of several highly correlated variables, the variable that causes the effect might not be the one selected for the model; the variable may have happened by chance to have had a somewhat higher correlation with the dependent variable. We did not search, however, for causative relationships, and knew beforehand that many metrics were highly correlated. This analysis was for exploratory purposes, and therefore a large number of candidate independent variables were included. Logistic regression was carried out with SPSS statistical software using forward selection, and F values for variable inclusion and exclusion were set at 0.10. The models were developed using only the calibration sites, and their predictive power was evaluated as the number of correctly classified sites after applying the models to the validation datasets.

Application of Indices to Assess Algal Community Condition at NAWQA Sampling Sites

The second goal of this study was to use the MMIs to assess the condition of algal communities at 1,071 sites sampled by the NAWQA Program (fig. 1). A distinct design feature of the NAWQA Program was the targeting of selected sites in key land-use and environmental settings. Accordingly, all non-reference sites were classified on the basis of land cover (Nakagaki and others, 2007) in their contributing watersheds using the criteria presented in table 4. Sites classified as "agricultural" were in watersheds dominated by agricultural land cover. Sites classified as "urban" were in watersheds with substantial urban land cover relative to agricultural or other land uses. The criteria by which basins were classified as "less developed" (table 4) were less stringent than the criteria used to classifiy sites as "reference" in this study, so sites classified as less developed represent an intermediate level of disturbance between reference sites and other land-use classifications (urban, agricultural, and mixed). In addition, sites classified as less developed may also be influenced by land use in close proximity to the sampling site even if most of the watershed is undeveloped. Sites classified as "mixed" are influenced by a mixture of urban and agricultural land uses.

Table 4. Land-cover criteria for classifying National Water-Quality Assessment Program sites into major land-use settings (from Gilliom and others, 2006).

[>, greater than; ≤, less than or equal to; %, percent]

Land-use classification	Watershed land-cover critieria
Agricultural	> 50% agricultural land and ≤ 5% urban land
Urban	>25% urban land and ≤ 25% agricultural land
Less developed	≤ 5% urban land and ≤ 25% agricultural land
Mixed use	All other combinations of urban, agricultural, and undeveloped land

Site values of MMIs were rescaled to facilitate interregional comparisons of algal condition by dividing the original MMI value at each site by the mean MMI value of the reference sites in the respective region (for example, Hawkins, 2006). This rescaling assumes that the expected MMI value at each non-reference site is the average of the reference site MMI values within the same region. The ratio of the observed MMI value and the expected MMI value can therefore be interpreted as a measure of disturbance, where ratios less than (<) or greater than (>) 1 indicate a substantial deviation of the MMI score at a site from the regional mean reference condition. All comparisons among land-use settings were made using an analysis of variance followed by multiple comparison tests.

Major Findings

Multi-metric indices based on different approaches to metric selection were evaluated for their ability to differentiate between reference and disturbed sites in independent, validation datasets. The best-performing MMI was applied to all sites and algal communities in streams influenced by varying land uses, which are assessed across the United States.

Algal Taxa Indicative of Reference and Disturbed Sites

A total of 1,179 algal taxa were identified at 1,071 sites; 829 of these were diatoms and 350 were non-diatoms. The 300 algal taxa identified by ISA and CCA as indicative of reference or disturbed sites in the five U.S. regions are listed in appendix 2. Most diatom species found to be indicative of disturbed sites in most regions (*Amphora pediculus, Cyclotella meneghiniana, Navicula gregaria, Navicula minima, Nitzschia amphibia, Nitzschia inconspicua, Planothidium lanceolatum, Rhoicosphenia abbreviata,* and *Sellaphora seminulum*) are known to increase in abundance with increasing nutrients, organic matter, and general pollution. Species indicative of reference sites in most regions (*Achnanthidium deflexum, Cymbella affinis, Encyonema minutum, Encyonopsis microcephala, Gomphonema angustatum, Staurosirella pinnata*) are those known to decrease in abundance with increasing organic and chemical nutrients. In contrast to the 266 diatom species found to be indicative of reference or disturbed sites, only 34 non-diatom taxa (or morphological groups) were found to be possible indicators of reference or disturbed sites (appendix 2).

Evaluation of Alternative Metrics

Few of the algal metrics examined were distinctly different between reference sites and disturbed sites. In most of the regional datasets, relatively few of the 124 metrics that passed the screening and Mann-Whitney tests had nonoverlapping interquartile ranges (a score of “3” in table 2). There were 26 such metrics in the Glaciated North but only 2 to 9 such metrics in the other regions. The metric that most often differentiated reference sites from disturbed sites in the validation data was the percentage of diatoms indicative of reference and disturbed sites within the calibration data. Other metrics with discriminatory ability included diatom metrics developed previously from the NAWQA dataset (for example, Potapova and Charles, 2007), such as percentage of diatoms indicative of nutrient enrichment and ionic strength. To increase the number of candidate metrics in the four regions other than the Glaciated North, metrics with scores of “2” and “1” were also considered (table 2).

The diatom metrics that were indicative of reference and disturbed sites in this study were also correlated with several chemical-specific metrics developed by other authors (table 5). Metrics for nutrient enrichment, organic pollution, and ionic strength were strongly (that is, absolute value Spearman rank ($|rho| > 0.7$) correlated with percentages of diatom species associated with disturbed or reference sites. In contrast, previously developed algal metrics indicative of ionic strength (Potapova and Charles, 2003), trophic state (Van Dam and others, 1994), and dissolved oxygen (Van Dam and others, 1994) were weakly correlated (that is, $|rho| < 0.6$) with metrics developed in this study.

Table 5. Spearman correlations between regional and national diatom metrics developed in this study. Diatoms associated with reference or disturbed sites and associated with several stressor-specific diatom metrics are shown.

[All correlations except one are significant at the 0.01 level; *, correlation significant at the 0.05 level. Regional metrics were calculated for corresponding sites. TP, total phosphorus; TN, total nitrogen]

Stressor-specific diatom metrics	Source	Metrics developed in this study			
		Diatoms associated with reference sites		Diatoms associated with disturbed sites	
		Regional	National	Regional	National
Brackish-fresh	Van Dam and others, 1994	-0.21	-0.61	0.51	0.61
High conductivity	Potapova and Charles, 2003	-0.23	-0.66	0.57	0.64
Low conductivity	Potapova and Charles, 2003	0.18	0.46	-0.42	-0.50
Eutraphentic+ hypereutraphentic	Van Dam and others, 1994	-0.33	-0.70	0.62	0.73
Oligo+ oligo-mesotraphentic	Van Dam and others, 1994	0.28	0.27	-0.27	-0.28
National low/high ratio TP indicators	Potapova and Charles, 2007	0.39	0.84	-0.62	-0.72
National low/high ratio TN indicators	Potapova and Charles, 2007	0.42	0.87	-0.70	-0.81
High oxygen-requiring taxa	Van Dam and others, 1994	0.46	0.79	-0.56	-0.63
Low oxygen-tolerant taxa	Van Dam and others, 1994	-0.07*	-0.35	0.27	0.33
Sum of α-mesosaprobic to polysaprobic	Van Dam and others, 1994	-0.36	-0.53	0.39	0.54

Multi-Metric Indices and Logistic Regression Models

In all five regions, the metrics developed in this study were most able to distinguish between reference and disturbed validation sites and were therefore selected for MMIs (table 5). Moreover, in all datasets except for the Western Mountains, the MMI with the most accurate classification was based on these metrics only. In the Southeastern Plains and Central and Western Plains, the MMIs with the highest CCR (82 percent and 88 percent, respectively) were those based on only two diatom metrics developed in this study: the regional lists of the diatom taxa associated with the reference sites and with the disturbed sites. The addition of existing diatom metrics or non-diatom metrics lowered the predictive power of the MMIs in these two regions. In the Glaciated North, the best MMIs were those including either diatom only or diatom and non-diatom metrics representing percentages of algae associated with reference and disturbed sites, and the CCR was exceptionally high (92 percent). In the Eastern Highlands, the MMI with the highest CCR for the validation dataset (79 percent) was based on diatom and non-diatom metrics representing percentages of algae associated with reference and disturbed sites. Only in the Western Mountains region did the most accurate MMI (CCR=73 percent) include metrics derived both from this and other studies. For the Western Mountains, inclusion of other metrics, such as proportion of diatoms known to be associated with organic pollution and dissolved oxygen concentration, helped to increase the classification accuracy of the index.

Logistic regression models and MMI-classified reference and disturbed sites with similar accuracy (60 to 90 percent) are shown in table 3. In all regions the models that had the highest CCR in validation datasets had only one predictor metric. In the Glaciated North, Southeastern Plains, and Central and Western Plains, the most accurate metric was the ratio of diatoms associated with reference sites to diatoms associated with disturbed sites. In the Eastern Highlands and Western Mountains, the most accurate metric was the percentage of diatoms associated with disturbed sites.

