New Jersey Periphyton Bioassessment Development Projects

Trophic Diatom Inference Models and Index Development for New Jersey Wadeable Streams

FINAL REPORTS
(2000 – 2005)

Submitted to
Thomas Belton, Project Manager

New Jersey Department of Environmental Protection
Division of Science, Research and Technology
401 East State Street, PO Box 409
Trenton New Jersey 08625

by
Karin Ponader and Donald Charles

Patrick Center for Environmental Research
The Academy of Natural Sciences
1900 Benjamin Franklin Parkway
Philadelphia, PA 19103-1195

Understanding the Relationship Between Natural Conditions and Loadings on Eutrophication: Algal Indicators of Eutrophication for New Jersey Streams

Final Report Year 4

Submitted to the

New Jersey Department of Environmental Protection
Division of Science, Research and Technology

by Karin Ponader and Donald Charles

Patrick Center for Environmental Research
The Academy of Natural Sciences
1900 Benjamin Franklin Parkway
Philadelphia, PA 19103-1195

November 2004

Executive Summary

The New Jersey Department of Environmental Protection has a need for development of algal indicators of stream and river eutrophication. This study evaluates the use of diatoms as indicators of nutrient conditions in the New Jersey streams and rivers of the Inner Coastal Plain. The results presented here are part of an ongoing study, initiated in July 2000. Years 1 through 3 were limited to development of algal indicators in northern and central New Jersey (Piedmont, Highlands and the Ridge and Valley physiographic provinces) (Ponader and Charles 2004). The fourth year of the study extended the development of indicators to the Inner Coastal Plain. Data from 28 sites studied during year 4 were used to develop and test additional indicator metrics to be used in the Inner Coastal Plain of southern New Jersey. All sites were sampled in 2003 for water chemistry, diatoms and algal biomass using artificial substrates (diatometers). Measurements of algal biomass, algal species composition, physical stream conditions and water chemistry were used to develop models and metrics for quantifying algal biomass and inferring nutrient concentrations from diatoms.

The following summarizes findings of the research presented in this report:

- The relationships between algal biomass measures (Chl *a*) and nutrient concentrations were not significant and weak, based on Spearman's rank-order correlations that included data from all the sites. However, variations in contents of Chl *a* can be explained through a combination of conductivity, NH3-N, %open canopy cover and TN.

- A total of 294 diatom taxa were found in the samples. Sample assemblages showed high species diversity and generally low average relative abundance of species. The most common species had moderate TP and TN WA-optima.

- Multivariate analysis of species and environmental variables showed that pH, total phosphorus (TP) and basin size explain significant differences in diatom assemblage composition. This finding provides statistical justification for developing diatom-based models and indices for these variables. No other nutrient variable than TP explained a significant amount of variation in diatom species composition.

- The nutrient inference model and index development for the Inner Coastal Plain was only possible for TP, and showed moderate performance when tested using cross validation. The model for inferring TP ($r^2_{(apparent)}$= 0.97; RMSEP $_{(jack)}$ = 0.31 log µg TP) developed using the complete 2003 dataset (n=25), showed low predictive ability with a jackknifed r^2=0.34. This finding stands in contrast to the results of previous study years (Piedmont, Highlands and Valley & Ridge), where TP and TN inference model development has been successful, using a large dataset (n=91). The decrease in model performance between apparent and tested r^2 in the Inner Coastal Plain dataset can be explained mainly by an uneven distribution of samples along the nutrient gradient, due the low number of samples (n=25). Therefore, we expect significant improvement of the TP model performance, when filling the data gaps using a larger dataset (Inner and Outer Coastal Plain datasets combined) during data analysis.

- Development of algal indicators of nutrient conditions in the Inner Coastal Plain is difficult, due to the overriding influence of pH and conductivity. Our results show that in the Inner Coastal Plain all three methods applied a) biomass metrics, b) inference model development and c) diversity metrics, showed limited applicability as measures of nutrient conditions.

- Other factors might be responsible for low model and metric performance in the Inner Coastal Plain rivers, compared with the northern and central part of the state. We will further investigate if the use of artificial substrate instead of natural substrate may have an influence on concentrations of algal biomass, diatom species composition and model performances. To answer this question more in detail we will compare inference model results based on samples that were collected from natural and artificial substrates during fieldwork in 2004 (study year 5- Outer Coastal Plain). We anticipate that the analysis of the combined Inner and Outer Coastal Plain datasets will provide more insight into these questions.

List of Tables

List of Figures

List of Abbreviations

AFDM	ash free dry mass
AMNET	Ambient Biomonitoring Network
ANSP	The Academy of Natural Sciences, Philadelphia
boot	bootstrapped
CCA	Canonical Correspondence Analysis
Chl *a*	chlorophyll *a*
DCA	Detrended Correspondence Analysis
IBI	index of biotic integrity
NAWQA	National Water-Quality Assessment
NJ DEP	New Jersey Department of Environmental Protection
O-P	orthophosphate
PCA	Principal Component Analysis
PCER	Patrick Center for Environmental Research
RBA	rapid bioassessment
RMSE	root mean square error
RMSEP	root mean square error of prediction
SWQS	Surface Water Quality Standards
TN	total nitrogen
TKN	total Kjeldahl nitrogen
TP	total phosphorus
USGS	United States Geological Survey
U.S. EPA	United States Environmental Protection Agency
WA	weighted averaging

Table of Contents

1. Introduction

The New Jersey Department of Environmental Protection has a need for development of algal indicators of stream and river eutrophication. These indicators will be used to assess relationships between extant water quality criteria (e.g., phosphorus and nitrogen concentrations) and overt signs of eutrophication. They will be applied in a regulatory context as secondary criteria for identifying nutrient impairment. These indicators should be based on an understanding of algal dynamics in New Jersey streams, and be able to distinguish between situations where nutrient concentrations are high due to natural environmental conditions and those that result from anthropogenic influences. Protocols are needed that describe procedures for sample collection and processing, analysis and presentation of data, and interpretation of results. For a detailed review of the problematic and study rationale and approach, please consult the year 3 final report (Ponader and Charles 2004) and the draft manuscript submitted for publication (Ponader et al., in review).

This study was initiated in July 2000. Years 1 and 2 of this project were limited to development of algae indicators in the Piedmont physiographic province in New Jersey. During the third year the study was expanded to include sites in the Highlands and the Ridge and Valley physiographic provinces. Data from sites studied during all three years were used to develop and test indicator metrics for the northern part of the State (Ponader et al., in review). The fourth year of the study extended the development of indicators to the Inner Coastal Plain. Data from sites studied during year 4 were used to develop and test additional indicator metrics to be used in the Coastal Plain of southern New Jersey and are presented here.

The fourth year uses the same general approach and specific methods as used in the first three years (Charles et al. 2000, PCER 2000). Nevertheless, because of the different geomorphology of the rivers (sand and clay river bottoms) in the Inner Coastal Plain, significant changes and adjustments in the sampling design and the methods for collection of algal samples were necessary. A new sampling protocol was developed for the use of artificial substrate (Ponader et al. 2003). Main differences in the sampling methods used this year are: a) the use of artificial substrate (diatometers) for collection of both diatom species composition and algal biomass, b) the establishment of only two

reaches instead of three per site (due to high cost of diatometers), c) the measurement of water color as a general indicator of DOC levels, and d) the sampling of water chemistry twice: once at diatometer deployment and once at removal.

2. Study area and sites

This study was conducted in southern New Jersey, in the Inner Coastal Plain ecoregion (Fig. 1). The Inner Coastal Plain is characterized by a generally flat to undulating topography, with unconsolidated sedimentary deposits consisting of cretaceous sand, silt and gravel (Tedrow 1986). Land-use in the Inner Coastal Plain is mainly agricultural and urban. The predominant forest type in the inner coastal Plain is a mix of oak, beech and red maple (Tedrow 1986).

During the late summer of 2003 we sampled 28 AMNET sites, 24 in the Inner Coastal Plain and 4 in the Piedmont (resampling sites) (Fig. 1). Table 1 presents site information including site name and location. The most important physical characteristics, water chemistry and other site information are summarized in Table 2.

3. Methods

3.1. Site selection

3.1.1. Preselection of sites

We worked with NJ DEP staff, mainly Tom Belton, to carefully select a set of study sites. Because a goal was to develop indicators of anthropogenic nutrient increases, it was important to select a suite of sites with relatively similar natural environmental conditions (e.g., geology, geomorphology), but with a wide range of nutrient concentrations. The sites were restricted to the Inner Coastal Plain physiographic province in southern New Jersey and have a limited range of hydrology, morphology and substrate type. We based our selection of sites on chemistry data available from the NJ

and USGS monitoring networks. All sites were part of the State's Ambient Surface Water Monitoring Network (AMNET) (NJDEP 2000). We selected sites with a range of known biological impairments ranging from unimpaired to severely-impaired, based on AMNET macroinvertebrate data collected between 1992 and 1999. We generally used the same selection criteria as in the previous years (Ponader and Charles 2001). An additional criteria used during year four was to avoid sites that were located within the tidal zone of the Delaware basin, for the following reasons: the taxonomic composition of the algae flora could potentially be influenced by a) high salinity, as well as b) by up-/downstream transport of algal specimens associated with tidal water level changes. For all sites, chemistry data were available through the NJ monitoring network program (SW sites, SS stations, or EWQ stations) or through the USGS. In addition Inner Coastal Plain sites, we resampled 4 stations that were sampled in 2000, 2001 and 2002 in the Piedmont province. This gave us a four-year record we could use to compare variability at individual sites among years.

3.1.2. Field reconnaissance

For several reasons, we rejected many potential candidate sites during field reconnaissance. To compensate, we added EWQ stations to increase the numbers of sites sampled. The main reasons for rejecting sites in the Inner Coastal Plain were the following. First, many sites were located in urban/recreational areas, where we expected disturbance or vandalism, a frequent problem with the use of diatometers. Another factor was that dams upstream from the sampling station often impacted the flow regime of the river at the site. Finally, some sites were rather deep (channel-like) or not accessible for fieldwork. Especially during the first week of September (1-5 September, 2004), the deployment of diatometers became more difficult or impossible at many rivers, because of high water levels due to local rainstorm events.

In the end, twenty-eight sites were sampled. Twenty-four of these were diatometer sites located in the Inner Coastal Plain. Four sites in the Piedmont were sampled for the fourth consecutive year; natural rock substrates were sampled.

3.2. Sampling period

Algae samples were collected by D. Winter, E. Hagan and K. Ponader from August 20th through October 8th, 2003. Samples were in late summer, when the influence of nutrient related water quality on algal assemblage composition is greatest. Water chemistry of samples collected during this time is also most directly comparable with sample data from other studies. Diatometers were deployed from 20 August through 11 September and were exposed for a period of between 14 to 18 days. They were removed in consecutive order from September 5 through 26. All diatometers except for the two deployed at site AN694 were collected before Hurricane Isabel, which affected the sampling area during September 18 and 19, 2003. The four resampling sites in the Piedmont were sampled on October 8, 2003. Water levels were generally higher during the summer of 2003, as compared to previous years. Nevertheless, despite increased rainfall and the generally higher water levels, no flood events occurred in the sampling area during the exposure time of the diatometers.

3.3. Sample/data collection

All algae samples were collected using techniques consistent with those used in the USGS NAWQA program (Moulton et al. 2002) and documented in PCER protocols (Charles et al. 2000 and Ponader et al. 2003). Water chemistry was measured twice in the field: once at the time of diatometer deployment and another time when the diatometers were removed. Conductivity, pH and temperature were measured using a OKATON deluxe pH/conductivity meter (model no. WD-35630-60). Samples were taken for laboratory analysis of soluble reactive phosphorus, total phosphorus, nitrate, ammonia, total nitrogen, chloride, total alkalinity, total hardness and conductivity (Velinsky 2000). Results of these analyses will supplement those collected by the NJ DEP. They may better represent conditions near the time that algae samples were collected, and will provide information of the nature and magnitude of variation in water chemistry.

3.4. Algal sample preparation and analysis

A total of 52 diatom samples were treated in the lab at the ANSP. Diatoms were permanently mounted on microscope slides. Per slide, 600 valves were identified to

lowest taxonomic level and counted. Because of the high content of clay in the rivers from which the diatom material was collected, the treatment of the material collected using diatometers and the analysis of the diatom assemblages under the microscope was considerably more time consuming than in previous years. High concentrations of clay in the majority of samples lead to increased debris on the microscope slides, which made diatom identification more difficult. Attempts to decrease clay contents in the samples, by diluting the concentration of the material used also decreased the density of diatom frustules, sometimes below acceptable limits (Charles et al. 2002). Therefore in many cases dilution of the samples led to very sparse diatom concentrations, which in turn increased the analysis time. For these reasons, the diatom sample from only one section (section 1 or 2, depending on the quality of the slide) at each site was analyzed. In total 29 diatom samples were analyzed. 48 samples were analyzed for algal biomass using standard protocols (Velinsky and DeAlteris 2000; Kiry et al. 1999). Fifty-two samples were measured for conductivity, alkalinity, hardness, chloride (Cl^-), dissolved nitrate + nitrite ($NO_3 + NO_2$, here referred to as NO_3-N), total Kjeldahl nitrogen (TKN), dissolved ammonia (NH_3-N), orthophosphate (O-P) and total phosphorus (TP) in the Patrick Center's Geochemistry Section using a Technicon AutoAnalyzer following USGS methods (Fishmann 1993).

3.5. Numerical Analysis

3.5.1. Algal biomass

3.5.1.1. Spearman's rank-order correlation (correlations between nutrients and algal biomass)

Because many of the data used in the correlation analysis were not measured on a continuous scale, we chose to run a Spearman's rank-order correlation using the program Sigma Stat 2.03. Included were data records associated with each of the 47 diatometer biomass samples collected in the Inner Coastal Plain during 2003. These data are from 24 diatometer sites with 2 sections each, except for site 169,where only one section could be sampled because of the depth of the river (e.g. 47 data records in total). The data analyzed

in this analysis were: Chl *a*, measured water chemistry parameters including nutrient measurements (TP, O-P, NH_3-N, NO_3-N, TKN , TN), physical site characteristics and percent open canopy cover and substrate type. In total, 25 variables were used in the analysis.

3.5.1.2. Forward stepwise regression (analysis of principal factors influencing algal biomass).

To help determine the principal factors influencing algal biomass, we examined correlations among algal biomass, nutrients, geomorphology and light conditions, running a Forward Stepwise regression with Sigma Stat 2.03. To reduce skewed distributions, all physical and chemical variables (except pH) were log10 transformed. All data expressed in percentages (land-use, substrate and %open canopy cover) were square-root transformed. The substrate categories were analyzed both separately and combined into different categories. We created two substrate categories, one including all bigger substrate (gravel only) and another combining all smaller and soft substrate (sand, silt and clay).

