Freshwater Biology

Freshwater Biology (2015) **60,** 267–281 doi:10.1111/fwb.12478

Are the landscape-level drivers of water column and surface sediment diatoms different?

A. K. WINEGARDNER*,†, B. E. BEISNER†,‡, P. LEGENDRE†,§ AND I. GREGORY-EAVES*,†

*Department of Biology, McGill University, Montréal, QC, Canada

†Québec Centre for Biodiversity Science, Montréal, QC, Canada

‡Department of Biological Sciences, University of Québec at Montréal, Montréal, QC, Canada

§Département de Sciences Biologiques, Université de Montréal, Montréal, QC, Canada

SUMMARY

1. Threats to biodiversity are fostering new collaboration between aquatic ecologists and palaeolimnologists, who have traditionally asked ecological questions on different time scales. While the differences between surface sediment and water column or snapshot sampling are well understood, less so are the consequences of comparing the predominant drivers of aquatic assemblages resulting from these two types of sampling.

2. Using diatom data from the 2007 USEPA National Lakes Assessment (NLA) program (468 lakes), we compared the main environmental and spatial drivers of diatom community composition between samples derived from the water column and surface sediments. We hypothesised that, in explaining community variation across the conterminous United States, the effect of environment would be stronger in diatom assemblages preserved in surface sediments because of the inclusion of benthic members and temporal integration. We used a combination of ordination overlays and variation partitioning to examine differences in community drivers between palaeolimnological (surface sediment) and water column sampling.

3. We found that these two types of sampling were significantly correlated with respect to the drivers of community composition in addition to having congruent patterns of ordination. Congruency between sampling methods further increased when the water column data were temporally integrated and may be explained by variation in seasonally dynamic taxa.

4. To our knowledge, this is the first study that has tested for differences in environmental structuring patterns between palaeolimnological and water column samples using such a highly replicated and landscape-level approach. On the basis of our results, we encourage ecologists to consider the joint analysis of these two types of data sets where data are available.

Keywords: community, lakes, large scale ecology, phytoplankton, plankton

Introduction

Globally, freshwater systems provide habitat for *c.* 10% of known species, despite covering <1% of the Earth's surface (Strayer & Dudgeon, 2010). To answer questions about how environmental change affects aquatic communities, scientists have adopted two key types of sampling methods: sampling of the water column (live organisms) and sampling lake sediment records (biotic indicators; palaeolimnology). Studies focusing on a diverse array of ecological questions make up the history of palaeolimnology, but recent attention has focused on historical reconstructions and the use of transfer functions to infer past conditions from subfossil assemblages (Smol, 2008; Birks *et al.*, 2012). Nonetheless, there has also been increasing interest in the long-term ecological perspective that palaeolimnology studies can offer (Flessa & Jackson, 2005; Heino, 2009).

Correspondence: A. K. Winegardner, Department of Biology, McGill University, 1205 Avenue Docteur Penfield, Montréal, QC, Canada H3A 1B1. E-mail: amanda.winegardner@mail.mcgill.ca

Gregory-Eaves & Beisner (2011) explicitly champion this approach and discuss the great potential of palaeolimnology to contribute to studies of aquatic biodiversity.

As a result of growing interest in using palaeolimnological data for new types of questions (Seddon *et al.*, 2014) and the movement towards coupling water column and palaeolimnological data sets (Battarbee *et al.*, 2005), there is a critical need to understand when they can effectively be used together (i.e. in joint analyses) and how conclusions drawn from these two data sources may differ (i.e. comparability). A recent example of this kind of study was reported by Levi *et al.* (2014) who quantified the macrophyte community of 35 Mediterranean lakes by surveying both the present-day vegetation and their remains in surface sediments. Their work moved beyond 'do we find the same species in present-day and sediment samples?' and asked instead 'what variables explain variation in the vegetative communities across these lakes?' and 'are the same dominant drivers of variation identified using data from both types of sampling?' This last question remains unanswered for diatom communities (as are other questions related to this type of comparison) and is especially important because while joint structuring of communities by both environmental and spatial variables is found in many water column studies (e.g. Cottenie, 2005), palaeolimnological studies often show stronger support for environmental structuring (e.g. Verleyen *et al.*, 2009).

Clearly, there are fundamental differences between water column samples and surface sediments, with the latter integrating habitats from across an entire lake and through time, and usually representing multiple years of sediment accumulation (Brothers, Vermaire & Gregory-Eaves, 2008). Some studies have used subfossil assemblages from surface sediment samples to ask about the relative contributions of environmental and spatial factors to community composition, though far fewer than with traditional water column data. For example, Verleyen *et al.* (2009) used diatom surface sediment calibration sets to show that factors related to local environment explained a median of 21% of diatom variation, whereas spatial variables explained a median of only 5.5%, while pure space (without any influence of environment) did not explain any significant variation in diatom communities. Bennett *et al.* (2010) provide another example of a diatom study conducted across a very large spatial scale, reporting that variables related to dispersal limitation were important at an intercontinental scale, but that pH exhibited an omnibus effect at regional spatial scales. More studies of this nature are needed to better understand drivers of community composition as preserved in sediment samples and to provide data for larger syntheses.

The main focus of our study was to compare the dominant drivers of diatom assemblages delineated from palaeolimnological (surface sediment) samples and two types of water column sampling methods: single-visit samples and temporally averaged values over repeated summer samplings. For this work, we relied on the USEPA National Lake Assessment from 2007 (USEPA, 2009) because it represented both a large sample size and included both water column and surface sediment sampling for many of the lakes. Our specific questions were:

1. How different are diatom assemblage compositions between surface sediment and single-visit water-column samples, as well as between surface sediment and temporally averaged water-column samples?
2. What are the dominant drivers of diatom variation across these lakes? Do conclusions about these drivers depend on whether surface sediment or water column data are used?
3. What are the implications of including only planktonic taxa when comparing the ecological conclusions drawn from these different sampling methods?

We hypothesised that because lake surface sediment samples have a degree of temporal integration, thus providing a longer interval for the immigration, emigration and colonisation of taxa than do water-column samples, assemblages within surface sediment would be more strongly explained by environmental gradients. However, we expected that when seasonally averaged water column samples are used, ecological patterns would more closely match those of surface sediments. We also hypothesised that excluding benthic species from sediment samples would increase the congruence between surface sediment and water column samples because only planktonic species are being compared between the two types of sampling. Surface sediment samples also integrate spatially across several zones in a lake (Smol, 2008) so that habitat type (benthic or planktonic species) may be an important structuring factor of diatom assemblages that becomes apparent when comparing surface sediment and water column samples. However, the removal of benthic species from surface sediment samples may decrease the amount of variation explained by environment relative to spatial drivers because benthic species are known to match closely to environmental conditions (Philibert & Prairie, 2002). To our knowledge, this is the most exhaustive analysis that addresses these questions with freshwater diatoms and with a common set of lakes for the two types of sampling methods.

Methods

Description of National Lakes Assessment data set

The NLA programme, administered by the United States Environment Protection Agency and partnered with state environment and resource agencies, is part of the National Aquatic Resource Surveys program and involves intensive sampling of lakes and reservoirs of the conterminous U.S.A. (the lower 48 states) every 5 years. A full description of the programme is available in the 'NLA Field Operations Manual' from both 2007 and 2012 (http://water.epa.gov/type/lakes/lakessurvey_index.cfm) and is further summarised in Beaulieu, Pick & Gregory-Eaves (2013). Metadata and raw data from the 2007 campaign is available from the USEPA: http://water.epa.gov/type/lakes/NLA_data.cfm. From the *c.* 1000 lakes sampled, we retained 468 sites that had diatom data for both water column and surface sediment samples.

Field teams from the NLA collected diatoms from the water column using an integrated water sampler (a PVC tube with a length of 2 m and diameter of 3.2 cm) over the entire depth of the euphotic zone (≥2 m) at the deepest point of each lake. The sampler was deployed twice, and the samples mixed together. One litre from the pooled sample was preserved in Lugol's solution for later enumeration (USEPA, 2011–2012). Because of the length of the integrated water sampler, 'water column' samples for the purpose of this study refer to two integrated samples of the top 2 m of lake water. Surface sediment was collected using a modified Kajak-Brinkhurst corer (Glew, 1989), again at the deepest point of the lake. Where possible, a 45 cm sediment core was collected and the top 1 cm section saved for diatom enumeration (minus a 1 cm^3 subsection from the centre of the sediment slice). Up to 500 diatom valves were enumerated using standardised methods from the U.S. Geological Survey National Water Quality Assessment (Charles, Knowles & Davis, 2003; USEPA, 2011–2012). Quality Control procedures involved re-identification of a random 10% subset of each sample by a second taxonomist to minimise differences in enumeration and taxonomic disagreement (USEPA, 2011–2012). An explanation of water-quality data collection is found in the 2007 field manual (USEPA, 2007).

