

Using change-point analysis and weighted averaging approaches to explore the relationships between common benthic diatoms and in-stream environmental variables in mid-atlantic highlands streams, USA


Christine L. Weilhoefer  Yangdong Pan

Received: 26 February 2008 / Revised: 3 June 2008 / Accepted: 16 June 2008 / Published online: 23 July 2008
© Springer Science+Business Media B.V. 2008

Abstract Diatoms are increasingly being used in the stream diatom assemblage. For the RT of taxa, the bioassessment of aquatic systems. However, where pH was an important predictor, higher pH autecological information for many common taxa is predicted higher relative abundances. In contrast, incomplete. We explored the potential of classification (CT) and regression tree (RT) approaches to some diatom taxa (e.g. *Achnanthes minutissimum* identify the hierarchical interaction among water (Kutz.) Czarnecki), while predicting higher relative quality variables in predicting the relative abundance abundances for other taxa (e.g. *Gomphonema olatum* (Brook.) Round & Bukht., *Gomphonema parvulum* (Kutz.) Kutz.). The environmental change used to identify environmental change points corresponding to major shifts in species abundances. We WA optima for all species. We suggest that RT also used traditional weighted-averaging approaches change point analysis can be used to complement (WA) to model taxon pH and total phosphorus (TP) traditional WA optima approaches, especially when optima. RT and WA approaches provided different, diatom taxa's abundances are affected by interactive yet complementary, information on the complex environmental factors, to provide more refined relationships between common stream diatoms and information on stream diatom environmental preferences. environmental variables. Both RT and CT highlighted the interaction of stream acidity (pH, acid neutralizing capacity (ANC)), and TP in structuring

Keywords Change point analysis · Classification and regression trees · Diatom · pH · Total phosphorus · Weighted-average

Handling editor: L. Naselli-Flores

C. L. Weilhoefer  · Y. Pan
Environmental Science and Resources, Portland State University, Portland, OR 97207, USA
e-mail: weilhoefer.christine@epamail.epa.gov

Present Address:

C. L. Weilhoefer
Pacific Coastal Ecology Branch, Western Ecology Division, US EPA, 2111 S.E. Marine Science Dr., Newport, OR 97365, USA

Introduction

Current knowledge of diatom autecology is incomplete, gleaned from studies which are not specifically designed to determine the environmental requirements of common species. Frequently cited environmental preferences are often based on qualitative best professional judgment, which were derived from studies within a limited geographic area or region (e.g.,

Lange-Bertalot, 1979; van Dam et al., 1994) or based on synthesis of studies with different study objectives, sampling designs, and various spatial and temporal scales (e.g., Lowd, 1974; Beaver, 1981; van Dam et al., 1994). As a consequence, species preference lists are often incomplete and inconsistent and autecological information about common species in many areas is lacking. More recently, numerical and modeling approaches have been employed to quantitatively characterize the relationships between diatoms and environmental variables in streams. While gradient analyses are commonly employed to elicit overall assemblage patterns (e.g., Biggs, 1990; Leland, 1995; Pan et al., 1996), species optima for key environmental variables are most often determined by weighted-averaging (WA) techniques (e.g., Pan et al., 1996; Leland & Porter, 2000; Winter & Duthie, 2000; Potapova & Charles, 2003).

WA techniques provide a quantitative evaluation of diatom autecology and in many cases expand our knowledge of diatom species preferences. In streams, WA models have been used to develop species optima and tolerances for conductivity (e.g., Leland, 1995; Leland et al., 2001; Munn et al., 2002; Potapova & Charles, 2003), pH (Kovács et al., 2006), phosphorus (e.g., Pan et al., 1996; Winter & Duthie, 2000; Sojinen & Niemela, 2002; Schenfelder et al., 2002; Potapova et al., 2004; Ponader et al., 2007), nitrogen (e.g., Leland, 1995; Ponader et al., 2007), sulfate (Potapova & Charles, 2003), major cations (Potapova & Charles, 2003), and dissolved inorganic carbon species (Schenfelder et al., 2002; Potapova & Charles, 2003). In addition, WA optima have successfully been used to reconstruct environmental conditions in lakes, including pH (e.g., Birks et al., 1990; Dixit et al., 1999) and total phosphorus (Dixit et al., 1992; Hall & Smol, 1992; Bennion, 1994), and wetlands (e.g., Gaiser & Taylor, 1995; Bunting et al., 1997; Cooper, 1999).

However, WA approaches suffer from the simplicity and assumptions of these models (see summary in Imbrie & Webb, 1981; Birks et al., 1990). Of primary concern when dealing with the relationships between diatom species and environmental variables is that WA modeling assumes that the variable of interest is the sole variable responsible for determining the species distribution. The importance of other environmental variables is implicitly included in the calculation of the WA optima.

However, WA cannot explicitly illustrate the interactions among environmental variables. Subsequently, environmental variables must be interpreted one at a time. In reality, most studies have displayed interactions among environmental predictors and stream algal assemblages. Interactive relationships between pH and nutrients were demonstrated in streams in the Illinois River Basin (Leland & Porter, 2000) and in Finland rivers (Sojinen & Niemela, 2002). The development of WA models, for example, nutrient models, has often been less-successful when other environmental variables, such as pH, are important in structuring the diatom assemblage (e.g. Hall & Smol, 1992; Reavie et al., 1995; Pan & Stevenson, 1996). Because WA cannot explicitly address other environmental variables that influence patterns of species abundances, the applicability of models developed in one area to other areas is questionable. A second assumption of WA modeling that often does not hold true for stream diatom assemblages is the idea that species abundance forms a unimodal relationship with the environmental variable of interest. While Gaussian responses of diatom abundance to physiological environmental variables (e.g., pH and salinity) have been reported (ter Braak & van Dam, 1989; Juggins, 1992), this relationship is often not the case for resource variables such as total phosphorus (Potapova et al., 2004). Recently, advanced regression techniques, such as generalized linear models (GLM) and generalized additive models (GAM), which place fewer assumptions and constraints on species–environmental relationships have been used to model species–environmental relationships (see Guisan et al., 2002 for review). Potapova et al. (2004) used GLM to model relationships between diatom relative abundance and total phosphorus for northern Piedmont streams. While these models implicitly incorporate biotic and environmental interactions (Guisan & Zimmerman, 2000), they are additive and still do not explicitly model interactions or incorporate hierarchical structure.

Relationships between diatoms and environmental variables in streams are complex, often with several variables operating through hierarchical interactions. Strong relationships between algal biomass and nutrients are not often shown in streams (e.g., Leland, 1995). Diatom autecology would benefit from approaches that allow interactions between variables to be explicitly incorporated. Increasingly, regression

tree approaches have been used to examine species–environmental relationships in plants and animals compared to classification models developed using (e.g., Iversen & Prasad 1998; DeOath & Fabricius 2000; O’Conner & Wagne 2004; Hershey et al. 2006). Regression trees (RT) and classification trees (CT) are useful for visually facilitating interpretation, revealing data structures, and displaying interactions (Clark & Pregibon 1993; DeOath & Fabricius 2000). RT and CT models performed better than analysis of variance and linear regression in predicting the abundance of coral taxa from environmental variables (DeOath & Fabricius 2000) and the authors suggest using regression tree approaches to select

more simple interaction terms for regression models.

CT models had higher predictive power than both GLM and GAM for modeling vegetation species (Franklin, 1998; Vayssières et al., 2000 in DeOath,

2002). Another advantage of RT analysis is that it can also be used to identify thresholds or change points

along environmental gradients, where species abun-

dances change from one state to another (Qian et al.,

2003). RT change point analysis was successfully

used to identify total phosphorus concentrations at

which the relative abundance of tolerant macroinver-

tebrate species shifted in wetlands (Qian et al. 2003)

and the level of wetland impairment at which

macrophytes, diatoms, and zooplankton abundances

shifted (Lougheed et al. 2007). While Pan et al.

(1999) used RT analysis to explore the hierarchical

relationships between algal biomass and environmen-

tal variables in streams, RT approaches have not been

used to explore individual diatom species distribution

in streams or to identify change points along envi-

ronmental gradients. RT and CT are non-parametric,

and therefore, they are well-suited for diatom species

relative abundance data that often contain many zero

values.

