CHARACTERIZATION OF SPATIAL AND ENVIRONMENTAL INFLUENCES ON STREAM DIATOMS AND CYANOBACTERIA

by

NICHOLAS OGDEN SCHULTE

B.S., Tennessee Technological University, 2012

M.S., Florida International University 2016

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of Doctor of Philosophy

Environmental Studies Program

2020

Committee Members:

Diane McKnight

Sarah Spaulding

Daren Carlisle

Noah Fierer

Patrick Kociolek

Schulte, Nicholas Ogden (Ph.D., Environmental Studies)

Characterization of spatial and environmental influences on stream diatoms and cyanobacteria

Thesis directed by Professor Diane McKnight

ABSTRACT

Primary producing algae form the basis of carbon fixation, oxygen production, and food webs in aquatic ecosystems globally. However, human activities disrupt climate and freshwater physicochemistry. These impacts alter the health of algal communities and the ecosystem services algae provide. Meanwhile, spatial processes like dispersal and landscape characteristics like geology also influence algal structure and function. Diatoms are indicators of stream health and are model organisms for understanding the processes underlying microbial biogeography. Benthic cyanobacteria present risks to human health through the proliferation of toxin-producing blooms. With this dissertation, I investigate the ecosystem processes that influence diatom and cyanobacterial community composition and taxon distributions. My goal is to advance the understanding of ecosystem controls on algal biogeography and to characterize taxon-specific autecology for use in environmental management. First, I measured the extent of wind-mediated dispersal of benthic diatoms across aquatic habitats to better understand how community composition is structured by spatial processes across the McMurdo Dry Valleys polar desert in Antarctica. I found that inter-habitat dispersal is common but less influential on community composition than intra-habitat factors such as environmental conditions. I then used non-linear, multivariable modeling to assess the relative influences of climate, watershed characteristics, and in-stream stressors on the relative abundances of 268 diatom taxa across gradients of human impact in the northeast United States. My results indicate diatom taxa are affected by different suites of environmental conditions but that taxa belong to ecological guilds based on shared responsiveness to environmental factors. Finally, I applied multivariable modeling towards understanding the effects of aquatic stressors, including herbicides and persistent organic pollutants, on the distributions of benthic cyanobacteria across northeast U.S. streams. I found that watershed characteristics, streamflow, and herbicides were more influential than light availability, water temperature, and nutrients on the distributions of potentially toxigenic cyanobacterial genera. Collectively, this research expands the knowledge of how benthic algal communities and taxon distributions are structured at large spatial scales along gradients of unimpacted and human-altered environmental conditions. I provide a novel modeling framework and taxon-specific autecological information that can be applied to environmental assessments of stream health and future algal research.

ACKNOWLEDGMENTS

I am deeply grateful for the support of my co-advisors Diane McKnight and Sarah Spaulding. Diane accommodated my wide-ranging scientific interests, introduced me to the allure of Antarctica, and encouraged me to pursue meaningful questions. She also makes the only pie I have ever looked forward to. Sarah introduced me to a variety of algae-related research and provided me the freedom and direction to discover what excited me most. Through New Jersey stream sampling, Utah stick-shifting, and daily COVID-19 virtual check-ins, Sarah showed me that self-care and compassion for others are integral parts of being a successful scientist.

I also thank my committee members whose scientific insight and guidance made this dissertation possible. Daren Carlisle was quick to respond when I had a modeling question or three and was always up for a discussion on gardening and classic rock. Noah Fierer brought me up to speed on DNA sequencing, and Pat Kociolek kept me thinking deeply about algae.

This research is built upon the dedicated work of innumerable members of the McMurdo Dry Valleys Long-Term Ecological Research (MCM LTER) algae, meteorology, soils, and streams teams that contributed to the field collections and laboratory analyses of the McMurdo algal records. Alia Khan, Emma Smith, and Byron Adams were integral collaborators for the research presented in Chapter II. I am also indebted to all those who collected and analyzed data for the U.S. Geological Survey (USGS) Northeast Stream Quality Assessment.

I am thankful for the institutional support provided by the University of Colorado Boulder, Environmental Studies Program (ENVS), and Institute of Arctic and Alpine Research (INSTAAR). This dissertation was funded by research assistantships through the MCM LTER (National Science Foundation Grant #OPP-1637708) and USGS (Cooperative Agreement #G15 AC00104), as well as a teaching assistantship through ENVS under the tutelage of Eve Hinckley.

Additional funding was provided by the Center for Water, Earth Science and Technology Aiken Endowed Memorial Graduate Research Fellowship and the American Water Resources Association Colorado Section Richard Herbert Memorial Scholarship.

I had the privilege of introducing three driven undergraduate students to the wonderful world of Antarctic diatoms: Maya Samuels, Emma Smith, and Jenna Nolder. With two lab groups, ENVS, and INSTAAR I was surrounded by a community of friends and colleagues that kept me focused and engaged, including Maggie Bowman, Jessie Egan, Anna Hermes, Tanya Petach, and Brian Straight. Penny Bates put up with my incessant questions and shared with me dog pictures and stories to get me through the days. Josh Darling and Eric Sokol were jam partners during much-needed Phycology Fridays. Ian Bishop and Meredith Tyree were always around to talk diatoms and life. Joel Singley's camaraderie and commiseration kept me going through lockdowns, elections, and the slow churn of graduate school. Fieldwork has a way of bringing people together, and I'm grateful for the bad song parodies, Tui, and games of Wizard I shared with members of the Lollygaggers and Algae Oops teams of 2016-17 in McMurdo.

I am also grateful for the friendship and grounding of David Burge, Raegan Guinn, Shelby Servais, and Angela Zoumplis. Facilitating philosophy discussions for children and adults with Kelly Cowling through Grey Havens Philosophy showed me the importance of promoting critical and creative thinking within my community. Finally, this journey was made possible by the love of my family, who have encouraged and supported me through everything up to this point: my hiking companion and dog, Bailey; my wife, Ashley; and my parents, Gail and Randy. Thank you all.

CONTENTS

CHAPTER

TABLES

Table

FIGURES

Figure

CHAPTER I

INTRODUCTION

Algae (eukaryotic algae and cyanobacteria) are biologically diverse, form the base of food webs, and produce much of the carbon and oxygen in aquatic ecosystems (Stevenson et al. 1996). However, human disturbances dramatically disrupt the climate, geomorphology, and biogeochemistry of freshwater streams on a global scale (Maloney and Weller 2011). These impacts alter the biotic health of and ecosystem services provided by algae. Environmental changes act alongside landscape characteristics, such as elevation, and spatial processes, such as dispersal, to influence algal communities and populations, but we know little about these interactions. With this dissertation, I characterize the relative influences of natural and human-impacted factors on algal community composition and taxon distributions at large spatial scales. My objective is to advance the understanding of algal biogeography and to develop novel techniques for the use of algae in assessments of environmental and human health.

Diatoms are single-celled eukaryotic algae that are abundant and highly diverse in stream biofilms worldwide. Within a diatom community, each taxon may be associated with specific environmental conditions, such as streamflow, pH, specific conductance, and nutrient concentrations (McCormick and Cairns 1994). Consequently, diatoms have served as markers of water quality degradation and stream health for decades in environmental assessments (Patrick 1973). Additionally, their high diversity, broad distribution, and overall environmental sensitivity have led to the use of diatoms as model organisms for studying the ecosystem processes that influence community composition and taxon distribution patterns (Soininen 2007).

Cyanobacteria (blue-green algae) are unicellular to colonial photosynthetic bacteria that are also common in streams but tend to be less diverse and abundant than diatoms outside of extreme habitats (Stevenson et al. 1996, Komárek 2006). Unlike freshwater diatoms, cyanobacteria present substantial risk to human, domestic animal, and wildlife health through the production of neurotoxins, hepatotoxins, cytotoxins, and endotoxins during events of rapid increases in population size and biomass known as blooms (cyanobacterial harmful algal blooms, or cyanoHABs; Paerl 2018). CyanoHABs are most common and severe in lentic waterbodies, but proliferation of toxigenic cyanobacteria in streams is increasing worldwide without a clear understanding of the causes (Wood et al. 2020). Together, diatoms and cyanobacteria share important roles as primary producers in stream ecosystem structure and function. However, each algal group has distinct applications to research on microbial biogeography and human impacts in streams.

Human impacts on streams such as increases in water temperature, streamflow diversions, and runoff of nutrients and road salt alter algal biodiversity, abundance, distributions, and toxin production (Lowe 1979, Allan 2004, Heisler et al. 2008). An abundance of work has been done to understand the effects of human-caused disturbances on algae at local, in-stream scales (Patrick 1973, Soininen 2014). However, we have yet to fully address how these multi-faceted alterations affect algal metacommunity (communities linked by dispersal) composition and taxon-specific distributions at spatial scales of multiple watersheds and larger (Soininen and Teittinen 2019).

Algae-based environmental assessments are founded on the observations that algal biogeography is strongly influenced by local environmental conditions (Heino 2013). In environmental assessments, patterns in biotic community composition and taxon distributions are used to determine the degree of ecosystem degradation caused by human influences (Bailey et al. 2004). Commonly measured environmental stressors like streamflow, major ions, and nutrients are well characterized in relation to algal patterns at local scales, but contaminants like pesticides, persistent organic pollutants, and metals are rarely incorporated into environmental assessments or research on algal biogeography, despite having known toxic effects on algae (Rimet and Bouchez 2011). Additionally, pesticides and persistent organic pollutants originate exclusively from human activities, making them useful measures of anthropogenic effects on algae in streams (Debenest et al. 2010).

While the influences of in-stream factors on algal metacommunities and taxa are commonly assessed, relatively little is known about the effects of larger-scale environmental processes, such as climate, or landscape characteristics, such as elevation and geology, on metacommunity composition or taxon distributions across streams, watersheds, regions, or continents (Heino et al. 2010). Regional increases in surface water temperature and decreases in streamflow because of drought are expected to increase the abundances of nuisance and cosmopolitan algae in streams (Whitehead et al. 2009). In Finland streams, climate-related factors account for more variation in metacommunity composition and distributions of diatoms than do local-scale factors and watershed characteristics (Pajunen et al. 2016, Jyrkänkallio-Mikkola et al. 2017). Overall, there is growing evidence that environmental processes that operate at large spatial scales, like climate, strongly affect algal metacommunities and individual taxa, but more work is needed to resolve these patterns across space.

In addition to environmental factors, many studies have shown that variation in algal metacommunity composition across disconnected aquatic systems is explained in part by spatial processes (Soininen and Teittinen 2019). Spatial controls are often attributed to taxon-specific dispersal abilities, but algal dispersal is rarely measured directly (Heino et al. 2015). Passive dispersal of algae among aquatic habitats that are not hydrologically connected occurs through animal movement and wind (Kristiansen 1996). Aeolian (wind-mediated) dispersal of algae is expected to be greatest in landscapes that are highly connected by wind-related processes and in which aquatic habitats are seasonally dry, when dehydrated biofilms and individual algae are likely to be most sensitive to lift by wind (Marshall and Chalmers 1997).

Landscape connectivity of intermittent streams in deserts worldwide is driven by aeolian-transported terrestrial inputs (Maher et al. 2010). In addition to abiotic material, biotic metacommunities of intermittent desert streams might be structured by high aeolian dispersal, in addition to strong environmental controls such as seasonal surface water flow (Stanish et al. 2012, Van Dam and Matzke 2016). Characterizing the potential aeolian dispersal of algae across a desert landscape can therefore be used to determine the extent to which dispersal-related processes can contribute to metacommunity structure in end-member systems like intermittent desert streams. Identifying the direct effects of algal dispersal on metacommunity structure may not only lead to better resolution of metacommunity patterns in the study area but may also provide important baseline information for mechanisms by which metacommunities can be structured across spatial scales and environment types.

Disentangling the relative importance of in-stream stressors, watershed characteristics, regional environmental processes like climate, and spatial processes like dispersal is a challenge for understanding algal metacommunity structure and taxon distributions (Holyoak et al. 2005). In particular, the variety of human-induced alterations to environmental conditions in streams has limited the effectiveness of measuring responses of algae to single environmental stressors. Therefore, multivariable techniques must be used to tease apart these complex relationships (Jyrkänkallio-Mikkola et al. 2017). Species distribution models (SDMs) are used to assess the dominant factors, of those measured and included in the model, that affect a taxon’s presence or abundance (Elith and Leathwick 2009). SDMs have been applied sparingly to freshwater algae across large spatial scales, as they are normally used to identify habitat suitability and invasion potential for taxa of special concern (Pearson 2007, Kumar et al. 2009). However, SDMs can also be applied to multiple taxa within a metacommunity to assess variability in taxon responses to environmental processes (Pajunen et al. 2016). Characterizing this variability provides the opportunity to assess complexity in ecosystem controls on algal biogeography and to identify taxa or groups of taxa that can be used to inform environmental assessments.

The objective of this dissertation is to advance the understanding of how algal community composition and taxon-specific patterns are influenced by environmental and spatial processes at large spatial scales. The research I conducted in pursuit of this goal is presented in the following three chapters, which are outlined below. Each of these chapters is formatted for submission to scientific journals. As such, in Chapters II - IV I use the first-person plural to present research that I conducted alongside the contributions and assistance of individuals who will be listed as co-authors and are recognized in the Acknowledgments section of this dissertation. The final chapter examines cross-cutting relationships among Chapters II – IV and presents recommendations for continued research on these topics.

Chapter II: Blowin' in the wind: transport, structure, and regional connectivity of aeolian diatom communities in the McMurdo Sound region, Antarctica. Few direct observations of algal dispersal have been integrated into understanding the spatial component of metacommunity structure, despite evidence indicating strong spatial structuring of certain algal communities. The spatial processes underlying metacommunity structure were investigated by assessing the associations between wind-mediated dispersal of diatoms with community composition of nearby aquatic habitats. Diatoms were enumerated and identified from aeolian material in an ecosystem in which landscape connectivity is driven in large part by wind, the McMurdo Dry Valleys polar desert of Antarctica. There are 50 known benthic freshwater diatom taxa in the McMurdo Dry Valleys, and they are abundant with well-described distributions across ephemeral glacial meltwater streams and ice-covered ponds and lakes. The composition of diatoms in these aquatic habitats was related to that of diatoms in aeolian material to determine the extent of wind dispersal and its effects on diatom metacommunity structure in an end-member ecosystem.

Chapter III: Modeling responses of diatom taxa to climate, watershed characteristics, and in-stream stressors in the northeast United States. Anthropogenic activity has large effects on climate and pollution in streams, and diatoms are commonly used to assess those impacts. However, relatively little is known about how multiple, co-occurrent environmental factors act together to affect the distributions of individual diatom taxa. Diatoms were identified by morphology from 92 samples collected from streams along urbanization and agricultural gradients in the northeast U.S. Species distribution models for 268 diatom taxa were developed using generalized additive modeling to determine the relative influences of climate, watershed characteristics, and in-stream stressors on each taxon in the regional diatom metacommunity.

Chapter IV: The life aquatic: multivariable effects of environmental stressors on benthic cyanobacterial presence and relative abundance across human-impacted streams. Bloom-forming and toxin-producing benthic cyanobacteria are increasing in abundance, frequency, and geographic distribution across freshwater bodies worldwide. These cyanobacteria decrease ecological condition, impair drinking water, harm human and animal health, and disrupt recreational use and economic benefits of freshwater ecosystems. However, little research has been conducted to identify the environmental factors that influence benthic cyanobacterial distributions in streams. Cyanobacteria were identified to genus, using microscopy, across the samples from the northeast U.S. that were analyzed for diatoms in Chapter III. To assess variability in taxon-specific environmental responses, multivariable distribution models were constructed from presence/absence and relative abundance of each potentially toxigenic and non-toxigenic cyanobacterial genus detected.

References

Allan, J. D. (2004). Landscapes and riverscapes: the influence of land use on stream ecosystems. *Annual Review of Ecology and Systematics*, *35*, 257-284.

Bailey, R. C., Norris, R. H., & Reynoldson, T. B. (2004). Bioassessment of freshwater ecosystems. In Bailey, R. C., Norris, R. H., Reynoldson, T. B. (Eds.), *Bioassessment of Freshwater Ecosystems* (pp. 1-15). Springer, Boston, MA.

Debenest, T., Silvestre, J., Coste, M., & Pinelli, E. (2010). Effects of pesticides on freshwater diatoms. *Reviews of Environmental Contamination and Toxicology*, *203*, 87-103.

Elith, J., & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 677-697.

Heino, J. (2013). The importance of metacommunity ecology for environmental assessment research in the freshwater realm. *Biological Reviews*, *88*, 166-178.

Heino, J., Bini, L. M., Karjalainen, S. M., Mykrä, H., Soininen, J., Vieira, L. C. G., & Diniz-Filho, J. A. F. (2010). Geographical patterns of micro-organismal community structure: are diatoms ubiquitously distributed across boreal streams? *Oikos*, *119*, 129-137.

Heino, J., Melo, A. S., Siqueira, T., Soininen, J., Valanko, S., & Bini, L. M. (2015). Metacommunity organisation, spatial extent and dispersal in aquatic systems: patterns, processes and prospects. *Freshwater Biology*, *60*, 845-869.

Heisler, J., Glibert, P. M., Burkholder, J. M., Anderson, D. M., Cochlan, W., Dennison, W. C., ... & Lewitus, A. (2008). Eutrophication and harmful algal blooms: a scientific consensus. *Harmful Algae*, *8*, 3-13.

Holyoak, M., Leibold, M. A., & Holt, R. D. (Eds.) (2005). *Metacommunities: Spatial Dynamics and Ecological Communities*. University of Chicago Press, Chicago, IL.

Jyrkänkallio-Mikkola, J., Meier, S., Heino, J., Laamanen, T., Pajunen, V., Tolonen, K. T., ... & Soininen, J. (2017). Disentangling multi-scale environmental effects on stream microbial communities. *Journal of Biogeography*, *44*, 1512-1523.

Komárek, J. (2006). Cyanobacterial taxonomy: current problems and prospects for the integration of traditional and molecular approaches. *Algae*, *21*, 349-375.

Kristiansen, J. (1996). Dispersal of freshwater algae—a review. *Hydrobiologia*, *336*, 151-157.

Kumar, S., Spaulding, S. A., Stohlgren, T. J., Hermann, K. A., Schmidt, T. S., & Bahls, L. L. (2009). Potential habitat distribution for the freshwater diatom *Didymosphenia geminata* in the continental US. *Frontiers in Ecology and the Environment*, *7*, 415-420.

Lowe, R. L. (1979). Phytobenthic ecology and regulated streams. In Ward, J. (Ed.), *The Ecology of Regulated Streams* (pp. 25-34). Springer, Boston, MA.

Maher, B. A., Prospero, J. M., Mackie, D., Gaiero, D., Hesse, P. P., & Balkanski, Y. (2010). Global connections between aeolian dust, climate and ocean biogeochemistry at the present day and at the last glacial maximum. *Earth-Science Reviews, 99*, 61-97.

Maloney, K. O., & Weller, D. E. (2011). Anthropogenic disturbance and streams: land use and land-use change affect stream ecosystems via multiple pathways. *Freshwater Biology*, *56*, 611-626.

Marshall, W. A., & Chalmers, M. O. (1997). Airborne dispersal of Antarctic terrestrial algae and cyanobacteria. *Ecography*, *20,* 585-594.

McCormick, P. V., & Cairns, J. (1994). Algae as indicators of environmental change. *Journal of Applied Phycology*, *6*, 509-526.

Paerl, H. W. (2018). Mitigating toxic planktonic cyanobacterial blooms in aquatic ecosystems facing increasing anthropogenic and climatic pressures. *Toxins*, *10*, 76.

Pajunen, V., Luoto, M., & Soininen, J. (2016). Climate is an important driver for stream diatom distributions. *Global Ecology and Biogeography*, *25*, 198-206.

Patrick, R. (1973). Use of algae, especially diatoms, in the assessment of water quality. In Cairns, J., Dickson, K. L. (Eds.), *Biological Methods for the Assessment of Water Quality* (pp. 76-95). ASTM International, Philadelphia, PA.

Pearson, R. G. (2007). Species' distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History*, *50*, 54-89.

Rimet, F., & Bouchez, A. (2011). Use of diatom life-forms and ecological guilds to assess pesticide contamination in rivers: lotic mesocosm approaches. *Ecological Indicators*, *11*, 489-499.

Soininen, J. (2007). Environmental and spatial control of freshwater diatoms—a review. *Diatom Research*, *22*, 473-490.

Soininen, J. (2014). A quantitative analysis of species sorting across organisms and ecosystems. *Ecology*, *95*, 3284-3292.

Soininen, J., & Teittinen, A. (2019). Fifteen important questions in the spatial ecology of diatoms. *Freshwater Biology*, *64*, 2071-2083.

Stanish, L. F., Kohler, T. J., Esposito, R. M., Simmons, B. L., Nielsen, U. N., Wall, D. H., ... & McKnight, D. M. (2012). Extreme streams: flow intermittency as a control on diatom communities in meltwater streams in the McMurdo Dry Valleys, Antarctica. *Canadian Journal of Fisheries and Aquatic Sciences*, *69*, 1405-1419.

Stevenson, R. J., Bothwell, M. L., Lowe, R. L., & Thorp, J. H. (Eds.) (1996). *Algal Ecology: Freshwater Benthic Ecosystem*. Academic Press, San Diego, CA.

Van Dam, M. H., & Matzke, N. J. (2016). Evaluating the influence of connectivity and distance on biogeographical patterns in the south-western deserts of North America. *Journal of Biogeography*, *43*, 1514-1532.

Whitehead, P. G., Wilby, R. L., Battarbee, R. W., Kernan, M., & Wade, A. J. (2009). A review of the potential impacts of climate change on surface water quality. *Hydrological Sciences Journal*, *54*, 101-123.

Wood, S. A., Kelly, L. T., Bouma-Gregson, K., Humbert, J. F., Laughinghouse IV, H. D., Lazorchak, J., ... & Quiblier, C. (2020). Toxic benthic freshwater cyanobacterial proliferations: Challenges and solutions for enhancing knowledge and improving monitoring and mitigation. *Freshwater Biology*, *65*, 1824-1842.

CHAPTER II

BLOWIN’ IN THE WIND: TRANSPORT, STRUCTURE, AND REGIONAL CONNECTIVITY OF AEOLIAN DIATOM COMMUNITIES IN THE MCMURDO SOUND REGION, ANTARCTICA

Abstract

Diatom metacommunities are structured by environmental factors, historical factors, and space-related factors that are often attributed to organism dispersal. In the McMurdo Sound region (MSR) of Antarctica, wind connects aquatic habitats (glacier cryoconite holes, intermittent streams, and ice-covered lakes and ponds) through delivery of inorganic and organic matter. Our objective was to determine whether diatoms were dispersed in aeolian material and, if so, characterize the extent of dispersal and its relation to the regional diatom metacommunity. We collected 16 aeolian samples across 6 sites over 3 sampling seasons. We compared aeolian diatom composition with that of 111 samples from 20 regional waterbodies. Of the 50 known taxa in the MSR, 72% were recovered in aeolian material. The most dominant taxa were the small and regionally widespread *Humidophila arcuata* var. *parallela* and *Mayamaea atomus*. We measured between 0 and $8.76 * 10^6$ diatom valves g^{-1} aeolian material. Most diatom cells were dispersed during winter seasons that correspond to the highest wind speeds. Up to 15% of whole cells were potentially viable, indicating that up to $3.43 * 10^4$ potentially viable individuals could be dispersed in a year to a single 2 cm^2 site. Community composition varied among aeolian sites, wherein sites atop glaciers differed from terrestrial sites. Aeolian communities were distinct from all other regional communities but were most similar to glacier cryoconite and benthic stream mats. Individual aeolian communities were both potential sources and sinks for local and regional communities. However, aeolian communities explained a relatively low proportion of composition in sink communities, suggesting aeolian dispersal is less important in explaining metacommunity structure than intra-habitat dispersal and environmental filtering. The present

study confirms diatoms are dispersed by wind across a landscape characterized by aeolian processes, integrating the regional flora and contributing to regional metacommunity structure and landscape connectivity.

Introduction

Diatom biogeography and metacommunity structure are defined, in part, by organism dispersal (Martiny et al. 2006). Some research has suggested microorganisms are not dispersal limited and that metacommunity structure is attributable to environmental selection alone (Finlay 2002). Growing evidence demonstrates that some diatom metacommunities reflect strong spatial structuring in addition to environmental filtering, which is often attributed to taxon-specific variability in dispersal abilities across space and time (Vyverman et al. 2007). However, few direct observations of dispersal across habitats have been integrated into understanding diatom metacommunity assembly and structure at the landscape scale.

In ecosystems with many types of aquatic habitats, benthic diatom dispersal among habitats is potentially mediated by landscape connectivity via surface water movement, biota, and wind (Kristiansen 1996). The relative contribution of wind-mediated (aeolian) dispersal is likely highest among aquatic habitats in arid regions, where seasonal dry-downs and little overland flow result in limited hydrologic connectivity (Morán-Ordóñez et al. 2015). In arid ecosystems, benthic diatom communities can become dehydrated during dry periods that frequently correspond with strong winds, potentially facilitating aeolian dispersal of individuals and communities across the landscape (Souffreau et al. 2010). Characterizing the extent of passive aeolian dispersal of diatoms can therefore explain potential contributions of inter-habitat dispersal to regional metacommunity structure in arid landscapes.

In polar deserts, winds are the dominant year-round vector for landscape-scale transport of ions, water-soluble nutrients, and microbiota (Witherow et al. 2006, Šabacká et al. 2012, Gillies et al. 2013, Diaz et al. 2018). Aerophilic diatoms tolerate desiccation and can occur in terrestrial habitats with limited hydrologic vectors for dispersal (Johansen 1999). Therefore, the prevalence of aerophilic benthic diatoms in intermittent aquatic habitats suggests the possibility of high rates of aeolian dispersal (Esposito et al. 2008). Relatively few direct observations have been made of wind-mediated transport of benthic algae (Sharma et al. 2007, Tesson et al. 2016). Aeolian algae have been identified by traditional microscope techniques and, more recently, high-throughput sequencing (Sherwood et al. 2017). Airborne algal communities tend to be dominated by cyanobacteria in tropical regions and by green algae in temperate, Arctic, and Antarctic regions (Brown 1971, Broady 1979, Chu et al. 2013). In the maritime Antarctic, aeolian material collected 1 m above the ground contained the cosmopolitan, aerophilic benthic diatom *Pinnularia borealis*, though all cells appeared dead (Marshall and Chalmers 1997). In these studies, organism size does not appear to limit aeolian dispersal. However, higher rates of aerosolization and viability of smaller taxa might be expected given that long-distance aeolian transport is highest for particles < 20 µm diameter (Maher et al. 2010). Additionally, smaller individuals generally have higher tensile strength that could prevent fragmentation during dispersal (Flower 1993).

The McMurdo Sound region (MSR) of Antarctica is a useful system in which to characterize aeolian diatom dispersal. In the MSR, strong winds connect waterbodies that are intermittently hydrated and have limited hydrologic connectivity to other habitats. Additionally, the diatom flora is well described, no higher organisms than mosses and nematodes are present, and long-term hydrologic and ecological research records exist. Across the MSR, winds connect the volcanically active Ross Island; partially ice-covered McMurdo Sound and Ross Sea; and the largest ice-free region in Antarctica, the McMurdo Dry Valleys (MDVs). The MDVs is a network of glaciers, ephemeral meltwater streams, perennially ice-covered lakes and ponds, and soils (Fountain et al. 1999). During the summer, from November to February annually, surface water is consistently present at low discharge in streams formed from glacial meltwater and in moats around ice-covered lakes and ponds. Within a lake basin, streams deliver water and organic and inorganic material to the lake. Ponds can be run-of-river or isolated features. Aeolian material transported atop glaciers can form cryoconite holes: ~20 cm^2 cylindrical features with benthic cryoconite, seasonally thawed water, and perennial ice lenses (Fountain et al. 2004). In the MDVs, cryoconite holes cover around 1% of glacier surface area, are transitory habitats that exist for no more than 10 years, and contribute no more than 1 L meltwater m^{-2} to streamflow in a season (Mass 2018). Windblown dust is delivered to all aquatic habitats in the MSR, primarily within lake basins and during the winter (March to October) when winds are strongest (Nylen et al. 2004). Soil invertebrates, aquatic cyanobacteria, and diatoms have each been observed in MDV aeolian material and aeolian-derived glacier cryoconite (Nkem et al. 2006, Šabacká et al. 2012, Stanish et al. 2013, Diaz et al. 2018).

In the MSR, diatoms are most commonly found in benthic microbial mats dominated in biomass by cyanobacteria (Esposito et al. 2008). There are 50 known diatom taxa in the MSR diatom metacommunity across 19 genera. Most taxa are endemic to Antarctica, with only a few considered cosmopolitan (Kociolek et al. 2017). Diatom taxa in the MSR range in length from 5.5 µm to 110 µm. In the MSR, taxonomic community diversity has been studied extensively, with streams generally having substantially higher diatom taxon richness than cryoconite, pond mat, or lake mat communities (Stanish et al. 2011).

The processes structuring diatom community composition have been investigated for the various aquatic habitats in the MSR. Diatom community composition and diversity in MDV streams is most strongly associated with flow intermittency (Stanish et al. 2012), as well as with macronutrient availability (Kohler et al. 2016) and water temperature (Esposito et al. 2006, Darling et al. 2017). However, simulations on a stream diatom metacommunity in the MDVs suggest in-stream drift and aeolian dispersal are needed to explain the observed structure (Sokol et al. 2020). Diatom composition in glacier cryoconite holes in the MDVs is also related to environmental conditions and diatom physiological tolerances but is most closely associated with spatial factors such as distance to streams and lakes - implicating aeolian dispersal as a driver of cryoconite hole composition (Stanish et al. 2013). Across ponds in the MSR, diatom communities are structured similarly by environmental and spatial processes (Sakaeva et al. 2016). Altogether, diatom research in the MSR indicates that environmental and spatial factors each influence metacommunity structure.

Aeolian dispersal has been implicated as a likely vector for diatom metacommunity connectivity in the MSR and elsewhere. However, no studies have thoroughly examined aeolian material for diatoms or related aeolian community composition to the regional metacommunity structure of benthic diatoms. Among their geochemical observations, Diaz et al. (2018) identified three diatom cells under scanning electron microscopy from three different aeolian collectors, constituting the first recorded observation of diatoms in wind-transported material in the MSR.

The present work sought to characterize the extent of aeolian transport of diatoms in the MDVs and its implications for metacommunity structure in the MSR. We posed three questions: (1) how many diatoms are present in aeolian material, (2) which diatom taxa are represented in aeolian material and how do they vary over space and season, and (3) how do aeolian communities relate to those of regional aquatic habitats? To address each question, we made the following hypotheses: (H_1) diatom concentrations in aeolian material are lower in summer seasons when winds are weaker and benthic mats are hydrated; (H_{2a}) aeolian community composition varies over space and season because of spatial and temporal variability in taxon distributions and wind intensity and (H_{2b}) smaller taxa are most prevalent in aeolian material because of a higher likelihood of lift and resistance to fragmentation; and (H_3) aeolian diatom communities are most similar to those of local aquatic habitats because aeolian transport of inorganic material is largely restricted to lake basins.

Methods

Study location

The MSR includes Ross Island (2460 km^2 land area), the Ross Sea and McMurdo Sound (2700 km^2 ice and open water area), the MDVs (4800 km^2 mostly ice-free land area), and surrounding continental features. On Ross Island, the mostly ice-free coastal region of Cape Royds is characterized by several lakes and ephemeral ponds, including Pony Lake near an Adélie penguin colony, Coast Lake adjacent to the coast, and the more inland Blue Lake. Approximately 80 km of McMurdo Sound separates Cape Royds and the MDVs.

The MDVs comprise the largest mostly ice-free region of Antarctica. In the MDVs, mean annual air temperature ranges from -15ºC to -30ºC (Doran et al. 2002). Taylor Valley is centrally located relative to the other MDVs and extends from the Taylor Glacier in the west, an outlet glacier of the East Antarctic Ice Sheet, down-valley to McMurdo Sound in the east. Taylor Valley is characterized by three major lake basins, each with distinct biogeochemical compositions that influence the environmental characteristics of the aquatic habitats (Fountain et al. 1999). Lake Bonney is a closed-basin lake that drains Taylor Glacier and other western glaciers through waterbodies including (west to east) Lawson Stream, the Wormherder Creek Wetland, and Priscu Stream (Lyons et al. 1999). Lake Fryxell is a coastal closed-basin lake that drains Canada and Commonwealth Glaciers, among others, through streams including Canada Stream, Bowles Creek, Von Guerard Stream, Aiken Creek, and Commonwealth Stream. The closed-basin Lake Hoare and the Nussbaum Riegel (rock-bar) separate Lake Bonney and Lake Fryxell basins. Miers Valley is among the southernmost MDVs and extends from the Royal Society Range in the northwest down-valley to McMurdo Sound in the southeast. Miers Valley contains Lake Miers, an open lake that drains Adams and Miers Glaciers through Adams and Miers Streams, respectively. Throughout the text, waterbodies are referred to in order from west to east in Taylor Valley, Miers Valley, and Cape Royds.

Winds in the MDVs are characterized as down-valley (from the polar plateau towards the coast), coastal (from the coast up-valley), and drainage (down valley sides) (Nylen et al. 2004). Wind speeds average between 2.5 - 5 m s^{-1} throughout the year, with down-valley winds dominant in the winter and up-valley winds dominant in the summer. Winter foehn winds originate from the polar plateau and move down-valley at speeds as low as 5 m s^{-1} and exceeding 30 m s^{-1} (Speirs et al. 2010). Thermally generated coastal winds in the summer can reach speeds up to 26 m s^{-1}, with reduced speeds further up-valley (Speirs et al. 2010). Down-valley foehn winds in the winter and high speed thermally generated up-valley winds in the summer are responsible for the majority of aeolian transport (Gillies et al. 2013). Wind speeds of 10 m s^{-1} are capable of moving sand-sized particles < 500 μm diameter in saltation at 4 m above the ground (Gillies et al. 2013), and speeds around 6 m s^{-1} are strong enough to resuspend combustion particles greater than 0.75 μm diameter deposited on the valley floor (Khan et al. 2018). Aeolian material is primarily transported within lake basins in the source valley (Šabacká et al. 2012, Deuerling et al. 2014), though there is some evidence for inter-basin transport at higher elevations and collection heights from the surface (Diaz et al. 2018).

Sampling

Coarse- and fine-grained aeolian material (< 2 mm diameter) was collected from 6 Big Spring Number Eight isokinetic passive wind samplers approximately 50 cm above the surface (Figure S1): 5 locations along a longitudinal transect through Taylor Valley and one location in Miers Valley (Figure 1). Of the collection sites, two were atop glaciers: Taylor Glacier (TG-G_B) and Commonwealth Glacier (CG-G_F), a coastal valley glacier in Taylor Valley. The remaining sites were terrestrial and located < 1 km from lakes, streams, or ponds for which they are named: in Taylor Valley, Lake Bonney (LB-T_B), Von Guerard Stream (VG-T_F), and Explorer's Cove (EC-T_F); and in Miers Valley, Miers Stream (MS-T_M). In sample abbreviations, the letter following the underscore represents the lake basin in which the collector was deployed: Lake Bonney (B), Lake Fryxell (F), and Lake Miers (M). A total of 16 samples were collected across three seasons: November 2013 - January 2014 (summer only), January 2014 - November 2014 (winter only), and January 2018 - January 2019 (full year). For each aeolian sampling location, the following wind characteristics were calculated from data collected from the closest meteorological station over the season during which the collector was deployed: mean daily wind speed, mean daily maximum wind speed, and mean daily wind direction.

Raw aeolian material samples were weighed at the McMurdo Station Crary Laboratory (wet mass, though only two samples contained a substantial amount of ice), shipped frozen, and stored at -20ºC until further analysis. At the University of Colorado Boulder, samples were manually homogenized and subsamples were taken, weighed, dried at 50ºC for at least 24 hours, and weighed again (dry mass, DM). Ash-free dry mass (AFDM) was calculated as the difference

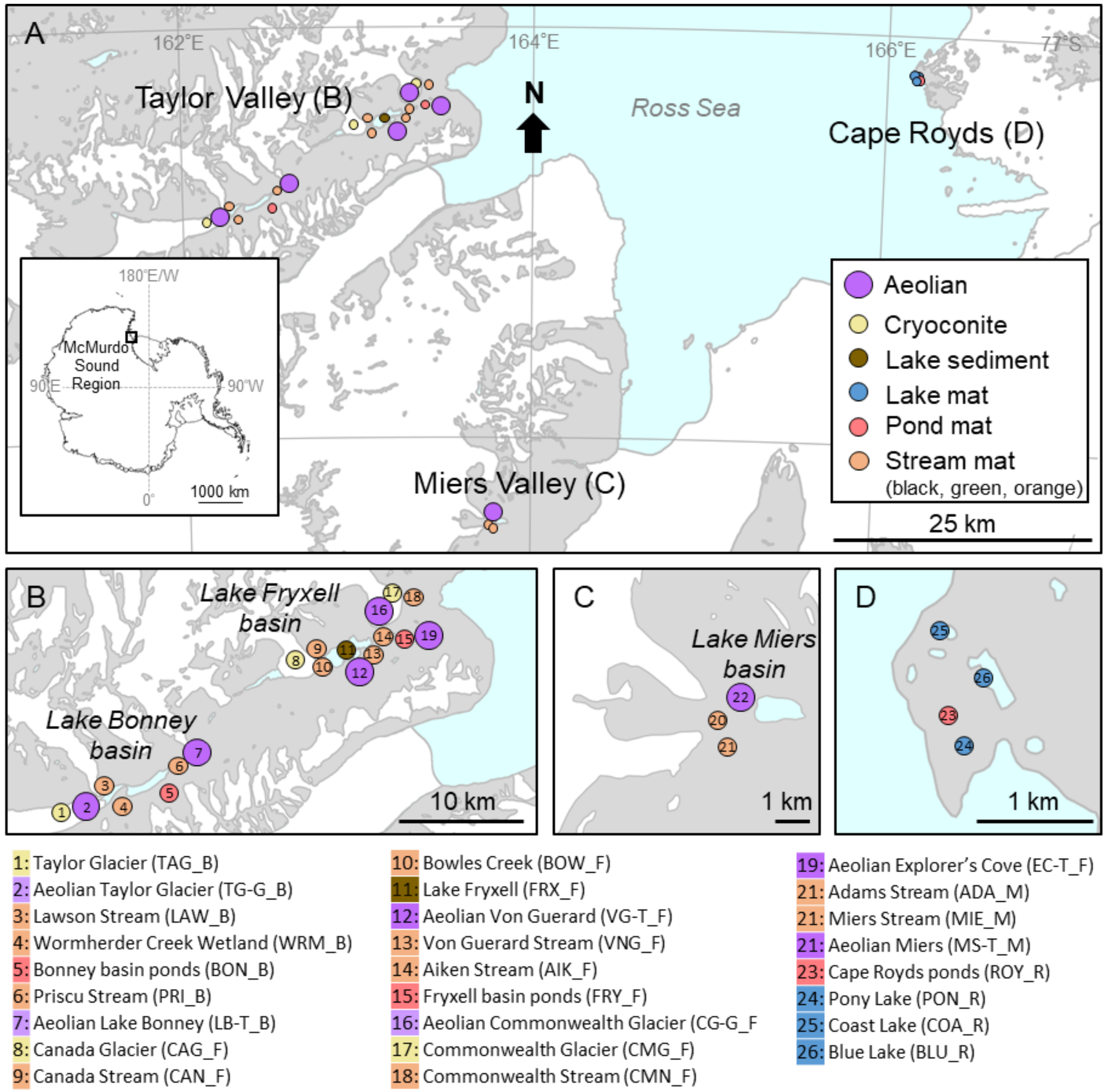


Figure 1. Map of aeolian and regional community locations in the study area, colored by habitat type. (A) McMurdo Sound region (inset: Antarctic continent) and detail of (B) Taylor Valley, (C) Miers Valley, and (D) Cape Royds.

between DM and the subsample mass after organic material was combusted in a muffle furnace at 550ºC. Whole sample DM was estimated by applying a conversion factor to the whole sample wet mass. Meteorological and aeolian material data are available at the McMurdo Dry Valleys Long-Term Ecological Research project (MCM LTER) website (https://mcm.lternet.edu/).

Separate subsamples from those for DM were taken for diatom analysis. Each subsample was passed through a 200 μm^2 pore size sieve using distilled water to remove large particles. The largest diatom in the MSR, *Hantzschia elongata*, does not exceed 175 μm length or 12 μm mid-breadth, so sieving did not bias against large taxa. Each sievate was centrifuged at 20,000 *g* and the supernatant removed. Each pellet was resuspended in distilled water to a known volume, and a known volume subsample was transferred to a cover slip and dried. After drying, the cover slip was mounted onto a glass microscope slide using the mounting medium Zrax (MicrAP Enterprises). At no point during the preparation were subsamples subjected to acidification or extreme heat so as not to destroy protoplasmic material. For each slide, up to 100 diatom valves were counted and identified to the lowest taxonomic resolution along a known area using differential interference contrast and a 100x oil immersion objective lens and condenser lens. All fragments greater than approximately 10% of a valve were counted, and both whole cells (two attached valves) and whole cells with identifiable protoplasm were enumerated. Valve fragments were counted in order to assess the maximum taxonomic diversity of a sample, given that whole cells could have been fragmented by the corrosive destruction of diatom valves by sediment inside the aeolian collectors. Diatoms per gram aeolian material DM (all valves and whole cells only) was calculated as the number of diatoms per subsample traversed while counting, converted and scaled to whole sample DM. Slides and cleaned material were archived in the INSTAAR Diatom Collection, University of Colorado. All scientific names and authorship follow the Antarctic Freshwater Diatoms website, which is where all count data are archived (http://huey.colorado.edu/diatoms/about/index.php).

Diatom taxa were assigned to size classes (sc) based on mean lengths, calculated from taxon-specific minimum and maximum lengths on the Antarctic Freshwater Diatoms website. Four size classes were established based on quartiles calculated from the mean lengths of all taxa found in the MSR: < 16 μm (sc1), 16 – 24.24 μm (sc2), 24.25 – 34 μm (sc3), and > 34 μm (sc4).

Aeolian diatom community compositions were compared to those of 7 regional diatom habitat types found in the MDVs: glacier cryoconite holes, lake sediment, lake algal mats, pond algal mats, and stream algal mats (separated by mat macroscopic color: black, green, and orange). Each sample was processed similarly to that of aeolian samples, except formalin-preserved subsamples were acidified to remove organic matter prior to counting. Additionally, up to 300 valves, including fragments, were counted for each sample, and neither whole cells nor potentially viable cells were recorded.

We compared aeolian communities to 111 regional samples, collected between 1992-2015 as part of the MCM LTER. Samples from lake and stream algal mats were chosen from collection year 2007 – 2015 and only from waterbodies in Taylor Valley, Miers Valley, and Cape Royds with at least 5 samples counted for diatoms. A random subsample of 5 samples for each waterbody and habitat type was used for analysis. Five randomized lake sediment samples were chosen from Lake Fryxell sediment cores collected in 1992. Fewer than 5 cryoconite samples from each glacier were available, so all glacier cryoconite samples were retained. Fewer than 5 samples were available for each pond. To reduce the number of distinct waterbodies used in the analysis, 5 samples were randomly selected from all ponds in each Lake Bonney basin, Lake Fryxell basin, and Cape Royds. Pond samples were therefore aggregated into 3 groups, with each assessed as a single waterbody: Lake Bonney ponds, Lake Fryxell ponds, and Cape Royds ponds. In total, 20 waterbodies were used in analyses, including the 3 pond groups. Three streams contained multiple habitat types because of the presence of multiple stream mat colors: Bowles Creek, Canada Stream, and Von Guerard Stream. Therefore, a total of 25 regional communities (waterbody and habitat type combinations) were compared to aeolian communities.

Counts for *Luticola muticopsis* were combined with those for *L. austroatlantica*, and *L. spainiae* with *L. laeta* because regional count data were not harmonized to the updated taxonomies presented in Kohler et al. (2015). *Luticola spainiae* but not *L. muticopsis* were found in aeolian samples. The aeolian collectors excluded microbial mat material larger than 2 mm^2, biasing our observations towards solitary individuals.

In the present study, we use the term "aeolian community" to represent the diatoms found co-occurrent in an aeolian sample, recognizing these individuals do not represent an active community of interacting organisms.

Data analysis

To determine the variability of diatom concentrations in aeolian material over space and time (H_1), diatom valve and whole cell concentrations were calculated for each sample site and sampling season. One-way analyses of variance (ANOVAs) were performed to determine if concentrations differed significantly by season or site. Simple linear regressions were performed between diatom concentrations and wind variables.

To assess aeolian diatom community composition (H_2), taxon richness and evenness were compared across space, sampling season, and valve size classes using ANOVAs and Tukey's post-hoc tests. Differences in whole community composition across aeolian samples were assessed using non-metric multidimensional scaling (NMDS) ordination, permutational analysis of variance (PERMANOVA), and tests of homogeneity of multivariate dispersion (PERMDISP) as a proxy for beta diversity on untransformed Bray-Curtis dissimilarities by sampling location, season, longitude, distance to coast, lake basin, elevation, and wind variables. Taxon-specific variations among sites were analyzed by calculating similarity percentages (SIMPER) of Bray-

Curtis dissimilarities followed by Kruskal-Wallis and Dunn post-hoc tests of taxon relative abundances against sites to quantify inter-site variability.

The connectivity of aeolian communities to regional habitat communities (H_3) was assessed using NMDS ordination and PERMANOVA by sampling location, season, and community (defined as waterbody [or aeolian collector] and habitat type – e.g., Von Guerard Stream orange mat). SIMPER, Kruskal-Wallis and post-hoc tests, and indicator species analysis were performed to determine taxon-specific differences among communities. To predict potential sources and sinks between regional and aeolian communities, the Bayesian algorithm SourceTracker (v. 0.9.8) was applied to the regional metacommunity count data (Knights et al. 2011). Briefly, SourceTracker applies a Bayesian probability model to determine the proportions of communities derived from source habitats. Gibbs sampling is used to estimate the uncertainty with the model output. Here, a community acts as a source when individuals emigrate from the community and as a sink when individuals immigrate into the community.

Aeolian communities are comprised of passively dispersing taxa and can represent both temporary source (taxa passively emigrating from) and sink (taxa passively immigrating to) communities. Therefore, SourceTracker was run with aeolian communities as potential sources to and potential sinks for regional communities. SourceTracker analyses were performed to infer community-specific connectivity between aeolian and regional samples. We used SourceTracker to explore the degree to which aeolian and regional sink community compositions were explained by potential sources at multiple spatial scales and potential dispersal abilities of whole communities from waterbodies. For aeolian sinks (receptors of individuals from regional communities), we considered (i) all regional sources except those from different valleys (i.e., Miers Valley sources were excluded for Taylor Valley sinks and vice versa) and (ii) only regional sources from the same lake basin as the aeolian collector. For regional sinks (receptors of individuals from aeolian and other regional communities), we considered (i) all aeolian and regional sources except those from different valleys, (ii) all same-valley aeolian sources and only regional sources from the same lake basin, and (iii) all same-valley aeolian sources and only regional sources from a hydrologically connected waterbody (e.g., orange mats as a source for green mats in the same stream, stream mats as a source for its drainage lake sediment). The mean proportion of sink community composition explained by each source was calculated for each waterbody and habitat type combination. A leave-one-out cross-validation analysis of source communities was performed to assess the ability to discriminate between potential sources.

All analyses were performed in R (R Core Team 2020): ‘vegan’ for community composition and diversity statistics (Oksanen et al. 2019), ‘labdsv’ for indicator species analysis (Roberts 2019), and ‘Sourcetracker.r’ for SourceTracker (Knights et al. 2011, https://github.com/danknights/sourcetracker/).

Results

Bulk aeolian transport of diatoms (H_1)

Across sites, wind direction was predominantly northwesterly during the summer sampling season and southeasterly during the winter and full-year sampling seasons (Figure S2-*A*). Mean daily and mean daily maximum wind speeds were highest in the summer for all terrestrial sites and higher in the winter and full year for both sites atop glaciers (Figure S2-*B*-*C*). However, the highest mean daily and mean daily maximum wind speeds for each site occurred during the winter or full-year sampling seasons. Wind variables substantially covaried with site (ANOVA $p = 0.045$), basin (ANOVA $p = 0.076$), and longitude (simple linear regression $R^2 = 0.48$ and $p = 0.004$).

Across all aeolian collectors, bulk aeolian material ranged between 0.036 and 367.61 g DM (Table 1). Diatom concentration in aeolian material varied between 0 valves g^{-1} DM and $6.89 * 10^5$ valves g^{-1} DM, which translated to up to $1.29 * 10^6$ diatom valves at a site in a single sampling season. In samples with at least 50 valves counted, whole cells comprised between 51 – 85% of valves $sample^{-1}$, of which between 0 – 31% contained visible protoplasm (Table 1).

Neither aeolian material DM nor valves g^{-1} DM was significantly explained by site or sampling season (ANOVA $p = 0.38$). However, Miers Valley aeolian samples (MS-T_M) had substantially higher sample DM than other sites, and winter samples consistently had higher valves g^{-1} DM across sites than other seasons (Table 1). In the one Commonwealth Glacier aeolian sample (CG-G_F) for which it could be measured, AFDM was much higher than other sites, which corresponded with the highest diatom valve concentration of any sample. Wind variables did not significantly explain variation in aeolian material AFDM or diatom cell concentration (simple linear regression $p > 0.12$).

Aeolian diatom community characteristics (H_2)

Thirty-six of the 50 known taxa in the MSR (72%) were found across all aeolian collectors combined (Figure 2). Of these, *Luticola austroatlantica* (size class 2 [sc2]), *Mayamaea atomus* (sc1), and *Humidophila arcuata* var. *parallela* (sc1) were the most common (present in 94%, 94%, and 88% of samples, respectively) and most abundant taxa (mean relative abundance across all samples of 7%, 30%, and 13%, respectively). Taxa comprising the smallest size class were the most common (present in all samples) and abundant (39.5% mean relative abundance) across all samples, followed in frequency and relative abundance by consecutively larger size classes (Figure 2).

Table 1. Aeolian sites, geographic information, wind-related variables, and aeolian material measurements by sampling season (n = 1 for each site by season).