Sediment cores were collected only once during the sampling period (May to mid-October 2007). Most water samples were collected at the same time as the sediment cores, but for a smaller subset of the 468 lakes, water column samples were collected (and enumerated for diatoms) both early in the sampling period (May–June; during the sediment coring) and later in the sampling period (August to October). We used these revisited sites for comparisons of surface sediment samples to water column samples, whereby averaging of counts was performed post-enumeration.

Data management and pre-processing of NLA data

Some pre-processing of the open access NLA data was required before statistical analyses could be conducted (Table S1). Diatoms were aggregated to species and genus levels, and we performed analyses using both of these resolutions. We removed species that did not reach at least a 5% relative abundance in a minimum of a single sample from the data set as a whole so that abundances were not influenced by rare species. We then transformed diatom species abundances to relativized values using the Hellinger transformation; this transformation is the square root of the relative abundance values per sample (Legendre & Gallagher, 2001). This transformation made the community composition data suitable for beta diversity study (Legendre & De Cáceres, 2013). After screening the environmental data of variables that were strongly collinear, the environment variables considered in our analyses were the following: pH (from the field), conductivity ($\mu S\ cm^{-1}$), turbidity (NTU), dissolved organic carbon [DOC ($mg\ L^{-1}$)], ammonium [NH_4 ($\mu eq\ L^{-1}$)], nitrate + nitrite by flow injection analysis (NO_3/NO_2 [$mg\ N\ L^{-1}$)], total nitrogen [TN ($\mu g\ L^{-1}$)], total phosphorus [TP ($\mu g\ L^{-1}$)], chloride [Cl^- ($\mu eq\ L^{-1}$)], sulphate [SO_4 ($\mu eq\ L^{-1}$)], calcium [Ca^+ ($\mu eq\ L^{-1}$)], magnesium [Mg^{+2}($\mu eq\ L^{-1}$)], colour (PCU), silica [SiO_2 ($mg\ L^{-1}$)], hydrogen ions (from pH measured in the lab; $\mu eq\ L^{-1}$), hydroxide (from pH measured in the lab; $\mu eq\ L^{-1}$), ion balance using acid neutralising capacity [ANC (%)], chlorophyll *a* concentration ($\mu g\ L^{-1}$) and mean Secchi depth (m) (Table 1). Bathymetric maps were not available for all sampled lakes, and so maximum depth (Z_{max}) was found using a depth finder. This approximate Z_{max} in metres was also included as a variable.

Statistical analyses

Broadly, we were interested in comparing diatom assemblages between sampling types and identifying the dominant drivers of diatom assemblage variation across the set of lakes and with each sampling type (see Fig. 1 for an overview of the statistical analyses with corresponding hypotheses). As such, we used redundancy

Table 1 Mean, median, range and standard deviation of (non-transformed) environmental variables measured from the integrated water column sample (n = 468)

Variable	Mean	Median	Range	Standard deviation
pH	8.1	8.2	4.7–10.3	0.8
Conductivity (μS cm^{-1})	470	255.4	12.9–9751	961.8
Turbidity (NTU)	13.2	3.4	0.3–312	31.7
DOC (mg L^{-1})	9.2	5.3	0.3–290.6	20.9
NH_4 (μeq L^{-1})	2.9	1.3	0.3–122.0	8.3
NO_3 + NO_2 (mg N L^{-1})	0.09	0.005	0–5.6	0.4
Total Nitrogen (μg L^{-1})	1185.1	543.5	70–26100	2481.1
Total Phosphorus (μg L^{-1})	107.9	24.5	1–2147	259.2
Cl (μeq L^{-1})	740.2	219.7	1.5–22890.4	1977.7
SO_4 (μeq L^{-1})	1840	203.5	2.5–133210.8	8414.9
Ca (μeq L^{-1})	1341.7	1201.0	61.0–17095.7	1357.3
Mg (μeq L^{-1})	1425.1	554.1	16.1–60703.9	3842.2
Colour (PCU)	16.6	11.0	0–93	15.5
SiO_2 (mg L^{-1} SiO_2)	8.6	5.4	0.03–91.9	10.6
H^+ (μeq L^{-1})	0.07	0.006	0–15.1	0.7
OH^- (μeq L^{-1})	3.2	1.7	0.001–123.0	8.2
Ion balance (ANC %)	−0.9	−1.2	−13.7 to 20.3	2.8
Chl *a* (μg L^{-1})	27.9	7.4	0.1–871.2	73.9
Secchi depth (m)	2.06	1.6	0.05–12.5	1.9
Z_{max} (m)	9.9	6.5	0.5–60.3	10.3

All variables included in selection procedures are included here.

analysis (RDA) to identify relationships between the diatom and predictor (i.e. environmental or spatial matrices) data sets and used co-inertia analysis to quantify the degree of common structure between the water column and surface sediment data sets (Legendre & Legendre, 2012).

Diatom species in the water column and surface sediment data sets. Diatom species from both the water column and surface sediment samples were classified as planktonic, benthic and tychoplanktonic using sources from both the primary literature and online databases (Table S2). Tychoplanktonic refers to species that are generally benthic, but that will also live in planktonic form if conditions allow (Wehr & Sheath, 2003). Generally, species found in the water column samples were only planktonic or tychoplanktonic, but species from the surface sediment samples are also often in the benthic class. We created two sets of diatom data: one with planktonic, benthic and tychoplanktonic species and the other with purely planktonic species (resulting in a total species richness reduction of 74%, i.e. 26% of the species remained).

To identify the main axes of variation across the species-by-site matrix, we performed principal components analysis (PCA) for the 468 lakes, for each of the following data sets: (i) the water column (all species) diatoms; (ii) the surface sediment (all species) diatoms; (iii) the planktonic-only water column diatoms; and (iv) planktonic-only surface sediment diatoms. To quantify correlation between the assemblage data, we then computed an RV coefficient between the first PCA axis of the water column diatom matrix and the surface sediment diatom matrix considering all species (n = 468), as well as between these respective matrices with only the planktonic species. For completeness, we also computed the RV coefficient for the full assemblage (see Picazo, Millán & Dolédec, 2012 for a similar approach). The RV coefficient is a multivariate generalisation of the Pearson correlation that correlates two matrices with corresponding rows (sites). It produces values between 0 (no correlation) and 1 (perfect correlation). The RV coefficient between two vectors of quantitative data is the square of the Pearson correlation; between two matrices, it is thus homologous to an R^2 (Legendre & Legendre, 2012).

Environmental and spatial drivers of diatom variation. We were interested in both identifying the most parsimonious set of environmental and spatial variables that explained the greatest variation in each of the diatom data sets and comparing these results between water column and the surface sediment assemblages. Given that the NLA data set included numerous environmental variables, we applied forward selection to the suite of potential predictors, after screening for collinearity. To test for the potential influence of variables related to dis-