3.5.2. Diatom species composition

3.5.2.1. Dataset and data transformations

The complete dataset used in the analyses using diatom species composition contained 25 samples including a 1 within-site replicate sample. This sample was a gravel sample taken instead of a diatometer sample. At this site, the diatometer sample from the same site (section 1) and this gravel (section 2) were both included to identify differences between the two substrates/sections. Because of the nature of the diatom assemblages (high diversity and high evenness) most diatom species were included in the analysis, even at low relative abundances. Species were included in the analysis if their maximum relative abundance was = 0.3%. All relative abundance data were square-root transformed. The water chemistry and physical parameters included in exploratory ordinations (PCA and DCA) were pH, conductivity, total alkalinity, total hardness, chloride, TP, O-P, TN, NO_3-N, NH_3-N, chl *a*, basin size, land-use (% forested, %

agriculture, % urban), % open canopy cover, flow estimate (slow, medium, fast), stream width, section length and percent substrate type. The following variables were averages of measurements for the dates diatometers were deployed and removed: pH, conductivity, total alkalinity, total hardness, chloride, TP, TN, TKN, NO3-N, NH3-N, flow estimate. A category of small grain size (SDSTCL) grouped the percentages of the three substrate types: % sand, silt, and clay. Environmental variables were transformed to reduce skewed distributions: all water chemistry and physical variables, except pH, were log10-transformed; all data expressed in percentages (land-use, substrate and open canopy cover) were square root transformed.

3.5.2.2. Ordination analysis

Ordinations were produced using Canoco for Windows (version 4.5, Microcomputer Power, Ithaca) (ter Braak and Šmilauer 2002). Principal components analysis (PCA) was performed to detect major gradients and principal patterns of variation within the environmental variables (ter Braak and Prentice 1988). The environmental variables were centered and standardized. In the same PCA, "outliers" or "rogues" were defined as samples with extreme sample scores on any of the first 4 axes of the PCA of the environmental data (Birks et al. 1990). Extreme sample scores were defined as scores falling above the 95% confidence interval of all sample score means (Winter and Duthie 2000). Detrended correspondence analysis (DCA) with detrending by segments and down-weighting of rare taxa was used to examine patterns in the diatom data, and to determine the maximum amount of variation within the species composition data (ter Braak 1995). We used the gradient length of the main DCA ordination axes to determine whether linear or unimodal techniques were to be applied for modeling the relationship between diatoms and environmental variables (ter Braak and Prentice 1988). Furthermore, DCA was used to determine outliers, e.g., samples that showed extreme sample scores on any of the first 4 axes of the DCA of the species data (Birks et al. 1990). Outliers were screened using the same criteria as applied in PCA. All samples determined as outliers by PCA and DCA were excluded from all subsequent ordinations and from the development of calibration models. A series of correspondence analyses (CA) and canonical correspondence analyses (CCAs) with down-weighting of rare taxa was

performed, in order to determine the variables that independently explained a significant amount of variation in diatom species composition (ter Braak 1995). First, a CA was run with passive environmental variables to determine the strength of the correlations among all 24 environmental variables and all 131 species, as well as to identify variables that were intercorrelated, based on weighted correlations and variance inflation factors (VIFs). Variables with VIFs >5 indicated strong co-linearity among environmental variables and were removed from all subsequent analyses. The data from the remaining 13 variables and all sites were included in a CCA analysis with forward selection in order to identify the minimal number of variables that explained the largest amount of variation in the diatom species data. Unrestricted Monte Carlo permutation tests (999 permutations) were used to assess the statistical significance of each forward selected variable ($p < 0.05$). As a last step, to assess the strength of the relationship between diatom species composition and the forward selected variables, we ran CCAs constrained to 1 variable at a time. A high ratio between the first (constrained) eigenvalue and the second (unconstrained) eigenvalue (e.g., $?_1/?_2 >0.5$) indicated strong influence of these variables on diatom species composition, and justified development of inference models (Bigler and Hall 2002, Winter and Duthie 2000).

3.5.2.3. Species optima and tolerances and inference models

Weighted averaging (WA) regression and calibration techniques were used to calculate diatom species optima and tolerances as well as to develop and test diatom inference models for TP. These were run using the program C^2, version 1.3 (Juggins 2003). Several models were developed, based on WA inverse and classical deshrinking (Birks et al. 1990) and on weighted-averaging partial least square regressions (WA-PLS) (ter Braak and Juggins 1993). Model error estimation was performed by bootstrapping with 1000 cycles and leave-one-out cross validation (jackknifing) (Birks 1995). We selected the best inference model based on a combination of the following characteristics: a) the highest prediction accuracy e.g. the lowest root mean square error of prediction (RMSEP) based on cross-validation; b) the highest coefficient of determination (r^2) between observed and inferred values and c) the minimum number of WA-PLS components (Birks 1995).

3.5.2.4. Diatom Index development

We created diatom TP indices based on values inferred for each diatom sample using the 2-component WA-PLS model. Index values are calculated by multiplying the inferred nutrient values (log_{10}) obtained for each sample by a constant that converts them to a scale from 0-100 (TP Index = 33.33 X inferred log_{10} TP). The 0-100 scales correspond to log_{10} TP from 1 to 1000 µg / L. The ranges for all 3 parameters include all values in our dataset. The indices are meant to provide a continuous measure of biological response to TP and TN. In addition to being indicators of response of species composition, the indices also serve as general indicators of trophic state conditions.

We divided the 0-100 scale into trophic state categories ranging from 1 (low) to 4 (very high) with respect to the range of values in the calibration dataset. The category boundaries are those suggested by Wetzel (2001) for trophic classification of temperate streams, based on the biomass-nutrient relationships established in Dodds et al. (1998). The 4 TP categories established were 1) less than 0.025, 2) 0.025-0.075, 3) 0.075-0.1, and 4) above 0.1 mg / L TP. These correspond to the following log_{10} TP values: 1) less than 1.4, 2) 1.4-1.9, 3) 1.9-2.0, and 4) above 2.0, and to the index scores of 1) 0-47, 2) 47-63, 3) 63-67, and 4) above 67. The lower boundary (0.1 mg / L TP) is also the nutrient criteria used by New Jersey. The technique used for developing indicators is the same as presented in Ponader et al. in review.

3.5.2.5. Diatom Diversity metrics

Diatom diversity indices and other simple metrics were calculated using the full dataset (*n*=25) diatom samples from both years, following Barbour et al. (1999). We used the program Phyco-AIDE version 2 (PCER 2003) to calculate the number of diatom taxa in the sample (# Taxa), the Shannon-Weiner diversity index (S-W Index), the percent of total diatom valves made up of taxa that occurred in >10% abundance (Percent Dominants), the percent of total diatom valves made up of taxa that occurred in >10% abundance (Percent Dominants), the percent abundance of the taxon *Achnanthidium minutissimum* (%Ach_min), the Centrales /Pennales ratio (C/P), and finally, the Siltation Index (% Siltation Index), which is the sum of the percent abundances of *Amphiprora,*

Anemautus, Biremis, Craticula, Cylindrotheca, Decussata, Entomoneis, Fallacia, Fistulifera, Geissleria, Gyrosigma, Hantzschia, Haslea, Hippodonta, Lyrella, Navicula, Nitzschia, Parlibellus, Placoneis, Plagiotropis, Pleurosigma, Proshkinia, Psammodictyon, Sellaphora, Stenopterobia, Surirella, and *Tryblionella.* These are genera of predominantly motile taxa that are able to maintain their positions on the substrate surface in depositional environments (Bahls 1993). We evaluated the use of these indices in conjunction with different chemical and physical site characteristics and types of landuse, running a Spearman's rank-order correlation using Sigma Stat 2.03.

4. Results

4.1 Environmental gradients

PCA detected two major gradients in the environmental data (Fig. 2). The first PCA axis (eigenvalue $?_1$= 0.53) reflected a physical gradient strongly influenced by sediment type (% gravel and % sand, silt and clay substrates) and basin size. The second axis (eigenvalue $?_2$= 0.24) revealed a land-use and chemical gradient driven mainly by % urban and agricultural land-use, pH and all inorganic nutrients measured. The first two PCA axes explained 77.7% of the cumulative percentage of variance in the environmental variables. In constrast to sites in the Northern Piedmont ecoregions, which were sampled in previous years of this study (Ponader and Charles 2004), high levels of inorganic nutrients in the Inner Coastal Plain seem to be correlated mainly with % agricultural land-use, whereas high loadings of conductivity and chloride are correlated more strongly with %urban land-use (Fig. 2). The nutrient gradient included in the dataset was large (TP: 0.02-0.56 mg/L and TN 0.46-4.82 mg/L). The pH gradient ranged from 5.9 to 7.2. Also, the PCA showed a great heterogeneity in combinations of environmental characteristics among the sites, expressed in the scattered distribution in the biplot (Fig.2).

4.2 Algal Biomass

4.2.1. Relationships between algal biomass measures and nutrient conditions (Spearman's rank-order correlation)

Relationships among all environmental variables and Chl *a* contents obtained from diatometers for the full dataset (*n*=47) was explored using a Spearman's rank order correlation matrix, two tailed test. The results are presented in Table 3. The following correlations were found to be significant at the 0.05 level. Chl *a* was significantly correlated with % open canopy cover (*r*=0.30), with conductivity (*r*=0.29), and with total hardness (*r*=0.32). Although these variables were significantly correlated with Chl *a*, the strength of the correlations were weak. No other variable, and especially no nutrient variable, showed significant correlation with measured levels of Chl *a*.

4.2.2. Analysis of principal factors influencing algal biomass (forward stepwise regression)

We analyzed measures of algal biomass (Chl *a*) and its relationship with nutrients (TP, O-P, NH3-N, NO3-N, TKN, TN) and other chemical and physical characteristics using forward stepwise regression (Appendix 1). Because many variables were strongly correlated, only conductivity, pH, Basin size, % SDSTCL, %open canopy cover, flow, TP, TN and NH3-N were included in the analysis. The results show that the dependent variable Chl *a* can be predicted from a linear combination of the independent variables conductivity, NH3-N, % open canopy cover and TN. The correlations are significant at the 0.001 level for conductivity and NH3-N and at the 0.05 level for %open canopy cover and TN.

4.3. Diatom species composition and species distribution

A total of 294 diatom species were identified. Sample assemblages showed high species diversity (mean S-W index of 3.2, mean # species per sample (species richness)= 62) and high evenness, due to generally low abundances (mean % abundance of most dominant species= 20 %). The 10 most abundant species, determined by high abundances and high numbers of occurrences (Hills N2), are commonly found in eutrophic waters

and had moderate TP and TN WA-optima (Table 4). These taxa were *Gomphonema parvulum* (Kützing) Kützing, *Sellaphora pupula* (Kützing) Meresckowsky, *Eunotia bilunaris* (Ehrenberg) Mills, *Nitzschia palea* (Kützing) W. Smith, *Navicula germainii* Wallace, *Navicula minima* (Grunow), *Achnanthidium minutissimum* (Kützing) Czarnecki, *Nitzschia palea* var. *debilis* (Kützing) Grunow, *Navicula tenelloides* Hustedt and *Navicula rhynchocephala* Kützing. After applying our selection criteria (= 0.3% abundance in = 1 sample), 253 taxa were included in the analyses. In the DCA of diatom data, the eigenvalues of the first two axes were $?_1$= 0.24 and $?_2$= 0.18, accounting for 18.8% of the variance in the species data **(**Fig. 3**)**. As the gradient length of the first two DCA axes of 2.9 and 2.3 standard deviation (SD) units were above 2 SD, we chose to use techniques based on assumptions of unimodal species distribution (ter Braak 1995, ter Braak and Prentice 1988) for further model development.

4.4. Development of nutrient inference models based on diatom species composition

4.4.1. Data screening

Three samples were identified as outliers, based on PCA and DCA, and were removed from CA, CCA analysis and inference model development. Outliers detected by PCA are samples with extreme values of environmental variables, whereas samples detected by DCA are samples with unusual species composition.

Two samples were determined to be outliers because of their high sample score means on the PCA axes: samples 9 and 10 (site 151 and 166) (Fig. 2). Site 166 had the highest % gravel substrate (45%) and the lowest % agricultural landuse (5.57%). Site 151 (North Branch Rancocas) had the biggest basin size, high % sand silt and clay (100%) and the lowest measured amount of hardness (21.9 mg/l), conductivity (70.4 mg/l) and NO3-N (0.159 mg/l) in the dataset. 5 samples were identified as outliers because of their high sample score means on the DCA axes: sample 6, 9, 14, 19 and 23 (sites 133, 151, 440, 488, 673 (section 2)). Sites 133 and 440 had nutrient concentrations in the lower range and site 488 had high-nutrient concentrations. Because it was important that sites in the calibration dataset have a wide nutrient gradient, we decided to exclude only the sample from site 673 (section 2) out of the 5 samples detected as outliers by DCA.

Despite the fact that this site represented the high end of the phosphorus gradient, we decided to exclude this sample, because it was a gravel sample and had a duplicate from section 1. This result is interesting in that the diatom species composition is significantly different in this sample taken from natural substrate (gravel) versus the sample taken from a diatometer at the same site. The fact that one sample is taken in section 1 and the other in section 2, might lead to the assumption that other factors than substrate might explain the difference in the diatom species assemblages. Figure 3 shows that the only difference responsible for the distance between the two sample points is % gravel. Due to its low pH and conductivity conditions, site 151 (Sample no. 9) had a diatom flora, distinctively different from the others, dominated by *Eunotia* ssp. and *Fragilaria* spp. and was excluded. In the end, samples 9 (site 151) and 10 (site 166) and sample 23 (site 673_2) were deleted from the dataset to be included in further analysis (Fig 3). Removing the three outlier samples reduced the dataset from 25 samples to 22.

4.4.2. Relationships between diatom assemblages and environmental variables

We used CA to determine the environmental variables explaining most of the variation in diatom species composition, and especially the importance of nutrients as compared with other environmental characteristics. Prior to running ordinations we used a Pearson correlation matrix including all 24 variables measured and all 22 sites to detect variables that were highly correlated (Table 5). Consequently, we excluded the twelve variables average river width, % gravel, % forested land-use, %urban land use, color, conductivity, hardness, alkalinity, O-P, NO3-N, TKN and NH3-N. Because TN, TKN and NO_3-N, as well as TP and O-P, were highly correlated, only 1 of the 2-3 variables, respectively, could be selected. Because TP and TN are more important with respect to management purposes, we chose to eliminate the variables O-P and NO_3-N and TKN. In a CA including all remaining 12 environmental variables as passive variables and diatom data for 22 sites, 21.0% of the total variance of the diatom data was explained by CA axis 1 ($?_1$ = 0.26) and CA axis 2 ($?_2$ = 0.19). The species-environment correlations were high for both axis 1 (r = 0.60) and axis 2 (r = 0.92), accounting for 20.1% of the variance in the species-environment relationships. This indicates a strong correlation between the 12 environmental variables and the 253 species included in the CA. Low weighted

correlations and low variance inflation factors (VIFs <5.5) indicated no strong co-linearity among environmental variables. CCA with forward selection including only these 12 variables in total, identified 3 environmental variables that significantly (p < 0.05) explained variance in diatom species composition. These were TP, pH and basin size (Fig. 4). The 3 variables together explained 30.7% of the total variance in the diatom data. Species-environment correlations of CCA axes were high for axis 1 (r = 0.95) and for axis 2 (r = 0.92) and accounted for 73.1% of the variance in the species-environment relationships. The correlations between pH and the first ordination axis were strong (r = -0.86) less strong for TP (r = -0.36), and the least strong for basin size (r =-0.15). TP was correlated strongest with the second axis (r = 0.54), followed by basin size (r = -0.43) and pH (r = -0.41).