Assessment of Algal Communities by Region and Land-Use Setting

The condition of algal communities varied substantially by region and land-use setting (fig. 3). In all regions, algal communities in agricultural, urban, and mixed-use watersheds had a reduced biological condition relative to reference sites. Algal community condition in urban watersheds of the Central and Western Plains was lower than that of all other land-use settings. In contrast, the biological condition of algal communities in less-developed watersheds was generally similar to that of the reference sites, except in the Glaciated North, where the number of sites was low (n=7).

Discussion of Index Performance and Metrics

The performance of MMIs developed in this study was similar at a national scale to MMIs developed by others at smaller regional scales. Most algal MMIs, including the indices developed in this study, do not incorporate metrics truly reflecting different aspects of the communities as envisioned by early developers of MMIs. Although the overall MMI developed in this study has limited application to other studies, the individual metrics identified herein provide useful insight into the development of future MMIs to assess algal community condition.

Metrics Selected for MMIs

The relatively high classification accuracy of MMIs developed in this study (66 to 92 percent of the validation sites were correctly classified as reference or disturbed sites) is comparable to the index of biotic integrity developed by Wang and others (2005) for a much more homogeneous dataset of the Interior Plateau ecoregion, where a CCR of 80 percent was achieved. Notably, the small number of calibration sites in the Western Mountains probably diminished the clarity with which indicator species could be identified in that region.

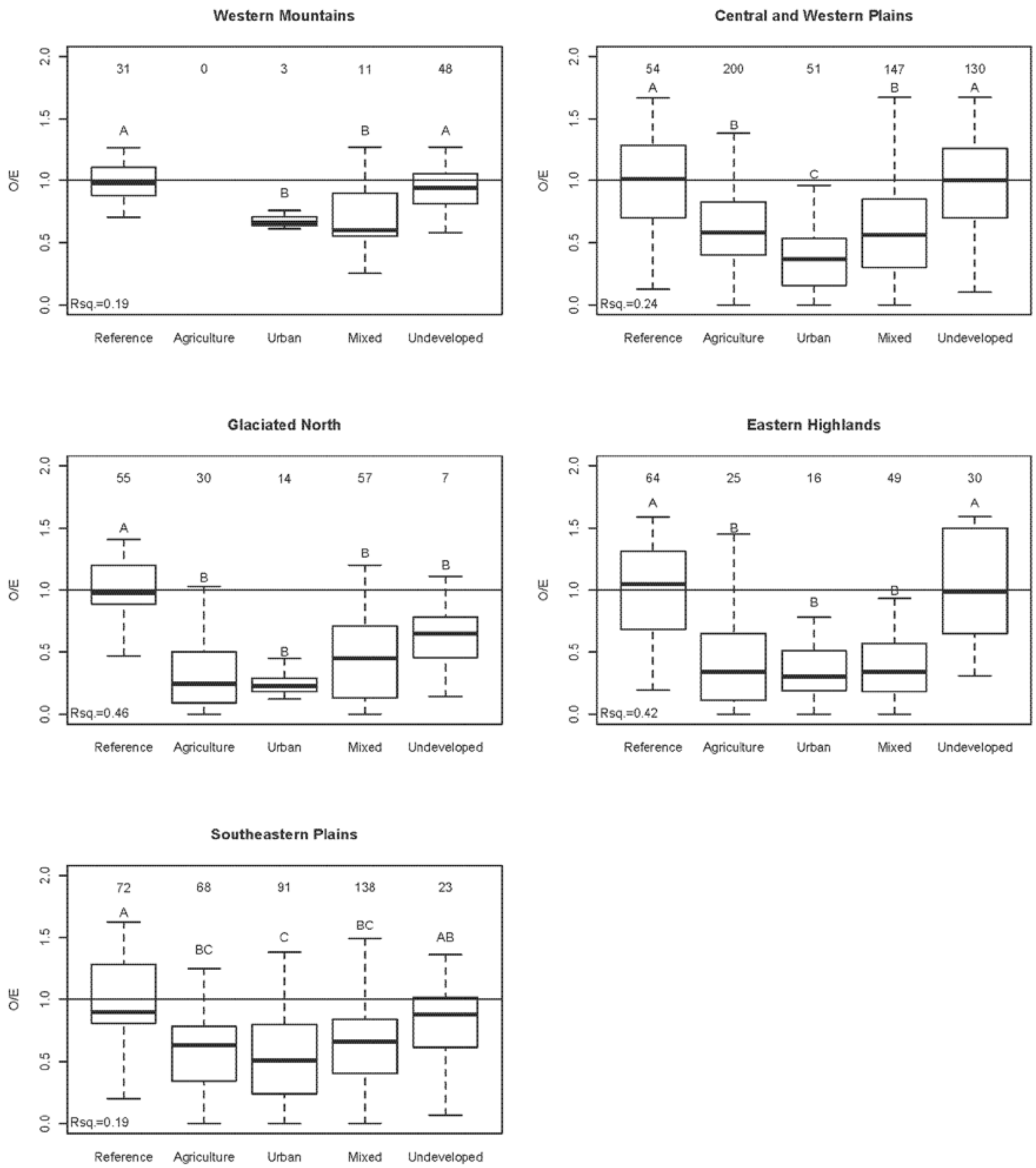

Figure 3. Values of multi-metric diatom indices in sites in the “Reference” group or grouped by land use settings. The box represents interquartile range, which contains 50 percent of values; the whiskers are lines that extend to the highest and lowest numbers excluding outliers, which are values outside 1.5 times the box length; the horizontal line across each box is the median. Boxes with the same superscript were not significantly different (P<0.05) in multiple comparison tests (bonferroni correction). The number of streams in each land-use category is provided above each box. Rsq.= adjusted r-squared value from anovas.

The original concept of combining different metrics into an index stems from the assumption that these metrics represent different structural and functional aspects of communities and ecosystems (Karr, 1981). For example, the major components of fish and macroinvertebrate indices of biotic integrity are usually some measures of community diversity and composition related to pollution tolerance and functional groups (Barbour and others, 1996; Fore and others, 2002). The metrics selected for algal MMIs in this study seem to represent different aspects of the community, but in fact these metrics are often highly redundant. Various diatom metrics related to trophic status, growth habit, organic pollution, and general pollution tolerance are actually based on similar lists of species compiled by others (Van Dam and others, 1994). They are usually correlated (table 2) because there is a uniform set of species known to favor eutrophic, silty, and otherwise anthropogenically disturbed waters—or conversely, oligotrophic, clear and undisturbed waters. For instance, 63 diatom taxa found in NAWQA samples were used to calculate the metric "percent brackish-water diatoms," while 43 of those were also classified as "eutraphentic diatoms" (both metrics are based on Van Dam and others, 1994). Similarly, 18 out of 37 "brackish," 51 out of 69 "alpha-mesosaprobic," and 15 out of 20 "polysaprobic-alpha-mesosaprobic" diatom taxa were also considered "eutraphentic." Further, 51 out of 69 taxa were also "eutraphentic." Combining various metrics into a single MMI essentially creates a new list of species that most accurately distinguish between impaired and non-impaired sites in a certain area. The same list can be more easily created by determining directly which species are commonly associated with either disturbed sites or reference sites.

Most algal MMIs, including the indices developed in this study, do not incorporate metrics truly reflecting different aspects of the communities as envisioned by Karr (1981). Hill and others (2000) did include metrics representing several distinct aspects of the community, such as measures of the diversity, productivity, biomass, and species composition, in their "periphyton index of biotic integrity," but the metric selection for this index was based on their expected responses rather than on their ability to distinguish between reference and disturbed sites. The diversity of algal communities usually increases at intermediate levels of watershed disturbance, and can be quite low in undisturbed sites as well as in heavily disturbed sites. For this reason, diversity measures are not considered reliable indicators of impairment. The percentage of malformed diatoms is rarely recorded during routine algal counts because in most areas malformed diatoms are rarely observed (but see Fore and Grafe, 2002).

Condition of Algal Communities

The MMIs developed in this study showed distinct differences in the condition of algal communities among four classified land uses (table 4). This result is not remarkable given that the index used to identify disturbed sites was partially based on land-cover characteristics. Pronounced alteration of algal communities in urban and agricultural streams was obvious in all five regions. Similar trends of degradation of algal communities with increased human activities were observed in other assessments using MMIs (Fore and Grafe, 2002; Wang and others, 2005). A unique aspect of this study, however, is that the condition of algal communities was assessed at a national spatial scale.

Use of Diatom MMIs and Individual Metrics

The MMIs developed in this study were intended to assess the condition of algal communities at sites sampled by the NAWQA Program, but with several caveats—the MMIs may be of use in other algal assessments. First, the MMIs developed in this study are limited to diatoms. Although metrics representing other algal groups were considered (such as percentages of specimens from certain ecological categories, total algal cell density and biovolume, and abundance of several algal divisions), they generally added little additional information to the assessment; that is, the predictive ability of MMIs improved by a negligible amount when diatom-based metrics were already included. Identification to species level is often difficult or impossible for non-diatom algae when using the traditional methodology based on light microscopy, and this might in part explain the lower success of non-diatom metrics. Because assessments of other algal groups are important to managers (for example, evaluating nuisance algae), new methods of characterizing non-diatom algal communities should be developed for environmental monitoring.