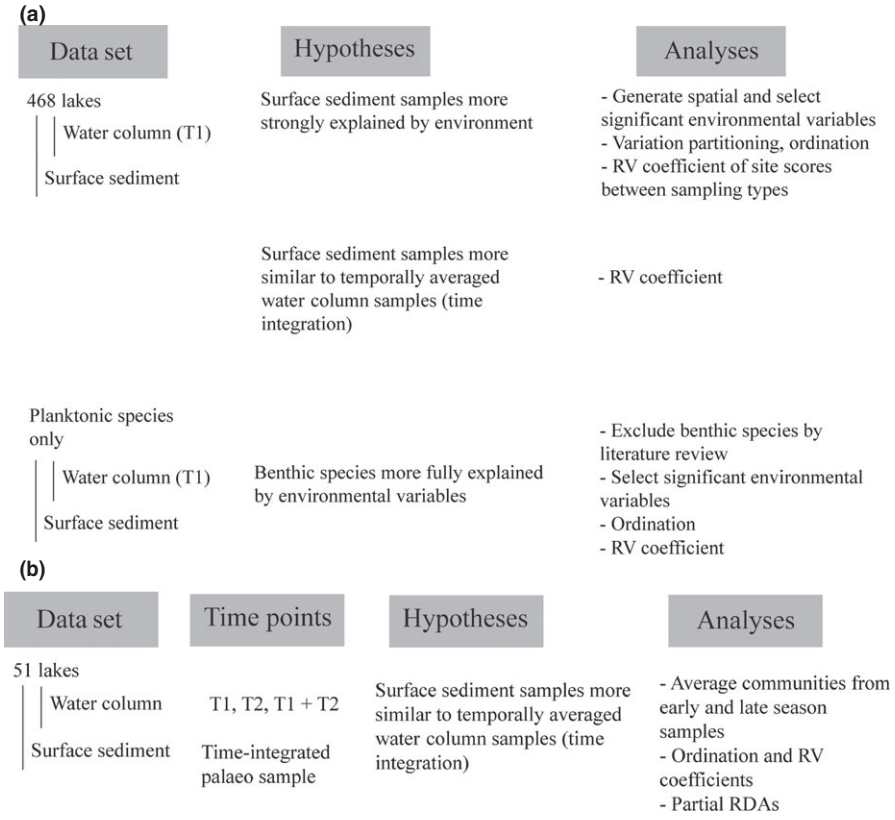

Fig. 1 Visual overview of statistical analyses and associated hypotheses. Overview of data set-up, hypotheses and associated statistical analyses for (a) analyses encompassing the entire lake data set (n = 468) and (b) a subset of 51 lakes with the inclusion of a second water column sampling visit. For both (a) and (b), samples were paired so that each lake was represented by a water column sample derived from an integrated water sample and a surface sediment sample derived from the top 1 cm of a sediment core. 'T1' and 'T2' refer to sampling time points included in each analysis, with 'T1' occurring in June or early July 2007 and 'T2' in August 2007. All analyses were conducted with both species- and genus-level community data.

persal, we generated spatial variables for this data set using the site coordinates and then selected significant variables using both the water column and surface sediment diatoms (see Software). The cut-off for variable retention within the context of the forward selection process was the adjusted R^2 of the model containing all variables; its value was 0.082 for the water column environmental variables, 0.055 for the water column spatial variables and 0.090 for the surface sediment spatial variables. Using these reduced sets of environmental and spatial variables, we independently partitioned the variation in water column or surface sediment assemblages into fractions that were uniquely explained by space (spatial variables) and environment, including shared fractions.

Given that diatom assemblages can be highly seasonal, we also wanted to know whether the important environmental variables driving diatom assemblage variation would change depending on when the lakes were sampled during the growing season. To examine this, we extracted a subset of 51 lakes having surface sediment samples (from a single core sample), as well as water column diatoms enumerated from a first sampling visit and a second set of water column data from a second sampling visit (see Table S1). For this section, we focused only on the environmental variables. We again ran forward selection on the full set of environmental variables, this time for these 51 sites, using the first visit (visit 1) set of water column diatom community data as well as the second visit (visit 2). We again used a cut-off criterion for forward selection that reflected the adjusted R^2 of the model containing all the variables, which was 0.084 for visit 1 and 0.071 for visit 2.

To quantify relationships between the water column diatom assemblages and our reduced set of environmental variables, we performed a RDA with the (Hellinger-transformed) water column diatom data and forward-selected environmental variables. To examine whether surface sediment diatoms showed similar patterns within the RDA and to examine relationships between the environment and surface sediment diatoms, we then performed a RDA with the Hellinger-transformed surface sediment diatom data and the same environmental variables. This reflects the approach taken in many palaeoecological studies using surface sediments, that is, to quantify subfossil organisms from surface sediments, while measuring environmental variables from the water column (e.g. Kurek, Weeber & Smol, 2011). This RDA approach was performed for all species (n = 468), as well as for the planktonic-only data

sets (n = 468), and for the visit 1 and visit 2 data sets, as well as an average of both visits (averaged post-enumeration) (n = 51). We then repeated these RDAs for the reduced set of spatial variables, again to identify relationships between spatial variables and diatom assemblages (for the n = 468 data sets only).

Correlation between environmental and spatial ordinations. After performing the RDAs, we quantified the resemblance of the water column site scores and surface sediment site scores from the various RDAs. To do this, we extracted the water column site scores, as well as the surface sediment site scores from each RDA. We then computed an RV coefficient for the water column versus surface sediment site scores for: (i) the environmental RDA including all species; (ii) the spatial RDA including all species; (iii) the environmental RDA including only planktonic species, (iv) the spatial RDA including only planktonic species; (v) the visit 1 water column (environmental) RDA including all species; (vi) the visit 2 water column (environmental) RDA including all species; and (vii) the mean of visit 1 and 2 water column (environmental) RDA including all species.

Comparing sampling types. We used partial RDA to find out if there were significant differences in diatom assemblage composition between the water column and the surface sediments. This form of analysis, which is the multivariate equivalent of a paired *t*-test, used the sampling methods as the explanatory variable (water column or surface sediment) and the lake identifiers as covariables. Contrary to co-inertia analysis, the two data sets to be compared were placed one on top of the other, matching the species columns, while keeping the lakes in the same order in the two parts of the combined data sets. We carried out this analysis for the matrix of 468 lakes with all species, the matrix of 468 lakes with only the planktonic species (sampling methods being either water column or surface sediment) and the matrix of 51 lakes with all species (one analysis where the sampling methods being water column visit 1, water column visit 2 and surface sediment, and another where the sampling methods were the mean of the water column visits and the surface sediment).

Software

For all statistical analyses and the majority of data preprocessing, we used R v. 3.0.2 (R Core Team, 2013). We tested environmental variables for normality using the Shapiro–Wilk test (*Shapiro.test*() (`stats`)) and *skewness*() in `moments` (Komsta & Novomestky, 2013). The Box-Cox transformation was used to normalise non-normal variables by applying *boxcox.fit*() (`geoR`) (Ribeiro & Diggle, 2013). Spatial variables were generated by developing a matrix of synthetic Moran's eigenvector maps (i.e. distance-based MEM) using the geographic coordinates of the lakes on a Cartesian plane and the appropriate functions from packages `PCNM`, `ade4` (Dray *et al.*, 2013), `spacemakeR` (Dray *et al.*, 2013) and `packfor` (Dray *et al.*, 2013). Third-order polynomials were also generated for the site coordinates to act as more simple spatial variables. Forward selection of both environmental and spatial variables was completed using *forward.sel*() in `packfor` and verified using forward-backward-stepwise selection using *ordistep*() in `vegan`. The water column diatom matrix was detrended for use in the selection of environmental variables. Detrended water column and surface sediment diatoms were used for the selection of spatial variables. We used *varpart*() in `vegan` (Oksanen *et al.*, 2013) to perform the variation partitioning analysis and used the *rda*(), *predict.rda*() and *scores*() functions of that same package for the ordination work. We computed RV coefficients using *coeffRV*() in package `FactoMineR` (Husson *et al.*, 2014).

Results

Diatom species in the water column and surface sediment data sets

We performed analyses at both the species and genus levels, but only the species-level results are shown, as genus-level results did not differ greatly. After removing species with low abundance from the data matrices (species with <5% relative abundance in any sample), the total number of diatom species used in analyses was 456 (338 benthic or tychoplanktonic forms and 118 planktonic species). Total species richness was 228 in the water column diatom matrix and 382 for the surface sediment. Including species from all habitats, the water column diatom matrix was significantly correlated with the surface sediment diatom matrix, with an RV coefficient of 0.23 for the first axis of variation in the PCA ($P < 0.001$; see Table 2). With planktonic-only species, the RV coefficient was 0.24 ($P < 0.001$). Independent PCAs of the two types of diatom data sets enabled us to ascertain the similarity in the distribution of taxa across sites. Qualitatively we observed that *Aulacoseira granulata* and *Fragilaria crotonensis* were the dominant taxa driving the first and second PC axes in both data sets (Fig. 2a,b).