Site	Latitude	Longitude	Distance to coast (km)	Lake basin	Elevation (m)	Season	Mean daily wind speed (m s^{-1})	Sample DM (g)	Sample AFDM (g)	AFDM g^{-1} DM	Valves sample^{-1} DM	Valves g^{-1} DM	Whole cells valves^{-1} (%)	Viable cells whole cells^{-1} (%)
TG-G_B	-77.74000	162.13135	38.5	Bonney	334	Summer	4.2 ± 1.7	0.04	na	na	0	0	na	na
						Winter	5.6 ± 2.8	5.28	0.20	0.038	13,041	2,470	33	100
LB-T_B	-77.69260	162.56233	27.0	Bonney	64	Summer	4.9 ± 1.6	0.78	0.12	na	1,036	1,337	62	0
						Winter	4.4 ± 0.9	40.68	0.65	0.016	1,209,191	29,726	57	4
						Full year	3.8 ± 3.2	130.96	0.73	0.006	90,852	694	98	17
VG-T_F	-77.60850	163.24870	8.5	Fryxell	19	Summer	4.8 ± 1.7	6.83	0.24	0.035	119,358	17,486	85	0
						Winter	3.7 ± 2.4	15.16	0.34	0.022	271,337	17,895	79	16
						Full year	4.0 ± 4.0	15.16	0.34	0.022	87,904	6,655	59	0
CG-G_F	-77.56460	163.28232	5.5	Fryxell	286	Winter	4.1 ± 1.0	0.64	0.10	0.157	440,080	688,965	54	0
						Full year	2.8 ± 2.1	0.09	na	na	8,528	91,679	51	0
EC-T_F	-77.58870	163.41752	3.5	Fryxell	24	Summer	2.8 ± 2.3	1.44	0.13	0.087	17,707	12,280	76	0
						Winter	4.0 ± 1.5	38.70	0.27	0.007	2,248,517	58,103	62	3
						Full year	4.4 ± 1.9	21.11	0.37	0.017	87,264	4,134	69	0
MS-T_M	-78.09810	163.79428	10.5	Miers	50	Summer	4.0 ± 1.4	36.09	0.47	0.013	135,269	3,748	83	2
						Winter	3.3 ± 2.6	211.82	1.29	0.006	8,764,129	41,375	79	0
						Full year	3.0 ± 2.3	367.61	2.72	0.007	1,290,301	3,510	68	15

Sites were located atop either a glacier (-G) or terrestrial (-T) surface
Years for each season: summer 2013-14, winter 2014, full year 2018-19
DM = dry mass, AFDM = ash-free dry mass
Viable cells were determined as cells containing visible protoplasm in light microscopy

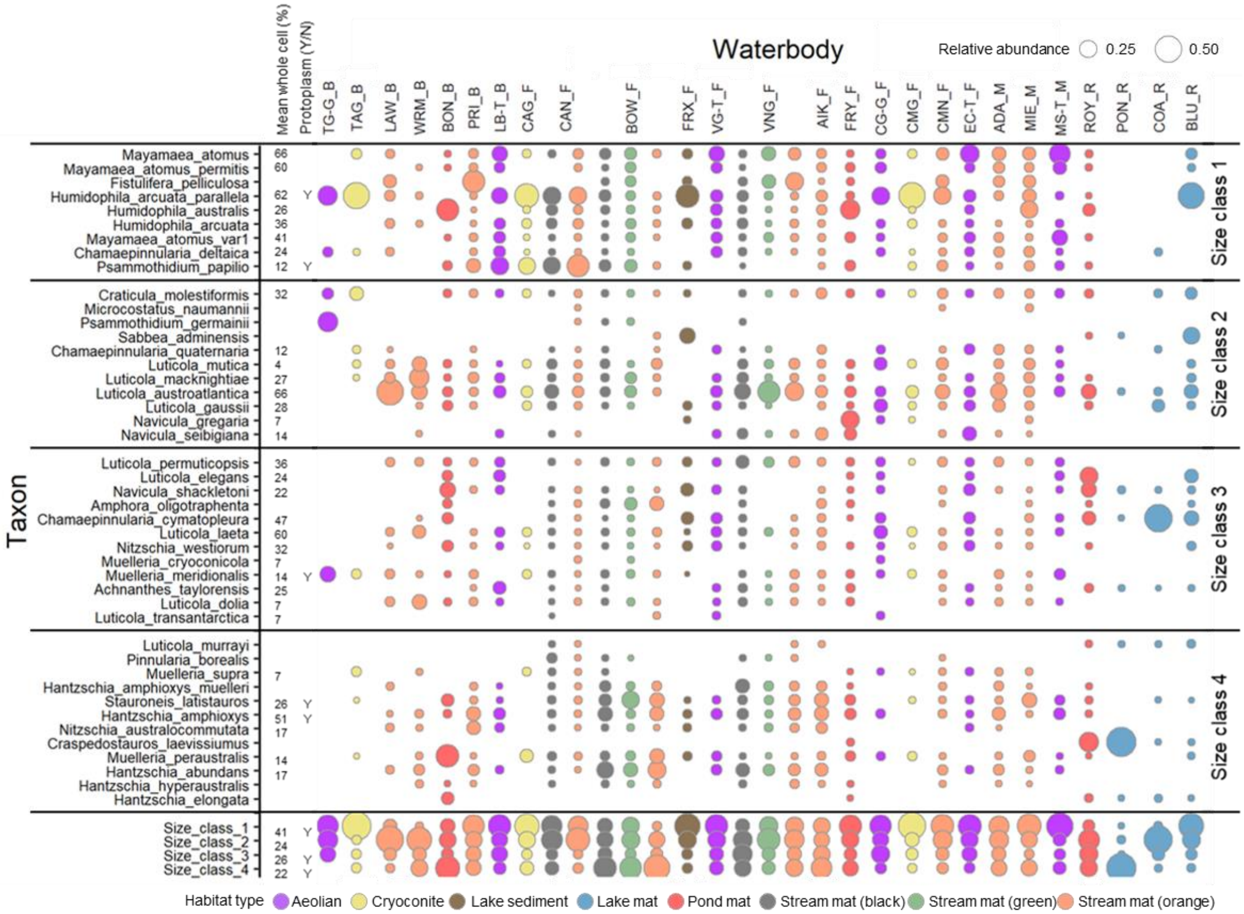


Figure 2. Mean taxon relative abundances across communities in the McMurdo Sound region. Dot sizes are proportional to relative abundances of each taxon for each community. Size class relative abundances were calculated by aggregating taxon counts within each size class.

Whole cells and whole cells with protoplasm (i.e., potentially viable cells) were detected in taxa across size classes (Figure 2). For several taxa, over half of all valves recorded across aeolian samples were whole cells: from smallest to largest mean length, *M. atomus*, *M. atomus* var. *permitis* (sc1, a more lightly silicified variety of *M. atomus*), *H. arcuata* var. *parallela*, *L. austroatlantica*, *L. laeta* (sc3), and *Hantzschia amphioxys* (sc4, a heavily silicified nitzschoid species). Only 5 taxa were identified with protoplasm: the small taxa *H. arcuata* var. *parallela* and *Psammothidium papilio* (sc1) as well as larger taxa *Muelleria meridionalis* (sc3), *Stauroneis latistauros* (sc3), and *H. amphioxys*.

Some general patterns in aeolian community composition were detectable in NMDS ordination space. Communities from the same site consistently clustered together, with high overlap between the Von Guerard (VG-T_F) and Explorer's Cove (EC-T_F) aeolian communities that was strongest between communities of the same season (Figure 3). The CG-G_F and MS-T_M communities each clustered closely to communities from the same site, while there was a larger spread in composition of the Lake Bonney (LB-T_B), VG-T_F, and EC-T_F communities. In general, community composition was more similar among sites over time than among sampling seasons across sites.

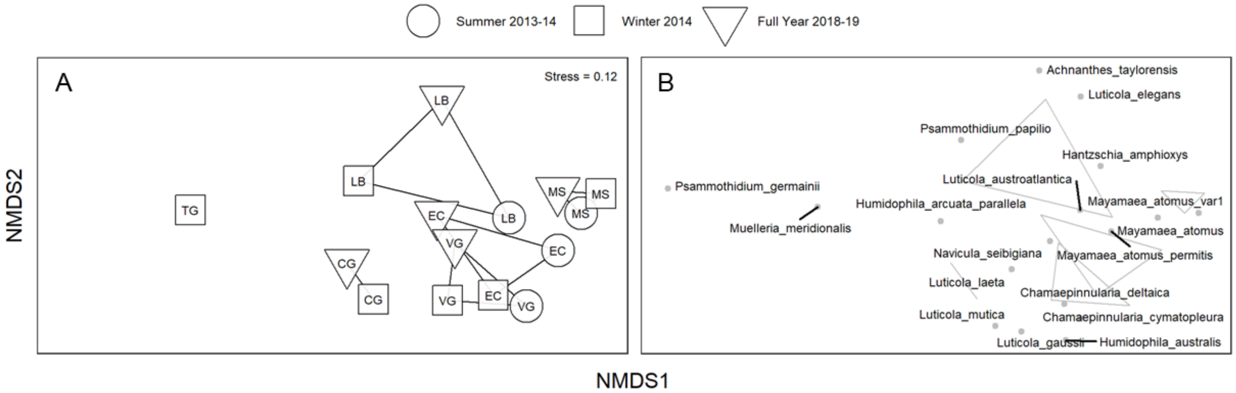


Figure 3. Non-metric multidimensional scaling (NMDS) ordination of 15 aeolian communities across three sampling seasons. (A) Centroids of communities and (B) centroids of diatom taxa with > 10% relative abundance in at least one sample. Hulls connect samples from the same aeolian collector, each of which is represented by a shortened abbreviation from that in Table 1.

Chamaepinnularia deltaica (sc1) and *H. arcuata* var. *parallela* were the only taxa found at all sites. Four taxa varied significantly in relative abundance across sites (Kruskal-Wallis and Dunn tests, $p < 0.05$): *H. arcuata* var. *parallela* was lower in MS-T_M than in Taylor Glacier (TG-G_B), LB-T_B, and CG-G_F communities; *Luticola gaussi* (sc2) was higher in CG-G_F than in TG-G_B, LB-T_B, and MS-T_M; *M. atomus* was lower in TG-G_B than in EC-T_F and MS-T_M and lower in CG-G_F than in EC-T_F and MS-T_M; and *Psammothidium germainii* (sc2) was only found in TG-G_B. There were significant indicator taxa for three sites: *Craticula* *molestiformis* (sc2) for TG-G_B, *L. gaussii* for CG-G_F, and *M. atomus* and *M. atomus* var. 1 (sc1) for MS-T_M.

Differences in the abundances of *H. arcuata* var. *parallela* and *M. atomus* were commonly among the highest contributors to pairwise site differences in taxon relative abundances, with mean contributions of 15% each (Table S1). In general, TG-G_B differed from other sites as the only community containing *P. germainii*, a high relative abundance of *M. meridionalis*, and low taxon richness (S = 5) (Figure S3). The LB-T_B communities differed from other aeolian communities in having high relative abundances of *P. papilio* (of which only EC-T_F contained small numbers). The VG-T_F and EC-T_F communities were the most similar of any other pair of sites, the only communities containing *Humidophila australis* (sc1), and characterized by high evenness and taxon richness (Figure S3). The CG-G_F communities differed from other sites by containing comparably high relative abundances of *H. arcuata* var. *parallela*, *L. gaussii*, and *L. laeta/spainiae* and low relative abundances of *M. atomus*. The MS-T_M communities were characterized by high relative abundances of all three *Mayamaea* taxa, contrasted with other communities that had high relative abundances of *M. atomus* only.

After controlling for each other, site and season significantly explained the variability in aeolian community composition (PERMANOVA by site: $R^2 = 0.63$, $p = 0.001$; by season: $R^2 = 0.13$, $p = 0.03$). Beta diversity (heterogeneity of multivariate dispersions) was lower at each site across seasons than among sites within a single season (PERMDISP ANOVA by site: $F = 15.55$, $p = 0.0003$; by season: $F = 4.59$, $p = 0.03$). Aeolian taxon richness was higher in Lake Fryxell basin aeolian communities than in those of Lake Bonney basin or Miers Valley (Tukey's post-hoc test, $p < 0.014$). Evenness was lower in Miers Valley communities than in Lake Bonney and Lake Fryxell basins ($p < 0.034$).

When controlling for covariation between wind and geography, wind variables did not explain a substantial amount of variation in aeolian community composition or diversity. When controlling for site, no other geographic variables significantly explained community variation.

Aeolian diatom community connectivity to regional diatom metacommunity (H_3)

In general, aeolian diatom communities were more similar to each other than to regional communities, clustering along NMDS axis 1 (Figure 4, Table 2). Waterbody, habitat type, season, and lake basin each significantly explained variation in metacommunity composition, but waterbody had an R^2 substantially greater than other factors (PERMANOVA by waterbody: R^2 = 0.73, $p = 0.001$). Individually, aeolian samples from atop glaciers (TG-G_B and CG-G_F) were most similar to local glacier cryoconite (Table 2). Terrestrial aeolian samples from both Taylor and Miers Valleys were less similar to local waterbody communities than to communities from Adams and Miers Streams in Miers Valley (each of which had high abundances of *M. atomus*), Commonwealth Stream, and glacier cryoconite (Table 2).

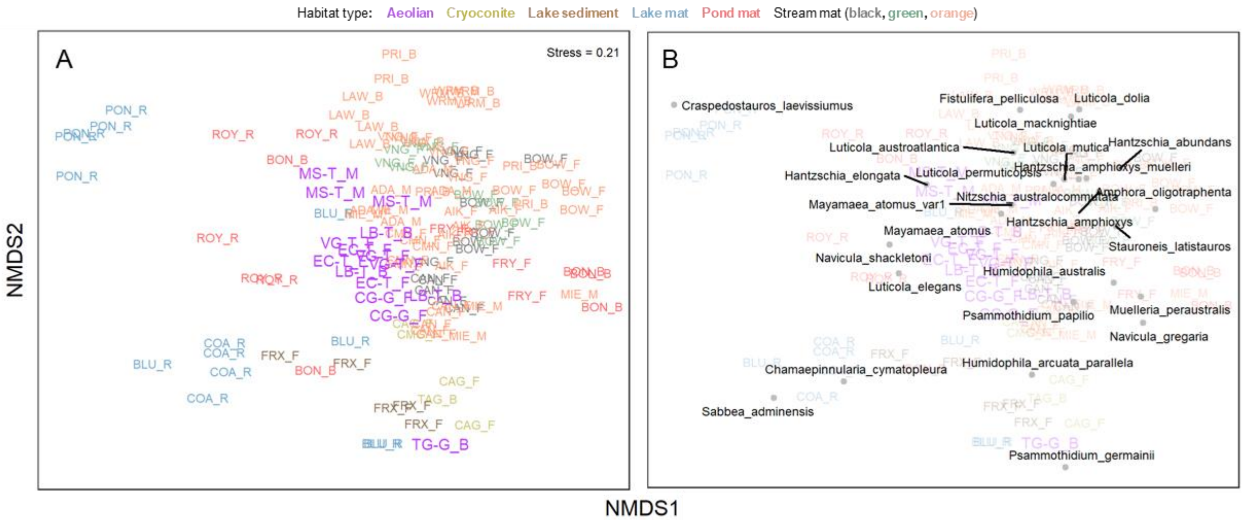


Figure 4. Non-metric multidimensional scaling (NMDS) ordination of 126 diatom communities from 32 waterbodies and aeolian collectors in the McMurdo Sound region. (A) Centroids of communities colored by habitat type and (B) centroids of diatom taxa with relative abundance > 20% in at least one sample.

Table 2. Ten lowest mean Bray-Curtis dissimilarities for pairwise comparisons of whole community composition between aeolian and regional communities. Lower mean dissimilarities for each aeolian site have darker backgrounds.

Longitude

	TG-G_B	LB-T_B	VG-T_F	CG-G_F	EC-T_F	MS-T_M
TG-G_B_aeolian	na	0.79	0.88	0.68	0.86	0.95
LB-T_B_aeolian	0.79	0.62	0.58	0.61	0.57	0.61
VG-T_F_aeolian	0.88	0.58	0.44	0.56	0.45	0.60
CG-G_F_aeolian	0.68	0.61	0.56	0.40	0.60	0.77
EC-T_F_aeolian	0.86	0.57	0.45	0.60	0.48	0.52
MS-T_M_aeolian	0.95	0.61	0.60	0.77	0.52	0.30
TAG_B_cryoconite	0.56	0.75	0.85	0.65	0.83	0.90
CAG_F_cryoconite	0.65	0.72	0.81	0.60	0.82	0.94
CAN_F_black	0.72	0.61	0.70	0.59	0.73	0.85
CAN_F_orange	0.75	0.62	0.73	0.63	0.74	0.85
FRX_F_sediment	0.73	0.79	0.78	0.66	0.79	0.95
VNG_F_green	0.97	0.74	0.71	0.82	0.71	0.72
VNG_F_orange	0.97	0.75	0.72	0.82	0.77	0.76
AIK_F_orange	0.88	0.72	0.64	0.72	0.68	0.75
CMG_F_cryoconite	0.64	0.68	0.72	0.54	0.75	0.86
FRY_F_pond	0.96	0.79	0.71	0.80	0.77	0.75
CMN_F_orange	0.69	0.58	0.62	0.52	0.63	0.72
ADA_M_orange	0.88	0.68	0.61	0.67	0.64	0.64
MIE_M_orange	0.83	0.75	0.69	0.69	0.70	0.78
BLU_R_lake	0.78	0.80	0.79	0.71	0.78	0.92

Longitude

Black text = 10 lowest pairwise dissimilarities for each aeolian site

Generally, aeolian communities contained higher abundances of *M. atomus* (and subspecies) and *H. arcuata* var. *parallela* than did regional communities (Figure 2). Additionally, lower abundances of *Hantzschia* spp., *Fistulifera pelliculosa* (sc1), *L. austroatlantica*, *P. papilio*, and habitat-dominant taxa (e.g., *Craspedostauros laevissimus* [sc4] in Pony Lake) differentiated aeolian communities from other regional communities. Notably, glacier cryoconite had higher abundances of *H. arcuata* var. *parallela* and lower abundances of *L. austroatlantica* and *Hantzschia* spp. than did aeolian communities. Miers and Adams Streams had the highest abundances of *M. atomus* across all regional communities but had lower abundances of *Mayamaea* taxa than aeolian communities. Individual aeolian sites did not have significant indicator taxa, but *M. atomus* (indicator value = 0.48), *C. deltaica* (0.45), *M. atomus* f. *permitis* (0.35), *M. atomus* var. 1 (0.32), and *L. laeta* (0.32) were indicators for aeolian communities altogether.

Aeolian communities did not differ significantly from same-valley regional communities in the relative abundance of taxa from the smallest size class (Tukey's post-hoc test, $p > 0.05$) (Figure 2). However, overall size structure of each aeolian community differed substantially from almost all regional communities but not from other aeolian communities (pairwise PERMANOVA, $p < 0.05$, Table S2). Individually, all Taylor Valley aeolian communities were most similar to Commonwealth Stream orange mats in size class structure (mean Bray-Curtis dissimilarity < 0.27) (Table S2). The MS-T_M communities in Miers Valley were most similar in size class structure to Commonwealth Glacier cryoconite (mean Bray-Curtis dissimilarity = 0.11).

Notable taxa absent from aeolian communities included *Craspedostauros laevissimus* (very abundant in Cape Royds lake and pond mats), *Fistulifera pelliculosa* (very lightly silicified and moderately abundant in Lake Fryxell basin black, green, and orange stream mats), *Hantzschia elongata* (sc4, largest diatom in the MSR, moderately abundant in pond mats), and *Sabbea adminensis* (sc2, very abundant in lake mats and lake sediment cores).

In general, mean aeolian taxon richness was lower than that of same-basin stream and pond mats and higher than same-basin glacier cryoconite, Lake Fryxell sediment, and Cape Royds lakes (Figure S3). However, TG-G_B had the lowest richness of any site. Mean aeolian community evenness was similar to that of stream mats and higher than that of glacier cryoconite and Cape Royds lake and pond mats (Figure S3).

In leave-one-out cross-validation analyses, most aeolian and regional communities were not strongly reliable as sources (< 50% of source composition explained by same-community validation samples; Supplementary Results, Figures S4 and S5). However, most communities were moderately reliable as sources in that the highest proportion explained by a single community was by the same-community validation samples.

Bearing in mind the limitations of SourceTracker analyses in this study, aeolian community composition was explained more by all potential regional community sources than by only same-basin sources for Lake Bonney basin communities but not Lake Fryxell basin or Miers Valley communities (Figure 5). Lake Bonney basin aeolian communities (TG-G_B and LB-T_B) were evenly explained by a variety of regional communities when considering the entire MSR, whereas roughly half that explained by all MSR sources was explained by same-basin orange stream mats. The contribution of sources to Lake Fryxell basin aeolian community composition was spread evenly among regional and same-basin waterbodies and habitat types. The MS-T_M aeolian community was strongly explained by the nearby Adams Stream orange mats. Over 50% of aeolian community variation was explained for Lake Bonney basin communities from all potential regional sources and for Lake Fryxell basin communities from all potential regional sources and same-basin sources. For all aeolian communities a substantial amount of community composition was unexplained by both all MSR regional community sources (20 – 70%) and same-basin sources (35 – 73%). High unexplained proportions likely result from the limitations in source reliability in addition to uncharacterized ecological processes.

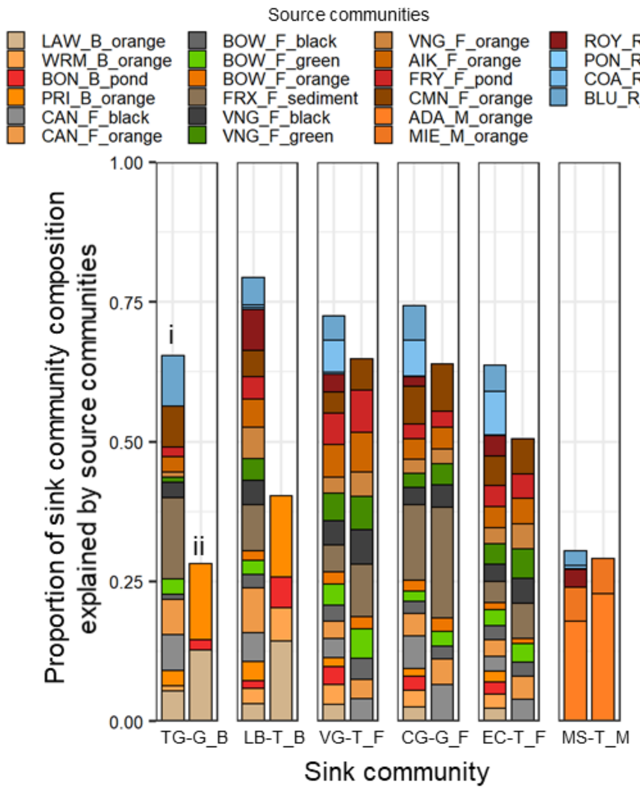


Figure 5. Proportions of aeolian sink community composition explained by potential source communities across the McMurdo Sound region using the SourceTracker algorithm. For each sink community panel, two bars are present that each represent different possible source communities. From left to right: (i) all possible regional sources except those from different valleys and (ii) all same-basin regional sources.

Regional sink community composition varied substantially in the relative contributions of aeolian, same-basin, and hydrologically connected sources (Figure 6). In general, aeolian communities explained a reasonable amount of composition in Taylor and Commonwealth Glacier cryoconite, Adams Stream in Miers Valley, Lake Fryxell lake sediment, and Lake Fryxell basin stream mats closest to the coast. For most sink communities, there was little difference between the proportion explained by all regional aeolian communities and same-basin aeolian communities. Other than for Taylor Glacier cryoconite and Adams Stream, no single aeolian community accounted for more than 10% of sink composition. However, cumulatively, aeolian communities often explained more than 10% of total sink composition.

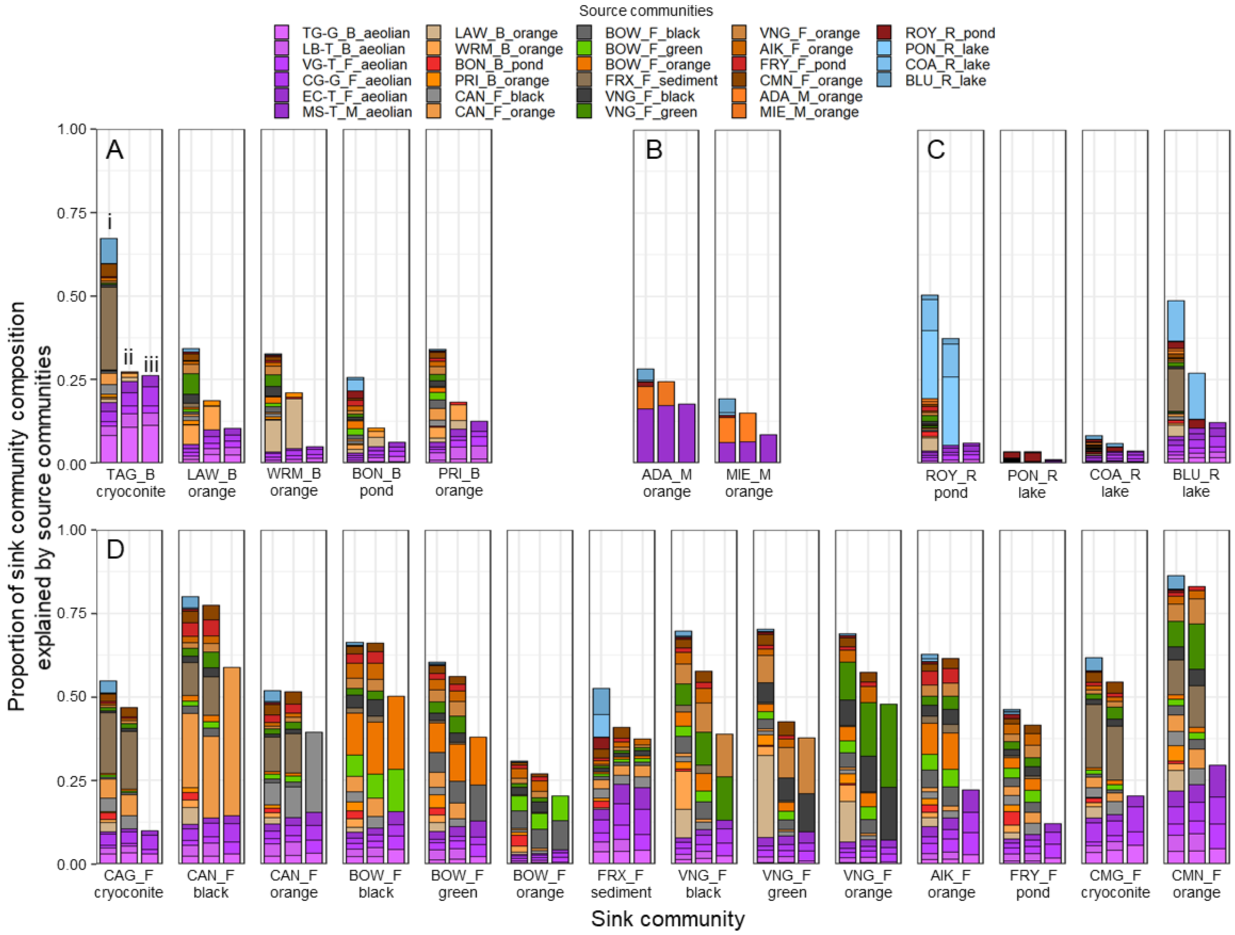


Figure 6. Proportions of regional sink community composition explained by potential source communities across the McMurdo Sound region using the SourceTracker algorithm. For each sink community panel, three bars are present that each represent different possible source communities. From left to right: (i) all possible aeolian and regional sources except those from different valleys, (ii) all same-valley aeolian and same-basin regional sources, and (iii) all same-valley aeolian and hydrologically connected regional sources. Sink waterbodies are grouped by basin: (A) Lake Bonney basin, (B) Miers Valley, (C) Cape Royds, and (D) Lake Fryxell basin.

Most regional communities outside of Lake Fryxell basin had less than 50% of their composition explained by either regional community or aeolian sources. For Lake Bonney basin sink communities, a substantially higher proportion of composition was explained from all regional sources than for same-basin or aeolian sources. In Miers Valley, Adams Stream orange mats (ADA_M orange) were explained more by Miers Valley aeolian sources than the nearby Miers Stream, and Miers Stream orange mats (MIE_M orange) were weakly explained nearby aeolian or stream sources. Pony Lake (PON_R_lake) and Coast Lake (COA_R_lake) communities in Cape Royds were barely explained by any MSR sources, while Cape Royds ponds (ROY_R_pond) and Blue Lake (BLU_R_lake) communities were explained by a balance of aeolian, MDV, and Cape Royds lakes sources. In Lake Fryxell basin, communities without hydrologically connected sources were better explained by all regional communities and same-basin regional communities than by aeolian communities. Nonetheless, aeolian communities regularly explained more than 10% of regional sink composition for both isolated and hydrologically connected communities. Across habitat types, Lake Fryxell basin sink communities had high proportions (>50%) of composition explained by a variety of MSR and same-basin regional and aeolian communities.

For all hydrologically connected stream communities (CAN_F, BOW_F, and VNG_F), hydrologically connected sources explained substantially more composition than did same-basin or regional aeolian sources. For example, Von Guerard Stream black mats (VNG_F black) had a higher explained proportion from Von Guerard Stream green (VNG_F green) and orange mats (VNG_F orange) than from regional or Lake Fryxell basin aeolian communities.

Discussion

Wind-mediated dispersal has often been hypothesized to be a driver of diatom metacommunity structure in landscapes with limited connectivity among freshwater habitats (Soininen 2007, Vanormelingen et al. 2008, Liu et al. 2013, Stanish et al. 2013, Sakaeva et al. 2016). The MSR of Antarctica contains glacier cryoconite holes, intermittent streams, lakes, and ponds that vary in hydrologic connectivity and geographic isolation and contain a well-described diatom flora. Strong winds deliver geological and biological material across the landscape year-round. Therefore, the MSR is a valuable system for characterizing the contribution of aeolian diatom dispersal to regional metacommunity structure. The present study confirms diatoms are passively dispersed through aeolian transport. Aeolian communities included taxa and displayed compositional patterns that suggest winds integrate individuals from across habitat types and waterbodies in the MSR.

We did not measure a strong effect of wind variables on diatom metacommunity structure, but the importance of strong winds in facilitating material transport is well documented (Kok et al. 2012). The highest wind speeds occurred during the winter and full year sampling seasons, which corresponded with the most aeolian material and diatom cells transported. These results indicate winter is responsible for the majority of aeolian diatom dispersal, supporting our H_1 that seasonality affects bulk diatom dispersal. These results are consistent with other work on aeolian material fluxes in the MDVs that demonstrate southeasterly, foehn wind events that dominate in the winter and are present in the summer are responsible for the majority of aeolian dust transport (Šabacká et al. 2012).

While the number of diatoms varied across seasons, community composition was consistent over time at each aeolian collector, and communities differed across space. Overall, our H_{2a} that aeolian composition varied over space and season was only partially supported because we did not observe substantial site-specific seasonal differences in aeolian community composition. These results indicate similar wind events occur in each season that consistently disperse diatoms from the same regional sources, just in varying absolute abundances of individuals.

Diatoms from all size classes were found in aeolian material, but small diatoms were most frequent and abundant. Despite this, aeolian communities did not have higher relative abundances of smaller taxa than did regional communities. Importantly, though, whole community size class structure differed between aeolian communities and regional communities because of an overall lack of abundant larger taxa. Additionally, the smallest diatom in the MSR, *M. atomus*, was more abundant in aeolian samples relative to regional communities, suggesting some taxon-specific size selection may occur in aeolian dispersal. Overall, we found mixed support for our H_{2b} that aeolian samples were overinflated with small taxa.

Long-term suspension in wind is most common with fine-grained material < 20 μm diameter, while short-term suspension and short distance saltation is more common with larger particles (Maher et al. 2010). During our sampling seasons, both down-valley and up-valley wind speeds in Taylor Valley exceeded the 10 m s^{-1} threshold for < 500 µm diameter particle saltation observed in the MDVs (Gillies et al. 2013), and high elevation winds can sustain long-distance suspension for smaller particles (Doran et al. 2002). Importantly, diatoms are substantially lighter than similarly sized inorganic particles, suggesting diatom saltation and long-distance suspension may occur at lower wind speeds than for inorganic material (Barton et al. 2013). Meanwhile, transportation for any distance is likely associated with high rates of cell fragmentation, given probable collision with inorganic particles during suspension and deposition. Repeated suspension and deposition over long distances would further increase the probability of cell fragmentation. In this study, recorded cell fragmentation may have been artificially high. Cells likely continued blow around and collide with other particles inside the collectors, whereas under natural conditions cells would likely settle in a more stable habitat. Even so, the two most abundant taxa in aeolian material, *H. arcuata* var. *parallela* and *M. atomus*, were also the two taxa most likely to be intact as whole cells. These taxa are small with relatively thick cell walls that could allow for long-distance transport with relatively little fragmentation, potentially explaining the high relative abundance of small taxa in aeolian material. In contrast, the similarly sized but more lightly silicified taxon *Fistulifera pelliculosa* is abundant in MSR streams but was not detected in aeolian material, suggesting silicification and strength of the frustule wall could affect aeolian dispersal.

We observed high diversity and absolute abundances of solitary diatoms in the absence of large dehydrated mat fragments, which suggests our estimates of aeolian diatom dispersal are conservative. In MSR freshwater habitats, diatoms almost exclusively occur within benthic microbial mats that desiccate in the absence of water and likely become positively buoyant because of large surface areas relative to mass (Hawes et al. 1992). Dehydrated mats have been observed in the MSR away from permanent aquatic features, likely moved by wind. For example, atop Commonwealth Glacier up to 2 cm^2 flakes of dehydrated black mats have been seen in some abundance (Sommers et al. 2019). Although the physiological mechanisms for diatom survival during dehydration are unknown among MSR taxa (*but see* Hawes et al. 1992, McKnight et al. 1999), work with other cyanobacterial biofilms has shown extracellular polymeric substances produced in microbial mat communities can provide a microenvironment that supports microbial life during extreme conditions (Lebre et al. 2017). Dehydrated mat fragments likely include concentrated and viable diatoms of all sizes that may have higher probabilities of colonization in new aquatic habitats than do solitary diatoms because of the microenvironment and interspecific interactions in the mat (Robson 2000). Dehydrated mat fragments are likely an important form of aeolian dispersal that our collectors excluded, particularly since more aeolian material was transported during winter and full-year sampling seasons when mats are dehydrated than in the summer when mats are active and attached.

The potentially viable cells we observed spanned genera and size classes, suggesting the possibility that a large number of diatom taxa in the MSR could survive aeolian transport. The abundance of solitary diatoms containing protoplasm suggest up to 15% of cells could be viable in at least some sites and seasons, though cell viability was not tested in the present study. The maximum percentage of potentially viable cells we observed is about half that of the mean percentage of potentially viable cells observed in two live stream mat communities in Taylor Valley (30 - 40%) (Kohler et al. 2016). Despite constituting a low percentage of valves observed, we estimated between 2,600 and 343,000 potentially viable cells could be transported at a single site in a single season or year. Even in those samples in which we did not observe any potentially viable cells, assuming only 0.1% cell viability would lead to between 1 and 4,400 viable cells distributed at a site. Considering Taylor Valley contains an approximate surface water area in streams alone of > 7 km^2 (compared to the aeolian collector area of 2 cm^2), there is the potential for a high number of aeolian transported diatoms to not only be viable over space and time but to be deposited in aquatic habitats. Additional studies are needed to better understand MSR diatom physiology, survival strategies, and colonization mechanisms in active and dispersing communities, but relationships between aquatic and aeolian diatom communities suggest viable cells are dispersed through wind.

We observed a roughly even mix of diatoms from regional and local sources in aeolian material, which was contrary to our H_3 that aeolian diatoms would be primarily sourced from local habitats. The one exception was the strong relationships between glacial aeolian material and cryoconite. These results are consistent with some geochemical evidence for inter-basin sources in aeolian samples superimposed upon the geochemical signatures of aeolian material from local sources (Diaz et al. 2018). Our statistical analyses indicated that aeolian diatom communities could be composed of taxa from all regional waterbodies and aquatic habitat types in relatively equal proportions. Notably, aeolian communities in the Lake Fryxell basin may be more affected by same-basin sources than communities in other lake basins, potentially reflecting the larger number of streams with abundant coverage of microbial mats in Lake Fryxell basin. Because the SourceTracker analyses were constrained by high compositional overlap among potential source communities and no geographic distance penalty in source assignments, our interpretation of the geographic scale of aeolian source communities remains tentative. Overall, compositional similarities between aeolian communities and regional communities across the MSR indicate there is some spatial variability in aeolian source material.

Although aeolian material is generally transported down-valley in Taylor Valley (Diaz et al. 2018), no clear down-valley trends in aeolian diatom community composition were observed. At each site, the diatoms we detected were representative of both local and regional sources, suggesting that even near surface winds carry diatoms from across the landscape, thereby connecting distant habitats and establishing a functional metacommunity across the MSR. However, the distinct composition of aeolian communities relative to regional communities suggests that some uncharacterized source or process, such as cell wall silicification, affects the community composition of aeolian diatoms.

We observed a high diversity of diatoms across space and time in a very small volume of aeolian material sampled relative to the region-wide fluxes, suggesting that many diatoms in the MSR are not dispersal limited. While aeolian communities were composed of taxa from multiple habitat types, the contribution of aeolian diatoms to aquatic community structure was low compared to intra-habitat and unknown sources. There are three plausible, likely interacting explanations for these patterns: (i) local dispersal within habitats is high, (ii) taxa are not dispersal limited and immigrant aeolian individuals are environmentally filtered so that only similar taxa to those already established survive, and (iii) immigrant aeolian taxa are excluded because of competition and priority effects (Verleyen et al. 2009). Given the regional scale of aeolian dispersal, our findings suggest biogeographic history (colonization and priority effects), mass effects (high dispersal among waterbodies), and species sorting (local environmental filtering and species interactions) interact to maintain regional metacommunity diversity in the MSR. This conclusion is consistent with work on aquatic invertebrates and zooplankton in ephemeral aquatic habitats that have found high local aeolian dispersal leads to control by species sorting processes (Vanschoenwinkel et al. 2008, Lopes et al. 2016) but that historical processes and regional dispersal are also needed to explain diversity patterns in some diatom metacommunities (Verleyen et al. 2009, Heino et al. 2010, Sakaeva et al. 2016).

Altogether, we propose a conceptual model for aeolian diatom dispersal in the MSR that incorporates our findings with respect to our initial hypotheses (Figure 7). As proposed in H_1 and H_{2a}, inter-basin transport is moderately high, affects local composition, and is driven by westerly foehn winds during the winter and easterly coastal winds during the summer. Diatoms dispersing across the polar plateau are scarce because of few potential source habitats and very long dispersal distances, despite the frequency of strong winds. Diatoms dispersing across the Ross Sea are scarce because of long dispersal distances despite many potential source habitats and strong seasonal winds. Furthermore, intra-basin transport is high and contributes to local aquatic community composition. In the context of H_{2b}, lower intensity winds result in short-distance dispersal that can result in cell fragmentation and death due to collision with inorganic particles, particularly for smaller, lightly silicified individuals. Higher intensity winds can also result in cell fragmentation and death because of inter-particle collision during initial suspension and deposition and exceedance of physiological tolerances. The finding that regional sources substantially influence aeolian communities did not support our H_3 that aeolian communities are most similar to local aquatic habitats and is reflected in the model as moderately high inter-basin dispersal. More detailed research could further assess the validity of and variance in this conceptual model.

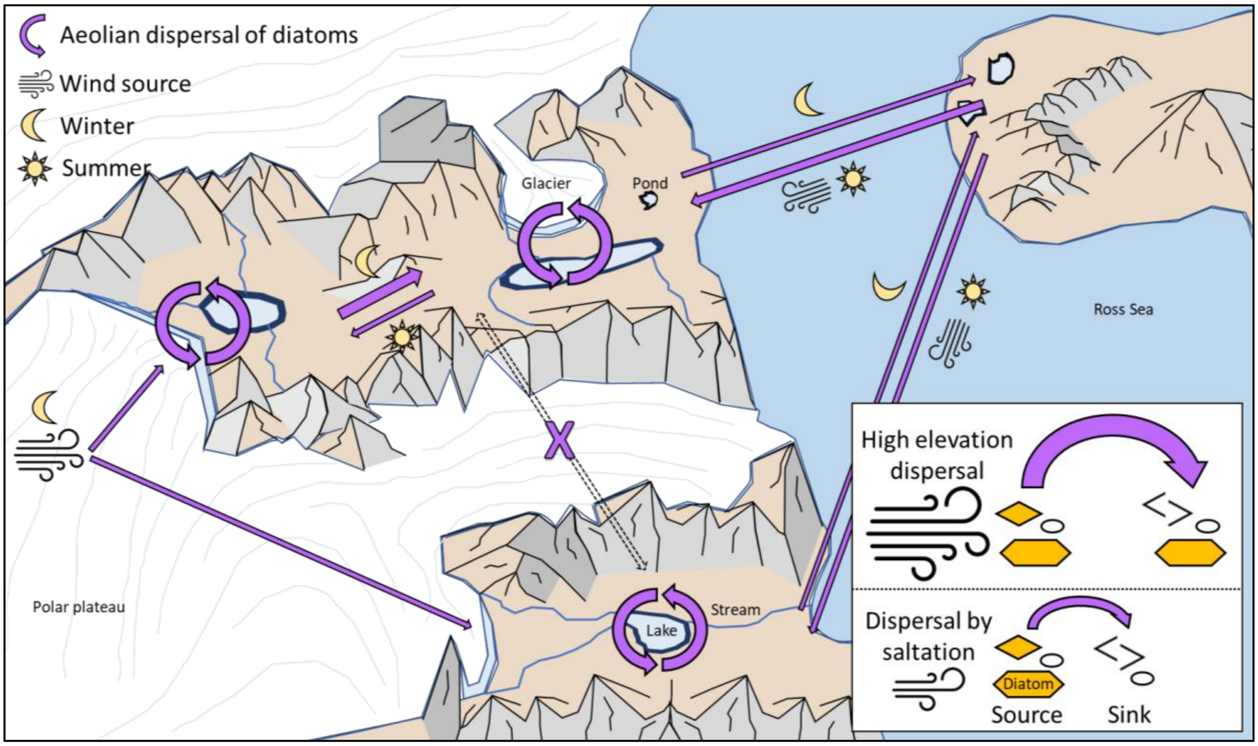


Figure 7. Conceptual model of aeolian diatom dispersal across the McMurdo Sound region. Dispersal occurs predominantly within lake basins with some seasonal inter-basin and regional dispersal. Inset: diatom transport between high intensity and low intensity winds. Stronger winds result in dispersal to high elevations and long distances; weaker winds result in low elevation and local, short-distance saltation. In both processes, living cells (orange) regardless of size in the source community can become fragmented or die due to the exceedance of physiological tolerances. Magnitude and frequency of dispersal is indicated by the width of directional arrows.

Our assessments of aeolian connectivity across the regional metacommunity were based upon community- and taxon-level comparisons. Aeolian diatoms do not represent an active, interacting community, and, for the solitary organisms we observed, dispersal and colonization likely occur at the individual level. Taxon-specific diatom populations are comprised of individuals from frequent asexual reproduction, occasional sexual reproduction, and variable immigration from aeolian or other regional sources. Characterizing intraspecific variation of each taxon could enable the inference of the relative contributions of reproduction, intra- and inter-habitat dispersal, and genetic drift to metapopulation structure. Because there is limited morphological variability within taxa across the MSR (Esposito et al. 2008, Kohler et al. 2015), determination of spatially explicit patterns in intraspecific variability would require molecular-based techniques including metabarcoding, metagenomics, restriction-site associated DNA sequencing (RAD-Seq), and whole genome sequencing (Rynearson et al. 2009, Elbrecht et al. 2018). Investigating the genetic structure of taxon-specific populations has the potential to reveal geographic- or waterbody-specific molecular signatures that could link wind-dispersed individuals to their source environments in the MSR.

In arid polar deserts like the MDVs, climate change is expected to increase landscape connectivity through increased glacial melt and intensified winds (Gooseff et al. 2011). Increased habitat connectivity is hypothesized to lead to homogenization of regional communities (Gooseff et al. 2016). While our results suggest bulk diatom aeolian transport would increase with higher wind speeds, we also found evidence for strong local control of metacommunity structure. Therefore, if sink habitats retain their environmental conditions despite climate-related disturbances, we might expect increased aeolian dispersal to only slightly affect metacommunity composition at the community scale and, rather, be more likely to affect taxon-specific metapopulation structure and intraspecific variation. However, increased surface water connectivity is likely to create novel conditions in many systems (Amoros and Bornette 2002). For example, long streams in the MDVs that regularly cease to flow in the summer are likely to become less intermittent and more resemble modern short streams (Wlostowski et al. 2016). The development of new environmental conditions could open niches that enable immigrant aeolian diatoms to establish and thrive in aquatic habitats, thus contributing to homogenization of the regional metacommunity given the taxon richness of aeolian material across space and time.

Integrating direct observations of regional dispersal into metacommunity assembly models could increase our ability to predict changes in biodiversity with environmental disturbance. Dispersal rates are a key variable in experimental and observational metacommunity studies, but few observations of passive aeolian dispersal have been integrated into assembly models (Logue et al. 2011). Simulations of diatom metacommunity structure in the MDVs using the program MCSim have resulted in well performing scenarios when high inter-habitat dispersal is considered (Sokol et al. 2020). Incorporating observed community and taxon-specific dispersal rates into simulations could improve assembly models and lead to better predictions of biodiversity outcomes with environmental disturbance (Sokol et al. 2017).

Conclusions

We detected and identified diatoms in wind-blown material from a polar desert characterized by ephemerally connected aquatic habitats, representing among the first confirmed recordings of aeolian dispersal as a mechanism for benthic diatom metacommunity assembly. A large number of diatom cells and potentially viable individuals were dispersed through the wind across space and time, with the majority of dispersal occurring during the winter when winds are strongest. Aeolian communities in the McMurdo Sound region of Antarctica have distinct diatom communities that appear to integrate taxa from multiple regional waterbodies and habitat types. However, despite widespread dispersal of 72% of taxa known in the region, aeolian dispersal explained relatively little of sink habitat biodiversity, suggesting intra-habitat dispersal, environmental factors, and historical legacies may be more dominant controls on metacommunity structure.

References

Amoros, C., & Bornette, G. (2002). Connectivity and biocomplexity in waterbodies of riverine floodplains. *Freshwater Biology*, *47*, 761-776.

Barton, A. D., Finkel, Z. V., Ward, B. A., Johns, D. G., & Follows, M. J. (2013). On the roles of cell size and trophic strategy in North Atlantic diatom and dinoflagellate communities. *Limnology and Oceanography, 58*, 254-266.

Broady, P. A. (1979). Wind dispersal of terrestrial algae at Signy Island, South Orkney Islands. *British Antarctic Survey Bulletin*, *48*, 99-102.

Brown, R. M. (1971). The distribution of airborne algae and fern spores across the island of Oahu, Hawaii. In Parker, B., Brown, R. M. (Eds), *Contributions in Phycology* (pp. 175-188). Allen Press, Lawrence, KS.

Chu, W. L., Tneh, S. Y., & Ambu, S. (2013). A survey of airborne algae and cyanobacteria within the indoor environment of an office building in Kuala Lumpur, Malaysia. *Grana*, *52,* 207-220.

Darling, J. P., Garland, D. D., Stanish, L. F., Esposito, R. M., Sokol, E. R., & McKnight, D. M. (2017). Thermal autecology describes the occurrence patterns of four benthic diatoms in McMurdo Dry Valley streams. *Polar Biology*, *40*, 2381-2396.

Deuerling, K. M., Lyons, W. B., Welch, S. A., & Welch, K. A. (2014). The characterization and role of aeolian deposition on water quality, McMurdo Dry Valleys, Antarctica. *Aeolian Research*, *13*, 7-17.

Diaz, M. A., Adams, B. J., Welch, K. A., Welch, S. A., Opiyo, S. O., Khan, A. L., ... & Lyons, W. B. (2018). Aeolian material transport and its role in landscape connectivity in the McMurdo Dry Valleys, Antarctica. *Journal of Geophysical Research: Earth Surface*, *123*, 3323-3337.

Doran, P. T., McKay, C. P., Clow, G. D., Dana, G. L., Fountain, A. G., Nylen, T., & Lyons, W. B. (2002). Valley floor climate observations from the McMurdo Dry Valleys, Antarctica, 1986–2000. *Journal of Geophysical Research: Atmospheres, 107*, ACL-13.

Elbrecht, V., Vamos, E. E., Steinke, D., & Leese, F. (2018). Estimating intraspecific genetic diversity from community DNA metabarcoding data. *PeerJ*, *6*, e4644.

Esposito, R. M. M., Horn, S. L., McKnight, D. M., Cox, M. J., Grant, M. C., Spaulding, S. A., ... & Cozzetto, K. D. (2006). Antarctic climate cooling and response of diatoms in glacial meltwater streams. *Geophysical Research Letters*, *33,* L07406.

Esposito, R. M. M., Spaulding, S. A., McKnight, D. M., Van de Vijver, B. A. R. T., Kopalová, K., Lubinski, D., ... & Whittaker, T. (2008). Inland diatoms from the McMurdo dry valleys and James Ross Island, Antarctica. *Botany*, *86*, 1378-1392.

Finlay, B. J. (2002). Global dispersal of free-living microbial eukaryote species. *Science*, *296*, 1061-1063.

Flower, R. J. (1993). Diatom preservation: experiments and observations on dissolution and breakage in modern and fossil material. *Hydrobiologia*, *269*, 473-484.

Fountain, A. G., Lyons, W. B., Burkins, M. B., Dana, G. L., Doran, P. T., Lewis, K. J., ... & Wall, D. H. (1999). Physical controls on the Taylor Valley ecosystem, Antarctica. *Bioscience, 49,* 961-971.

Fountain, A. G., Tranter, M., Nylen, T. H., Lewis, K. J., & Mueller, D. R. (2004). Evolution of cryoconite holes and their contribution to meltwater runoff from glaciers in the McMurdo Dry Valleys, Antarctica. *Journal of Glaciology*, *50,* 35-45.

Gillies, J. A., Nickling, W. G., & Tilson, M. (2013). Frequency, magnitude, and characteristics of aeolian sediment transport: McMurdo Dry Valleys, Antarctica. *Journal of Geophysical Research: Earth Surface*, *118*, 461-479.

Gooseff, M. N., McKnight, D. M., Doran, P., Fountain, A. G., & Lyons, W. B. (2011). Hydrological connectivity of the landscape of the McMurdo Dry Valleys, Antarctica. *Geography Compass*, *5,* 666-681.

Gooseff, M. N., McKnight, D., Wall, D., Virginia, R., & Fountain, A. (2016). McMurdo Dry Valleys Long Term Ecological Research. *National Science Foundation Proposal*.

Hawes, I., Howard-Williams, C., & Vincent, W. F. (1992). Desiccation and recovery of Antarctic cyanobacterial mats. *Polar Biology*, *12*, 587-594.

Heino, J., Bini, L. M., Karjalainen, S. M., Mykrä, H., Soininen, J., Vieira, L. C. G., & Diniz-Filho, J. A. F. (2010). Geographical patterns of micro-organismal community structure: are diatoms ubiquitously distributed across boreal streams? *Oikos*, *119*, 129-137.

Johansen, J. R. (1999). Diatoms of Aerial Habitats. In Stoermer, E. F., Smol, J. P. (Eds.), *The Diatoms: Applications for the Environmental and Earth Sciences* (pp. 264-274). Cambridge University Press, United Kingdom.

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., ... & Kelley, S. T. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nature Methods, 8*, 761.

Khan, A. L., McMeeking, G. R., Schwarz, J. P., Xian, P., Welch, K. A., Berry Lyons, W., & McKnight, D. M. (2018). Near-Surface Refractory Black Carbon Observations in the Atmosphere and Snow in the McMurdo Dry Valleys, Antarctica, and Potential Impacts of Foehn Winds. *Journal of Geophysical Research: Atmospheres*, *123,* 2877-2887.

Kociolek, J. P., Kopalová, K., Hamsher, S. E., Kohler, T. J., Van de Vijver, B., Convey, P., & McKnight, D. M. (2017). Freshwater diatom biogeography and the genus *Luticola*: an extreme case of endemism in Antarctica. *Polar Biology, 40,* 1185-1196.

Kohler, T. J., Kopalová, K., Van De Vijver, B., & Kociolek, J. P. (2015). The genus *Luticola* DG Mann (Bacillariophyta) from the McMurdo Sound Region, Antarctica, with the description of four new species. *Phytotaxa*, *208*, 103-134.

Kohler, T. J., Van Horn, D. J., Darling, J. P., Takacs-Vesbach, C. D., & McKnight, D. M. (2016). Nutrient treatments alter microbial mat colonization in two glacial meltwater streams from the McMurdo Dry Valleys, Antarctica. *FEMS Microbiology Ecology, 92*, fiw049.

Kok, J. F., Parteli, E. J., Michaels, T. I., & Karam, D. B. (2012). The physics of wind-blown sand and dust. *Reports on Progress in Physics*, *75*, 106901.

Kristiansen, J. (1996). Dispersal of freshwater algae—a review. *Hydrobiologia*, *336*, 151-157.

Lebre, P. H., De Maayer, P., & Cowan, D. A. (2017). Xerotolerant bacteria: surviving through a dry spell. *Nature Reviews Microbiology, 15*, 285.

Liu, J., Soininen, J., Han, B. P., & Declerck, S. A. (2013). Effects of connectivity, dispersal directionality and functional traits on the metacommunity structure of river benthic diatoms. *Journal of Biogeography*, *40*, 2238-2248.

Logue, J. B., Mouquet, N., Peter, H., Hillebrand, H., & Metacommunity Working Group. (2011). Empirical approaches to metacommunities: a review and comparison with theory. *Trends in Ecology & Evolution*, *26*, 482-491.

Lopes, P. M., Bozelli, R., Bini, L. M., Santangelo, J. M., & Declerck, S. A. (2016). Contributions of airborne dispersal and dormant propagule recruitment to the assembly of rotifer and crustacean zooplankton communities in temporary ponds. *Freshwater Biology*, *61*, 658-669.

Lyons, W. B., Frape, S. K., & Welch, K. A. (1999). History of McMurdo Dry Valley lakes, Antarctica, from stable chlorine isotope data. *Geology*, *27*, 527-530.