An additional caveat in using the MMIs developed in this study is the varying quality and relatively limited number of reference sites sampled in some regions. The difficulty in locating and sampling minimally disturbed watersheds in the Central and Western Plains region, for instance, means that the reference sites in this region may have experienced substantial anthropogenic disturbance. Such an unavoidable limitation should not, however, preclude the use of the MMIs or predictive models to assess impairment. It has been recognized (such as by Fore and Grafe, 2002) that quantifying human disturbance requires regional approaches; therefore, the assessment of impairment should also be region-specific (Stoddard and others, 2006). A more serious concern is the relatively small number of reference sites sampled in regions such as the Western Mountains. Reliable assessments depend on a set of reference sites that are environmentally similar to sites being evaluated. MMIs developed for regions that are environmentally heterogenous over small spatial scales (for example, elevation gradients in the Western Mountains region) require many more reference sites than were available for this study.

A final caveat for the use of these MMIs concerns the thresholds. The thresholds for MMI scoring used in this study are entirely dependent on the distribution of MMI values at NAWQA sites within each region. Assessments of sites sampled as part of other local or regional studies may or may not be appropriate, depending largely on their environmental similarity to sites sampled by NAWQA.

In addition to the MMIs, the individual metrics developed in this study may be reliable for assessing general water-quality conditions. Because NAWQA sampling was targeted toward streams influenced by intensive land use, diatom metrics identified in this study that are indicative of disturbed conditions are probably a reliable indicator of generally poor water quality. In contrast, diatom metrics that were indicative of undisturbed conditions are probably indicative of minimal human disturbance within most regions. Finally, metrics developed in many previous studies are known to be reliable indicators of specific chemical stressors and should continue to provide useful information regarding possible causes of algal community disturbance.

Summary

The National Water-Quality Assessment (NAWQA) Program of the U.S. Geological Survey developed multi-metric indices for use in assessing the condition of benthic algal communities in U.S. streams and rivers that encompass five geographic regions across the conterminous United States. The indices were developed by evaluating whether 124 alternative metrics of algal communities were able to discriminate between streams in disturbed and undisturbed watersheds. The study assessed the condition of algal communities at 1,071 NAWQA Program sampling sites. Algal communities in agricultural, urban, and mixed use settings in all regions had a reduced biological condition relative to undisturbed reference sites. The algal data and algal taxa developed for this study are available for use in the development of additional multi-metric indices.

Acknowledgments

This report was prepared as part of a cooperative research agreement between the Academy of Natural Sciences Patrick Center for Environmental Research and the U.S. Geological Survey National Water-Quality Assessment Program. The authors acknowledge Yangdong Pan, Portland State University, and Wade Bryant and James Coles, U.S. Geological Survey, for thoughtful reviews that improved the clarity of the report.

References Cited

Bahls, L.L., 1993, Periphyton bioassessment methods for Montana streams: Helena, Montana, Montana Department of Health and Environmental Sciences, accessed February 2, 2011, from http://www.archive.org/details/periphytonbioass00bahl.

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 macroinvertebtrates: Journal of the North American Benthological Society, v. 15, no. 2, p. 185–211.

Carlisle, D.M., and Meador, M.R., 2007, A biological assessment of streams in the eastern United States using a predictive model for macroinvertebrate assemblages: Journal of the American Water Resources Association, v. 43, no. 5, p. 1194–1207.

Carlisle, D.M., Hawkins, C.P., Meador, M.R., Potapova, Mariana, and Falcone, James, 2008, Biological assessments of Appalachian streams based on predictive models for fish, macroinvertebrate, and diatom assemblages: Journal of the North American Benthological Society, v. 27, no. 1, p. 16–37.

Charles, D.F., Knowles, Candia, and Davis, R.S., eds., 2002, Protocols for the analysis of algal samples collected as part of the U.S. Geological Survey National Water-Quality Assessment Program: Academy of Natural Sciences Report No. 02–06, 124 p.

Dufrêne, Marc, and Legendre, Pierre, 1997, Species assemblages and indicator species—The need for a flexible asymmetrical approach: Ecological Monographs, v. 67, no. 3, p. 345–366.

Falcone, J.A., Carlisle, D.M., and Weber, L.C., 2010. Quantifying human disturbance in watersheds—Variable selection and performance of a GIS-based disturbance index for predicting the biological condition of perennial streams: Ecological Indicators, v. 10, no. 2, p. 264–273.

Fore, L.S., 2002, Response of diatom assemblages to human disturbance—Development and testing of a multimetric index for the Mid-Atlantic region (USA), *in* Simon, T. P. (ed.), Biological response signatures—Indicator patterns using aquatic communities: Boca Raton, Fla., CRC Press, p. 445–480.

Fore, L.S., Karr, J.R., and Wisseman, R.W., 1996, Assessing invertebrate responses to human activities—Evaluating alternative approaches: Journal of the North American Benthological Society, v. 15, no. 2, p. 212–231.

Fore, L.S., and Grafe, Cynthia, 2002, Using diatoms to assess the biological condition of large rivers in Idaho (U.S.A.): Freshwater Biology, v. 47, no. 10, p. 2015–2037.

GeoLytics, 2001, Census 2000 and street 2000: East Brunswick, N.J., GeoLytics, Inc., 2 CD-ROMs.

Gilliom, R.J., Barbash, J.E., Crawford, C.G., Hamilton, P.A., Martin, J.D., Nakagaki, Naomi, Nowell, L.H., Scott, J.C., Stackelberg, P.E., Thelin, G.P., and Woolock, D.M., 2006, The quality of our nation's waters—Pesticides in the nation's streams and ground water, 1992–2001: U.S. Geological Survey Circular 1291, 172 p. (Also available at http://pubs.usgs.gov/circ/2005/1291/.)

Hawkins, C.P., 2006, Taxonomic completeness as a measure of biological integrity. Ecological Applications, v. 16, p. 1277-1294.

Hill, B.H., Herlihy, A.T., Kaufmann, P.R., Stevenson, R.J., McCorkick, F.H., and Burch Johnson, C., 2000, Use of periphyton assemblage data as an index of biotic integrity: Journal of the North American Benthological Society, v. 19, no. 1, p. 50–67.

Horizon Systems Corporation, 2006, National Hydrography Dataset Plus (NHDPlus) Home: Horizon Systems Corporation, accessed February 9, 2011, at *http://www.horizon-systems.com/nhdplus/*.

Karr, J.R., 1981, Assessment of biotic integrity using fish communities: Fisheries, v. 6, no. 6, p. 21–27.

Karr, J.R., and Chu, E.W., 1999, Restoring life in running waters—Better biological monitoring: Washington, D.C., Island Press, 206 p.

Kelly, M.G., and Whitton, B.A., 1995, The trophic diatom index—A new index for monitoring eutrophication in rivers: Journal of Applied Phycology, v. 7, no. 4, p. 433–444.

Kerans, B.L., and Karr, J.R., 1994, A benthic index of biotic integrity (B–IBI) for rivers of the Tennessee Valley: Ecological Applications, v. 4, p. 768–785.

Kentucky Division of Water, 1993, Methods for assessing biological integrity of surface waters: Frankfort, Kentucky, Kentucky Department of Environmental Protection. (Available from the Kentucky Department of Environmental Protection, 14 Reilly Road, Frankfort, Kentucky 40601 USA.)

Lange-Bertalot, H., 1979, Pollution tolerance of diatoms as a criterion for water quality estimation: Nova Hedwigia, v. 64, p. 285–304.

Moulton, S.R., II, Kennen, J.G., Goldstein, R.M., and Hambrook, J.A., 2002, Revised protocols for sampling algal, invertebrate, and fish communities as part of the National Water-Quality Assessment Program: U.S. Geological Survey Open-File Report 02–150, 87 p. (Available at *http://pubs.usgs.gov/of/2002/ofr-02-150/index.html*.)

Nakagaki, N., Price, C.P., Falcone, J.A., Hitt, K.J., and Ruddy, B.C., 2007, Enhanced National Land Cover Data 1992 (NLCDe 92), Edition 1.0, [digital map], accessed December 1, 2007, at *http://water.usgs.gov/lookup/getspatial?nlcde92.*

Porter, S.D., Cuffney, T.F., Gurtz, M.E., Meador, M.R., 1993, Methods for collecting algal samples as part of the National Water-Quality Assessment Program: U.S. Geological Survey Open-File Report 93–409. (This report has been superseded by U.S. Geological Survey Open-File Report 02–150, http://pubs.usgs.gov/of/2002/ofr-02-150/index.html.)

Potapova, Marina, and Charles, D.F., 2003, Distribution of benthic diatoms in U.S. rivers in relation to conductivity and ionic composition: Freshwater Biology, v. 48, no. 8, p. 1311–1328.