CCAs constrained to 1 variable at a time were run on the remaining 3 variables to assess the strength of their relationship with diatom species composition, and to determine which variables had a strong enough relationship with diatom assemblages to justify development of inference models. Of the 3 variables, pH had the highest $?_1/?_2$ ratio (0.61), capturing 7.3% of the variation in the species data (Table 6). TP had a $?_1/?_2$ ratio of 0.51, capturing 6.0% of the variation in species data. The strength of the correlation between pH and the first CCA axis indicates that pH is the strongest factor influencing diatom species composition in the Inner Coastal Plain dataset. To verify the independent relationship between the variables pH and TP, a CCA was run constrained to TP with pH as covariables. When pH was entered as a covariable the ratio $?_1/?_2$ changed only slightly from 0.510 to 0.508 (Table 6). This confirmed that TP had a significant and independent influence on the diatom assemblages and that development of inference models for this variable was justified, despite the strong pH gradient. The variable TN did not have a significant influence to justify the development of an inference model. We did two test runs to see if TKN or NO3-N would have significant influence on diatom species composition, by including them instead of TN as variables in the CCA with forward selection. The CCA with forward selection showed that neither TN, TKN or NO3-N had a significant influence, therefore development of an inference model for any measured form of nitrogen was not possible. Nonetheless, even if TP significantly and independently influenced diatom species composition, the strong pH (e.g. reflecting

alkalinity and hardness) gradient should be kept in mind when analyzing this dataset for nutrient response.

4.4.3. Species WA-optima and tolerances

TP and TN WA species optima were calculated using the reduced dataset (*n*=22). Species apparent WA TP optima for all species ranged from 24 to 560 µg L^{-1}, Weighted-average TN optima ranged from 457 to 14671µg L^{-1}. Because the calculated optima are reliable only for abundant species, optima are presented for the species with effective numbers of occurrences (Hill's N2) > 10 (Table 4). The TP and TN species optima obtained were higher in the Inner Coastal Plain, but generally comparable to those calculated for northern New Jersey (Ponader and Charles 2004, Ponader et al., in review). Species with high TP optima were *Luticola mutica* (Kutz.) Mann, *Surirella angusta* Kützing, *Sellaphora seminulum* (Grun.) Mann, *Tryblionella debilis* Arnott, *Hippodonta capitata* (Ehrenberg) Lange-Bertalot, Metzeltin et Witkowski, *Placoneis clementis* (Grun) Cox, *Caloneis bacillum* (Grunow) Cleve, *Nitzschia capitellata* Hustedt, *Fragilaria vaucheriae* (Kützing) Petersen, *Sellaphora pupula* (Kützing) Meresckowsky and *Planothidium lanceolatum* (Brébisson ex Kützing) L.-B. Species with low TP optima were *Eunotia exigua* (Brébisson ex Kützing) Rabenhorst, *Neidium alpinum* Hustedt, *Encyonema minutum* (Hilse) Mann, *Achnanthidium minutissimum* (Kützing) Czarnecki,. *Navicula rostellata* Kützing, *Frustulia vulgaris* (Thwaites) DeT., *Gomphonema gracile* Ehr. emend. V. H., *Navicula rhynchocephala* Kützing, *Navicula longicephala* Hustedt, *Navicula tenelloides* Hustedt.

4.4.4. Weighted averaging - nutrient inference models

Models to infer TP ($\log_{10}$ µg/L) were developed and tested using WA and WA-PLS. The models were run on the full (*n*=25) and the reduced (*n*=22) calibration dataset, including 253 taxa. Table 7 shows performance measures for all inference models developed. The best TP model was a two-component WA-PLS model using the full dataset (*n*=25), showing the highest r^2_{jack} (0.34) and the lowest $RMSE_{jack}$ (0.31). For all modeling techniques, the full dataset always gave better results as compared to when

outliers were removed. In the reduced and the full model dataset, the TP gradient was 23.5-559.5 µg/L.

When comparing all TP models developed using different techniques, the difference between apparent and bootstrapped or jackknifed r^2 was generally high (Table 7, Fig. 5) indicating that the models showed weak performance when tested. We investigated, if model performance could be improved using different data transformations or by creating data subsets. For that reason, we created and tested additional WA-PLS models using the full (n=25) dataset on which we performed the following changes: a) we removed both samples from site 673 (highest TP site), in order to eliminate the large data gap at the higher end of the TP gradient; b) we used the untransformed diatom data set (% relative abundance), as well as c) the full diatom dataset containing all 294 species to verify whether the previously applied data transformation (square root) and elimination of rare species (< 0.3 %) might have decreased the species response; d) because of the strong pH gradient, we split the dataset into two subsets based on pH; the first contained all sites with pH > 6.1, and e) the second included all sites with pH values <6.8. The results showed, that none of the performed dataset transformations did significantly improve the model performance (Table 7), suggesting that other factors must be considered when investigating the causes for the decrease in apparent versus bootstrapped or jackknifed r^2.

The graph of the observed versus inferred values (Fig. 5), as well as the residuals plot (Fig. 6) for the best TP model, using WA-PLS (n=25) shows that the model overestimates values <80 µg/L (1.9 log10 TP) and underestimates TP values >150 µg/L (2.18 log10 TP) when tested using leave-one-out cross validation. This trend commonly occurs with models developed on samples with an uneven distribution along the TP gradient (Soininen and Niemela 2002, Reavie et al. 1995), and also occurred during the analysis of the NJ Piedmont, Highlands and Ridge and Valley (study year 1-3) dataset (Ponader and Charles 2004, Ponader et al. 2004). It is often referred to as “edge-effect” (Hall and Smol 1999), and causes the WA estimates of species optima to be biased. However, compared to the models developed in northern New Jersey, this effect is much stronger in the Inner Coastal Plain dataset. Examination of the residuals plot (Fig. 5) and the observed versus inferred TP graph (Fig. 6) show the following patterns in the

distribution of TP along the phosphorus gradient. First, despite the rather wide range of TP included in the model, sites with low and high TP are underrepresented. Second, the dataset shows two important gaps, one between the levels of 1.9-2.1 log10 TP (80-125µg/L TP) and a larger gap between 2.3-2.75 log10 TP (200-560µg/L TP). In summary, the Inner Coastal Plain TP model is built mainly on sites with moderate phosphorus concentrations, and generally shows an uneven distribution of sites along the TP gradient. This explains the decrease of the model performance at both ends of the TP gradient, when tested using cross-validation techniques. More detailed explanations for differences between the performances of the TP models obtained for the Inner Coastal Plain dataset versus the northern New Jersey dataset are provided in the discussion (chapter 5).

4.4.5. Diatom TP index

Because they are rescaled values of inferred nutrient concentrations, the TP diatom index corresponds directly with those concentrations (Fig. 5). All measures of the performance of inference models are also relevant to the relationships between the indices and measured nutrient concentrations.

To provide a measure of the error, we evaluated how accurately index values, calculated using bootstrapped inferred nutrient concentrations, could be assigned to the nutrient categories. The Diatom TP Index correctly assigned 44% of the samples, but placed 66% of the samples into neighboring categories. The TP categories 2 and 4 had the highest proportion of samples correctly assigned, whereas categories 1 and 3 had the highest proportion of samples assigned to a neighboring category. The index showed least accuracy placing samples in the TP categories 1 (< 0.025 mg / L) and 3 (0.075-0.1 mg / L), which are the categories with the lowest number of samples.

4.5 Diversity metrics

Six different diatom diversity metrics were calculated using the full dataset (n=25). A Spearman's rank-order correlation was run to evaluate how strongly these metrics were correlated with different environmental variables. The main goal of this analysis was to identify metrics that can be used to assess nutrient impairment in NJ Inner

Coastal Plain rivers. The results are summarized in Table 8. Considering the indirect relationship between the metrics and the variables they were correlated with, we consider that any correlations with an *r* greater than 0.5 are strong, that correlations between *r* = 0.2 and 0.5 are moderate and that any correlations below an *r* of 0.2 are weak. The following indices showed significant (at 0.01 level) and strong correlations. Percent Dominant Diatom Taxa was strongly negatively correlated with pH and conductivity. Similarly the Siltation Index was strongly correlated with pH, conductivity and hardness as well as TP. # of Diatom Taxa showed strong correlation with chl a. Correlations at the (0.05 level) were generally moderate to strong, but showed that out of all diversity indices, only % *Achnanthidum minutissimun* and the Shannon-Wiener Index were moderately correlated with direct or indirect measures of nutrients such as TN and % Urban and Chl a. Except for the Siltation Index, no other significant and strong relationship was found with measures of nutrients. Strong correlation of several indices with pH, alkalinity, and hardness suggested that in the Inner Coastal Plain these diversity metrics are better indicators of pH conditions than of nutrient conditions.

5. Discussion and Conclusion

5.1 Evaluation of Inner Coastal Plain models/metrics

5.1.1 Biomass metrics, principal factors influencing algal biomass

Algal biomass was measured by analyzing the contents of Chl *a* contained in the composite biomass samples collected from diatometer slides. Spearman's rank-order correlation between nutrient variables and the biomass data (Chl *a* contents) were obtained for the full dataset (*n*=47). Significant correlations were found between Chl *a* and % open canopy cover, conductivity, and total hardness but not between biomass and nutrients. However, variations in contents of Chl *a* can be explained through a combination of conductivity, NH3-N, %open canopy cover and TN, as shown through forward stepwise regression. % Open canopy cover is the factor that was significantly correlated with biomass in both datasets, the Inner Coastal Plain and the Piedmont

datasets (Ponader and Charles 2004). In the Piedmont, variations of biomass were explained through a combination of light, nutrient concentrations and larger sized substrate. Because biomass in the Inner Coastal Plain was measured from algae on artificial substrata (diatometers), and most rivers had predominantly fine-grained substrate (sand silt and clay), substrate could not have had an influence on among-sample difference in biomass. The results from the Inner Coastal Plain data show that assessing nutrient conditions through measuring Chl *a* from diatometer slides was difficult, due to overriding factors like light conditions (%open canopy cover) and conductivity. Also, high contents of clay particles in the water causing increased turbidity (Hill 1996), might have influenced algal biomass production and may limit the use of this method in the Inner Coastal Plain. In addition, the time of deployment of the diatometers could have had an influence on biomass concentrations. The diatometers were collected after 14-18 days, to reduce the risk of scouring of diatom mats from the slides. On the other hand, because of relatively low %percent open canopy cover at the sites sampled (mean percent canopy cover at all Inner Coastal Plain sites = 13 %), development of algal biomass might have been reduced due to high shading. We will investigate this question further during the year 5 study year in the Outer Coastal Plain: during fieldwork in 2004, the diatometers were collected after an increased deployment time of 21-24 days. In addition, biomass samples were taken from diatometer slides as well as from natural substrate (sand, silt, clay) using a new field method (Ponader and Winter 2004). More detailed recommendations for the collection of biomass measures (Chl *a*) in the NJ Coastal Plain and the use of natural substrate versus diatometers, exposure times etc. will be given at the end of study year 5. Current results from the Inner Coastal Plain suggest that biomass measures obtained from diatometer slides do not reflect nutrient conditions well.

5.1.2 Inference model performance: Comparison of results between the Inner Coastal Plain data set and Piedmont, Highlands and Valley & Ridge data set.

The diatom-based TP inference models developed for the full dataset performed moderately and could not be improved by removing outliers, or when different transformations were applied to the dataset (Table 7). Compared to those developed for streams in northern and central NJ (Ponader and Charles 2004, Ponader et al. in review)

the TP models presented here (*n*=25 and *n*=22) show lower apparent and bootstrapped or jackknifed r^2 and higher apparent and bootstrapped or jackknifed RMSE (Table 7). Differences in model performances between the Inner Coastal Plain study and the previous studies may be explained by several factors.

First, due to lower sample size (*n*= 25) the dataset shows significant data gaps, despite the fact that a wide range of concentrations along the nutrient gradients was captured. The Inner Coastal Plain dataset included a TP range of 24-560 µg L^{-1} (*n*= 25) and the range in TN was 457-14671 µg L^{-1} (*n*= 25). Two gaps exist in the TP range, one from 80-125µg/L TP and a larger gap from 200-560µg/L TP. TN generally shows even distribution at the lower end of the gradient, but the data is sparse at the higher end leading to a very large gap from 4829-14971µg/L TN. Uneven distribution of samples along the nutrient gradient generally leads to lower model performance (Reavie et al. 1995, Soininen and Niemelä 2002). In comparison, the northern and central NJ dataset showed more even distribution of samples along the gradient (TP : 6-732 µg L^{-1}; TN: 170-8547 µg L^{-1} (*n*=101), leading to increased model performance.

Second, in contrast to the central and northern NJ dataset, the Inner Coastal Plain dataset showed that pH and the correlated variables alkalinity and hardness and conductivity were the most important variables influencing diatom species composition. Development of nutrient models is known to be difficult for datasets where pH is a variable strongly influencing diatom species composition (Reavie and Smol, 2001). As in previous years, sites in this study were selected based on knowledge of their chemistry to avoid wide ranges of pH, alkalinity and conductivity. The pH gradient included in the Inner Coastal Plain study was relatively short 5.9-7.2, but on average the rivers were more acidic (pH 6.5), less alkaline (27 mg/L) and had lower values of conductivity (198 µS cm/L) than in the central and northern NJ dataset (mean: pH 7.6; alkalinity 70 mg/L, conductivity: 327 mg/L). The northern and central NJ Dataset showed a pH range of 6.5-9 (*n*=101) and 6.9-9 (*n*=91), and excluded acidic sites. Therefore, although the diatom flora in the Inner Coastal Plain is typical of eutrophic waters, it also is influenced by the generally lower pH of the rivers. In contrast, because in central and northern NJ the rivers have a generally circumneutral pH, but a strong gradient in nutrient loadings, the diatom species composition was influenced to a higher degree by nutrients.

Third, it is possible that the amount of clay in the water column of most of the Inner Coastal Plain rivers sampled might play a role in the availability of nutrients. Because of their cation exchange capacity, clay particles can bind and precipitate nutrients (Wetzel 2001), which in turn can makes nutrients less available for uptake by algal communities. Another negative impact of high amounts of clay in the water column could be less availability of light, one of the most important factors influencing algal growth (Hill 1996). More detailed analysis of this relationship would be needed, but surpasses the scope of this study, as not enough data are available to verify the impact of suspended clay particles in the water.

Finally, there are no published records of inference models, which were developed using diatom data from diatometer slides. It is possible that the diatom species composition was influenced by the use of artificial substrate. For example, insufficient exposure time, would mean that the algal assemblage would be biased towards early 'immigrator' species and therefore may have had an impact on the success of nutrient model development. To verify this, and we took diatom samples from both, artificial and additional natural substrate and increased the exposure time of the diatometers, during the 5th project year (Outer Coastal Plain). We will investigate and summarize potential differences in the model performances between nutrient models developed using diatometers and natural silt/sand substrate in the year 5 final report.