The objective of this study was to use both

weighted-average and regression tree approaches to

explore the relationships between environmental

variables and common diatom taxa in Mid-Atlantic

Highlands (MAHA) streams. In addition, RT analysis

was used to identify change points along environ-

mental gradients (pH and total phosphorus), where

taxa shift from low to high relative abundances. TP

and pH were selected for change point analysis

because these two variables have been shown to be

most influential in controlling MAHA stream diatom

assemblages (Pan et al. 1996). Regression tree

Models developed on relative abundance data were compared to classification models developed using both presence–absence and relative abundance categories to determine if different data transformations reveal different relationships between species and environmental variables. Change points identified by RT analysis were compared to optima derived from WA methods to determine if change point analysis can classify diatom taxa’s environmental preferences more precisely. This work will further our understanding of the autecology of common stream diatom taxa and of how environmental variables interact to determine their abundance.

Materials and methods

Study area

Streams were located in the Mid-Atlantic Highlands

region of the U.S.A. (Fig. 1). The region has been

delineated into several ecoregions (Omernik 1987

used to identify total phosphorus concentrations at

Level III ecoregion classification), which include the

Northern Appalachian Plateau and Uplands, the

North Central Appalachian Mountains, the Blue

Ridge Mountains, the Central Appalachian Ridges

and Valleys, the Central Appalachian Mountains, and

the Western Allegheny Plateau. Topography of the

ecoregions included a mixture of mountains, plateaus,

plains, and valleys. Data were collected as part of the

U. S. Environmental Protection Agency’s Environ-

mental Monitoring and Assessment Program

(EMAP). Stream sites in the region were selected

using a stratified random sampling design (Herlihy

et al., 1998). The stream population in the region was

sampled on the basis of digitized versions of U.S.

Geological Survey’s 1:100,000-scale topographic

maps. In order to get a more equitable distribution

of stream sizes, sample probabilities were set so that

roughly equal numbers of first-, second-, and third-

order streams would be sampled. A total of 256

unique sites, sampled for periphyton and water

chemistry once from late April to early July in

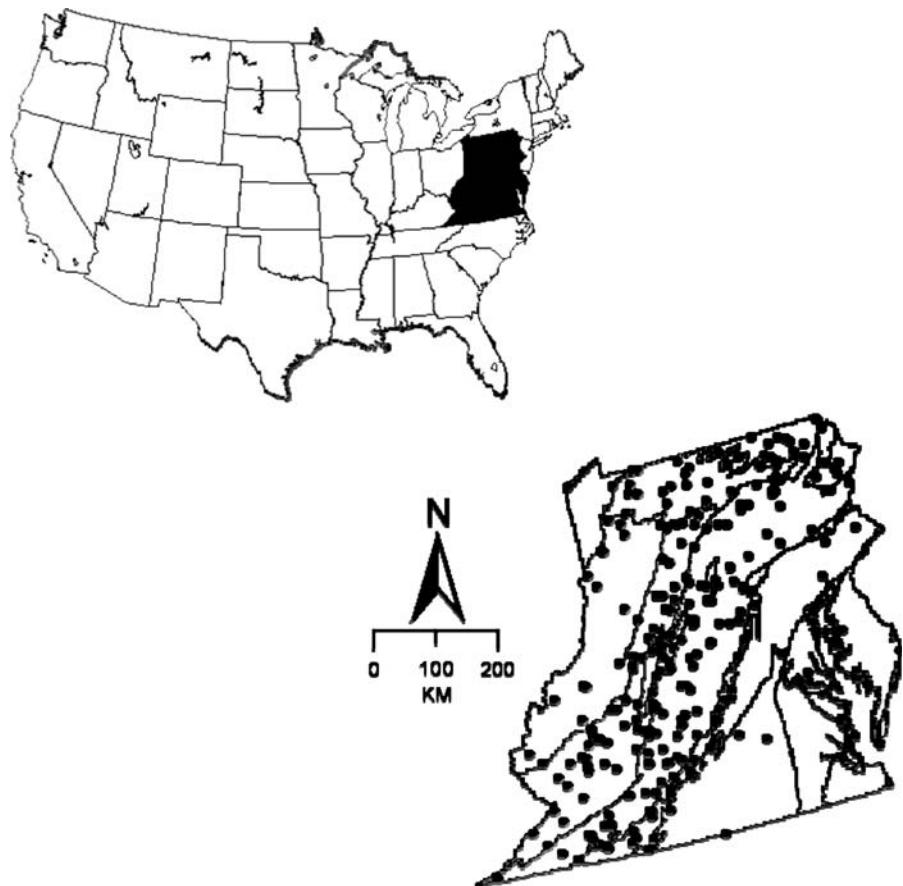
1993 and 1994, were used for this analysis.

Sampling design

A study reach was established around either side of

the selected sample sites with a total length equal to

Fig. 1 Map showing location of the Mid-Atlantic Highland region, U.S.A. and location of study sites. The solid lines are ecoregional boundaries



40 times the average wetted channel width (minimum length 150 m). Each study reach was evenly divided into 10 equal length intervals and 11 cross-section transects were set up (including one transect at the start and end of each reach).

Field sampling

Periphyton samples were collected from erosional habitats at each of the 11 transects and combined into a composite sample. Transects with visible water movement were considered erosional habitat. At each transect, periphyton was collected from a 12²-cm² area of the stream bed using a 1.5-cm long piece of 3.9-cm diameter PVC pipe as a template. Periphyton was scraped from the upper surfaces of cobbles with a toothbrush and rinsed with stream water. Composite periphyton samples were then preserved with formalin. Stream water samples were taken near the middle of the stream in flowing water. Detailed methodology can be found in Pan et al. (1996).

Detailed information on the analytical procedures used for each of the analyses can be found in U. S. Environmental Protection Agency (1987).

Laboratory analyses

An aliquot of homogenized algal suspension was acid-cleaned and mounted in HYRAX[®] to identify and enumerate diatom species (Patrick and Reimer, 1966). A minimum of 500 diatom valves were counted at 1,000 magnification. Diatom taxonomy followed mainly Krammer & Lange-Bertalot (1986, 1988, 1991a, b) and Patrick & Reimer (1966, 1975).

Data analysis

Weighted-average models: Weighted-averaging regression and calibration were used to quantify relationships between individual diatom taxon^Ōs relative abundance and environmental variables (Birks et al., 1990). The taxon^Ōs optima was calculated as

the mean of the measured environmental variables and additive models for regression problems. RT can be weighted by the abundance of this taxon in all sites. identify a set of important predictors (both numeric and categorical), automatically handle interactions between predictor variables, and illustrate hierarchical relationships among predictor variables (Venables and Ripley, 2002). Tolerance values were corrected for bias by taking into account the effective number of occurrences and using inverse de-shrinking methods. Cross-validation with leave-one-out jack-knifing was used to validate the models. In WA regression and calibration and model validation were performed using R v. 1.4 (Juggins 2003).