Potapova, Marina, and Charles, D.F., 2007, Diatom metrics for monitoring eutrophication in rivers of the United States: Ecological Indicators, v. 7, p. 48–70.

Price, C.V., Nakagaki, Naomi, Hitt, K.J., and Clawges, R.M., 2006, Enhanced historical land-use and land-cover data sets of the U.S. Geological Survey: U.S. Geological Survey Data Series 240, with digital maps. (Available at http://pubs.usgs.gov/ds/2006/240.)

Prygiel, J., Lévéque, L., and Iserentant, R., 1996, Un nouvel indice diatomique pratique pour l'évaluation de la qualité des eaux en réseau de surveilliance: Revue des Sciences de l'Eau, v. 9, no. 1, p. 97–113.

Ruddy, B.C., Lorenz, D.L., and Mueller, D.K., 2006, County-level estimates of nutrient inputs to the land surface of the conterminous United States, 1982–2001: U.S. Geological Survey Scientific Investigations Report 2006–5012, 17 p. (Available online from http://pubs.usgs.gov/sir/2006/5012/.)

Sládeček, V., 1986, Diatoms as indicators of organic pollution: Acta Hydrochimica et Hydrobiologica, v. 14, no. 5, p. 555–566.

Steinberg, C., and Schiefele, S., 1988, Indication of trophy and pollution in running waters: Zeitschrift Fur Wasser-Abwasser-Forschung, v. 21, p. 227–234.

Stoddard, J.L., Larsen, D.P., Hawkins, C.P., Johnson, R.K., and Norris, R.H., 2006, Setting expectations for the ecological condition of running waters—The concept of reference conditions: Ecological Applications, v. 16, p. 1267–1276.

U.S. Army Corps of Engineers, 2006, National Inventory of Dams: U.S. Army Corps of Engineers, accessed in July 2006 at *http://crunch.tec.army.mil/nidpublic/webpages/nid.cfm*.

U.S. Geological Survey, 2006, National atlas home page: U.S. Geological Survey, accessed in November 2006 at *http://www-atlas.usgs.gov*.

U.S. Geological Survey, 2007, Grids of agricultural pesticide use in the conterminous United States, 1997: U.S. Geological Survey Web page, accessed in June 2007 at *http://water.usgs.gov/GIS/metadata/usgswrd/XML/agpest97grd.xml.*

Van Dam, Herman, Mertens, Adrienne, and Sinkeldam, Jos, 1994, A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands: Aquatic Ecology, v. 28, no. 1, p. 117–133.

Wang, Yi-Kuang, Stevenson, R.J., and Metzmeier, Lythia, 2005, Development and evaluation of a diatom-based Index of Biotic Integrity for the Interior Plateau Ecoregion, USA: Journal of the North American Benthological Society, v. 24, no. 4, p. 990–1008.

Zuur, Alain, Ieno, E.N., and Smith, G.M., 2007, Analysing ecological data (Statistics for biology and health): New York, Springer, 672 p.

Appendix 1. Algal metrics commonly included in a multi-metric index. A total of 124 metrics are grouped into 8 categories. Abbreviations: TN, total nitrogen; TP, total phosphorous; µS/cm, microsiemens per centimeter; mg/L, milligrams per liter; %, percent; <, less than; >, greater than.

General category (total number of metrics per category)	Most common algal metrics	Algal metric abbreviation (if used in this report)	Source
Total algal abundance (2)	Total algal biovolume		
	Total algal cell density	CD	
Broad taxa groups (28)	Red algae, absolute and relative biovolume or cell density		
	Green algae, absolute and relative biovolume or cell density		
	Diatoms, absolute and relative biovolume or cell density		
	Cyanobacteria, absolute and relative biovolume or cell density		
	Xanthophytes, absolute and relative biovolume or cell density		
	Euglenophytes, absolute and relative biovolume or cell density		
	Cryptophytes, absolute and relative biovolume or cell density		
Diversity measures (6)	Number of all algal species or diatom species only		
	Percent of the valves in the diatom count made by the dominant diatom taxon and by the ten dominant diatom taxa		
	Shannon-Wiener diversity index based on proportions of cell densities of all algal taxa and diatoms only		
Growth habit of diatoms (2)	Ratio of centric to pennate diatoms	CP	
	Motile diatoms	MOTILE	Bahls 1993
Nutrient status indicators (24)	Relative abundance of oligotraphentic diatoms		Van Dam and others (1994)
	Relative abundance of oligo-mesotraphentic diatoms		Van Dam and others (1994)
	Relative abundance of oligotraphentic and oligo-mesotraphentic diatoms		Van Dam and others (1994)

Appendix 1. Algal metrics commonly included in a multi-metric index. A total of 124 metrics are grouped into 8 categories. Abbreviations: TN, total nitrogen; TP, total phosphorous; µS/cm, microsiemens per centimeter; mg/L, milligrams per liter; %, percent; <, less than; >, greater than.—Continued.

General category (total number of metrics per category)	Most common algal metrics	Algal metric abbreviation (if used in this report)	Source
Nutrient status indicators (24)—continued	Relative abundance of mesotraphentic diatoms	MESOTR	Van Dam and others (1994)
	Relative abundance of meso-eutraphentic diatoms		Van Dam and others (1994)
	Relative abundance of eutraphentic diatoms	EUTR	Van Dam and others (1994)
	Relative abundance of hypereutraphentic diatoms		Van Dam and others (1994)
	Relative abundance of eutraphentic and hypereutraphentic diatoms	EUTR and H	Van Dam and others (1994)
	Relative abundance of nitrogen-autotrophic diatoms tolerating very small concentrations of organic nitrogen		Van Dam and others (1994)
	Relative abundance of nitrogen-autotrophic diatoms tolerating elevated concentrations of organic nitrogen		Van Dam and others (1994)
	Relative abundance of facultatively nitrogen-heterotrophic diatoms needing periodically elevated concentrations of organic nitrogen	Nhf	Van Dam and others (1994)
	Relative abundance of obligately nitrogen-heterotrophic diatoms needing continuously elevated concentrations of organic nitrogen		Van Dam and others (1994)
	Relative abundance of diatoms–indicators of low total phosphorus at national scale	TPl	Potapova and Charles (2007)
	Relative abundance of diatoms–indicators of low total phosphorus at regional scale	TPl_r	Potapova and Charles (2007)
	Relative abundance of diatoms–indicators of high total phosphorus at national scale	TPh	Potapova and Charles (2007)

Appendix 1. Algal metrics commonly included in a multi-metric index. A total of 124 metrics are grouped into 8 categories. Abbreviations: TN, total nitrogen; TP, total phosphorous; µS/cm, microsiemens per centimeter; mg/L, milligrams per liter; %, percent; <, less than; >, greater than.—Continued.

General category (total number of metrics per category)	Most common algal metrics	Algal metric abbreviation (if used in this report)	Source
Nutrient status indicators (24)—continued	Relative abundance of diatoms–indicators of high total phosphorus at regional scale	TPh_r	Potapova and Charles (2007)
	Ratio of low to sum of high and low TP diatom indicators at national scale	TP	Potapova and Charles (2007)
	Ratio of low to sum of high and low TP diatom indicators at regional scale	TP_r	Potapova and Charles (2007)
	Relative abundance of diatoms–indicators of low total nitrogen at national scale	TNl	Potapova and Charles (2007)
	Relative abundance of diatoms–indicators of low total nitrogen at regional scale	TNl_r	Potapova and Charles (2007)
	Relative abundance of diatoms–indicators of high total nitrogen at national scale	TNh	Potapova and Charles (2007)
	Relative abundance of diatoms–indicators of high total nitrogen at regional scale	TNh_r	Potapova and Charles (2007)
	Ratio of low to sum of high and low TN diatom indicators at national scale	TN	Potapova and Charles (2007)
	Ratio of low to sum of high and low TN diatom indicators at regional scale	TN_r	Potapova and Charles (2007)
Organic pollution indicators (11)	Relative abundance of oligosaprobic diatoms	SP-OL	Van Dam and others (1994)
	Relative abundance of β-mesosaprobic diatoms		Van Dam and others (1994)
	Relative abundance of α-mesosaprobic diatoms	A_MESO	Van Dam and others (1994)
	Relative abundance of α -meso- to polysaprobic diatoms	AP	Van Dam and others (1994)
	Relative abundance of polysaprobic diatoms		Van Dam and others (1994)

Appendix 1. Algal metrics commonly included in a multi-metric index. A total of 124 metrics are grouped into 8 categories. Abbreviations: TN, total nitrogen; TP, total phosphorous; µS/cm, microsiemens per centimeter; mg/L, milligrams per liter; %, percent; <, less than; >, greater than.—Continued.