Maher, B. A., Prospero, J. M., Mackie, D., Gaiero, D., Hesse, P. P., & Balkanski, Y. (2010). Global connections between aeolian dust, climate and ocean biogeochemistry at the present day and at the last glacial maximum. *Earth-Science Reviews*, *99*, 61-97.

Marshall, W. A., & Chalmers, M. O. (1997). Airborne dispersal of Antarctic terrestrial algae and cyanobacteria. *Ecography*, *20*, 585-594.

Martiny, J. B. H., Bohannan, B. J., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., ... & Morin, P. J. (2006). Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, *4*, 102-112.

Mass, A. Q. (2018). *Biogeochemistry, Contaminant Transport, and Atmospheric Exchange In Glacial Cryoconite Meltwater of the McMurdo Dry Valleys, Antarctica*. Doctoral dissertation, University of Colorado at Boulder.

McKnight, D. M., Niyogi, D. K., Alger, A. S., Bomblies, A., Conovitz, P. A., & Tate, C. M. (1999). Dry valley streams in Antarctica: ecosystems waiting for water. *Bioscience*, *49*, 985-995.

Morán-Ordóñez, A., Pavlova, A., Pinder, A. M., Sim, L., Sunnucks, P., Thompson, R. M., & Davis, J. (2015). Aquatic communities in arid landscapes: local conditions, dispersal traits and landscape configuration determine local biodiversity. *Diversity and Distributions*, *21*, 1230-1241.

Nkem, J. N., Wall, D. H., Virginia, R. A., Barrett, J. E., Broos, E. J., Porazinska, D. L., & Adams, B. J. (2006). Wind dispersal of soil invertebrates in the McMurdo Dry Valleys, Antarctica. *Polar Biology*, *29*, 346-352.

Nylen, T. H., Fountain, A. G., & Doran, P. T. (2004). Climatology of katabatic winds in the McMurdo dry valleys, southern Victoria Land, Antarctica. *Journal of Geophysical Research: Atmospheres*, *109*, D03114.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., ... & Wagner, H. (2019). vegan: Community Ecology Package. R package version 2.5-6. Available at https://CRAN.R-project.org/package=vegan.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org/.

Roberts, D. W. (2019). labdsv: Ordination and Multivariate Analysis for Ecology. R package version 2.0-1. Available at https://CRAN.R-project.org/package=labdsv.

Robson, B. J. (2000). Role of residual biofilm in the recolonization of rocky intermittent streams by benthic algae. *Marine and Freshwater Research*, *51*, 725-732.

Rynearson, T. A., Lin, E. O., & Armbrust, E. V. (2009). Metapopulation structure in the planktonic diatom *Ditylum brightwellii* (Bacillariophyceae). *Protist*, *160*, 111-121.

Šabacká, M., Priscu, J. C., Basagic, H. J., Fountain, A. G., Wall, D. H., Virginia, R. A., & Greenwood, M. C. (2012). Aeolian flux of biotic and abiotic material in Taylor Valley, Antarctica. *Geomorphology*, *155*, 102-111.

Sakaeva, A., Sokol, E. R., Kohler, T. J., Stanish, L. F., Spaulding, S. A., Howkins, A., ... & McKnight, D. M. (2016). Evidence for dispersal and habitat controls on pond diatom communities from the McMurdo Sound Region of Antarctica. *Polar Biology*, *39*, 2441-2456.

Sharma, N. K., Rai, A. K., Singh, S., & Brown, R. M. (2007). Airborne algae: their present status and relevance. *Journal of Phycology*, *43*, 615-627.

Sherwood, A. R., Dittbern, M. N., Johnston, E. T., & Conklin, K. Y. (2017). A metabarcoding comparison of windward and leeward airborne algal diversity across the Ko'olau mountain range on the island of O'ahu, Hawai'i. *Journal of Phycology*, *53*, 437-445.

Soininen, J. (2007). Environmental and spatial control of freshwater diatoms—a review. *Diatom Research*, *22*, 473-490.

Sokol, E. R., Brown, B. L., & Barrett, J. E. (2017). A simulation-based approach to understand how metacommunity characteristics influence emergent biodiversity patterns. *Oikos*, *126*, 723-737.

Sokol, E. R., Barrett, J. E., Kohler, T. J., McKnight, D. M., Salvatore, M. R., & Stanish, L. F. (2020). Evaluating Alternative Metacommunity Hypotheses for Diatoms in the McMurdo Dry Valleys Using Simulations and Remote Sensing Data. *Frontiers in Ecology and Evolution, 8*, 295.

Sommers, P., Darcy, J. L., Porazinska, D. L., Gendron, E., Fountain, A., Zamora, F., ... & Ryder, J. (2019). Comparison of microbial communities in the sediments and water columns of frozen cryoconite holes in the McMurdo Dry Valleys, Antarctica. *Frontiers in Microbiology*, *10,* 65.

Souffreau, C., Vanormelingen, P., Verleyen, E., Sabbe, K., & Vyverman, W. (2010). Tolerance of benthic diatoms from temperate aquatic and terrestrial habitats to experimental desiccation and temperature stress. *Phycologia*, *49*, 309-324.

Speirs, J. C., Steinhoff, D. F., McGowan, H. A., Bromwich, D. H., & Monaghan, A. J. (2010). Foehn winds in the McMurdo Dry Valleys, Antarctica: The origin of extreme warming events. *Journal of Climate*, *23,* 3577-3598.

Stanish, L. F., Nemergut, D. R., & McKnight, D. M. (2011). Hydrologic processes influence diatom community composition in Dry Valley streams. *Journal of the North American Benthological Society*, *30*, 1057-1073.

Stanish, L. F., Kohler, T. J., Esposito, R. M., Simmons, B. L., Nielsen, U. N., Wall, D. H., ... & McKnight, D. M. (2012). Extreme streams: flow intermittency as a control on diatom communities in meltwater streams in the McMurdo Dry Valleys, Antarctica. *Canadian Journal of Fisheries and Aquatic Sciences, 69*, 1405-1419.

Stanish, L. F., Bagshaw, E. A., McKnight, D. M., Fountain, A. G., & Tranter, M. (2013). Environmental factors influencing diatom communities in Antarctic cryoconite holes. *Environmental Research Letters*, *8,* 045006.

Tesson, S. V., Skjøth, C. A., Šantl-Temkiv, T., & Löndahl, J. (2016). Airborne microalgae: Insights, opportunities, and challenges. *Applied Environmental Microbiology, 82*, 1978-1991.

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

Vanschoenwinkel, B., Gielen, S., Seaman, M., & Brendonck, L. (2008). Any way the wind blows – frequent wind dispersal drives species sorting in ephemeral aquatic communities. *Oikos*, *117*, 125-134.

Verleyen, E., Vyverman, W., Sterken, M., Hodgson, D. A., De Wever, A., Juggins, S., ... & Flower, R. (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., ... & Flower, R. (2007). Historical processes constrain patterns in global diatom diversity. *Ecology*, *88*, 1924-1931.

Witherow, R. A., Lyons, W. B., Bertler, N. A., Welch, K. A., Mayewski, P. A., Sneed, S. B., ... & Fountain, A. (2006). The aeolian flux of calcium, chloride and nitrate to the McMurdo Dry Valleys landscape: evidence from snow pit analysis. *Antarctic Science*, *18*, 497-505.

Wlostowski, A. N., Gooseff, M. N., McKnight, D. M., Jaros, C., & Lyons, W. B. (2016). Patterns of hydrologic connectivity in the McMurdo Dry Valleys, Antarctica: a synthesis of 20 years of hydrologic data. *Hydrological Processes*, *30*, 2958-2975.

CHAPTER III

MODELING RESPONSES OF DIATOM TAXA TO CLIMATE, WATERSHED CHARACTERISTICS, AND IN-STREAM STRESSORS IN THE NORTHEAST UNITED STATES

Abstract

Benthic diatoms are important indicators of stream health and are strongly influenced by climate, watershed characteristics, and in-stream conditions. However, few attempts have been made to assess the relative importance of co-occurring stressors on diatom taxon abundance at regional spatial scales, particularly for infrequently measured stressors like pesticides and persistent organic pollutants. As climate change and human disturbances to freshwater streams alter ecosystem structure and function in multi-faceted ways, it is critical to integrate multivariable environmental effects into diatom-based environmental assessments and ecological research. We present over 600 generalized additive models that explain the relative abundances of 268 benthic diatom taxa, constructed from a pool of 27 distinct environmental factors measured across 92 sites along urban and agricultural gradients in the northeast United States. To our knowledge, these models constitute the largest effort to date to identify which among a suite of environmental variables most influence diatom taxon patterns along gradients of human impact. We found that watershed-scale factors, stream geomorphology, streamflow, water temperature, dissolved oxygen, pH, specific conductance, nutrients, herbicides, fungicides, polychlorinated biphenyls (PCBs), and metals each strongly affected diatom taxa across morphological groups and evolutionary lineages. The relative importance of watershed and in-stream factors varied substantially by taxon. Our models were consistent with previously described autecological preferences for many taxa regarding commonly measured stressors. Meanwhile, the present study is the first to describe the responses of many taxa to less well studied stressors like herbicides, PCBs, and metals. For each taxon, measures of model composition, performance, and

deviance explained are presented alongside partial effects of each variable as autecological data that can be applied to water quality assessments and further research on diatom ecology.

Introduction

Environmental processes have large effects on biotic species, particularly along environmental gradients (Cottenie 2005). For benthic diatoms at the base of stream ecosystem function, taxon-specific variability is often tightly associated with local environmental changes (Heino et al. 2014). Strong relationships between local environmental conditions and diatom taxon patterns have led to the use of diatoms as indicators of stream health for decades (Patrick 1973, Pan et al. 1996). However, almost all environmental assessments that use diatom taxon-specific patterns are based on single stressor responses, derived from techniques such as weighted averaging to determine univariate optima and tolerances (Stevenson et al. 2008). Meanwhile, anthropogenic disturbances commonly have multi-faceted effects on environmental conditions (Maloney and Weller 2011, Pickett et al. 2011). In particular, the effects of contaminants such as pesticides, persistent organic pollutants, and metals on diatom taxa have rarely been assessed alongside more frequently measured stressors outside of distinctive situations such as acid mine drainage (Roubeix et al. 2011). Additionally, spatially structured environmental variables like climate, watershed topography, and geology can confound, contribute to, or dominate over diatom responses to in-stream conditions and should be included when assessing diatom-environmental relationships (Soininen 2007, Soininen and Weckström 2009). To better understand taxon-specific biogeography and inform environmental assessment, researchers need to consider diatom responses to the relative and additive effects of multiple variables that operate at local (in-stream) and regional (watershed and inter-watershed) scales.

Non-linear, multivariable modeling provides the opportunity to determine the relative importance, explanatory power, and shape of effect that multiple environmental variables have upon taxon-specific patterns. While many studies have applied multivariable models to diatom community structure (Soininen and Teittinen 2019), few have modeled multivariable environmental influences on taxon-specific responses of multiple diatom taxa at large spatial scales (Pajunen et al. 2016, Soininen et al. 2019). Correlative species distribution models (SDMs) incorporate observed occurrence or abundance (absolute or relative) of a taxon as a function of environmental conditions (Elith and Leathwick 2009). SDMs of stream diatom taxa have largely been constructed to determine potential habitat suitability for single species of special interest, like the invasive *Didymosphenia geminata* (Kumar et al. 2009). By applying the multivariable framework of SDMs to multiple species, researchers can identify similarities in environmental responses among taxa as well as the relative importance of different environmental controls (Pajunen et al. 2016). Altogether, this information can help to determine autecological characteristics of individual and groups of taxa that can advance the development of biotic metrics for environmental assessment. Additionally, this modeling approach provides insight into the variability of diatom responses to environmental variables that affect taxa at multiple spatial scales, thereby expanding our understanding of the controls on diatom biogeography.

Our objectives were to (1) assess the relative importance of climate, watershed topography, geology, stream geomorphology, hydrology, and water chemistry (particularly understudied contaminants) on the relative abundances of diatom taxa in streams of the northeast United States; (2) identify groups of taxa influenced by similar environmental conditions; and (3) determine taxon response curves to influential environmental variables. We identified diatoms from 92 stream biofilms collected during the summer across 8 states in the northeast U.S. We used an all combinations approach to generalized additive modeling to assess the influences of 27 distinct watershed- and local-scale variables on diatom taxon relative abundance. The resulting autecological information for 268 benthic diatom taxa is provided to inform water quality assessments and general understanding of diatom ecology.

Methods

Sample collection

Between July and August 2016, 92 wadeable stream sites were sampled across 8 states in the northeast U.S. (150,000 km^2) as part of the U.S. Geological Survey (USGS) National Water-Quality Assessment (NAWQA) Northeast Stream Quality Assessment (NESQA; Coles et al. 2019; Figure 1).

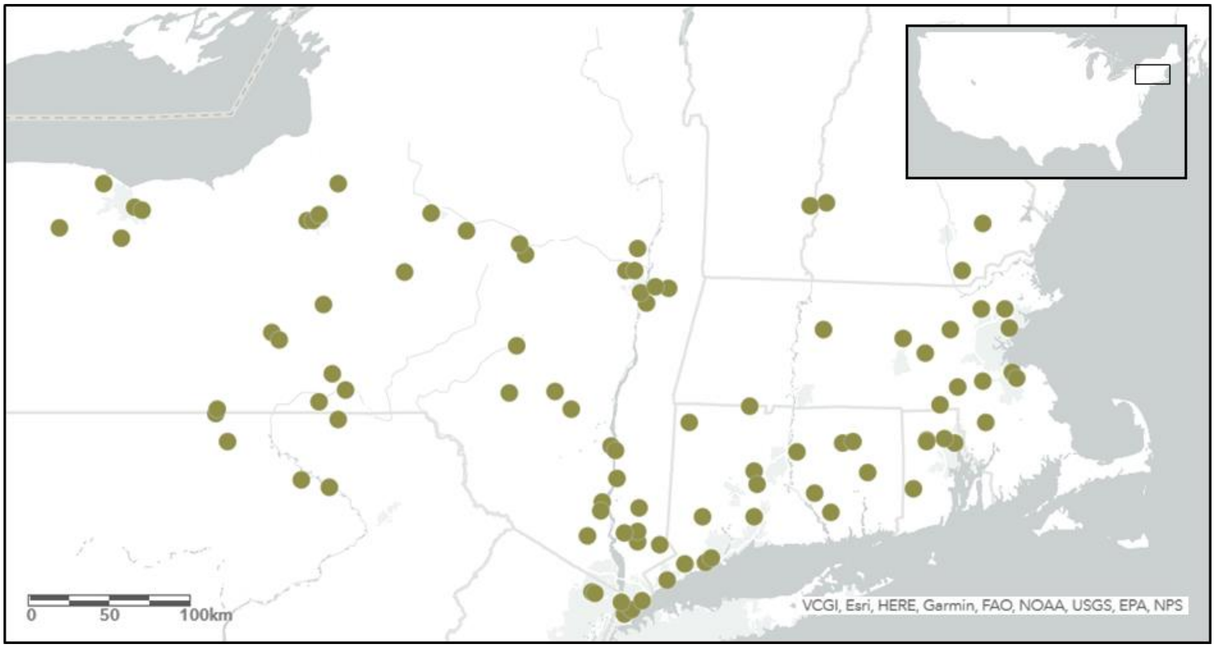


Figure 1. Map of sites in the northeast United States used in the analysis.

Sites were selected along urban and agricultural gradients according to watershed land use from the 2011 National Land Cover Database (NLCD; Homer et al. 2015). Of all sites, 69%

were in watersheds in which urbanization was considered the main human development (urban land cover exceeded 1% and agricultural cover did not exceed 5%), 17% were most characterized by agriculture (cultivated row crop land cover exceeded 5%), 2% by mixed urbanization and agriculture (urban and agricultural land cover each exceeded 20%), and 12% by undeveloped land (urban and agricultural land cover did not exceed 1% and 5%, respectively). Urban-dominated sites were incrementally distributed across 5 tiers of intensity based on percent of urban land cover: tier 1 (1 – 10% urban land cover), tier 2 (10 – 20%), tier 3 (20 – 37.5%), tier 4 (37.5 – 50%), and tier 5 (> 50%; Coles et al. 2019). For each site, watershed-scale factors were measured from NLCD data, including drainage area, mean elevation, mean annual precipitation, annual runoff, and watershed geology and soil characteristics.

At each site, local, in-stream factors were measured during 4- to 9-week sampling periods, with longer sampling for watersheds with more development. During the sampling period, surface water was analyzed across a 150-m reach for key field parameters and in-stream stressors following the methods in Coles et al. (2019). Briefly, dissolved oxygen (DO), pH, and specific conductance were measured weekly with YSI field meters from 5 locations within each stream reach. Weekly discrete surface water samples were collected as 10 subsamples across the reach and composited. Discrete samples were analyzed for major ions, nutrients, and pesticides. Dissolved cation concentrations were determined by inductively coupled plasma-atomic emissions spectroscopy (Fishman 1993), and dissolved anion concentrations by ion chromatography (Fishman and Friedman 1989). Nitrogen (ammonia, nitrate, nitrite, total) and phosphorus (orthophosphate, total) concentrations were determined by colorimetry (Fishman 1993, O'Dell 1993, Patton and Kryskalla 2011). From filtered water samples, 225 pesticides and pesticide degradates were analyzed by direct aqueous injection liquid chromatography tandem mass spectrometry (Sandstrom et al. 2016).

For each site, one or two streambed bulk sediment samples were collected during the sampling period from multiple collections at depositional areas along the reach (Shelton and Capel 1994, Radtke 2005). Subsamples were analyzed for grain size, persistent organic pollutants (including polychlorinated biphenyls [PCBs]), major and trace metals, and other contaminants. PCBs were measured using accelerated solvent extraction followed by solid-phase extraction cleanup and analysis by electron-capture negative ion mode (Mahler et al. 2009, Wagner et al. 2016). For each sample, a probable effect concentration quotient (PECQ) was derived for PCBs by dividing the PCB concentration by its probable effect concentration as determined in MacDonald et al. (2000). A PECQ was derived from organic carbon-normalized concentrations of 7 metals (As, Cd, Cr, Cu, Pb, Ni, Zn) that were pooled together (Moran et al. 2017).

Physical habitat was assessed and aquatic biota were collected at the end of the sampling period following standard USGS protocols (Fitzpatrick et al. 1998). Each 150-m stream reach was segmented with 11 transects separated by 15 m intervals. Stream geomorphology was characterized by measuring canopy cover (spherical densiometer at the center of each transect), channel wetted width and bankfull height (together converted to channel width:depth), riffle percentage, and substrate size (at 5 equidistant locations along each transect). At each transect, 2-cm^2 rock scrapes for benthic biofilms were made, alternating among left, center, and right sampling locations with each new transect. All biofilm collections along a reach were combined. Composite biofilm samples were homogenized, subsampled with known volumes, and preserved in 2.5% formalin for microscopic identification. Streamflow and water temperature were measured continuously for up to one year prior to physical habitat and biofilm sampling. At sites with USGS streamflow gages, stream stage and discharge were recorded at 15-minute intervals. At non-gaged sites, water-level loggers (HOBO Water Temp Pro v2 U22) recorded stream stage and water temperature at hourly intervals. Environmental data are available at the NESQA website (https://webapps.usgs.gov/rsqa/#!/region/NESQA).

Diatom identification

Organic matter was removed from biofilm subsamples by digestion, using concentrated nitric acid within an Anton Paar microwave digester (Bishop et al. 2017). Samples were serially centrifuged, decanted, and rinsed with distilled water. Then, a known volume was dried onto four replicate cover slips in Battarbee chambers (Battarbee 1973). Cover slips were mounted onto glass microscope slides using Naphrax (Brunel Microscopes Ltd) or Zrax (MicrAP Enterprises) mounting medium. Slides and cleaned material were archived in the INSTAAR Diatom Collection, University of Colorado.

One replicate slide from each site was examined using light microscopy with differential interference contrast and a 100x oil immersion (1.4) objective lens and condenser lens. Before slides were analyzed for diatoms, a working flora was created following the methods outlined in Bishop et al. (2017). Approximately 80% of the slides were examined, and each distinct taxon was imaged across its size variation and assigned a provisional code independent of species-level taxonomy. Slides were assigned to two analysts randomly, with 10% of slides self-recounted and 10% cross-recounted (Tyree et al. 2020). Each of the analysts identified and counted 600 valves from each slide, following the provisional code for each taxon in the working flora. A valve was only counted if at least 60% of the valve was visible within the field of view. Taxa identified during formal counts but not during the creation of the initial flora were imaged, given new provisional codes, and added to the working flora. After all slides were counted, the analysts worked together to determine formal species names for each identified taxon using Krammer (2000), Cantonati et al. (2017), Spaulding et al. (2019), and other resources. All scientific names and authorship follow USGS BioData, and all count data are archived in the BioData database (U.S. Geological Survey 2019). Provisional names were assigned to taxa for which no formal name could be definitively determined. In the present study, diatom taxa are reported as relative abundance by site.

Environmental variable selection

From over 400 measured variables, 36 watershed-scale variables and environmental stressors were selected *a priori* as factors that have been shown to or were deemed likely to affect diatom taxon relative abundance. Spearman rank correlations were calculated across these 36 variables. Of variable pairs with $\rho > |0.70|$, only the more inclusive or more ecologically interpretable variable was retained. In total, 27 uncorrelated variables were selected for use in further analysis (Table 1). These variables were organized into two broad categories: watershed-scale variables and local, in-stream variables. Watershed-scale variables were drainage area, elevation, mean annual precipitation, runoff, sedimentary geology, percent SiO, and soil permeability. Local, in-stream variables were grouped into 9 variable classes: stream geomorphology, stream discharge, water temperature, DO, pH, ions, nutrients, herbicides, and other pollutants (fungicides, total metals, and PCBs). In total, variables were assigned to 10 classes (watershed-scale as one class and the 9 local-scale classes). The median value was used for each variable with multiple measurements over the sampling period. Each variable was transformed to reduce skewness and kurtosis.

To verify that the selected variables represented environmental gradients across sites, principal components analysis (PCA) was performed. Variables that were correlated (Pearson's *r* > |0.20|) with a principal component that explained > 5% of inter-site variation were considered to contribute to overall environmental variability across the study region.

Table 1. Descriptive statistics for each of the 27 environmental variables used in models. For each variable, correlation with one of the principal components explaining > 5% of inter-sample variation is listed.

Class	Variable	Min	Med	Mean	Max	PC	Loading
Watershed	Drainage area (km^2)	6.1	67.27	75.14	263.26	PC5	-0.37
	Elevation (m)	31.89	158.9	216.07	708.49	PC1	-0.3
	Mean annual precipitation (mm)	85.2	120.9	116.41	135.1	PC2	0.41
	Mean annual runoff (mm)	264.8	359.9	369.65	547.75	PC6	-0.34
	Sedimentary geology (%)	0	100	78.74	100	PC2	-0.3
	SiO (%)	32.82	62.67	60.52	75.03	PC2	0.33
	Soil permeability (%)	0	9	11.14	80	PC6	0.35
Geomorphology	Canopy (%)	0.53	70.05	63.99	100	PC6	0.41
	Channel width:depth	4.23	13.65	15.02	69.02	PC3	-0.28
	Riffles (%)	0	28.57	30.21	100	PC3	-0.29
	Substrate D50 (%)	0.1	4.63	5.32	25	PC3	-0.25
Discharge	Base flow index (%)	34.96	47.67	48.64	62.43	PC4	0.28
	Peak flow interval (days)	4.63	17.09	19.81	100	PC5	-0.27
Temperature	Water temperature (°C)	15.5	21.7	21.63	27.16	PC3	-0.34
DO	Dissolved oxygen (DO) (mg L^{-1})	2.99	8.21	7.95	9.89	PC5	0.4
pH	pH	6.23	7.61	7.56	8.72	PC2	-0.38
Ions	Specific conductance (µS cm^{-1})	55.5	476.6	612.64	2834	PC1	0.33
Nutrients	Dissolved inorganic nitrogen (DIN) (mg L^{-1})	0	0.34	0.67	7.46	PC1	0.33
	Total phosphorus (TP) (mg L^{-1})	0	0.04	0.09	2.07	PC4	-0.33
	Total nitrogen:total phosphorus (TN:TP)	3	15.24	21.05	155.34	PC4	0.56
Herbicides	Acetanilide and amide herbicides (ng L^{-1})	0	0	35.14	501.25	PC2	-0.36
	Acid herbicides (ng L^{-1})	0	0	18.75	256.62	PC1	0.28
	Sulfonylurea and urea herbicides (ng L^{-1})	0	0	5.26	65.59	PC3	-0.28
	Total herbicides (ng L^{-1})	0	35.58	104.53	872.6	PC2	-0.29
Pollutants	Total fungicides (ng L^{-1})	0	0	42.99	1535.29	PC1	0.33
	Polychlorinated biphenyls (PCBs) (PECQ)	0	0	0.01	0.87	PC1	0.23
	Total metals (PECQ)	0.01	0.12	0.17	1.3	PC3	-0.33

Min = minimum; Med = median; Max = maximum

PC = principal component explaining > 5% of inter-sample variation onto which the variable has the highest loading (from principal components analysis)

Loading = loading (Pearson's *r*) between the variable and the principal component in column PC

PECQ = probable effect concentration quotient

Generalized additive modeling

To assess the relative importance and magnitude of effect of each environmental variable on each taxon, multivariable modeling was conducted using generalized additive models (GAMs). GAMs assess the additive effects of multiple, co-occurring variables and can be used to identify which most strongly influence the response variable (e.g., taxon; Wood 2006). GAMs are non-parametric extensions of generalized linear models that use observed data to estimate the shape of the relationship between response and explanatory variables, assuming additive effects among a smoothed function of the explanatory variables (Yee and Mitchell 1991). GAMs can be robust to zero-inflated data, and observations of response variable absence (i.e., 0) contribute to the degrees of freedom used to determine the number of explanatory variables a model can contain (Dunn and Smyth 2008). General practice is that a minimum of one explanatory variable can be included in a GAM for every 10 observations (zero and non-zero; Wood 2006).

GAMs were constructed for the relative abundance of each diatom taxon using a partially unsupervised, all variable combinations approach (Marra and Wood 2011). The number of non-zero observations for each taxon ranged from 2 – 89. Computer processing power limited the construction of taxon-specific GAMs from all subsets of the 27 environmental variables. Therefore, for each taxon a subset of the 27 variables was selected for use in GAM construction using a variable selection procedure based on random forests. Briefly, random forests were computed, and environmental variables were sorted according to their mean variable importance. A threshold was computed as the minimum predicted value of a pruned classification and regression tree (CART) fitted to the standard deviations of all variable importance values. Only variables with a mean variable importance larger than the threshold were retained (Genuer et al. 2019). On average, 13 explanatory variables were selected at this step for inclusion in GAMs for each taxon. For quality control, we constructed GAMs for all subsets of the entire 27 variable pool for 5 taxa. In only one case did a variable that was excluded by random forest selection perform as well as – but not better than – GAMs constructed from the pool of variables selected by random forest filtering.

Based on the reduced set of variables, all one-, two-, three-, and four-variable combinations were formed, and each combination was used to construct a GAM. A maximum of four component variables was set to ensure sufficient degrees of freedom for each model. Each variable was included with a penalized regression spline and maximum knots (k) = 5 to ensure appropriate degrees of freedom in the model (Wood 2006). Random effects maximum likelihood (REML) smoothness selection was conducted, in which estimation was penalized by maximum likelihood estimates to avoid overfitting (Wood 2004). Heavier penalization to avoid overfitting was conducted by constructing an additional penalty on the null space of each smoothed term, with gamma = 1.4 on each model degree of freedom to increase penalization (Marra and Wood 2011). Tweedie distributions in which the power parameter was estimated between 1 and 2 were used to accommodate distributions that varied between generalized Poisson and gamma distributions and were zero-inflated (Tweedie 1984).

Nonsignificant, unexplanatory, and poorly performing models were sequentially removed from the pool of models. First, models containing variables with pairwise Spearman's $\rho > |0.50|$ or overlapping ecological interpretations (e.g., acid herbicides and sulfonylurea and urea herbicides) were removed. Second, models were retained in which each component variable was significant ($p < 0.05$) and overall model deviance explained ≥ 0.15. Next, Moran's I was calculated to estimate spatial autocorrelation (Cliff and Ord 1970). For any model with significant Moran's I ($p < 0.05$), the model was re-run with an interaction term for geographic position (latitude/longitude). If the re-run model passed the previous model quality checks, it was retained. Models with estimated concurvity (the non-parametric analog to multicollinearity, occurring when a smooth term in a model can be approximated by a different smooth term in the model) ≥ 0.70 were removed. The concurvity cutoff of 0.70 was set to match the Spearman correlation threshold used for initial variable. To reduce oversmoothing caused by too low of k for a smoothed term, any model indicating significantly low k for a term ($p < 0.05$) was discarded. Additionally, any model that did not converge (where convergence is defined as estimated parameters changing minimally between iterations) was removed.

For the models passing the above criteria, several additional computations were performed. First, Akaike Information Criterion (AIC) was used to select among models (Akaike 1974). Akaike weights and differences in AIC (ΔAIC) from the lowest AIC model were calculated. Akaike weights and ΔAIC represent the likelihood of a model relative to the other models passing the aforementioned criteria. Only models with ΔAIC ≤ 2 were retained, consistent with an evidence ratio of 2.7 that indicates the best model is more than 2.7 times more likely to be the right fit compared to any model with ΔAIC > 2 (Burnham and Anderson 2004). When examining each set of models, we found that ΔAIC ≤ 2 was consistent with retaining all models with Akaike weights ≥ 0.10, with few exceptions. Next, residual plots for each GAM were examined to ensure that the model did not violate any assumptions. Primarily, plots of residuals against observations of the explanatory variable were checked for random scatter around residuals (y) = 0. Finally, the deviance explained by each component variable was estimated by calculating the reduction in deviance explained when fitting alternative models without each term and normalizing the variable deviance explained to that of the whole model.

To identify groups of taxa that were influenced by similar environmental conditions, *k*-means cluster analysis was performed on the deviance explained of each explanatory variable across all GAMs and taxa. *K*-means clustering is an unsupervised machine learning algorithm that identifies *k* centroids (centers of clusters) and allocates each data point to the nearest centroid (MacQueen 1967). Optimal numbers of clusters were determined using the Calinski-Harabasz index, which assesses cluster validity by the average inter- and intra-cluster sum of squares (Liu et al. 2010). The maximum number of clusters permitted was 25% of the number of taxa with models that passed quality filtering. Clustering was performed across 1000 iterations to arrive at consensus clusters.

To assess the shapes of taxon responses to environmental variables, partial effects plots were created for each cluster of models and for taxa with consistent responses across multiple models. Partial plots represent the effect of a single explanatory variable on the modeled response of taxon relative abundance, in which $y = 0$ is the mean taxon relative abundance.

All analyses were performed in R v. 4.0.2 (R Core Team 2020). The PCA was conducted using ‘prcomp’ and visualized using ‘ggbiplot’ (Vu 2011), variable selection was performed using ‘VSURF’ default parameters (Genuer et al. 2019), GAMs were constructed using ‘mgcv’ (Wood 2020), Moran’s I was calculated using ‘ape’ (Paradis et al. 2004), partial effects plots were created using ‘mgcViz’ (Fasiolo et al. 2020), and color palettes were determined using ‘wesanderson’ (Ram and Wickham 2018).

Results

Environmental gradients

Study sites were distributed across gradients of watershed land use: 0 – 93% urban (mean 27%) and 0 – 77% agricultural (mean 13%). Across the sampling region, sites varied strongly in watershed-scale and in-stream variables (Table 1). Watersheds ranged in size across more than an order of magnitude (6 – 263 km^2). Watersheds spanned the low elevation Eastern Great Lakes Lowlands and Northeastern Coastal Zone to the higher Northeastern Highlands and Northern Allegheny Plateau. The majority of watersheds were dominated by sedimentary geology, but soil permeability and percent SiO varied dramatically.

Streams were sampled during overall low flows in the summer, but peak flows across the year prior to sampling varied from very frequent (5 days) to highly spread out (100 days). Sampled streams varied from low (3 mg L^{-1}) to high DO (10 mg L^{-1}). While most sites were circumneutral, some sites were strongly acidic (pH 6.2) and others alkaline (pH 8.7). Specific conductance was generally high for freshwater but ranged from fresh (55 µS cm^{-1}) to oligohaline (2800 µS cm^{-1}). Concentrations of each nutrient, herbicide, and other pollutant varied by orders of magnitude.

No single stressor or group of stressors explained a majority of inter-site variance in the PCA (Figure S1). Rather, each stressor explained portions of variability among sites, as indicated by variable loadings on different principal components (Table 1, Figure S1). These results suggest that selected variables represented environmental gradients over which diatom responses could be measured.

Relative importance of environmental variables

A total of 458 taxa were identified, and GAMs that passed quality filtering modeled the relative abundances of 268 taxa. The composition, explanatory power, and performance of each model are presented in Figure S2. Most models were highly explanatory, with a mean deviance explained of 52% (range 15 – 99%). The 190 taxa for which no models passed quality filtering were characterized by low frequency (present in < 5 sites), low relative abundance (mean < 5%

where present), or both. Over half (56%) of the unmodeled taxa were found in only one site, and 17% were present in only two sites. The frequency and mean relative abundance of each modeled taxon is presented in Table S1 and of each unmodeled taxon in Table S2.

Four or fewer variables provided the primary influence for the majority of taxa. Almost half of all modeled taxa had only one model that passed quality filtering (126 taxa). Of the 142 taxa with multiple models, 67 had a majority of models with similar explanatory variable composition based on cluster analysis (*see* next section).

Overall, model composition was highly taxon-specific. No two explanatory variables significantly covaried across all models (minimum Spearman's $\rho = 0.15$). However, a few variables were present in substantially more models and for more taxa than others (Figure 2). pH was a significant explanatory variable in over half of all models and for almost one-quarter of all taxa, followed by elevation, peak flow interval, specific conductance, and mean annual precipitation.

Each of the 27 environmental variables except total nitrogen:total phosphorus (TN:TP) strongly affected at least one taxon. Only 15 modeled taxa were affected by spatial autocorrelation and passed quality filtering with GAMs containing the latitude/longitude term. On average, each variable explained 40% of the total model deviance explained. However, the proportion of model deviance explained by each variable varied considerably within and across variables (Figure 3).

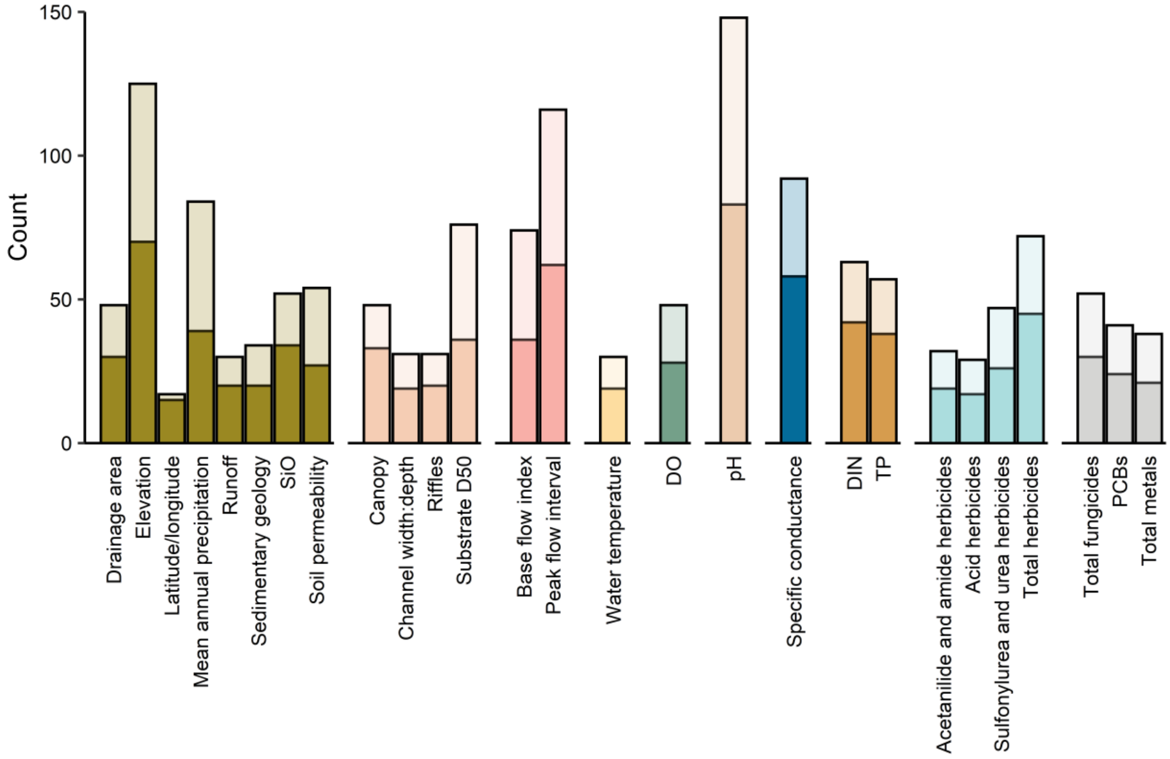


Figure 2. Number of taxa (dark shading) and models (dark + light shading) for which each environmental factor was a significant explanatory variable.

Taxon clusters

Models were grouped into 26 clusters based on common environmental variable composition and deviance explained (Figures 4 and S2). Each cluster was characterized by one environmental variable regardless of the modeled directional responses of relative abundance. At least one variable from each of the 10 environmental classes characterized a single model cluster. Elevation, peak flow interval, and pH characterized two clusters each, while latitude/longitude, runoff, and riffle percentage did not distinguish any cluster. A mean of 17 taxa comprised each cluster. Only 22 clusters contained taxa in which all or a majority of their models were grouped in the same cluster.

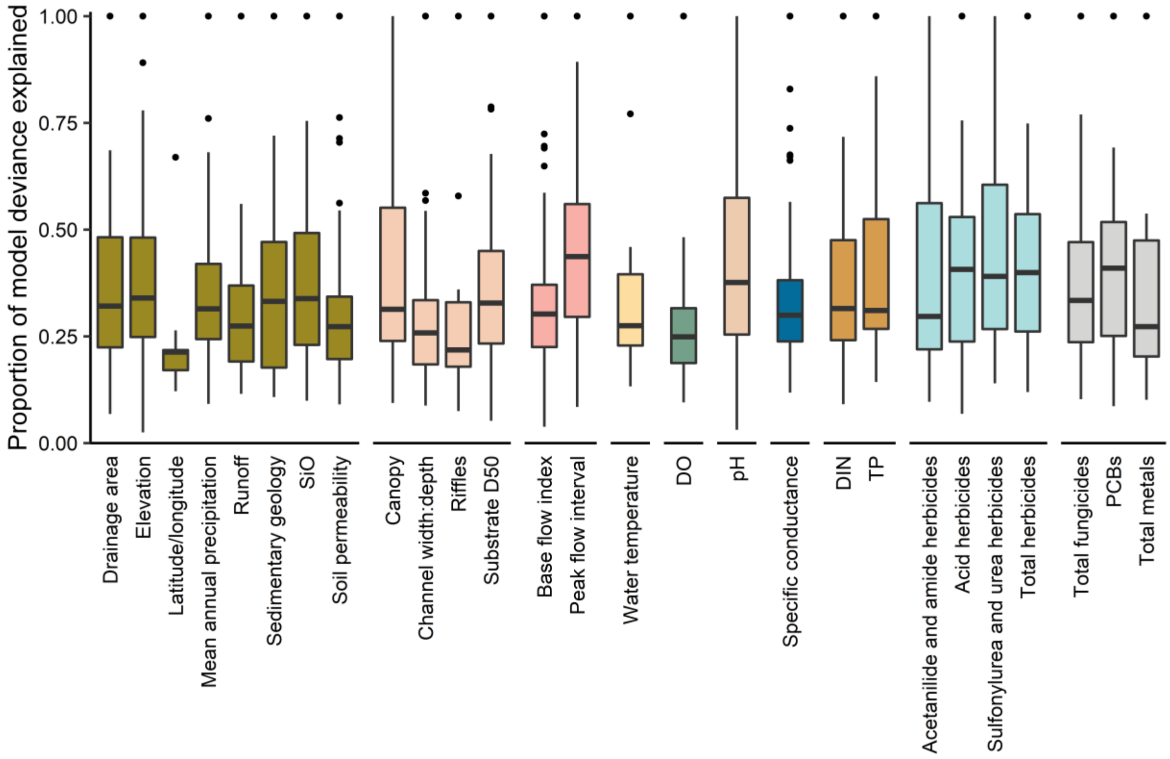


Figure 3. Proportion of model deviance explained by each environmental variable for which the variable was present in the model.

Some diatoms considered to be acidophilous were clustered together by pH (Potapova and Charles 2003), supporting the basis of our species distribution modeling approach. The influences of other stressors could not be generalized across taxa in clusters given previously identified autecological characteristics.

Partial effects of environmental variables

Partial effects of environmental variables on taxon relative abundance varied in shape and deviance explained across taxa, including within clusters (Figures 4 and S2). However, some clusters contained models in which relative abundance was influenced similarly across all models. For example, in most models clustered by watershed-scale factors and stream geomorphology (clusters 4, 6, 7, 8, 10), taxon relative abundance changed only slightly across each environmental gradient. The relative abundances of most taxa decreased at high specific

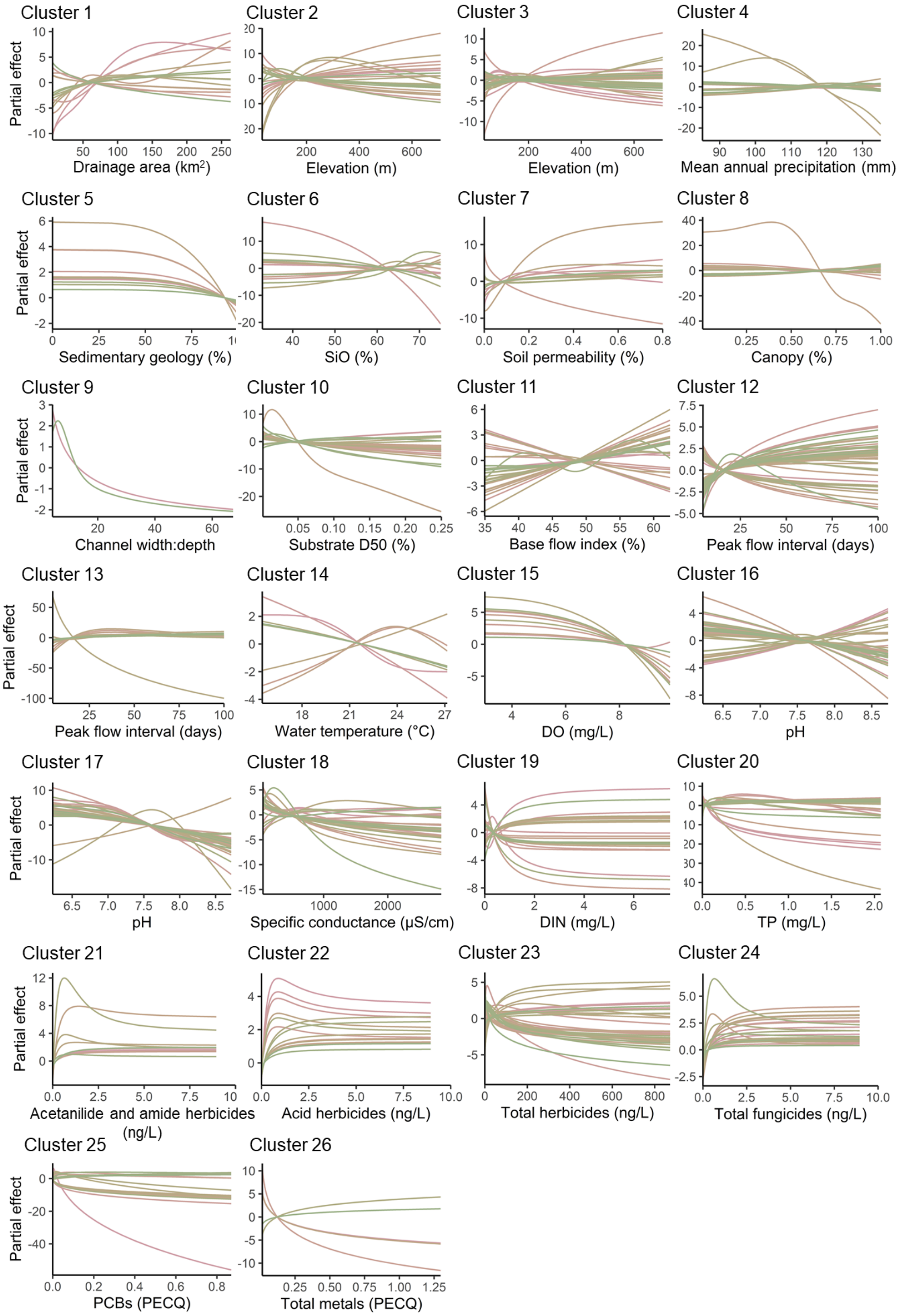


Figure 4. Partial plots of the effects of the dominant environmental variable on the predicted relative abundance of diatom taxa in each model cluster. Each line represents a single model. Refer to Figure S2 for which models and taxa are represented in each cluster.

conductance (cluster 18). Relative abundances both increased and decreased at low concentrations of dissolved inorganic nitrogen (DIN), but all responses reached an asymptote at high concentrations (cluster 19). Of the taxa with models most affected by TP, most modeled relative abundances decreased with increasing TP (cluster 20). Most modeled relative abundances increased sharply before a strong decrease at very low concentrations of herbicides and fungicides (clusters 21, 22, 24), but there were no effects on relative abundance at higher concentrations – with the exception of total herbicides that showed more variability along the gradient (cluster 23). PCBs both slightly increased or decreased modeled relative abundances of most taxa (cluster 25). Half of the taxa clustered by total metals increased logarithmically with increasing metal concentrations, and the other half decreased (cluster 26).

Discussion

Our models demonstrate that, at a large spatial scale, the environmental factors that influence diatom distributions are multi-faceted and taxon-specific across multiple environmental gradients. By selecting for models that best explained taxon relative abundance, we found that the relative influences of climate, watershed topography, geology, stream geomorphology, hydrology, and surface water chemistry differed dramatically among taxa. Despite that variability, we also showed that taxa were clustered together based on shared responses to highly explanatory variables, supporting the concept of diatom ecological guilds that are defined by shared environmental responsiveness. Most taxa varied in directional response to dominant environmental influences and were best modeled by including additive effects of environmental variables of different classes, indicating taxon-specificity within guilds along environmental gradients. Our results demonstrate that diatom taxa vary in responsiveness to environmental factors that function at multiple spatial scales and to different types of in-stream stress.

Environmental gradients

The models and partial effects of watershed-scale and in-stream factors on diatom taxon relative abundances presented in this study are derived from sites distributed along strong environmental gradients of anthropogenic impact and natural, regional variability. Our results provide evidence that anthropogenic disturbances affect taxa disparately. Taxa responded to different combinations of both in-stream factors that are affected by human disturbance and watershed factors that are less directly impacted by anthropogenic activity. For example, some taxa were almost entirely influenced by watershed-scale variables like elevation, and others by variables with exclusively anthropogenic signatures (e.g., herbicides and PCBs). Our results are consistent with recently developed SDMs for taxa across anthropogenically impacted and unimpacted streams in Finland (Pajunen et al. 2019). These models showed that influential environmental variables differed between impacted and unimpacted sites, and that taxa found in both were affected by different factors according to site impactedness.

Relative importance of environmental variables

Given the high variability in model composition and explanatory power observed in the present study, we conclude that diatom taxa respond to environmental conditions in taxon-specific ways and to a variety of variables at once. Even within taxa, multiple models frequently differed in interpretation, suggesting some taxa are strongly affected by different co-occurring environmental gradients. Our findings are consistent with recent modeling efforts that have used SDMs to identify patterns in the relative importance of environmental variables across multiple diatom taxa. Pajunen et al. (2016) observed high variability in the predictability of species distributions, and Pajunen et al. (2019) found variability within and among species responses to individual environmental factors when comparing unimpacted and impacted streams at a regional scale. Together, these results suggest diatoms have context-dependent, individualistic responses to the environment.

A notable difference in our approach relative to other benthic diatom SDMs is the modeled response of taxa as presence/absence (Kumar et al. 2009; Pajunen et al. 2016, 2019) and relative abundance (the present study). SDMs are most commonly constructed using taxon presence/absence or presence only to determine potential habitat suitability (Pearson 2007). This approach is effective in assessing geographic range and niche breadth (i.e., environmental tolerances) but can provide only limited information on a taxon's realized distribution (Araujo and Guisan 2006). Modeling relative abundance, meanwhile, provides information on a taxon's estimated probability distribution and environmental optima (i.e., preferences; Gomes et al. 2018). As a result, the factors influencing taxon presence can differ from those affecting relative abundance (Howard et al. 2014).

Relative abundance-based autecological metrics and indicator species are standard in diatom-based environmental assessments because of their approximation of community structure and environmental optima (Stevenson et al. 2008). Therefore, applying relative abundance-based SDMs to diatom taxa provides consistency with commonly used assessment methods while providing integral information about taxa's optimal environmental conditions (Howard et al. 2014). Alongside recent studies by Pajunen et al. (2016, 2019), our results show that the application of SDMs to all diatom taxa in a community offers novel perspectives on the assessment of environmental factors that influence the realized niches of different diatoms.

In addition to environmental influences, diatom taxa exhibit distinct biogeographical patterns that are partially driven by spatial structuring (Soininen et al. 2004). Generally, as spatial scale increases, the importance of environmental variables decreases (Martiny et al. 2006). Taxon patterns at the regional scale tend to be influenced by combinations of local environmental, spatially structured environmental, and spatial process (Astorga et al. 2012). While integrating space-based factors may improve modeled taxon relative abundance patterns (Boulangeat et al. 2012), spatial processes generally act alongside environmental controls in an additive manner (Astorga et al. 2012). Therefore, the autecological responses of taxa along environmental gradients reported in this study stand alone as insights into the realized niches of each taxon.

Taxon clusters

While we detected high variability among taxon responses to environmental conditions, cluster analyses indicated that groups of taxa responded to similar environmental factors. An ecological guild is a group of taxa that share structural characteristics (e.g., habit), use resources in similar ways, or respond to environmental conditions in similar ways (Simberloff and Dayan 1991, Wilson 1999). In the present study, taxa were not clustered by taxonomy, morphological group, or structural characteristics but by the most influential variable in their GAMs. Therefore, the taxon groups represent guilds of shared environmental responsiveness across the regional metacommunity (Carrick et al. 1988). For example, 9 taxa were grouped together by specific conductance, suggesting a shared responsiveness to that variable across taxa. Importantly, partial effects analyses showed that taxa within each cluster did not necessarily share directional responses. In the example above, 7 taxa decreased with increasing specific conductance, while 2 decreased. Therefore, the environmental clusters identified in the present study could be further deconstructed into more restrictive guilds based on overlapping realized niches, as determined by directional responses to the dominant environmental variable. This analysis highlights the power of SDMs applied to multiple taxa to reveal novel similarities and overlapping environmental responses among taxa.

Partial effects of environmental variables

In addition to the high variability in the presence and explanatory power of each environmental variable across models and taxa, partial effects analysis showed that taxa vary widely in response to different environmental variables. While watershed-scale factors, stream geomorphology, streamflow, water temperature, and DO are commonly measured in stream ecological studies and assessments, little information exists for taxon-specific relationships to these factors at non-local scales. Therefore, it is difficult to make generalizations about taxon responses to these variables modeled by GAMs based on previous autecological understanding. For example, factors like SiO, canopy, and substrate size are measures of habitat availability for taxa with different growth habits, yet unattached, attached, epilithic, epipelic, and epipsammic taxa each demonstrated little positive or negative change with these variables. In contrast, taxa responded positively and negatively to peak flow interval (a measure of bed stability), but taxon directional responses could not be differentiated by habit or genus.

Watershed-scale factors can provide insight into climate-related effects when mean annual precipitation is a significant and explanatory component of GAMs. Recent research suggests that climatic factors such as precipitation and growing season duration may be stronger drivers of diatom taxon distributions than local variables across streams in Finland (Pajunen et al. 2016). We found support for the strong effect of climate, as measured by precipitation, at the scale of the northeast U.S. for 17 taxa, including the widespread taxa *Achnanthidium crassum* (identified together with *A. rivulare* and *A. deflexum* in the present study), *Amphora copulata*, *Cocconeis pediculus*, *Encyonema silesiacum*, *Gomphonema parvulum*, and *Nitzschia gracilis*. These taxa might be considered potential indicators for climate change and highlight the importance of watershed-scale variables on broadly distributed taxa.