Regression tree change point analyses: RT analysis was used to explore the relationships between each of the 10 most commonly occurring benthic diatom taxa and in-stream environmental gradients. TP and pH change points were identified as the first or second splits of the RT for each diatom taxa. RT provides an alternative to linear

Table 1 Environmental variables (median (med), minimum (min), and maximum (max) values) used in predicting relative abundance of 10 most frequently occurring diatom taxa in Mid-Atlantic Highland streams. Frequency of occurrence of each variable in the regression trees (RT) and classification trees (CT) based on presence-absence data of all 10 species is also provided

Code	Variable	Med	Min	Max	Frequency	
					RT	CT
AM	Ammonium ($\mu\text{eq l}^{-1}$)	0	0	17	0	0
ANC	Acid neutralizing capacity ($\mu\text{eq l}^{-1}$)	271	0	5520	4	1
Ca	Calcium ($\mu\text{eq l}^{-1}$)	359	28	5240	0	0
Cl	Chloride ($\mu\text{eq l}^{-1}$)	69	8	716	1	3
COL	Color (PCU)	6	0	62	0	0
COND	Conductivity ($\mu\text{S cm}^{-1}$)	84	12	612	1	0
DIC	Dissolved inorganic carbon (mg l^{-1})	4	0	66	0	0
DOC	Dissolved organic carbon (mg l^{-1})	2	0	6	0	0
K	Potassium ($\mu\text{eq l}^{-1}$)	28	4	98	0	0
Mg	Magnesium ($\mu\text{eq l}^{-1}$)	183	19	2599	0	0
Na	Sodium ($\mu\text{eq l}^{-1}$)	106	7	1179	0	0
NN	Nitrate + Nitrite ($\mu\text{g l}^{-1}$)	18	0	722	1	2
pH	pH	7.2	3.4	8.4	5	1
SI	Silica (mg l^{-1})	6	1	19	1	1
SO4	Sulfate ($\mu\text{eq l}^{-1}$)	191	21	4103	2	0
TN	Total nitrogen ($\mu\text{g l}^{-1}$)	440	27	16260	0	0
TP	Total phosphorus ($\mu\text{g l}^{-1}$)	13	1	108	3	2
TSS	Total suspended solids (mg l^{-1})	4.6	0	206	0	0
TURB	Turbidity (NTU)	2.1	0.2	35	0	0

PCU: platinum-cobalt units, Conductivity reference temperature = 25°C, NTU: nephelometric turbidity units

Table 2 Summary data for 10 most frequently occurring diatom taxa (mean and maximum (max) relative abundance, and number sites) in Mid-Atlantic Highlands streams

Code	Taxon	Mean (%)	Max (%)	No. of sites
AB	<i>Achnanthydium biasolettianum</i> (Grun.) Bukht.	10	71	149
AM	<i>Achnanthydium minutissimum</i> (Kütz.) Czarnecki	25	80	239
EM	<i>Encyonema minutum</i> (Hilse ex Rab.) D. Mann	4	55	182
FC	<i>Fragilaria capucina</i> Desm.	5	29	189
GP	<i>Gomphonema parvulum</i> (Kütz.) Kütz.	2	13	134
ND	<i>Nitzschia dissipata</i> (Kütz.) Grun.	2	24	106
NP	<i>Nitzschia palea</i> (Kütz.) W. Sm.	1	13	93
PL	<i>Planorhynchium lanceolatum</i> (Bréb.) Round & Bukht.	3	32	126
RS	<i>Reimeria sinuata</i> (Greg.) Koc. & Stoerm.	1	14	106
SU	<i>Synedra ulna</i> (Nitz.) Ehr.	2	16	126

into two subsets or child nodes based on the decisionpath from root to leaf and attributes may be encountered more than once. For each split, alternative group homogeneity or purity. Decision rules are then splitting variables are presented and evaluated using applied to each group separately until either all an improvement index. The improvement index is classes have been sorted or the tree has recalculated as the number of sites in the branch times maximum complexity. The effectiveness of the the impurity index. It is the relative size of the decision rule at each split is evaluated as a function improvement index that gives an indication of the of the decrease in impurity achieved by dividing the utility of a variable in splitting the data rather than its sample according to that rule. In our study, impurity absolute value (Therneau & Atkinson1997).

within each child node was measured as deviance by A cross-validation procedure was used to determine the Gini Index (Therneau & Atkinson1997). The when to stop partitioning the data. Cross-validation deviation after each split is calculated as occurs by dividing the original data into several, mutually exclusive datasets, and producing trees of different sizes (numbers of splits). The dataset not used to build the series of trees is then used to evaluate the predictions of trees. The final tree size is selected by examining the plots of relative predictive error versus number of splits. For this study, the 1-SE stopping rule was used (Therneau & Atkinson1997). For this rule, any tree that is within one standard error of the tree with the lowest relative predictive error is considered as being equivalent to this tree and the simplest model (fewest number of splits) among those within 1-SE is selected. If splitting the relative abundance data based on measured environmental variables did not result in reduced predictive error, no tree was developed for that taxon.

$$D_{i\text{ child}} = D_{i,L} + D_{i,R}$$

where $D_{i,L}$ is the deviance in the left child node, $D_{i,R}$ is the deviance in the right child node, and $D_{i\text{ child}}$ is the total deviance in the split (Brieman et al1984). For any given node, the decision rule that maximizes the reduction in deviance ($\Delta D = D_i - D_{i\text{ child}}$) is selected. For regression trees, this reduction in deviance is equivalent to maximizing the between group sum of squares (Therneau & Atkinson1997).

For CT, the end point of each tree is characterized by the distribution of objects in each of the classes selected. For RT, the end point of each tree is the predicted mean of the response variable that taxon.

(e.g., relative abundance of *Achnanthydium minutissimum*), number of objects in each group, and the variables, allowing regression tree models to be hierarchical set of decision rules (response variables) applied directly. For the regression models, species that define it. Thus, the decision tree can be used to data were double-square-root transformed to stabilize variance in the species data. Due to the species-rich subset of attributes may be encountered on a particular nature of diatom data and fixed-count methodologies,

the relative abundance of a taxon is often zero at many sites, resulting in left-skewed data. As an attempt to deal with left-skewed data, we created trees for data transformed into presence-absence (a special case of classification with 2 groups). RT and CT analyses were performed using the rpart package for R (Therneau & Atkinson 1997; Atkinson & Therneau, 2000).

Results

Diatom species and environmental characteristics

Water chemistry variables are presented in Table 1. Stream water pH ranged between 3.4 and 8.4. Total phosphorus ranged between 1 and $108 \mu\text{g l}^{-1}$. A total of 619 diatom taxa were identified from the 256 stream sites. Average taxa richness at a site was 28 (range: 6–68). The relative abundances of the ten most common taxa (based on frequency) are presented in Table 2. *Achnanthydium minutissimum* (Kütz.) Czarnecki was the most common taxa, with a mean relative abundance of 25% and was present at 239 sites. This taxon dominated the diatom assemblage at most sites, having a relative abundance greater than 25% at 110 sites and a relative abundance greater than 10% at 179 sites. *Achnanthydium minutissimum* (Kütz.) Czarnecki, a small, stalked taxon, similar in morphology to *A. minutissimum*, was also very common, having a relative abundance greater than 10% at 68 sites. Other common taxa were much less dominant in a sample. Of the remaining 10 most common taxa, the number of sites in which their relative abundance was greater than 10% ranged from 2 to 35.

Weighted-average models

WA pH models had relatively high predictive power (WA $r^2 = 0.70$, WA jack-knifed $r^2 = 0.64$) and low root-mean squared error of prediction (WA RMSE = 0.44, WA jack-knifed RMSE = 0.49). Performance of WA TP models was lower (WA $r^2 = 0.33$, WA jack-knifed $r^2 = 0.30$, RMSE WA = $2.1 \mu\text{g l}^{-1}$, RMSE = $2.3 \mu\text{g l}^{-1}$). The pH optima for the ten most commonly occurring taxa were all approximately neutral to slightly alkaline, ranging from 7.0 (*Gomphonema parvulum* (Kütz.)

Table 3 Regression tree change points (CP) predicting higher relative abundance for pH and acid-neutralizing capacity (ANC), weighted-average pH optima, and van Dam et al. (1994) pH classification for 10 most frequently occurring diatom taxa in Mid-Atlantic Highlands streams

Taxon	CP	Optima	van Dam et al. (1994)
AB	pH > 6.7	7.3	Alkaliphilous
AM	pH > 6.1	7.2	Circumneutral
EM	ANC > 944	7.3	Circumneutral
FC	pH > 5.7	7.1	Circumneutral
GP	ANC > 1585	7.0	Circumneutral
ND	ANC > 1259	7.5	Alkaliphilous
NP		7.4	Circumneutral
PL	pH > 7.0	7.3	Alkaliphilous
RS	pH > 7.1	7.5	Circumneutral
SU		7.2	Alkaliphilous

Alkaliphilous taxa occur at mainly pH > 7, Circumneutral taxa occur mainly pH = 7

(Kütz.) to 7.5 (*Nitzschia dissipata* (Kütz.) Grun.; Table 3; Fig. 2). Species TP optima ranged from oligotrophic to mesotrophic based on the suggested trophic boundaries in streams presented in Dodds et al. (1998, Table 4; Fig. 3). *A. minutissimum* had the lowest TP optima ($11 \mu\text{g l}^{-1}$), while *Planorhynchium lanceolatum* (Breb.) Round & Bukht. had the highest TP optima ($30 \mu\text{g l}^{-1}$).