General category (total number of metrics per category)	Most common algal metrics	Algal metric abbreviation (if used in this report)	Source
Organic pollution indicators (11) —continued	Sum of relative abundances of α-mesosaprobic, α -meso- to polysaprobic, and polysaprobic diatoms		Van Dam and others (1994)
	Relative abundance of diatoms requiring high dissolved oxygen (100% saturation)	O_h	Van Dam and others (1994)
	Relative abundance of diatoms requiring fairly high dissolved oxygen (above 75% saturation)		Van Dam and others (1994)
	Relative abundance of diatoms requiring moderate dissolved oxygen (above 50% saturation)		Van Dam and others (1994)
	Relative abundance of diatoms tolerating low dissolved oxygen (30% saturation)	O_l	Van Dam and others (1994)
	Relative abundance of diatoms tolerating very low dissolved oxygen (about 10% saturation)	O_vl	Van Dam and others (1994)
Water ionic content indicators (15)	Relative abundance of acidobiontic diatoms (pH optima <5.5)		Van Dam and others (1994)
	Relative abundance of acidophilous diatoms (mainly occurring at pH <7)		Van Dam and others (1994)
	Relative abundance of circumneutral diatoms (mainly occurring at pH about 7)		Van Dam and others (1994)
	Relative abundance of alkalibiontic diatoms (exclusively occurring at pH>7)		Van Dam and others (1994)
	Relative abundance of alkaliphilous diatoms (mainly occurring at pH >7)		Van Dam and others (1994)
	Relative abundance of diatoms indicating fresh water		Van Dam and others (1994)
	Relative abundance of diatoms indicating fresh to brackish water	FB	Van Dam and others (1994)

Appendix 1. Algal metrics commonly included in a multi-metric index. A total of 124 metrics are grouped into 8 categories. Abbreviations: TN, total nitrogen; TP, total phosphorous; µS/cm, microsiemens per centimeter; mg/L, milligrams per liter; %, percent; <, less than; >, greater than.—Continued.

General category (total number of metrics per category)	Most common algal metrics	Algal metric abbreviation (if used in this report)	Source
Water ionic content indicators (15)—continued	Relative abundance of diatoms indicating brackish to fresh water	BF	Van Dam and others (1994)
	Relative abundance of diatoms indicating brackish water		Van Dam and others (1994)
	Relative abundance of diatoms with conductivity optima below 200 µS/cm		Potapova and Charles (2003)
	Relative abundance of diatoms with conductivity optima above 400 µS/cm	CONDh	Potapova and Charles (2003)
	Relative abundance of diatoms with chloride optima below 8 mg/L	CHLl	Potapova and Charles (2003)
	Relative abundance of diatoms with chloride optima above 17 mg/L	CHLh	Potapova and Charles (2003)
	Relative abundance of diatoms with sulfate optima below 8 mg/L	SULFl	Potapova and Charles (2003)
	Relative abundance of diatoms with sulfate optima above 24 mg/L	SULFh	Potapova and Charles (2003)
Diatoms indicating general impairment (36)	Relative abundance of most sensitive diatoms		Kentucky Division of Water (1993)
	Relative abundance of sensitive diatoms		Kentucky Division of Water (1993)
	Relative abundance of tolerant diatoms		Kentucky Division of Water (1993)
	Relative abundance of most tolerant diatoms		Kentucky Division of Water (1993)
	Relative abundance of very tolerant diatoms		Lange_Bertalot (1997)
	Relative abundance of tolerant diatoms, 2a category		Lange_Bertalot (1997)
	Relative abundance of tolerant diatoms, 2b category		Lange_Bertalot (1997)
	Relative abundance of less tolerant diatoms, 3a category		Lange_Bertalot (1997)

Appendix 1. Algal metrics commonly included in a multi-metric index. A total of 124 metrics are grouped into 8 categories. Abbreviations: TN, total nitrogen; TP, total phosphorous; µS/cm, microsiemens per centimeter; mg/L, milligrams per liter; %, percent; <, less than; >, greater than.—Continued.

General category (total number of metrics per category)	Most common algal metrics	Algal metric abbreviation (if used in this report)	Source
Diatoms indicating general impairment (36)—continued	Relative abundance of less tolerant diatoms, 3b category		Lange_Bertalot (1997)
	Relative abundance of most tolerant diatoms	Bt	Bahls (1993)
	Relative abundance of less tolerant diatoms		Bahls (1993)
	Relative abundance of sensitive diatoms		Bahls (1993)
	Relative abundance of diatoms associated with reference sites at regional scale	DiaR	This study
	Relative abundance of diatoms associated with disturbed sites at regional scale	DiaD	This study
	Ratio: DiaR/(DiaR+DiaD)	Dia	This study
	Relative abundance of diatoms associated with reference sites at national scale	DiaR_n	This study
	Relative abundance of diatoms associated with disturbed sites at national scale	DiaD_n	This study
	Ratio: DiaR_n/(DiaR_n+DiaD_n)	Dia_n	This study
	Relative density of non-diatom algae associated with reference sites at regional scale	SAR	This study
	Relative density of non-diatom algae associated with disturbed sites at regional scale	SAD	This study
	Ratio: SAR/(SAR+SAD)		This study
	Relative density of non-diatom algae associated with reference sites at national scale		This study
	Relative density of non-diatom algae associated with disturbed sites at national scale	SAD_n	This study

Appendix 1. Algal metrics commonly included in a multi-metric index. A total of 124 metrics are grouped into 8 categories. Abbreviations: TN, total nitrogen; TP, total phosphorous; µS/cm, microsiemens per centimeter; mg/L, milligrams per liter; %, percent; <, less than; >, greater than.—Continued.

General category (total number of metrics per category)	Most common algal metrics	Algal metric abbreviation (if used in this report)	Source
Diatoms indicating general impairment (36)—continued	Ratio: SAR_n /(SAR_n + SAD_n)		This study
	Cell density of non-diatom algae associated with reference sites at regional scale		This study
	Cell density of non-diatom algae associated with disturbed sites at regional scale		This study
	Ratio: SARcd/(SARcd+SADcd)		This study
	Cell density of non-diatom algae associated with reference sites at national scale		This study
	Cell density of non-diatom algae associated with disturbed sites at regional and national scale		This study
	Ratio: SARcd_n/(SARcd_n+SADcd_n)		This study
	Biovolume of non-diatom algae associated with reference sites at regional scale		This study
	Biovolume of non-diatom algae associated with disturbed sites at regional scale		This study
	Ratio: SARb/(SARb+SADb)		This study
	Biovolume of non-diatom algae associated with reference sites at national scale		This study
	Biovolume of non-diatom algae associated with disturbed sites at national scale		This study
	Ratio: SARb_n/(SARb_n+SADb_n)		This study

Appendix 2. Algal taxa associated with reference (+) or disturbed (-) sites. Taxa were included in this list as a result of two analyses. First, indicator species analysis (Dufrêne and Legendre, 1997) determined species indicator values in groups of reference or disturbed sites. Taxa with indicator values greater than 5 (P<0.1) are listed here and their indicator values are shown. Second, the canonical correspondence analysis of species relative abundance data and a single constraining variable "reference" or "disturbed" site was carried out. The taxa with extreme (upper or lower quartile) scores along the first Canonical Correspondence Analyses axis were included in the list and marked by asterisks (*). All analyses were carried out using the five regional datasets and the national dataset. WM, Western Mountains; CWP, Central and Western Plains; GN, Glaciated North; EH, Eastern Highlands; SP, Southeastern Plains; NAT, national dataset.

Species or morphological category	WM	CWP	GN	EH	SP	NAT
Diatoms, followed by source(s)						
Achnanthes conspicua Mayer		-19	-38	-14	-20*	-21*
Achnanthes lemmermannii Hustedt		+20*				+*
Achnanthes subhudsonis var. *kraeuselii* (Cholnoky) Cholnoky	-8*					
Achnanthidium deflexum (Reimer) Kingston	+24	+10	+45*	+55		+29
Achnanthidium exiguum (Grunow) Czarnecki		-16				
Achnanthidium minutissimum (Kützing) Czarnecki	+51		+77		+55	+54
Achnanthidium rivulare Potapova et Ponader			+39	+26*		+18
Adlafia minuscula (Grunow) Lange-Bertalot				+*		
Amphora inariensis Krammer			-30	-23*	-26*	-22*
Amphora montana Krasske				-20*		-10
Amphora ovalis (Kützing) Kützing		-12		-*	-12	-10*
Amphora pediculus (Kützing) Grunow	-50*	-50	-78	-65		-51*
Amphora veneta Kützing				-*		
Aulacoseira ambigua (Grunow) Simonsen			-*			
Aulacoseira granulata (Ehrenberg) Simonsen		-11	-*			-11*
Aulacoseira italica (Ehrenberg) Simonsen	-33*					
Aulacoseira muzzanensis (Meister) Krammer						-*
Bacillaria paradoxa Gmelin	-22*			-*		
Biremis circumtexta (Meister ex Hustedt) Lange-Bertalot et Witkowski		-*				-*
Brachysira brebissonii Ross					+18	+6
Brachysira microcephala (Grunow) Compère				+*	+31	+12
Caloneis amphisbaena (Bory) Cleve			-*			
Caloneis bacillum (Grunow) Cleve		+34	-39*	-30		
Caloneis hyalina Hustedt					+28*	
Caloneis silicula (Ehrenberg) Cleve					+11*	
Chamaepinnularia mediocris (Krasske) Lange-Bertalot					+9*	
Cocconeis fluviatilis Wallace					-*	
Cocconeis neodiminuta Krammer				-14*		
Cocconeis neothumensis Krammer			+30			+8
Cocconeis pediculus Ehrenberg		+42			-24*	
Cocconeis placentula Ehrenberg		+69			-50	
Cocconeis placentula var. *pseudolineata* Geitler		+*				
Craticula molestiformis (Hustedt) Lange-Bertalot		+35			-13*	
Ctenophora pulchella (Ralfs ex Kützing) Williams et Round		-*	-11		-*	-7*
Ctenophora pulchella var. *lacerata* (Hustedt) Bukhtiyarova					-*	-*
Cyclostephanos dubius (Frick) Round		-10				
Cyclotella atomus Hustedt			-17*			