Table 2 RV coefficients from comparisons of diatom assemblage matrices and lake positions within redundancy analysis (RDAs)

Ordination overlay	RV coefficient of 1st axis of fitted scores (*P*-value)	RV coefficient of (full set) of fitted scores (*P*-value)
Matrix A: Site scores from WC diatom assemblage PCA Matrix B: Site scores from SSed diatom assemblage PCA (all species)	0.23 ($P < 0.001$)	0.63 ($P < 0.001$)
Matrix A: Site scores from WC diatom assemblage PCA Matrix B: Site scores SSed diatom assemblage PCA (planktonic only)	0.24 ($P < 0.001$)	0.23 ($P < 0.001$)
Matrix A: Site scores from WC environmental RDA Matrix B: Site scores from SSed environmental RDA (all species)	0.54 ($P < 0.001$)	0.17 ($P < 0.001$)
Matrix A: Site scores from WC spatial RDA Matrix B: Site scores from SSed spatial RDA(all species)	0.54 ($P < 0.001$)	0.16 ($P < 0.001$)
Matrix A: Site scores from WC environmental RDA Matrix B: Site scores from SSed environmental RDA (planktonic only)	0.50 ($P < 0.001$)	0.22 ($P < 0.001$)
Matrix A: Site scores from WC spatial RDA Matrix B: Site scores from SSed spatial RDA (planktonic)	0.53 ($P < 0.001$)	0.10 ($P < 0.001$)
Matrix A: Site scores from WC environmental RDA (visit 1, 51 sites) Matrix B: Site scores from SSed environmental RDA (51 sites) (all species)	0.03 ($P = 0.3$)	0.38 ($P < 0.001$)
Matrix A: Site scores from WC environmental RDA (visit 2, 51 sites) Matrix B: Site scores from SSed environmental RDA (51 sites) (all species)	0.25 ($P = 0.0002$)	0.07 ($P < 0.001$)
Matrix A: Site scores from WC environmental RDA (mean visits, 51 sites) Matrix B: Site scores from SSed environmental RDA (51 sites) (all species)	0.47 ($P < 0.001$)	0.5 ($P < 0.001$)

'WC' refers to water column samples and 'SSed' to surface sediment samples. The 'Ordination overlay' column lists the two matrices, of which their structure is compared *symmetrically* using an RV coefficient. The column for the RV coefficient of the full set of fitted scores refers to the comparison of scores from all of the axes within an ordination versus the RV coefficient of the 1st axis scores, which is the correlation between the main axis of variation in one matrix and the main axis of variation in another (synonymous with Ordinary Least Squares Regression).

Environmental and spatial drivers of diatom variation

From the original set of 20 environmental variables, 14 were retained by the selection procedure to explain diatom community structure in the water column data set: mean Secchi, Z_{max}, SO_4, DOC, Mg, conductivity, Ca, TP, NH_4, Cl, TN, turbidity, chl *a* and colour (see Table S3 for information on the Box-Cox transformed environmental variables). A similar set of environmental predictors were identified when the surface sediment data set was used as the response matrix. Forward selection using subsets of the water column data resulted in few significant environmental predictors. When only the 118 planktonic species from the water column data set were considered, the significant environmental variables were conductivity, Ca, mean Secchi, Cl, chlorophyll *a* and observed Z_{max}. The forward-selected variables identified using only visit 1 water column diatoms, or only visit 2 water-column diatoms, yielded only three significant variables: TP, conductivity and turbidity (the same for each of the visit data).

Diatom assemblages from the water column were mostly explained by productivity-related variables (i.e. mean Secchi depth and chl *a* with the water column samples) as well as Z_{max} (RDA1 = 0.43 variation explained) and to a lesser extent by variables related to lake identity or catchment chemistry (RDA2 = 0.14) (Fig. 3a). Diatom assemblages from the surface sediment were also mostly explained by productivity-related variables (RDA1 = 0.41 variation explained) (Fig. 3b). Planktonic-restricted RDA biplots showed similar sorting patterns to the complete diatom assemblage plots with a primary axis related to chlorophyll *a*, colour and mean Secchi for both the water column diatoms and surface sediment diatoms (figures not shown). The first and second RDA axes explained 60 and 16% of variation for the water column diatoms, and 57 and 17% for the surface sediment diatoms. RDAs using the significant variables for the 51-lake data set of the visit 1, visit 2, averaged visit samples and surface sediment samples showed turbidity along the primary axis of all four ordinations (figures not shown). RDA1 values were 0.48, 0.57, 0.41 and 0.45 for visit 1, visit 2, mean visits and surface sediment, respectively.

The spatial RDAs were similar to the environmental RDAs, in that the surface sediment site scores (Fig. 4a) displayed a similar pattern in the ordination to the water column site scores (Fig. 4b), for both the RDA with all species and the planktonic-only RDA. The first RDA values were 0.35 and 0.44 for the water column spatial RDA and the surface sediment spatial RDA, respectively, with the second RDA values being 0.17 and

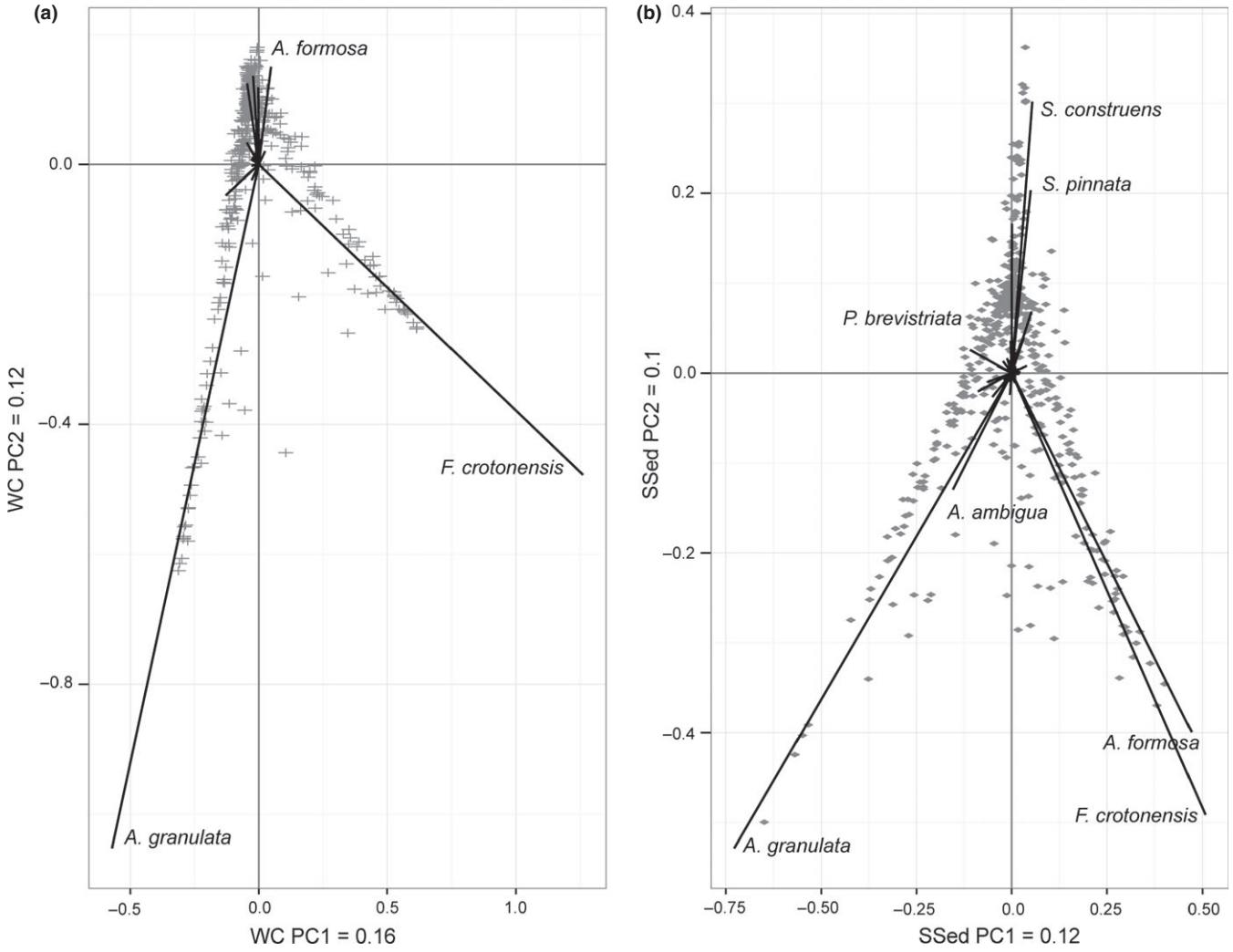

Fig. 2 PCA biplots of (a) water column diatom species and (b) surface sediment diatoms. Species shown in the PCA biplots are those with vectors greater than or approaching 0.2 units: *Asterionella formosa*, *Aulacoseira ambigua*, *Aulacoseira granulata*, *Fragilaria crotonensis*, *Staurosira construens* and *Staurosirella pinnata*. Water column samples are represented by '+' symbols, while the surface sediment samples are represented by filled shapes.