Environmental optima, tolerances, and indicator species have been determined for many taxa in the U.S. relative to pH (Weilhoefer and Pan 2008), conductivity and major ions (Potapova and Charles 2003), and nutrients (Potapova and Charles 2007). The models constructed in the present study share only a few obvious similarities with these previously determined autecological attributes. For example, several taxa commonly considered to be acidophilous (*Eunotia bilunaris, E. exigua, E. minor*/*implicata, E. mucophila, E. tenella,* and *Tabellaria flocculosa*) each contained pH as a strong explanatory variable in their GAMs. Some taxa with previously determined low conductivity optima (*Gomphonema olivaceoides* var. *hutchinsoniana* and *Psammothidium helveticum*) along with taxa frequently considered to be halophilous (*Amphora ovalis, Synedra pulchella* var. *flexella*) were strongly affected by specific conductance in GAMs. However, other oligohaline taxa (*Diatoma moniliformis, Brachysira microcephala*) had no relationship with specific conductance in GAMs. Potapova and Charles (2007) identified 8 common taxa as indicators of TP or TN in the U.S., and only one of those taxa (*Achnanthidium minutissimum*/*eutrophilum*) was positively associated with nutrients (DIN) in the GAMs constructed in the present study. Of the taxa most strongly associated with TP and DIN in GAMs, none were identified as strong nutrient indicators in the Eastern Highlands or Glaciated North in Potapova and Charles (2007). Together, our results indicate novel relationships between many taxa and nutrients that may be attributable to the inclusion of other additive variables in the taxon models.

Herbicides inhibit Photosystem II activity and other cellular pathways critical to diatom metabolism (Dorigo and Leboulanger 2001). As taxa sensitive to herbicides decline, more tolerant taxa can expand in absolute and relative abundance – the process of competitive release (Pandey et al. 2017). Wood et al. (2019) compiled herbicide sensitivity data for 57 benthic stream taxa, 19 of which were detected in the present study. However, only one sensitive species (*Planothidium lanceolatum*) and two tolerant species (*Cocconeis placentula* and *Nitzschia palea*), as characterized by Wood et al. (2019), were significantly associated with herbicides in GAMs. In contrast to its delineation as sensitive, *P. lanceolatum* relative abundance increased sharply at low concentrations of total herbicides and decreased logarithmically at higher concentrations – though never as low as without herbicides. Meanwhile, the reportedly sensitive *C. placentula* relative abundance decreased dramatically with low concentrations of acetanilide and amide herbicides. Consistent with its assignment as tolerant from Wood et al. (2019), *N. palea* relative abundance increased at low concentrations of sulfonylurea and urea herbicides – potentially due to competitive release.

Several taxa were consistently affected by herbicides in GAMs across multiple models, but for all except *Cocconeis placentula* each taxon increased in relative abundance with any herbicide presence before decreasing and leveling off at higher concentrations. Therefore, most of the GAMs in the present study indicate effects of competitive release from sensitive taxa with herbicide contamination rather than identifying the sensitive taxa that decline. Previous research suggests centric and araphid diatoms are most sensitive to herbicides (Larras et al. 2014). In the present study, almost all the taxa demonstrating a positive response to herbicides in GAMs were raphid taxa. These results provide some support for past observations that herbicides have stronger toxic effects on centric and araphid taxa. Importantly, our observation of potential competitive release caused by herbicide toxicity could come from non-diatom constituents of the biofilm whose potential declines could open niche space for tolerant diatoms.

Very few diatom species sensitivity distributions have been determined for other pollutants. Like herbicides, fungicides can be toxic for diatoms by inducing oxidative stress or inhibiting photosynthesis (Rimet and Bouchez 2011, Du et al. 2019). The 7 taxa consistently responsive to fungicides increased sharply before decreasing to an asymptote, suggesting, as with herbicides, competitive release and overall tolerance rather than a dominant toxic effect.

PCBs bioaccumulate in periphyton (Hobbs et al. 2019), though the growth rates of some eukaryotic algae are not inhibited by PCBs in laboratory experiments (Kostel et al. 1999). However, PCB toxicity has been documented for a few diatom species, and toxicity increases as tolerance limits for other niche-determining factors are approached (Fisher and Wurster 1973). Five taxa were consistently responsive to modeled effects of PCBs, each of which had a differently shaped response, varying among a logarithmic increase, increase to an asymptote, ∩-shape, and exponential decrease. The most conspicuous species affected by PCBs was *Cyclotella meneghiniana*, one of the few centric diatoms common in streams, particularly those enriched in nutrients (Lowe and Kheiri 2015). However, in the one GAM for *C. meneghiniana*, only PCBs, fungicides, DO, and SiO were significant – suggesting a previously undetected relationship with organic pollutants in urban streams independent of or more dominant than previously described relationships with nutrients. The variety of taxon responses to PCBs supports the little research that has documented ranges of sensitivity to PCBs for benthic algae.

In contrast to other toxic stressors, almost all taxa affected by metals were influenced consistently across multiple models – indicating metals are a dominant stressor on those taxa that are affected in any way. Previous work has identified taxa sensitive and tolerant to total and individual metals across the U.S (Rushforth et al. 1981, Morin et al. 2012). Of 17 taxa with metals in their GAMs, only *Nitzschia dissipata* var. *media* (negative response) has been recorded as being both tolerant (Rushforth et al. 1981) and sensitive (Duong et al. 2008) to heavy metals. Importantly, we measured taxon responses to total metals, so that some variability in modeled responses may be attributable to taxon responses to different metals (Morin et al. 2012).

Applications

GAMs, like any multivariable model, are limited by the empirical observations from which they are constructed. As a result, GAMs represent biotic responses and realized niches that integrate unmeasured processes including indirect environmental effects, local biotic interactions, spatial effects, and stochasticity. Modeled relationships between taxa and environmental factors are inherently limited to the environmental gradients, taxa, and stochasticity in the sampled region. Additionally, growing evidence demonstrates diatom community and taxon patterns are context dependent, being a function of the biotic and abiotic pressures at each site, watershed, and region (Heino et al. 2012). As such, the models constructed from data in the northeast U.S. may not be directly applicable to patterns of the same taxon in other watersheds, regions, or worldwide. Importantly, we did not measure the predictive ability of our GAMs given sample size constraints. Therefore, the models produced in the present study are intended to be strictly explanatory. However, the relationships described in the present study provide novel insights into potential taxon-specific interactions with local and regional environmental conditions that can serve as context for studies within and outside the study region. While the autecological information presented in this study can be broadly used, it should be applied with caution and understanding of its limitations.

For practical application of the GAMs from this study to ecological assessments, model clusters can be used to determine taxa with similar responses to environmental conditions. Taxon-specific sensitivity and tolerance could also be estimated by the explanatory power and direction of partial effects of variables of interest for use in autecological metrics (Yuan 2004). Likewise, indicator species analyses could be conducted on those taxa most influenced by variables of interest. Importantly, the relatively low number of sites sampled in this study is insufficient for robust metric development, particularly given the high environmental variance across sites. However, the modeling framework presented here can be applied to surveys with larger sample sizes to determine regional and continental metrics. Priorities for ecological research include determining sensitivity distributions for taxa identified as responsive to herbicides, fungicides, PCBs, and metals. Integrating estimates of spatial processes like dispersal and environmental history could improve model performance and shed light on taxon-specific variability regarding environmental and spatial controls on metapopulation structure.

Species distribution modeling for all taxa within a community can be used as a tool to both better understand the relative influences of different environmental controls on diatoms and to assess taxon-specific differences in realized niches. The autecological attributes presented in the GAMs from this study provide a necessary step towards integrating multivariable effects in diatom-based ecological assessments and research.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716-723.

Araujo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, *33*, 1677-1688.

Astorga, A., Oksanen, J., Luoto, M., Soininen, J., Virtanen, R., & Muotka, T. (2012). Distance decay of similarity in freshwater communities: do macro-and microorganisms follow the same rules? *Global Ecology and Biogeography*, *21*, 365-375.

Battarbee, R. W. (1973). A new method for the estimation of absolute microfossil numbers, with reference especially to diatoms. *Limnology and Oceanography, 18*, 647-653.

Bishop, I. W., Esposito, R. M., Tyree, M., & Spaulding, S. A. (2017). A diatom voucher flora from selected southeast rivers (USA). *Phytotaxa*, *332*, 101-140.

Boulangeat, I., Gravel, D., & Thuiller, W. (2012). Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecology Letters*, *15*, 584-593.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*, 261-304.

Cantonati, M., Kelly, M. G., & Lange-Bertalot, H. (Eds.) (2017). *Freshwater Benthic Diatoms of Central Europe: Over 800 Common Species Used in Ecological Assessment*. Koeltz Botanical Books, Schmitten-Oberreifenberg.

Carrick, H. J., Lowe, R. L., & Rotenberry, J. T. (1988). Guilds of benthic algae along nutrient gradients: relationships to algal community diversity. *Journal of the North American Benthological Society*, *7*, 117-128.

Cliff, A. D., & Ord, K. (1970). Spatial autocorrelation: a review of existing and new measures with applications. *Economic Geography*, *46*, 269-292.

Coles, J. F., Riva-Murray, K., Van Metre, P. C., Button, D. T., Bell, A. H., Qi, S. L., ... & Sheibley, R. W. (2019). Design and methods of the US Geological Survey Northeast Stream Quality Assessment (NESQA), 2016 (No. 2018-1183). U.S. Geological Survey.

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

Dorigo, U., & Leboulanger, C. (2001). A pulse-amplitude modulated fluorescence-based method for assessing the effects of photosystem II herbicides on freshwater periphyton. *Journal of Applied Phycology*, *13*, 509-515.

Du, B., Zhang, Z., Liu, W., Ye, Y., Lu, T., Zhou, Z., ... & Qian, H. (2019). Acute toxicity of the fungicide azoxystrobin on the diatom *Phaeodactylum tricornutum*. *Ecotoxicology and Environmental Safety*, *168*, 72-79.

Dunn, P. K., & Smyth, G. K. (2008). Evaluation of Tweedie exponential dispersion model densities by Fourier inversion. *Statistics and Computing*, *18*, 73-86.

Duong, T. T., Morin, S., Herlory, O., Feurtet-Mazel, A., Coste, M., & Boudou, A. (2008). Seasonal effects of cadmium accumulation in periphytic diatom communities of freshwater biofilms. *Aquatic Toxicology*, *90*, 19-28.

Elith, J., & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 677-697.

Fasiolo, M., Nedellec, R., Goude, Y., & Wood, S. N. (2020). Scalable visualization methods for modern generalized additive models. *Journal of Computational and Graphical Statistics*, *29*, 78-86.

Fisher, N. S., & Wurster, C. F. (1973). Individual and combined effects of temperature and polychlorinated biphenyls on the growth of three species of phytoplankton. *Environmental Pollution*, *5*, 205-212.

Fishman, M. J. (1993). Methods of analysis by the US Geological Survey National Water Quality Laboratory: Determination of inorganic and organic constituents in water and fluvial sediments. U.S. Geological Survey.

Fishman, M. J., & Friedman, L. C. (1989). Ion chromatography, methods for determination of inorganic substances in water and fluvial Sediments. US Geological Survey Techniques of Water Resources Investigations (5A1). U.S. Geological Survey

Fitzpatrick, F. A., White, I. R., D'Arconte, P. J., Meador, M. R., Maupin, M. A., & Gurtz, M. E. (1998). Revised methods for characterizing stream habitat in the National Water-Quality Assessment Program (Report 98-4052). U.S. Geological Survey.

Genuer, R., Poggi, J. M., Tuleau-Malot, C., & Genuer, M. R. (2019). Package 'vsurf'. *Pattern Recognition Letters*, *31*, 2225-2236.

Gomes, V. H., IJff, S. D., Raes, N., Amaral, I. L., Salomão, R. P., de Souza Coelho, L., ... & Guevara, J. E. (2018). Species Distribution Modelling: Contrasting presence-only models with plot abundance data. *Scientific Reports*, *8*, 1-12.

Heino, J., Grönroos, M., Soininen, J., Virtanen, R., & Muotka, T. (2012). Context dependency and metacommunity structuring in boreal headwater streams. *Oikos*, *121*, 537-544.

Heino, J., Tolkkinen, M., Pirttilä, A. M., Aisala, H., & Mykrä, H. (2014). Microbial diversity and community–environment relationships in boreal streams. *Journal of Biogeography*, *41*, 2234-2244.

Hobbs, W. O., Collyard, S. A., Larson, C., Carey, A. J., & O'Neill, S. M. (2019). Toxic burdens of freshwater biofilms and use as a source tracking tool in rivers and streams. *Environmental Science & Technology*, *53*, 11102-11111.

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., ... & Megown, K. (2015). Completion of the 2011 National Land Cover Database for the conterminous United States–

representing a decade of land cover change information. *Photogrammetric Engineering & Remote Sensing*, *81*, 345-354.

Howard, C., Stephens, P. A., Pearce-Higgins, J. W., Gregory, R. D., & Willis, S. G. (2014). Improving species distribution models: the value of data on abundance. *Methods in Ecology and Evolution*, *5*, 506-513.

Kostel, J. A., Wang, H., Amand, A. L. S., & Gray, K. A. (1999). 1. Use of a novel laboratory stream system to study the ecological impact of PCB exposure in a periphytic biolayer. *Water Research*, *33*, 3735-3748.

Krammer, K. (2000). *Diatoms of Europe: Diatoms of the European Inland Waters and Comparable Habitats. Volume 1: The Genus Pinnularia*. ARG Gantner Varlag, Ruggell, Germany.

Kumar, S., Spaulding, S. A., Stohlgren, T. J., Hermann, K. A., Schmidt, T. S., & Bahls, L. L. (2009). Potential habitat distribution for the freshwater diatom *Didymosphenia geminata* in the continental US. *Frontiers in Ecology and the Environment*, *7*, 415-420.

Larras, F., Keck, F., Montuelle, B., Rimet, F., & Bouchez, A. (2014). Linking diatom sensitivity to herbicides to phylogeny: a step forward for biomonitoring? *Environmental Science & Technology*, *48*, 1921-1930.

Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010). Understanding of internal clustering validation measures. In *2010 IEEE International Conference on Data Mining* (pp. 911-916). IEEE.

Lowe, R., & Kheiri, S. (2015). *Cyclotella meneghiniana*. In Diatoms of North America. Retrieved November 3, 2020, from https://diatoms.org/species/cyclotella_meneghiniana

MacDonald, D. D., Ingersoll, C. G., & Berger, T. A. (2000). Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems. *Archives of Environmental Contamination and Toxicology*, *39*, 20-31.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281-297).

Mahler, B. J., Van Metre, P. C., Wilson, J. T., Musgrove, M., Zaugg, S. D., & Burkhardt, M. R. (2009). Fipronil and its degradates in indoor and outdoor dust. *Environmental Science & Technology*, *43*, 5665–5670.

Maloney, K. O., & Weller, D. E. (2011). Anthropogenic disturbance and streams: land use and land-use change affect stream ecosystems via multiple pathways. *Freshwater Biology*, *56*, 611-626.

Marra, G., & Wood, S. N. (2011). Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis*, *55*, 2372-2387.

Martiny, J. B. H., Bohannan, B. J., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., ... & Morin, P. J. (2006). Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, *4*, 102-112.

Moran, P. W., Nowell, L. H., Kemble, N. E., Mahler, B. J., Waite, I. R., & Van Metre, P. C. (2017). Influence of sediment chemistry and sediment toxicity on macroinvertebrate communities across 99 wadable streams of the Midwestern USA. *Science of the Total Environment, 599,* 1469-1478.

Morin, S., Cordonier, A., Lavoie, I., Arini, A., Blanco, S., Duong, T. T., ... & Laviale, M. (2012). Consistency in diatom response to metal-contaminated environments. In Guasch, H., Ginebreda, A., Geiszinger, A. (Eds.), *Emerging and Priority Pollutants in Rivers* (pp. 117-146). Springer, Berlin, Germany.

O'Dell, J. W. (1993). *EPA Method 365.1—Determination of phosphorus by semi-automated colorimetry*. U.S. Environmental Protection Agency, Office of Research and Development.

Pajunen, V., Jyrkänkallio-Mikkola, J., Luoto, M., & Soininen, J. (2019). Are drivers of microbial diatom distributions context dependent in human-impacted and pristine environments? *Ecological Applications*, *29,* e01917.

Pajunen, V., Luoto, M., & Soininen, J. (2016). Climate is an important driver for stream diatom distributions. *Global Ecology and Biogeography, 25*, 198-206.

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

Pandey, L. K., Bergey, E. A., Lyu, J., Park, J., Choi, S., Lee, H., ... & Han, T. (2017). The use of diatoms in ecotoxicology and bioassessment: insights, advances and challenges. *Water Research*, *118,* 39-58.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20,* 289-290.

Patrick, R. (1973). Use of algae, especially diatoms, in the assessment of water quality. In Cairns, J., Dickson, K. L. (Eds.), *Biological Methods for the Assessment of Water Quality* (pp. 76-95). ASTM International, Philadelphia, PA.

Patton, C. J., & Kryskalla, J. R. (2011). Colorimetric determination of nitrate plus nitrite in water by enzymatic reduction, automated discrete analyzer methods. *U.S. Geological Survey Techniques and Methods, 5,* B8.

Pearson, R. G. (2007). Species' distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History, 50*, 54-89.

Pickett, S. T., Cadenasso, M. L., Grove, J. M., Boone, C. G., Groffman, P. M., Irwin, E., ... & Pouyat, R. V. (2011). Urban ecological systems: Scientific foundations and a decade of progress. *Journal of Environmental Management*, *92*, 331-362.

Potapova, M., & Charles, D. F. (2003). Distribution of benthic diatoms in US rivers in relation to conductivity and ionic composition. *Freshwater Biology*, *48*, 1311-1328.

Potapova, M., & Charles, D. F. (2007). Diatom metrics for monitoring eutrophication in rivers of the United States. *Ecological Indicators*, *7*, 48-70.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org/.

Radtke, D. B. (2005). *Bottom-material samples: U.S. Geological Survey Techniques of Water-Resources Investigations* (Vol. 9, No. A8). U.S. Geological Survey.

Ram, K., & Wickham, H. (2018). wesanderson: a Wes Anderson palette generator. R package version 0.3.6. Available at https://cran.r-project.org/package=wesanderson.

Rimet, F., & Bouchez, A. (2011). Use of diatom life-forms and ecological guilds to assess pesticide contamination in rivers: lotic mesocosm approaches. *Ecological Indicators*, *11*, 489-499.

Roubeix, V., Mazzella, N., Schouler, L., Fauvelle, V., Morin, S., Coste, M., ... & Margoum, C. (2011). Variations of periphytic diatom sensitivity to the herbicide diuron and relation to species distribution in a contamination gradient: implications for biomonitoring. *Journal of Environmental Monitoring*, *13*, 1768-1774.

Rushforth, S. R., Brotherson, J. D., Fungladda, N., & Evenson, W. E. (1981). The effects of dissolved heavy metals on attached diatoms in the Uintah Basin of Utah, USA. *Hydrobiologia*, *83*, 313-323.

Sandstrom, M. W., Kanagy, L. K., Anderson, C. A., & Kanagy, C. J. (2016). Determination of pesticides and pesticide degradates in filtered water by direct aqueous-injection liquid chromatography-tandem mass spectrometry (No. 5-B11). U.S. Geological Survey.

Shelton, L. R., & Capel, P. D. (1994). Guidelines for collecting and processing samples of stream bed sediment for analysis of trace elements and organic contaminants for the National Water-Quality Assessment Program (Vol. 94, No. 458). U.S. Geological Survey.

Simberloff, D., & Dayan, T. (1991). The guild concept and the structure of ecological communities. *Annual Review of Ecology and Systematics*, *22*, 115-143.

Soininen, J. (2007). Environmental and spatial control of freshwater diatoms—a review. *Diatom Research*, *22*, 473-490.

Soininen, J., & Teittinen, A. (2019). Fifteen important questions in the spatial ecology of diatoms. *Freshwater Biology*, *64*, 2071-2083.

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

Soininen, J., Jamoneau, A., Rosebery, J., Leboucher, T., Wang, J., Kokociński, M., & Passy, S. I. (2019). Stream diatoms exhibit weak niche conservation along global environmental and climatic gradients. *Ecography, 42*, 346-353.

Soininen, J., Paavola, R., & Muotka, T. (2004). Benthic diatom communities in boreal streams: community structure in relation to environmental and spatial gradients. *Ecography, 27,* 330-342.

Spaulding, S. A., Bishop, I. W., Edlund, M. B., Lee, S., Furey, P., Jovanovska, E., & Potapova, M. (2019). Diatoms of North America. Available at https://diatoms.org/.

Stevenson, R. J., Pan, Y., Manoylov, K. M., Parker, C. A., Larsen, D. P., & Herlihy, A. T. (2008). Development of diatom indicators of ecological conditions for streams of the western US. *Journal of the North American Benthological Society*, *27,* 1000-1016.

Tweedie, M. C. (1984). An index which distinguishes between some important exponential families. In Ghosh, J. K., Roy, J. (Eds.), *Statistics: Applications and new directions: Proc. Indian statistical institute golden Jubilee International conference* (pp. 579-604). Indian Statistical Institute, Calcutta.

Tyree, M. A., Bishop, I. W., Hawkins, C. P., Mitchell, R., & Spaulding, S. A. (2020). Reduction of taxonomic bias in diatom species data. *Limnology and Oceanography: Methods, 18*, 271-279.

U.S. Geological Survey. (2019). BioData—Aquatic bioassessment data for the nation available on the world wide web. Accessed October 10, 2019. doi:10.5066/F77W698B

Vu, V. Q. (2011). ggbiplot: A ggplot2 based biplot. R package version 0.5. Available at http://github.com/vqv/ggbiplot.

Wagner, R. J., Moran, P. W., Zaugg, S. D., Sevigny, J. M., & Pope, J. M. (2016). Contaminants of emerging concern in the lower Stillaguamish River Basin, Washington, 2008–11 (Open-File Report 2014-1028). U.S. Geological Survey.

Weilhoefer, C. L., & Pan, Y. (2008). 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. *Hydrobiologia*, *614*, 259-274.

Wilson, J. B. (1999). Guilds, functional types and ecological groups. *Oikos*, 507-522.

Wood, R. J., Mitrovic, S. M., Lim, R. P., Warne, M. S. J., Dunlop, J., & Kefford, B. J. (2019). Benthic diatoms as indicators of herbicide toxicity in rivers–A new SPEcies At Risk (SPEARherbicides) index. *Ecological Indicators*, *99*, 203-213.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, *99,* 673-686.

Wood, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, *62*, 1025-1036.

Wood, S. N. (2020). mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. R package version 1.8-33. Available at https://cran.r-project.org/package=mgcv.

Yee, T. W., & Mitchell, N. D. (1991). Generalized additive models in plant ecology. *Journal of Vegetation Science*, *2*, 587-602.

Yuan, L. L. (2004). Assigning macroinvertebrate tolerance classifications using generalised additive models. *Freshwater Biology*, *49*, 662-677.

CHAPTER IV

THE LIFE AQUATIC: MULTIVARIABLE EFFECTS OF ENVIRONMENTAL STRESSORS ON BENTHIC CYANOBACTERIAL PRESENCE AND RELATIVE ABUNDANCE ACROSS HUMAN-IMPACTED STREAMS

Abstract

Cyanobacteria are expanding in distribution and abundance in freshwater bodies worldwide. These microorganisms can present a risk to human health and are sensitive to local environmental conditions and watershed characteristics. Most research relating cyanobacteria to physicochemical conditions in streams has considered community and taxon responses to single or a limited number of environmental stressors. Few studies have investigated the extent to which contaminants such as herbicides, persistent organic pollutants, and metals function alongside other known environmental stressors at large spatial scales to affect the presence/absence and relative abundance of cyanobacterial taxa. We measured watershed topography, geology, and climate alongside stream geomorphology, hydrology, pH, specific conductance, nutrients, and other contaminants from 92 sites along an urban gradient in the northeast United States. We used generalized additive modeling to identify the suites of watershed-scale factors and in-stream stressors that most influenced the distributions and relative abundances of 13 cyanobacterial genera as identified by microscopy. Model deviance explained varied from 15% to 97%, with a mean of 44%. We found that each genus responded to a distinct suite of environmental variables, with light availability, water temperature, and nutrients having little to no modeled effect on each genus. Low concentrations of herbicides increased the modeled presence and/or relative abundances of *Calothrix*, *Chroococcus*, *Leptolyngbya*, *Merismopedia*, and *Phormidium*, with no effect at high concentrations. For these genera, herbicides appear to have a threshold effect of potential competitive release from the loss of less tolerant taxa. Increased polychlorinated biphenyl concentrations decreased the likelihood of

Calothrix presence and the relative abundances of *Homoeothrix* and *Phormidium*, demonstrating novel toxic effects of persistent organic pollutants on cyanobacterial genera. For half of the modelled genera, presence/absence and relative abundance of the same genus were explained by different environmental factors, indicating the need for different management strategies based on the objective of cyanobacterial control. Overall, we found that multivariable modeling can be used to uncover previously undetected relationships between benthic cyanobacterial taxa and infrequently measured contaminants. Additionally, taxon-specific, multivariable models expand upon recent work describing strong relationships between cyanobacteria and commonly measured stressors. This research contributes to our growing understanding of the environmental conditions that affect potentially harmful or nuisance cyanobacteria.

Introduction

Bloom-forming and toxin-producing benthic cyanobacteria are increasing in range and abundance across freshwater streams, rivers, ponds, and lakes worldwide (Wood et al. 2020). These cyanobacteria can form harmful algal blooms (cyanoHABs) that disturb food webs and disrupt ecosystem structure and function (Paerl et al. 2001). Additionally, cyanoHABs can produce neurotoxins, hepatotoxins, cytotoxins, and endotoxins that present a risk to human health by degrading drinking water quality, contaminating agricultural products, and restricting recreational opportunities (Carmichael 2001). In order to assess freshwater cyanoHAB proliferation, it is critical to understand how local, in-stream stressors and watershed characteristics affect the distribution and abundance of cyanobacteria.

The spread of harmful benthic cyanobacteria in freshwater streams has been associated with alterations to water temperature, nutrient concentrations, sediment deposition, streamflow dynamics, light availability, and substrate composition (McAllister et al. 2018). However, our understanding of the physicochemical factors that affect benthic cyanobacterial distribution and abundance is limited to a few taxa, habitats, and geographic regions (Wood et al. 2020). Growing evidence suggests that some cyanobacteria are affected by less well-studied contaminants such as herbicides (Powell et al. 1991, Pérez et al. 2007), persistent organic pollutants (Harris and Smith 2016), and metals (Kupper and Kroneck 2005). However, these studies infrequently consider cyanobacteria in streams.

Environmental stressors have additive and synergistic effects on aquatic microorganisms, so that measuring only univariate responses excludes potentially important biotic-environmental interactions (James and McCulloch 1990). While some researchers have investigated the effects of multiple, coincident stressors on cyanobacterial distributions (Wood et al. 2017, McAllister et al. 2018), few have considered a broad suite of potential stressors alongside watershed characteristics that each vary across large spatial scales and environmental gradients. Species distribution modeling is a non-linear, multivariable modeling approach that provides the opportunity to identify the relative influence of different environmental characteristics on individual taxon presence and abundance (Pearson 2007).

The extent and severity of pollution and physical habitat alteration in streams have increased worldwide near urban and agricultural areas. In the United States, nutrient and road salt loading to streams have increased nationwide for over a century, particularly in the northeast (Van Sickle et al. 2004, Dodds and Smith 2016, Kaushal et al. 2018). Stream discharge has decreased and flow intermittency has increased nationally because of climate change and human water use (Lake 2003). Surface water temperature has increased by up to 0.08ºC per year for streams in the U.S. since the mid-1900s (Kaushal et al. 2010). Since the mid-1990s, herbicide applications to agricultural fields have increased over 15-fold (Benbrook 2016). While polychlorinated biphenyls (PCBs), a persistent organic pollutant, were banned for use in 1979, PCBs bioaccumulate and are commonly detected in stream fish tissue across the U.S. today in concentrations that can exceed human health standards (Batt et al. 2017). These alterations of environmental conditions have the potential to affect benthic cyanobacteria through changes in geographic ranges, increased production of biomass, and synthesis of toxins (Scott and Marcarelli 2012, Wood et al. 2020).

Environmental conditions can affect the presence and abundance of taxa differently. Assessments and management of cyanobacteria in streams are based upon the detection of potential invasion, identification of potential blooms, or both. Taxon presence/absence is a measure of taxon occupancy across space and environmental conditions, which is used to assess environmental tolerances (niche breadth) and invasion potential (Heino and Soininen 2006). In contrast, relative abundance can be used as an estimate of optimal environmental conditions for population growth and potential for bloom formation (McCormick and Stevenson 1991). Each of these methods of characterizing cyanobacteria is most often analyzed in isolation. Understanding how environmental influences on taxon presence/absence and relative abundance differ is important to inform the management of geographic spread and blooms of cyanobacteria.

We used multivariable generalized additive modeling to characterize the relationships between cyanobacteria and environmental conditions from biofilms collected from 92 streams in the northeast U.S. We aimed to (1) determine the suites of environmental factors that affect the presence/absence and relative abundances of benthic cyanobacterial genera along gradients of human impact and (2) assess how environmental factors influence genus presence/absence versus relative abundance. We highlight the responses of potentially toxigenic cyanobacteria because of their relevance to human health and environmental management. We also emphasize the influences of herbicides and persistent organic pollutants because of their overall lack of study regarding cyanobacteria in streams.

Methods

Sample collection

In summer 2016, 92 wadeable stream sites were sampled across the northeast U.S. (150,000 km^2) as part of the U.S. Geological Survey (USGS) National Water-Quality Assessment (NAWQA) Northeast Stream Quality Assessment (NESQA; Coles et al. 2019; Figure 1). Sites were selected along urban and agricultural gradients according to watershed land use from the 2011 National Land Cover Database (NLCD; Homer et al. 2015). Most sites were sampled in watersheds in which urban development was the dominant human land use, ranging from 1 – 93% urban land cover in sites with < 5% agricultural land cover.

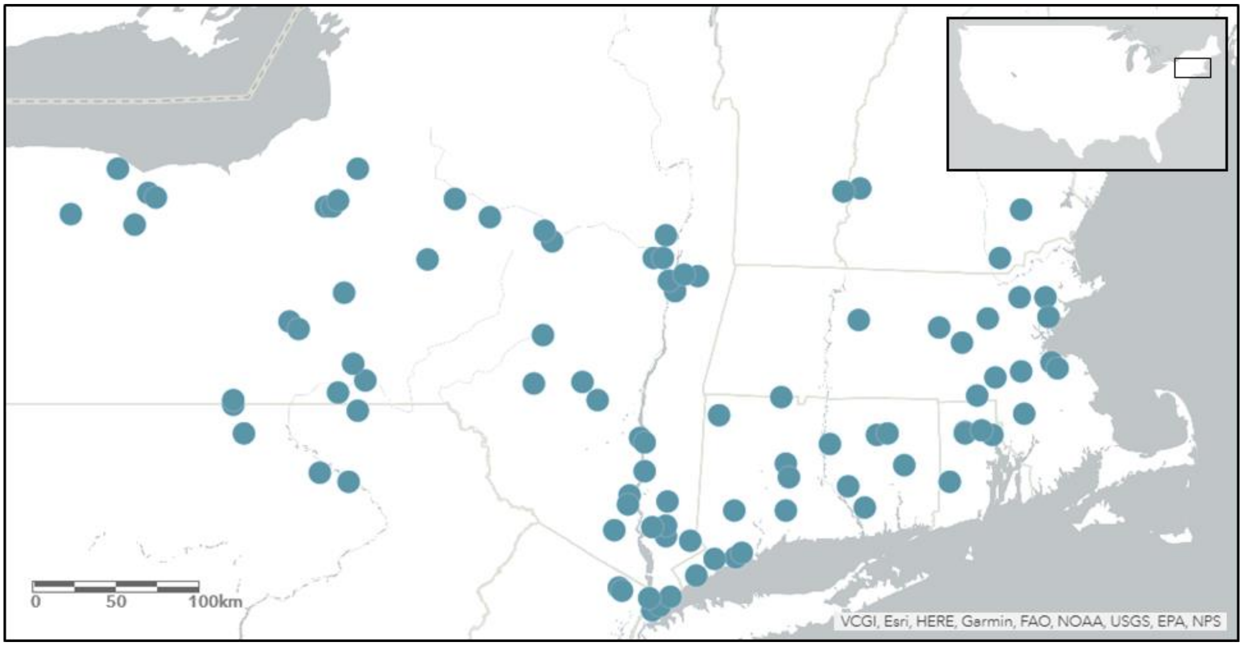


Figure 1. Map of sites in the northeast United States used in the analysis.

Detailed environmental sampling and analytical methods for NESQA are described in Chapter III and in Coles et al. (2019). Briefly, for each site, watershed-scale factors were measured from NLCD data, including drainage area, elevation, precipitation, runoff, and geology. Local scale, in-stream variables were assessed at different time points depending upon the target measurement. Streamflow and water temperature were measured continuously for up to one year before all other sampling was completed. Between 4 to 9 weekly surface water collections were analyzed for dissolved oxygen (DO), pH, specific conductance, nutrients (total phosphorus [TP] and dissolved inorganic nitrogen [DIN]), herbicides, and fungicides. Pooled concentrations of 7 metals (As, Cd, Cr, Cu, Pb, Ni, Zn) and PCBs were measured from one-time collections of streambed sediment, and probable effect concentration quotients (PECQs) were calculated for each (MacDonald et al. 2000). During final site visits, stream geomorphology was characterized as canopy cover, channel width:depth, riffle percentage, and substrate particle size across 11 transects spaced 15 m apart along a 150-m stream reach. Additionally, 2-cm^2 rock scrapes for biofilms were collected at each transect and combined. Composite biofilm samples were homogenized, subsampled with known volumes, and preserved in 2.5% formalin for microscopic identification of cyanobacteria. Environmental data are available at the NESQA website (https://webapps.usgs.gov/rsqa/#!/region/NESQA).

Cyanobacterial identification

Eukaryotic algae were identified to phylum and cyanobacteria to genus using 400x magnification light microscopy in Palmer-Maloney counting chambers. Taxa were quantified by identification and enumeration of 300 natural units (Moulton et al. 2002), with cell number concurrently recorded for each individual. All algal scientific names and authorship follow USGS BioData, and all algal count data are archived in the BioData database (U.S. Geological Survey 2019). Composite biofilm samples are not archived.

For genus-specific modeling, the cell number of each cyanobacterial genus was normalized by the cell number of all cyanobacterial genera combined – excluding eukaryotic algal cell counts. We report cyanobacterial abundance relative to only the cyanobacterial community, and cyanobacterial genus-specific counts are also presented as presence/absence. We only considered cyanobacterial counts relative to eukaryotic algal counts outside of modeling to highlight the samples in which cyanobacteria were more abundant than eukaryotic algae.

Generalized additive modeling

The procedure used to construct generalized additive models (GAMs) is described in Chapter III. Briefly, 27 uncorrelated (Spearman $\rho < |0.70|$) watershed- and local-scale variables were selected for inclusion in models (Table 1). Watershed-scale variables were drainage area, elevation, mean annual precipitation, runoff, sedimentary geology, percent SiO, and soil permeability. Local, in-stream variables were grouped into 9 variable classes: stream geomorphology, stream discharge, water temperature, DO, pH, ions, nutrients, herbicides, and other pollutants (fungicides, PCBs, and metals). In total, variables were assigned to 10 classes (watershed-scale factors as one class and the 9 local-scale classes). The median value was used for all variables with multiple measurements. Variables were transformed to reduce skewness and kurtosis. To verify that the selected variables represented sufficient environmental variability across sites, principal components analysis (PCA) was performed.

GAMs were constructed for each genus for both presence/absence and relative abundance datasets using an all combinations approach (Wood 2006). First, a subset of important variables for each genus was determined using random forest mean variable importance values. All one-, two-, three-, and four-variable combinations of these variables were formed, and each combination of explanatory variables was used to construct a GAM. Each GAM was penalized to avoid overfitting and oversmoothing. Binomial and Tweedie distributions were used to construct GAMs for genus presence/absence and relative abundance, respectively.

Table 1. Descriptive statistics for each of the 27 environmental variables used in models. For each variable, correlation with one of the principal components explaining > 5% of inter-sample variation is listed.

Class	Variable	Min	Med	Mean	Max	PC	Loading
Watershed	Drainage area (km^2)	6.1	67.27	75.14	263.26	PC5	-0.37
	Elevation (m)	31.89	158.9	216.07	708.49	PC1	-0.3
	Mean annual precipitation (mm)	85.2	120.9	116.41	135.1	PC2	0.41
	Mean annual runoff (mm)	264.8	359.9	369.65	547.75	PC6	-0.34
	Sedimentary geology (%)	0	100	78.74	100	PC2	-0.3
	SiO (%)	32.82	62.67	60.52	75.03	PC2	0.33
	Soil permeability (%)	0	9	11.14	80	PC6	0.35
Geomorphology	Canopy (%)	0.53	70.05	63.99	100	PC6	0.41
	Channel width:depth	4.23	13.65	15.02	69.02	PC3	-0.28
	Riffles (%)	0	28.57	30.21	100	PC3	-0.29
	Substrate D50 (%)	0.1	4.63	5.32	25	PC3	-0.25
Discharge	Base flow index (%)	34.96	47.67	48.64	62.43	PC4	0.28
	Peak flow interval (days)	4.63	17.09	19.81	100	PC5	-0.27
Temperature	Water temperature (°C)	15.5	21.7	21.63	27.16	PC3	-0.34
DO	Dissolved oxygen (DO) (mg L^{-1})	2.99	8.21	7.95	9.89	PC5	0.4
pH	pH	6.23	7.61	7.56	8.72	PC2	-0.38
Ions	Specific conductance (µS cm^{-1})	55.5	476.6	612.64	2834	PC1	0.33
Nutrients	Dissolved inorganic nitrogen (DIN) (mg L^{-1})	0	0.34	0.67	7.46	PC1	0.33
	Total phosphorus (TP) (mg L^{-1})	0	0.04	0.09	2.07	PC4	-0.33
	Total nitrogen:total phosphorus (TN:TP)	3	15.24	21.05	155.34	PC4	0.56
Herbicides	Acetanilide and amide herbicides (ng L^{-1})	0	0	35.14	501.25	PC2	-0.36
	Acid herbicides (ng L^{-1})	0	0	18.75	256.62	PC1	0.28
	Sulfonylurea and urea herbicides (ng L^{-1})	0	0	5.26	65.59	PC3	-0.28
	Total herbicides (ng L^{-1})	0	35.58	104.53	872.6	PC2	-0.29
Pollutants	Total fungicides (ng L^{-1})	0	0	42.99	1535.29	PC1	0.33
	Polychlorinated biphenyls (PCBs) (PECQ)	0	0	0.01	0.87	PC1	0.23
	Total metals (PECQ)	0.01	0.12	0.17	1.3	PC3	-0.33

Min = minimum; Med = median; Max = maximum

PC = principal component explaining > 5% of inter-sample variation onto which the variable has the highest loading (from principal components analysis)

Loading = loading (Pearson’s *r*) between the variable and the principal component in column PC

PECQ = probable effect concentration quotient

Nonsignificant, unexplanatory, and poorly performing GAMs were removed from the resultant pool of models. Only GAMs with all significant explanatory variables, model deviance explained ≥ 15%, no spatial autocorrelation (i.e., nonsignificant Moran's I), no concurvity, differences in Akaike Information Criterion (ΔAIC) ≤ 2 from the top performing model, and residual plots with more or less random scatter around $y = 0$ were used. Akaike weights were calculated for all models prior to removing all models with ΔAIC > 2. For each model passing the above criteria, the deviance explained by each explanatory variable was estimated by calculating the reduction in overall model deviance explained when fitting alternative models without each term.

Heatmaps were constructed to visualize model composition, performance, and explanatory power. Partial effects plots were created to represent the effect of each modeled variable on the genus response (presence/absence or relative abundance), where $y = 0$ is the mean genus response. All analyses were performed in R v. 4.0.2 (R Core Team 2020) using the packages 'prcomp,' 'ggbiplot' (Vu 2011), 'VSURF' (Genuer et al. 2019), 'mgcv' (Wood 2020), 'ape' (Paradis et al. 2004), 'mgcViz' (Fasiolo et al. 2020), and 'wesanderson' (Ram and Wickham 2018).

Results

Environmental characteristics

Surveyed sites were distributed across gradients of watershed land use: 0 – 93% urban (mean 27%) and 0 – 77% agricultural (mean 13%). Drainage area averaged 75 km^2 (range 6 – 263 km^2) across elevations from 32 – 708 m (mean 216 m) (Table 1). Mean annual precipitation varied between 85 – 135 cm (mean 116 cm). Most sites were characterized by sedimentary geology. Streams varied in canopy cover (1 – 100%), size (channel width:depth between 4 – 69), and bed composition (0 – 100% riffles).

In-stream stressors spanned ranges that represented large gradients over which cyanobacterial response could be measured (Table 1). Time since peak water flow, a measure of the interval of likely bed sediment mobilization, varied from 4 to 100 days. Dissolved oxygen ranged between 2.99 – 9.89 mg L^{-1}. Most sites were circumneutral (mean pH 7.56), but some sites were slightly acidic (minimum 6.23) and others alkaline (maximum 8.72). Specific conductance was generally high, with 29% of sites characteristic of oligohaline conditions (> 800 µS cm^{-1}). Overall nutrient concentrations were low, with TP and DIN averaging 0.67 mg L^{-1} and 0.09 mg L^{-1}, respectively. However, two sites exceeded 1.4 mg L^{-1} TP, and four sites were over 3.9 mg L^{-1} DIN. Herbicides were detectable in 82% of streams and varied between 1.01 – 873 ng L^{-1} where present (total herbicides). Acetanilide and amide herbicides were the most common and concentrated class of herbicides. PCBs were measurable in 88% of sites, with one site containing dramatically higher concentrations than the rest (0.87 ng L^{-1} compared to the mean of 0.01 ng L^{-1}). Pooled metal concentrations varied across two orders of magnitude from 0.01 – 1.30 ng L^{-1}.

In the PCA, each environmental factor loading was between |0.23| - |0.56| on at least one of the 6 principal components that explained ≥ 5% of inter-site variability (Tables 1 and S1, Figure S1).

Genus frequency and relative abundance

Eighteen cyanobacterial genera were identified, 11 of which contain potentially toxigenic species (Table 2). *Chamaesiphon*, *Homoeothrix*, and *Phormidium* were the most common genera with the highest relative abundances based on cell number, while *Anabaena*, *Komvophoron*, and

Spirulina were the least common with the lowest relative abundance. The number of genera in each sample ranged between 0 – 7 (mean of 4). Mean genus frequency varied from 1 – 70% of sites and mean relative abundance (of the sites where present) from 0.3 – 52%. Mean frequency and mean relative abundance were positively correlated (Spearman's $\rho = 0.75$, $p < 0.001$).

Table 2. Cyanobacterial genus frequency (percent of sites detected) and mean relative abundance (from sites where present, calculated from cell number).

	Mean frequency (%)	Mean relative abundance where present (%)	Contains potential benthic toxigenic species (Y/N)[+]	Approximate number of benthic species in North America*	Approximate number of species worldwide^	Morphological group
Anabaena	4.35	5.6	Y	5 (~)	80	Filamentous
Aphanocapsa±	21.74	19.52	N	5	50	Coccoid
Aphanothece	3.26	11.21	Y	7	40	Coccoid
Calothrix	23.91	20.48	Y	36	60	Filamentous
Chamaesiphon±	63.04	22.12	N	4	30	Coccoid
Chroococcus	5.43	3.38	N	7	60	Coccoid
Heteroleibleinia	26.09	35.48	N	5 (~)	30	Filamentous
Homoeothrix	69.57	52.14	N	3 (~)	5 (~)	Filamentous
Komvophoron±	3.26	0.89	N	5 (~)	15	Filamentous
Leptolyngbya	16.3	39.15	Y	10 (~)	140	Filamentous
Lyngbya±	1.09	13.89	Y	30	60	Filamentous
Merismopedia	9.78	3.18	N	6	30	Coccoid
Phormidium	55.43	20.5	Y	20 (~)	30	Filamentous
Pleurocapsa	31.52	18.81	Y	5 (~)	20	Coccoid
Pseudanabaena±	23.91	17.68	Y	5 (~)	30	Filamentous
Schizothrix	5.43	10.52	Y	5	70	Filamentous
Spirulina	4.35	0.31	Y	5 (~)	50	Filamentous
Synechococcus	11.96	4.41	Y	5 (~)	30	Coccoid

[+] From Wood et al. (2020)
*From Komárek and Johansen (2015a, b)
^Inclusive of freshwater and marine, benthic and planktonic. From Komárek and Johansen (2015a, b)
(~) refers to less confidence about species numbers because of lack of study or taxonomic revision
Genus± = No models passed quality filtering

There was no clear pattern among genus frequency, mean relative abundance, morphological group (coccoid or filamentous), and the number of species per genus known to be found in the benthos of North American freshwater bodies (Komárek and Johansen 2015a, b;

Table 2). Only three genera with > 10 benthic species in North America were detected: C*alothrix*, *Lyngbya*, and *Phormidium*. While *Calothrix* and *Phormidium* had high frequency and mean relative abundances that could be associated with potentially high numbers of species, other genera with < 5 North American benthic species were also common with high relative abundances: *Aphanocapsa*, *Chamaesiphon*, *Heteroleibleinia*, *Homoeothrix*, and *Pleurocapsa*.

Most samples were dominated by diatoms, in terms of natural algal units, and by cyanobacteria, in terms of algal cell number (Table S2). With the exception of the relatively common and abundant *Homoeothrix*, most of the cyanobacterial genera that were abundant in the present study are characterized by small, many celled species (Komárek and Johansen 2015a, b; Table 2). Therefore, relative abundance based on natural units may best represent the relative dominance by algal groups. In only three sites did cyanobacterial natural units exceed those of diatoms. Of these, one site was dominated by *Heteroleibleinia* and *Leptolyngbya* (55% of all natural algal units), another by *Homoeothrix* (50%), and the other by *Phormidium* (42%).

Environmental factors influencing genera

Aphanocapsa, *Chamaesiphon*, *Komvophoron*, *Lyngbya*, and *Pseudanabaena* had no models that passed quality filtering. Of the modellable genera, 10 were modeled for presence/absence and 12 for relative abundance (Figures 2 and 4). A total of 44 models passed quality filtering (15 for presence/absence, 29 for relative abundance).

Variables within each environmental class explained a high relative proportion of deviance for the presence/absence or relative abundance of at least one genus. Variables or variable classes did not cluster together in any clear patterns across genera. Only the worst performing model for *Synechococcus* relative abundance contained the latitude/longitude term. No other models were affected by significant spatial autocorrelation. Across all genera, mean model deviance explained was not significantly correlated with mean genus frequency (Spearman's ρ = -0.19, *p* = 0.60) or relative abundance (Spearman's ρ = -0.28, *p* = 0.43).

Presence/absence

Model deviance explained for genus presence/absence varied between 15% (the lowest bound allowed to pass quality filtering) and 63%, with a median of 29% and mean of 33% (Figure 2). The two models for *Anabaena* had the highest deviance explained (mean of 59%) of all presence/absence models. Models for *Phormidium*, *Leptolyngbya*, and *Aphanothece* had the lowest deviance explained (15 – 17%). Multiple similarly performing models for *Anabaena*, *Chroococcus*, and *Heteroleibleinia* each had at least one explanatory variable in common.

For cases in which the same variable affected multiple genera, the general shape of the partial effect of that variable was approximately the same among genera (Figure 3). There were some exceptions to this pattern for drainage area, elevation, and mean annual precipitation. In general, the modeled presence of each genus was most strongly affected by one or two variables, as determined by consistent occurrence across multiple models and high deviance explained compared to other variables (Figure 2). Of the in-stream stressors, canopy cover, percent riffles, water temperature, DIN, TP, and acetanilide and amide herbicides did not have a dominant influence on any genera. Only two genera were affected strongly by the same class of variables: *Calothrix* and *Pleurocapsa* by pH (opposite directional responses) and *Chroococcus* and *Leptolyngbya* by herbicides (same responses).

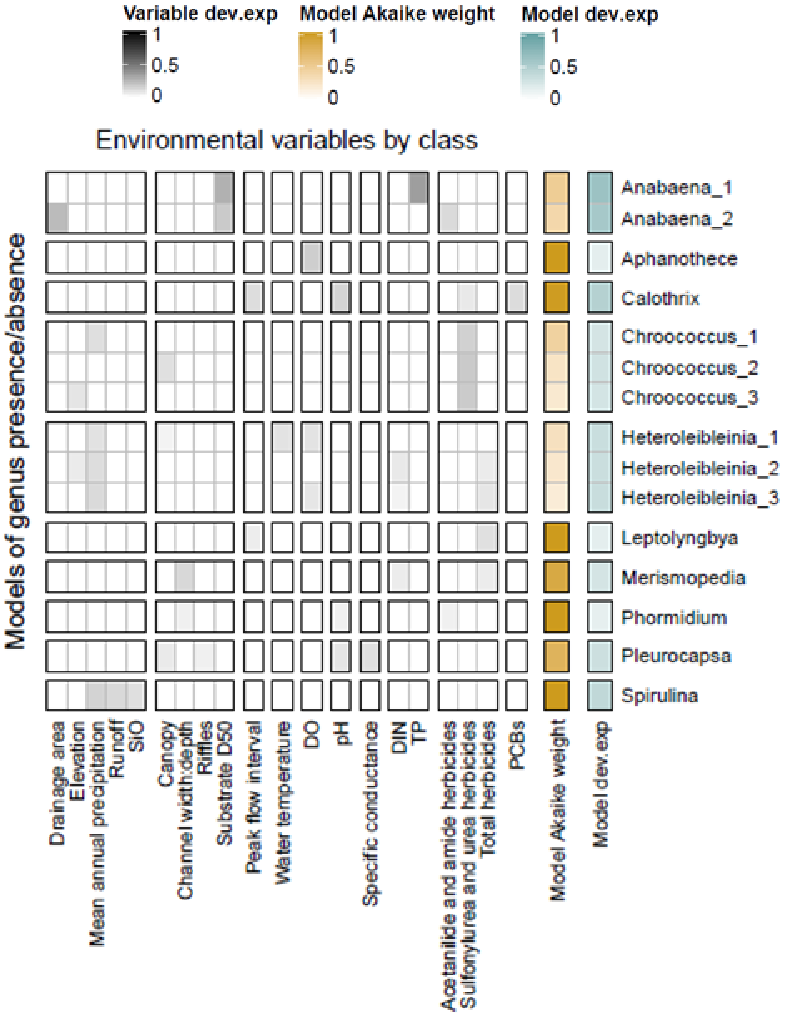


Figure 2. Variable deviance explained (dev.exp), Akaike weights, and model deviance explained of filtered generalized additive models for genus presence/absence. Multiple models for the same genus are sorted in descending order by Akaike weight.

In addition to the effects of their dominant variables, the modeled presence of *Calothrix*, *Chroococcus*, *Leptolyngbya*, *Merismopedia*, and *Phormidium* each increased at low concentrations of herbicides before leveling off at higher concentrations (Figure 3). Additionally, *Calothrix* presence was also modeled to decrease with increasing PCB concentrations.