Appendix 2. Algal taxa associated with reference (+) or disturbed (-) sites. Taxa were included in this list as a result of two analyses. First, indicator species analysis (Dufrêne and Legendre, 1997) determined species indicator values in groups of reference or disturbed sites. Taxa with indicator values greater than 5 (P<0.1) are listed here and their indicator values are shown. Second, the canonical correspondence analysis of species relative abundance data and a single constraining variable “reference” or “disturbed” site was carried out. The taxa with extreme (upper or lower quartile) scores along the first Canonical Correspondence Analyses axis were included in the list and marked by asterisks (*). All analyses were carried out using the five regional datasets and the national dataset. WM, Western Mountains; CWP, Central and Western Plains; GN, Glaciated North; EH, Eastern Highlands; SP, Southeastern Plains; NAT, national dataset.—Continued.

Species or morphological category	WM	CWP	GN	EH	SP	NAT
Diatoms, followed by source(s)—continued						
Cyclotella meneghiniana Kützing	-52*	-35	-50	-18*		-30
Cyclotella ocellata Pantocsek						-*
Cyclotella pseudostelligera Hustedt	-22*				-15*	
Cyclotella stelligera (Cleve et Grunow) Van Heurck		-11*				
Cymatopleura solea (Brébisson) Smith				-12*		
Cymbella affinis Kützing		+23	+54	+51		+33
Cymbella cistula (Ehrenberg) Kirchner			+14*			+*
Cymbella cymbiformis Agardh				+23*		+6
Cymbella delicatula Kützing			+17*	+52*		+16
Cymbella hustedtii Krasske				+17*		+*
Cymbella mesiana Cholnoky					+11	
Cymbella naviculiformis Auerswald ex Héribaud					+19	
Cymbella subcuspidata Krammer					+11*	
Cymbella subturgidula Krammer			+17*			+7*
Cymbella tumida (Brébisson ex Kützing) Van Heurck			+30			
Cymbellonitzschia diluviana Hustedt			+*			
Denticula elegans Kützing				+10*		+*
Diadesmis confervacea Kützing		-10	-17		-12*	-11*
Diadesmis contenta (Grunow ex Van Heurck) Mann					+27	
Diatoma mesodon (Ehrenberg) Kützing	+21*	+10*				+*
Diatoma vulgaris Bory	-23	+29			-15*	
Encyonema auerswaldii Rabenhorst						+7
Encyonema lunatum (Smith) Van Heurck					+29	+10
Encyonema minutum (Hilse) Mann		+23	+71		+52	+46
Encyonema muelleri (Hustedt) Mann		+12*				+*
Encyonema prostratum (Berkeley) Kützing			+14*			
Encyonema silesiacum (Bleisch) Mann	-20*			+20	+34	
Encyonopsis cesatii (Rabhenhorst) Krammer	+13*					
Encyonopsis microcephala (Grunow) Krammer	+13*	+9	+23	+25*		+16
Epithemia reichelti var. 1 ANS OZRK				+*		
Epithemia sorex Kützing		+17*				+9
Epithemia turgida (Ehrenberg) Kützing		+11		+10*		+7
Eunotia bilunaris (Ehrenberg) Mills					+49	+14
Eunotia exigua (Brébisson ex Kützing) Rabenhorst					+26	
Eunotia flexuosa Brébisson ex Kützing					+33	+9
Eunotia formica Ehrenberg					+30	

Appendix 2. Algal taxa associated with reference (+) or disturbed (-) sites. Taxa were included in this list as a result of two analyses. First, indicator species analysis (Dufrêne and Legendre, 1997) determined species indicator values in groups of reference or disturbed sites. Taxa with indicator values greater than 5 (P<0.1) are listed here and their indicator values are shown. Second, the canonical correspondence analysis of species relative abundance data and a single constraining variable “reference” or “disturbed” site was carried out. The taxa with extreme (upper or lower quartile) scores along the first Canonical Correspondence Analyses axis were included in the list and marked by asterisks (*). All analyses were carried out using the five regional datasets and the national dataset. WM, Western Mountains; CWP, Central and Western Plains; GN, Glaciated North; EH, Eastern Highlands; SP, Southeastern Plains; NAT, national dataset.—Continued.

Species or morphological category	WM	CWP	GN	EH	SP	NAT
Diatoms, followed by source(s)—continued						
Eunotia incisa Smith ex Gregory			+*		+29	
Eunotia minor (Kützing) Grunow					+17	+*
Eunotia monodon Ehrenberg					+13*	+*
Eunotia muscicola var. *tridentula* Nörpel et Lange-Bertalot					+11*	
Eunotia naegelii Migula					+33*	+9
Eunotia pectinalis var. *undulata* (Ralfs) Rabenhorst			+*		+33	+10
Eunotia praerupta Ehrenberg					+15*	+*
Eunotia rhomboidea Hustedt					+26	+7
Eunotia sp. 9 NAWQA EAM					+20*	+5*
Eunotia tenella (Grunow) Cleve	+13*					+*
Fallacia monoculata (Hustedt) Mann			-*			
Fallacia omissa (Hustedt) Mann					-10*	-*
Fallacia subhamulata (Grunow) Mann				-12*		
Fallacia tenera (Hustedt) Mann			-*			
Fistulifera pelliculosa (Brébisson ex Kützing) Lange-Bertalot	+13*					
Fragilaria aff. amphicephala ANS NAWQA EAM					+11	
Fragilaria capucina var. *gracilis* (Østrup) Hustedt				+10*	+20	+10
Fragilaria capucina var. *mesolepta* Rabenhorst		-	-15		-9*	-8*
Fragilaria crotonensis Kitton		+	-11*	+10 *		
Fragilaria nanana Lange-Bertalot					+25	+7
Fragilaria pinnata var. *acuminata* Mayer		+*				
Fragilaria pinnata var. *lancettula* (Schumann) Hustedt			+19*			+*
Fragilaria vaucheriae (Kützing) Petersen		+32	+40			+30
Frustulia amphipleuroides (Grunow) Cleve-Euler					+20	+9
Frustulia crassinervia (Brebisson) Lange-Bertalot et Krammer	+13*				+40	+13
Frustulia rhomboides (Ehrenberg) De Toni					+41	+12
Frustulia vulgaris (Thwaites) deToni		-9		-15*		
Frustulia weinholdii Hustedt					+16	
Geissleria acceptata (Hustedt) Lange-Bertalot et Metzeltin	-16*					
Gomphoneis eriense (Grunow) Skvortzow et Meyer		+12*				
Gomphoneis herculeana (Ehrenberg) Cleve		+12*	+12*			+7
Gomphonema acuminatum Ehrenberg				+*	+12	
Gomphonema angustatum (Kützing) Rabenhorst	+23		+30		+43	+29
Gomphonema angustatum var. *intermedia* Grunow				+*		
Gomphonema apuncto Wallace			+*	+21*		+7
Gomphonema gracile Ehrenberg emend Van Heurck	+17*					

Appendix 2. Algal taxa associated with reference (+) or disturbed (-) sites. Taxa were included in this list as a result of two analyses. First, indicator species analysis (Dufrêne and Legendre, 1997) determined species indicator values in groups of reference or disturbed sites. Taxa with indicator values greater than 5 (P<0.1) are listed here and their indicator values are shown. Second, the canonical correspondence analysis of species relative abundance data and a single constraining variable "reference" or "disturbed" site was carried out. The taxa with extreme (upper or lower quartile) scores along the first Canonical Correspondence Analyses axis were included in the list and marked by asterisks (*). All analyses were carried out using the five regional datasets and the national dataset. WM, Western Mountains; CWP, Central and Western Plains; GN, Glaciated North; EH, Eastern Highlands; SP, Southeastern Plains; NAT, national dataset.—Continued.