Regression tree change point analysis

Regression trees were developed for nine of the ten most commonly occurring taxa, with predictive power ranging between 0.18 and 0.40 (Fig. 4). Based on cross-validation, no RT could be developed for *Synedra ulna* (Nitz.) Ehr. We present detailed results of the RT for *A. minutissimum* relative abundance data as an illustrative example. The final RT had two splits (Fig. 4b, $r^2 = 0.34$). Further splits did not increase predictive power or reduce relative predictive error. The explanatory variables used were pH and TP. The first split was on pH, with low pH predicting low *A. minutissimum* relative abundances. Alternatively, the first split could have been on ANC, as the improvement index for this variable was only slightly lower than that of pH (Table 5). Twenty-six sites had a pH < 6.1 and a low (0%–28%) predicted relative abundance of *A. minutissimum* (left node). The cross-validation indicates that this node was relatively homogeneous and not further divided. For sites with

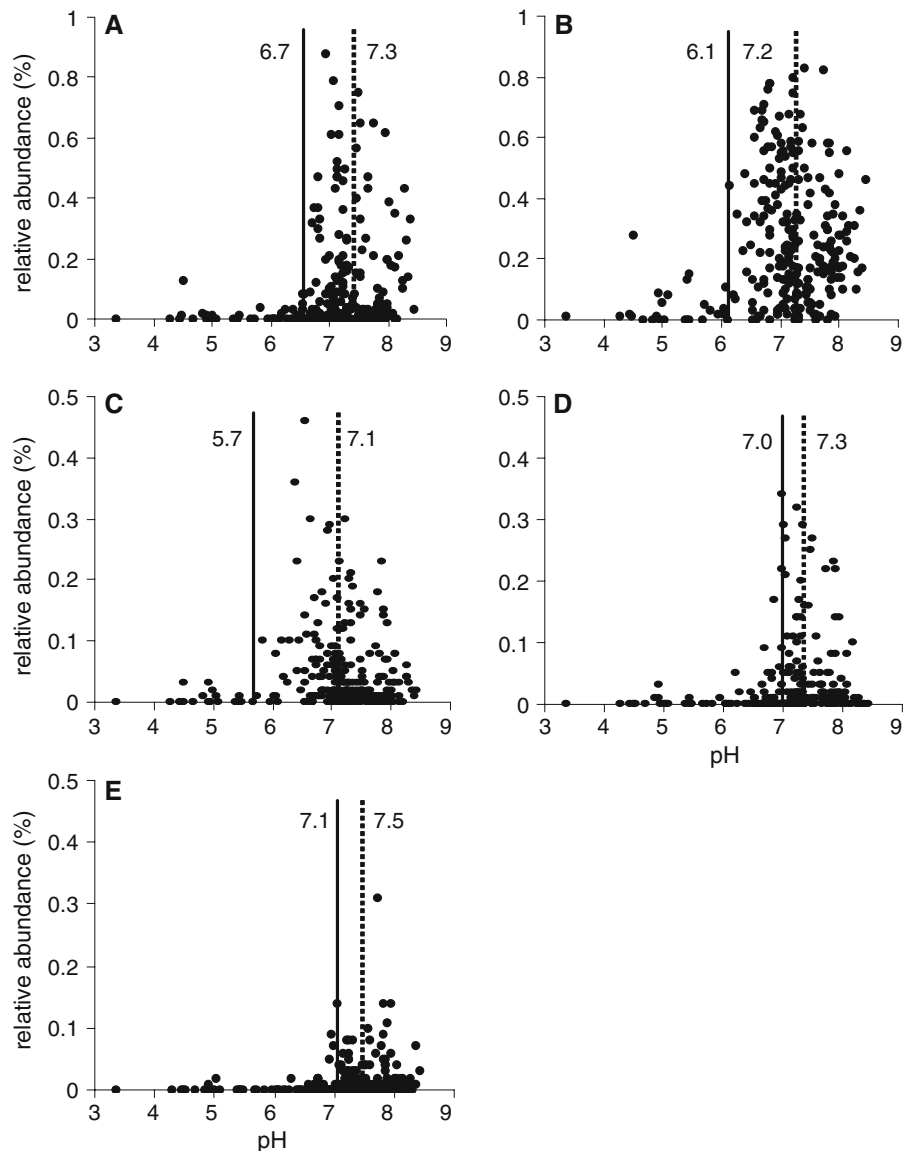


Fig. 2 Relationship between pH and relative abundance for (A) *Achnanthyidium biasolettianum*, (B) *Achnanthyidium minutissimum*, (C) *Fragilaria capucina*, (D) *Planothidium lanceolatum*, and (E) *Reimeria sinuata* at all sites. Solid lines indicate regression tree change points for pH. Dashed lines indicate weighted-average pH optima. Only taxa where pH was an important predictor in the RT and/or CT were included

pH > 6.1 (right node), the second split was on TP. Higher relative abundance was predicted for sites with lower TP concentrations. For the 177 sites with TP < 28 $\mu\text{g l}^{-1}$, highest relative abundance (0%–83%) was predicted. These nodes were not split further.

Overall, the variability in relative abundance of common diatom species within the MAHA region was determined primarily by pH/buffering capacity

and secondarily by nutrient concentration, particularly TP (Table 5). The first split was on pH for the RT of *A. biasolettianum*, *A. minutissimum*, *Fragilaria capucina* Desm., and *Reimeria sinuata* (Greg.) Koc. & Stoerm. and on acid-neutralizing capacity (ANC) for the RT models of *Encyonema minutum* (Hilse ex Rab.) D. Mann, *G. parvulum*, and *N. dissipata* (Fig. 4). In addition, pH and/or ANC were important predictors in the RT of all species, except *palea*

Table 4 Regression tree change points (CP) predicting higher relative abundance for total phosphorus (TP $\mu\text{g l}^{-1}$), weighted-average (WA) TP optima, published WA optima (with study range TP in parenthesis), and van Dam et al. 1994 trophic classification for 10 most frequently occurring diatom taxa in Mid-Atlantic Highlands streams

Taxon	CP	WA Pan et al. 1996 (3D472)	Schoenfelder et al. 2002 (9D1687)	Soininen & Niemela 2002 (2D190)	Potapova et al. 2004 (6D761)	Ponader et al. 2007 (6D731)	Winter & Duthie 2000 (5D215)	Leland & Porter 2000 (40D1200)	van Dam et al. 1994	
AB		14							Meso	
AM	<28	11	17	21	27	33	33	35	Oligo-eutra	
EM		18				32	35			
FC		12				32	35		Meso	
GP	>6	16		38	33	50	52	57	220	Eutra
ND		21			53	35	63	50	110	Meso-eutr
NP		28	62		93	58	70	35	180	Hyper
PL	>18	29	47		86	58	52	32	210	Eutra
RS			21			50	47	57		Meso
SU		9		123	33	59	48	42	220	Oligo-eutra

Meso = Mesotraphentic, Meso-eutr = meso-eutrathentic, Eutra = eutrathentic, Hyper = hypereutrathentic, Oligo-eutra = oligo-eutrathentic

(Kutz.) W. Sm. For all species where pH was an important predictor, lower pH predicted lower abundances. TP concentration was an important predictor of relative abundance for several species, including *A. minutissimum*, *P. lanceolatum*, and *G. parvulum* (Table 5). TP was an alternative predictor for the trees of both *F. capucina* and *N. palea*, with improvement indices very similar to the primary variable chosen for the split (Table 5). Eleven of the nineteen environmental variables were not selected in any of the cross-validated regression trees. These included measures of dissolved cations, total suspended solids, turbidity, total nitrogen, ammonium, and both dissolved organic and inorganic carbon. For most RT, two or three splits were able to predict taxon relative abundance (based on cross-validation r^2 and relative predictive error) as well as trees with more splits. Thus, for most species, pH and TP values are enough to predict relative abundances. High predictive power of RT using all sites was found for *A. minutissimum* ($r^2 = 0.34$) and *P. lanceolatum* ($r^2 = 0.40$; Fig. 4), two taxa with wide ranges in relative abundances. The RT for species with narrower ranges in relative abundances, including *N. palea* and *G. parvulum*, were poor or no successful tree was built (e.g. *S. ulna*).