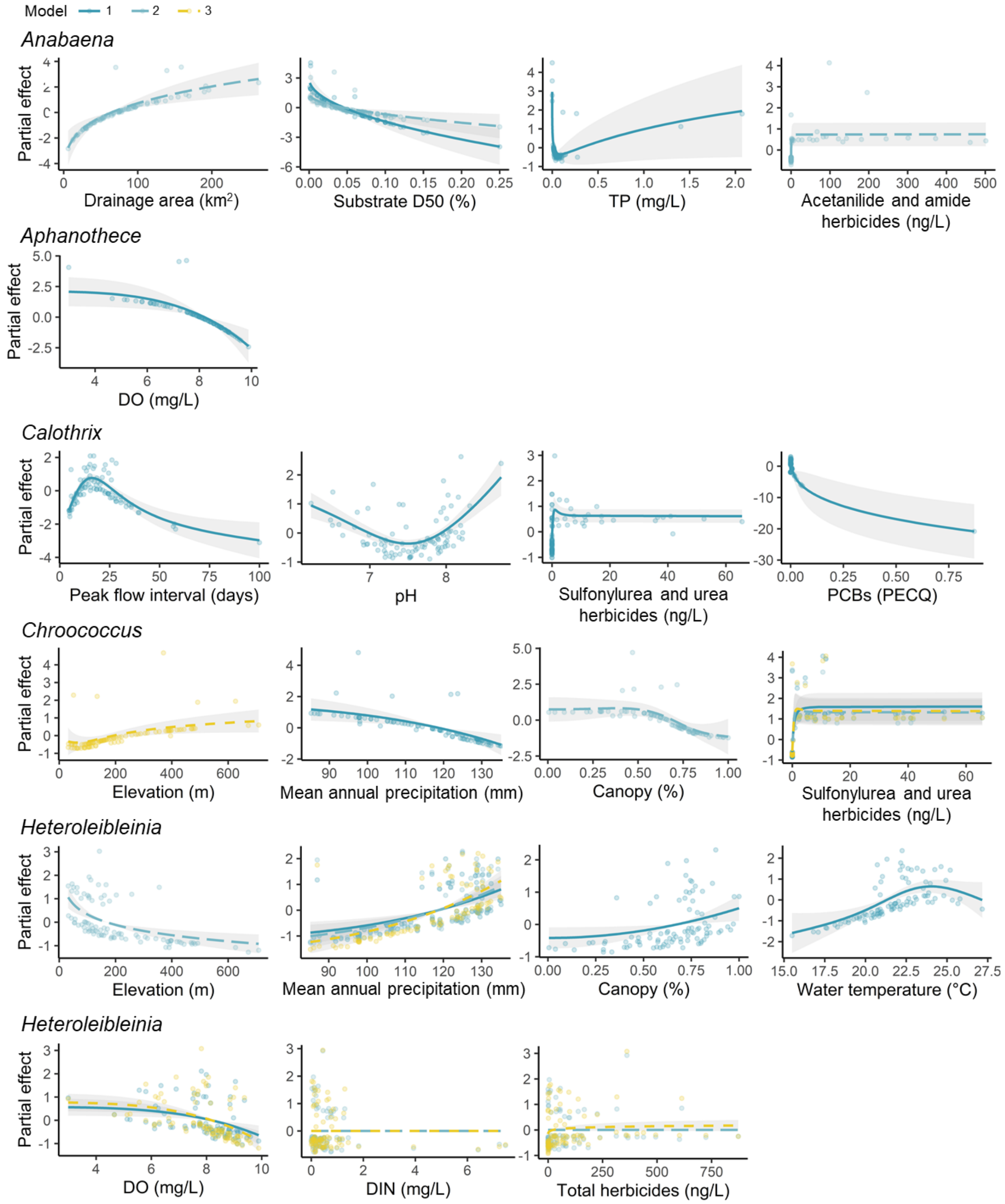


Figure 3. Partial plots of the effects of environmental variables on the presence/absence of cyanobacterial genera. Models containing the same environmental variable for the same genus are represented on a single plot and correspond to the model numbers in Figure 2. Variable abbreviations match those in Table 1. (1 of 2)

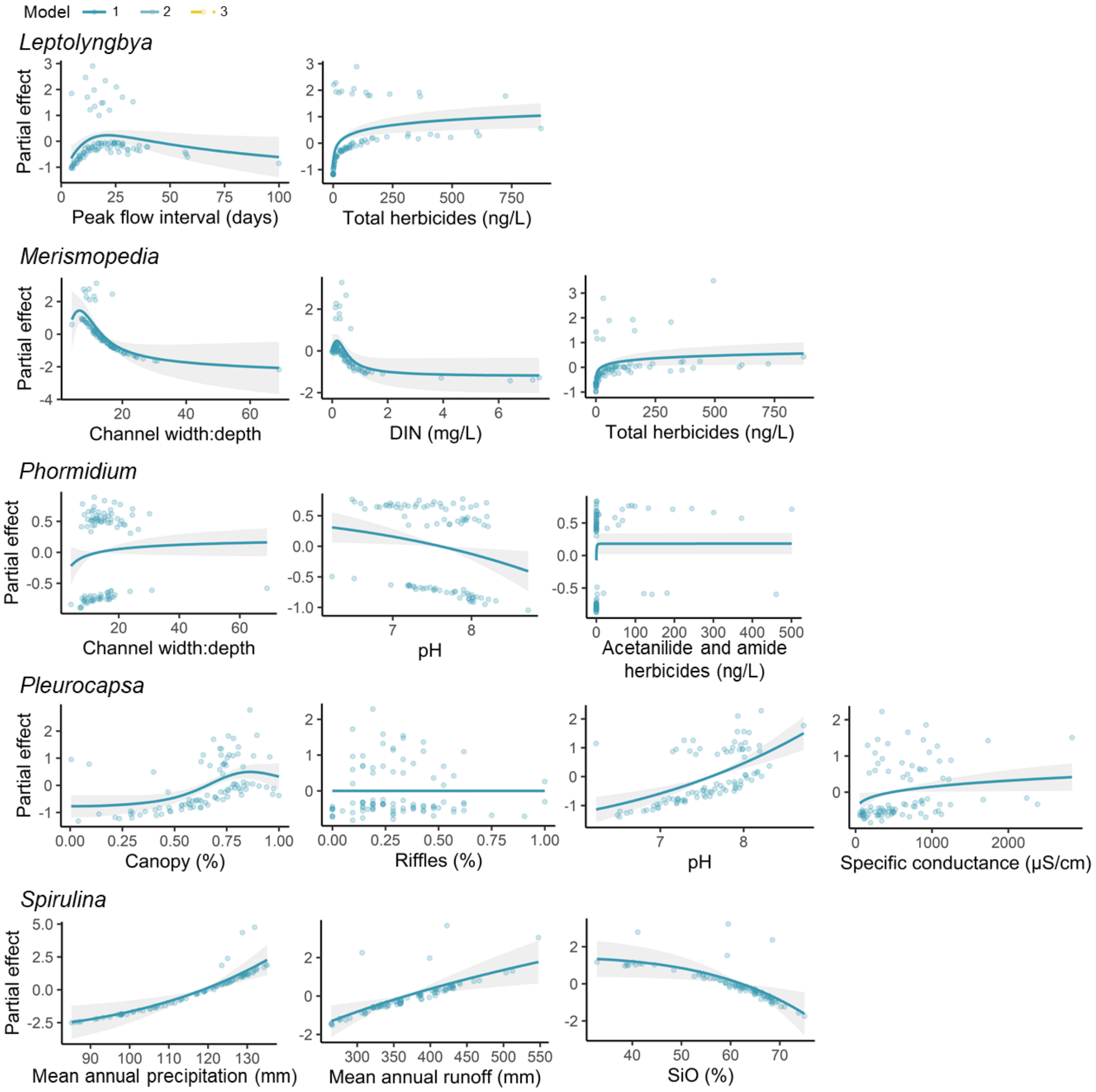


Figure 3 (continued). (2 of 2)

Relative abundance

Among models for relative abundance, deviance explained varied between 15 – 97% with a median of 26% and mean of 47% (Figure 4). Seven of the 12 genera had multiple similarly performing models. For three of those genera (*Anabaena*, *Schizothrix*, and *Synechococcus*), the multiple models did not share any explanatory variables, and the variables present in alternative models were uncorrelated across only samples containing counts of that genus. As with presence/absence, the modeled relative abundance of each genus was most strongly and/or consistently affected by one or two variables, mainly drainage area, substrate size, peak flow interval, DO, pH, DIN, and herbicides (Figures 4 and 5). Model composition was distinct among all genera.

The relative abundances of *Anabaena*, *Chroococcus*, *Pleurocapsa*, *Schizothrix*, and *Spirulina* were influenced by herbicide concentrations. However, for all genera but *Chroococcus*, other models that did not include herbicides passed quality filtering. Only *Spirulina* relative abundance was modeled to decrease with increases in herbicides. All other genera were modeled to increase sharply at low concentrations and reach an asymptote at higher concentrations, as with presence/absence responses. Increasing PCB concentrations resulted in decreased modeled relative abundances of *Homoeothrix* and *Phormidium*. Metals did not consistently affect any genus.

Presence/absence vs. relative abundance

Overall, fewer models for presence/absence passed quality filtering than did those for relative abundance. Additionally, presence/absence models explained less deviance, on average, than those for relative abundance. Of the 9 genera modeled for both characterization method, 4 were influenced by the same environmental variables for both presence/absence and relative abundance: *Aphanothece* decreased with increasing DO, *Chroococcus* increased initially and leveled off with increasing sulfonylurea and urea herbicides (though relative abundance was also strongly influenced by substrate size), *Phormidium* decreased with pH, and *Pleurocapsa* increased with pH. Explanatory variables did not meaningfully overlap between characterization methods for *Anabaena*, *Calothrix*, *Heteroleibleinia*, *Merismopedia*, and *Spirulina*.

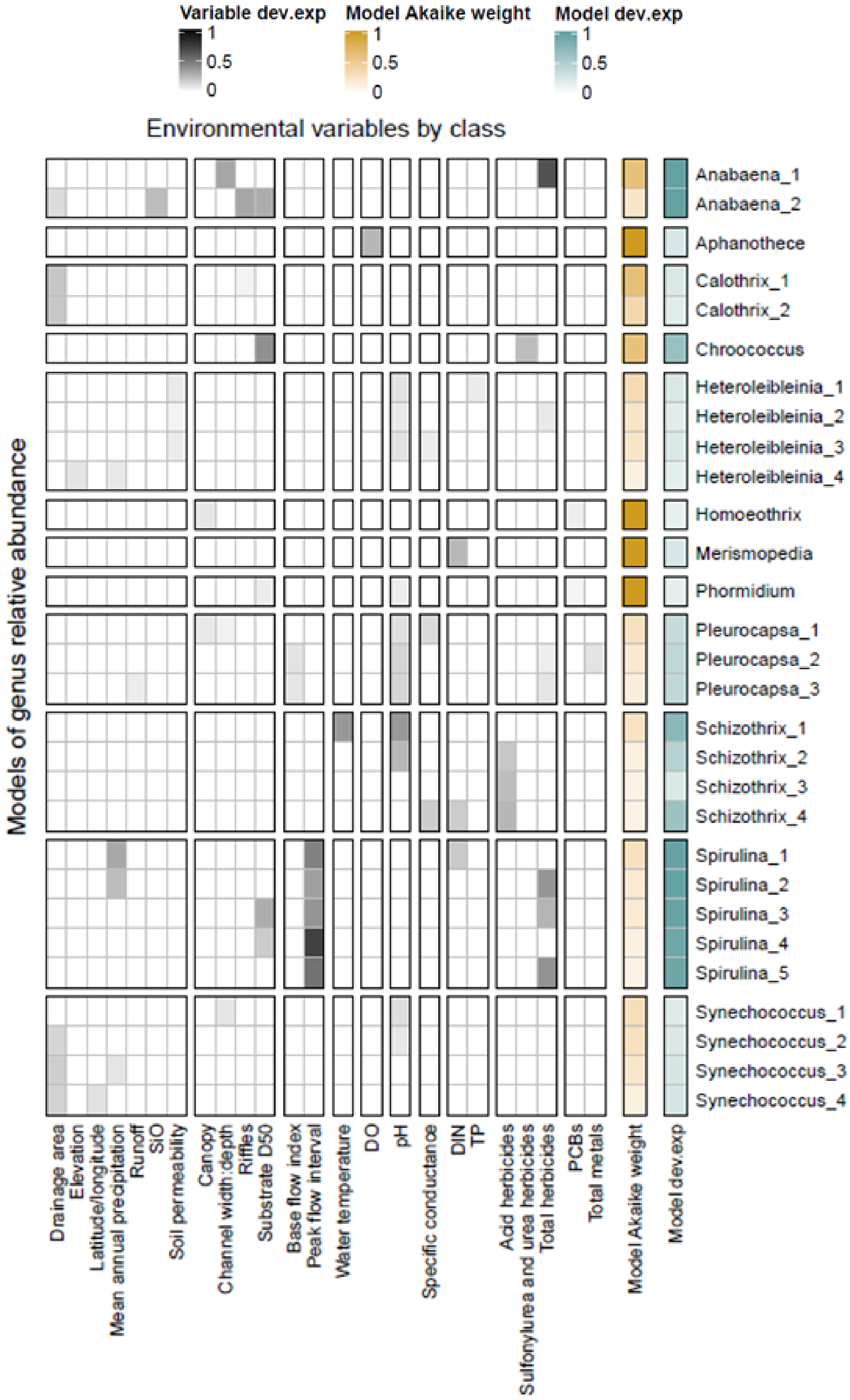


Figure 4. Variable deviance explained (dev.exp), Akaike weights, and model deviance explained of filtered generalized additive models for genus relative abundance. Multiple models for the same genus are sorted in descending order by Akaike weight.

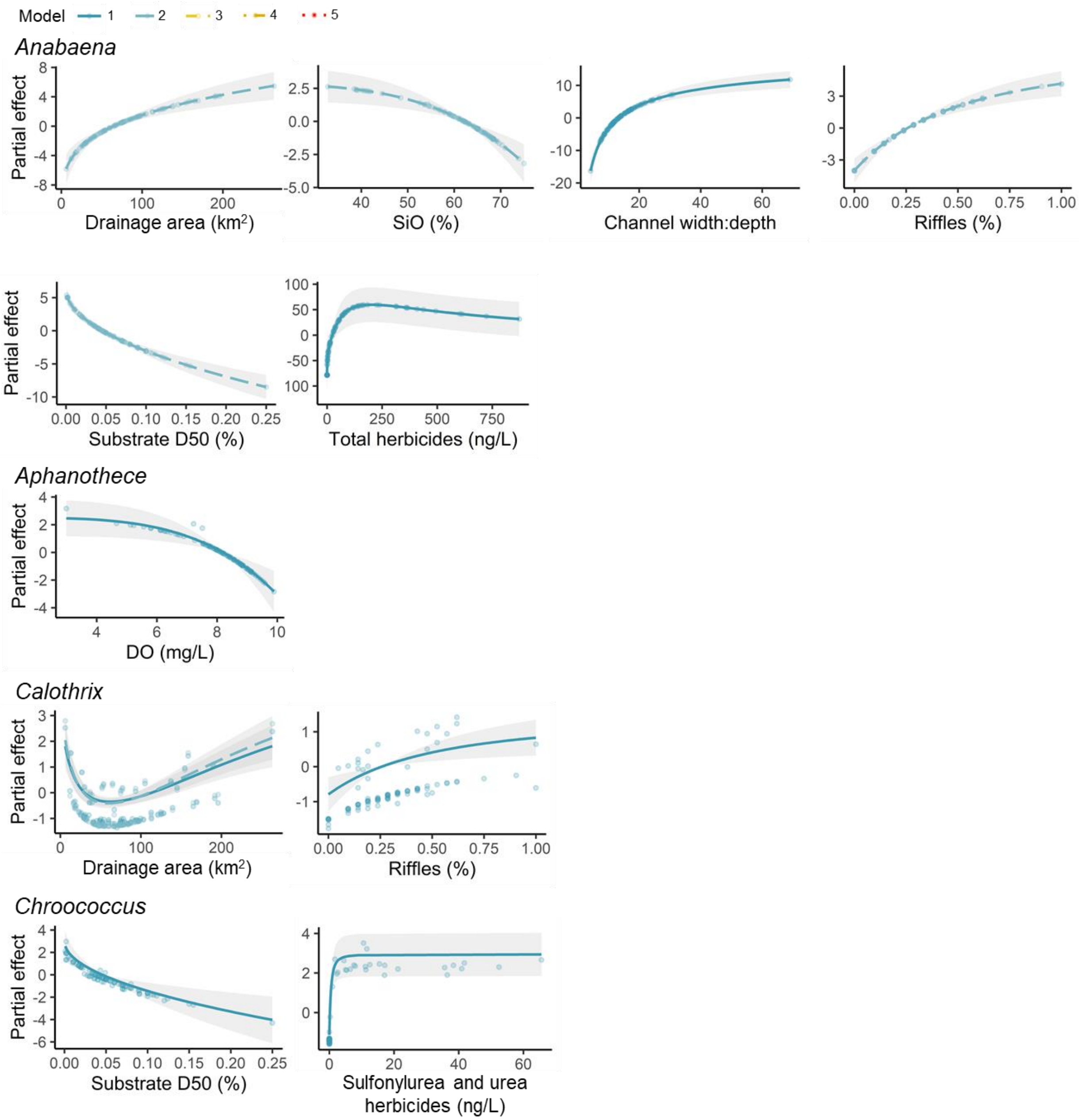


Figure 5. Partial plots of the effects of environmental variables on the relative abundance of cyanobacterial genera. Models containing the same environmental variable for the same genus are represented on a single plot and correspond to the model numbers in Figure 4. Variable abbreviations match those in Table 1. (1 of 4)

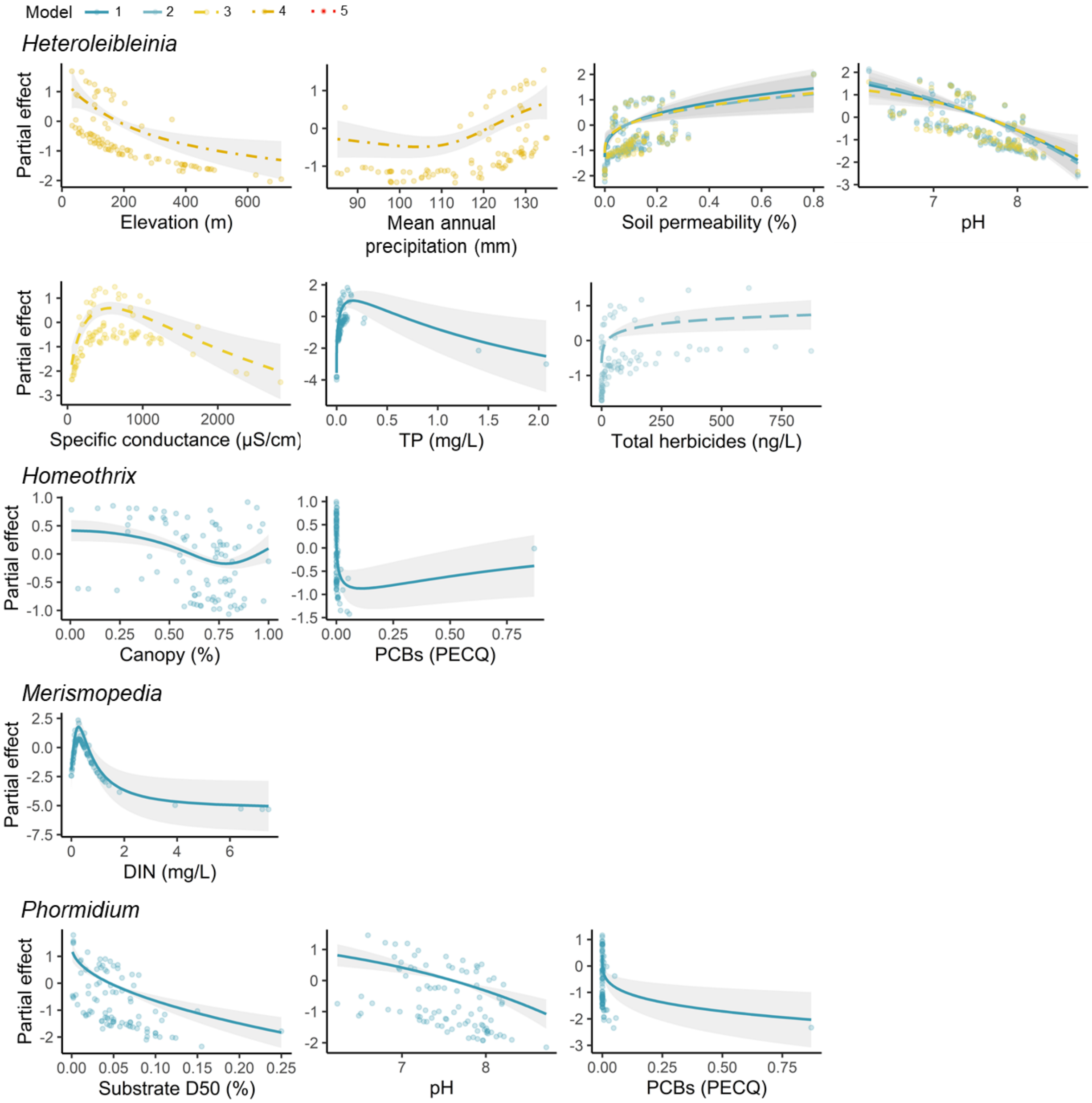


Figure 5 (continued). (2 of 4)

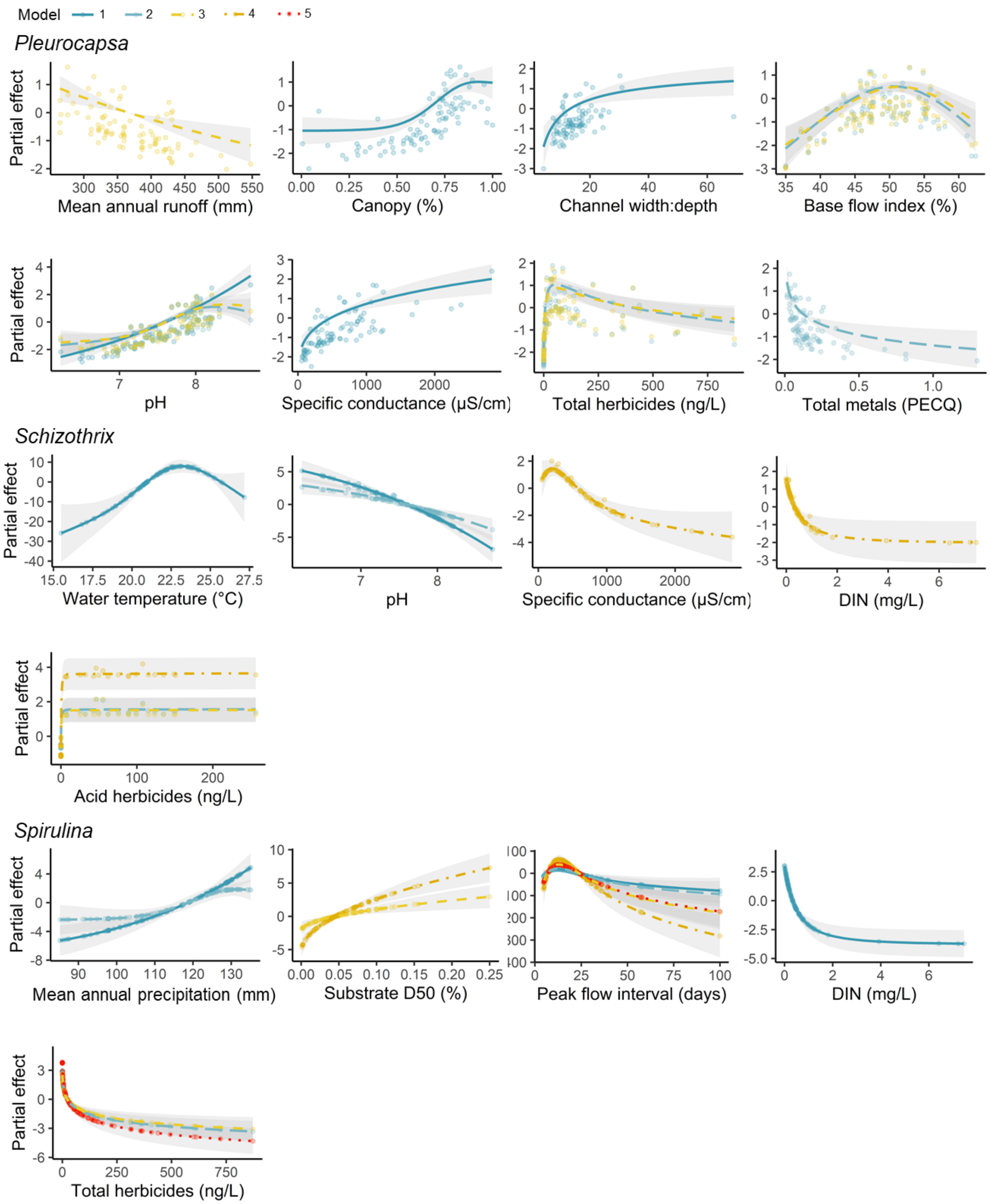


Figure 5 (continued). (3 of 4)

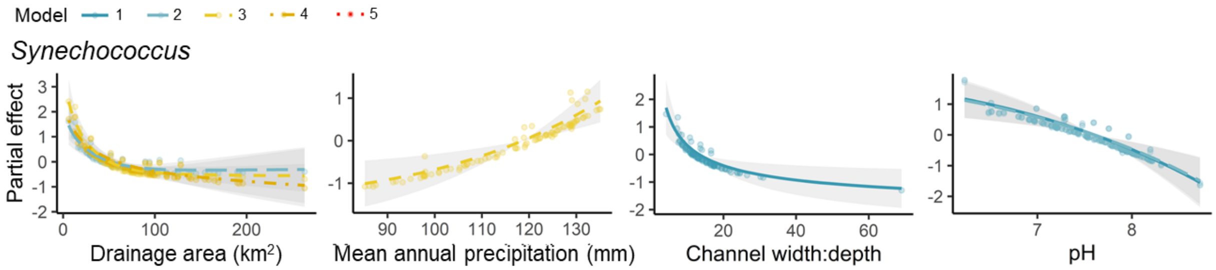


Figure 5 (continued). (4 of 4)

Discussion

Environmental characteristics

Measuring stressor:response relationships requires that environmental stressors are distributed across levels below and above biotic responsiveness (Munn et al. 2018). Across large spatial scales with widely distributed taxa, the larger the gradient in environmental conditions the greater the possibility of measuring biotic responses (Turak et al. 1999). Furthermore, when assessing multivariable stressor effects across multiple taxa, ensuring no one variable defines the environmental gradient is key. In this study, we observed large ranges in watershed and in-stream variables of all environmental classes. While little work has been done to establish minimum and maximum levels of environmental stressors that elicit responses in cyanobacterial taxa, work on eukaryotic algae and overall stream biofilm responses suggests we captured biologically meaningful gradients in environmental variables (Dodds et al. 2002, Stevenson et al. 2008, Potapova and Charles 2003). We also found that no one variable, class of variables, or combination of variables explained a majority of inter-site variability. These results reinforce our conclusions that this field-based study spanned adequate environmental gradients over which to measure cyanobacterial responses.

Genus frequency and relative abundance

The number of cyanobacterial genera per site (4) and the number of genera across all sites (18) was low given the spatial, land use, and environmental extents captured in this study. In temperate regions, the number of cyanobacterial genera in a single stream often exceeds 15 genera, and stream surveys of some U.S. states have detected over 25 genera in a sampling season (Komárek and Johansen 2015a). Noticeably absent were the genera *Nostoc*, *Oscillatoria*, and *Scytonema* that are common and widespread in North America (Komárek and Johansen 2015b). Regardless, high frequency and relative abundance of genera such as *Calothrix*, *Chamaesiphon*, *Heteroleibleinia*, *Homoeothrix*, and *Phormidium* are consistent with other studies that show these taxa often dominate the cyanobacterial communities in temperate streams (Douterelo et al. 2004, Komárek and Johansen 2015b). We found that no rare genera had high relative abundances where present, suggesting that there were no blooms from genera with narrow niche breadths at the time of sampling.

Environmental factors influencing genera

Watershed-scale variables and in-stream stressors of all types affected cyanobacteria from multiple evolutionary and functional groups. There was little overlap in the individual or combination of influential environmental variables among genera. These results indicate there may be substantial niche partitioning among cyanobacterial genera across the large spatial extent of the northeast U.S. While many genera co-occurred, our results suggest each genus is distinct in the specific stressors that characterize its presence/absence and relative abundance across the spatial extent and environmental gradients considered. An exhaustive summary of all the environmental influences on each genus is beyond the scope of the present discussion, as these relationships were found to be nuanced and genus-specific. However, we have presented a framework for the interpretation of genera of interest. Additionally, some notable results stand out in light of past work on some genera and stressors of broad concern.

Over half of the genera identified contain known or potential toxigenic species. Most of these genera were most responsive to watershed variables, stream geomorphology, streamflow, pH, and/or herbicides – with only *Anabaena* influenced by nutrients, *Schizothrix* by water temperature, and *Pleurocapsa* by canopy (a proxy for light availability). Except for *Spirulina*, no toxigenic genus was influenced by climate, as measured by mean annual precipitation. These results alone suggest the need to consider other variables in addition to those related to climate change, water temperature, and nutrients in order to understand freshwater benthic cyanoHAB patterns.

Of the potentially toxigenic genera in the present study, *Anabaena*, *Lyngbya* (of which some toxigenic species have been transferred to *Microseira*), and *Phormidium* (toxigenic species transferred to *Kamptonema* or *Microcoleus*) most commonly contain toxigenic species found in stream biofilms (Wood et al. 2020). *Lyngbya* was unable to be modeled, possibly due to its presence in few samples. *Anabaena* was rare and low in relative abundance across the study region but was able to be modeled with high deviance explained for both presence/absence and relative abundance. Large proliferations of *Anabaena*-dominated benthic mats have been detected in the slow-moving, warm, and nitrogen-rich Eel River, northern California where fine sediments are consistently loaded into the stream (Bouma-Gregson et al. 2017, 2018). In these studies, water temperature, light, and low phosphorus were tentatively attributed to mat proliferation. In contrast, our models indicated stream geomorphology, possibly TP, and possibly herbicides most affected *Anabaena*, while increasing percentages of fine bed substrate led to decreases in expected *Anabaena* presence and relative abundance. One model did indicate a U-shaped response of *Anabaena* presence to TP concentrations and another showed increasing *Anabaena* relative abundance with increasing stream size (a potential proxy for light availability), suggesting some consistency with Eel River environmental controls of low TP and high light. Notably, our models reflect low relative abundances of *Anabaena*, as opposed to the large blooms observed in the Eel River – suggesting environmental controls may differ between small and large cyanobacterial populations or have regional context dependency.

Most multi-stressor work on benthic cyanobacteria has focused on toxigenic *Phormidium*-dominated mats in New Zealand streams (Wood et al. 2020). In these streams, *Phormidium* mat cover is positively associated with low nitrogen and phosphorus concentrations, water temperature > 15ºC, and increasing conductivity (Wood et al. 2017). Our models for *Phormidium* presence/absence and relative abundance performed relatively weakly (< 17% deviance explained), but each explanatory variable was significant – suggesting some information between *Phormidium* and environmental influences in the northeast U.S. can be gained. In models for presence/absence and relative abundance, stream geomorphology (stream size and substrate), pH, and pollutants (herbicides and PCBs) each affected *Phormidium*. None of these stressors were considered in generalized additive mixed models for *Phormidium* in New Zealand streams, but nutrients, water temperature, and specific conductance were notably absent in our models (Wood et al. 2017, McAllister et al. 2018).

Caution should be used in comparing New Zealand and northeast U.S. *Phormidium* patterns because of the differences in response variable (aerial cover versus presence/absence and relative abundance, respectively) and in relative dominance by *Phormidium* of the whole benthic community. In the present study, *Phormidium* was among the most common and abundant cyanobacteria identified. However, most sites were dominated by diatoms, as opposed to the predominantly cyanobacterial mats in New Zealand streams. Notably, though, one sample was comprised of over 50% *Phormidium*, which suggests that GAMs for *Phormidium* integrated the environmental responses of a biofilm that was perhaps naturally high in cyanobacterial relative abundance or experiencing a *Phormidium* bloom. Regardless, these results highlight the need to consider a broad suite of environmental stressors in all research modeling cyanobacterial patterns, as well as suggest that different stressors may affect the same taxon given region-specific conditions.

Herbicides inhibit photosynthesis and other critical cellular functions in eukaryotic algae and cyanobacteria (Aslim and Ozturk 2009). Across algal groups, herbicides reduce population growth in single-taxon cultures (Singh et al. 2016). However, in phytoplanktonic and benthic mixed algal communities, cyanobacteria outcompete eukaryotic algae in the presence of herbicides (Harris and Smith 2016). Among cyanobacteria, taxa vary in tolerance to varying herbicide concentrations, with some strains having a high tolerance to glyphosate in particular (Powell et al. 1991, Forlani et al. 2008). In the present study, 9 of the 13 modeled genera had at least one high performing model containing herbicides as a significant explanatory variable for either presence/absence or relative abundance. For every genus except *Spirulina*, low concentrations of herbicides increased the modeled presence or relative abundance of the genus. These modeled responses to herbicides among cyanobacterial genera are consistent with the principle of population growth as a result of competitive release, in which the loss of one taxon allows its competitor to expand in population size (Fleeger et al. 2003). Positive, genus-specific responses from competitive release could be attributable to decreases of more sensitive cyanobacterial or eukaryotic algal taxa, increases in the tolerant cyanobacterial genera that have a positive response to herbicides in their GAMs, or a combination. Notably, herbicides were one of the most highly represented classes of environmental stressors in all GAMs.

Similar to herbicides, PCBs and polycyclic aromatic hydrocarbons (PAHs) have been shown to increase cyanobacterial abundance relative to eukaryotic algae in laboratory experiments because of higher toxicity to eukaryotes. Zhang et al. (2015) even showed that a strain of *Anabaena* can biodegrade PCBs in culture. However, unlike herbicides and in contrast to these previous findings, in our GAMs PCBs had small but negative effects on *Calothrix*, *Homoeothrix*, and *Phormidium*. PCBs accumulate in benthic biofilms (Wang et al. 1999). The negative effects of PCBs we observed may be attributable to one site with very high concentrations that may exceed those for which PCBs are competitively advantageous for cyanobacteria. Overall, more studies must directly address PCB effects on benthic biofilm community structure before conclusions can be made about the population- and community-scale effects of PCBs on cyanobacteria.

Taxon presence/absence and relative abundance measure different aspects of communities, but one or the other is often used to characterize taxon-specific patterns in aquatic microbial research. For half of the genera modeled, there was little to no overlap of influential explanatory variables between presence/absence and relative abundance. These results indicate that, for some taxa, different processes control whether a genus is present and how relatively abundant that genus is where present. This finding has repercussions for managing cyanobacteria of concern. If the management goal is prevention or eradication of a taxon, the stressors affecting genus presence/absence should be prioritized. Conversely, if the objective is population control and bloom prevention, addressing the stressors influencing relative abundance is more critical.

Across all genera and between presence/absence and relative abundance, models with the highest deviance explained were for genera with the lowest mean frequency and relative abundance. In contrast, the most common and abundant genera were modeled with < 30% deviance explained or not able to be modeled at all. Several explanations for these results are possible.

First, broad distribution and highly variable relative abundance across environmental conditions often result in poor model performance. As such, common and high relative abundance genera like *Heteroleibleinia*, *Homoeothrix*, and *Phormidium* might best be characterized as generalists that are responsive to multiple stressors depending on the site. In contrast, the more restricted distributions of genera like *Anabaena*, *Chroococcus*, and *Spirulina* indicate potential specialization and narrow niche breadth.

The present study is also limited by the identification of cyanobacteria to genus, rather than to species or lower taxonomic resolution. Cyanobacterial communities are commonly identified to genus in environmental monitoring and ecological research due to ambiguity in species-level morphology (Johansen and Casamatta 2005). In eukaryotic algae, species and populations within a genus often respond differently to myriad environmental conditions (Growns 1999, Hill et al. 2001), which is likely the case with cyanobacteria. For some cyanobacterial species, only some strains are toxigenic, and their presence in the environment can depend on environmental conditions (Davis et al. 2009). Altogether, multiple species or strains with different environmental preferences might lead to high variability in genus distribution and relative abundance, resulting in weakly performing GAMs. Importantly, though, we observed no clear relationships among the frequency, relative abundance, and number of species per genus. These results suggest that intra-generic variability in taxonomy or environmental responsiveness may not have had a strong effect upon the results of this study. Regardless, identifying cyanobacteria to species, strains, or unique DNA sequences could lead to finding stronger and more novel relationships between cyanobacteria and environmental stressors.

Finally, while we considered a broad suite of potential watershed characteristics and local environmental variables, some potential stressors and controls on cyanobacteria were not accounted for. First, pharmaceutical and personal care products (PPCPs) and plastics are contaminants of emerging concern that have the potential to affect cyanobacteria in addition to or instead of the stressors measured in the present study. Each of these stressors has been shown to both negatively and positively affect different groups of eukaryotic algae and cyanobacteria (Wilson et al. 2003). Antibiotics, in particular, are more toxic to cyanobacteria than to eukaryotic algae (Välitalo et al. 2017), likely because of physiological similarities to other gram-negative bacteria targeted by certain antibiotics (Robinson et al. 2005). Second, while local and regional environmental conditions are known to strongly affect eukaryotic algae and cyanobacteria, many benthic microorganisms demonstrate distinct biogeographical patterns that result from dispersal limitation or historical environments that are not characterized by measurements of current conditions (Ribeiro et al. 2018). While we removed spatial autocorrelation signals from our models, we did not directly integrate spatial variables like site or eigenvector maps based on latitude and longitude that can identify spatial structuring and indicate dispersal extent. Finally, grazing pressure is a strong control of cyanobacterial abundance in many temperate streams (Scott and Marcarelli 2012). The lower edibility of cyanobacteria relative to diatoms and, to a lesser extent, green algae can result in increased cyanobacterial absolute and relative abundance and diversity when herbivory is high (Rosemond et al. 1993). Furthermore, variability in habit, chemical defenses, and nitrogen content can affect the herbivory rate of taxa within the cyanobacterial community (Monica and Agrawal 2011). Grazing is difficult to account for in field-based, observational studies. However, metrics of potential grazing pressure could possibly be developed based on local herbivore community composition and integrated into GAMs. Additionally, pesticides that directly affect herbivore abundance might be used as indicators of relative grazing pressure on eukaryotic algae and cyanobacteria. However, given that some cyanobacteria can biodegrade insecticides (Zhang et al. 2012), this application may be difficult. Altogether, each of these outstanding, potential influences on cyanobacteria should be incorporated going forward for truly holistic models that describe cyanobacterial distributions and abundances.

The use of a partially unsupervised, all subsets procedure for model selection allowed us to identify novel cyanobacterial-environmental relationships using GAMs. However, the results from these modeling exercises should always be interpreted with context and caution. In multivariable modeling, it is important to consider both relative model performance and overall model information. For roughly half of the genera modeled, we uncovered multiple similarly performing models that led to disparate ecological conclusions. The intent in this study was to characterize cyanobacterial-environmental relationships, rather than predict them. Therefore, it is important to consider the ecological implications of different models rather than accept the single best performing model, even when there is resulting ambiguity in the interpretation. Additionally, in the presentation of our results we focused on dominant stressors for each genus. However, we emphasize that interpretation of the stressors in GAMs is contingent upon their co-occurrent explanatory variables. While some dominant stressors may be significant and explanatory by themselves, the addition of other explanatory variables substantially improved model quality and ecological characterization. Therefore, the entire suite of explanatory variables should be considered ecologically relevant.

Conclusions

To our knowledge, the 44 multi-stressor models constructed for 13 cyanobacterial genera in this study comprise the most comprehensive effort to date to model taxon-level benthic cyanobacterial responses to conditions along environmental gradients. We detected genus-specific patterns in environmental relationships across the northeast United States. Watershed-scale variables, stream geomorphology, flow, water temperature, DO, pH, conductivity, nutrients, herbicides, and PCBs each substantially affected at least one cyanobacterial genus. Genera that contain potential toxin-producing species were responsive to variables of all classes, though rarely to climate, water temperature, or nutrients. The presence and/or relative abundance of *Calothrix*, *Chroococcus*, *Leptolyngbya*, *Merismopedia*, *Phormidium*, and *Schizothrix* each increased at low concentrations of herbicides, representing a threshold effect of possible competitive release from less tolerant cyanobacterial taxa. Meanwhile, *Calothrix*, *Homoeothrix*, and *Phormidium* decreased with increasing concentrations of PCBs, indicating previously undescribed PCB toxicity to these genera. For about half of the modeled genera, the environmental factors affecting genus presence/absence differed from those influencing the relative abundance of the same genus, suggesting management strategies that target preventing the geographic spread of cyanobacteria and those focused on population control may need to address different environmental stressors. Altogether, the present study demonstrates that a range of watershed and in-stream environmental factors affect cyanobacteria in streams, including understudied contaminants that are uniquely human-caused. This work provides a study design and modeling framework for future field-based research on cyanobacterial presence/absence and relative abundance across aquatic habitats.

References

Aslim, B., & Ozturk, S. (2009). Toxicity of herbicides to cyanobacterial isolates. *Journal of Environmental Biology*, *30*, 381-384.

Batt, A. L., Wathen, J. B., Lazorchak, J. M., Olsen, A. R., & Kincaid, T. M. (2017). Statistical survey of persistent organic pollutants: risk estimations to humans and wildlife through consumption of fish from US rivers. *Environmental Science & Technology*, *51*, 3021-3031.

Benbrook, C. M. (2016). Trends in glyphosate herbicide use in the United States and globally. *Environmental Sciences Europe*, *28*, 3.

Bouma-Gregson, K., Kudela, R. M., & Power, M. E. (2018). Widespread anatoxin-a detection in benthic cyanobacterial mats throughout a river network. *PLoS One*, *13*, e0197669.

Bouma-Gregson, K., Power, M. E., & Bormans, M. (2017). Rise and fall of toxic benthic freshwater cyanobacteria (*Anabaena* spp.) in the Eel river: Buoyancy and dispersal. *Harmful Algae*, *66*, 79-87.

Carmichael, W. W. (2001). Health effects of toxin-producing cyanobacteria: “The CyanoHABs”. *Human and Ecological Risk Assessment: An International Journal*, *7*, 1393-1407.

Coles, J. F., Riva-Murray, K., Van Metre, P. C., Button, D. T., Bell, A. H., Qi, S. L., ... & Sheibley, R. W. (2019). Design and methods of the US Geological Survey Northeast Stream Quality Assessment (NESQA), 2016 (No. 2018-1183). U.S. Geological Survey.

Davis, T. W., Berry, D. L., Boyer, G. L., & Gobler, C. J. (2009). The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of *Microcystis* during cyanobacteria blooms. *Harmful Algae*, *8*, 715-725.

Dodds, W. K., & Smith, V. H. (2016). Nitrogen, phosphorus, and eutrophication in streams. *Inland Waters*, *6*, 155-164.

Dodds, W. K., López, A. J., Bowden, W. B., Gregory, S., Grimm, N. B., Hamilton, S. K., ... & Morrall, D. (2002). N uptake as a function of concentration in streams. *Journal of the North American Benthological Society*, *21*, 206-220.

Douterelo, I., Perona, E., & Mateo, P. (2004). Use of cyanobacteria to assess water quality in running waters. *Environmental Pollution*, *127*, 377-384.

Fasiolo, M., Nedellec, R., Goude, Y., & Wood, S. N. (2020). Scalable visualization methods for modern generalized additive models. *Journal of Computational and Graphical Statistics*, *29*, 78-86.

Fleeger, J. W., Carman, K. R., & Nisbet, R. M. (2003). Indirect effects of contaminants in aquatic ecosystems. *Science of the Total Environment, 317*, 207-233.

Forlani, G., Pavan, M., Gramek, M., Kafarski, P., & Lipok, J. (2008). Biochemical bases for a widespread tolerance of cyanobacteria to the phosphonate herbicide glyphosate. *Plant and Cell Physiology*, *49*, 443-456.

Genuer, R., Poggi, J. M., Tuleau-Malot, C., & Genuer, M. R. (2019). Package 'vsurf'. *Pattern Recognition Letters*, *31*, 2225-2236.

Growns, I. (1999). Is genus or species identification of periphytic diatoms required to determine the impacts of river regulation? *Journal of Applied Phycology*, *11*, 273-283.

Harris, T. D., & Smith, V. H. (2016). Do persistent organic pollutants stimulate cyanobacterial blooms? *Inland Waters*, *6*, 124-130.

Heino, J., & Soininen, J. (2006). Regional occupancy in unicellular eukaryotes: a reflection of niche breadth, habitat availability or size-related dispersal capacity? *Freshwater Biology*, *51*, 672-685.

Hill, B. H., Stevenson, R. J., Pan, Y., Herlihy, A. T., Kaufmann, P. R., & Johnson, C. B. (2001). Comparison of correlations between environmental characteristics and stream diatom assemblages characterized at genus and species levels. *Journal of the North American Benthological Society*, *20*, 299-310.

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., ... & Megown, K. (2015). Completion of the 2011 National Land Cover Database for the conterminous United States–representing a decade of land cover change information. *Photogrammetric Engineering & Remote Sensing*, *81*, 345-354.

James, F. C., & McCulloch, C. E. (1990). Multivariate analysis in ecology and systematics: panacea or Pandora's box? *Annual Review of Ecology and Systematics*, *21*, 129-166.

Johansen, J. R., & Casamatta, D. A. (2005). Recognizing cyanobacterial diversity through adoption of a new species paradigm. *Algological Studies*, *117*, 71-93.

Kaushal, S. S., Likens, G. E., Jaworski, N. A., Pace, M. L., Sides, A. M., Seekell, D., ... & Wingate, R. L. (2010). Rising stream and river temperatures in the United States. *Frontiers in Ecology and the Environment*, *8*, 461-466.

Kaushal, S. S., Likens, G. E., Pace, M. L., Utz, R. M., Haq, S., Gorman, J., & Grese, M. (2018). Freshwater salinization syndrome on a continental scale. *Proceedings of the National Academy of Sciences*, *115*, E574-E583.

Komárek, J., & Johansen, J. R. (2015a). Coccoid cyanobacteria. In Wehr, J. D., Sheath, R. G., Kociolek, J. P. (Eds.), *Freshwater Algae of North America: Ecology and Classification* (pp. 75-135). Academic Press, Amsterdam.

Komárek, J., & Johansen, J. R. (2015b). Filamentous cyanobacteria. In Wehr, J. D., Sheath, R. G., Kociolek, J. P. (Eds.), *Freshwater Algae of North America: Ecology and Classification* (pp. 135-235). Academic Press, San Diego, CA.

Kupper, H., & Kroneck, P. M. (2005). Heavy metal uptake by plants and cyanobacteria. *Metal Ions in Biological Systems*, *44*, 97.

Lake, P. S. (2003). Ecological effects of perturbation by drought in flowing waters. *Freshwater Biology*, *48*, 1161-1172.

MacDonald, D. D., Ingersoll, C. G., & Berger, T. A. (2000). Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems. *Archives of Environmental Contamination and Toxicology*, *39*, 20-31.

McAllister, T. G., Wood, S. A., Atalah, J., & Hawes, I. (2018). Spatiotemporal dynamics of *Phormidium* cover and anatoxin concentrations in eight New Zealand rivers with contrasting nutrient and flow regimes. *Science of the Total Environment*, *612*, 71-80.

McCormick, P. V., & Stevenson, R. J. (1991). Mechanisms of benthic algal succession in lotic environments. *Ecology*, *72*, 1835-1848.

Monica, A., & Agrawal, M. K. (2011). Cyanobacteria-herbivore interaction in freshwater ecosystem. *Journal of Microbiology and Biotechnology Research*, *1*, 52-66.

Moulton, S. R., Kennen, J. G., Goldstein, R. M., & Hambrook, J. A. (2002). Revised protocols for sampling algal, invertebrate, and fish communities as part of the National Water-Quality Assessment Program (No. 2002-150). Geological Survey (US).

Munn, M. D., Waite, I., & Konrad, C. P. (2018). Assessing the influence of multiple stressors on stream diatom metrics in the upper Midwest, USA. *Ecological Indicators*, *85*, 1239-1248.

Paerl, H. W., Fulton, R. S., Moisander, P. H., & Dyble, J. (2001). Harmful freshwater algal blooms, with an emphasis on cyanobacteria. *The Scientific World Journal*, *1*, 76-113.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*, 289-290.

Pearson, R. G. (2007). Species' distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History*, *50*, 54-89.

Pérez, G. L., Torremorell, A., Mugni, H., Rodriguez, P., Vera, M. S., Nascimento, M. D., ... & Izaguirre, I. (2007). Effects of the herbicide Roundup on freshwater microbial communities: a mesocosm study. *Ecological Applications*, *17*, 2310-2322.

Potapova, M., & Charles, D. F. (2003). Distribution of benthic diatoms in US rivers in relation to conductivity and ionic composition. *Freshwater Biology*, *48*, 1311-1328.

Powell, H. A., Kerbby, N. W., & Rowell, P. (1991). Natural tolerance of cyanobacteria to the herbicide glyphosate. *New Phytologist*, *119*, 421-426.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org/.

Ram, K., & Wickham, H. (2018). wesanderson: a Wes Anderson palette generator. R package version 0.3.6. Available at https://cran.r-project.org/package=wesanderson.

Ribeiro, K. F., Duarte, L., & Crossetti, L. O. (2018). Everything is not everywhere: a tale on the biogeography of cyanobacteria. *Hydrobiologia*, *820*, 23-48.

Robinson, A. A., Belden, J. B., & Lydy, M. J. (2005). Toxicity of fluoroquinolone antibiotics to aquatic organisms. *Environmental Toxicology and Chemistry: An International Journal*, *24*, 423-430.

Rosemond, A. D., Mulholland, P. J., & Elwood, J. W. (1993). Top-down and bottom-up control of stream periphyton: effects of nutrients and herbivores. *Ecology*, *74*, 1264-1280.

Scott, J. T., & Marcarelli, A. M. (2012). Cyanobacteria in freshwater benthic environments. In Whitton, B. A., Potts, M. (Eds.), *The Ecology of the Cyanobacteria II* (pp. 271-289). Springer, Dordrecht.

Singh, D. P., Khattar, J. I. S., Kaur, G., & Singh, Y. (2016). Toxicological impact of herbicides on cyanobacteria. *Annual Research & Review in Biology*, *9*, 1-39.

Stevenson, R. J., Hill, B. H., Herlihy, A. T., Yuan, L. L., & Norton, S. B. (2008). Algae–P relationships, thresholds, and frequency distributions guide nutrient criterion development. *Journal of the North American Benthological Society*, *27,* 783-799.

Turak, E., Flack, L. K., Norris, R. H., Simpson, J., & Waddell, N. (1999). Assessment of river condition at a large spatial scale using predictive models. *Freshwater Biology*, *41*, 283-298.

U.S. Geological Survey. (2019). BioData—Aquatic bioassessment data for the nation available on the world wide web. Accessed October 10, 2019. doi:10.5066/F77W698B

Välitalo, P., Kruglova, A., Mikola, A., & Vahala, R. (2017). Toxicological impacts of antibiotics on aquatic micro-organisms: a mini-review. *International Journal of Hygiene and Environmental Health*, *220*, 558-569.

Van Sickle, J., Baker, J., … & Li, J. (2004). Projecting the biological condition of streams under alternative scenarios of human land use. *Ecological Applications*, *14*, 368-380.

Vu, V. Q. (2011). ggbiplot: A ggplot2 based biplot. R package version 0.5. Available at http://github.com/vqv/ggbiplot.

Wang, H., Kostel, J. A., Amand, A. L. S., & Gray, K. A. (1999). 2. The response of a laboratory stream system to PCB exposure: study of periphytic and sediment accumulation patterns. *Water Research*, *33*, 3749-3761.

Wilson, B. A., Smith, V. H., deNoyelles, F., & Larive, C. K. (2003). Effects of three pharmaceutical and personal care products on natural freshwater algal assemblages. *Environmental Science & Technology*, *37*, 1713-1719.

Wood, S. A., Atalah, J., Wagenhoff, A., Brown, L., Doehring, K., Young, R. G., & Hawes, I. (2017). Effect of river flow, temperature, and water chemistry on proliferations of the benthic anatoxin-producing cyanobacterium *Phormidium*. *Freshwater Science*, *36*, 63-76.

Wood, S. A., Kelly, L. T., Bouma-Gregson, K., Humbert, J. F., Laughinghouse IV, H. D., Lazorchak, J., ... & Quiblier, C. (2020). Toxic benthic freshwater cyanobacterial proliferations: Challenges and solutions for enhancing knowledge and improving monitoring and mitigation. *Freshwater Biology*, *65*, 1824-1842.

Wood, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, *62*, 1025-1036.

Wood, S. N. (2020). mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. R package version 1.8-33. Available at https://cran.r-project.org/package=mgcv.

Zhang, H., Hu, C., Jia, X., Xu, Y., Wu, C., Chen, L., & Wang, F. (2012). Characteristics of γ-hexachlorocyclohexane biodegradation by a nitrogen-fixing cyanobacterium, *Anabaena azotica*. *Journal of Applied Phycology*, *24*, 221-225.

Zhang, H., Jiang, X., Lu, L., & Xiao, W. (2015). Biodegradation of polychlorinated biphenyls (PCBs) by the novel identified cyanobacterium *Anabaena* PD-1. *PLoS One*, *10*, e0131450.