Species or morphological category	WM	CWP	GN	EH	SP	NAT
Diatoms, followed by source(s)—continued						
Gomphonema insigne Gregory		-12	-*			
Gomphonema intricatum Kützing	-16*	+9				
Gomphonema kobayasii Kociolek et Kingston			-33		-34	
Gomphonema lagenula Kützing		+14				
Gomphonema mehleri Camburn				+13		
Gomphonema micropus Kützing			-11*			-*
Gomphonema minutum (Agardh) Agardh	-42		-55			-22
Gomphonema olivaceoides var. *hutchinsoniana* Patrick			+21			+5
Gomphonema olivaceum (Lyngbye) Kützing	+33	+22			-13*	
Gomphonema parvulum (Kützing) Kützing			+59	-60		
Gomphonema patrickii Kociolek et Stoermer					+12	+*
Gomphonema pumilum (Grunow) Reichardt et Lange-Bertalot	+42*		+45			+18
Gomphonema rhombicum Fricke	+17*					
Gomphonema sp. 32 NAWQA EAM					+20*	+5*
Gomphonema sphaerophorum Ehrenberg				+17*		+7
Gyrosigma acuminatum (Kützing) Rabenhorst				-17*		
Gyrosigma attenuatum (Kützing) Rabenhorst			-*	-20*		-7*
Gyrosigma nodiferum (Grunow) Reimer				-20*		-9*
Gyrosigma obtusatum (Sullivant et Wormley) Boyer				-*		
Hannaea arcus (Ehrenberg) Patrick	+37					+6
Hantzschia amphioxys (Ehrenberg) Grunow		+16				
Hippodonta capitata (Ehrenberg) Lange-Bertalot, Metzeltin et Witkowski	-8		-46	-19*		-19
Hippodonta lueneburgensis (Grunow) Lange-Bertalot, Metzeltin et Witkowski					+11*	
Karayevia clevei (Grunow) Bukhtiyarova		+20*	+40			+17
Karayevia laterostrata (Hantzsch) Bukhtiyarova						-*
Kolbesia ploenensis (Hustedt) Kingston			-11*			
Lemnicola hungarica (Grunow) Round et Basson		-10				
Luticola goeppertiana (Bleisch) Mann				-27*		-12*
Luticola mutica (Kützing) Mann				-*		
Mayamaea atomus (Kützing) Lange-Bertalot	-21*		-17*		-19*	
Mayamaea atomus var. *permitis* (Hustedt) Lange-Bertalot		-10			-*	-6*
Melosira varians Agardh	-55*			-34		-31
Meridion circulare (Greville) Agardh	+17			-*		
Navicula aff. subminuscula ANS NAWQA EAM Manguin				-23*		-6*
Navicula angusta Grunow					+16	

Appendix 2. Algal taxa associated with reference (+) or disturbed (-) sites. Taxa were included in this list as a result of two analyses. First, indicator species analysis (Dufrêne and Legendre, 1997) determined species indicator values in groups of reference or disturbed sites. Taxa with indicator values greater than 5 (P<0.1) are listed here and their indicator values are shown. Second, the canonical correspondence analysis of species relative abundance data and a single constraining variable “reference” or “disturbed” site was carried out. The taxa with extreme (upper or lower quartile) scores along the first Canonical Correspondence Analyses axis were included in the list and marked by asterisks (*). All analyses were carried out using the five regional datasets and the national dataset. WM, Western Mountains; CWP, Central and Western Plains; GN, Glaciated North; EH, Eastern Highlands; SP, Southeastern Plains; NAT, national dataset.—Continued.

Species or morphological category	WM	CWP	GN	EH	SP	NAT
Diatoms, followed by source(s)—continued						
Navicula antonii Lange-Bertalot		+8	-18			
Navicula arvensis Hustedt		+11				
Navicula canalis Patrick		+8		-*		
Navicula capitatoradiata Germain		+31		-48	-21	
Navicula caterva Hohn et Hellermann	-26*					
Navicula cf. acceptata CODY Hustedt		+10*				
Navicula cf. kriegerii NAWQA KM Krasske					+11*	
Navicula cryptocephala Kützing	-52*	-15			+49	
Navicula cryptotenella Lange-Bertalot ex Krammer et Lange-Bertalot			-51	-49	-30	-38
Navicula erifuga Lange-Bertalot			-*	-*		
Navicula exilis Kützing			-21			-8
Navicula festiva Krasske					+9	
Navicula germainii Wallace	-19*		-37	-17		
Navicula gregaria Donkin	-55*	-36	-66		-42	-41*
Navicula harderii Hustedt			-14*			-*
Navicula ingenua Hustedt		-16*				-7*
Navicula lanceolata (Agardh) Ehrenberg		+17	-*		-19*	
Navicula lateropunctata Wallace					+25	+7
Navicula longicephala Hustedt					+18	+
Navicula menisculus Schumann		+28		-21		
Navicula menisculus var. *upsaliensis* (Grunow) Grunow			-*			
Navicula meniscus Schumann			-18			-*
Navicula minima Grunow		-65	-70	-59	-50	-58*
Navicula notha Wallace					+45	+14
Navicula perminuta Grunow	-21*					
Navicula pseudoventralis Hustedt		-*				
Navicula recens Lange-Bertalot		-30	-*	-*		-10
Navicula reichardtiana Lange-Bertalot			-*			
Navicula rostellata Kützing	-31*		-18	-23		
Navicula schroeteri var. *escambia* Patrick				-20*		
Navicula sp. 47 NAWQA HAMSHER			-29*			-7*
Navicula subminuscula Manguin					-30*	-17
Navicula submuralis Hustedt					-18*	-6*
Navicula symmetrica Patrick		+19	-23			-10
Navicula tantula Hustedt	-19*	-13				-9*
Navicula tenelloides Hustedt			-22			

Appendix 2. Algal taxa associated with reference (+) or disturbed (-) sites. Taxa were included in this list as a result of two analyses. First, indicator species analysis (Dufrêne and Legendre, 1997) determined species indicator values in groups of reference or disturbed sites. Taxa with indicator values greater than 5 ($P<0.1$) are listed here and their indicator values are shown. Second, the canonical correspondence analysis of species relative abundance data and a single constraining variable "reference" or "disturbed" site was carried out. The taxa with extreme (upper or lower quartile) scores along the first Canonical Correspondence Analyses axis were included in the list and marked by asterisks (*). All analyses were carried out using the five regional datasets and the national dataset. WM, Western Mountains; CWP, Central and Western Plains; GN, Glaciated North; EH, Eastern Highlands; SP, Southeastern Plains; NAT, national dataset.—Continued.

Species or morphological category	WM	CWP	GN	EH	SP	NAT
Diatoms, followed by source(s)—continued						
Navicula tripunctata (Müller) Bory		+37	-69	-56		-36
Navicula trivialis Lange-Bertalot		-20	-38	-28*		-19*
Navicula veneta Kützing			-21			
Navicula viridula (Kützing) Kützing emend. Van Heurck				-17*	*-	-8*
Navicula viridulacalcis (Hustedt) Lange-Bertalot					+17	
Navicula wallacei Reimer					+9*	
Nitzschia amphibia Grunow	-19*	-52	-64	-61	-58	-55*
Nitzschia archibaldii Lange-Bertalot	-7		+31	-19		
Nitzschia capitellata Hustedt		+22	-17	-*	+13	
Nitzschia clausii Hantzsch		-8*				-5*
Nitzschia desertorum Hustedt		-9				-*
Nitzschia dissipata (Kützing) Grunow			-60	-43	-31	-39
Nitzschia dissipata var. *media* (Hantzsch) Grunow		+14	-*		+11*	
Nitzschia fonticola Grunow		+30				
Nitzschia frustulum (Kützing) Grunow		-41				
Nitzschia gracilis Hantzsch ex Rabenhorst					+35*	+9
Nitzschia heufleriana Grunow			-11*			
Nitzschia inconspicua Grunow	-64*	-37	-48		-49*	-40*
Nitzschia intermedia Hantzsch ex Cleve et Grunow		+23			+14	
Nitzschia linearis (Agardh ex Smith) Smith	-41*			-15		-11
Nitzschia linearis var. *tenuis* (Smith) Grunow ex Cleve et Grunow				-*		
Nitzschia nana Grunow ex Van Heurck					-*	
Nitzschia palea (Kützing) Smith	-35*	-38		-36		-35
Nitzschia palea var. *debilis* (Kützing) Grunow		-31	-31			-14
Nitzschia palea var. *tenuirostris* Grunow			-*			
Nitzschia paleacea Grunow ex Van Heurck		+12*				
Nitzschia pusilla Grunow	-33*					
Nitzschia recta Hantzsch ex Rabenhorst	-28*	+16	-27		+25	
Nitzschia reversa Smith		-7				
Nitzschia sigmoidea (Nitzsch) Ehrenberg		+25				+7
Nitzschia sinuata var. *delognei* (Grunow) Lange-Bertalot					-9	-*
Nitzschia sinuata var. *tabellaria* (Grunow) Grunow			+12*	-*		
Nitzschia sociabilis Hustedt			-20	-17		-11
Nitzschia solita Hustedt			-*			
Nitzschia subtilis Grunow					+25	
Nupela carolina Potapova et Clason					+10	

Appendix 2. Algal taxa associated with reference (+) or disturbed (-) sites. Taxa were included in this list as a result of two analyses. First, indicator species analysis (Dufrêne and Legendre, 1997) determined species indicator values in groups of reference or disturbed sites. Taxa with indicator values greater than 5 (P<0.1) are listed here and their indicator values are shown. Second, the canonical correspondence analysis of species relative abundance data and a single constraining variable “reference” or “disturbed” site was carried out. The taxa with extreme (upper or lower quartile) scores along the first Canonical Correspondence Analyses axis were included in the list and marked by asterisks (*). All analyses were carried out using the five regional datasets and the national dataset. WM, Western Mountains; CWP, Central and Western Plains; GN, Glaciated North; EH, Eastern Highlands; SP, Southeastern Plains; NAT, national dataset.—Continued.