Classification trees based on presence/absence data with high predictive power and low misclassification rates (range: 22%–30%) were developed for for pH > 6.1. The TP change point was $18 \mu\text{g l}^{-1}$ for

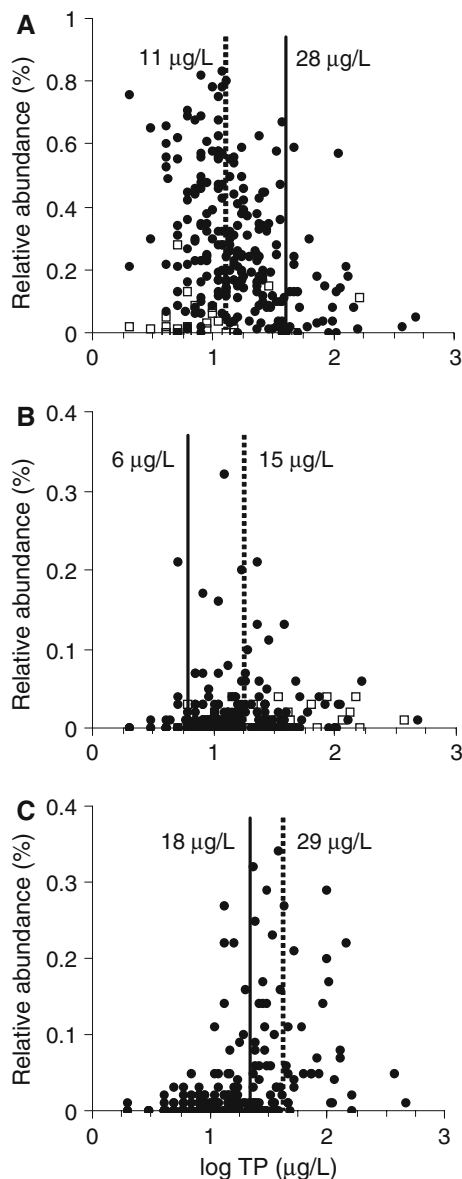


Fig. 3 Relationship between total phosphorus (TP) and relative abundance for **A**) *Achnanthidium minutissimum*, **B**) *Gomphonema parvulum*, and **C**) *Planothidium lanceolatum* at all sites. Solid lines indicate regression tree change points for TP. Dashed lines indicate weighted-average TP optima. Solid circles indicate sites used in change point calculation. Open squares indicate sites not used in change point calculation because they fell out at the first RT break. Only taxa where TP was an important predictor in the RT and/or CT were included

sites with $\text{pH} > 6.1$, with shifts from high relative abundance at sites below this concentration to medium relative abundance at sites with TP greater

than $28 \mu\text{g l}^{-1}$. For *G. parvulum*, the TP change point only occurred at sites with low ANC. For sites with $\text{ANC} < 955 \mu\text{eq/l}$, $\text{TP} > 6 \mu\text{g l}^{-1}$ predicted high relative abundance. Change points for pH were not identified for *E. minutum*, *G. parvulum*, and *N. dissipata*. Change points for TP were not identified for *A. biasolettianum*, *E. minutum*, *F. capucina*, or *N. dissipata*.

Discussion

Regression tree and weighted-averaging approaches provided different, yet complementary, information on the complex relationships between common stream diatoms and environmental variables. While WA provides information on the optimal conditions for a taxon for a single environmental variable, RT both highlights the interactive effects of multiple predictors and can identify breakpoints at which taxa's abundance change from one state to another. In our study, change points identified by regression trees highlighted the interaction between stream acidity (pH and ANC) and TP in shaping the relative abundance of common diatoms in MAHA streams. Stream water pH and ANC were important determinants of diatom species composition in another study of MAHA streams (Pan et al., 1996). Acid mine drainage affects approximately 4% of the streams in the MAHA region (US EPA, 2000) and has been shown to have profound effects on stream periphyton communities (Verb & Vis, 2000; Brake et al., 2004). For most common taxa, RT illustrated that TP is only an important variable under circumneutral to alkaline conditions. The performance of our WA models also reflected the importance of pH in these streams. While the WA pH model performed well ($r^2 = 0.70$), performance of the WA TP model was poor ($r^2 = 0.33$). Nutrients and pH often interact to structure the stream diatom assemblages (e.g., Leland & Porter, 2000; Sojininen & Niemata, 2002). WA nutrient optima models have been less-successful when pH gradients are important (e.g., Hall & Smol, 1992; Reavie et al., 1995; Pan & Stevenson, 1996). Environmental change points identified by RT take into account hierarchical relationships between environmental variables, and therefore, might provide more accurate information on where diatom species abundances shift along environmental gradients.

Fig. 4 Regression tree structure for relative abundance of a) *Achnanthydium biasolettianum*, (b) *Achnanthydium minutissimum*, (c) *Encyonema minutum*, (d) *Gomphonema parvulum*, (e) *Fragilaria capucina*, (f) *Nitzschia dissipata*, (g) *Planothydium lanceolatum*, (h) *Nitzschia palea*, and (i) *Reimeria sinuta*. Terminal nodes give maximum relative abundance for that branch and number of sites in that group (in parenthesis). The values beside each split represent the critical threshold of given variables, which provide the basis for that split. Only taxa where relative predictive power of the RT was >0 were included