CHAPTER V

CONCLUSION

Insights from the dissertation

The composition and geographic distributions of benthic algal communities and populations are influenced by spatial and environmental processes that are not fully understood (Heino et al. 2015, Soininen and Teittinen 2019). Anthropogenic activities alter the climate, geomorphology, and water chemistry of streams in multi-faceted ways that are rarely assessed together (Pickett et al. 2011). Meanwhile, landscape features, such as elevation, and spatial processes, including dispersal, also affect algal structure and function (Soininen 2007). With this dissertation, I investigated the ecosystem processes that influence algal composition and distribution to disentangle the relative influences of ecological controls on algal biogeography and to characterize the autecology of individual taxa for use in future research and environmental assessments. I detected novel relationships between stream algae and ecological factors by relating direct measurements of dispersal to algal metacommunity structure and by applying non-linear, multivariable species distribution modeling to all taxa found across a human-impacted region.

Chapter II demonstrates that dispersal of diatoms via wind is common in a landscape that is characterized by aeolian processes. In the McMurdo Sound region (MSR) of Antarctica, wind-mediated dispersal of diatoms contributes to metacommunity structure in addition to processes including physicochemical conditions and intra-habitat dispersal via surface water flow, though to a lesser extent. Previous research has shown that spatial factors, in general, are important in controlling algal community dynamics (Soininen 2007, Sakaeva et al. 2016). My results reveal that direct measurement of dispersal, as opposed to inference from geographic distance, allows for a nuanced interpretation of the importance of spatial processes on microbial metacommunity composition. Despite the large fluxes of aeolian material in the MSR, the diatom metacommunity appears to be predominantly affected by intra-habitat controls rather than inter-habitat dispersal.

The results from Chapter II highlighted the importance of local, intra-habitat processes on diatom metacommunity composition. In Chapter III, I show that regional gradients in environmental conditions caused by human activities lead to taxon-specific responses of diatoms to distinct suites of environmental conditions in the northeast United States. My results show that (1) the relative importance of different environmental stressors and watershed characteristics differ dramatically by diatom taxon, (2) herbicides appear to have an indirect effect of competitive release on those diatoms most influenced by them, (3) diatom taxa belong to ecological guilds based on shared responsiveness to environmental variables, and (4) species distribution modeling applied across all taxa in a metacommunity is an effective tool to distinguish variability in biogeographic controls and autecological characteristics among diatoms.

While diatoms are important indicators of stream health and are model organisms for studying microbial biogeography, benthic cyanobacteria affect human health, recreation, and economies through the proliferation of toxin-producing blooms. In Chapter IV, I investigate the environmental influences on potentially toxigenic cyanobacterial genera that are distributed along multiple environmental gradients in the northeast U.S. I provide evidence that benthic cyanobacteria are less influenced by light availability, water temperature, and nutrients than they are by watershed characteristics, stream substrate, streamflow, and herbicides across a large spatial scale. These results highlight the need to consider a variety of environmental factors when assessing the potential for benthic cyanobacterial proliferation and blooms in streams.

Given these findings, some relationships across the chapters are evident. I found an overall small effect of inter-habitat dispersal by wind in a landscape characterized by wind (the MSR) alongside strong influences of environmental conditions for many diatom taxa in the human-impacted northeast U.S. Together, these results suggest there is limited evidence for contemporary wind dispersal being a primary control on metacommunity structure or taxon distributions for diatoms at landscape to regional spatial scales. While most research characterizing spatial processes affecting benthic diatoms attributes spatial control to dispersal (Heino et al. 2010), my results suggest factors such as historical dispersal, environmental history, and evolutionary history may more accurately represent previously observed distance effects. The ability to relate processes in the MSR to other regions characterized by different environments and human impacts highlights the utility of researching biogeographic processes in end-member systems like the MSR.

In the northeast U.S., diatoms were highly diverse taxonomically and in environmental responsiveness when compared to cyanobacteria. Unlike diatoms, few cyanobacteria were responsive to factors such as pH, specific conductance, and metals when modeled across the same environmental gradients. While some of these differences are likely attributable to different taxonomic resolution, there is striking evidence that cyanobacteria are not as affected by some common in-stream stressors as are diatoms. These results highlight the need to consider multiple algal groups and taxa in ecological research and environmental assessments.

Future directions

The work described in this dissertation supports previous research demonstrating that algal communities and taxon distributions are influenced by spatial processes, like dispersal, in addition to environmental factors that operate at multiple spatial scales. However, many known influences on algal biogeography have neither been studied in detail nor been integrated with measures of environmental or spatial controls, including historical environmental conditions like glacial periods (Brown and Lomolino 2000); evolutionary history and gene flow (Keck et al. 2016); intraspecific variability in responses to environmental factors (Rynearson and Armbrust 2004); and biotic interactions such as herbivory (Fukumori et al. 2015, Passy and Larson 2019), competition (Matthiessen et al. 2010), or co-occurrence (Heino and Grönroos 2013). The importance of each process is likely to vary by spatial scale, from strong environmental control at small scales (intra-habitat); mixed environmental, spatial, evolutionary, and historical control at intermediate scales (inter-habitat); and predominantly spatial, evolutionary, and historical controls at large scales (continents;). A substantial amount of information has yet to be gained about how algal biogeography varies in response to integrated measures of biogeographic controls at different spatial scales.

In the study areas considered in this dissertation, there are several opportunities for further integration of different influences on algal biogeography. In the MSR, diatom taxon co-occurrence (Stanish et al. 2013), intra-habitat dispersal (Stanish 2011), habitat physicochemical conditions (Stanish et al. 2012, Kohler et al. 2016, Darling et al. 2017), and landscape characteristics (Doran et al. 2002) have each been measured extensively. Meanwhile, models have been constructed to predict observed metacommunity structure given inputs of environment, geography, and dispersal (Sokol et al. 2020). By integrating each of these measurements and tools with the inter-habitat, taxon-specific dispersal rates detailed in Chapter II, a thorough landscape model for the processes controlling diatom metacommunity structure in the MSR can be constructed. Importantly, even with the detailed inputs of such a model, assembly processes like geologic history, historical dispersal, genetic drift, herbivory, and competition would still be unaccounted for – highlighting both the difficulty in developing a truly comprehensive understanding of the processes that influence algal biogeography and the need for more ecological measurements.

In this dissertation, I integrated watershed characteristics, climate, and in-stream environmental conditions and accounted for spatial autocorrelation in species distribution models for diatoms and cyanobacteria in the northeast U.S. However, the models did not include explanatory variables that could account for non-environmental processes. Future models could more thoroughly respond to the objective of comparing the influences of different ecosystem processes by including explanatory variables that characterize distance effects (e.g., site, latitude, longitude, principal coordinates of neighbor matrices, direct measurements of aeolian dispersal; Borcard and Legendre 2002) or biotic interactions (e.g., herbivore abundance, edibility, insecticide concentrations). The effects of evolutionary history or gene flow can be assessed by using as a response variable such as the genetic distance among taxon-specific populations instead of taxon relative abundance or presence (Cushman et al. 2006). Assessment of potential indirect effects of environmental factors (e.g., competitive release, herbivory) can be aided by mesocosm experiments that assess taxon-specific and community-scale responses to simulated environmental stressors (Fukumori et al. 2015).

Notably, the goal of fundamental ecological research differs from that of modeling and environmental assessments. While the motivation of basic research is to strive towards complete ecological information, modeling and assessments balance ecological complexity with statistical performance, interpretability, and feasibility (Barbour et al. 1999, Johnson and Omland 2004). Therefore, while it is critical to continue to understand the different ecosystem processes that influence algal biogeography, integrating each of these processes may lead to poorly performing models for use in further research or assessments. As such, the limitations of an approach resulting from the underlying ecological measurements and model assumptions should always be clearly represented. Continued development of methods that allow for the integration of multiple processes will allow for a more holistic picture of algal biogeography to be formed and applied beyond fundamental research.

Conclusions

Water resources worldwide are currently being degraded by anthropogenic impacts at a rapid pace, caused by activities ranging from climate change to pollution from runoff to physical alteration from development. There is an urgent need to conserve inland waters for both human use and biological, chemical, and physical integrity. Given unabated human development and limited environmental protections of freshwater bodies worldwide, changes in algal community composition and taxon distributions can be expected to continue. Understanding the factors that alter algal community and population dynamics is critical to inform management practices aimed at maintaining the current structure and function of aquatic ecosystems and the ecosystem services they provide.

References

Barbour, M. T., Gerritsen, J., Snyder, B. D., & Stribling, J. B. (1999). *Rapid bioassessment protocols for use in streams and wadeable rivers.* U.S. Environmental Protection Agency.

Borcard, D., & Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, *153*, 51-68.

Brown, J. H., & Lomolino, M. V. (2000). Concluding remarks: historical perspective and the future of island biogeography theory. *Global Ecology and Biogeography*, *9*, 87-92.

Cushman, S. A., McKelvey, K. S., Hayden, J., & Schwartz, M. K. (2006). Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *The American Naturalist*, *168*, 486-499.

Darling, J. P., Garland, D. D., Stanish, L. F., Esposito, R. M., Sokol, E. R., & McKnight, D. M. (2017). Thermal autecology describes the occurrence patterns of four benthic diatoms in McMurdo Dry Valley streams. *Polar Biology*, *40*, 2381-2396.

Doran, P. T., McKay, C. P., Clow, G. D., Dana, G. L., Fountain, A. G., Nylen, T., & Lyons, W. B. (2002). Valley floor climate observations from the McMurdo Dry Valleys, Antarctica, 1986–2000. *Journal of Geophysical Research: Atmospheres*, *107*, ACL-13.

Fukumori, K., Livingston, G., & Leibold, M. A. (2015). Disturbance-mediated colonization–extinction dynamics in experimental protist metacommunities. *Ecology*, *96*, 3234-3242.

Heino, J., & Grönroos, M. (2013). Does environmental heterogeneity affect species co-occurrence in ecological guilds across stream macroinvertebrate metacommunities? *Ecography*, *36*, 926-936.

Heino, J., Bini, L. M., Karjalainen, S. M., Mykrä, H., Soininen, J., Vieira, L. C. G., & Diniz-Filho, J. A. F. (2010). Geographical patterns of micro-organismal community structure: are diatoms ubiquitously distributed across boreal streams? *Oikos*, *119*, 129-137.

Heino, J., Melo, A. S., Siqueira, T., Soininen, J., Valanko, S., & Bini, L. M. (2015). Metacommunity organisation, spatial extent and dispersal in aquatic systems: patterns, processes and prospects. *Freshwater Biology*, *60*, 845-869.

Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, *19*, 101-108.

Keck, F., Rimet, F., Franc, A., & Bouchez, A. (2016). Phylogenetic signal in diatom ecology: perspectives for aquatic ecosystems biomonitoring. *Ecological Applications*, *26*, 861-872.

Kohler, T. J., Van Horn, D. J., Darling, J. P., Takacs-Vesbach, C. D., & McKnight, D. M. (2016). Nutrient treatments alter microbial mat colonization in two glacial meltwater streams from the McMurdo Dry Valleys, Antarctica. *FEMS Microbiology Ecology*, *92*, fiw049.

Matthiessen, B., Mielke, E., & Sommer, U. (2010). Dispersal decreases diversity in heterogeneous metacommunities by enhancing regional competition. *Ecology*, *91*, 2022-2033.

Passy, S. I., & Larson, C. A. (2019). Niche dimensionality and herbivory control stream algal biomass via shifts in guild composition, richness, and evenness. *Ecology*, *100*, e02831.

Pickett, S. T., Cadenasso, M. L., Grove, J. M., Boone, C. G., Groffman, P. M., Irwin, E., ... & Pouyat, R. V. (2011). Urban ecological systems: Scientific foundations and a decade of progress. *Journal of Environmental Management*, *92*, 331-362.

Rynearson, T. A., & Armbrust, E. V. (2004). Genetic differentiation among populations of the planktonic marine diatom *Ditylum brightwellii* (Bacillariophyceae). *Journal of Phycology*, *40*, 34-43.

Sakaeva, A., Sokol, E. R., Kohler, T. J., Stanish, L. F., Spaulding, S. A., Howkins, A., ... & McKnight, D. M. (2016). Evidence for dispersal and habitat controls on pond diatom communities from the McMurdo Sound Region of Antarctica. *Polar Biology*, *39*, 2441-2456.

Soininen, J. (2007). Environmental and spatial control of freshwater diatoms—a review. *Diatom Research*, *22*, 473-490.

Soininen, J., & Teittinen, A. (2019). Fifteen important questions in the spatial ecology of diatoms. *Freshwater Biology*, *64*, 2071-2083.

Sokol, E. R., Barrett, J. E., Kohler, T. J., McKnight, D. M., Salvatore, M. R., & Stanish, L. F. (2020). Evaluating Alternative Metacommunity Hypotheses for Diatoms in the McMurdo Dry Valleys Using Simulations and Remote Sensing Data. *Frontiers in Ecology and Evolution*, *8*, 295.

Stanish, L. F. (2011). *Ecological controls on stream diatom communities in the McMurdo Dry Valleys, Antarctica*. Doctoral dissertation, University of Colorado at Boulder.

Stanish, L. F., Kohler, T. J., Esposito, R. M., Simmons, B. L., Nielsen, U. N., Wall, D. H., ... & McKnight, D. M. (2012). Extreme streams: flow intermittency as a control on diatom communities in meltwater streams in the McMurdo Dry Valleys, Antarctica. *Canadian Journal of Fisheries and Aquatic Sciences*, *69*, 1405-1419.

Stanish, L. F., O'Neill, S. P., Gonzalez, A., Legg, T. M., Knelman, J., McKnight, D. M., ... & Nemergut, D. R. (2013). Bacteria and diatom co-occurrence patterns in microbial mats from polar desert streams. *Environmental Microbiology*, *15*, 1115-1131.

BIBLIOGRAPHY

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716-723.

Allan, J. D. (2004). Landscapes and riverscapes: the influence of land use on stream ecosystems. *Annual Review of Ecology and Systematics*, *35*, 257-284.

Amoros, C., & Bornette, G. (2002). Connectivity and biocomplexity in waterbodies of riverine floodplains. *Freshwater Biology*, *47*, 761-776.

Araujo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, *33*, 1677-1688.

Aslim, B., & Ozturk, S. (2009). Toxicity of herbicides to cyanobacterial isolates. *Journal of Environmental Biology*, *30*, 381-384.

Astorga, A., Oksanen, J., Luoto, M., Soininen, J., Virtanen, R., & Muotka, T. (2012). Distance decay of similarity in freshwater communities: do macro-and microorganisms follow the same rules? *Global Ecology and Biogeography*, *21*, 365-375.

Bailey, R. C., Norris, R. H., & Reynoldson, T. B. (2004). Bioassessment of freshwater ecosystems. In Bailey, R. C., Norris, R. H., Reynoldson, T. B. (Eds.), *Bioassessment of Freshwater Ecosystems* (pp. 1-15). Springer, Boston, MA.

Barbour, M. T., Gerritsen, J., Snyder, B. D., & Stribling, J. B. (1999). *Rapid bioassessment protocols for use in streams and wadeable rivers.* U.S. Environmental Protection Agency.

Barton, A. D., Finkel, Z. V., Ward, B. A., Johns, D. G., & Follows, M. J. (2013). On the roles of cell size and trophic strategy in North Atlantic diatom and dinoflagellate communities. *Limnology and Oceanography*, *58*, 254-266.

Batt, A. L., Wathen, J. B., Lazorchak, J. M., Olsen, A. R., & Kincaid, T. M. (2017). Statistical survey of persistent organic pollutants: risk estimations to humans and wildlife through consumption of fish from US rivers. *Environmental Science & Technology*, *51*, 3021-3031.

Battarbee, R. W. (1973). A new method for the estimation of absolute microfossil numbers, with reference especially to diatoms. *Limnology and Oceanography*, *18*, 647-653.

Benbrook, C. M. (2016). Trends in glyphosate herbicide use in the United States and globally. *Environmental Sciences Europe*, *28*, 3.

Bishop, I. W., Esposito, R. M., Tyree, M., & Spaulding, S. A. (2017). A diatom voucher flora from selected southeast rivers (USA). *Phytotaxa*, *332*, 101-140.

Borcard, D., & Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, *153*, 51-68.

Boulangeat, I., Gravel, D., & Thuiller, W. (2012). Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecology Letters*, *15*, 584-593.

Bouma-Gregson, K., Kudela, R. M., & Power, M. E. (2018). Widespread anatoxin-a detection in benthic cyanobacterial mats throughout a river network. *PLoS One*, *13*, e0197669.

Bouma-Gregson, K., Power, M. E., & Bormans, M. (2017). Rise and fall of toxic benthic freshwater cyanobacteria (*Anabaena* spp.) in the Eel river: Buoyancy and dispersal. *Harmful Algae*, *66*, 79-87.

Broady, P. A. (1979). Wind dispersal of terrestrial algae at Signy Island, South Orkney Islands. *British Antarctic Survey Bulletin, 48*, 99-102.

Brown, J. H., & Lomolino, M. V. (2000). Concluding remarks: historical perspective and the future of island biogeography theory. *Global Ecology and Biogeography*, *9*, 87-92.

Brown, R. M. (1971). The distribution of airborne algae and fern spores across the island of Oahu, Hawaii. In Parker, B., Brown, R. M. (Eds), *Contributions in Phycology* (pp. 175-188). Allen Press, Lawrence, KS.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*, 261-304.

Cantonati, M., Kelly, M. G., & Lange-Bertalot, H. (Eds.) (2017). *Freshwater Benthic Diatoms of Central Europe: Over 800 Common Species Used in Ecological Assessment*. Koeltz Botanical Books, Schmitten-Oberreifenberg.

Carmichael, W. W. (2001). Health effects of toxin-producing cyanobacteria: "The CyanoHABs". *Human and Ecological Risk Assessment: An International Journal*, *7*, 1393-1407.

Carrick, H. J., Lowe, R. L., & Rotenberry, J. T. (1988). Guilds of benthic algae along nutrient gradients: relationships to algal community diversity. *Journal of the North American Benthological Society*, *7*, 117-128.

Chu, W. L., Tneh, S. Y., & Ambu, S. (2013). A survey of airborne algae and cyanobacteria within the indoor environment of an office building in Kuala Lumpur, Malaysia. *Grana*, *52*, 207-220.

Cliff, A. D., & Ord, K. (1970). Spatial autocorrelation: a review of existing and new measures with applications. *Economic Geography*, *46*, 269-292.

Coles, J. F., Riva-Murray, K., Van Metre, P. C., Button, D. T., Bell, A. H., Qi, S. L., ... & Sheibley, R. W. (2019). Design and methods of the US Geological Survey Northeast Stream Quality Assessment (NESQA), 2016 (No. 2018-1183). U.S. Geological Survey.

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

Cushman, S. A., McKelvey, K. S., Hayden, J., & Schwartz, M. K. (2006). Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *The American Naturalist*, *168*, 486-499.

Darling, J. P., Garland, D. D., Stanish, L. F., Esposito, R. M., Sokol, E. R., & McKnight, D. M. (2017). Thermal autecology describes the occurrence patterns of four benthic diatoms in McMurdo Dry Valley streams. *Polar Biology*, *40*, 2381-2396.

Davis, T. W., Berry, D. L., Boyer, G. L., & Gobler, C. J. (2009). The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of *Microcystis* during cyanobacteria blooms. *Harmful Algae*, *8*, 715-725.

Debenest, T., Silvestre, J., Coste, M., & Pinelli, E. (2010). Effects of pesticides on freshwater diatoms. *Reviews of Environmental Contamination and Toxicology*, *203*, 87-103.

Deuerling, K. M., Lyons, W. B., Welch, S. A., & Welch, K. A. (2014). The characterization and role of aeolian deposition on water quality, McMurdo Dry Valleys, Antarctica. *Aeolian Research*, *13*, 7-17.

Diaz, M. A., Adams, B. J., Welch, K. A., Welch, S. A., Opiyo, S. O., Khan, A. L., ... & Lyons, W. B. (2018). Aeolian material transport and its role in landscape connectivity in the McMurdo Dry Valleys, Antarctica. *Journal of Geophysical Research: Earth Surface*, *123*, 3323-3337.

Dodds, W. K., & Smith, V. H. (2016). Nitrogen, phosphorus, and eutrophication in streams. *Inland Waters*, *6*, 155-164.

Dodds, W. K., López, A. J., Bowden, W. B., Gregory, S., Grimm, N. B., Hamilton, S. K., ... & Morrall, D. (2002). N uptake as a function of concentration in streams. *Journal of the North American Benthological Society*, *21*, 206-220.

Doran, P. T., McKay, C. P., Clow, G. D., Dana, G. L., Fountain, A. G., Nylen, T., & Lyons, W. B. (2002). Valley floor climate observations from the McMurdo Dry Valleys, Antarctica, 1986–2000. *Journal of Geophysical Research: Atmospheres*, *107*, ACL-13.

Dorigo, U., & Leboulanger, C. (2001). A pulse-amplitude modulated fluorescence-based method for assessing the effects of photosystem II herbicides on freshwater periphyton. *Journal of Applied Phycology*, *13*, 509-515.

Douterelo, I., Perona, E., & Mateo, P. (2004). Use of cyanobacteria to assess water quality in running waters. *Environmental Pollution*, *127*, 377-384.

Du, B., Zhang, Z., Liu, W., Ye, Y., Lu, T., Zhou, Z., ... & Qian, H. (2019). Acute toxicity of the fungicide azoxystrobin on the diatom *Phaeodactylum tricornutum*. *Ecotoxicology and Environmental Safety*, *168*, 72-79.

Dunn, P. K., & Smyth, G. K. (2008). Evaluation of Tweedie exponential dispersion model densities by Fourier inversion. *Statistics and Computing*, *18*, 73-86.

Duong, T. T., Morin, S., Herlory, O., Feurtet-Mazel, A., Coste, M., & Boudou, A. (2008). Seasonal effects of cadmium accumulation in periphytic diatom communities of freshwater biofilms. *Aquatic Toxicology*, *90*, 19-28.

Elbrecht, V., Vamos, E. E., Steinke, D., & Leese, F. (2018). Estimating intraspecific genetic diversity from community DNA metabarcoding data. *PeerJ*, *6*, e4644.

Elith, J., & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 677-697.

Esposito, R. M. M., Horn, S. L., McKnight, D. M., Cox, M. J., Grant, M. C., Spaulding, S. A., ... & Cozzetto, K. D. (2006). Antarctic climate cooling and response of diatoms in glacial meltwater streams. *Geophysical Research Letters*, *33*, L07406.

Esposito, R. M. M., Spaulding, S. A., McKnight, D. M., Van de Vijver, B. A. R. T., Kopalová, K., Lubinski, D., ... & Whittaker, T. (2008). Inland diatoms from the McMurdo dry valleys and James Ross Island, Antarctica. *Botany*, *86*, 1378-1392.

Fasiolo, M., Nedellec, R., Goude, Y., & Wood, S. N. (2020). Scalable visualization methods for modern generalized additive models. *Journal of Computational and Graphical Statistics*, *29*, 78-86.

Finlay, B. J. (2002). Global dispersal of free-living microbial eukaryote species. *Science*, *296*, 1061-1063.

Fisher, N. S., & Wurster, C. F. (1973). Individual and combined effects of temperature and polychlorinated biphenyls on the growth of three species of phytoplankton. *Environmental Pollution*, *5*, 205-212.

Fishman, M. J. (1993). Methods of analysis by the US Geological Survey National Water Quality Laboratory: Determination of inorganic and organic constituents in water and fluvial sediments. U.S. Geological Survey.

Fishman, M. J., & Friedman, L. C. (1989). Ion chromatography, methods for determination of inorganic substances in water and fluvial Sediments. US Geological Survey Techniques of Water Resources Investigations (5A1). U.S. Geological Survey

Fitzpatrick, F. A., White, I. R., D'Arconte, P. J., Meador, M. R., Maupin, M. A., & Gurtz, M. E. (1998). Revised methods for characterizing stream habitat in the National Water-Quality Assessment Program (Report 98-4052). U.S. Geological Survey.

Fleeger, J. W., Carman, K. R., & Nisbet, R. M. (2003). Indirect effects of contaminants in aquatic ecosystems. *Science of the Total Environment*, *317*, 207-233.

Flower, R. J. (1993). Diatom preservation: experiments and observations on dissolution and breakage in modern and fossil material. *Hydrobiologia*, *269*, 473-484.

Forlani, G., Pavan, M., Gramek, M., Kafarski, P., & Lipok, J. (2008). Biochemical bases for a widespread tolerance of cyanobacteria to the phosphonate herbicide glyphosate. *Plant and Cell Physiology*, *49*, 443-456.

Fountain, A. G., Lyons, W. B., Burkins, M. B., Dana, G. L., Doran, P. T., Lewis, K. J., ... & Wall, D. H. (1999). Physical controls on the Taylor Valley ecosystem, Antarctica. *Bioscience, 49,* 961-971.

Fountain, A. G., Tranter, M., Nylen, T. H., Lewis, K. J., & Mueller, D. R. (2004). Evolution of cryoconite holes and their contribution to meltwater runoff from glaciers in the McMurdo Dry Valleys, Antarctica. *Journal of Glaciology, 50,* 35-45.

Fukumori, K., Livingston, G., & Leibold, M. A. (2015). Disturbance-mediated colonization–extinction dynamics in experimental protist metacommunities. *Ecology, 96,* 3234-3242.

Genuer, R., Poggi, J. M., Tuleau-Malot, C., & Genuer, M. R. (2019). Package 'vsurf'. *Pattern Recognition Letters*, *31*, 2225-2236.

Gillies, J. A., Nickling, W. G., & Tilson, M. (2013). Frequency, magnitude, and characteristics of aeolian sediment transport: McMurdo Dry Valleys, Antarctica. *Journal of Geophysical Research: Earth Surface, 118,* 461-479.

Gomes, V. H., IJff, S. D., Raes, N., Amaral, I. L., Salomão, R. P., de Souza Coelho, L., ... & Guevara, J. E. (2018). Species Distribution Modelling: Contrasting presence-only models with plot abundance data. *Scientific Reports, 8,* 1-12.

Gooseff, M. N., McKnight, D. M., Doran, P., Fountain, A. G., & Lyons, W. B. (2011). Hydrological connectivity of the landscape of the McMurdo Dry Valleys, Antarctica. *Geography Compass*, *5,* 666-681.

Gooseff, M. N., McKnight, D., Wall, D., Virginia, R., & Fountain, A. (2016). McMurdo Dry Valleys Long Term Ecological Research. *National Science Foundation Proposal*.

Growns, I. (1999). Is genus or species identification of periphytic diatoms required to determine the impacts of river regulation? *Journal of Applied Phycology*, *11*, 273-283.

Harris, T. D., & Smith, V. H. (2016). Do persistent organic pollutants stimulate cyanobacterial blooms? *Inland Waters*, *6*, 124-130.

Hawes, I., Howard-Williams, C., & Vincent, W. F. (1992). Desiccation and recovery of Antarctic cyanobacterial mats. *Polar Biology*, *12*, 587-594.

Heino, J. (2013). The importance of metacommunity ecology for environmental assessment research in the freshwater realm. *Biological Reviews, 88,* 166-178.

Heino, J., & Grönroos, M. (2013). Does environmental heterogeneity affect species co-occurrence in ecological guilds across stream macroinvertebrate metacommunities? *Ecography*, *36*, 926-936.

Heino, J., & Soininen, J. (2006). Regional occupancy in unicellular eukaryotes: a reflection of niche breadth, habitat availability or size-related dispersal capacity? *Freshwater Biology*, *51*, 672-685.

Heino, J., Bini, L. M., Karjalainen, S. M., Mykrä, H., Soininen, J., Vieira, L. C. G., & Diniz-Filho, J. A. F. (2010). Geographical patterns of micro-organismal community structure: are diatoms ubiquitously distributed across boreal streams? *Oikos*, *119,* 129-137.

Heino, J., Grönroos, M., Soininen, J., Virtanen, R., & Muotka, T. (2012). Context dependency and metacommunity structuring in boreal headwater streams. *Oikos*, *121*, 537-544.

Heino, J., Melo, A. S., Siqueira, T., Soininen, J., Valanko, S., & Bini, L. M. (2015). Metacommunity organisation, spatial extent and dispersal in aquatic systems: patterns, processes and prospects. *Freshwater Biology*, *60,* 845-869.

Heino, J., Tolkkinen, M., Pirttilä, A. M., Aisala, H., & Mykrä, H. (2014). Microbial diversity and community–environment relationships in boreal streams. *Journal of Biogeography*, *41*, 2234-2244.

Heisler, J., Glibert, P. M., Burkholder, J. M., Anderson, D. M., Cochlan, W., Dennison, W. C., ... & Lewitus, A. (2008). Eutrophication and harmful algal blooms: a scientific consensus. *Harmful Algae*, *8*, 3-13.

Hill, B. H., Stevenson, R. J., Pan, Y., Herlihy, A. T., Kaufmann, P. R., & Johnson, C. B. (2001). Comparison of correlations between environmental characteristics and stream diatom assemblages characterized at genus and species levels. *Journal of the North American Benthological Society*, *20,* 299-310.

Hobbs, W. O., Collyard, S. A., Larson, C., Carey, A. J., & O'Neill, S. M. (2019). Toxic burdens of freshwater biofilms and use as a source tracking tool in rivers and streams. *Environmental Science & Technology*, *53,* 11102-11111.

Holyoak, M., Leibold, M. A., & Holt, R. D. (Eds.) (2005). *Metacommunities: Spatial Dynamics and Ecological Communities*. University of Chicago Press, Chicago, IL.

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., ... & Megown, K. (2015). Completion of the 2011 National Land Cover Database for the conterminous United States–representing a decade of land cover change information. *Photogrammetric Engineering & Remote Sensing*, *81,* 345-354.

Howard, C., Stephens, P. A., Pearce-Higgins, J. W., Gregory, R. D., & Willis, S. G. (2014). Improving species distribution models: the value of data on abundance. *Methods in Ecology and Evolution*, *5,* 506-513.

James, F. C., & McCulloch, C. E. (1990). Multivariate analysis in ecology and systematics: panacea or Pandora's box? *Annual Review of Ecology and Systematics*, *21*, 129-166.

Johansen, J. R. (1999). Diatoms of Aerial Habitats. In Stoermer, E. F., Smol, J. P. (Eds.), *The Diatoms: Applications for the Environmental and Earth Sciences* (pp. 264-274). Cambridge University Press, United Kingdom.

Johansen, J. R., & Casamatta, D. A. (2005). Recognizing cyanobacterial diversity through adoption of a new species paradigm. *Algological Studies*, *117*, 71-93.

Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, *19*, 101-108.

Jyrkänkallio-Mikkola, J., Meier, S., Heino, J., Laamanen, T., Pajunen, V., Tolonen, K. T., ... & Soininen, J. (2017). Disentangling multi-scale environmental effects on stream microbial communities. *Journal of Biogeography*, *44*, 1512-1523.

Kaushal, S. S., Likens, G. E., Jaworski, N. A., Pace, M. L., Sides, A. M., Seekell, D., ... & Wingate, R. L. (2010). Rising stream and river temperatures in the United States. *Frontiers in Ecology and the Environment*, *8*, 461-466.

Kaushal, S. S., Likens, G. E., Pace, M. L., Utz, R. M., Haq, S., Gorman, J., & Grese, M. (2018). Freshwater salinization syndrome on a continental scale. *Proceedings of the National Academy of Sciences*, *115*, E574-E583.

Keck, F., Rimet, F., Franc, A., & Bouchez, A. (2016). Phylogenetic signal in diatom ecology: perspectives for aquatic ecosystems biomonitoring. *Ecological Applications*, *26*, 861-872.

Khan, A. L., McMeeking, G. R., Schwarz, J. P., Xian, P., Welch, K. A., Berry Lyons, W., & McKnight, D. M. (2018). Near-Surface Refractory Black Carbon Observations in the Atmosphere and Snow in the McMurdo Dry Valleys, Antarctica, and Potential Impacts of Foehn Winds. *Journal of Geophysical Research: Atmospheres*, *123*, 2877-2887.

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., ... & Kelley, S. T. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nature Methods*, *8*, 761.

Kociolek, J. P., Kopalová, K., Hamsher, S. E., Kohler, T. J., Van de Vijver, B., Convey, P., & McKnight, D. M. (2017). Freshwater diatom biogeography and the genus *Luticola*: an extreme case of endemism in Antarctica. *Polar Biology*, *40*, 1185-1196.

Kohler, T. J., Kopalová, K., Van De Vijver, B., & Kociolek, J. P. (2015). The genus *Luticola* DG Mann (Bacillariophyta) from the McMurdo Sound Region, Antarctica, with the description of four new species. *Phytotaxa*, *208*, 103-134.

Kohler, T. J., Van Horn, D. J., Darling, J. P., Takacs-Vesbach, C. D., & McKnight, D. M. (2016). Nutrient treatments alter microbial mat colonization in two glacial meltwater streams from the McMurdo Dry Valleys, Antarctica. *FEMS Microbiology Ecology*, *92*, fiw049.

Kok, J. F., Parteli, E. J., Michaels, T. I., & Karam, D. B. (2012). The physics of wind-blown sand and dust. *Reports on Progress in Physics*, *75*, 106901.

Komárek, J. (2006). Cyanobacterial taxonomy: current problems and prospects for the integration of traditional and molecular approaches. *Algae*, *21*, 349-375.

Komárek, J., & Johansen, J. R. (2015a). Coccoid cyanobacteria. In Wehr, J. D., Sheath, R. G., Kociolek, J. P. (Eds.), *Freshwater Algae of North America: Ecology and Classification* (pp. 75-135). Academic Press, Amsterdam.

Komárek, J., & Johansen, J. R. (2015b). Filamentous cyanobacteria. In Wehr, J. D., Sheath, R. G., Kociolek, J. P. (Eds.), *Freshwater Algae of North America: Ecology and Classification* (pp. 135-235). Academic Press, San Diego, CA.

Kostel, J. A., Wang, H., Amand, A. L. S., & Gray, K. A. (1999). 1. Use of a novel laboratory stream system to study the ecological impact of PCB exposure in a periphytic biolayer. *Water Research*, *33*, 3735-3748.

Krammer, K. (2000). *Diatoms of Europe: Diatoms of the European Inland Waters and Comparable Habitats. Volume 1: The Genus Pinnularia*. ARG Gantner Varlag, Ruggell, Germany.

Kristiansen, J. (1996). Dispersal of freshwater algae—a review. *Hydrobiologia*, *336*, 151-157.

Kumar, S., Spaulding, S. A., Stohlgren, T. J., Hermann, K. A., Schmidt, T. S., & Bahls, L. L. (2009). Potential habitat distribution for the freshwater diatom *Didymosphenia geminata* in the continental US. *Frontiers in Ecology and the Environment*, *7*, 415-420.

Kupper, H., & Kroneck, P. M. (2005). Heavy metal uptake by plants and cyanobacteria. *Metal Ions in Biological Systems*, *44*, 97.

Lake, P. S. (2003). Ecological effects of perturbation by drought in flowing waters. *Freshwater Biology*, *48*, 1161-1172.

Larras, F., Keck, F., Montuelle, B., Rimet, F., & Bouchez, A. (2014). Linking diatom sensitivity to herbicides to phylogeny: a step forward for biomonitoring? *Environmental Science & Technology*, *48*, 1921-1930.

Lebre, P. H., De Maayer, P., & Cowan, D. A. (2017). Xerotolerant bacteria: surviving through a dry spell. *Nature Reviews Microbiology*, *15*, 285.

Liu, J., Soininen, J., Han, B. P., & Declerck, S. A. (2013). Effects of connectivity, dispersal directionality and functional traits on the metacommunity structure of river benthic diatoms. *Journal of Biogeography*, *40*, 2238-2248.

Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010). Understanding of internal clustering validation measures. In *2010 IEEE International Conference on Data Mining* (pp. 911-916). IEEE.

Logue, J. B., Mouquet, N., Peter, H., Hillebrand, H., & Metacommunity Working Group. (2011). Empirical approaches to metacommunities: a review and comparison with theory. *Trends in Ecology & Evolution*, *26*, 482-491.

Lopes, P. M., Bozelli, R., Bini, L. M., Santangelo, J. M., & Declerck, S. A. (2016). Contributions of airborne dispersal and dormant propagule recruitment to the assembly of rotifer and crustacean zooplankton communities in temporary ponds. *Freshwater Biology*, *61*, 658-669.

Lowe, R. L. (1979). Phytobenthic ecology and regulated streams. In Ward, J. (Ed.), *The Ecology of Regulated Streams* (pp. 25-34). Springer, Boston, MA.

Lowe, R., & Kheiri, S. (2015). *Cyclotella meneghiniana*. In Diatoms of North America. Retrieved November 3, 2020, from https://diatoms.org/species/cyclotella_meneghiniana

Lyons, W. B., Frape, S. K., & Welch, K. A. (1999). History of McMurdo Dry Valley lakes, Antarctica, from stable chlorine isotope data. *Geology*, *27*, 527-530.

MacDonald, D. D., Ingersoll, C. G., & Berger, T. A. (2000). Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems. *Archives of Environmental Contamination and Toxicology*, *39*, 20-31.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281-297).

Maher, B. A., Prospero, J. M., Mackie, D., Gaiero, D., Hesse, P. P., & Balkanski, Y. (2010). Global connections between aeolian dust, climate and ocean biogeochemistry at the present day and at the last glacial maximum. *Earth-Science Reviews*, *99*, 61-97.

Mahler, B. J., Van Metre, P. C., Wilson, J. T., Musgrove, M., Zaugg, S. D., & Burkhardt, M. R. (2009). Fipronil and its degradates in indoor and outdoor dust. *Environmental Science & Technology*, *43*, 5665–5670.

Maloney, K. O., & Weller, D. E. (2011). Anthropogenic disturbance and streams: land use and land-use change affect stream ecosystems via multiple pathways. *Freshwater Biology*, *56*, 611-626.

Marra, G., & Wood, S. N. (2011). Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis*, *55*, 2372-2387.

Marshall, W. A., & Chalmers, M. O. (1997). Airborne dispersal of Antarctic terrestrial algae and cyanobacteria. *Ecography*, *20*, 585-594.

Martiny, J. B. H., Bohannan, B. J., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., ... & Morin, P. J. (2006). Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, *4*, 102-112.

Mass, A. Q. (2018). *Biogeochemistry, Contaminant Transport, and Atmospheric Exchange In Glacial Cryoconite Meltwater of the McMurdo Dry Valleys, Antarctica*. Doctoral dissertation, University of Colorado at Boulder.

Matthiessen, B., Mielke, E., & Sommer, U. (2010). Dispersal decreases diversity in heterogeneous metacommunities by enhancing regional competition. *Ecology, 91*, 2022-2033.

McAllister, T. G., Wood, S. A., Atalah, J., & Hawes, I. (2018). Spatiotemporal dynamics of *Phormidium* cover and anatoxin concentrations in eight New Zealand rivers with contrasting nutrient and flow regimes. *Science of the Total Environment, 612*, 71-80.

McCormick, P. V., & Cairns, J. (1994). Algae as indicators of environmental change. *Journal of Applied Phycology, 6*, 509-526.

McCormick, P. V., & Stevenson, R. J. (1991). Mechanisms of benthic algal succession in lotic environments. *Ecology, 72*, 1835-1848.

McKnight, D. M., Niyogi, D. K., Alger, A. S., Bomblies, A., Conovitz, P. A., & Tate, C. M. (1999). Dry valley streams in Antarctica: ecosystems waiting for water. *Bioscience, 49*, 985-995.

Monica, A., & Agrawal, M. K. (2011). Cyanobacteria-herbivore interaction in freshwater ecosystem. *Journal of Microbiology and Biotechnology Research, 1*, 52-66.

Moran, P. W., Nowell, L. H., Kemble, N. E., Mahler, B. J., Waite, I. R., & Van Metre, P. C. (2017). Influence of sediment chemistry and sediment toxicity on macroinvertebrate communities across 99 wadable streams of the Midwestern USA. *Science of the Total Environment, 599*, 1469-1478.

Morán-Ordóñez, A., Pavlova, A., Pinder, A. M., Sim, L., Sunnucks, P., Thompson, R. M., & Davis, J. (2015). Aquatic communities in arid landscapes: local conditions, dispersal traits and landscape configuration determine local biodiversity. *Diversity and Distributions, 21*, 1230-1241.

Morin, S., Cordonier, A., Lavoie, I., Arini, A., Blanco, S., Duong, T. T., ... & Laviale, M. (2012). Consistency in diatom response to metal-contaminated environments. In Guasch, H., Ginebreda, A., Geiszinger, A. (Eds.), *Emerging and Priority Pollutants in Rivers* (pp. 117-146). Springer, Berlin, Germany.

Moulton, S. R., Kennen, J. G., Goldstein, R. M., & Hambrook, J. A. (2002). Revised protocols for sampling algal, invertebrate, and fish communities as part of the National Water-Quality Assessment Program (No. 2002-150). Geological Survey (US).

Munn, M. D., Waite, I., & Konrad, C. P. (2018). Assessing the influence of multiple stressors on stream diatom metrics in the upper Midwest, USA. *Ecological Indicators, 85*, 1239-1248.

Nkem, J. N., Wall, D. H., Virginia, R. A., Barrett, J. E., Broos, E. J., Porazinska, D. L., & Adams, B. J. (2006). Wind dispersal of soil invertebrates in the McMurdo Dry Valleys, Antarctica. *Polar Biology, 29*, 346-352.

Nylen, T. H., Fountain, A. G., & Doran, P. T. (2004). Climatology of katabatic winds in the McMurdo dry valleys, southern Victoria Land, Antarctica. *Journal of Geophysical Research: Atmospheres*, *109*, D03114.

O'Dell, J. W. (1993). *EPA Method 365.1—Determination of phosphorus by semi-automated colorimetry.* U.S. Environmental Protection Agency, Office of Research and Development.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., ... & Wagner, H. (2019). vegan: Community Ecology Package. R package version 2.5-6. Available at https://CRAN.R-project.org/package=vegan.

Paerl, H. W. (2018). Mitigating toxic planktonic cyanobacterial blooms in aquatic ecosystems facing increasing anthropogenic and climatic pressures. *Toxins*, *10,* 76.

Paerl, H. W., Fulton, R. S., Moisander, P. H., & Dyble, J. (2001). Harmful freshwater algal blooms, with an emphasis on cyanobacteria. *The Scientific World Journal*, *1*, 76-113.

Pajunen, V., Jyrkänkallio-Mikkola, J., Luoto, M., & Soininen, J. (2019). Are drivers of microbial diatom distributions context dependent in human-impacted and pristine environments? *Ecological Applications, 29*, e01917.

Pajunen, V., Luoto, M., & Soininen, J. (2016). Climate is an important driver for stream diatom distributions. *Global Ecology and Biogeography*, *25,* 198-206.

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

Pandey, L. K., Bergey, E. A., Lyu, J., Park, J., Choi, S., Lee, H., ... & Han, T. (2017). The use of diatoms in ecotoxicology and bioassessment: insights, advances and challenges. *Water Research*, *118,* 39-58.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*, 289-290.

Passy, S. I., & Larson, C. A. (2019). Niche dimensionality and herbivory control stream algal biomass via shifts in guild composition, richness, and evenness. *Ecology*, *100,* e02831.

Patrick, R. (1973). Use of algae, especially diatoms, in the assessment of water quality. In Cairns, J., Dickson, K. L. (Eds.), *Biological Methods for the Assessment of Water Quality* (pp. 76-95). ASTM International, Philadelphia, PA.

Patton, C. J., & Kryskalla, J. R. (2011). Colorimetric determination of nitrate plus nitrite in water by enzymatic reduction, automated discrete analyzer methods. *U.S. Geological Survey Techniques and Methods*, *5,* B8.

Pearson, R. G. (2007). Species' distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History*, *50,* 54-89.

Pérez, G. L., Torremorell, A., Mugni, H., Rodriguez, P., Vera, M. S., Nascimento, M. D., ... & Izaguirre, I. (2007). Effects of the herbicide Roundup on freshwater microbial communities: a mesocosm study. *Ecological Applications*, *17*, 2310-2322.

Pickett, S. T., Cadenasso, M. L., Grove, J. M., Boone, C. G., Groffman, P. M., Irwin, E., ... & Pouyat, R. V. (2011). Urban ecological systems: Scientific foundations and a decade of progress. *Journal of Environmental Management*, *92*, 331-362.

Potapova, M., & Charles, D. F. (2003). Distribution of benthic diatoms in US rivers in relation to conductivity and ionic composition. *Freshwater Biology*, *48*, 1311-1328.

Potapova, M., & Charles, D. F. (2007). Diatom metrics for monitoring eutrophication in rivers of the United States. *Ecological Indicators*, *7*, 48-70.

Powell, H. A., Kerbby, N. W., & Rowell, P. (1991). Natural tolerance of cyanobacteria to the herbicide glyphosate. *New Phytologist*, *119*, 421-426.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org/.

Radtke, D. B. (2005). *Bottom-material samples: U.S. Geological Survey Techniques of Water-Resources Investigations* (Vol. 9, No. A8). U.S. Geological Survey.

Ram, K., & Wickham, H. (2018). wesanderson: a Wes Anderson palette generator. R package version 0.3.6. Available at https://cran.r-project.org/package=wesanderson.

Ribeiro, K. F., Duarte, L., & Crossetti, L. O. (2018). Everything is not everywhere: a tale on the biogeography of cyanobacteria. *Hydrobiologia*, *820*, 23-48.

Rimet, F., & Bouchez, A. (2011). Use of diatom life-forms and ecological guilds to assess pesticide contamination in rivers: lotic mesocosm approaches. *Ecological Indicators*, *11*, 489-499.

Roberts, D. W. (2019). labdsv: Ordination and Multivariate Analysis for Ecology. R package version 2.0-1. Available at https://CRAN.R-project.org/package=labdsv.

Robinson, A. A., Belden, J. B., & Lydy, M. J. (2005). Toxicity of fluoroquinolone antibiotics to aquatic organisms. *Environmental Toxicology and Chemistry: An International Journal*, *24*, 423-430.

Robson, B. J. (2000). Role of residual biofilm in the recolonization of rocky intermittent streams by benthic algae. *Marine and Freshwater Research*, *51*, 725-732.

Rosemond, A. D., Mulholland, P. J., & Elwood, J. W. (1993). Top-down and bottom-up control of stream periphyton: effects of nutrients and herbivores. *Ecology*, *74*, 1264-1280.

Roubeix, V., Mazzella, N., Schouler, L., Fauvelle, V., Morin, S., Coste, M., ... & Margoum, C. (2011). Variations of periphytic diatom sensitivity to the herbicide diuron and relation to species distribution in a contamination gradient: implications for biomonitoring. *Journal of Environmental Monitoring*, *13*, 1768-1774.

Rushforth, S. R., Brotherson, J. D., Fungladda, N., & Evenson, W. E. (1981). The effects of dissolved heavy metals on attached diatoms in the Uintah Basin of Utah, USA. *Hydrobiologia*, *83*, 313-323.

Rynearson, T. A., & Armbrust, E. V. (2004). Genetic differentiation among populations of the planktonic marine diatom *Ditylum brightwellii* (Bacillariophyceae). *Journal of Phycology*, *40*, 34-43.

Rynearson, T. A., Lin, E. O., & Armbrust, E. V. (2009). Metapopulation structure in the planktonic diatom *Ditylum brightwellii* (Bacillariophyceae). *Protist*, *160*, 111-121.

Šabacká, M., Priscu, J. C., Basagic, H. J., Fountain, A. G., Wall, D. H., Virginia, R. A., & Greenwood, M. C. (2012). Aeolian flux of biotic and abiotic material in Taylor Valley, Antarctica. *Geomorphology*, *155*, 102-111.

Sakaeva, A., Sokol, E. R., Kohler, T. J., Stanish, L. F., Spaulding, S. A., Howkins, A., ... & McKnight, D. M. (2016). Evidence for dispersal and habitat controls on pond diatom communities from the McMurdo Sound Region of Antarctica. *Polar Biology*, *39*, 2441-2456.

Sandstrom, M. W., Kanagy, L. K., Anderson, C. A., & Kanagy, C. J. (2016). Determination of pesticides and pesticide degradates in filtered water by direct aqueous-injection liquid chromatography-tandem mass spectrometry (No. 5-B11). U.S. Geological Survey.

Scott, J. T., & Marcarelli, A. M. (2012). Cyanobacteria in freshwater benthic environments. In Whitton, B. A., Potts, M. (Eds.), *The Ecology of the Cyanobacteria II* (pp. 271-289). Springer, Dordrecht.

Sharma, N. K., Rai, A. K., Singh, S., & Brown, R. M. (2007). Airborne algae: their present status and relevance. *Journal of Phycology*, *43*, 615-627.

Shelton, L. R., & Capel, P. D. (1994). Guidelines for collecting and processing samples of stream bed sediment for analysis of trace elements and organic contaminants for the National Water-Quality Assessment Program (Vol. 94, No. 458). U.S. Geological Survey.

Sherwood, A. R., Dittbern, M. N., Johnston, E. T., & Conklin, K. Y. (2017). A metabarcoding comparison of windward and leeward airborne algal diversity across the Ko'olau mountain range on the island of O'ahu, Hawai'i. *Journal of Phycology*, *53*, 437-445.

Simberloff, D., & Dayan, T. (1991). The guild concept and the structure of ecological communities. *Annual Review of Ecology and Systematics*, *22*, 115-143.

Singh, D. P., Khattar, J. I. S., Kaur, G., & Singh, Y. (2016). Toxicological impact of herbicides on cyanobacteria. *Annual Research & Review in Biology*, *9*, 1-39.

Soininen, J. (2007). Environmental and spatial control of freshwater diatoms—a review. *Diatom Research*, *22*, 473-490.

Soininen, J. (2014). A quantitative analysis of species sorting across organisms and ecosystems. *Ecology*, *95*, 3284-3292.

Soininen, J., & Teittinen, A. (2019). Fifteen important questions in the spatial ecology of diatoms. *Freshwater Biology*, *64*, 2071-2083.

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

Soininen, J., Jamoneau, A., Rosebery, J., Leboucher, T., Wang, J., Kokociński, M., & Passy, S. I. (2019). Stream diatoms exhibit weak niche conservation along global environmental and climatic gradients. *Ecography*, *42*, 346-353.

Soininen, J., Paavola, R., & Muotka, T. (2004). Benthic diatom communities in boreal streams: community structure in relation to environmental and spatial gradients. *Ecography*, *27*, 330-342.

Sokol, E. R., Barrett, J. E., Kohler, T. J., McKnight, D. M., Salvatore, M. R., & Stanish, L. F. (2020). Evaluating Alternative Metacommunity Hypotheses for Diatoms in the McMurdo Dry Valleys Using Simulations and Remote Sensing Data. *Frontiers in Ecology and Evolution*, *8*, 295.

Sokol, E. R., Brown, B. L., & Barrett, J. E. (2017). A simulation-based approach to understand how metacommunity characteristics influence emergent biodiversity patterns. *Oikos*, *126*, 723-737.

Sommers, P., Darcy, J. L., Porazinska, D. L., Gendron, E., Fountain, A., Zamora, F., ... & Ryder, J. (2019). Comparison of microbial communities in the sediments and water columns of frozen cryoconite holes in the McMurdo Dry Valleys, Antarctica. *Frontiers in Microbiology*, *10*, 65.

Souffreau, C., Vanormelingen, P., Verleyen, E., Sabbe, K., & Vyverman, W. (2010). Tolerance of benthic diatoms from temperate aquatic and terrestrial habitats to experimental desiccation and temperature stress. *Phycologia*, *49*, 309-324.

Spaulding, S. A., Bishop, I. W., Edlund, M. B., Lee, S., Furey, P., Jovanovska, E., & Potapova, M. (2019). Diatoms of North America. Available at https://diatoms.org/.

Speirs, J. C., Steinhoff, D. F., McGowan, H. A., Bromwich, D. H., & Monaghan, A. J. (2010). Foehn winds in the McMurdo Dry Valleys, Antarctica: The origin of extreme warming events. *Journal of Climate*, *23*, 3577-3598.

Stanish, L. F. (2011). *Ecological controls on stream diatom communities in the McMurdo Dry Valleys, Antarctica*. Doctoral dissertation, University of Colorado at Boulder.

Stanish, L. F., Bagshaw, E. A., McKnight, D. M., Fountain, A. G., & Tranter, M. (2013). Environmental factors influencing diatom communities in Antarctic cryoconite holes. *Environmental Research Letters*, *8*, 045006.