0.20 (RDAs with all species). For planktonic species only, the first and second RDA values were 0.35 and 0.18 for the water column RDA, and 0.49 and 0.19 for the surface sediment RDA. As evident from both the environmental and spatial RDAs, sites appear to be structured in the same way for both surface sediment and water column samples across both types of variables.

Contrary to expectations, the amount of variation explained by space did not differ substantially between the water column and surface sediment samples. Pure space explained *c.* 3.8% of variation in the water column diatoms and *c.* 5.6% in surface sediment diatoms. Pure environment explained *c.* 4.9% of variation in water column diatoms and *c.* 5.4% in surface sediment diatoms, with 88% of variation being unexplained for water column diatoms and 85% for surface sediment diatoms. Removing benthic species from the diatom matrices (such that only 118 planktonic or tychoplanktonic species remained) did not change the proportion of total variation explained by either environmental or spatial variables.

Correlation between environmental and spatial ordinations

The first axis of variation in the environmental RDA explained 43% of variation in the water column assemblage and 41% of the surface sediment assemblage. The RV between site scores of the water column diatom assemblage and the site scores from the surface sediment RDA for this first axis was 0.54 ($P < 0.001$;

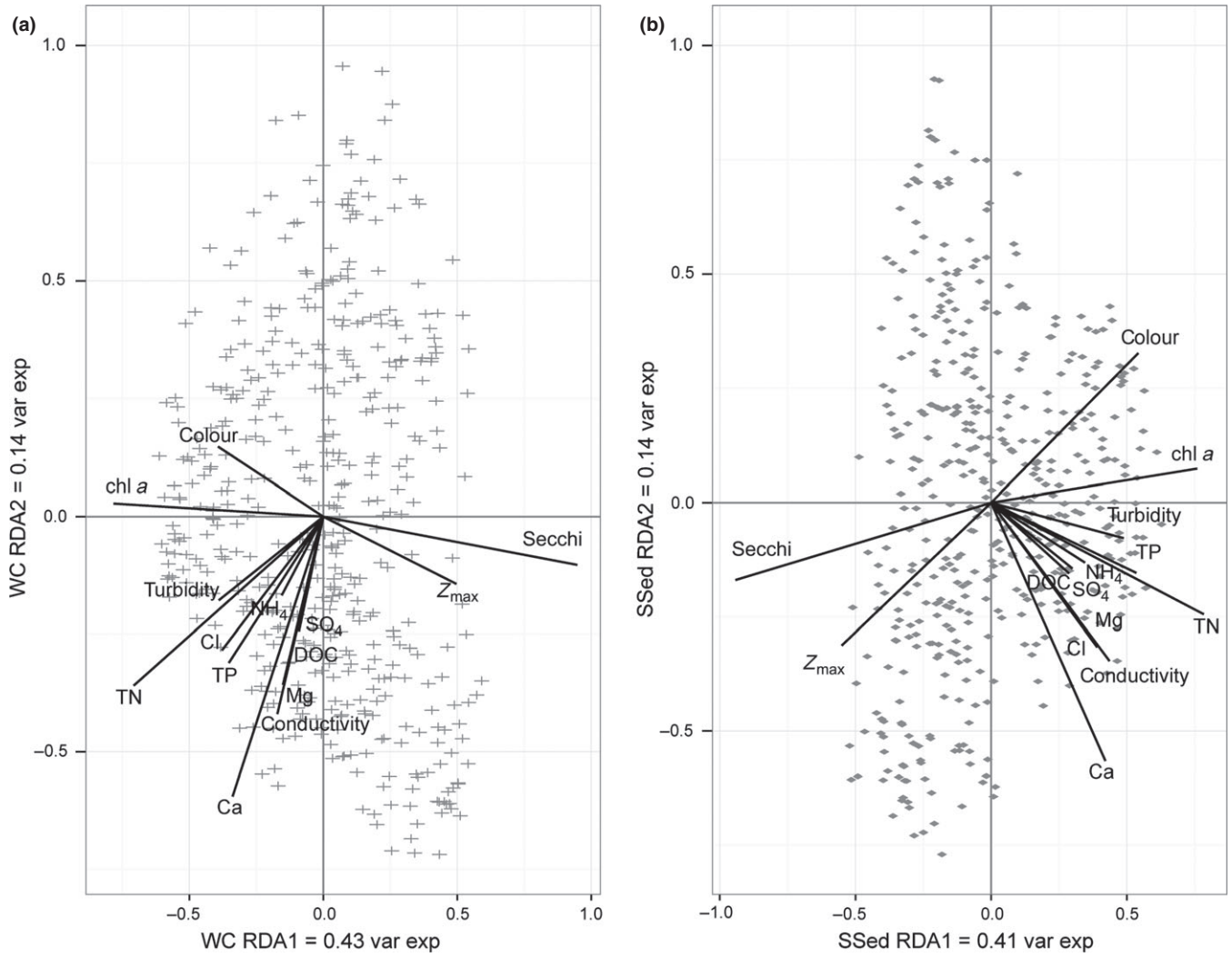

Fig. 3 Biplots of the first two axes from redundancy analysis (RDA) of environmental variables using diatom species from the 468 study lakes: (a) using water column diatoms and (b) using surface sediment diatoms. Water column samples are represented by '+' symbols, while the surface sediment samples represented are by filled shapes. The environmental variables depicted by the arrows were selected using forward selection. Variables left untransformed were the following: Z_{max} (i.e. maximum observed depth), conductivity, Ca, Mg, SO_4, NH_4, TP, turbidity, DOC and colour (PCU). Box-Cox transformed variables were TN, Chl *a* and Secchi.

RV = 0.17 for full set of scores). As we found with the environmental matrix, there was consistency between the spatial predictors when RDAs were performed using either the water column or surface sediment data sets (Fig. 4). The RV coefficient value for the correlation between water column RDA1 (35% variance explained) site scores and surface sediment RDA1 (44% variance explained) site scores was 0.54 for the spatial RDA ($P < 0.001$; 0.16 for the full set of scores). As such, there appears to be quantifiable congruence between the species-by-site data from both the water column and surface sediment samples.

For the planktonic-only analyses, RV coefficients between RDA1 water column site scores and RDA1 surface sediment site scores were 0.50 for the environmental RDA (using both actual surface sediment site scores and predicted) and 0.53 for the spatial RDA 1 [$P < 0.001$; RV = 0.22 (env) and 0.10 (spatial) for full set of scores]. The strength of the correlation between averaged water column samples and surface sediment samples was stronger than the correlation between single snapshot and surface sediment samples. For the water column visit 1 comparison to surface sediment, the RV coefficient was 0.38 for the first axis and 0.03 for all the axes ($P < 0.001$). The correlation was weaker for the water column visit 2 comparison with surface sediment with an RV coefficient of 0.07 ($P = 0.001$; 0.25 for the full set of axes). However, the strength of the correlation