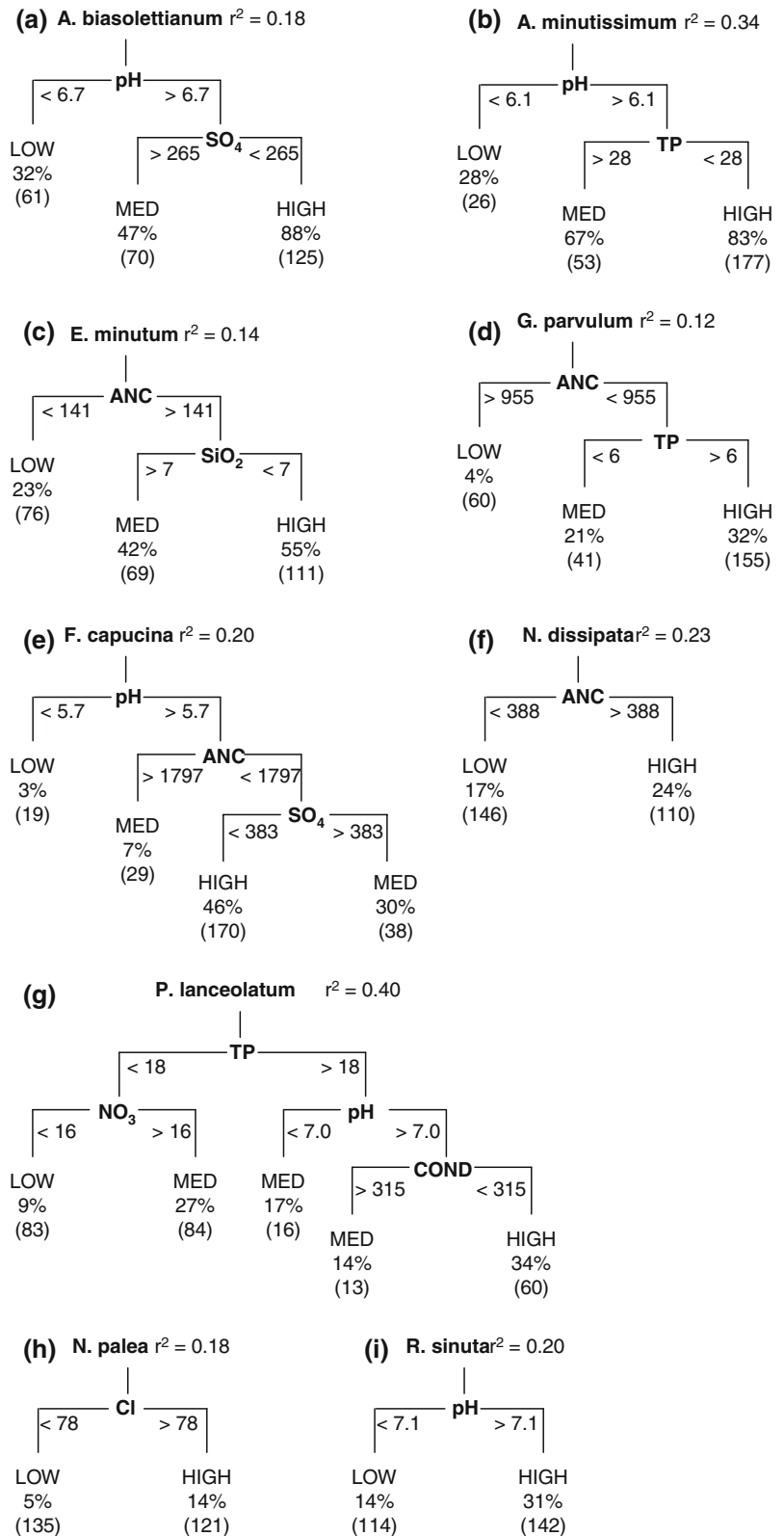


Table 5 Predictor variables for each split and their improvement index (in parenthesis) and alternative split variables and their improvement index (in parenthesis) for regression tree (RT) models

Taxon	Split	RT	Alternative splits	
AB	1st	pH (0.11)	ANC (0.06)	SQ (0.04)
	2nd	SQ (0.09)	COND (0.06)	TURB (0.06)
AM	1st	pH (0.23)	ANC (0.22)	TP (0.08)
	2nd	TP (0.16)	NTL (0.10)	NQ (0.09)
EM	1st	ANC (0.08)	COND (0.06)	pH (0.06)
	2nd	SiO ₂ (0.08)	SQ (0.07)	COND (0.05)
FC	1st	pH (0.08)	ANC (0.06)	TP (0.03)
	2nd	ANC (0.07)	TP (0.05)	SQ (0.04)
	3rd	SQ (0.07)	TP (0.05)	ANC (0.03)
GP	1st	ANC (0.05)	TP (0.04)	COND (0.04)
	2nd	TP (0.09)	NQ (0.07)	TN (0.06)
ND	1st	ANC (0.23)	pH (0.22)	COND (0.19)
NP	1st	CI (0.18)	TP (0.17)	TN (0.15)
PL	1st	TP (0.22)	NQ (0.11)	TN (0.11)
	2nd	NO ₃ (0.19)	TN (0.11)	CI (0.09)
	3rd	pH (0.19)	ANC (0.14)	TP (0.09)
	4th	COND (0.15)	ANC (0.11)	SQ (0.10)
RS	1st	pH (0.24)	ANC (0.19)	COND (0.12)

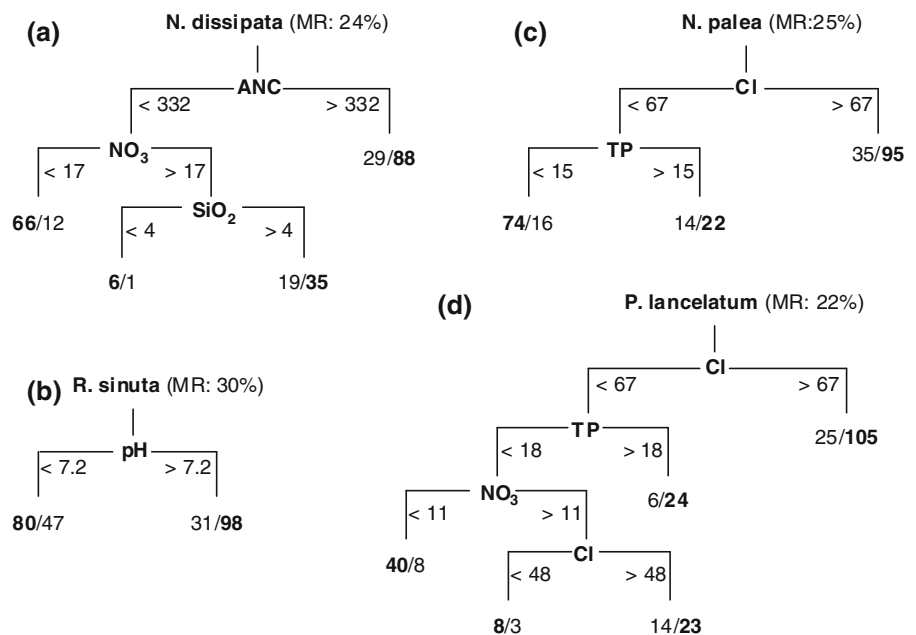
The improvement index is calculated as the number of sites in the branch times the change in impurity index. The number of splits to retain for each species was based on cross-validation results and 1NSE stopping rule

Table 6 Predictor variables for each split and their improvement index (in parenthesis) and alternative split variables and their improvement index (in parenthesis) for classification tree (CT) models. The improvement index is calculated as the number of sites in the branch times the change in impurity index. The number of splits to retain for each species was based on cross-validation results and 1NSE stopping rule

Taxon	Split	CT	Alternative splits	
ND	1st	ANC (21)	pH (20)	TN (20)
	2nd	NO ₃ (13)	TN (11)	SQ (5)
	3rd	SiO ₂ (3)	NO ₃ (3)	SQ (3)
NP	1st	CI (24)	TP (17)	COND (15)
	2nd	TP (10)	SiO ₂ (4)	TURB (4)
PL	1st	CI (15)	NQ (15)	TN (14)
	2nd	TP (9)	TN (6)	NQ (3)
RS	1st	pH (19)	ANC (18)	COND (13)

The autecological information gained through our study augments previous work on diatom species environmental relationships in streams. The WA pH optima for the 10 most commonly occurring taxa were circumneutral to alkaline (7.0–7.5), agreeing with WA optima developed for Swedish and Hungarian streams (Kovács et al., 2006). WA pH optima of *A. biasolettianum*, *G. parvulum*, *N. dissipata*, *P. lanceolatum*, and *S. ulna* agreed with van Dam et al. (1994) autecological classification (Table 1).

Fig. 5 Classification tree structure based on presence/absence for (a) *Nitzschia dissipata*, (b) *Nitzschia palea*, (c) *Reimeria sinuta*, and (d) *Planothidium lanceolatum*. Terminal nodes give number of sites in each class (presence/absence). Predicted class for each branch shown in bold. The values beside each split represent the critical threshold of given variables, which provide the basis for that split. MR = misclassification rate. Only taxa where relative predictive power of the CT was >0 were included



RT change point analysis predicted higher relative abundance at higher pH for all taxa, where pH was an important predictor variable. The change point from low to high abundance identified for all species were lower than their WA optima (Fig. 2). While WA provides an idea of the environmental conditions where abundance is maximized, change point analysis provides information as to where along the environmental gradient the onset of major shifts in abundance occur.

RT change point analysis appeared to characterize species throughout the study region and how well their relationship with TP better than WA approaches. Performance of the WA TP model was weak, indicating that species optima derived from this method might not be accurate. The WA TP optima developed in our study tend to be lower than published optima (Table 4). This might potentially be due to the lower maximum TP concentration in streams in our study compared with many other studies (Table 4). An assumption of WA optima calculations is that the species distribution is unimodal with respect to the environmental variable of interest and the optima will be calculated as the weighted midpoint of this distribution. Consequently, if the maximum TP concentration in a study is high, calculated TP optima will tend to be higher. For both *P. lanceolatum* and *G. parvulum*, change points predicting a shift to high relative abundance were at lower TP concentrations than optima generated by WA in this study. While both of these taxa are considered eutrathentic by van Dam et al. (1994), indicating tolerance to elevated TP, our results suggest that increases in abundance may occur at relatively low TP concentrations ($6 \mu\text{g l}^{-1}$ and $18 \mu\text{g l}^{-1}$ for *G. parvulum* and *P. lanceolatum*, respectively). In contrast, for *A. minutissimum*, the change point indicating a shift to lower relative abundance occurred at a higher TP concentration than our calculated WA optima. This species is characterized as oligotrophic to eutrathentic (van Dam et al., 1994), indicating that it can tolerate a wide range of nutrient conditions. Because change point analysis identifies where species shift relative abundance states, it may provide more refined environmental preferences than WA optima, which is solely based on a mathematical average. Based on our change point analysis, *A. minutissimum* may be considered to tolerate oligotrophic to mesotrophic conditions.