Many other factors might be responsible for low model performances of the Inner Coastal Plain models compared to those developed in previous study years. These include the influence of temporal variability in nutrient concentrations in streams (Cattaneo and Prairie 1994, Pan et al. 1997). So far, only a few TP and TN diatom inference models have been developed for rivers in the eastern US, and future studies are necessary to help understand the influence of these factors on diatom species composition in relationship with nutrients.

The Inner Coastal Plain TP inference model has moderate predictive power and showed relatively high uncertainty (e.g. inferred nutrient concentrations) when tested using bootstrapping, and especially when compared to nutrient models developed from central and northern NJ (Ponader and Charles 2004, Ponader et al. 2004). Consequently, the diatom TP index also showed lower performance, and assigned the majority of the

samples to the neighboring categories. Despite the wide ranges in TP and TN captured in our models, model performance would need to be improved by adding samples to the dataset. It would be especially important to fill the existing data gaps in the nutrient gradient. We will merge the Inner Coastal Plain dataset with the Outer Coastal Plain dataset to increase the size of the dataset and to fill the existing data gaps. The development of nutrient inference models should give more reliable results, when using a bigger dataset with more even distribution of the samples along the nutrient gradient.

5.1.3 Diatom Diversity metrics

The usefulness of the six diversity metrics was evaluated by comparing their correlations with nutrient impairment measures. Out of all diversity indices only % *Achnanthidum minutissimum* and the Shannon Wiener Index were moderately correlated with direct or indirect measures of nutrients such as TN and % Urban and Chl *a*. Strong correlation of several indices with pH, alkalinity and hardness, suggested that in the Inner Coastal Plain these diversity metrics are better indicators of pH conditions than of nutrient conditions. Therefore, it is possible to use these diatom indicators to monitor river impairment, but the results must be interpreted with caution, especially with respect to pH conditions. Nevertheless, the formulas for the indices are simple and calculation is easy. We therefore recommend using simple metrics as complementary to inference models and biomass metrics.

5.2 Recommendations for use of algal indicators in the Inner Coastal Plain

Our results show that a) biomass, b) inference model development and c) diversity metrics, showed limited applicability as measures of nutrient conditions in the Inner Coastal Plain. The chemical characteristics of the Inner Coastal Plain rivers, such as high nutrient loadings combined with generally lower pH and conductivity make it difficult model the effect of nutrients on algae communities. In addition, uneven distribution of the samples along the nutrient gradient makes development of reliable diatom inference TP models difficult. Finally, we are currently investigating if artificial substrate (diatometers) may have influenced algal biomass and species composition.

We expect significant improvement of the TP model performance, when using a larger dataset (Inner and Outer Coastal Plain datasets combined) during data analysis of the current study year 5. Also, increased diatometer exposure time as well as additional biomass and diatom samples taken from natural substrate during year 5 (Outer Coastal Plain) should help understanding the main factors that influence diatom species composition in the Coastal Plain. We will continue to further improve and adapt indicators of nutrient conditions to the specific nature of the rivers of the Coastal Plain physiographic province during year 5.

References

American Public Health Association, American Water Works Association and Water Pollution Control Federation (APHA, AWWA and WPCF). 1992. Standard Methods for the Examination of Water and Wastewater, 19th ed. Washington, DC.

Bahls, L. L. 1993. Periphyton bioassessment methods for Montana streams. Montana Water Quality Bureau, Department of Health and Environmental Science, Helena, MT.

Barbour, M.T., J. Gerritsen, B. D. Snyder, and J. B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: Periphyton, benthic macroinvertebrates, and fish, Second Edition. EPA 841-B-99-002, Washington, D.C.,U.S. Environmental Protection Agency, Office of Water.

Bigler, C., and R. I. Hall. 2002. Diatoms as indicators of climatic and limnological change in Swedish Lapland: a 100-lake calibration set and its validation for paleoecological reconstructions. Journal of Paleolimnology 27:97-115.

Birks, H.J.B. 1995. Quantitative paleoenvironmental reconstructions. Pages 161-254. In: Statistical Modelling of Quarternary Science Data. D. Maddy and J.S.Brew. Technical Guide 5, Quarternary Research Association, Cambridge.

Birks, H. J. B., J. M. Line, S. Juggins, A. C. Stevenson, and C. J. F. ter Braak. 1990. Diatoms and pH reconstructions. Philosophical Transactions of the Royal Society of London, series B 327:263-278.

Cattaneo, A., and Y.T. Prairie. 1995. Temporal variability in the chemical characteristics along the Rivière de l'Achigan: how many samples are necessary to describe stream chemistry. Can. J. Fish. Aquat. Sci. 52: 828-835.

Charles, D., D. Winter, and M. Hoffman. 2000. Field Sampling procedures for the New Jersey Algae Indicators project. PCER Procedure P-13-64. Patrick Center for Environmental Research, Academy of Natural Sciences, Philadelphia, PA. 18 Pages.

Charles, D.F., C. Knowles, R. Davis (eds.). 2002. Protocols for the analysis of algal samples collected as part of the U.S. Geological Survey, National Water-Quality Assessment Program. Report No. 02-06. The Academy of Natural Sciences. Patrick Center for Environmental Research, Philadelphia, PA. 124 pp.

Dodds, W. K., J. R. Jones, and E. B. Welch. 1998. Suggested classification for stream trophic state: distributions of temperate stream types by chlorophyll, total N and P. Water Research 32: 1455-1462.

Fishman, M.J. 1993. Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory- Determination of inorganic and organic constituents in the water and fluvial sediments: U.S. Geological Survey Open-File Report 93-125.

Hall, R.I., and J.P Smol. 1999. Diatoms as indicators of lake eutrophication. In: Stoermer E. F. and J. P. Smol (eds.). The Diatoms: Applications for the Environmental and Earth Sciences. Cambridge University Press, Cambridge, UK. Pp.128-168.

Hill, W. 1996. Effects of Light. In: Algal Ecology, Freshwater Benthic Ecosystems. R. J. Stevenson, M. L. Bothwell, and R. L. Lowe (eds.). Academic Press, NY.

Juggins, S. 2003. C^2 User guide. Software for ecological and paleoecological data analysis and visualization. University of Newcastle, Newcastle upon Tyne, UK. 69 pp.

Kiry, P., D. Velinsky and A.-M. Compton. 1999. Determination of dry weight and percent organic matter for sediments, tissues and benthic algae. PCER Procedure P-16-113. Patrick Center for Environmental Research, Academy of Natural Sciences, Philadelphia, PA. 3 Pages.

Moulton, S. R., J. G. Kennen, R. M. Goldstein, and J. A. Hambrook. 2002. Revised Protocols for Sampling Algal, Invertebrate and Fish Communities as Part of the National Water-Quality Assessment Program. Open-File Report 02-150. US Geological Survey, Reston, Virginia. 75 Pages.

NJDEP. 2000. Water Quality Monitoring Networks 2000. New Jersey Department of Environmental Protection. Division of Watershed Management. Trenton, NJ

Pan, Y., R. J. Stevenson, B.H. Hill, A.T. Herlihy, G.B. Collins. 1996. Using diatoms as indicators of ecological conditions in lotic systems: a regional assessment. Journal of the North American Benthological Society 15: 481-495.

Patrick Center for Environmental Research. 2000. Quality Assurance Project Plan for the project "Understanding the Relationship Between Natural Conditions and Loadings on Eutrophication: Algal Indicators of Eutrophication for New Jersey Streams", Academy of Natural Sciences, Philadelphia, PA. 21 Pages.

Patrick Center for Environmental Research. 2003. Requirements for Phycological Algal Indicators and Data Exploration (Phyco-AIDE) Version 2. Program Manual submitted to USGS (NAWQA) by The Academy of Natural Sciences, Philadelphia, PA.

Ponader, K., and D. Charles. 2001. Understanding the Relationship Between Natural Conditions and Loadings on Eutrophication: Algal Indicators of Eutrophication for New Jersey Streams - Final Report Year 1. PCER Report 01-26. Patrick Center for Environmental Research, Academy of Natural Sciences, Philadelphia, PA. 23 Pages.

Ponader, K., and D. Charles. 2003. Understanding the Relationship Between Natural Conditions and Loadings on Eutrophication: Algal Indicators of Eutrophication for New Jersey Streams - Final Report Year 2. Report No. 03-04. Patrick Center for Environmental Research, The Academy of Natural Sciences, Philadelphia, PA. 74 Pages.

Ponader, K., D. Winter, and D. Charles. 2003. Field Sampling procedures for the New Jersey Algae Indicators project: Use of Diatometer Artificial Substrates. PCER Procedure P-13-66. Patrick Center for Environmental Research, Academy of Natural Sciences, Philadelphia, PA. 32 Pages.

Ponader, K., and D. Charles. 2004. Understanding the Relationship Between Natural Conditions and Loadings on Eutrophication: Algal Indicators of Eutrophication for New Jersey Streams - Final Report Year 3. Patrick Center for Environmental Research, The Academy of Natural Sciences, Philadelphia, PA. 56 Pages.

Ponader, K. and D. Winter. 2004. Field Sampling procedures for the New Jersey Algae Indicators project: Sampling method for collection of qualitative diatom samples and algal biomass samples from epipsammic/epipelic habitat in the field. PCER Procedure P-13-67. Patrick Center for Environmental Research, Academy of Natural Sciences, Philadelphia, PA. 8 Pages.

Ponader, K., D. Charles, and T. Belton. Diatom-based TP and TN inference models and indices for monitoring enrichment of New Jersey Streams. Manuscript submitted for peer-reviewed publication in 2004, currently in review.

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

Reavie, E.D., R.I. Hall, and J. P. Smol. 1995. An expanded weighted-averaging model for inferring past total phosphorus concentrations from diatom assemblages in eutrophic British Columbia (Canada) lakes. Journal of Paleolimnology 14:49-67.

Soininen, J., and P. Niemelä. 2002. Inferring the phosphorus levels of rivers from benthic diatoms using weighted averaging. Archiv für Hydrobiologie 154:1-18.

Tedrow, J. C. F. 1986. Soils of New Jersey. Robert E. Krieger Publishing Company, Inc., Malabar, FL.

ter Braak, C.J.F. 1995. Ordination. Pages 91-169. In: Data Analysis in Community and Landscape Ecology. R. H. G. Jongman, C. J. F. ter Braak, and O. F. R. van Tongeren (editors). Cambridge University Press, Cambridge.

ter Braak, C. J. F., and S. Juggins. 1993. Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. Hydrobiolgia: 269/270: 485-502.

ter Braak, C. J. F., and I. C. Prentice, 1988. A theory of gradient analysis. Advances in Ecological Research 18: 221-317.

ter Braak, C.J.F., and P. Šmilauer, 2002. CANOCO reference manual and CanoDraw for Windows. User's guide: software for Canonical community ordination. Version 4.5. Microcoputer Power, Ithaca, NY, USA.

Velinsky, D. 2000. Syringe water sampling and filtration for the collection of filtered nutrient samples and unfiltered nutrient samples. PCER Procedure P-16-119. Patrick Center for Environmental Research, Academy of Natural Sciences, Philadelphia, PA. 4 Pages.

Velinsky, D., and J. DeAlteris. 2000. Benthic algae and sediment chlorophyll *a* preparation and analysis. PCER Procedure P-16-117. Patrick Center for Environmental Research, Academy of Natural Sciences, Philadelphia, PA. 4 Pages.

Wetzel, R. G. Limnology. 2001. Lake and River Ecosystems. Third Edition. Academic Press, San Diego. 1006 pp.

Winter, J. G., and H. C. Duthie. 2000. Epilithic diatoms as indicators of stream total N and total P concentration. Journal of the North American Benthological Society 19: 32-49.

Table 1. List of sites sampled in 2003 - Inner Coastal Plain

Date sampled	NJ Site ID	Waterbody	Location	Impairment1	Impairment2
Inner Coastal Plain diatometer sites					
9/16/03	AN0109	Assunpink Ck	Rt 535 Old Trenton Rd	Non-Impaired	Moderate
9/5/03	AN0121	Crosswicks Ck	Rt 537	Moderate	Moderate
9/16/03	AN0124	Lahaway Ck	New Egypt-Allentown Rd (Holms Mill Rd.)	Moderate	Moderate
9/9/03	AN0129	Doctors Ck	Breza Rd	Moderate	Moderate
9/5/03	AN0132	Blacks Ck	Chesterfield-Georgetown Rd	Moderate	Moderate
9/9/03	AN0133	Bacons Run	White Pine Rd Nr Kuser Pond	Moderate	Moderate
9/16/03	AN0136	Crafts Ck	Island Rd	Moderate	Severe
9/5/03	AN0139	Annaricken Bk	Island Rd	Moderate	Moderate
9/5/03	AN0151	North Br Rancocas Ck	Birmingham Rd.	Moderate	Severe
9/4/03	AN0166	Barton Run	Tuckerton Rd & Christopher Mill Rd	Severe	Moderate
9/5/03	AN0169	Southwest Br Rancocas Ck	Rt 70	Moderate	Moderate
9/16/03	AN0384	Bear Bk	Cranbury Rd. (Rt. 615)	Moderate	Non-impaired
9/10/03	AN0439	Manalapan Brook	Federal Rd	Moderate	Moderate
9/10/03	AN0440	Manalapan Brook	Old Forge Rd	Moderate	Moderate
9/15/03	AN0448	Matchaponix Brook	Rt 527	Moderate	Moderate
9/10/03	AN0451	Matchaponix Brook	Texas Rd	Severe	Moderate
9/15/03	AN0466	Hop Brook	Willow Brook Rd	Moderate	Moderate
9/10/03	AN0470	Big Brook	Cross Rd	Moderate	Moderate
9/9/03	AN0488	UNT to Manasquan River (Killtime Bk)	Strickland Rd	Moderate	Moderate
9/9/03	AN0489	Manasquan River	Rt. 9	Moderate	Moderate
9/15/03	AN0490	Manasquan River	West Farms Rd	Moderate	Moderate
9/4/03	AN0673	Edwards Run	Pitman-Jefferson Rd	Severe	-
9/11/03	AN0683	Raccoon Ck	Tomlin Station Rd (USGS gage)	Moderate	-
9/29/03	AN0694	Major Run	Pointers-Sharptown Rd	Severe	-
Piedmont resampling sites (rock substrate)					
10/8/03	AN0115	Miry Run	Rt 533 (Quakerbridge Rd)	Moderate	Moderate
10/8/03	AN0234	Whippany River	Ridgedale Ave W of I-287	Severe	Non-impaired
10/8/03	AN0374	N Br Raritan River	Rt. 202	Non-Impaired	Non-impaired
10/8/03	AN0405	Pike Run	Rt 533	Moderate	Severe

Table 2. Values of selected environmental variables for each sample collected during 2003. AvgW: average stream width, Open: % open canopy cover, GR: % gravel, SDSTCL: % sand silt and clay, URB: %urban land-use, AG: %agriculture land-use, FOR: %forest land-use, BASIN: basin size, Cond: Conductivity, Water color, Flow (estimate: 1=slow, 2=moderate, 3=fast), Alk: total alkalinity, Cl: chloride, Hardn: Total hardness, NO3-N: dissolved nitrate, NH3-N: dissolved ammonia, TKN: total Kjeldahl nitrogen, TN: Total Nitrogen, O-P: orthophosphate, TP: total phosphorus, Chl a: Chlorophyll a.