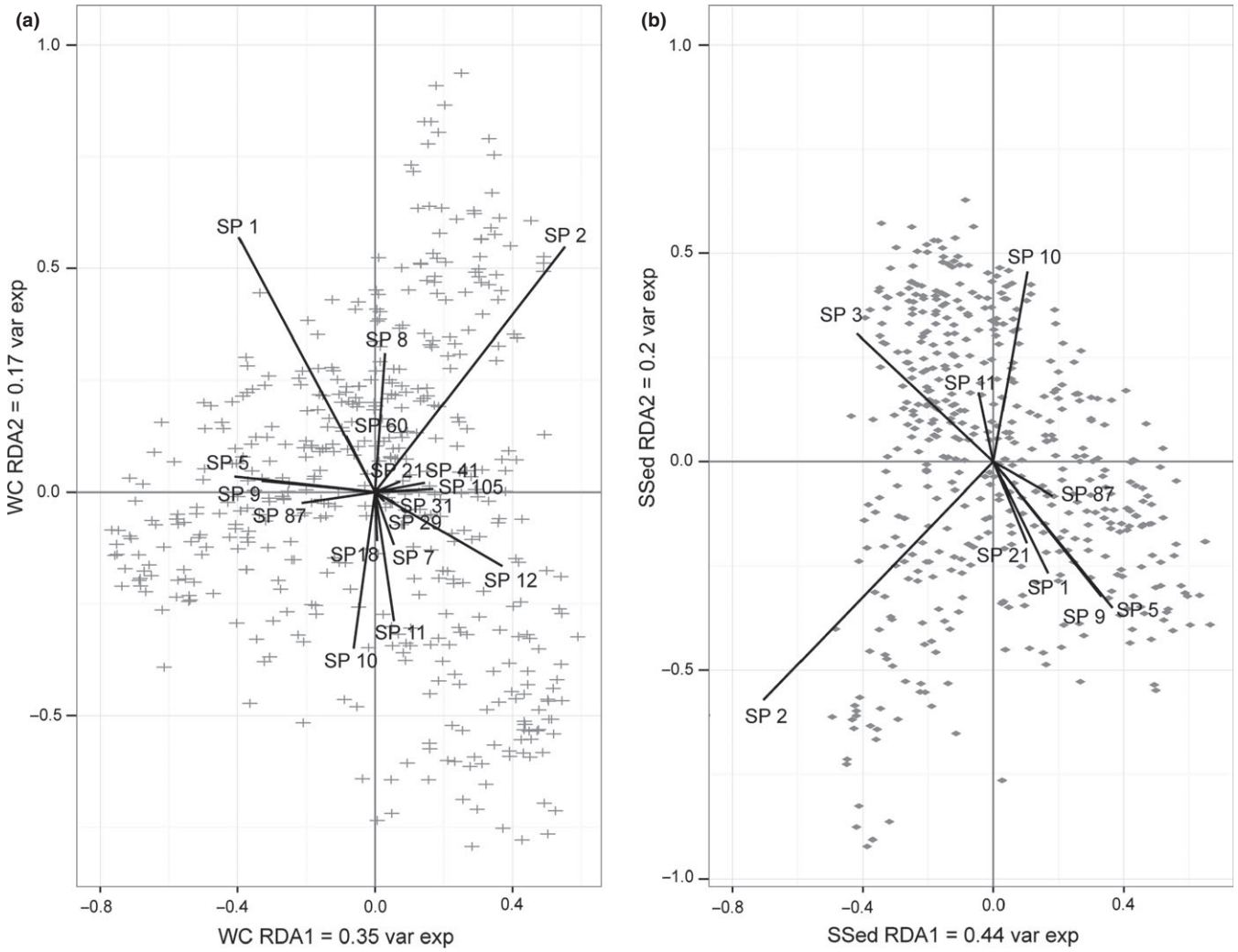

Fig. 4 Biplots of (a) the redundancy analysis (RDA) of 17 spatial PCNM predictors from the 468 study lakes, using the water column diatom data and (b) the 9 spatial PCNM predictors from the 468 study lakes, using the surface sediment diatom data. The spatial variables consisted of PCNM predictors ('SP') selected using forward selection from 105 PCNM predictors.

increased when comparing the averaged water column data to the surface sediment data, RV = 0.5 ($P < 0.001$; 0.47 for the full set of axes).

Comparing sampling types

Partial RDAs constrained diatom data to sampling method while controlling for the variation among lakes, as the lake sites were the same for both the water column and surface sediment samples. The proportion of variation explained by sampling type (water column or surface sediment) for the $n = 468$ data set of all diatom taxa was 0.015 (adj. $R^2 = 0.014$; pseudo F-value = 13.9; $P = 0.005$). The proportion of variation explained by sampling type with planktonic species only was 0.013 (adj. $R^2 = 0.012$; pseudo F-value = 12.7; $P = 0.005$). The proportion of variation explained by sampling type when considering water column visit 1, water-column visit 2 and surface sediment was also 0.013 (adj. $R^2 = 0.010$; pseudo F-value = 3.2; $P = 0.005$). The proportion of variation explained by sampling type when considering the mean of the water column visits and surface sediment was 0.009 (adj. $R^2 = 0.010$; pseudo F-value = 3.2; $P = 0.015$). These results were consistent with our other analyses, in that there was a negligible effect associated with the sampling method.

Discussion

Palaeolimnology has been used extensively in tracking long-term environmental change. However, there are also numerous examples of how palaeolimnological

approaches can be applied to ecological questions on more contemporary time scales: response to nutrient reduction (Battarbee *et al.*, 2005), tracking invasive species (Hawryshyn *et al.*, 2012), space for time substitutions (Blois *et al.*, 2013) and questions related to human–environment interactions, biogeochemical cycling and combining multiple records (Seddon *et al.*, 2014). An awareness of both the shared attributes of these sampling types and their differences is crucial when using these data in concert. Perhaps more critically, an understanding of where there is the potential to draw different conclusions about the effects of environmental variation on aquatic community composition is necessary, especially as data from these two different sampling methods are increasingly being integrated into ecological research (e.g., Gregory-Eaves & Beisner, 2011; Velghe & Gregory-Eaves, 2013).

We found that both types of data sets yielded similar relationships with the environmental and spatial predictors, despite a low amount of explained variation. The most prominent environmental variables related to this 468 lake data set were mean Secchi depth, Z_{max}, chlorophyll *a* and colour, irrespective of whether benthic taxa were included or whether water column or surface sediment samples were considered. When a smaller subset of lakes with multiple sampling dates was considered, the main environmental variables were conductivity, total phosphorus and turbidity. Thus, with this large set of study lakes, researchers analysing environmental data would have drawn a similar (RV = 0.54, $P < 0.001$ relationship between water-column and surface sediment RDA scores) conclusion, regardless of whether they had access to surface sediment or water column diatom counts; diatom variation was mainly structured by lake primary productivity. It is worth noting that RV coefficients were generally lower when looking at the correlative structure amongst matrices representing the full set of RDA scores, but still significant. We also found that variation across spatial variables was similar between surface sediment and water column diatoms and that the significant spatial structure identified for both types of sampling was reflective of regional scale processes. This result echoes findings from a few other studies that have identified the importance of space across larger scales (Verleyen *et al.*, 2009; Bennett *et al.*, 2010).

Our original hypotheses emphasised differences between water column and surface sediment samples, mostly with respect to the integration across habitat types and time scales (seasonal or even annual). Our rationale was that environmental variables would more fully explain variation in diatom communities preserved in surface sediment than captured from the water column for two primary reasons: first because benthic diatom species more closely track environmental conditions than do planktonic species (Philibert & Prairie, 2002), and second because palaeolimnological samples integrate over a longer time period. This means that surface sediment diatom assemblages would reflect communities observed over a longer period of time (at least an entire growing season), capturing species that are temporally transient or may show a patchy distribution in a system, thereby resulting in more complete species sorting across environmental gradients. We found only an approximately similar relationship between the water column environmental RDA and surface sediment environmental RDA when including only planktonic species (RV = 0.50 versus RV = 0.54 with all species); however, our study did provide an insight into why we instead found congruence between these data sets despite differences in species, and this information could be useful when planning sampling methods or combining data, a main goal of the study.

Interestingly, variation explained by the environment was slightly greater in surface sediments than in water column sediments. Variation explained by space was also slightly greater in surface sediments when compared to water column sediments, resulting in overall lower unexplained variation in diatom assemblages when using surface sediment samples. These differences were minor though (e.g. 88% unexplained variation for water-column diatoms, 85% unexplained variation for surface sediment diatoms). This means that there was but weak support for our hypothesis that diatoms from surface sediment samples would be more strongly structured by environmental variables than by spatial eigenfunctions (as studied using dbMEM spatial variables). While the large amount of unexplained variation in the variation partitioning analyses necessitates a cautious interpretation of these results, the amount of variation explained is not disproportionate to other large surveys.

We did find evidence that the relationships of diatom assemblages from surface sediment samples to environmental variation was significantly more similar to seasonally averaged water column samples than to individual snapshot samples (regardless of time in the growing season). In particular, we saw a higher RV coefficient between scores from an environmental RDA where surface sediment diatoms were compared to water column samples where diatom species collected from two visits from early and later in the growing season were averaged. This could relate to an aspect of time integration that we did not consider in our initial

hypotheses. In particular, surface sediment samples (and time-integrated water column samples) are probably more similar to each other because both more accurately reflect cyclical changes of abundance amongst diatom species. This is true even if their variation is not explained in a significantly higher proportion by the environmental variables.