Stanish, L. F., Kohler, T. J., Esposito, R. M., Simmons, B. L., Nielsen, U. N., Wall, D. H., ... & McKnight, D. M. (2012). Extreme streams: flow intermittency as a control on diatom communities in meltwater streams in the McMurdo Dry Valleys, Antarctica. *Canadian Journal of Fisheries and Aquatic Sciences, 69*, 1405-1419.

Stanish, L. F., Nemergut, D. R., & McKnight, D. M. (2011). Hydrologic processes influence diatom community composition in Dry Valley streams. *Journal of the North American Benthological Society*, *30*, 1057-1073.

Stanish, L. F., O'Neill, S. P., Gonzalez, A., Legg, T. M., Knelman, J., McKnight, D. M., ... & Nemergut, D. R. (2013). Bacteria and diatom co-occurrence patterns in microbial mats from polar desert streams. *Environmental Microbiology*, *15*, 1115-1131.

Stevenson, R. J., Bothwell, M. L., Lowe, R. L., & Thorp, J. H. (Eds.) (1996). *Algal Ecology: Freshwater Benthic Ecosystem*. Academic Press, San Diego, CA.

Stevenson, R. J., Hill, B. H., Herlihy, A. T., Yuan, L. L., & Norton, S. B. (2008). Algae–P relationships, thresholds, and frequency distributions guide nutrient criterion development. *Journal of the North American Benthological Society*, *27*, 783-799.

Stevenson, R. J., Pan, Y., Manoylov, K. M., Parker, C. A., Larsen, D. P., & Herlihy, A. T. (2008). Development of diatom indicators of ecological conditions for streams of the western US. *Journal of the North American Benthological Society*, *27*, 1000-1016.

Tesson, S. V., Skjøth, C. A., Šantl-Temkiv, T., & Löndahl, J. (2016). Airborne microalgae: Insights, opportunities, and challenges. *Applied Environmental Microbiology, 82*, 1978-1991.

Turak, E., Flack, L. K., Norris, R. H., Simpson, J., & Waddell, N. (1999). Assessment of river condition at a large spatial scale using predictive models. *Freshwater Biology, 41*, 283-298.

Tweedie, M. C. (1984). An index which distinguishes between some important exponential families. In Ghosh, J. K., Roy, J. (Eds.), *Statistics: Applications and new directions: Proc. Indian statistical institute golden Jubilee International conference* (pp. 579-604). Indian Statistical Institute, Calcutta.

Tyree, M. A., Bishop, I. W., Hawkins, C. P., Mitchell, R., & Spaulding, S. A. (2020). Reduction of taxonomic bias in diatom species data. *Limnology and Oceanography: Methods, 18*, 271-279.

U.S. Geological Survey. (2019). BioData—Aquatic bioassessment data for the nation available on the world wide web. Accessed October 10, 2019. doi:10.5066/F77W698B

Välitalo, P., Kruglova, A., Mikola, A., & Vahala, R. (2017). Toxicological impacts of antibiotics on aquatic micro-organisms: a mini-review. *International Journal of Hygiene and Environmental Health*, *220*, 558-569.

Van Dam, M. H., & Matzke, N. J. (2016). Evaluating the influence of connectivity and distance on biogeographical patterns in the south-western deserts of North America. *Journal of Biogeography*, *43*, 1514-1532.

Van Sickle, J., Baker, J., … & Li, J. (2004). Projecting the biological condition of streams under alternative scenarios of human land use. *Ecological Applications*, *14*, 368-380.

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

Vanschoenwinkel, B., Gielen, S., Seaman, M., & Brendonck, L. (2008). Any way the wind blows – frequent wind dispersal drives species sorting in ephemeral aquatic communities. *Oikos*, *117*, 125-134.

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

Vu, V. Q. (2011). ggbiplot: A ggplot2 based biplot. R package version 0.5. Available at http://github.com/vqv/ggbiplot.

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

Wagner, R. J., Moran, P. W., Zaugg, S. D., Sevigny, J. M., & Pope, J. M. (2016). Contaminants of emerging concern in the lower Stillaguamish River Basin, Washington, 2008–11 (Open-File Report 2014-1028). U.S. Geological Survey.

Wang, H., Kostel, J. A., Amand, A. L. S., & Gray, K. A. (1999). 2. The response of a laboratory stream system to PCB exposure: study of periphytic and sediment accumulation patterns. *Water Research*, *33*, 3749-3761.

Weilhoefer, C. L., & Pan, Y. (2008). 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. *Hydrobiologia*, *614*, 259-274.

Whitehead, P. G., Wilby, R. L., Battarbee, R. W., Kernan, M., & Wade, A. J. (2009). A review of the potential impacts of climate change on surface water quality. *Hydrological Sciences Journal*, *54*, 101-123.

Wilson, B. A., Smith, V. H., deNoyelles, F., & Larive, C. K. (2003). Effects of three pharmaceutical and personal care products on natural freshwater algal assemblages. *Environmental Science & Technology*, *37*, 1713-1719.

Wilson, J. B. (1999). Guilds, functional types and ecological groups. *Oikos*, 507-522.

Witherow, R. A., Lyons, W. B., Bertler, N. A., Welch, K. A., Mayewski, P. A., Sneed, S. B., ... & Fountain, A. (2006). The aeolian flux of calcium, chloride and nitrate to the McMurdo Dry Valleys landscape: evidence from snow pit analysis. *Antarctic Science*, *18*, 497-505.

Wlostowski, A. N., Gooseff, M. N., McKnight, D. M., Jaros, C., & Lyons, W. B. (2016). Patterns of hydrologic connectivity in the McMurdo Dry Valleys, Antarctica: a synthesis of 20 years of hydrologic data. *Hydrological Processes*, *30*, 2958-2975.

Wood, R. J., Mitrovic, S. M., Lim, R. P., Warne, M. S. J., Dunlop, J., & Kefford, B. J. (2019). Benthic diatoms as indicators of herbicide toxicity in rivers–A new SPEcies At Risk (SPEARherbicides) index. *Ecological Indicators*, *99*, 203-213.

Wood, S. A., Atalah, J., Wagenhoff, A., Brown, L., Doehring, K., Young, R. G., & Hawes, I. (2017). Effect of river flow, temperature, and water chemistry on proliferations of the benthic anatoxin-producing cyanobacterium *Phormidium*. *Freshwater Science*, *36*, 63-76.

Wood, S. A., Kelly, L. T., Bouma-Gregson, K., Humbert, J. F., Laughinghouse IV, H. D., Lazorchak, J., ... & Quiblier, C. (2020). Toxic benthic freshwater cyanobacterial proliferations: Challenges and solutions for enhancing knowledge and improving monitoring and mitigation. *Freshwater Biology*, *65*, 1824-1842.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, *99*, 673-686.

Wood, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, *62*, 1025-1036.

Wood, S. N. (2020). mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. R package version 1.8-33. Available at https://cran.r-project.org/package=mgcv.

Yee, T. W., & Mitchell, N. D. (1991). Generalized additive models in plant ecology. *Journal of Vegetation Science*, *2*, 587-602.

Yuan, L. L. (2004). Assigning macroinvertebrate tolerance classifications using generalised additive models. *Freshwater Biology*, *49*, 662-677.

Zhang, H., Hu, C., Jia, X., Xu, Y., Wu, C., Chen, L., & Wang, F. (2012). Characteristics of γ-hexachlorocyclohexane biodegradation by a nitrogen-fixing cyanobacterium, *Anabaena azotica*. *Journal of Applied Phycology*, *24*, 221-225.

Zhang, H., Jiang, X., Lu, L., & Xiao, W. (2015). Biodegradation of polychlorinated biphenyls (PCBs) by the novel identified cyanobacterium *Anabaena* PD-1. *PLoS One*, *10*, e0131450.

APPENDIX

Chapter II Supplementary Material

Table S1. Similarity percentage (SIMPER, presented as proportion) of taxon relative abundance differences in pairwise site comparisons. Underlined and bolded taxa differed in relative abundance substantially across all sites (SIMPER permutation significance test).

	TG-G_B		LB-T_B		VG-T_F		CG-G_F		EC-T_F	
	Taxon	Contr.	Taxon	Contr.	Taxon	Contr.	Taxon	Contr.	Taxon	Contr.
LB-T_B	*P. germainii*	0.24								
	M. meridionalis	0.36								
	M. atomus	0.47								
VG-T_F	*P. germainii*	0.21	*M. atomus*	0.09						
	H. arcuata var. *par.*	0.35	*H. arcuata* var. *par.*	0.18						
	M. meridionalis	0.46	*P. papilio*	0.26						
CG-G_F	*P. germainii*	0.28	*M. atomus*	0.11	*H. arcuata var. par.*	0.16				
	M. meridionalis	0.40	***L. gaussi***	0.20	*M. atomus*	0.28				
	L. laeta	0.49	***L. laeta***	0.28	*L. laeta*	0.35				
EC-T_F	*P. germainii*	0.22	*M. atomus*	0.14	*M. atomus*	0.14	*M. atomus*	0.22		
	M. atomus	0.39	*H. arcuata* var. *par.*	0.24	*H. arcuata* var. *par.*	0.23	*H. arcuata* var. *par.*	0.36		
	H. arcuata* var. *par.	0.53	*P. papilio*	0.32	*H. amphioxys*	0.29	***L. laeta***	0.45		
MS-T_M	***M. atomus***	0.22	*M. atomus*	0.20	*M. atomus*	0.20	***M. atomus***	0.25	***M. atomus var. 1***	0.15
	P. germainii	0.42	*H. arcuata* var. *par.*	0.35	***M. atomus* var. 1**	0.32	***H. arcuata* var. *par.***	0.42	*M. atomus*	0.29
	H. arcuata* var. *par.	0.59	***M. atomus* var. 1**	0.47	*H. arcuata* var. *par.*	0.39	***M. atomus* var. 1**	0.53	*H. arcuata* var. *par.*	0.37

Contr. = cumulative proportion (contribution) of pairwise site differences explained by each taxon

H. arcuata var. *parallela* abbreviated to *H. arcuata* var. *par.*

$p < 0.1$, **$p < 0.05$**

Table S2. Ten lowest mean Bray-Curtis dissimilarities for pairwise comparisons of size class composition between aeolian and regional communities. Lower mean dissimilarities for each aeolian site have darker backgrounds.

Longitude →

Longitude ↓

	TG-G_B	LB-T_B	VG-T_F	CG-G_F	EC-T_F	MS-T_M
TG-G_B_aeolian	na	0.25	0.19	0.19	0.17	0.40
LB-T_B_aeolian	0.25	0.37	0.25	0.23	0.24	0.30
VG-T_F_aeolian	0.19	0.25	0.12	0.21	0.15	0.27
CG-G_F_aeolian	0.19	0.23	0.21	0.20	0.20	0.35
EC-T_F_aeolian	0.17	0.24	0.15	0.20	0.17	0.25
MS-T_M_aeolian	0.40	0.30	0.27	0.35	0.25	0.18
TAG_B_cryoconite	0.50	0.39	0.40	0.46	0.37	0.17
CAG_F_cryoconite	0.33	0.33	0.24	0.34	0.26	0.25
CAN_F_black	0.18	0.31	0.20	0.29	0.26	0.40
CAN_F_orange	0.24	0.40	0.32	0.38	0.35	0.49
FRX_F_sediment	0.35	0.37	0.34	0.40	0.32	0.27
CMG_F_cryoconite	0.37	0.29	0.28	0.35	0.24	0.11
FRY_F_pond	0.28	0.29	0.16	0.29	0.22	0.29
CMN_F_orange	0.16	0.27	0.15	0.25	0.16	0.29
ADA_M_orange	0.27	0.33	0.25	0.32	0.29	0.41
MIE_M_orange	0.35	0.34	0.25	0.35	0.27	0.26
BLU_R_lake	0.29	0.34	0.30	0.36	0.28	0.29

Black text = 10 lowest pairwise dissimilarities for each aeolian site

Figure S1. Image of a Big Spring Number Eight isokinetic passive wind sampler from Explorer's Cove, Taylor Valley. Arrow indicates height from which aeolian samples were analyzed.

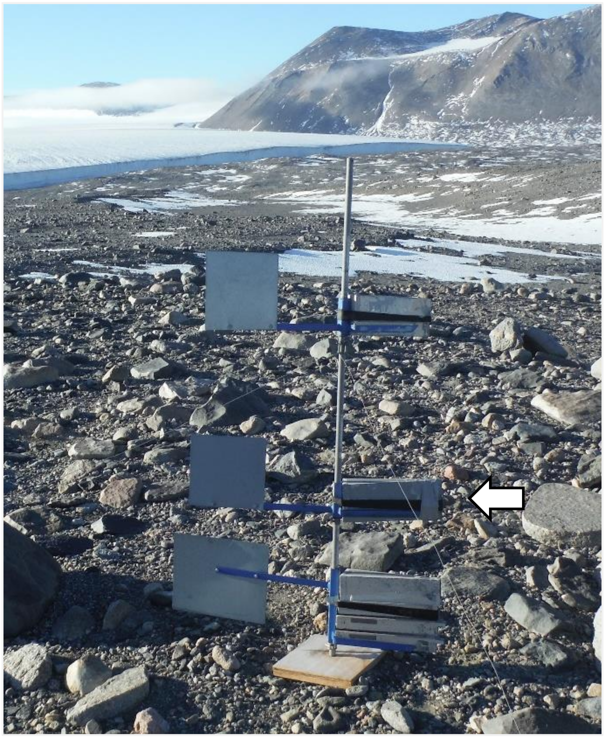


Figure S2. Box plots of wind-related variables across sampling sites and seasons. (A) Mean daily wind speed, (B) mean daily maximum wind speed, and (C) mean daily wind direction.

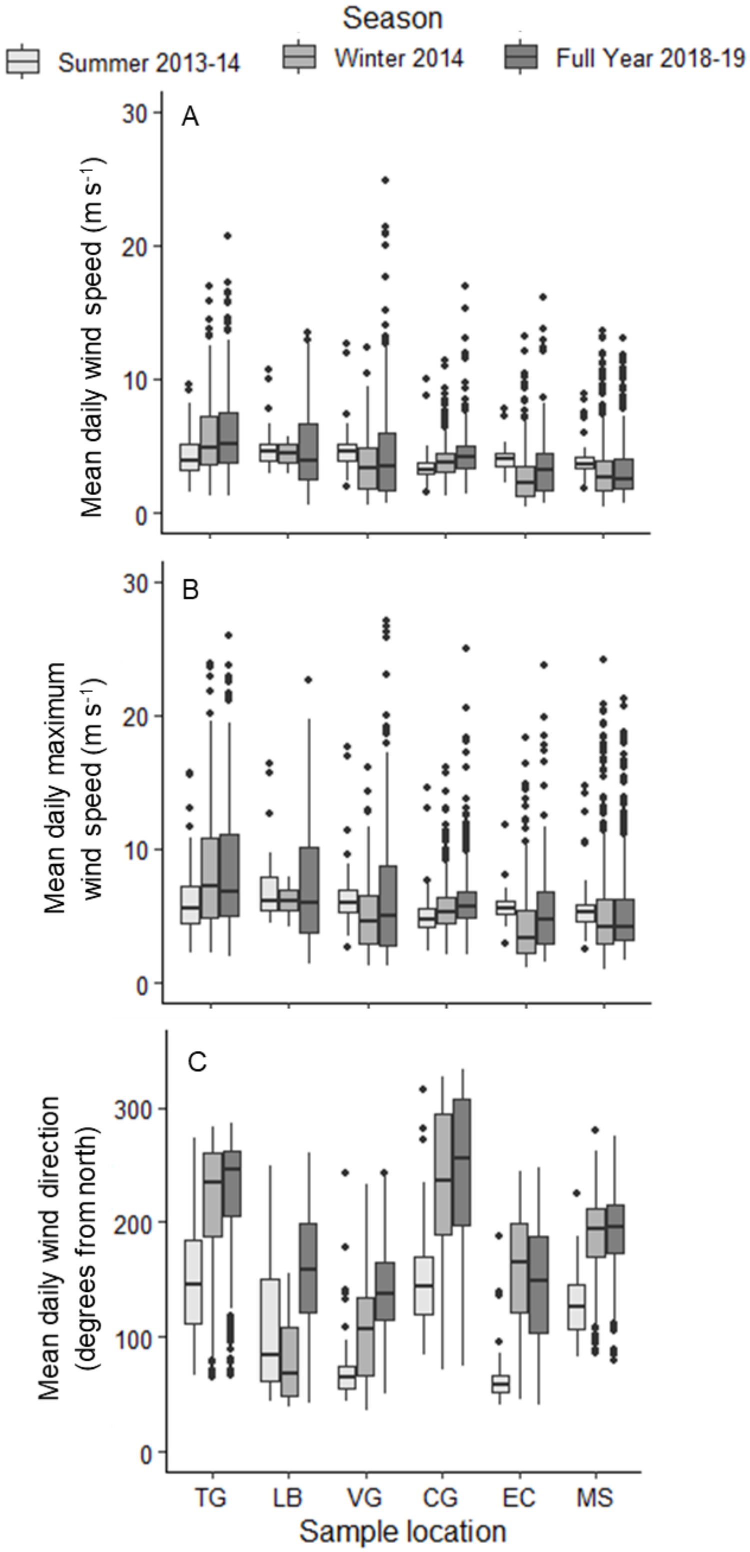
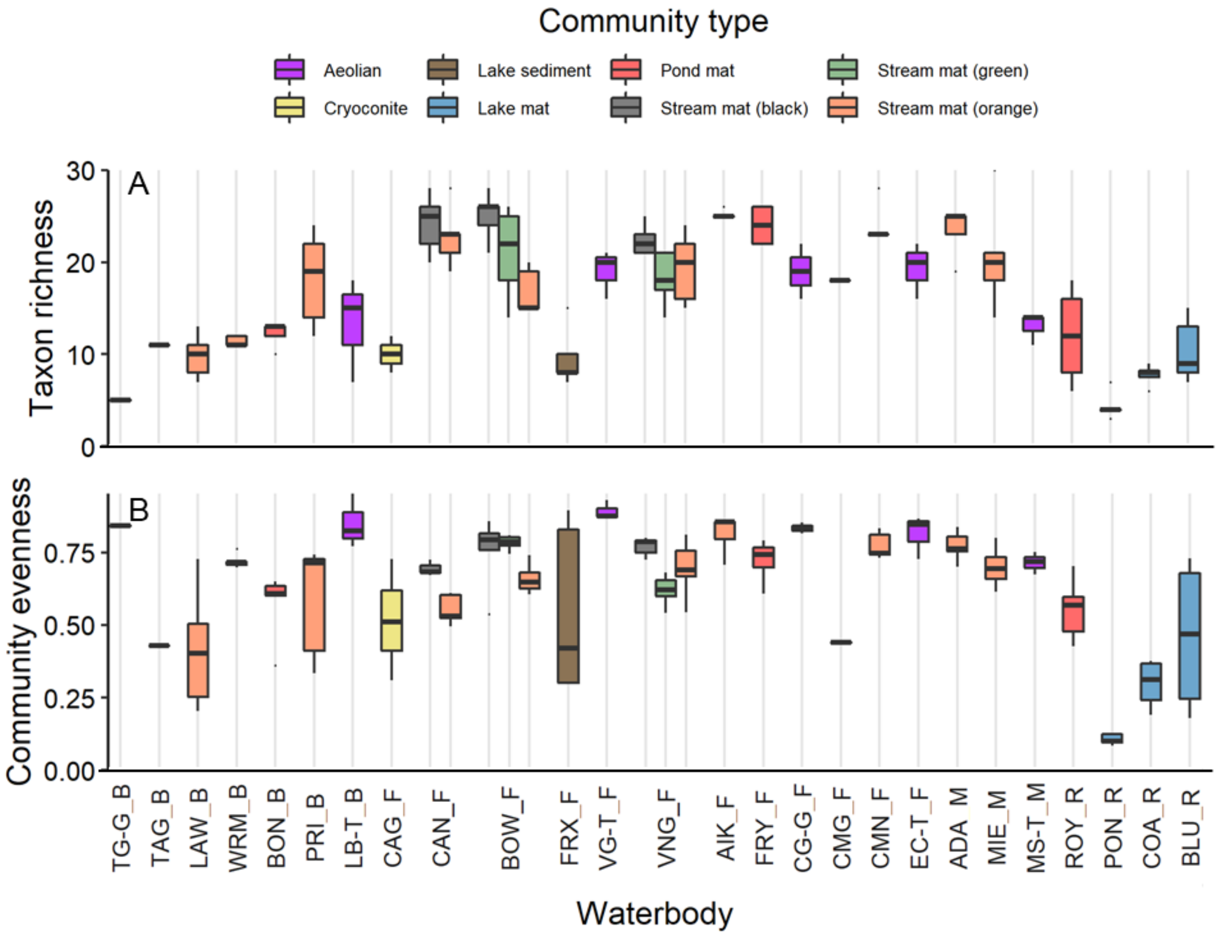


Figure S3. (A) Mean taxon richness and (B) mean community evenness for each waterbody used in the analysis.



Supplementary Results

In leave-one-out cross-validation analyses, aeolian and regional sources varied in specificity to their validation sources. We considered over 50% explained (a majority) by the same-community validation source as a strongly reliable source and having the highest proportion explained by the same-community validation source (a plurality), even if under 50%, as a moderately reliable source. Generally, each aeolian community sample was not explained by other samples from the same community any better than by other aeolian communities or some regional communities (Figure S4). The main exceptions to this pattern were CG-G_F and MS-T_M, which had moderately high specificity to same-community samples (i.e., the highest proportion was explained by the same community). However, the plurality of aeolian source composition explained was from all aeolian samples together. The validation results for aeolian samples indicated individual aeolian communities may not be reliably assigned as sources. However, given relatively high compositional similarity among aeolian communities, the cumulative signal of all aeolian communities together may still be prescribed to a general aeolian source rather than from regional material.

In cross-validation, regional communities also showed relatively low source specificity, with the exceptions of the strong source-specific signals of the Wormherder Creek Wetland, Pony Lake, and Coast Lake (Figure S5). However, a plurality of most source community composition was explained by validation sources from the same respective communities. In source communities with potential hydrologically connected sources (e.g., black and green mats for orange mats in the same stream), the validation sources generally explained the most composition, but hydrologically connected sources were also explanatory. Communities with potential hydrologically connected sources were used as possible sources in SourceTracker given intra-habitat connectivity was one process being tested. Overall, cross-validation analyses of aeolian and regional sources indicated few source communities were strongly reliable, likely because of (a) a large number of sources, (b) relatively low sample size, and (c) high compositional similarity among some source communities. Altogether, SourceTracker results should be interpreted with caution. We presented results from the SourceTracker analyses to provide coarse estimates of potential geographic scales of diatom connectivity across the MSR.

Figure S4. Leave-one-out cross-validation of potential aeolian source communities calculated using SourceTracker. A higher proportion (> 50% or the highest explained proportion of potential sources) of the source community explained by the same validation source community indicates the source community can be used more reliably in SourceTracker. For each source community panel, two bars are present that each represent different possible source communities. From left to right: (i) all possible aeolian and regional sources except those from different valleys and (ii) all same-valley aeolian and same-basin regional sources.

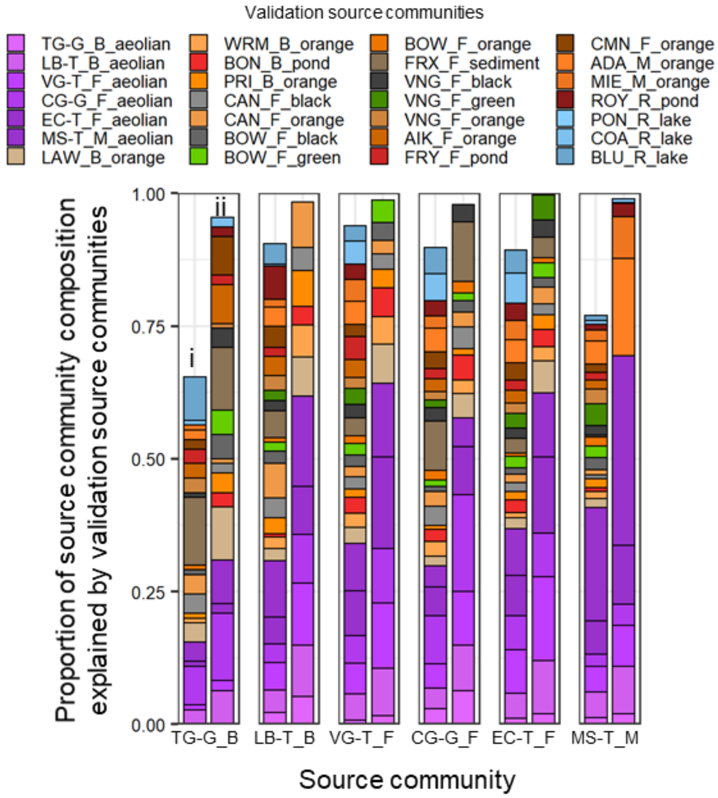
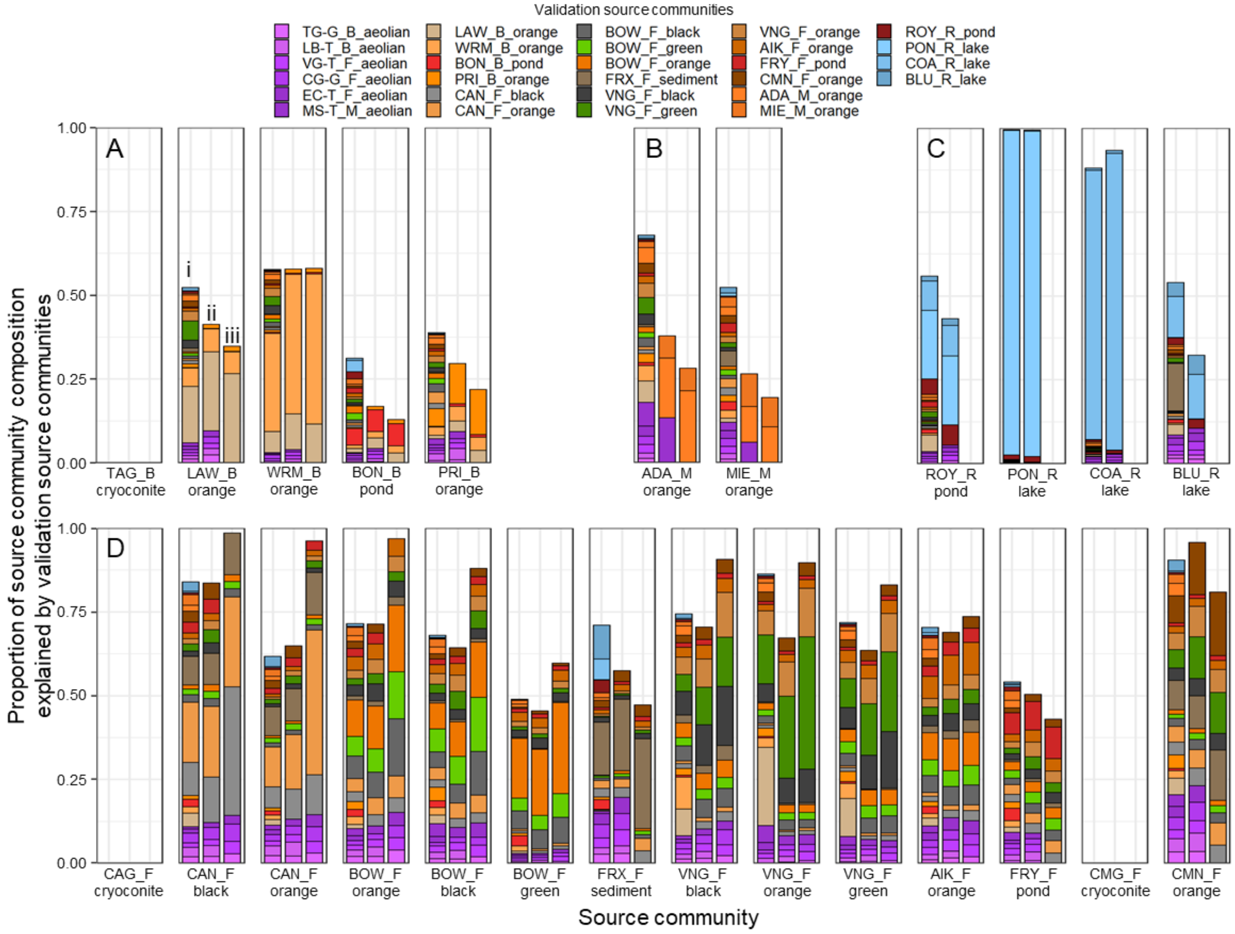


Figure S5. Leave-one-out cross-validation of potential regional source communities calculated using SourceTracker. A higher proportion (> 50% or the highest explained proportion of potential sources) of the source community explained by the same validation source community indicates the source community can be used reliably in SourceTracker. For each source community panel, three bars are present that each represent different possible source communities. From left to right: (i) all possible aeolian and regional sources except those from different valleys, (ii) all same-valley aeolian and same-basin regional sources, and (iii) all same-valley aeolian and hydrologically connected regional sources. Sink waterbodies are grouped by basin: (A) Lake Bonney basin, (B) Miers Valley, (C) Cape Royds, and (D) Lake Fryxell basin.



Chapter III Supplementary Material

Table S1. Frequency (percent of sites detected) and mean relative abundance (from sites where present) of diatom taxa with models that passed quality filtering.

Taxon	Frequency (%)	Mean relative abundance (%)	Taxon	Frequency (%)	Mean relative abundance (%)
Achnanthidium crassum / Achnanthidium rivulare / Achnanthidium deflexum	70.65	12.11	*Luticola goeppertiana*	3.26	0.33
Achnanthidium delmontii	14.13	5.47	*Mayamaea permitis*	21.74	0.40
Achnanthidium druartii	4.35	0.62	*Mayamaea* sp. 1 NESQA2016	3.26	2.44
Achnanthidium exiguum	17.39	0.55	*Meridion alansmithii*	3.26	0.22
Achnanthidium gracillimum	14.13	3.01	*Meridion circulare*	15.22	0.36
Achnanthidium kranzii	6.52	0.92	*Navicula amphiceropsis*	29.35	0.82
Achnanthidium minutissimum / Achnanthidium eutrophilum	96.74	21.02	*Navicula antverpiensis*	3.26	1.01
Achnanthidium pyrenaicum	28.26	17.52	*Navicula capitatoradiata*	25.00	1.12
Achnanthidium reimeri	8.70	0.39	*Navicula caterva*	15.22	1.01
Achnanthidium sp.	7.61	0.32	*Navicula cryptocephala*	21.74	0.43
Achnanthidium subhudsonis var. *kraeuselii*	27.17	7.62	*Navicula cryptofallax*	19.57	1.31
Adlafia bryophila	5.43	0.30	*Navicula cryptotenella*	57.61	1.82
Amphipleura pellucida	11.96	0.32	*Navicula escambia*	18.48	1.80
Amphora bicapitata	5.43	0.23	*Navicula freesei*	15.22	0.60
Amphora copulata	18.48	0.43	*Navicula genovefae*	7.61	0.48
Amphora ovalis	21.74	0.43	*Navicula germainii*	35.87	0.70
Amphora pediculus	58.70	13.28	*Navicula gregaria / Navicula canalis*	67.39	1.48
Aulacoseira ambigua	15.22	0.72	*Navicula lanceolata*	20.65	0.60
Aulacoseira nivaloides	3.26	0.27	*Navicula microcari*	9.78	0.74
Aulacoseira perglabra	2.17	0.41	*Navicula moskalii*	21.74	0.67
Aulacoseira sp.	3.26	0.17	*Navicula notha*	10.87	0.51
Aulacoseira subarctica	4.35	0.21	*Navicula oppugnata*	5.43	1.23
Aulacoseira tenella	7.61	0.33	*Navicula pseudoreinhardtii*	43.48	0.62
Bacillaria paxillifera	6.52	0.63	*Navicula radiosa*	2.17	0.25
Brachysira intermedia	3.26	2.17	*Navicula recens*	15.22	2.28
Brachysira microcephala	9.78	1.39	*Navicula rhynchocephala*	14.13	0.32
Caloneis sp. 3 SESQA	29.35	1.57	*Navicula rostellata*	3.26	0.72
Cocconeis fluviatilis	7.61	1.74	*Navicula* sp.	19.57	0.38
Cocconeis pediculus	40.22	0.81	*Navicula* sp. 4 NESQA2014	2.17	0.17
Cocconeis placentula	89.13	6.40	*Navicula* sp. 209 NESQA2016	3.26	0.94
Cocconeis pseudothumensis	4.35	0.46	*Navicula symmetrica*	19.57	0.91
Craticula subminuscula	6.52	0.72	*Navicula tripunctata*	45.65	1.93
Ctenophora pulchella	11.96	0.44	*Navicula trivialis*	6.52	0.25
Cyclotella distinguenda	4.35	0.78	*Nitzschia*	2.17	0.25
Cyclotella meneghiniana	54.35	1.06	*Nitzschia acicularis*	6.52	3.81

Taxon	Frequency (%)	Mean relative abundance (%)	Taxon	Frequency (%)	Mean relative abundance (%)
Cymbella affinis	11.96	4.12	*Nitzschia acidoclinata*	46.74	1.79
Cymbella helvetica	13.04	0.85	*Nitzschia amphibia*	40.22	1.94
Cymbella sp.	4.35	0.42	*Nitzschia brevissima*	2.17	0.42
Cymbella subturgidula	14.13	2.40	*Nitzschia clausii*	7.61	0.24
Cymbella tumida	31.52	0.48	*Nitzschia dissipata*	38.04	1.94
Cymbopleura naviculiformis	4.35	0.25	*Nitzschia dissipata* var. *media*	28.26	0.63
Delicata sp. 2 NESQA2014 / *Delicata* sp. 3 NESQA2014 / *Delicata* sp. 4 NESQA2014	3.26	0.61	*Nitzschia diversa*	2.17	0.17
Diadesmis confervacea	5.43	0.27	*Nitzschia fonticola*	32.61	1.50
Diatoma moniliformis	4.35	0.83	*Nitzschia fossilis*	18.48	2.76
Diatoma polonica	4.35	0.25	*Nitzschia frequens*	3.26	0.22
Diatoma tenuis	7.61	0.82	*Nitzschia gracilis*	6.52	0.61
Diatoma vulgaris	28.26	0.47	*Nitzschia heufleriana*	8.70	0.40
Diploneis krammeri	5.43	0.47	*Nitzschia microcephala*	18.48	0.54
Discostella stelligera	4.35	0.21	*Nitzschia oligotraphenta*	25.00	1.33
Encyonema appalachianum	4.35	0.54	*Nitzschia oregona*	6.52	0.31
Encyonema auerswaldii	3.26	0.44	*Nitzschia palea*	16.30	0.47
Encyonema lange-bertalotii	65.22	1.17	*Nitzschia palea* var. *debilis* / *Nitzschia palea* var. *tenuirostris*	66.30	1.41
Encyonema minutum	7.61	1.02	*Nitzschia recta*	9.78	0.26
Encyonema reichardtii	7.61	0.50	*Nitzschia sinuata* var. *tabellaria*	2.17	0.25
Encyonema silesiacum	14.13	0.92	*Nitzschia sociabilis*	14.13	0.58
Encyonema sp.	6.52	0.42	*Nitzschia soratensis / Nitzschia inconspicua / Nitzschia perminuta*	58.70	5.94
Encyonema vulgare	2.17	0.17	*Nitzschia* sp.	15.22	0.45
Encyonopsis minuta	5.43	9.10	*Nitzschia* sp. 10 NESQA2014	34.78	0.61
Encyonopsis subminuta	3.26	2.61	*Nitzschia* sp. 18 NESQA2014	15.22	0.49
Epithemia sorex	2.17	3.50	*Nitzschia subacicularis*	4.35	0.29
Eucocconeis laevis	7.61	0.31	*Nupela impexiformis*	5.43	0.30
Eunotia bilunaris	5.43	0.30	*Nupela lapidosa*	13.04	1.39
Eunotia dianae stitinensis	2.17	0.25	*Nupela vitiosa*	4.35	0.37
Eunotia exigua	5.43	0.37	*Nupela wellneri*	3.26	0.33
Eunotia formica	6.52	0.89	*Pinnularia erratica* var. *fossilis*	2.17	0.17
Eunotia incisa	7.61	0.33	*Pinnularia subgibba* var. *sublinearis*	3.26	0.33
Eunotia metamonodon	7.61	1.20	*Placoneis* sp. 3 NESQA2016	2.17	0.25
Eunotia minor / Eunotia implicata	25.00	2.12	*Placoneis symmetrica*	2.17	0.17
Eunotia mucophila	6.52	0.52	*Planothidium daui*	6.52	0.69
Eunotia naegelii	6.52	0.33	*Planothidium delicatulum*	6.52	0.22
Eunotia pectinalis	5.43	0.51	*Planothidium frequentissimum*	51.09	0.87
Eunotia pirla	3.26	0.28	*Planothidium hauckianum*	6.52	3.03
Eunotia rhomboidea	4.35	0.50	*Planothidium lanceolatum*	26.09	0.47
Eunotia sp.	13.04	0.50	*Planothidium robustum*	6.52	0.22
Eunotia sp. 15 NE Lakes	4.35	0.33	*Planothidium rostratum*	2.17	0.17

Taxon	Frequency (%)	Mean relative abundance (%)	Taxon	Frequency (%)	Mean relative abundance (%)
Eunotia subarcuatoides	5.43	0.63	*Planothidium* sp.	2.17	0.17
Eunotia tenella	2.17	0.17	*Platessa bahlsii*	25.00	2.03
Fallacia sp. 2 NESQA2016	3.26	0.72	*Platessa conspicua*	8.70	0.62
Fallacia subhamulata	8.70	0.54	*Platessa* sp. 8 NESQA2016	2.17	1.08
Fragilaria gracilis	42.39	1.15	*Platessa stewartii*	3.26	0.22
Fragilaria mesolepta	10.87	0.20	*Psammothidium alpinum*	2.17	0.33
Fragilaria sp.	21.74	0.70	*Psammothidium chlidanos*	4.35	0.25
Fragilaria sp. 6 NESQA2014	5.43	0.53	*Psammothidium grischunum*	4.35	1.00
Fragilaria sp. 7 NESQA2014	23.91	0.33	*Psammothidium helveticum*	8.70	0.44
Fragilaria sp. 8 NESQA2014	9.78	0.48	*Psammothidium lauenburgianum*	4.35	0.42
Fragilaria sp. 10 NESQA2014 / *Fragilaria* sp. 11 NESQA2014	32.61	0.84	*Psammothidium marginulatum*	10.87	0.37
Fragilaria sp. 15 NESQA2014	7.61	1.07	*Psammothidium rossii*	4.35	0.62
Fragilaria sp. 17 NESQA2014	16.30	0.66	*Psammothidium subatomoides*	18.48	1.02
Fragilaria sp. 18 SESQA	10.87	0.44	*Pseudofallacia monoculata*	6.52	0.25
Fragilaria sp. 35 SESQA	3.26	1.28	*Pseudostaurosira elliptica*	2.17	0.17
Fragilaria sp. 201 NESQA2016	2.17	0.50	*Reimeria sinuata*	57.61	1.21
Fragilaria subconstricta	2.17	0.38	*Reimeria uniseriata*	19.57	0.53
Fragilaria tenera	16.30	0.56	*Rhoicosphenia abbreviata*	63.04	5.92
Fragilaria vaucheriae	33.70	0.97	*Rossithidium anastasiae*	8.70	0.27
Fragilariforma constricta	2.17	0.25	*Sellaphora atomoides*	14.13	0.68
Fragilariforma marylandica	2.17	0.50	*Sellaphora difficillima*	6.52	0.58
Frustulia amphipleuroides	7.61	0.31	*Sellaphora elorantana*	3.26	0.61
Frustulia crassinervia	4.35	0.33	*Sellaphora hustedtii*	4.35	0.21
Frustulia krammeri	4.35	0.17	*Sellaphora nigri / Sellaphora saugerresii*	65.22	4.35
Frustulia neomundana	3.26	0.39	*Sellaphora pseudopupula*	3.26	0.17
Frustulia saxonica	6.52	0.47	*Sellaphora rexii / Sellaphora meridionalis*	14.13	0.22
Geissleria acceptata	9.78	0.37	*Sellaphora* sp. 16 NESQA2016	10.87	3.02
Gomphoneis minuta	3.26	0.61	*Sellaphora* sp. 19 SESQA	31.52	0.71
Gomphoneis olivaceum	10.87	0.35	*Sellaphora* sp. 20 NE Lakes	15.22	0.71
Gomphonema apuncto	3.26	6.78	*Sellaphora* sp. 24 NE Lakes	3.26	0.22
Gomphonema caperatum	19.57	5.74	*Sellaphora* sp. 30 NE Lakes	3.26	0.28
Gomphonema celatum	5.43	0.27	*Sellaphora* sp. 41 NE Lakes	4.35	0.38
Gomphonema gracile	7.61	1.86	*Simonsenia delognei*	3.26	0.72
Gomphonema louisianum	6.52	1.75	*Stauroforma exiguiformis*	6.52	0.75
Gomphonema minutum / Gomphonema kobayasii	64.13	3.92	*Stauroneis kriegeri*	3.26	0.28
Gomphonema olivaceoides var. *hutchinsoniana*	14.13	0.33	*Stauroneis thermicola*	2.17	0.17
Gomphonema parvulum	82.61	1.47	*Staurosira construens*	7.61	0.50
Gomphonema sp.	39.13	0.97	*Staurosira construens* var. *pumila*	18.48	1.27
Gomphonema sp. 9 SESQA	3.26	0.39	*Staurosira construens* var. *venter*	19.57	1.33
Gomphonema sp. 10 NESQA2014	6.52	0.31	*Staurosira* sp.	9.78	1.20

Taxon	Frequency (%)	Mean relative abundance (%)	Taxon	Frequency (%)	Mean relative abundance (%)
Gomphonema sp. 16 NESQA2014	7.61	0.40	*Staurosira* sp. 1 NESQA2014	4.35	2.83
Gomphonema sp. 20 SESQA	19.57	0.81	*Staurosira* sp. 12 NESQA2014	14.13	0.78
Gomphonema sp. 26 SESQA / sp. 40 SESQA	5.43	0.67	*Staurosira* sp. 16 NESQA2016	3.26	1.22
Gomphonema sp. 47 NESQA2016	6.52	2.83	*Staurosirella oldenburgiana*	6.52	0.72
Gomphonema sp. 49 NESQA2016	2.17	0.42	*Staurosirella pinnata*	19.57	0.74
Gomphonema turgidum	5.43	0.43	*Staurosirella sp.*	10.87	0.78
Gomphonema variostriatum	2.17	1.50	*Surirella amphioxys*	7.61	0.26
Gyrosigma attenuatum	5.43	0.30	*Surirella brebissonii*	8.70	0.78
Gyrosigma reimeri	9.78	0.37	*Synedra pulchella* var. *flexella*	13.04	0.75
Halamphora montana	4.35	0.54	*Tabellaria flocculosa*	8.70	0.85
Hippodonta capitata	21.74	0.33	*Thalassiosira weissflogii*	19.57	0.58
Hippodonta capitata subsp. *iberoamericana*	7.61	0.48	*Tryblionella apiculata*	6.52	0.31
Hippodonta pseudacceptata	13.04	1.19	*Tryblionella debilis*	4.35	0.31
Humidophila contenta	5.43	0.20	*Ulnaria acus*	11.96	0.24
Humidophila schmassmanni	3.26	0.28	*Ulnaria contracta*	21.74	0.67
Karayevia clevei	7.61	0.67	*Ulnaria ramesi*	4.35	0.21
Karayevia oblongella	13.04	4.00	*Ulnaria rostrata*	2.17	0.25
Karayevia ploenensis var. *gessneri*	19.57	1.21	*Ulnaria* sp.	15.22	0.44
Karayevia suchlandtii	3.26	0.55	*Ulnaria* sp. 10 NESQA2016	22.83	0.47

Table S2. Frequency (percent of sites detected) and mean relative abundance (from sites where present) of diatom taxa with no models that passed quality filtering. Gray text for taxon names distinguishes the unmodeled taxa in this table from the modeled taxa in Table S1.

Taxon	Frequency (%)	Mean relative abundance (%)	Taxon	Frequency (%)	Mean relative abundance (%)
Achnanthidium atomus	3.26	0.67	*Microcostatus maceria*	1.09	3.50
Achnanthidium latecephalum	5.43	0.40	*Navicula angusta*	2.17	0.17
Adlafia minuscula	1.09	0.17	*Navicula antonii / Navicula chiarae*	72.83	2.70
Adlafia sp. 5 SESQA	2.17	0.33	*Navicula erifuga*	2.17	0.33
Adlafia sp. 1 NE Lakes	1.09	0.33	*Navicula ingenua*	1.09	0.33
Amphora eximia	1.09	0.17	*Navicula slevicensis*	2.17	0.25
Amphora sp.	2.17	0.25	*Navicula splendicula*	1.09	0.33
Asterionella formosa	3.26	0.59	*Navicula supergregaria*	1.09	0.33
Aulacoseira granulata var. *angustissima*	1.09	0.17	*Navicula vilaplanii*	6.52	0.31
Aulacoseira lacustris	1.09	0.33	*Navicula viridula*	5.43	0.42
Aulacoseira nygaardii	2.17	0.42	*Navicula viridulacalcis* subsp. *neomundana*	2.17	0.17
Aulacoseira sp. 4 NE Lakes	1.09	0.17	*Neidium bisulcatum*	1.09	0.17
Aulacoseira sp. 10 NE Lakes	1.09	0.17	*Neidium* sp.	1.09	0.17
Aulacoseira sp. 200 NESQA2016	1.09	2.50	*Nitzschia acicularioides*	2.17	0.33
Brachysira brebissonii	1.09	1.17	*Nitzschia capitellata*	1.09	0.17
Brachysira sp. 300 NESQA2016	1.09	0.67	*Nitzschia draveillensis*	8.70	0.71
Caloneis amphisbaena	1.09	0.17	*Nitzschia fruticosa*	8.70	0.44
Caloneis bacillum	2.17	0.83	*Nitzschia homburgiensis*	1.09	0.33
Caloneis sp.	7.61	0.81	*Nitzschia montanestris*	8.70	0.38
Caloneis sp. 1 NE Lakes	1.09	0.17	*Nitzschia palustris*	1.09	0.33
Caloneis sp. 200 NESQA2016	1.09	0.33	*Nitzschia pusilla*	3.26	0.28
Caloneis vasileyevae	1.09	0.33	*Nitzschia reversa*	1.09	0.17
Capartogramma crucicula	1.09	0.33	*Nitzschia semirobusta*	2.17	0.83
Cavinula cocconeiformis	1.09	0.50	*Nitzschia tubicola*	1.09	0.50
Cavinula vincentii	2.17	0.33	*Nupela decipiens*	1.09	0.50
Craticula buderi	2.17	0.42	*Nupela neglecta*	1.09	0.33
Craticula molestiformis	2.17	0.17	*Pinnularia acrosphaeria* var. *parva*	1.09	0.50
Ctenophora sp. 201 NESQA2016	1.09	0.17	*Pinnularia divergens* var. *media*	1.09	0.33
Cymatopleura solea	1.09	0.17	*Pinnularia obscura* morphotype I	1.09	0.33
Cymbella hustedtii	1.09	0.33	*Pinnularia perirrorata*	1.09	0.17
Cymbella lanceolata	1.09	0.17	*Pinnularia saprophila*	1.09	0.33
Cymbella neocistula	1.09	0.17	*Pinnularia schoenfelderi*	2.17	0.17
Cymbella subleptoceros	2.17	0.25	*Pinnularia* sp.	3.26	0.33
Cymbella tropica	1.09	0.17	*Placoneis pseudanglica*	1.09	0.33
Decussata placenta	1.09	0.17	*Placoneis* sp.	1.09	0.33
Denticula sp. 1 NE Lakes	1.09	0.17	*Plagiotropis lepidoptera* var. *proboscidea*	1.09	0.17
Diatoma ehrenbergii	2.17	0.33	*Planothidium apiculatum*	1.09	0.17

Taxon	Frequency (%)	Mean relative abundance (%)
Didymosphenia geminata	1.09	0.33
Diploneis marginestriata	3.26	0.78
Discostella pseudostelligera	5.43	0.59
Encyonema leibleinii	3.26	0.28
Encyonema pankowii	1.09	0.17
Encyonema silesiacum var. *lata*	4.35	0.41
Encyonopsis sp.	1.09	0.17
Encyonopsis sp. 3 NE Lakes	1.09	0.33
Entomoneis paludosa	2.17	0.25
Eolimna aboensis	2.17	0.33
Eunotia curtagrunowi	1.09	0.17
Eunotia macroglossa	1.09	0.17
Eunotia meisteri	2.17	0.42
Eunotia neocompacta var. *vixcompacta*	1.09	0.33
Eunotia perpusilla	1.09	0.50
Eunotia sp. 31 NESQA2014	1.09	0.33
Eunotia sp. 204 NESQA2016	1.09	0.33
Eunotia sudetica	1.09	0.17
Eunotia ursamaioris	1.09	0.17
Eunotia varioundulata	2.17	0.33
Fallacia insociabilis	1.09	0.67
Fallacia pygmaea ssp. *subpygmaea*	2.17	0.33
Fallacia tenera	2.17	0.50
Fragilaria crotonensis	1.09	5.17
Fragilaria grunowii	2.17	0.67
Fragilaria perminuta	6.52	0.22
Fragilaria sp. 11 NESQA2014	3.26	0.22
Fragilaria sp. 12 NESQA2014	1.09	0.17
Fragilaria sp. 13 NESQA2014	1.09	0.17
Fragilaria sp. 22 NESQA2014	1.09	0.33
Fragilaria sp. 203 NESQA2016	4.35	0.50
Fragilariforma bicapitata	1.09	0.17
Fragilariforma sp. 4 NE Lakes	3.26	0.44
Frustulia sp.	1.09	0.43
Frustulia vulgaris	6.52	0.17
Geissleria decussis	13.04	0.47
Gomphoneis sp. 200 NESQA2016	1.09	0.33
Gomphonema affine var. *rhombicum*	1.09	0.67
Gomphonema brebissonii	1.09	0.33
Gomphonema coronatum	5.43	0.33
Gomphonema intricatum	1.09	0.33
Gomphonema manubrium	1.09	0.33
Planothidium incuriatum	3.26	0.44
Planothidium oestrupii	1.09	0.17
Planothidium pumilum	1.09	0.83
Platessa hustedtii	7.61	1.05
Platessa kingstonii	1.09	3.00
Platessa sp.	2.17	0.25
Platessa sp. 6 NESQA2016	1.09	0.17
Pleurosira laevis	2.17	0.42
Psammothidium bioretii	2.17	0.25
Psammothidium harveyi	1.09	0.17
Psammothidium pennsyvanicum	1.09	0.17
Psammothidium scoticum	1.09	0.33
Psammothidium sp.	1.09	0.67
Psammothidium sp. 400 NESQA2016	1.09	0.50
Psammothidium sp. 402 NESQA2016	1.09	0.33
Pseudostaurosira parasitica	5.43	0.27
Pseudostaurosira sp.	3.26	0.78
Pseudostaurosira sp. 3 NE Lakes	1.09	0.17
Pseudostaurosira sp. 201 NESQA2014	1.09	1.00
Rhopalodia gibba	1.09	0.17
Rhopalodia gibberula	1.09	0.17
Sellaphora auldreekie	4.35	1.79
Sellaphora bacillum	1.09	0.33
Sellaphora capitata	8.70	0.21
Sellaphora disjuncta	1.09	0.17
Sellaphora pseudoventralis	1.09	0.33
Sellaphora raederae	1.09	0.85
Sellaphora saprotolerans	4.35	0.46
Sellaphora sp.	3.26	0.33
Sellaphora sp. 15 NESQA2016	1.09	0.33
Sellaphora sp. 23 NESQA2016	1.09	0.17
Sellaphora sp. 37 NE Lakes	1.09	0.17
Sellaphora sp. 301 NESQA2016	1.09	0.17
Sellaphora stroemii	3.26	1.44
Sellaphora subbacillum	3.26	1.00
Sellaphora vitabunda	4.35	0.33
Sellaphora wallacei	5.43	0.30
Stauroneis sp. 3 NE Lakes	1.09	0.17
Staurosira sp. 6 NE Lakes	1.09	0.17
Staurosira sp. 401 NESQA2016	1.09	0.17
Staurosira subsalina	1.09	0.17
Staurosirella anasta	1.09	0.17

Taxon	Frequency (%)	Mean relative abundance (%)	Taxon	Frequency (%)	Mean relative abundance (%)
Gomphonema sp. 7 SESQA	2.17	0.33	*Staurosirella leptostauron*	2.17	0.33
Gomphonema sp. 18 NESQA2014	2.17	0.33	*Stephanodiscus binderanus*	1.09	0.50
Gomphonema sp. 201 NESQA2016	1.09	0.33	*Surirella angusta*	1.09	0.17
Gomphonema sp. / *haerophorum*	2.17	0.17	*Surirella* sp.	5.43	0.23
Gomphonema truncatum	3.26	0.17	*Surirella stalagma*	1.09	0.17
Gyrosigma scalproides	3.26	1.44	*Tabellaria fenestrata*	2.17	0.25
Halamphora veneta	2.17	0.25	*Tabularia fasciculata*	6.52	0.44
Hippodonta sp.	1.09	0.50	*Thalassiosira lacustris*	1.09	0.21
Hippodonta sp. 400 NESQA2016	1.09	0.17	*Tryblionella calida*	1.09	0.50
Humidophila perpusilla	1.09	0.17	*Ulnaria delicatissima*	5.43	0.23
Humidophila sp. 1 NE Lakes	1.09	0.17	*Ulnaria* sp. 2 NESQA2014	2.17	0.17
Karayevia laterostrata	4.35	0.29	*Ulnaria* sp. 3 NESQA2016	2.17	0.19
Lindavia intermedia	1.09	0.33	*Ulnaria* sp. 5 NESQA2014	1.09	2.83
Lindavia ocellata	3.26	0.39	*Ulnaria* sp. 11 NESQA2016	1.09	0.17
Luticola acidoclinata	5.43	0.20	*Ulnaria ulna*	7.61	0.55
Melosira varians	68.48	1.61	Unknown pennate	8.70	0.54

Figure S1. Two-dimensional ordination of the scaled principal components analysis of inter-site similarity as defined by environmental factors.