Sample ID	*Diatom Sample #*	*Section*	*AvgW (m)*	*Open (%)*	*GR (%)*	*SDSTCL (%)*	*URB (%)*	*FOR (%)*	*AG (%)*	*BASIN (km2)*	*pH*	*Cond (µS/cm)*	*Color*	*Flow*	*Alk (mg/L)*	*Cl-(mg/L)*	*Hardn (mg/L)*	*NO3-N (mg/l)*	*NH3-N (mg/l)*	*TKN (mg/l)*	*TN (mg/l)*	*O-P (µg/L)*	*TP (µg/L)*	*Chl a (mg/m2)*
Inner Coastal Plain diatometer sites																								
AN109		1	8.0	11	30	70	14	12	50	69	6.39	145	5.0	1.75	25.2	21.9	44.83	0.38	0.05	0.51	0.90	17.5	58.5	0.32
AN109	1	2	8.0	10	20	80	14	12	50	69	6.39	145	5.0	2	25.2	21.9	44.83	0.38	0.05	0.51	0.90	17.5	58.5	0.33
AN121	2	1	8.0	30	0	100	17	24	28	107	6.51	138	31.5	2.5	29.7	16.7	55.00	0.38	0.11	0.79	1.17	19.5	130	2.90
AN121		2	8.0	23	0	100	17	24	28	107	6.51	138	31.5	2.5	29.7	16.7	55.00	0.38	0.11	0.79	1.17	19.5	130	2.87
AN124		1	10.0	9	0	100	12	39	28	55	6.46	111	22.5	2.5	14.2	16.4	31.30	0.28	0.05	0.59	0.87	19.5	236	0.22
AN124	3	2	10.0	21	0	100	12	39	28	55	6.46	111	22.5	2.5	14.2	16.4	31.30	0.28	0.05	0.59	0.87	19.5	236	0.83
AN129	4	1	7.0	16	30	70	19	11	57	62	6.82	175	7.5	1	33.4	26.3	57.33	0.57	0.52	0.88	1.45	11.0	42.8	1.85
AN129		2	7.0	8	20	80	19	11	57	62	6.82	175	7.5	1.25	33.4	26.3	57.33	0.57	0.52	0.88	1.45	11.0	42.8	1.43
AN132	5	1	20.0	10	40	60	9	18	61	23	6.70	154	22.5	2	39.8	16.9	58.50	0.76	0.11	0.70	1.46	30.5	161	1.38
AN132		2	20.0	19	15	85	9	18	61	23	6.70	154	22.5	2	39.8	16.9	58.50	0.76	0.11	0.70	1.46	30.5	161	0.87
AN133		1	2.5	7	60	40	12	7	61	11	6.14	224	7.5	2	3.4	38.3	60.85	0.44	0.13	0.33	0.77	1.0	23.5	0.60
AN133	6	2	2.5	2	10	90	12	7	61	11	6.14	224	7.5	1.75	3.4	38.3	60.85	0.44	0.13	0.33	0.77	1.0	23.5	1.58
AN136	7	1	3.5	13	0	100	11	7	58	12	6.01	127	20.0	2.25	12.2	13.0	36.80	0.84	0.07	0.85	1.69	27.5	195	0.25
AN136		2	3.5	18	0	100	11	7	58	12	6.01	127	20.0	2.5	12.2	13.0	36.80	0.84	0.07	0.85	1.69	27.5	195	0.32
AN139	8	1	3.0	21	10	90	10	8	69	8	5.90	170	15.0	2	6.8	21.0	58.10	0.98	0.10	0.56	1.53	19.0	154	2.32
AN139		2	3.0	4	20	80	10	8	69	8	5.90	170	15.0	2	6.8	21.0	58.10	0.98	0.10	0.56	1.53	19.0	154	1.58
AN151	9	1	17.5	7	0	100	13	52	10	375	6.07	70	27.5	2.5	7.4	13.2	21.90	0.16	0.35	0.87	1.03	13.5	81.0	0.17
AN151		2	17.5	11	0	100	13	52	10	375	6.07	70	27.5	2.5	7.4	13.2	21.90	0.16	0.35	0.87	1.03	13.5	81.0	1.23
AN166	10	1	3.5	8	45	55	39	20	6	31	6.41	129	17.5	2	24.2	19.4	41.00	0.20	0.11	0.67	0.88	18.0	75.5	0.58
AN166		2	3.5	6	0	100	39	20	6	31	6.41	129	17.5	1.5	24.2	19.4	41.00	0.20	0.11	0.67	0.88	18.0	75.5	0.21
AN169	11	1	12.0	4	0	100	40	27	7	127	6.48	119	20.0	1.5	17.4	21.5	27.50	0.22	0.08	0.55	0.78	19.0	63.5	0.71

AN384	12	1	7.0	9	20	80	49	5	28	29	6.14	168	2.5	2.25	22.2	27.6	39.35	0.82	0.06	0.51	1.33	16.0	47.0	0.54
AN384		2	7.0	10	40	60	49	5	28	29	6.14	168	2.5	2.25	22.2	27.6	39.35	0.82	0.06	0.51	1.33	16.0	47.0	0.71
AN439	13	1	6.0	9	5	95	30	14	31	53	6.58	209	5.0	2	9.8	48.6	44.85	0.71	0.05	0.31	1.02	8.5	38.5	2.77
AN439		2	6.0	15	0	100	30	14	31	53	6.58	209	5.0	1.75	9.8	48.6	44.85	0.71	0.05	0.31	1.02	8.5	38.5	2.64
AN440	14	1	5.0	13	0	100	31	22	25	98	6.01	179	12.5	1.75	6.8	38.0	42.70	0.59	0.06	0.45	1.04	2.0	27.5	2.08
AN440		2	5.0	10	20	80	31	22	25	98	6.01	179	12.5	2	6.8	38.0	42.70	0.59	0.06	0.45	1.04	2.0	27.5	1.68
AN448	15	1	6.0	10	5	95	54	13	15	62	6.36	227	7.5	1.25	19.0	43.9	56.80	0.52	0.07	0.38	0.90	6.0	31.0	0.91
AN448		2	6.0	11	0	100	54	13	15	62	6.36	227	7.5	1.25	19.0	43.9	56.80	0.52	0.07	0.38	0.90	6.0	31.0	0.78
AN451		1	10.0	10	0	90	46	15	14	109	6.61	405	4.0	2	16.8	56.8	108.5	14.0	0.11	0.70	14.7	2.0	65.5	0.44
AN451	16	2	10.0	14	0	100	46	15	14	109	6.61	405	4.0	2	16.8	56.8	108.5	14.0	0.11	0.70	14.7	2.0	65.5	1.18
AN466		1	2.0	15	0	100	50	16	19	16	7.12	303	2.5	1.5	55.7	45.5	109.9	0.73	0.03	0.18	0.92	21.0	45.5	0.70
AN466	17	2	2.0	11	10	90	50	16	19	16	7.12	303	2.5	1.75	55.7	45.5	109.9	0.73	0.03	0.18	0.92	21.0	45.5	0.85
AN470	18	1	2.5	17	0	100	47	11	25	26	7.18	334	1.0	2	49.8	48.3	117.5	0.90	0.01	0.21	1.11	29.0	62.0	0.55
AN470		2	2.5	14	0	100	47	11	25	26	7.18	334	1.0	1.75	49.8	48.3	117.5	0.90	0.01	0.21	1.11	29.0	62.0	0.42
AN488	19	1	1.8	14	40	60	37	11	37	5	6.88	233	12.5	1.5	27.6	36.2	53.78	0.16	0.06	0.39	0.56	57.8	204	0.86
AN488		2	1.8	7	30	70	37	11	37	5	6.88	233	12.5	1.25	27.6	36.2	53.78	0.16	0.06	0.39	0.56	57.8	204	1.72
AN489	20	1	4.0	22	0	100	44	14	23	57	6.67	225	7.5	1.5	21.7	38.0	66.00	0.35	0.05	0.20	0.55	11.0	53.5	2.57
AN489		2	4.0	23	5	95	44	14	23	57	6.67	225	7.5	1.5	21.7	38.0	66.00	0.35	0.05	0.20	0.55	11.0	53.5	2.17
AN490	21	1	12.0	13	10	90	39	16	22	86	6.71	231	7.5	2	40.1	26.5	83.68	0.29	0.02	0.16	0.46	1.0	34.0	1.43
AN490		2	12.0	16	50	50	39	16	22	86	6.71	231	7.5	2.25	40.1	26.5	83.68	0.29	0.02	0.16	0.46	1.0	34.0	5.07
AN673	22	1	2.0	6	40	60	16	14	61	7	6.64	202	10.0	2	38.5	22.4	67.65	1.44	0.20	1.62	3.06	219	560	0.95
AN673	23	2	2.0	3	30	70	16	14	61	7	6.64	202	10.0	2	38.5	22.4	67.65	1.44	0.20	1.62	3.06	219	560	1.58
AN683	24	1	8.0	18	5	95	20	18	51	66	6.77	194	10.0	2	40.0	22.3	69.15	1.08	0.36	0.75	1.83	90.0	159	1.58
AN683		2	8.0	14	5	95	20	18	51	66	6.77	194	10.0	2	40.0	22.3	69.15	1.08	0.36	0.75	1.83	90.0	159	1.69
AN694	25	1	2.0	27	30	70	6	9	72	7	7.07	286	22.5	1.5	74.2	27.9	102.0	2.85	0.29	1.98	4.83	44.5	197	5.80
AN694		2	2.0	7	20	80	6	9	72	7	7.07	286	22.5	1.5	74.2	27.9	102.0	2.85	0.29	1.98	4.83	44.5	197	10.1
Piedmont resampling sites (rock substrate)																								
AN115		2	4.0	31	25	73	47	4	28	28	6.22	186	5.0	1.5	21.6	32.5	46.00	1.86	0.08	0.76	2.62	6.0	24.0	19.8
AN234		2	9.0	10	30	15	48	43	3	73	7.40	419	0.0	2.0	58.8	87.5	120.3	1.64	0.05	0.29	1.93	17.0	33.0	32.4
AN374		2	35.0	100	35	20	32	39	22	485	7.87	272	10.0	3.0	59.6	39.8	92.30	1.26	0.01	0.21	1.47	22.0	30.0	24.2
AN405		2	12.0	35	50	25	32	28	27	56	7.67	387	0.0	1.5	66.0	35.2	130.0	4.93	0.01	0.41	5.33	45.0	52.0	80.3

Table 3. Spearman's rank-order correlation between Chl *a* and chemical and physical site characteristics (*n*=47).
AvgW: average stream width, Open: % open canopy cover, GR: % gravel, SD: % sand, ST: % silt; CL: %clay, SDSTCL% sand silt and clay, URB: %urban land-use, AG: %agriculture land-use, FOR: %forest land-use, BASIN: basin size, Cond: Conductivity, Water color, Flow (estimate: 1=slow, 2=moderate, 3=fast), Alk: total alkalinity, Cl: chloride, Hardn: Total hardness, NO3-N: dissolved nitrate, NH3-N: dissolved ammonia, TKN: total Kjeldahl nitrogen, TN: Total Nitrogen, O-P: orthophosphate, TP: total phosphorus, Chl a: Chlorophyll a.

r = correlation coefficient
p- value: ** correlation significant at the 0.01 level (2-tailed test)
* correlation significant at the 0.05 level (2-tailed test)

	Chl a	
	r	*p-value*
AvgW	-0.10	0.50
%Open	0.31	0.03*
GR	0.19	0.21
SDSTCL	-0.16	0.30
BASIN	-0.02	0.89
Temp	-0.19	0.21
pH	0.25	0.09
Cond	0.29	0.05*
Color	0.16	0.27
Flow	-0.20	0.18
Alk	0.14	0.35
Cl-	0.20	0.18
Hardn	0.32	0.03*
NO3-N	0.14	0.36
NH3-N	0.17	0.26
TKN	0.02	0.91
TN	0.09	0.54
O-P	-0.05	0.74
TP	-0.07	0.64

Table 4. Species apparent optima and tolerances estimated by WA.Only species with effective numbers of occurrence (Hill's N2) > 10 are shown. Species are sorted by increasing TP optima.

	Taxon Name	*N2*	*TP µg / L*		*TN µg / L*	
			Opt	Tol	Opt	Tol
56	Eunotia exigua (Bréb. Kütz.) Rabenhorst	10	54.40	2.10	1235	2.93
150	Neidium alpinum Hustedt	10	58.57	1.78	1441	3.15
240	Encyonema minutum (Hilse) Mann	11	65.35	1.94	1132	2.13
1	Achnanthidium minutissimum (Kütz.) Czar.	16	68.31	2.02	1155	2.52
147	Navicula rostellata Kützing	12	68.91	2.22	1314	2.35
86	Frustulia vulgaris (Thwaites) DeT.	14	69.30	2.05	1327	2.25
93	Gomphonema gracile Ehr. emend. V. H.	13	70.63	2.31	1052	1.72
118	Navicula rhynchocephala Kützing	15	72.87	2.09	1144	2.29
130	Navicula longicephala Hustedt	13	74.87	2.14	1436	2.49
125	Navicula tenelloides Hustedt	15	75.26	2.06	1099	2.35
209	Stauroneis smithii Grunow	10	78.43	2.34	1410	2.25
73	Eunotia bilunaris (Her.) Mills	17	80.88	2.25	1160	1.96
228	Synedra ulna (Nitz.) Ehr.	15	81.88	1.96	1194	2.23
183	Nitzschia tubicola Grun. in Cl. et Grun.	12	82.15	2.07	1674	2.40
108	Meridion circulare (Grev.) Ag.	10	82.38	2.01	1166	2.03
135	Navicula germainii Wallace	17	85.04	2.44	1244	2.09
177	Nitzschia palea var. debilis (Kützing) Grun.	15	85.13	2.34	1312	2.24
113	Navicula cryptocephala Kützing	15	85.69	2.19	1298	2.12
94	Gomphonema parvulum (Kütz.) Kütz.	20	86.13	2.42	1185	2.08
81	Fragilaria capucina var. gracilis (Oest.) Hustdt.	14	87.42	2.21	1125	2.01
116	Navicula minima Grunow	16	89.03	2.13	1301	2.18
160	Nitzschia palea (Kütz.) Smith	17	90.58	2.30	1219	2.03
251	Planothidium frequentissimum (L.-B.) L.-B.	13	93.37	2.40	1135	1.82
195	Pinnularia microstauron (Ehr.) Cl.	15	94.21	2.25	1437	2.54
176	Nitzschia sociabilis Hustedt	13	95.83	2.04	1472	2.45
249	Planothidium lanceolatum (Bréb. Kütz.) L.-B.	10	96.08	2.34	1204	1.74
254	Sellaphora pupula (Kütz.) Meresckowsky	18	96.13	2.21	1240	2.10
78	Fragilaria vaucheriae (Kütz.) Petersen	13	97.19	2.17	1172	2.21
155	Nitzschia capitellata Hustedt	12	98.59	2.11	1270	2.24
25	Caloneis bacillum (Grun.) Cleve	13	99.58	2.39	1351	1.87
276	Placoneis clementis (Grun) Cox	11	100.21	2.20	1835	3.03
292	Hippodonta capitata (Ehr.) L-B., Metz. et Witk.	12	100.97	2.64	1180	1.68
260	Tryblionella debilis Arnott	10	104.12	2.40	1261	1.55
255	Sellaphora seminulum (Grun.) Mann	14	112.39	2.62	1219	1.93
217	Surirella angusta Kützing	11	113.30	2.26	1128	1.64
246	Luticola mutica (Kütz.) Mann	10	115.44	2.22	1274	2.58

Table 5. Pearson correlation between all chemical and physical site characteristics (*n*=47).
AvgW: average stream width, Open: % open canopy cover, GR: % gravel, SD: % sand, ST: % silt; CL: %clay, SDSTCL% sand silt and clay, URB: %urban land-use, AG: %agriculture land-use, FOR: %forest land-use, BASIN: basin size, Cond: Conductivity, Water color, Flow (estimate: 1=slow, 2=moderate, 3=fast), Alk: total alkalinity, Cl: chloride, Hardn: Total hardness, NO3-N: dissolved nitrate, NH3-N: dissolved ammonia, TKN: total Kjeldahl nitrogen, TN: Total Nitrogen, O-P: orthophosphate, TP: total phosphorus, Chl a: Chlorophyll a.