Diatom species are highly dynamic and generally display two peaks in abundance throughout the growing season, in spring and early autumn. It is well known that environmental variables related to lake productivity are also seasonally cyclical. This family of variables (e.g. Secchi depth, chlorophyll *a*) most effectively represented variation in our diatom communities and thus may explain why integrating over the growing season results in a closer match between surface sediment samples and temporally averaged water-column samples. Our PCAs of diatom assemblages identified *Cyclotella* spp. and *Fragilaria* spp. (*sensu stricto*) as key diatom taxa. Both of these genera contain species known to show large seasonal peaks (at least in ponds), with autumn being an important month in temperate systems for some of these species (Köster & Pienitz, 2006). *Cyclotella* spp. also show periodicity in palaeolimnological records (Saros & Anderson, 2014). The set of 468 lakes used for the main analysis and the first visit samples for the smaller subset of 51 lakes, which consisted of water column samples collected in May or June, would only capture at best one of the large seasonal peaks of phytoplankton, whereas the surface sediment samples collected at the same time would have included diatoms from peaks in abundance of the previous growing season. For genera like these two examples, timing of water column sampling can result in different community compositions, altering the conclusions drawn about metacommunity composition. As a result, while both water column and surface sediment samples yielded the same environmental signals, their assemblage resemblance appears to depend on the timing of sampling and can be enhanced by comparing surface sediment samples to averaged data from the water column over multiple sampling points.

Recent studies of environmental drivers of water column diatom composition have found that many variables are significant contributors, including lake productivity, longitude, nitrate, nitrate/nitrite levels, pH, phosphate, silica, stratification, TP and per cent surrounding vegetation (Vanormelingen, Verleyen & Vyverman, 2008; Soininen & Weckström, 2009; Ptacnik *et al*., 2010; Gottschalk & Kahlert, 2012). A similarly wide set of environmental variables have been found to affect diatom surface sediment composition (core samples or sediment traps), including Ca, chlorophyll *a*, Cl, conductivity, elevation, K, lake circulation, Mg, Na, pH, surface area, TN and TP (Dixit *et al*., 1999; Köster & Pienitz, 2006; Hausmann & Pienitz, 2009; Leira *et al*., 2009; Verleyen *et al*., 2009; Bennett *et al*., 2010; Hájek *et al*., 2011). In some cases, the connectivity of habitats and variables related to dispersal limitation has been identified as important predictors at certain scales (Vyverman *et al*., 2007; Vanormelingen *et al*., 2008). With such a wealth of knowledge present in the literature, the challenge is not in finding studies to corroborate the importance of a candidate variable, but in realising that with different gradients and different measured environmental variables, many outcomes are possible with respect to drivers of diatom community composition. The relevance to this study is that, for any of these different study examples, the same conclusion about the important environmental variables could probably have been reached regardless of using water column or surface sediment diatom samples. This is shown by correlated ordination structure, but also by forward selection of environmental variables using the different sampling types.

Unlike many other sampling programmes where a particular environmental gradient is targeted, the primary goal of the NLA survey was to randomly sample from all lakes in the continental U.S.A. that were deeper than 1 m and larger than 1 ha in surface area. As such, the relatively high-nutrient status evident in this data set (i.e. the median values for total phosphorus is indicative of eutrophic conditions and mesotrophic based chlorophyll *a*) is reflective of the average trophic state of most US lakes. Previous research in more oligotrophic (nutrient poor) systems has shown lake pH to be a dominant structuring variable for diatoms (e.g. Ginn, Cumming & Smol, 2007; Valois, Keller & Ramcharan, 2011), with acidification resulting in a loss of planktonic taxa (Battarbee *et al*., 1984) and close tracking of benthic taxa to environmental gradients. If pH had been a more important variable for diatom communities across the lakes in this study, we may not have come to the same conclusions about diatom variation with both the water column and surface sediment samples; instead, the different types of samples may have yielded different results (not in taxonomic composition, but in the key structuring variables identified). While we recognise that different regions with particularly low or high pH levels may see a significant effect on diatom communities, pH may not be as important at the continental scale as it was in the 1980s (e.g. Wigington *et al*., 1992).

A further impetus behind our study was to provide insights into some of the perceived challenges associated

with comparing contemporary and palaeolimnological studies, including their joint use. We think this type of work is central to collaborative research in the aquatic sciences. The general message from our analyses is that there are broadly similar patterns from the analyses of diatom communities as captured by the surface sediments and water column samples, although the greatest similarity is evident when water column samples are pooled across time to reflect a time-integrated sample. Data sharing is one way in which these two branches of aquatic ecology can work together more concretely, and this is especially important as both the availability of data and requirements for data storage evolve in ecology (Hampton *et al*., 2013). Nonetheless, there are many reasons why researchers may choose one type of sampling over another for a given study. For example, palaeolimnology studies have been very useful in quantifying environmental change in a large number of lakes (i.e. upwards of 50) through analysing pre-industrial (pre-1850 CE) and surface sediments (e.g., Dixit *et al*., 1999). On the other hand, direct water column sampling can allow for a more thorough representation of total algal community diversity (as opposed to just diatoms which are often the target of palaeolimnological studies).

Future research directions and potential implications for monitoring programmes

This study was singular in its focus on diatom taxa. However, as is evident in many of the works cited herein, this question is also important for other organismal groups. As such, follow-up studies could conduct similar analyses with zooplankton subfossils and compare results to those presented in this work and with studies utilising sediment traps and net samples to track seasonal changes in zooplankton communities (e.g. see Nykänen *et al*., 2009 and Alric & Perga, 2011). While our study did focus on the effect of averaging water column samples across a season, we were not able to compare time series data from more than one year to a full core sediment record. This would be a logical extension to our work where we have shown comparisons between early visit and later visit water column sampling to surface sediment samples.

Acknowledgments

The authors wish to thank Amina Pollard of the NLA programme for help with data access and interpretation, as well as Mihaela Enache and Marina Popatova at The Patrick Center for Environmental Research (Academy of Natural Sciences Philadelphia; ANSP) for diatom information and help procuring samples. Members of the Gregory-Eaves lab at McGill contributed greatly to the development of this work. We also thank two anonymous reviewers for their comments on an earlier version of this manuscript. Funding for this project was made possible by an FRQNT Equipe grant awarded to IGE, BB and PL, as well as an NSERC PGS-D scholarship awarded to AKW. Travel to collect samples from the ANSP was supported by a McGill International Travel Grant. This project has also benefitted from the support of a Québec Centre for Biodiversity Science excellence award to AKW. This study is a contribution to the FQRNT-funded Québec Centre for Biodiversity Science (QCBS-CSBQ).

References

Alric B. & Perga M.E. (2011) Effects of production, sedimentation and taphonomic processes on the composition and size structure of sedimenting cladoceran remains in a large deep subalpine lake: paleo-ecological implications. *Hydrobiologia*, **676**, 101–116.

Battarbee R.W., Anderson N.J., Jeppesen E. & Leavitt P.R. (2005) Combining paleolimnological and limnological approaches in assessing lake ecosystem response to nutrient reduction. *Freshwater Biology*, **50**, 1772–1780.

Battarbee R.W., Thrush B.A., Clymo R.S., Le Cren E.D., Goldsmith P., Mellanby K. *et al*. (1984) Diatom analysis and acidification of lakes. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, **305**, 451–477.

Beaulieu M., Pick F. & Gregory-Eaves I. (2013) Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set. *Limnology and Oceanography*, **58**, 1736–1746.

Bennett J.R., Cumming B.F., Ginn B.K. & Smol J.P. (2010) Broad-scale environmental response and niche conservation in lacustrine diatom communities. *Global Ecology and Biogeography*, **19**, 724–732.

Birks H.H., Jones V.J., Brooks S.J., Birks H.J.B., Telford R.J.S. & Peglar S.M. (2012) From cold to cool in northernmost Norway: lateglacial and early Holocene multi-proxy environmental and climate reconstructions from Jansvatnet, Hammerfest. *Quaternary Science Reviews*, **33**, 100–120.

Blois J.L., Williams J.W., Fitzpatrick M.C., Jackson S.T. & Ferrier S. (2013) Space can substitute for time in predicting climate change effects on biodiversity. *Proceedings of the National Academy of Sciences*, **110**, 9374–9379.

Brothers S., Vermaire J.C. & Gregory-Eaves I. (2008) Empirical models for describing recent sedimentation rates in lakes distributed across broad spatial scales. *Journal of Paleolimnology*, **40**, 1003–1019.

Charles D.F., Knowles C. & Davis R.S. (2003) Protocols for the analysis of algal samples as collected as part of the U.S. Geological Survey National Water-Quality Assessment program. Patrick Center for Environmental Research, The Academy of Natural Sciences. Report No. 02-06.

Cottenie K. (2005) Integrating environmental and spatial processes in ecological community dynamics. *Ecology Letters*, **8**, 1175–1182.

Dixit S.S., Smol J.P., Charles C.F., Hughes R.M., Paulsen S.G. & Collins G.B. (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.