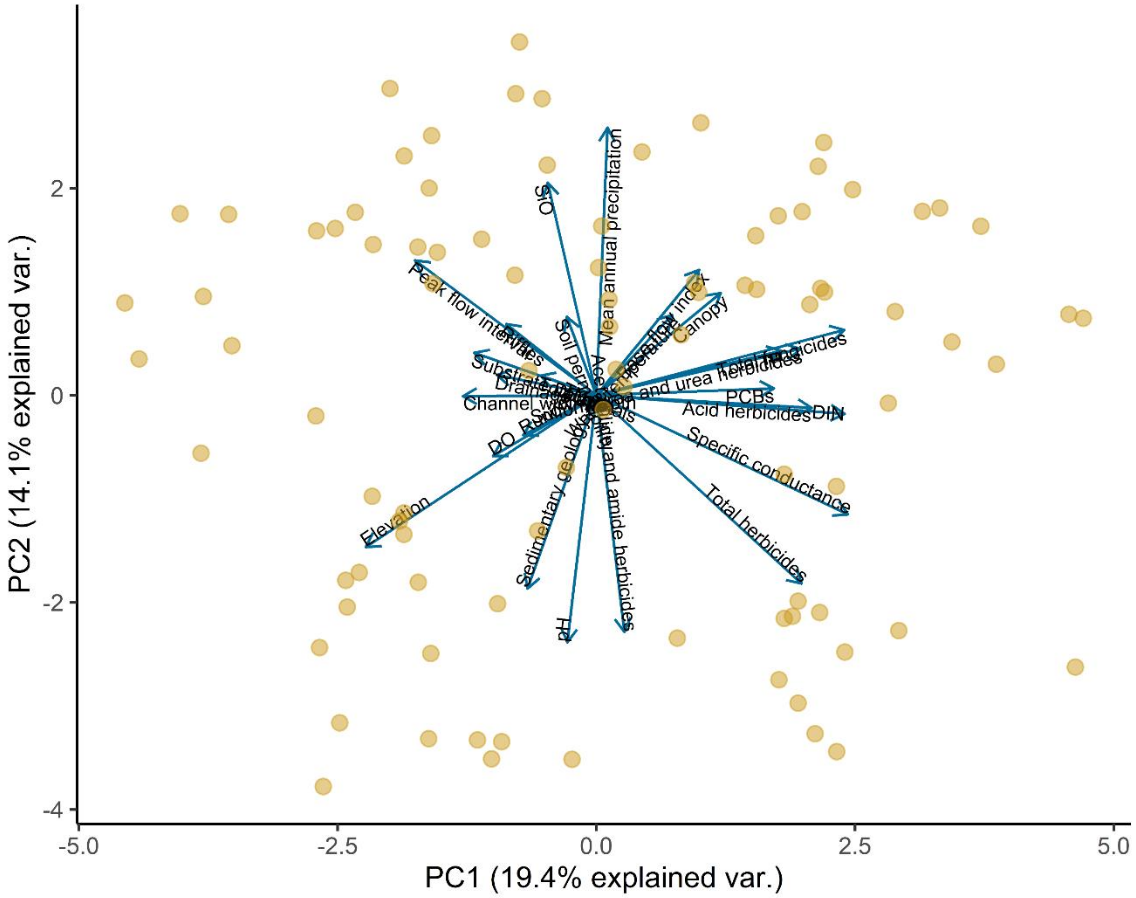


Figure S2. Variable deviance explained (dev.exp), Akaike weights, and model deviance explained of filtered generalized additive models for taxon relative abundance. Akaike weights were calculated relative to models of the same taxon. Models are organized alphabetically by taxon within each cluster. For each model, numbers following the taxon name (e.g., 1of2) represent the model identity relative to all the models for a taxon, in descending order from best performing model by Akaike weight. Clusters were determined from *k*-means cluster analysis across deviance explained of model variables. (1 of 12)

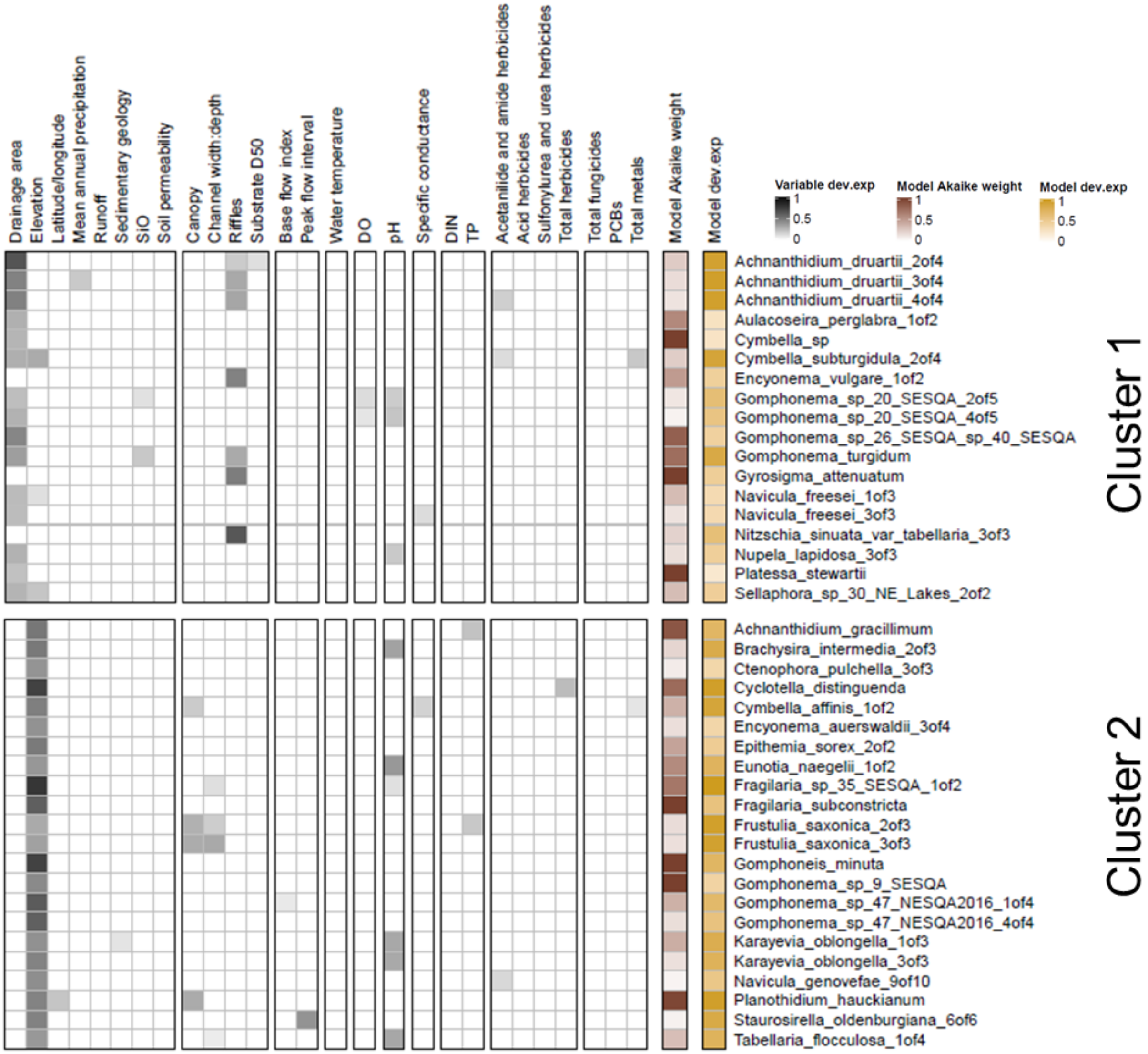


Figure S2 (continued). (2 of 12)

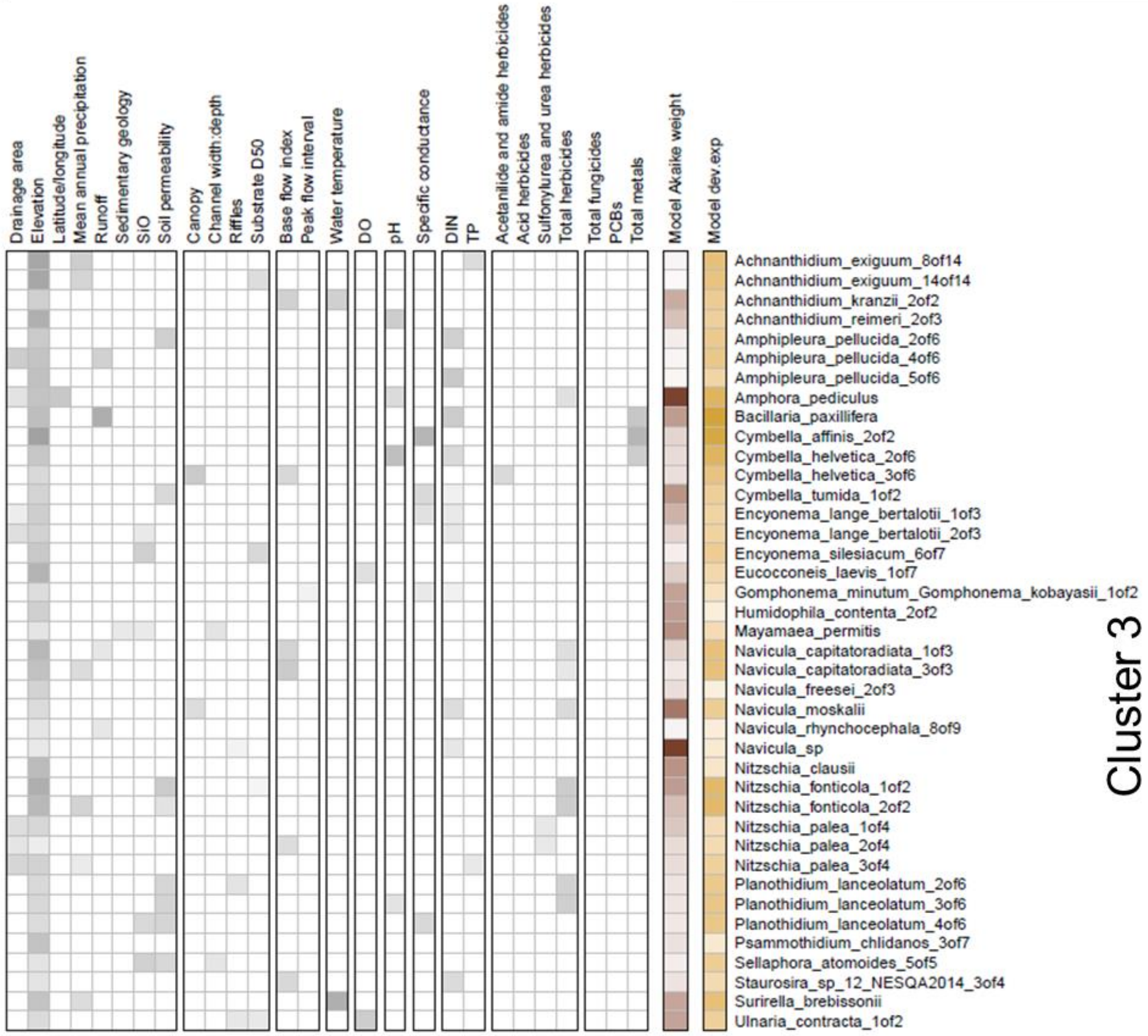


Figure S2 (continued). (3 of 12)

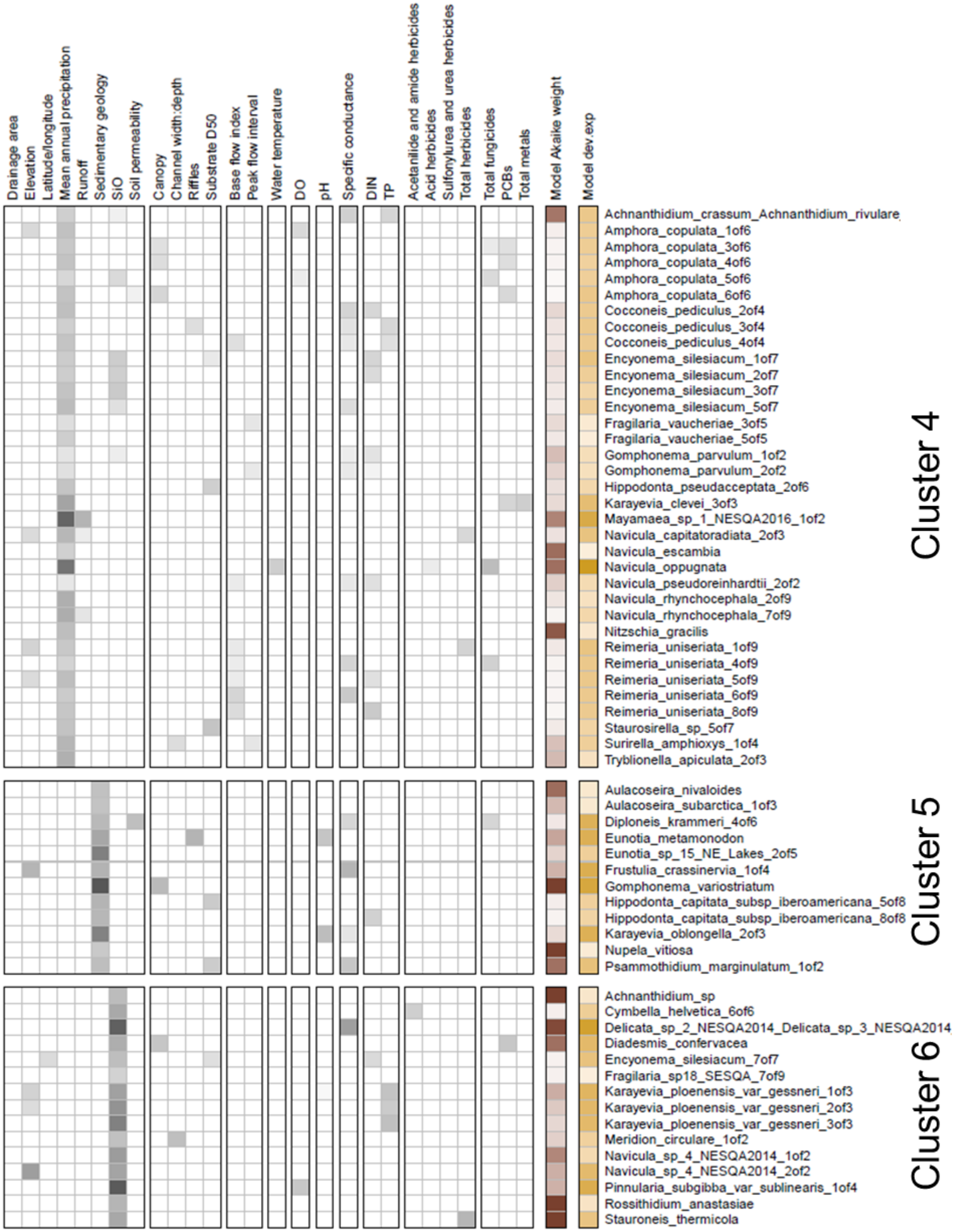


Figure S2 (continued). (4 of 12)

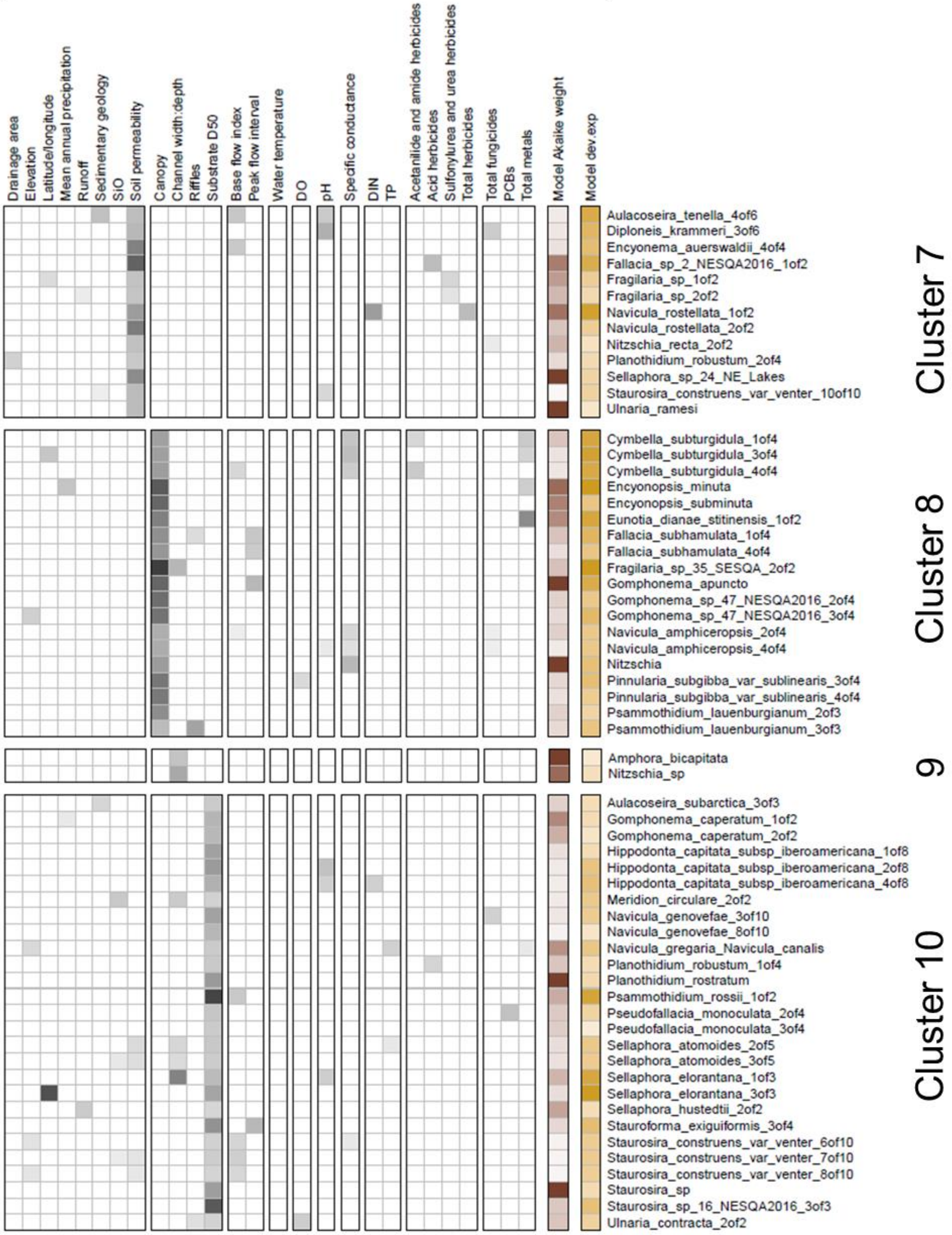


Figure S2 (continued). (5 of 12)

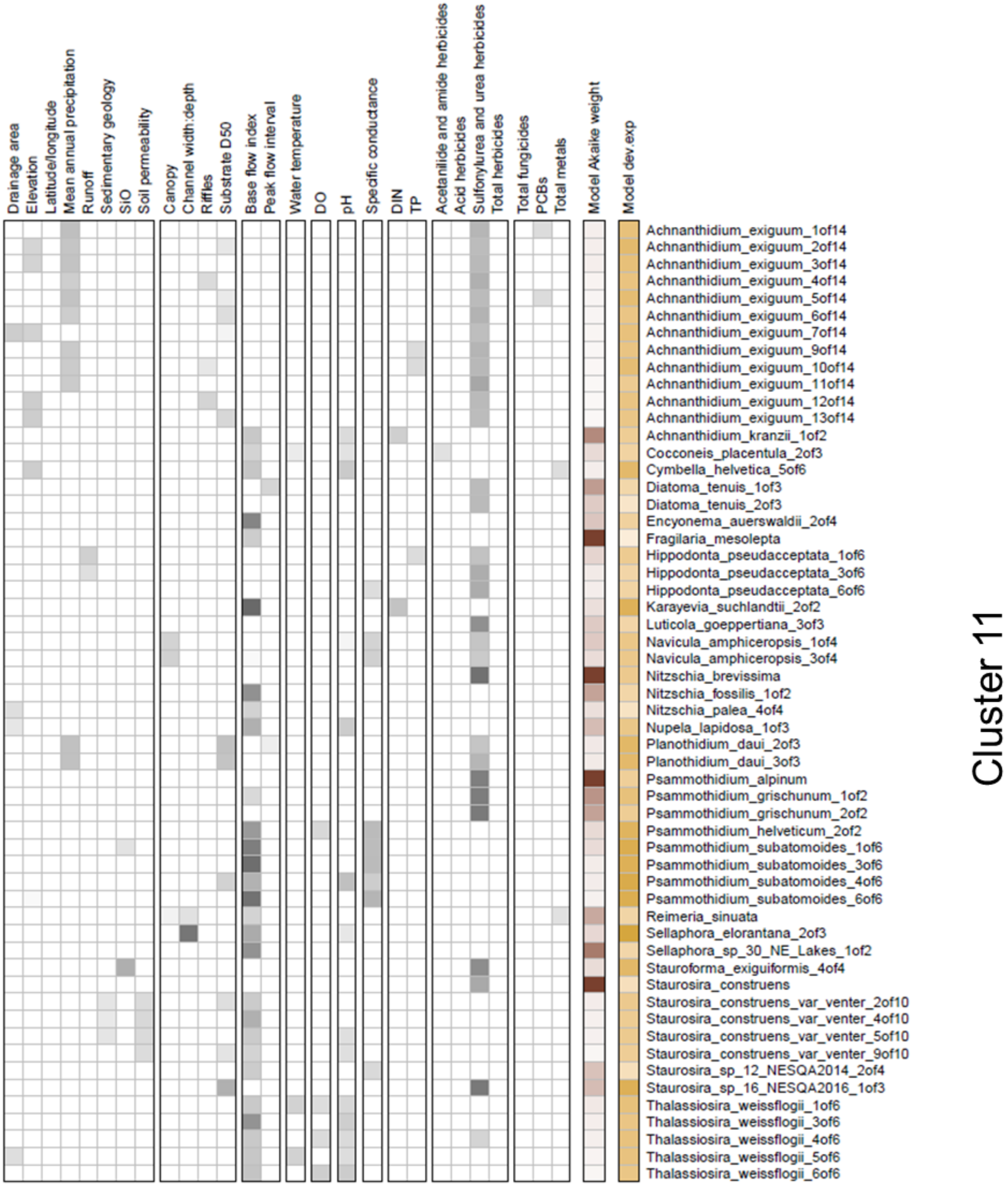


Figure S2 (continued). (6 of 12)

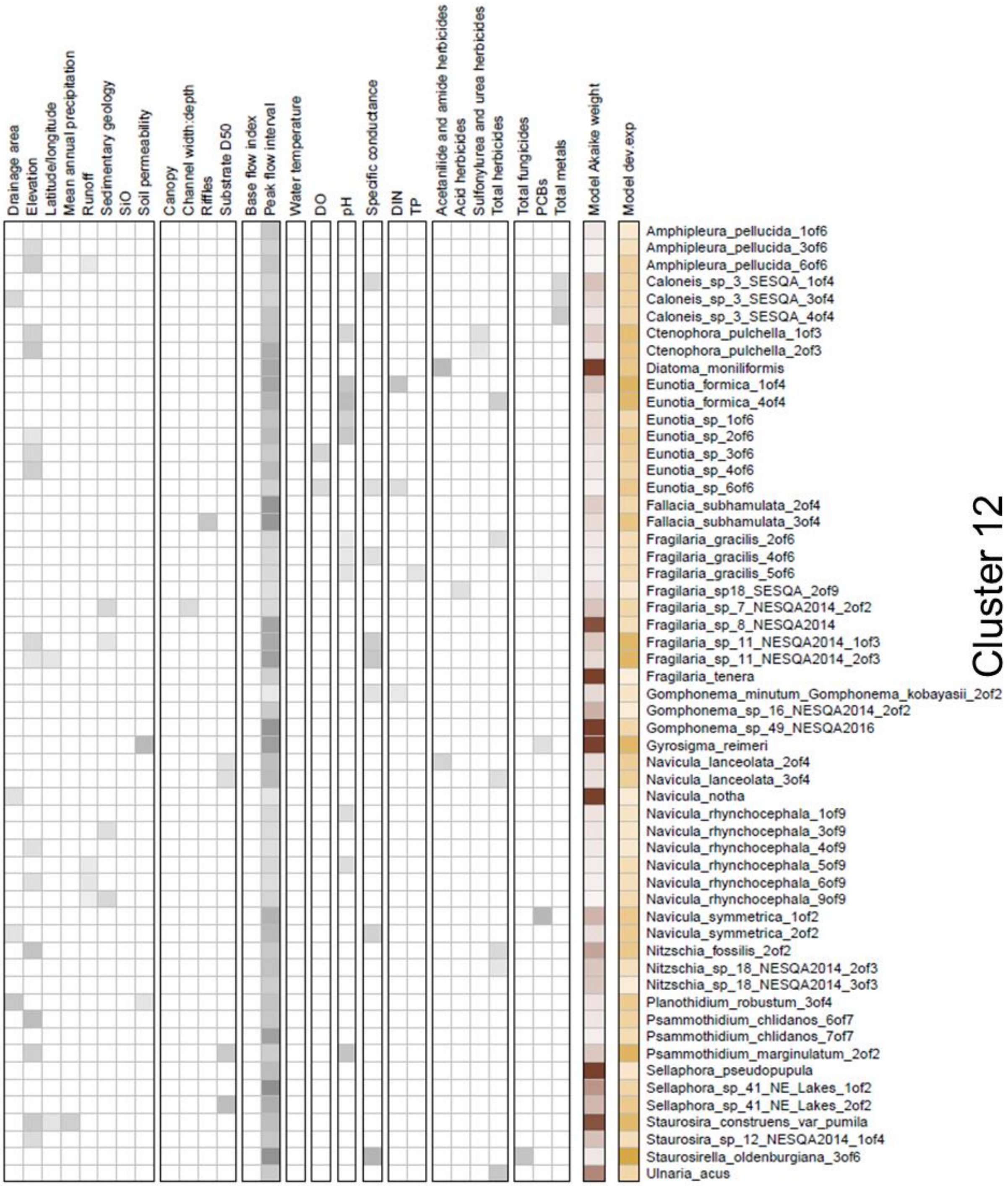


Figure S2 (continued). (7 of 12)

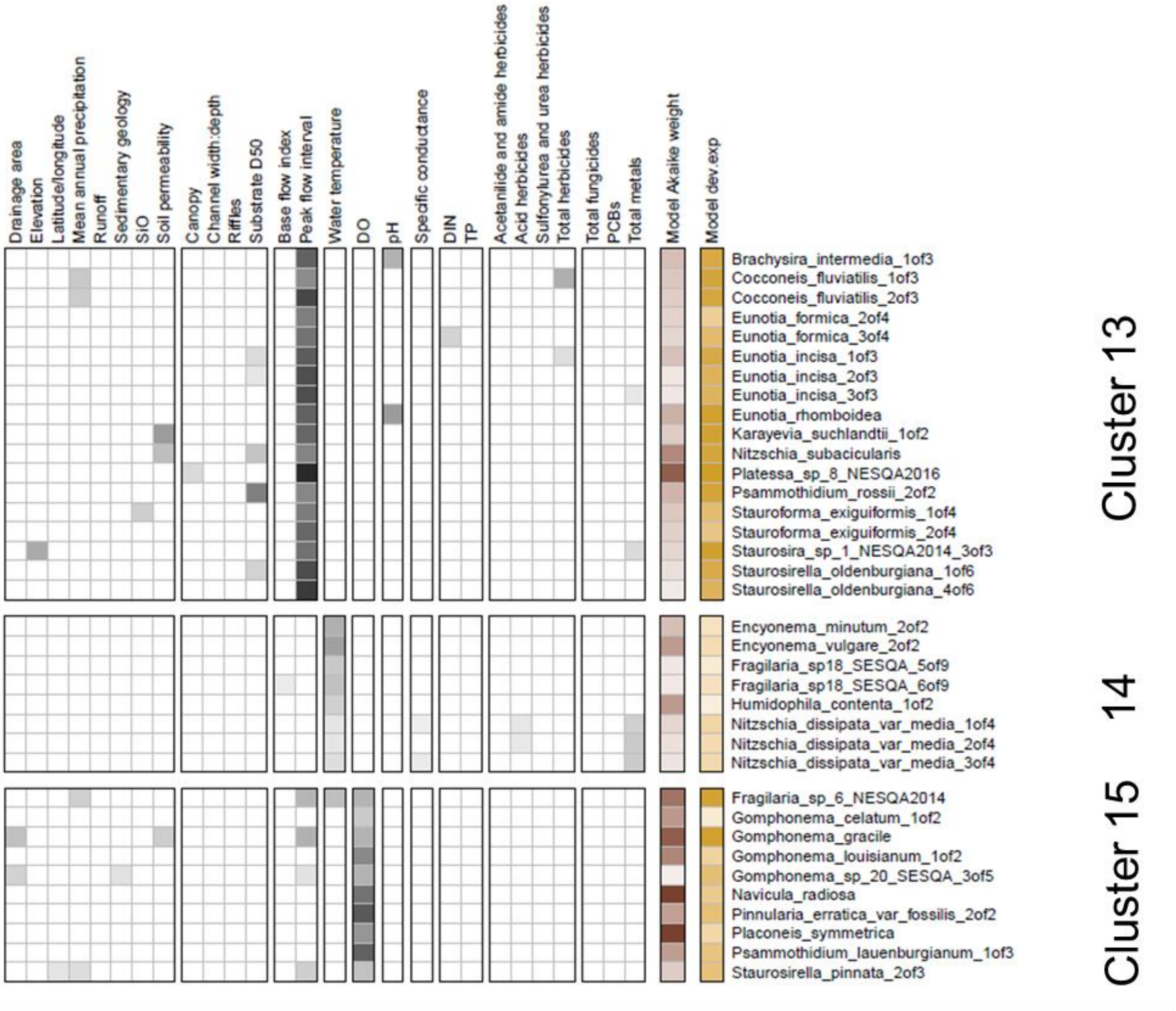


Figure S2 (continued). (8 of 12)

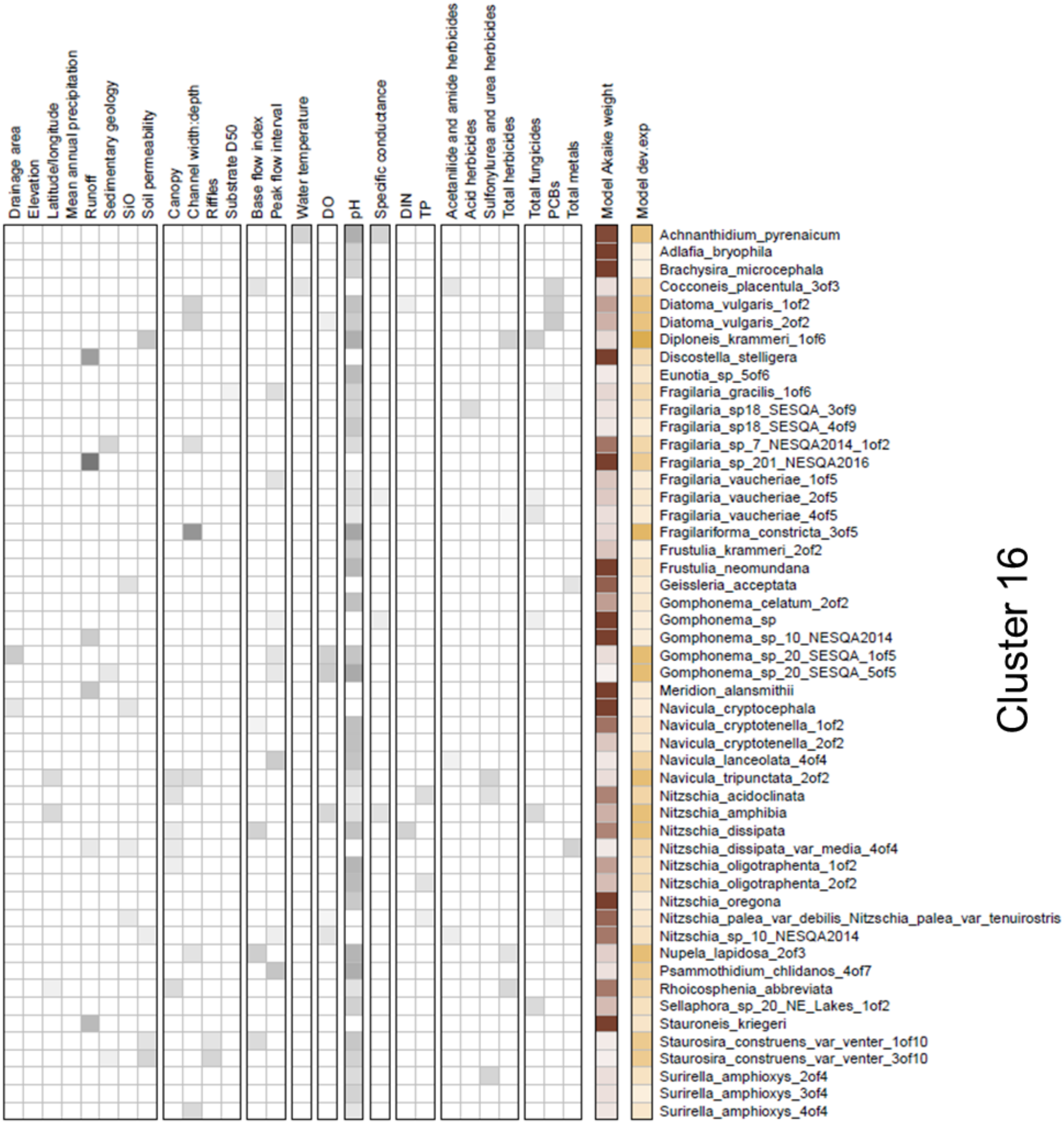


Figure S2 (continued). (9 of 12)

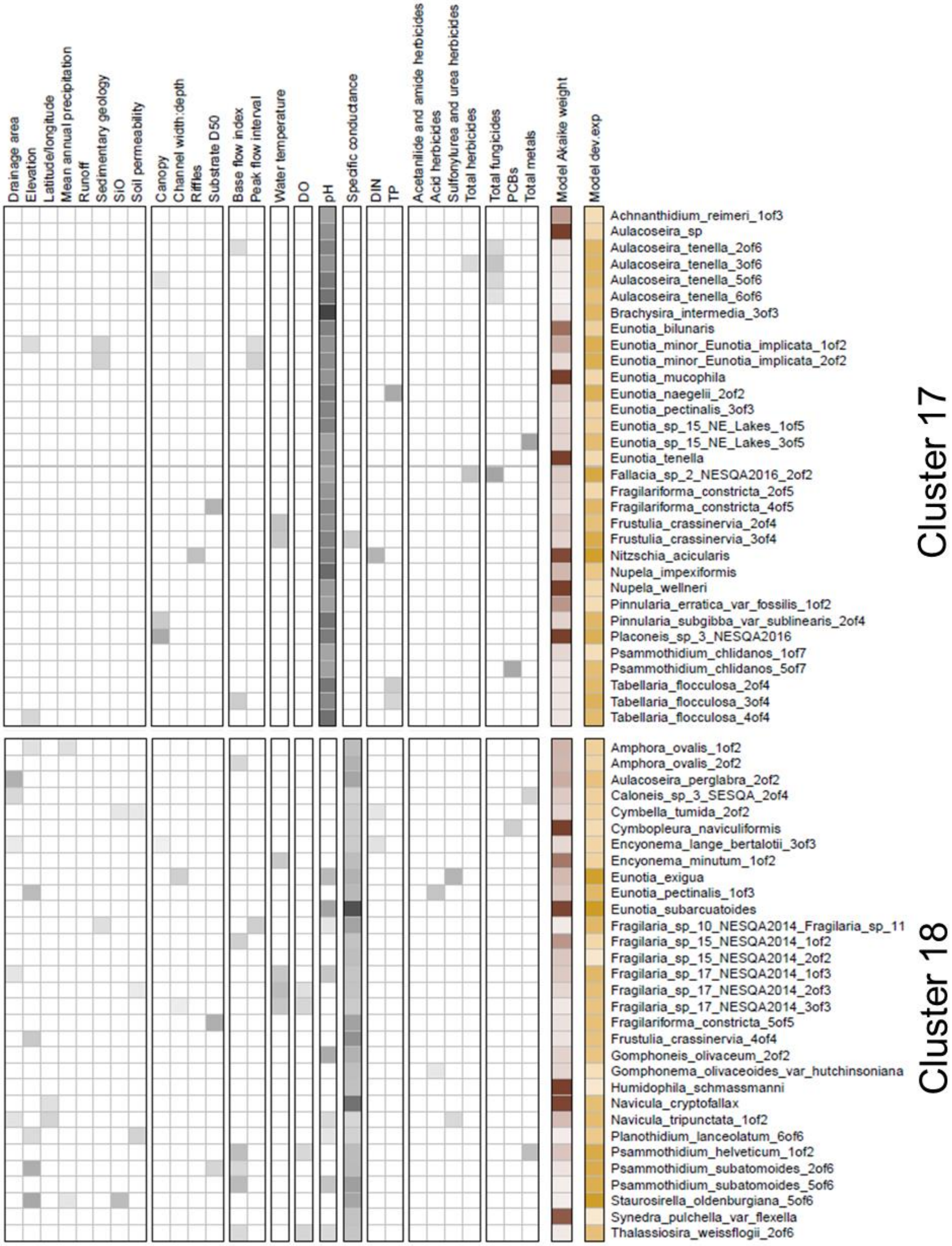


Figure S2 (continued). (10 of 12)

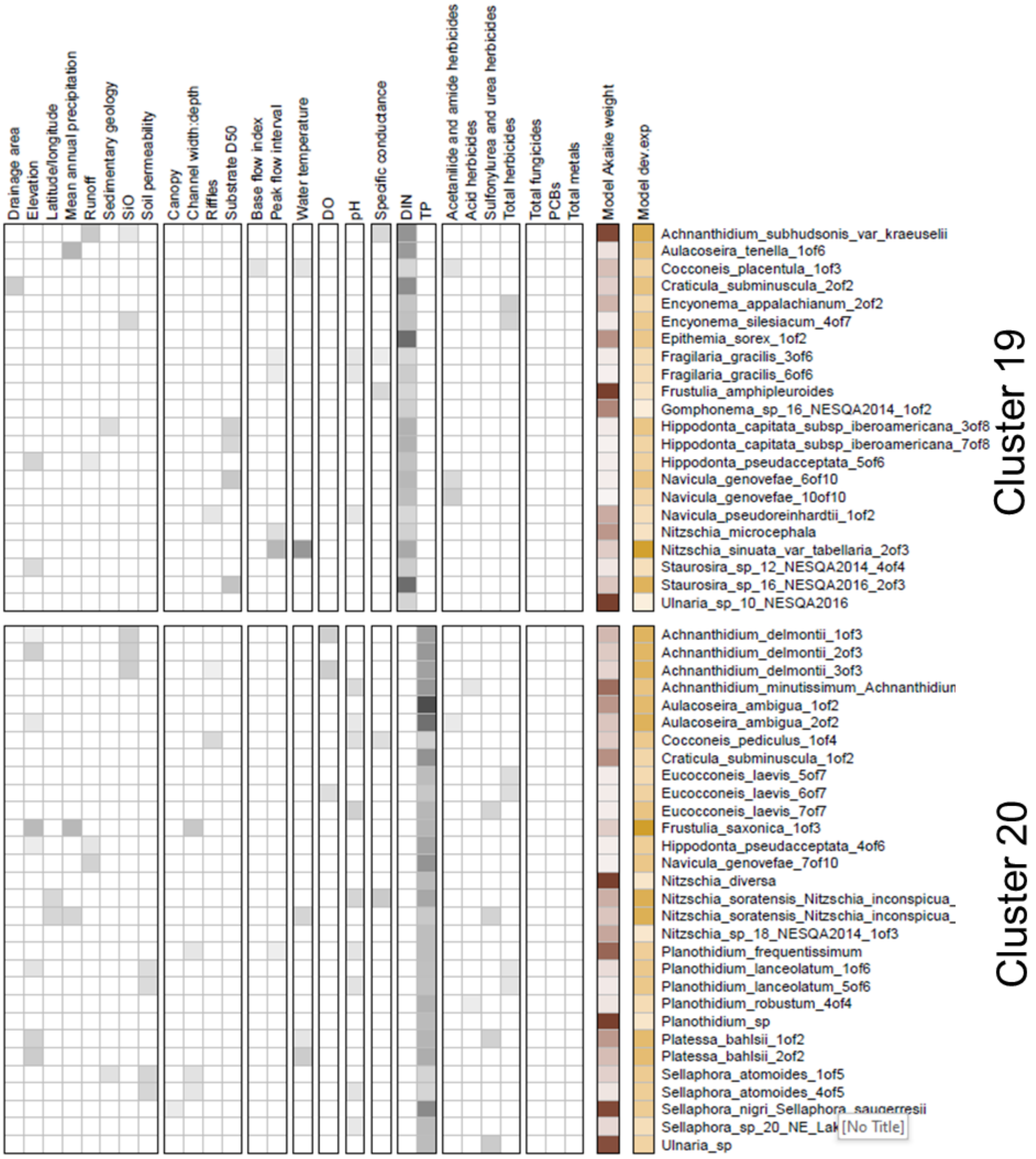


Figure S2 (continued). (11 of 12)

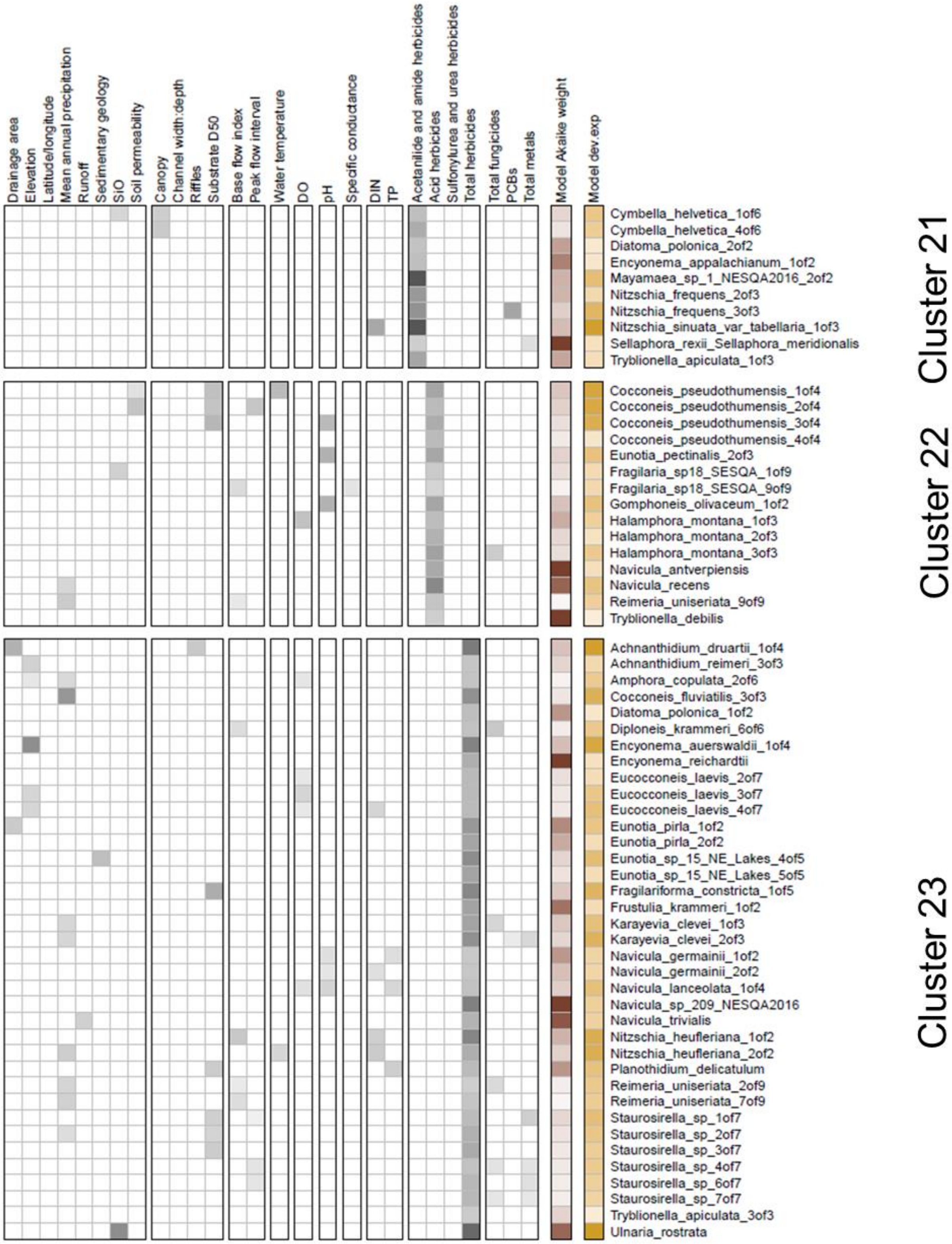
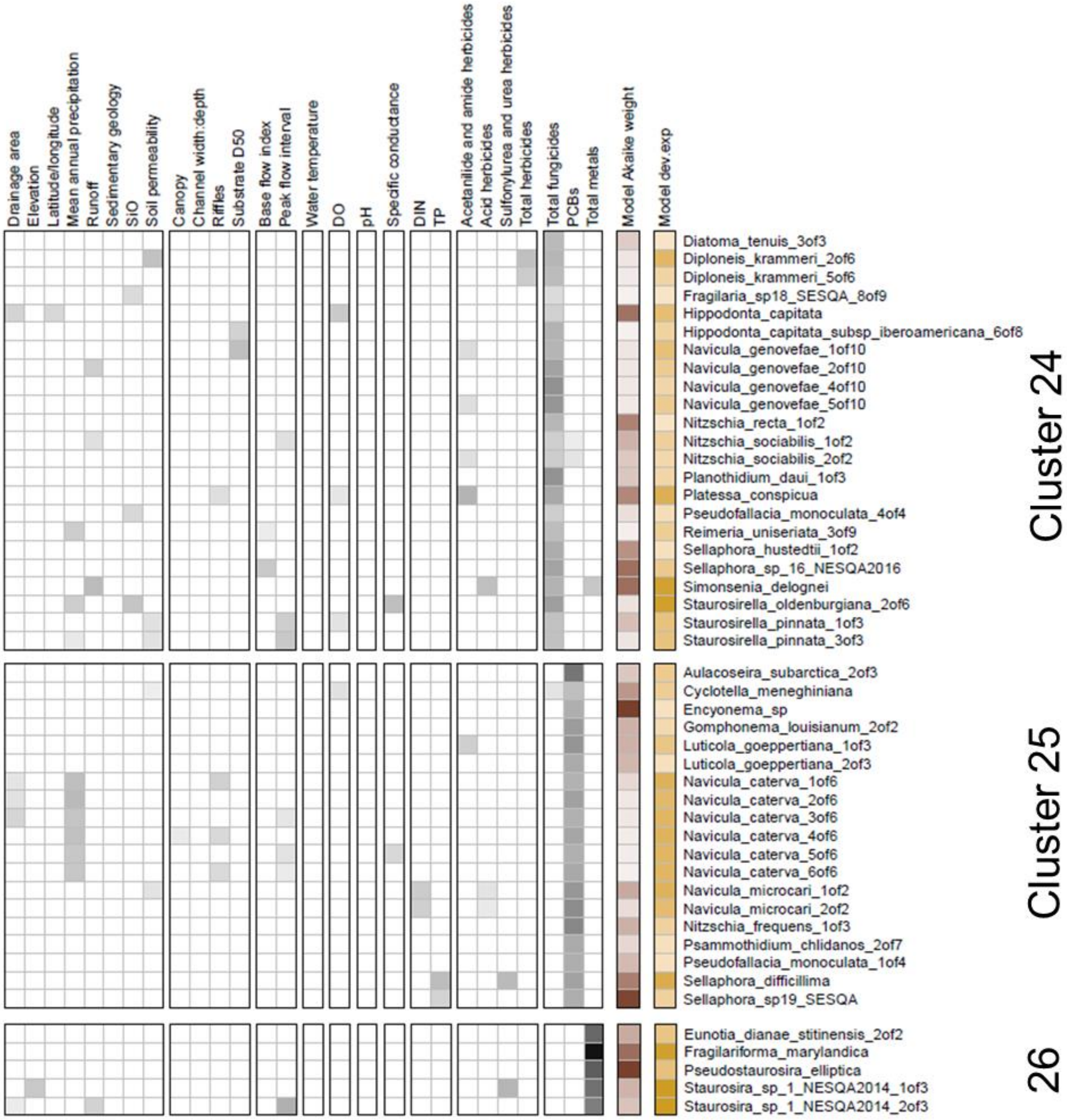


Figure S2 (continued). (12 of 12)



Chapter IV Supplementary Material

Table S1. Variance explained by each principal component in the principal components analysis of inter-site similarity as defined by environmental factors.

Principal component	Standard deviation	Proportion of Variance	Cumulative Proportion
PC1	**2.287**	**0.194**	**0.194**
PC2	**1.952**	**0.141**	**0.335**
PC3	**1.664**	**0.103**	**0.437**
PC4	**1.384**	**0.071**	**0.508**
PC5	**1.310**	**0.064**	**0.572**
PC6	**1.180**	**0.052**	**0.624**
PC7	1.060	0.042	0.665
PC8	1.024	0.039	0.704
PC9	0.987	0.036	0.740
PC10	0.933	0.032	0.772
PC11	0.912	0.031	0.803
PC12	0.835	0.026	0.829
PC13	0.793	0.023	0.852
PC14	0.774	0.022	0.875
PC15	0.733	0.020	0.894
PC16	0.678	0.017	0.911
PC17	0.639	0.015	0.927
PC18	0.625	0.014	0.941
PC19	0.566	0.012	0.953
PC20	0.491	0.009	0.962
PC21	0.467	0.008	0.970
PC22	0.458	0.008	0.978
PC23	0.432	0.007	0.985
PC24	0.405	0.006	0.991
PC25	0.334	0.004	0.995
PC26	0.310	0.004	0.998
PC27	0.213	0.002	1.000

Bold = PC explaining > 0.05 proportion of variance

Table S2. Descriptive statistics of relative abundances of algal groups, calculated as percent of total natural units or cell numbers across algal groups per site.

	Cyano-bacteria (%)	Diatoms (%)	Non-diatom eukaryotic algae (%)
Natural units			
Min	0	22	0
Med	11	83	3
Mean	16	78	6
Max	67	99	45
Cell number			
Min	0	1	0
Med	67	27	4
Mean	61	31	8
Max	98	95	68

Min = minimum; Med = median; Max = maximum

Figure S1. Two-dimensional ordination of the principal components analysis of inter-site similarity as defined by environmental factors.