	Avg W	*Open*	*%GR*	*SdSt Cl*	*Urb %*	*For %*	*Ag %*	*Basin*	*Temp*	*pH*	*Cond*	*Color*	*Flow*	*Alk*	*Cl*	*Hardn*	*NO3-N*	*NH3-N*	*TKN*	*TN*	*O-P*	*TP*
Open	-0.07	1.00																				
%GR	-0.13	-0.22	1.00																			
SdStCl	0.13	0.22	-1.00	1.00																		
Urb%	-0.14	-0.13	-0.19	0.19	1.00																	
For%	0.61	-0.01	-0.30	0.30	-0.16	1.00																
Ag %	-0.27	0.09	0.39	-0.39	-0.74	-0.48	1.00															
Basin	0.61	-0.11	-0.39	0.39	-0.04	0.80	-0.48	1.00														
Temp	0.26	0.08	0.24	-0.24	-0.17	0.13	-0.01	0.08	1.00													
pH	-0.15	0.24	0.27	-0.27	0.24	-0.16	0.01	-0.28	-0.21	1.00												
Cond	-0.37	0.12	-0.06	0.06	0.46	-0.48	-0.02	-0.34	-0.50	0.58	1.00											
Color	0.39	0.32	0.01	-0.01	-0.56	0.58	0.05	0.37	0.38	-0.26	-0.61	1.00										
Flow	0.30	0.06	-0.26	0.26	-0.30	0.43	-0.09	0.30	-0.07	-0.34	-0.33	0.36	1.00									
Alk	-0.17	0.28	0.40	-0.40	0.02	-0.24	0.25	-0.32	-0.29	0.83	0.43	-0.08	-0.21	1.00								
Cl	-0.34	-0.05	-0.24	0.24	0.66	-0.36	-0.30	-0.20	-0.39	0.35	0.83	-0.71	-0.41	0.04	1.00							
Hardn	-0.32	0.25	0.01	-0.01	0.28	-0.41	0.11	-0.34	-0.48	0.71	0.91	-0.46	-0.22	0.68	0.58	1.00						
NO3N	0.06	0.06	-0.13	0.13	0.18	-0.10	-0.09	0.05	-0.13	0.11	0.63	-0.21	0.03	0.01	0.46	0.47	1.00					
NH3N	0.13	0.05	0.25	-0.25	-0.45	0.17	0.36	0.31	0.30	0.11	-0.17	0.21	-0.23	0.20	-0.31	-0.05	0.03	1.00				
TKN	-0.13	0.03	0.41	-0.41	-0.57	0.02	0.54	-0.01	0.01	0.11	-0.07	0.38	0.04	0.41	-0.41	0.06	0.17	0.56	1.00			
TN	0.04	0.06	-0.06	0.06	0.08	-0.10	0.00	0.05	-0.12	0.13	0.60	-0.14	0.04	0.08	0.37	0.46	0.99	0.12	0.33	1.00		
O-P	-0.30	-0.30	0.39	-0.39	-0.28	-0.09	0.44	-0.17	-0.15	0.22	0.01	-0.03	0.08	0.34	-0.24	0.11	-0.02	0.25	0.67	0.09	1.00	
TP	-0.25	-0.16	0.38	-0.38	-0.44	0.00	0.50	-0.19	-0.08	0.12	-0.08	0.19	0.21	0.26	-0.37	0.02	0.01	0.19	0.73	0.12	0.94	1.00

Table 6. Results of CCA's constrained to TP, pH and basin size. All analyses included the three variables only (TP, pH and basin size) that significantly ($p < 0.05$) explained the variance in diatom species composition in the reduced ($n = 23$) dataset. To verify independent relationship between the TP and pH and basin size respectively, a CCA was run constrained to TP with pH and basin size as covariables at a time, to verify if the ratio $?_1/?_2$ would decrease . $?_1$: eigenvalue axis 1, $?_2$

variable	covariable	$?_1$	$?_2$	$?_1/?_2$	permutation test (999 permutations)	% variance explained by axis 1
TP	none	0.126	0.247	0.510	0.048	6.0
	pH	0.125	0.246	0.508	0.001	6.4
	Basin size	0.135	0.245	0.551	0.046	6.7
pH	none	0.155	0.253	0.613	0.001	7.3
Basin size	none	0.117	0.253	0.462	0.046	5.5

Table 7. Performance of weighted-averaging (WA) and weighted-averaging partial least (WA-PLS) squares regression and calibration models developed for the full (*n*=25) and the reduced (*n*=22) sample dataset. All TP units are in log10 µg L^{-1}. WA-inv.desh.: WA-inverse deshrinking, WA-class.desh.: WA-classical deshrinking. The r^2 and RMSE in parentheses were derived from bootstrapping ($_{boot}$) and jackknifing ($_{jack}$). *n* = Number of samples included in model. Numbers in bold indicate the model with the highest performance (r^2_{jack}) and the lowest error ($RMSE_{jack}$). For comparison the performance of the best WA-PLS (2 component) TP model developed using the northern Piedmont dataset (*n*=91) was r^2 = 0.87, $r^2_{(boot)}$= 0.72, RMSE: 0.15; $RMSEP_{(boot)}$= 0.23, Mean $bias_{(boot)}$= -0.010, max bias $_{(boot)}$= 0.45.

n	method	variable	r^2 (r^2_{boot})	RMSE (RMSEP $_{boot}$)	r^2 (r^2_{jack})	RMSE (RMSEP $_{jack}$)	Mean bias $_{boot}$	Max bias $_{boot}$	Mean bias $_{jack}$	Max bias $_{jack}$
n= 25	WA-inv. desh.	TP	0.87 (0.19)	0.14 (0.36)	0.87 (0.27)	0.14 (0.33)	0.003	0.71	0.005	0.63
n= 25	WA-class. desh.	TP	0.87 (0.21)	0.15 (0.36)	0.87 (0.28)	0.15 (0.32)	0.002	0.69	0.005	0.60
n= 22	WA-inv. desh.	TP	0.89 (0.09)	0.12 (0.36)	0.89 (0.14)	0.12 (0.34)	0.006	0.87	0.009	0.86
n= 22	WA-class. desh.	TP	0.89 (0.10)	0.13 (0.36)	0.89 (0.13)	0.13 (0.33)	0.005	0.87	0.010	0.86
n= 25	WA-PLS-2 components	TP	0.97 (0.27)	0.06 (0.35)	**0.97 (0.34)**	**0.06 (0.31)**	0.012	0.69	0.017	0.61
n= 22	WA-PLS-3 components	TP	0.99 (0.21)	0.03 (0.34)	0.99 (0.27)	0.03 (0.31)	0.009	0.81	0.013	0.76
	Attempts to improve model performance using WA-PLS bootstrapping (see chapter 4.4.4)									
a) site 673 samples (highest TP site) taken out										
n= 23	WA-PLS-2 components	TP	0.97 (0.24)	0.06 (0.28)			-0.009	0.39		
b) all species > 0.3 % (% realtive abundances - no data transformation)										
n= 25	WA-PLS-2 components	TP	0.86 (0.14)	0.11 (0.31)			-0.007	0.38		
c) all species (square root data transformation)										
n= 25	WA-PLS-2 components	TP	0.98 (0.25)	0.06 (0.35)			-0.015	0.71		
d) all species > 0.3 % (square root data transformation), all samples < pH 6.8										
n= 20	WA-PLS-2 components	TP	0.97 (0.33)	0.07 (0.35)			-0.001	0.64		
e) all species > 0.3 % (square root data transformation), all samples > pH 6.1										
n= 20	WA-PLS-2 components	TP	0.98 (0.33)	0.05 (0.34)			-0.027	0.39		

Table 8. Spearman's rank-order correlation between diversity metrics and chemical and physical site characteristics (*n*=25).

Diversity metrics: C/P: the ratio Centrales/Pennales, # Taxa: number of diatom taxa in the sample, Percent Dominants: percent of total diatom valves made up of taxa that occurred in >10% abundance, S-W Index: Shannon-Weiner diversity index, %Ach_min: the percent total abundance of the taxon Achnanthidium minutissimum, % Siltation Index: Siltation Index (see chapter 3.5.2.5 for explanation).
Chemical and physical variables: AvgW: average stream width, Open: % open canopy cover, GR: % gravel, SD: % sand, ST: % silt; CL: %clay, SDSTCL% sand silt and clay, URB: %urban land-use, AG: %agriculture land-use, FOR: %forest land-use, BASIN: basin size, Cond: Conductivity, Water color, Flow (estimate: 1=slow, 2=moderate, 3=fast), Alk: total alkalinity, Cl^-: chloride, Hardn: Total hardness, NO3-N: dissolved nitrate, NH3-N: dissolved ammonia, TKN: total Kjeldahl nitrogen, TN: Total Nitrogen, O-P: orthophosphate, TP: total phosphorus, Chl a: Chlorophyll a.

r = correlation coefficient
p- value: ** correlation significant at the 0.01 level (2-tailed test)
* correlation significant at the 0.05 level (2-tailed test)

	C/P		# Taxa		Percent Dominants		% Ach_min		S-W Index		% Siltation Index	
	r	*p-value*	*r*	*p-value*	*r*	*p-value*	*r*	*p-value*	*r*	*p-value*	*r*	*p-value*
AvgW	0.24	0.24	0.40	0.05*	0.04	0.86	0.16	0.45	0.23	0.26	-0.29	0.15
%Open	0.13	0.54	-0.16	0.44	-0.07	0.75	0.11	0.60	0.12	0.58	0.30	0.14
%GR	-0.07	0.72	0.06	0.79	-0.28	0.18	0.05	0.82	0.09	0.67	0.31	0.14
%SD	0.35	0.09	0.02	0.91	-0.04	0.85	-0.09	0.65	0.11	0.59	0.14	0.49
%ST	0.24	0.24	0.34	0.09	-0.14	0.51	0.36	0.08	0.19	0.36	-0.11	0.61
%CL	-0.23	0.27	-0.05	0.82	0.38	0.06	-0.07	0.73	-0.22	0.30	-0.50	0.01*
%STCL	-0.15	0.47	-0.01	0.98	0.35	0.09	0.02	0.92	-0.20	0.34	-0.50	0.01*
SDSTCL	0.07	0.72	-0.06	0.79	0.28	0.18	-0.05	0.82	-0.09	0.67	-0.31	0.14
% URB	0.36	0.08	0.20	0.34	-0.23	0.27	0.48	0.01*	0.27	0.20	-0.04	0.83
% FOR	0.31	0.13	0.02	0.92	-0.04	0.86	-0.17	0.41	0.07	0.73	-0.10	0.64
% AG	-0.46	0.02*	-0.35	0.09	0.19	0.35	-0.32	0.12	-0.32	0.12	0.31	0.13
BASIN	0.47	0.02*	0.30	0.15	0.04	0.83	0.22	0.29	0.21	0.31	-0.45	0.02*
Temp	-0.09	0.65	0.50	0.01*	-0.29	0.16	0.07	0.73	0.33	0.10	-0.25	0.23
pH	0.17	0.41	0.02	0.93	-0.57	0.00**	0.17	0.41	0.44	0.03*	0.78	0.00**
Cond	0.16	0.43	-0.35	0.09	-0.10	0.64	0.21	0.30	0.01	0.96	0.46	0.02*
Color	0.05	0.83	-0.08	0.69	0.04	0.85	-0.47	0.02*	-0.14	0.52	-0.11	0.60
Flow	-0.49	0.01*	0.10	0.64	0.37	0.07	-0.20	0.35	-0.23	0.26	-0.31	0.13
Alk	0.18	0.39	0.20	0.34	-0.56	0.00**	0.06	0.77	0.44	0.03*	0.69	0.00**
Cl^-	0.18	0.38	-0.33	0.10	0.02	0.93	0.27	0.19	-0.05	0.79	0.16	0.45
Hardn	0.05	0.82	-0.24	0.25	-0.17	0.40	0.01	0.95	0.09	0.66	0.66	0.00**
NO3_N	-0.13	0.55	-0.23	0.28	0.10	0.62	-0.30	0.14	-0.18	0.40	0.30	0.14
NH3-N	0.10	0.63	-0.01	0.96	-0.01	0.97	-0.37	0.07	-0.10	0.64	0.00	0.99
TKN	0.01	0.96	0.08	0.72	-0.05	0.83	-0.38	0.06	-0.05	0.81	0.03	0.88
TN	-0.07	0.75	-0.07	0.73	0.08	0.70	-0.44	0.03*	-0.15	0.48	0.13	0.52
O-P	-0.20	0.33	0.04	0.86	-0.39	0.05	-0.35	0.09	0.18	0.40	0.53	0.01**
TP	-0.19	0.37	-0.04	0.86	-0.19	0.35	-0.25	0.23	0.04	0.86	0.35	0.08
Chl_a	0.03	0.88	-0.67	0.00**	0.28	0.18	-0.10	0.63	-0.40	0.05*	0.17	0.41