Species or morphological category	WM	CWP	GN	EH	SP	NAT
Diatoms, followed by source(s)—continued						
Nupela sp. 1 ANS NEW JERSEY KCP					-*	-*
Opephora cf. schwartzii NAWQA EAM (Grunow) Petit					+15*	
Opephora martyi Héribaud		+22*	+28			+11
Pinnularia acidophila Hofmann et Krammer					+19	+
Pinnularia divergens Smith					+24	+6
Pinnularia gibba Ehrenberg					+14	
Pinnularia interrupta Smith					+16	
Pinnularia mesogongyla Ehrenberg						+*
Planothidium apiculatum (Patrick) Lange-Bertalot					+28*	+8*
Planothidium frequentissimum (Lange-Bertalot) Lange-Bertalot		+40			-33	
Planothidium lanceolatum (Brébisson ex Kützing) Lange-Bertalot	-48	-38	-39		-54*	-34
Planothidium peragalli Brun et Héribaud			+*			
Planothidium robustum (Hustedt) Lange-Bertalot		+12*				
Pleurosira laevis (Ehrenberg) Compère		+19				
Psammothidium grischunum *fo. daonensis* (Lange-Bertalot) Bukhtiyarova et Round					+13*	+*
Psammothidium helveticum (Hustedt) Bukhtiyarova et Round					+9*	
Pseudostaurosira brevistriata (Grunow) Williams et Round	-15			-*	-17	
Reimeria sinuata (Gregory) Kociolek et Stoermer	+45	+40			-30*	
Reimeria sinuata fo. antiqua Kociolek et Stoermer		+10*	-11			
Rhoicosphenia abbreviata (Agardh) Lange-Bertalot	-55	-52	-86	-67	-50*	-62*
Rhopalodia gibba (Ehrenberg) Müller		+8				+*
Sellaphora laevissima (Kützing) Mann					+20*	
Sellaphora pupula (Kützing) Meresckowsky		+25		-35*	-32	
Sellaphora seminulum (Grunow) Mann	-8	-27	-29	-29	-34	-31
Sellaphora sp. 6 NAWQA EAM					+9*	
Simonsenia delognei (Grunow) Lange-Bertalot		+19	-36			-9
Stauroneis kriegeri Patrick		-7				
Stauroneis smithii Grunow			-11*		+18	
Stauroneis thermicola (Petersen) Lund					+13*	+*
Staurosira construens Ehrenberg	-22	+15			-12	
Staurosira construens var. *binodis* (Ehrenberg) Hamilton		+20				
Staurosira construens var. *venter* (Ehrenberg) Hamilton		+39	+37		-31	
Staurosirella leptostauron (Ehrenberg) Williams et Round		+17*	+23	-22*		
Staurosirella pinnata (Ehrenberg) Williams et Round	+27	+28	+53			+23
Stephanodiscus niagarae Ehrenberg			-11			-*

Appendix 2. Algal taxa associated with reference (+) or disturbed (-) sites. Taxa were included in this list as a result of two analyses. First, indicator species analysis (Dufrêne and Legendre, 1997) determined species indicator values in groups of reference or disturbed sites. Taxa with indicator values greater than 5 (P<0.1) are listed here and their indicator values are shown. Second, the canonical correspondence analysis of species relative abundance data and a single constraining variable "reference" or "disturbed" site was carried out. The taxa with extreme (upper or lower quartile) scores along the first Canonical Correspondence Analyses axis were included in the list and marked by asterisks (*). All analyses were carried out using the five regional datasets and the national dataset. WM, Western Mountains; CWP, Central and Western Plains; GN, Glaciated North; EH, Eastern Highlands; SP, Southeastern Plains; NAT, national dataset.—Continued.

Species or morphological category	WM	CWP	GN	EH	SP	NAT
Diatoms, followed by source(s)—continued						
Stephanodiscus tenuis Hustedt						
Surirella angusta Kützing	-7					
Surirella biseriata Brébisson		+10*				
Surirella brebissonii Krammer et Lange-Bertalot		+15*	-*			
Surirella brebissonii var. *kuetzingii* Krammer et Lange-Bertalot		-*	-17*			-6*
Surirella minuta Brébisson	-11*		-*			-12
Synedra acus Kützing					+22	+8
Synedra delicatissima Smith					+11	
Synedra ulna (Nitzsch) Ehrenberg	+39		+56			+40
Tabellaria flocculosa (Roth) Kützing	+17*		+*	+*	+40	+15
Tabularia fasciculata (Agardh) Williams et Round			-*			
Tabularia tabulata (Agardh) Snoeijs		-10				
Thalassiosira weissflogii (Grunow) Fryxell et Hasle			-*	-*		
Tryblionella apiculata Gregory			-16	-19*	-10*	-12
Tryblionella hungarica (Grunow) Mann			-15			
Green algae						
Ankistrodesmus falcatus (Corda) Ralfs	-10*	+17			-13	
Oedogonium sp.	-24*	+12*				
Protoderma viride Kützing						-*
Scenedesmus acuminatus (Lagerheim) Chodat			-*			
Scenedesmus acutus Meyen				-11		
Scenedesmus ecornis (Ralfs) Chodat	-10*					
Scenedesmus quadricauda (Turpin) Brébisson		-22*		-12		-13
Scenedesmus spinosus Chodat		-9*				
Unknown Chlorophyte flagellate		+12*				
Cyanobacteria						
Anabaena sp.	+21*					
Blennothrix brebissonii (Kützing ex Gomont) Anagnostidis et Komárek	-20*	+31*	+17*	-35		
Calothrix fusca (Kützing) Bornet et Flahault						+*
Calothrix parietina (Nägeli) Thuret	-26*	+20*	+26*			+10
Calothrix sp.	+21*			+43	-19	
Chroococcus limneticus Lemmermann		-*				
Heteroleibleinia sp.			-23*			-
Homoeothrix janthina (Bornet et Flahault) Starmach	-31*		+30		-24	
Jaaginema *pseudogeminatum* (Schmidle) Anagnostidis et Komárek			-17			-

Appendix 2. Algal taxa associated with reference (+) or disturbed (-) sites. Taxa were included in this list as a result of two analyses. First, indicator species analysis (Dufrêne and Legendre, 1997) determined species indicator values in groups of reference or disturbed sites. Taxa with indicator values greater than 5 (P<0.1) are listed here and their indicator values are shown. Second, the canonical correspondence analysis of species relative abundance data and a single constraining variable "reference" or "disturbed" site was carried out. The taxa with extreme (upper or lower quartile) scores along the first Canonical Correspondence Analyses axis were included in the list and marked by asterisks (*). All analyses were carried out using the five regional datasets and the national dataset. WM, Western Mountains; CWP, Central and Western Plains; GN, Glaciated North; EH, Eastern Highlands; SP, Southeastern Plains; NAT, national dataset.—Continued.

Species or morphological category	WM	CWP	GN	EH	SP	NAT
Cyanobacteria)—continued						
Leptolyngbya sp.					+29	+8
Leptolyngbya tenuis (Gomont) Anagnostidis et Komárek					+11	
Lyngbya sp.	-10*	+*				
Oscillatoria angustissima West et West			-11*			
Oscillatoria fremyii De Toni		-*				
Oscillatoria sp. 1?	+29*	-8*				-*
Phormidium granulatum		-*				
Phormidium laetevirens			-11*			-
Planktolyngbya subtilis (Lemmermann) Anagnostidis et Komárek			-17*			-*
Schizothrix friesii (Agardh) Gomont				-11		
Unknown Cyanophyte coccoid (1-3µ)		+15*		+32		+14
Unknown Cyanophyte coccoid (3-5µ)	+25*	+15				
Unknown Cyanophyte Oscillatoriales without sheath	+63*	+34	+30			+29
Unknown Cyanophyte Oscillatoriales with sheath	+33*		+28	+38		+20
Red algae						
Unknown Rhodophyte *Florideophycidae (chantransia)*				-34		
Unknown alga			+23			