While WA optima approaches are common in the diatom autecological literature, the utility of regression tree approaches in exploring species environmental preferences has yet to be demonstrated.

In this study, RT and CT with high predictive power and low misclassification rates were developed for several of the common diatom species, including *A. minutissimum*, *P. lanceolatum*, *R. sinuta*, *N. dissipata*, and *N. palea*. We feel that the strength of RT and CT analysis may depend on the distribution of individual species throughout the study region and how well their abundance is characterized at any given site. In our study, RT had the highest predictive power for species that were dominant within samples and had a wide range of relative abundances throughout the study sites (e.g., *A. minutissimum*). We feel that RT works well in these instances because the abundance of these species has been well-characterized within the sample and can therefore be accurately modeled. Transformation of relative abundance data into categories and subsequent CT analysis might provide an alternative to RT for less common species whose relative abundance might not be well-characterized by the fixed count methodologies commonly employed in diatom studies. Our CT analyses had misclassification rates comparable to CT of Psh abundance (misclassification rate 22–25%; Hershey et al. 2006) and tree species (misclassification rate 20–71%; Iverson & Prasad 1998). *N. dissipata*, *N. palea*, and *R. sinuata* have low ranges in relative abundance and were found at less than half of the sites in the study. While presence-absence abundance categories seem to provide a successful alternative when developing trees for less-well represented and characterized diatom species, there are a few issues of concern with these types of data transformations. For *N. dissipata*, *P. lanceolatum*, and *R. sinuta*, CT predicted presence better than absence. For example, the CT for *N. dissipata* predicted presence correctly 90% of the time, while only predicted the absence class correctly 60% of the time. This finding contrasts CT for Psh species abundance, where misclassification of Psh absence was only slightly higher (25%) than misclassification of species presence (22%; Hershey et al. 2006). For diatom relative abundances generated through fixed-count methods, zero relative abundance does not necessarily mean that a species is not present at a site, but rather that it was not encountered during the fixed count, and thus an absence in the count does not

necessarily equate to being truly absent from the site. For larger species, such as *P. sh.*, where counts are usually generated by trapping and identifying most of the individuals within the stream reach, presence-absence data probably reflects true presence-absence.

In conclusion, RT analysis illustrated the importance of the interaction between pH and TP in shaping the abundance of common diatom species in the MAHA region and complemented traditional weighted-averaging approaches to modeling species-environmental relationships. Change points identified by RT analysis provided more refined information on where relative abundance of common diatom species shifted along pH and TP gradients. Several authors have put forth hierarchical models of variables structuring stream algal distribution and abundance (e.g., Biggs, 1996; Stevenson, 1997). While our study, and most other diatom autecological studies, focused only on water quality variables, we feel that regression tree approaches have the potential to increase our understanding of how interactions among environmental variables shape stream diatom assemblages. In addition, we feel that tree approaches could be used in conjunction with other modeling approaches (e.g., Generalized additive models) to provide insight into interaction terms and also as a framework for developing and interpreting WA optima.

Acknowledgments We thank Brian Hill, Jan Stevenson, Alan Herlihy and the EMAP field crews. This manuscript was improved by the comments of two anonymous referees.

References

- Atkinson, E. J. & T. M. Therneau, 2000. An Introduction to Recursive Partitioning Using the RPART Routines. Mayo Foundation.
- Beaver, J., 1981. Apparent ecological characteristics of some common freshwater diatoms. Ontario Ministry of the Environment, Technical Support Sections, Central Region, Don Mills.
- Bennion, H., 1994. A diatom-phosphorus transfer function for shallow, eutrophic ponds in southeast England. *Hydrobiologia* 275/276: 391–410.
- Biggs, B. F. J., 1990. Periphyton communities and their environments in New Zealand rivers. *New Zealand Journal of Marine and Freshwater Research* 24: 367–386.
- Biggs, B. F. J., 1996. Patterns of benthic algae in streams. In Stevenson, R. J., M. L. Bothwell & R. L. Lowe (eds), *Algal Ecology Freshwater Benthic Ecosystems*. Academic Press, San Diego, CA.
- Birks, H. J. B., J. M. Line, S. Juggins & A. C. Stevenson, 1990. Diatoms and pH reconstruction. *Philosophical Transactions of the Royal Society of London Series B Biological Science* 327: 263–278.
- Brake, S. S., S. T. Hasiotis & H. K. Dannelly, 2004. Diatoms in acid mine drainage and their role in the formation of iron-rich stromatolites. *Geomicrobiology Journal* 21: 331–340.
- Brieman, L., J. H. Friedman, R. A. Olshen & C. J. Stone, 1984. *Classification and Regression Trees*. Wadsworth and Brooks/Cole, Pacific Grove, California.
- Bunting, M. J., H. C. Duthie, D. R. Campbell, B. G. Warner & L. J. Turner, 1997. A paleoecological record of recent environmental changes at Big Creek Marsh, Long Point, Lake Erie. *Journal of Great Lakes Research* 23: 349–368.
- Clark, L. A. & D. Pregibon, 1993. Tree-based models. In Chambers, S. J. M. & T. J. Hastie (eds), *Statistical Models*. Chapman and Hall, London, UK.
- Cooper, S. R., 1999. Estuarine paleoenvironmental reconstructions using diatoms. In Stoermer, E. F. & J. P. Smol (eds), *The Diatoms: Applications for the Environmental and Earth Sciences*. Cambridge University Press, Cambridge, UK.
- DeOath, G., 2002. Multivariate regression trees: a new technique for modeling species-environmental relationships. *Ecology* 83: 1105–1117.
- DeOath, G. & K. W. Fabricius, 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81: 3178–3192.
- Dixit, A. S., S. S. Dixit & J. P. Smol, 1992. Long-term trends in lake water pH and metal concentrations inferred from diatoms and chrysophytes in three lakes near Sudbury, Ontario. *Canadian Journal of Fisheries and Aquatic Sciences* 49: 17–24.
- Dixit, S. S., J. P. Smol, D. F. Charles, R. M. Hughes, S. G. Paulsen & G. B. Collins, 1999. Assessing water quality changes in the lakes of the northeastern United States using sediment diatoms. *Canadian Journal of Fisheries and Aquatic Sciences* 56: 131–152.
- Dodds, W. K., J. R. Jones & E. B. Welch, 1998. Suggested criteria for stream trophic state: distributions of temperate streams types by chlorophyll, total nitrogen and phosphorus. *Water Research* 32: 1455–1462.
- Franklin, J., 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *Journal of Vegetation Science* 9: 733–748.
- Gaiser, E. E. & B. E. Taylor, 1995. Paleocological reconstruction of Holocene environments in wetland ponds of the Upper Atlantic Coastal Plain. *Bulletin of the Ecological Society of America* 76: 89.
- Guisan, A. & N. E. Zimmerman, 2000. Predictive habitat distribution models in ecology. *Ecological Modeling* 135: 147–186.
- Guisan, A., T. Edwards Jr. & T. Hastie, 2002. General linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modeling* 157: 89–100.
- Hall, R. I. & J. P. Smol, 1992. A weighted-averaging regression and calibration model for inferring total phosphorus concentration from diatoms in British Columbia (Canada) lakes. *Freshwater Biology* 27: 417–434.