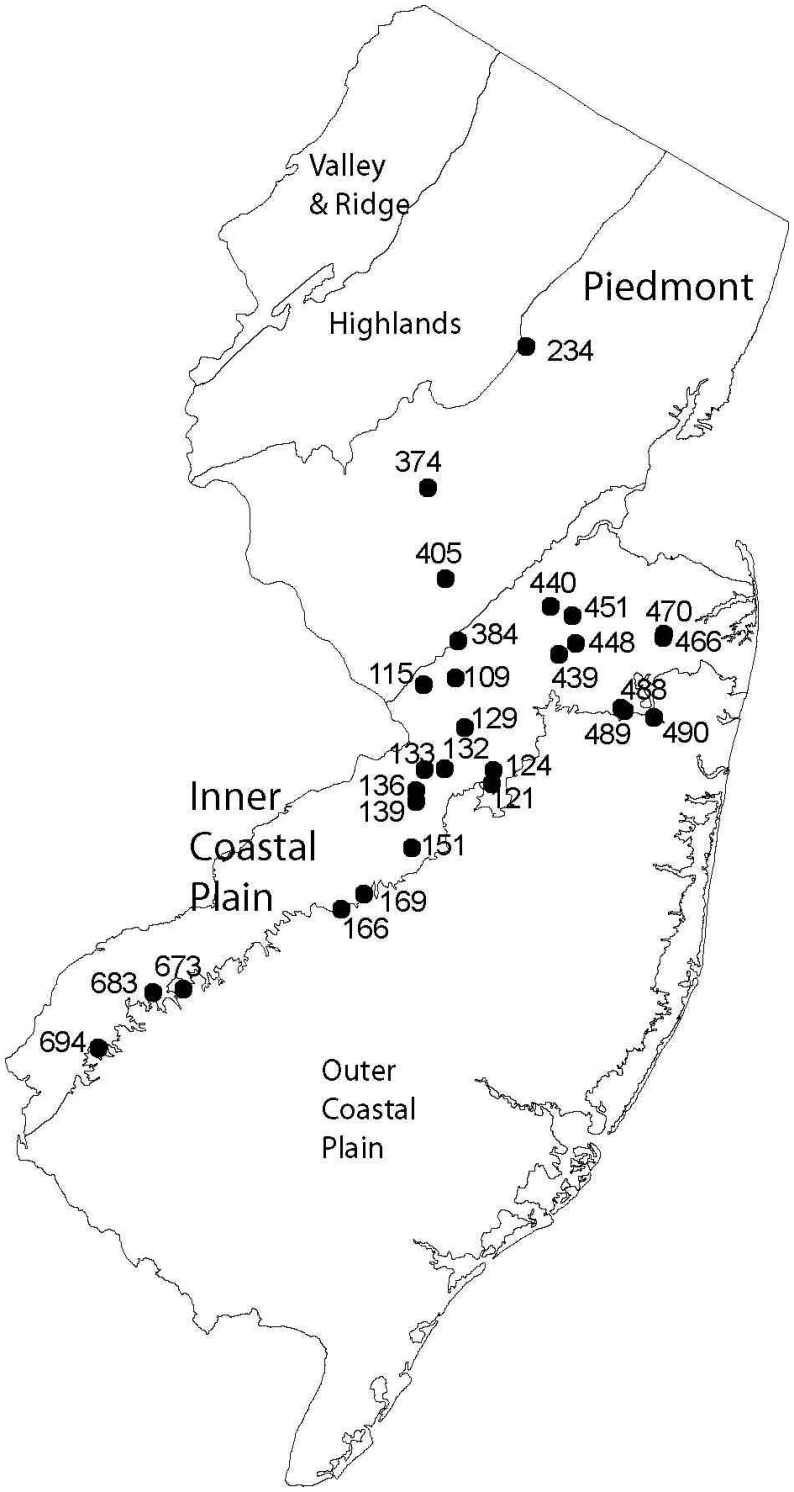

Figure 1. Site locations sampled during project year 4 in the Inner Coastal Plain (diatometer sites) and Piedmont (resampling sites) physiographic provinces of NJ. Site numbers correspond to New Jersey AMNET site location ID's. See Table 1 for site names and locations.

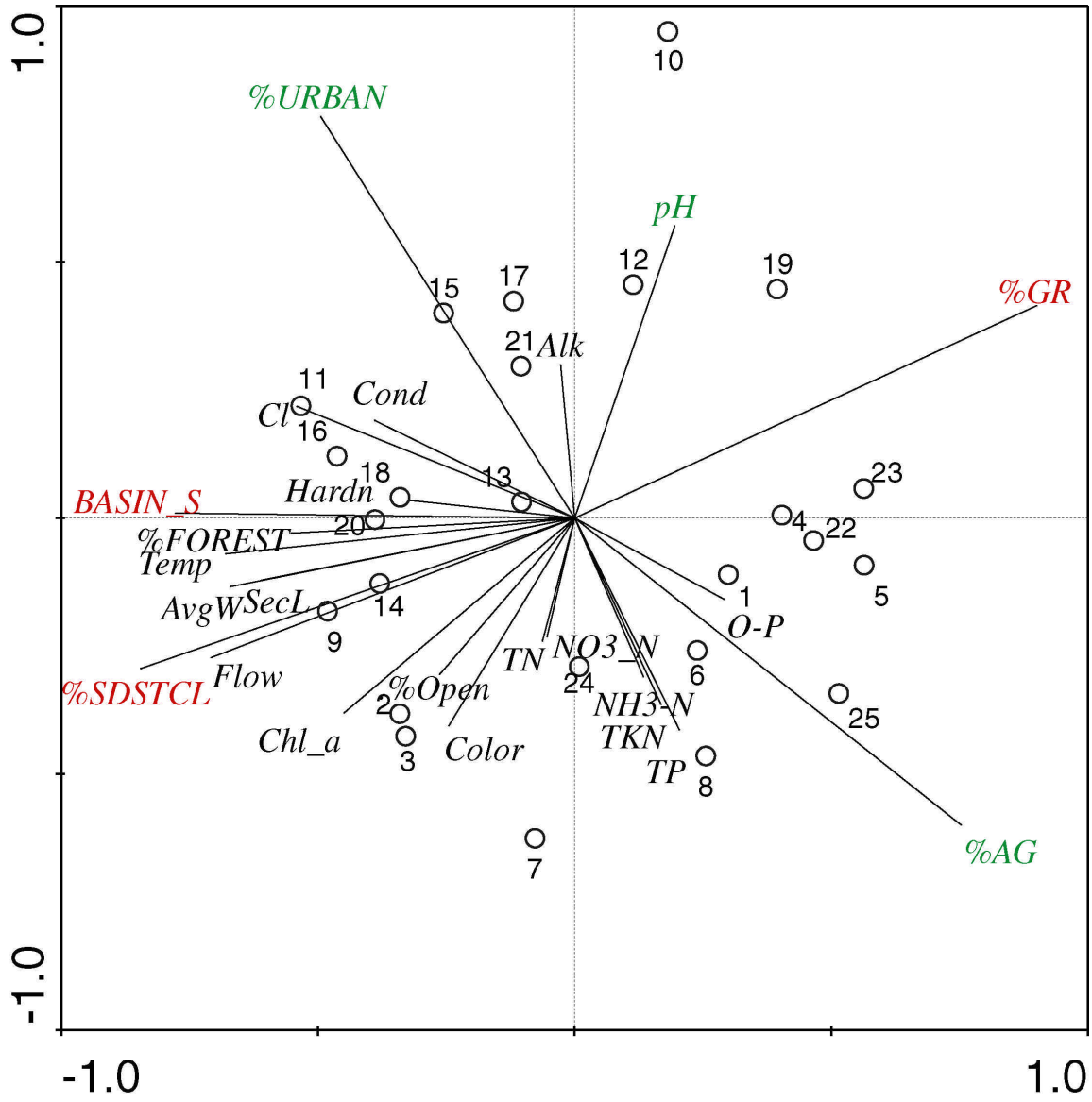

Figure 2. Principal components analysis (PCA) biplot showing 30 environmental variables (arrows) and 25 sites (circles). Horizontal axis = PCA axis 1; vertical axis = PCA axis 2. Abbreviations used for environmental variables and site numbers can be found in Table 2. The length of each arrow expresses the 'strength' of the influence of the variable on site distribution. Each axis is determined by a combination of variables. The three most important variables for the first axis are %gravel, %sand, silt and clay and basin size and for the second axis are %urban land-use, % agricultural land-use and pH.

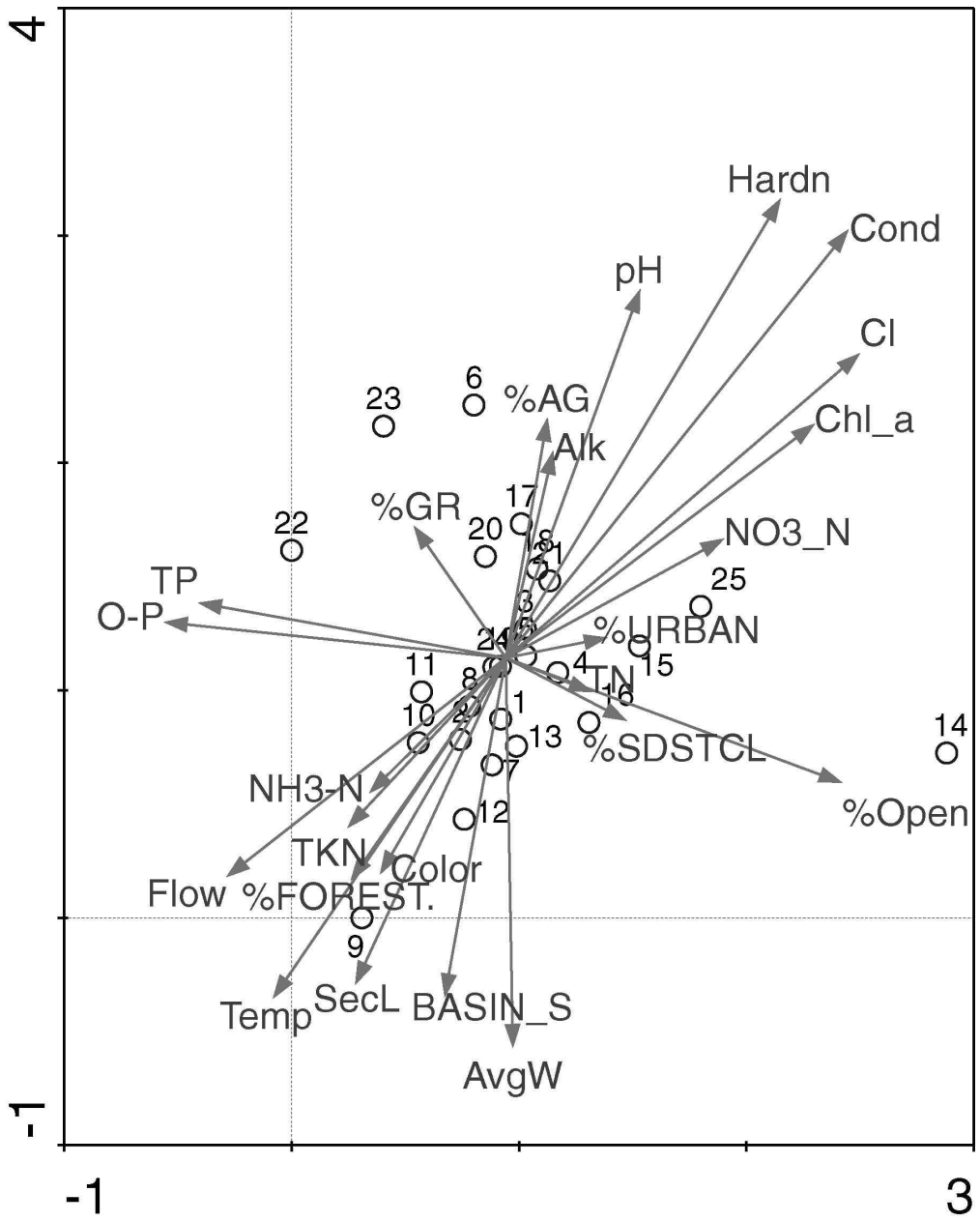

Figure 3. Detrended correspondence analysis (DCA) biplot showing samples and passive environmental variables. Horizontal axis = DCA axis 1; vertical axis = DCA axis 2. Abbreviations used for environmental variables and site numbers can be found in Table 2.

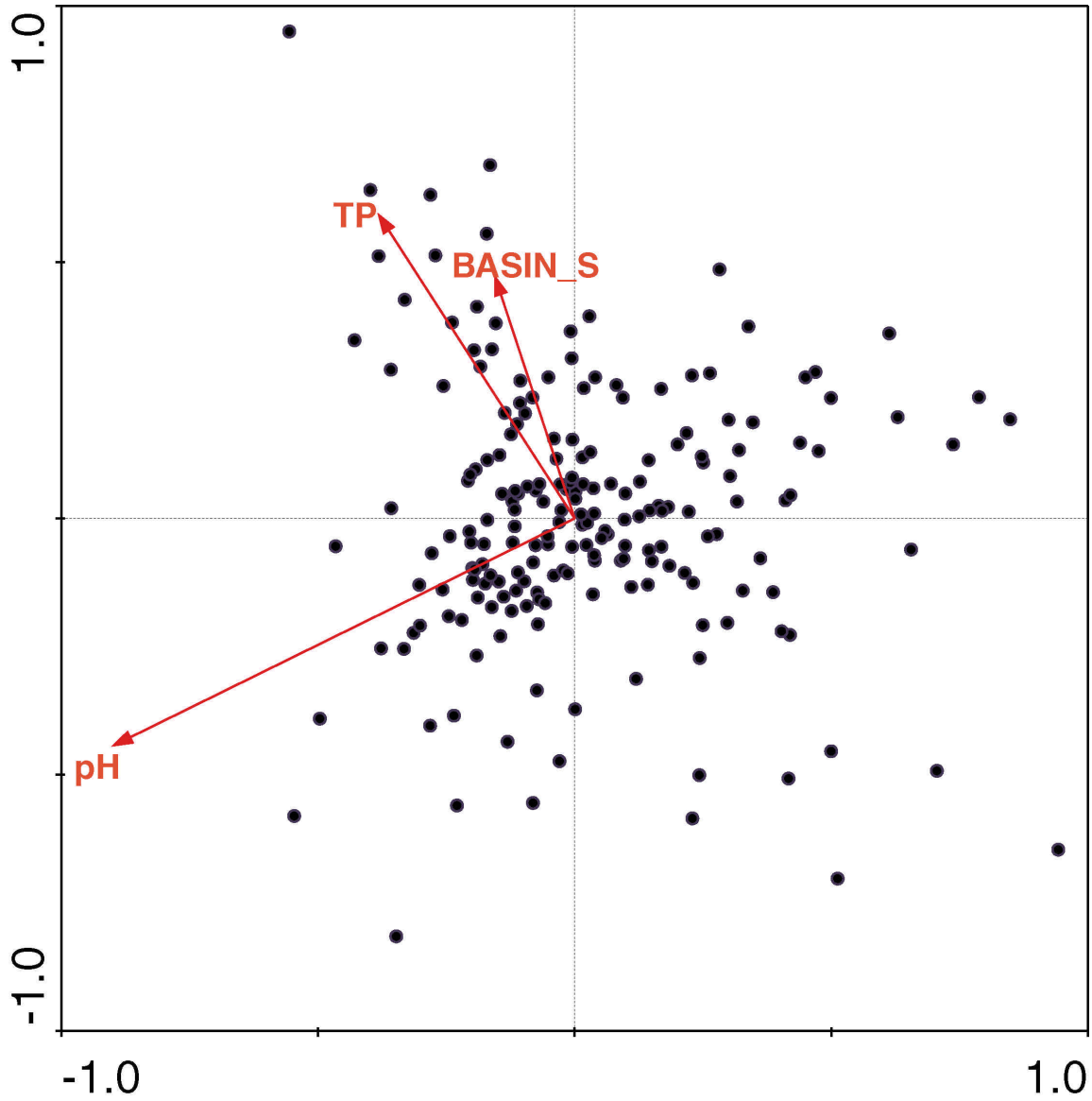

Figure 4. Canoncial Correspondence Analysis (CCA) biplot showing diatom taxa (filled circles) and the 3 variables with significant influence on diatom species composition (arrows). Horizontal axis = CCA axis 1; vertical axis = CCA axis 2.