Dray S., Legendre P., Oksanen J., Blanchet F.G., Solymos P., Thioulouse J. *et al.* (2013) Spatial ecological data analysis with R. sedaR project, R-Forge. Available at: sedar.r-forge.r-project.org.

Flessa K.W. & Jackson S.T. (2005) Forging a common agenda for ecology and paleoecology. *BioScience*, **55**, 1030–1031.

Ginn B.K., Cumming B.F. & Smol J.P. (2007) Diatom-based environmental inferences and model comparisons from 494 northeastern North-American Lakes. *Journal of Phycology*, **43**, 647–661.

Glew J. (1989) A new trigger mechanism for sediment samplers. *Journal of Paleolimnology*, **2**, 241–243.

Gottschalk S. & Kahlert M. (2012) Shifts in taxonomical and guild composition of littoral diatom assemblages along environmental gradients. *Hydrobiologia*, **694**, 41–56.

Gregory-Eaves I. & Beisner B.E. (2011) Paleolimnological insights for biodiversity science: an emerging field. *Freshwater Biology*, **56**, 2653–2661.

Hájek M., Roleček J., Cottenie K., Kintrová K., Horsák M., Poulíčková A. *et al.* (2011) Environmental and spatial controls of biotic assemblages in a discrete semi-terrestrial habitat: comparison of organisms with different dispersal abilities sampled in the same plots. *Journal of Biogeography*, **38**, 1683–1693.

Hampton S.E., Strasser C.A., Tewksbury J.J., Gram W.K., Budden A.E., Batcheller A.L. *et al.* (2013) Big data and the future of ecology. *Frontiers in Ecology and the Environment*, **11**, 156–162.

Hausmann S. & Pienitz R. (2009) Seasonal water chemistry and diatom changes in six boreal lakes of the Laurentian Mountains (Québec, Canada): impacts of climate and timber harvesting. *Hydrobiologia*, **635**, 1–14.

Hawryshyn J., Rühland K.M., Julius M. & Smol J.P. (2012) Absence of evidence is not evidence of absence: is *Stephanodiscus binderanus* (Bacillariophyceae) an exotic species in the Great Lakes region? *Journal of Phycology*, **48**, 270–274.

Heino J. (2009) Species co-occurrence, nestedness and guild-environment relationships in stream macroinvertebrates. *Freshwater Biology*, **54**, 1947–1959.

Husson F., Josse J., Le S. & Mazet J. (2014) Package 'FactoMineR'. Available at: http://cran.r-project.org/web/packages/FactoMineR/FactoMineR.pdf.

Komsta L. & Novomestky F. (2013) Package 'moments'. Available at: www.komsta.net.

Köster D. & Pienitz R. (2006) Seasonal diatom variability and paleolimnological inferences – a case study. *Journal of Paleolimnology*, **35**, 395–416.

Kurek J., Weeber R.C. & Smol J.P. (2011) Environment trumps predation and spatial factors in structuring cladoceran communities from Boreal Shield lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **68**, 1408–1419.

Legendre P. & De Cáceres M. (2013) Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecology Letters*, **16**, 951–963.

Legendre P. & Gallagher E.D. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia*, **129**, 271–280.

Legendre P. & Legendre L. (2012) *Numerical Ecology*, 3rd edn. Elsevier Science BV, Amsterdam, The Netherlands.

Leira M., Chen G., Dalton C., Irvine K. & Taylor D. (2009) Patterns in freshwater diatom taxonomic distinctness along an eutrophication gradient. *Freshwater Biology*, **54**, 1–14.

Levi E.E., Çakiroğlu Aİ., Bucak T., Odgaard B.V., Davidson T.A., Jeppesen E. *et al.* (2014) Similarity between contemporary vegetation and plant remains in the surface sediment in Mediterranean lakes. *Freshwater Biology*, **59**, 724–736.

Nykänen M., Vakkilainen K., Liukkonen M. & Kairesalo T. (2009) Cladoceran remains in lake sediments: a comparison between plankton counts and sediment records. *Journal of Paleolimnology*, **42**, 551–570.

Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B. *et al.* (2013) Package 'vegan'. Available at: cran.r-project.org/web/packages/vegan/vegan.pdf.

Philibert A. & Prairie Y.T. (2002) Diatom-based transfer functions for western Quebec lakes (Abitibi and Haute Mauricie): the possible role of epilimnetic CO_2 concentration in influencing diatom assemblages. *Journal of Paleolimnology*, **27**, 465–480.

Picazo F., Millán A. & Dolédec S. (2012) Are patterns in the taxonomic, biological and ecological traits of water beetles congruent in Mediterranean ecosystems? *Freshwater Biology*, **57**, 2192–2210.

Ptacnik R., Andersen T., Brettum P., Lepistö L. & Willén E. (2010) Regional species pools control community saturation in lake phytoplankton. *Proceedings of the Royal Society of London. Series B*, **277**, 3755–3764.

R Core Team. (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www/r-project.org.

Ribeiro P.J. & Diggle P.J. (2013) Package 'geoR'. Available at: http://cran.r-project.org/web/packages/geoR/geoR.pdf.

Saros J.E. & Anderson N.J. (2014) The ecology of the planktonic diatom *Cyclotella* and its implications for global environmental change studies. *Biological Reviews of the Cambridge Philosophical Society*, http://onlinelibrary.wiley.com/doi/10.1111/brv.12120/full.

Seddon A.W.R., Mackay A.W., Baker A.G., Birks H.J.B., Breman E., Buck C.E., Ellis E.C. *et al.* (2014) Looking forward through the past: identification of 50 priority research questions in paleoecology. *Journal of Ecology*, **102**, 256–267.

Smol J.P. (2008) *Pollution of Lakes and Rivers: A Paleoenvironmental Perspective*, 2nd edn. Blackwell Publishing, Oxford, United Kingdom.

Soininen J. & Weckström J. (2009) Diatom community structure along environmental and spatial gradients in lakes and streams. *Fundamental and Applied Limnology*, **174**, 205–213.

Strayer D.L. & Dudgeon D. (2010) Freshwater biodiversity conservation: recent progress and future challenges. *Journal of the North American Benthological Society*, **29**, 344–358.

USEPA. (2007) *Survey of the Nation's Lakes. Field Operations Manual. EPA 841-B-07-004*. U.S. Environmental Protection Agency,Washington, DC, USA.

USEPA. (2009) *National Lakes Assessment: A Collaborative Survey of the Nation's Lakes. EPA 841-R-09-001*. U.S. Environmental Protection Agency, Office of Water and Office of Research and Development, Washington, DC, USA.

USEPA. (2011–2012) *National Lakes Assessment. Field Operations Manual. EPA 841-B-11-003*. U.S. Environmental Protection Agency, Washington, DC, USA.

Valois A.E., Keller W.B. & Ramcharan C.W. (2011) Recovery in a multiple stressor environment: using the reference condition approach to examine zooplankton community change along opposing gradients. *Journal of Plankton Research*, **33**, 1417–1429.

Vanormelingen P., Verleyen E. & Vyverman W. (2008) The diversity and distribution of diatoms: from cosmopolitanism to narrow endemism. *Biodiversity and Conservation*, **17**, 393–405.

Velghe K. & Gregory-Eaves I. (2013) Body size is a significant predictor of congruency in species richness patterns: a meta-analysis. *PLoS ONE*, **8**, e57019.

Verleyen E., Vyverman W., Sterken M., Hodgson D.A., De Wever A., Juggins S. *et al.* (2009) The importance of dispersal related and local factors in shaping the taxonomic structure of diatom metacommunities. *Oikos*, **118**, 1239–1249.

Vyverman W., Verleyen E., Sabbe K., Vanhoutte K., Sterken M., Hodgson D.A. *et al.* (2007) Historical processes constrain patterns in global diatom diversity. *Ecology*, **88**, 1924–1931.

Wehr J.D. & Sheath R.G. (2003) *Freshwater Algae of North America: Ecology and Classification*. Academic Press, Elsevier Science, New York, NY, USA.

Wigington P.J. Jr, Davies T.D., Tranter M. & Eshleman K.N. (1992) Comparison of episodic acidification in Canada, Europe and the United States. *Environmental Pollution*, **78**, 29–35.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Data matrices from metadata used to create the data matrices for statistical analyses.

Table S2. Classification of diatom taxa as benthic, planktontic and tychoplanktonic.

Table S3. Transformation parameters for environmental variables transformed using Box-Cox transformation, rounded to two decimal places.

(Manuscript accepted 9 September 2014)