- Herlihy, A. T., J. L. Stoddard & C. B. Johnson, 1998. The relationship between stream chemistry and watershed land cover data in the Mid-Atlantic region. *U.S. Water Air and Soil Pollution* 105: 379–386.
- Hershey, A. E., S. Beaty, K. Fortino, M. Keyse, P. P. Mou, W. J. O'Brien, A. J. Ulseth, G. A. Gettel, P. W. Lienesch, C. Lueke, M. E. McDonald, C. Mayer, M. Miller, C. Richards, J. A. Schuldt & S. C. Whalen, 2006. Effect of landscape factors on fish distribution in arctic Alaska lakes. *Freshwater Biology* 51: 39–55.
- Hill, M. O., 1973. Diversity and evenness: a unifying notation and its consequences. *Ecology* 54: 427–432.
- Imbrie, J. & T. Webb III, 1981. Transfer functions: calibrating micropaleontological data in climatic terms. In Berger, A. (ed.), *Climatic Variations and Variability: Facts and Theories*. Springer, Dordrecht, Netherlands.
- Iverson, L. R. & A. M. Prasad, 1998. Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecological Monographs* 68: 465–485.
- Juggins, S., 1992. Diatoms in the Thames Estuary, England: ecology, paleoecology, and salinity transfer function. *Bibliotheca Diatomologica* 25: 1–216.
- Juggins, S., 2003. *User Guide. Software for Ecological and Palaeoecological Data Analysis and Visualization*. University of Newcastle, Newcastle upon Tyne, UK.
- Kovács, C., M. Kahlert & J. Padisák, 2006. Benthic diatom communities along pH and TP gradients in Hungarian and Swedish streams. *Journal of Applied Phycology* 18: 105–117.
- Krammer, K. & H. Lange-Bertalot, 1986. *Bacillariophyceae, Teil 1. Naviculaceae*. Spektrum Akademischer Verlag, Heidelberg, Germany.
- Krammer, K. & H. Lange-Bertalot, 1988. *Bacillariophyceae, Teil 2. Epithemiaceae, Bacillariophyceae, Surirellaceae*. Spektrum Akademischer Verlag, Heidelberg, Germany.
- Krammer, K. & H. Lange-Bertalot, 1991a. *Bacillariophyceae, Teil 3. Centrales, Fragilariaceae, Eunotiaceae, Achnantheaceae*. Spektrum Akademischer Verlag, Heidelberg, Germany.
- Krammer, K. & H. Lange-Bertalot, 1991b. *Bacillariophyceae, Teil 4. Achnantheaceae, Kritische Ergänzungen zu Navicula (Lineolate) und Gomphonema*. Spektrum Akademischer Verlag, Heidelberg, Germany.
- Lange-Bertalot, H., 1979. Pollution tolerance of diatoms as a criterion for water quality estimation. *Nova Hedwigia* 64: 285–304.
- Leland, H. V., 1995. Distribution of phytobenthos in the Yakima River basin, Washington, in relation to geology, land use, and other environmental factors. *Canadian Journal of Fisheries and Aquatic Sciences* 52: 1108–1129.
- Leland, H. V. & S. D. Porter, 2000. Distribution of benthic algae in the upper Illinois River basin in relation to geology and land use. *Freshwater Biology* 44: 279–301.
- Leland, H. V., L. R. Brown & D. K. Mueller, 2001. Distribution of algae in the San Joaquin River, California, in relation to nutrient supply, salinity, and other environmental factors. *Freshwater Biology* 46: 1139–1167.
- Lougheed, V. L., C. A. Parker & R. J. Stevenson, 2007. Using non-linear responses of multiple taxonomic groups to establish criteria indicative of wetland biological condition. *Wetlands* 27: 96–109.
- Lowe, R. L., 1974. Environmental requirements and pollution tolerance of freshwater diatoms. National Environmental Research Center, Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati.
- Munn, M. D., R. W. Black & S. J. Gruber, 2002. Response of benthic algae to environmental gradients in an agriculturally dominated landscape. *Journal of the North American Benthological Society* 21: 221–237.
- O'Connor, R. J. & T. L. Wagner, 2004. A test of a regression tree model of species distribution. *Auk* 121: 604–609.
- Omernik, J. M., 1987. Ecoregions of the conterminous United States. Map (scale 1:7, 500, 000). *Annals of the Association of American Geographers* 77: 118–125.
- Pan, Y. & R. J. Stevenson, 1996. Gradient analysis of diatom communities in western Kentucky wetlands. *Journal of Phycology* 32: 201–212.
- Pan, Y., R. J. Stevenson, B. Hill, A. Herlihy & G. 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.
- Pan, Y., R. J. Stevenson, B. H. Hill, P. R. Kaufmann & A. T. Herlihy, 1999. Spatial patterns and ecological determinants of benthic algal assemblages in mid-Atlantic streams, USA. *Journal of Phycology* 35: 460–468.
- Patrick, R. & C. W. Reimer, 1966. *The Diatoms of the United States, Vol. 1. Monographs of the Academy of Natural Sciences of Philadelphia No. 13*. Philadelphia, PA, USA.
- Patrick, R. & C. W. Reimer, 1975. *The Diatoms of the United States, Vol. 2. Monographs of the Academy of Natural Sciences of Philadelphia No. 13*. Philadelphia, PA, USA.
- Ponader, K. C., D. F. Charles & T. J. Belton, 2007. Diatom-based TP and TN inference models and indices for monitoring nutrient enrichment of New Jersey Streams. *Ecological Indicators* 7: 79–93.
- Potapova, M. G. & D. F. Charles, 2003. Distribution of benthic diatoms in U.S. rivers in relation to conductivity and ionic composition. *Freshwater Biology* 48: 1311–1328.
- Potapova, M. G., D. F. Charles, K. C. Ponader & D. M. Winter, 2004. Quantifying species indicator values for trophic diatom indices: a comparison of approaches. *Hydrobiologia* 517: 25–41.
- Qian, S. S., R. S. King & C. J. Richardson, 2003. Two statistical methods for the detection of environmental thresholds. *Ecological Modelling* 166: 87–97.
- Quinlan, J. R., 1986. Induction of decision trees. *Machine Learning* 1: 81–106.
- Reavie, E. D., R. I. Hall & J. P. Smol, 1995. An expanded weighted-averaging model for inferring past total phosphorus concentrations in eutrophic British Columbia (Canada) lakes. *Journal of Paleolimnology* 14: 49–67.
- Schoofelder, I., J. Gelbrecht, J. Schoofelder & C. E. W. Steinberg, 2002. Relationships between littoral diatoms and their chemical environment in northeastern German lakes and rivers. *Journal of Phycology* 38: 66–82.
- Soininen, J. & P. Niemelä, 2002. Inferring the phosphorus levels of rivers from benthic diatoms using weighted averaging. *Archiv für Hydrobiologia* 154: 1–18.
- Stevenson, R. J., 1997. Scale-dependent determinants and consequences of benthic algal heterogeneity. *Journal of the North American Benthological Society* 16: 248–262.
- ter Braak, C. F. J. & H. van Dam, 1989. Inferring pH from diatoms: a comparison of old and new calibration methods. *Hydrobiologia* 178: 209–223.

- Therneau, T. M. & E. J. Atkinson, 1997. An Introduction to Recursive Partitioning using the RPart Routines. Mayo Foundation.
- U.S. EPA. 1987. Handbook of Methods for Acid Deposition Studies, Laboratory Analyses for Surface Water Chemistry. EPA 600/4-87/026. U.S. Environmental Protection Agency, Office of Research and Development. Washington, D.C.
- U.S. EPA. 2000. Mid-Atlantic Highlands Streams Assessment. EPA/903/R-00/015. U.S. Environmental Protection Agency, Region 3. Philadelphia, PA.
- van Dam, H., A. Mertens & J. Sinkeldam, 1994. A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands. *Netherlands Journal of Aquatic Ecology* 28: 117–133.
- Vayssières, M. P., R. E. Plant & B. H. Allen-Diaz, 2000. Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science* 11: 679–694.
- Venables, W. N. & B. D. Ripley, 2002. *Modern and Applied Statistics with S-Plus*. Springer-Verlag, New York.
- Verb, R. G. & M. L. Vis, 2000. Comparison of benthic diatom assemblages from streams draining abandoned and reclaimed coal mines and nonimpacted sites. *Journal of the North American Benthological Society* 19: 274–288.
- Winter, J. G. & H. C. Duthie, 2000. Epithic diatoms as indicators of stream total N and total P concentration. *Journal of the North American Benthological Society* 19: 32–49.