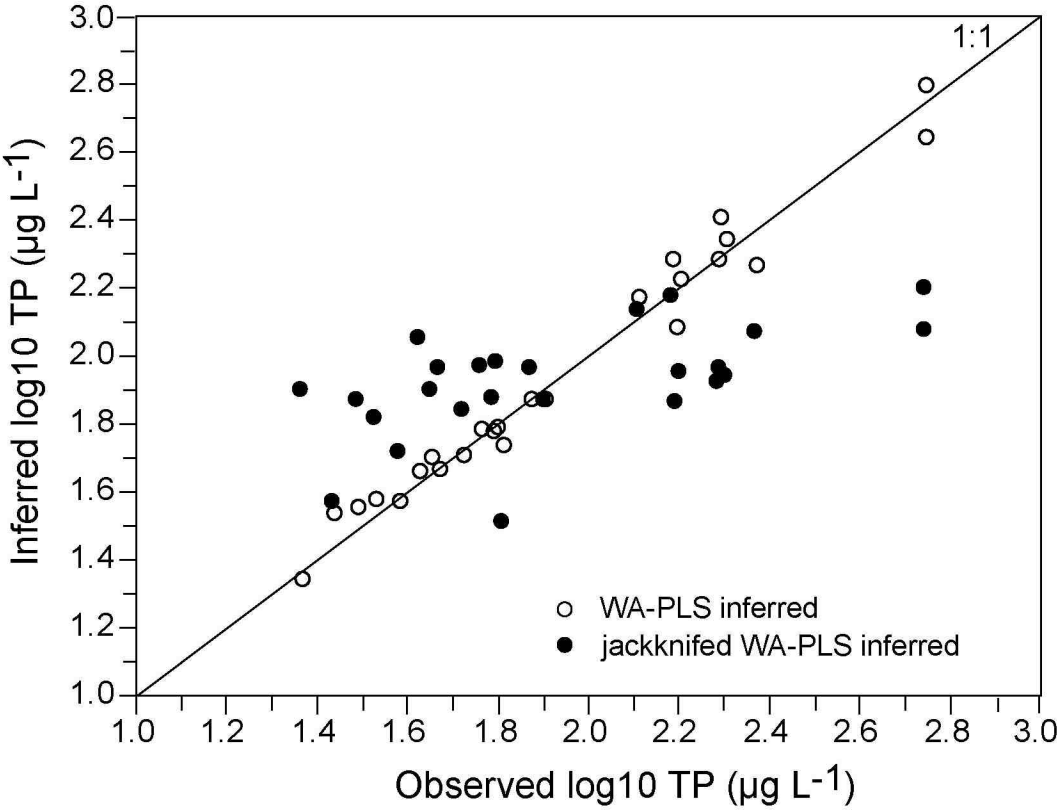

Figure 5. Plot of the best WA-PLS inference model (*n*=25; 2 components) showing diatom-inferred log 10 TP values versus observed log 10 TP values. The diagonal line is a 1:1 line. The Diatom TP Index score is shown for comparison.

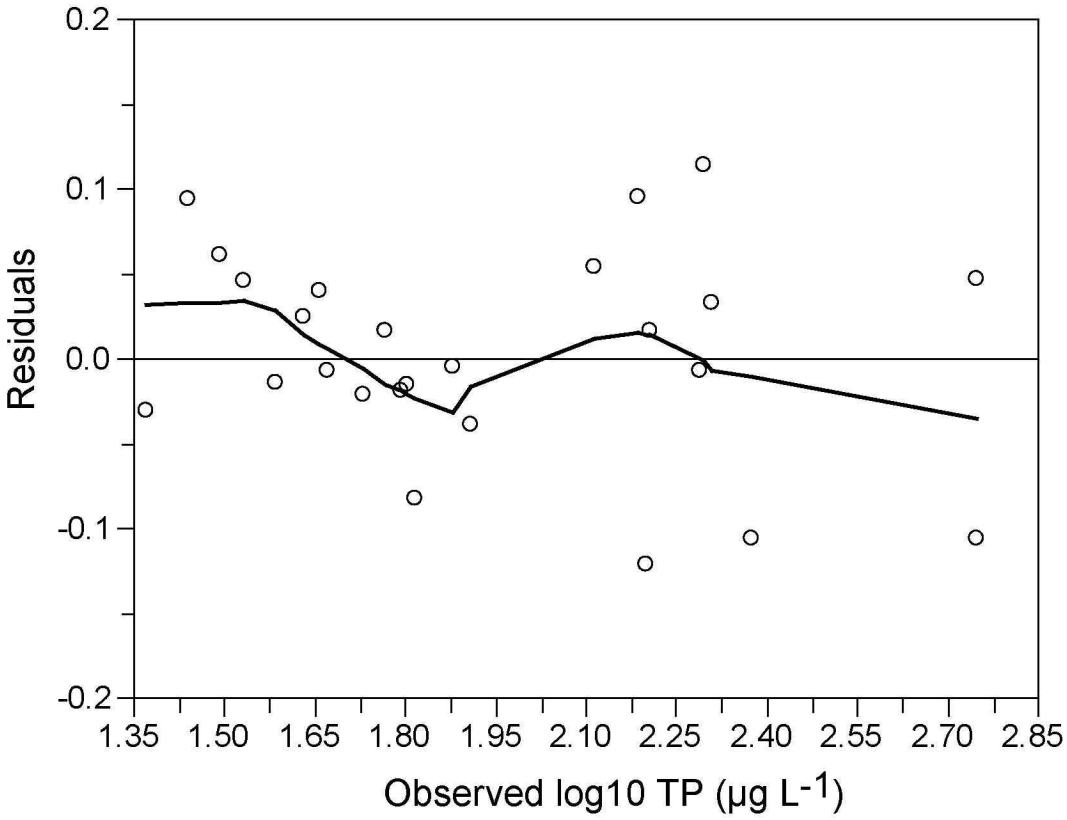

Figure 6. Plots of WA-PLS inference model showing residuals versus observed log 10 TP values. The solid line shows a LOESS scatter plot smoother (span = 0.45).

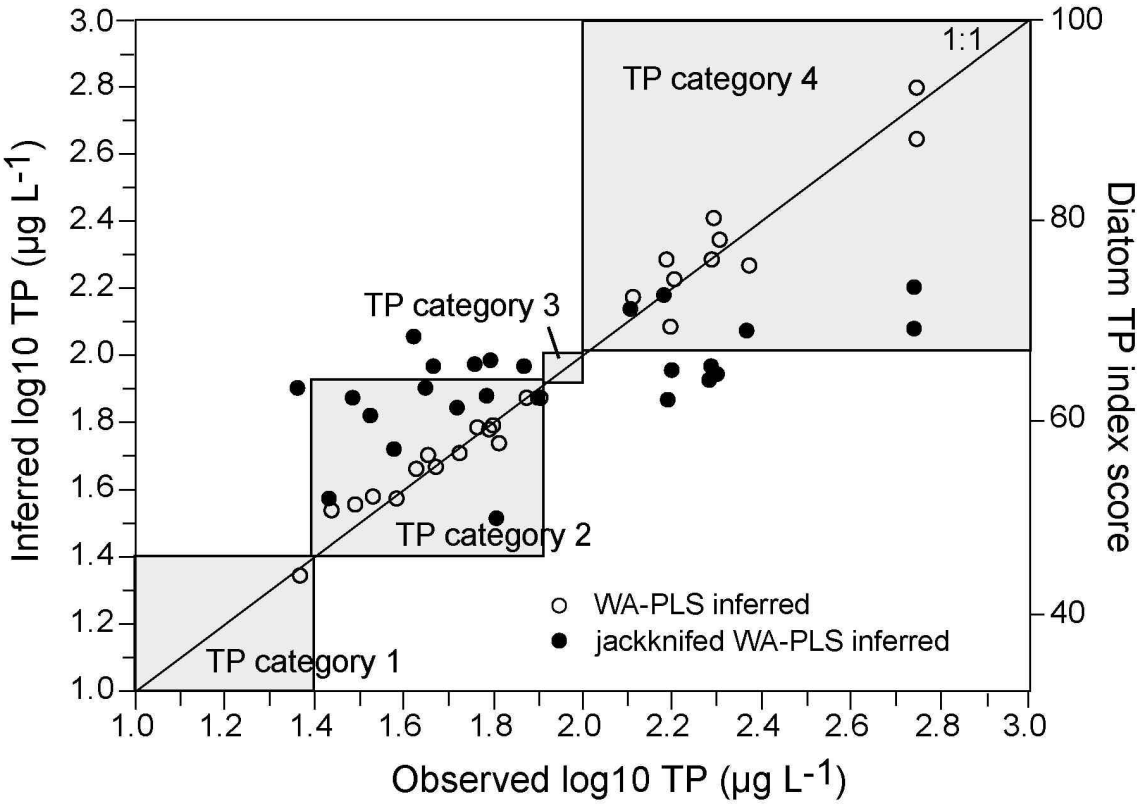

Figure 7. Plots of inferred bootstrapped versus observed TP and TN and the assignment of the Diatom TP Index scores to the corresponding nutrient categories. Shaded rectangles correspond to the water chemistry categories established for TP. TP categories are: 1 (< 0.025), 2 (0.025--0.075), 3 (0.075--0.1), 4 (> 0.1) mg / L TP.

Appendix 1: Results of Forward stepwise regression for the dependent variable Chl *a* and AFDM. Abbreviations used for variables are the same as listed under Table 2.

Forward Stepwise Regression: **Dependent Variable Chl a**

F-to-Enter: 4.000 P = 0.051
F-to-Remove: 3.900 P = 0.054

Step 0:
Standard Error of Estimate = 0.390
Analysis of Variance:

Group	DF	SS	MS	F	P
Residual	46	6.992	0.152		

Variables in Model

Group	Coef.	Std. Coeff.	Std. Error	F-to-Remove	P
Constant	3.027		0.0569		

Variables not in Model

Group	F-to-Enter	P
TP	0.000986	0.975
TN	0.380	0.541
NH3-N	1.905	0.174
pH	3.536	0.066
BASIN	0.562	0.457
SDSTCL	0.818	0.370
Flow	2.293	0.137
Cond	5.491	0.023
Open (%)	3.882	0.055

Step 1: Cond Entered

R = 0.330 Rsqr = 0.109 Adj Rsqr = 0.089
Standard Error of Estimate = 0.372

Analysis of Variance:

Group	DF	SS	MS	F	P
Regression	1	0.760	0.760	5.491	0.024
Residual	45	6.232	0.138		

Variables in Model

Group	Coef.	Std. Coeff.	Std. Error	F-to-Remove	P
Constant	1.282	0.747			
Cond	0.769	0.330	0.328	5.491	0.024

Variables not in Model

Group	F-to-Enter	P
TP	0.286	0.595
TN	0.0200	0.888
NH3-N	5.867	0.020
pH	0.434	0.514
BASIN	0.00131	0.971
SDSTCL	0.503	0.482
Flow	0.426	0.517
Open (%)-	3.429	0.071

Step 2: NH3-N Entered

R = 0.462 Rsqr = 0.214 Adj Rsqr = 0.178
Standard Error of Estimate = 0.354

Analysis of Variance:

Group	DF	SS	MS	F	P
Regression	2	1.494	0.747	5.976	0.005
Residual	44	5.499	0.125		

Variables in Model

Group	Coef.	Std. Coeff.	Std. Error	F-to-Remove	P
Constant	0.0561	0.871			
NH3-N	0.336	0.342	0.139	5.867	0.020
Cond	1.023	0.439	0.329	9.679	0.003

Variables not in Model

Group	F-to-Enter	P
TP	0.00560	0.941
TN	4.144	0.048
pH	0.301	0.586
BASIN	0.0387	0.845
SDSTCL	0.101	0.752
Flow	0.000104	0.992
Open (%)	5.705	0.021

Step 3: Open (%)- Entered

R = 0.553 Rsqr = 0.306 Adj Rsqr = 0.257
Standard Error of Estimate = 0.336

Analysis of Variance:

Group	DF	SS	MS	F	P
Regression	3	2.138	0.713	6.312	0.001
Residual	43	4.855	0.113		

Variables in Model

Group	Coef.	Std. Coeff.	Std. Error	F-to-Remove	P
Constant	-0.432	0.853			
NH3-N	0.381	0.388	0.133	8.208	0.006
Cond	0.994	0.427	0.313	10.113	0.003
Open (%)	0.134	0.308	0.0563	5.705	0.021

Variables not in Model

Group	F-to-Enter	P
TP	0.155	0.696
TN	5.235	0.027
pH	0.0000808	0.993
BASIN	0.0990	0.755
SDSTCL	0.922	0.342
Flow	0.0895	0.766

Step 4: TN Entered

R = 0.619 Rsqr = 0.383 Adj Rsqr = 0.324
Standard Error of Estimate = 0.321

Analysis of Variance:

Group	DF	SS	MS	F	P
Regression	4	2.676	0.669	6.509	<0.001
Residual	42	4.317	0.103		

Variables in Model

Group	Coef.	Std. Coeff.	Std. Error	F-to-Remove	P
Constant-	0.488	0.814			
TN	0.447	-0.366	0.195	5.235	0.027
NH3-N	0.595	0.605	0.158	14.246	<0.001
Cond	1.439	0.618	0.356	16.334	<0.001
Open (%)	0.140	0.321	0.0538	6.799	0.013

Variables not in Model

Group	F-to-Enter	P
TP	0.226	0.637
pH	0.292	0.592
BASIN	0.0275	0.869
SDSTCL	0.231	0.633
Flow	1.531	0.223

Summary Table

Step #	Vars. Entered	R	RSqr	Delta RSqr	Vars in Model
1	Cond	0.330	0.109	0.109	1
2	NH3-N	0.462	0.214	0.105	2
3	Open (%)-	0.553	0.306	0.0921	3
4	TN	0.619	0.383	0.0769	4

The dependent variable log10(-Chl a (m can be predicted from a linear combination of the independent variables:

	P
TN	0.027
NH3-N	<0.001
Cond	<0.001
Open (%)	0.013

The following variables did not significantly add to the ability of the equation to predict Chl a, and were not included in the final equation: TP, pH, BASIN, SDSTCL, Flow.

Normality Test: Passed (P = 0.163)

Constant Variance Test: Passed (P = 0.430)

Power of performed test with alpha = 0.050: 0.998