IMPROVING DIATOM ENUMERATION METHODS FOR USE IN PREDICTIVE BIOASSESSMENT MODELS

by

MEREDITH A. TYREE

B.S., University of Central Arkansas, 2014

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

Master of Science

Department of Environmental Studies

2017

This thesis entitled:
Improving diatom enumeration methods for use in predictive bioassessment models
written by Meredith A. Tyree
has been approved for the Department of Environmental Studies

Diane M. McKnight, Ph.D.

Sarah A. Spaulding, Ph.D.

Daren M. Carlisle, Ph.D.

Date

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

ABSTRACT

Tyree, Meredith A. (M.S., Environmental Studies)
Improving diatom enumeration methods for use in predictive bioassessment models
Thesis directed by Professor Diane M. McKnight

Diatoms are routinely sampled in biological assessments of water quality, but the method traditionally used to characterize diatom communities does not adequately capture species richness for use in most assessment applications. The traditional enumeration method of 300 cell (or 600 valve) fixed counts was designed to characterize the relative abundance only of dominant taxa, making it inappropriate for common bioassessment applications such as observed/expected (O/E) models, which rely on species richness. We analyzed the nature of diatom communities in reference sites of varying diversity using a measure of counting efficiency, which revealed that 600 valve fixed counts did not consistently characterize high diversity sites compared to low diversity sites. To address this problem, we compared the fixed count method to a stratified method, which captures both abundance and richness, and a timed presence method, which captures richness. The stratified and timed presence methods captured greater species richness compared to fixed counts. We then evaluated the performance of these methods in O/E models using genus and species-level data. The timed presence method produced more sensitive and precise models than the fixed method at both the genus- and species-level. A timed presence method could thus improve measurements of stream health while expediting analyses and saving effort.

ACKNOWLEDGEMENTS

I wish to thank my supervisor and mentor Dr. Sarah Spaulding for offering support and guidance throughout this process, while consistently allowing me to explore my own paths and make this thesis my own. She always provided calm reassurance when I most needed it, and her confidence in me helped me to instill confidence in myself.

I would also like to thank Dr. Diane McKnight, who, along with providing valuable insight into the contents of this thesis, dauntlessly advocated on my behalf to the department when scheduling conflicts arose.

My sincere thanks also go to Dr. Daren Carlisle, who devoted much time to instructing me in the workings of the statistical models used in this thesis. His insight and outside perspective greatly enhanced the direction of this research.

I further wish to thank my colleagues Ian Bishop and Nicholas Schulte, who generously provided feedback throughout my research and whose suggestions significantly impacted the direction of this thesis.

Finally, I must express my profound gratitude to my loved ones who have unfailingly supported me throughout my studies. Although baffled by my fascination with “scum,” my family, friends, and partner have given their unwavering encouragement even when hundreds of miles away.

Thank you all.

CONTENTS

TABLES

Table

FIGURES

Figure

CHAPTER 1

REGIONAL BIOASSESSMENTS: POLICY AND HISTORICAL FOUNDATIONS

WATER QUALITY SERVICES

As recently as fifty years ago, streams in the United States regularly caught fire due to the massive amounts of organic pollution they contained. Our nation's waters are no longer flammable, thanks to important regulatory progress and technological advances. Nonetheless, the degradation of our streams remains serious, despite being less visually striking (Andreen 2013, USEPA 2013). The establishment of the Clean Water Act (CWA) in 1972 has been monumental in improving and maintaining the quality of U.S. water resources through regulation of point sources of pollution, such as wastewater treatment plants. However, the Act does not directly regulate nonpoint sources of pollution, such as agriculture (USEPA 1972). Nutrients from nonpoint sources are now the main source of impairment for streams nationally (USEPA 2013), threatening biological integrity, human health and livelihood, and recreation.

Eutrophication of streams threatens the biological integrity of aquatic life both directly and indirectly. Species in most major trophic levels (e.g. fungi, algae, invertebrates, and fish) exhibit sensitivity to nutrient pollution, and communities can become less diverse under conditions of eutrophication (Hillebrand and Sommer 2000, Gulis et al. 2006, Weijters et al. 2009, Dunck et al. 2015). In addition, nutrient runoff from lawns and agricultural fields fertilize algae, causing them to grow and proliferate rapidly in what is called an algal bloom. When the algae die, the decomposition process consumes massive amounts of oxygen in the water, which can kill fish and threaten fish populations.

Degradation of stream biological communities can result in economic losses that threaten human livelihoods. Due to increasing incidence of algal blooms, many communities throughout the US are experiencing economic losses due to devastated fish populations. For example, the Chesapeake Bay and Albemarle-Pamlico Estuary, the two largest estuaries in the U.S., have seen steadily declining fish catches since the early 1980s (Whitehead et al. 2000, Iho et al. 2015). Commercial fishing and tourism driven by recreational fishing in these estuaries support tens of thousands of jobs and bring in billions of dollars of revenue to neighboring regions each year. As extensive fish kills continue, these jobs and revenue are compromised (CBF 2012, Hadley and Wiegand 2014, APNEP 2015).

In addition to declining fish populations, eutrophication of US waterbodies is also increasing the incidence of harmful algal blooms (HABs). HABs include blooms of cyanobacteria capable of producing toxins harmful to human health and livelihood (Paerl et al. 2001). Consequences of cyanotoxins for human life are substantial, threatening human health. HABs can also negatively impact economies through livestock poisoning and impaired drinking water supplies (Falconer 2001). Toxins kill thousands of livestock animals each year, causing large economic losses for ranchers (Falconer 2001). Furthermore, elevated cyanotoxins can significantly increase drinking water treatment costs (Sklenar et al. 2016). The additional cost of treatment to remove cyanotoxins from US waterbodies in 2009 was an estimated $8.5 million for Charleston, IL, $15 million for Mattoon, IL, and $31.8 million for Fairmont, MN (Hamilton et al. 2014).

Not only does the quality of freshwaters have serious implications for biological integrity and human economics, water quality is also tied to an array of nonmaterial recreational

benefits (Satz et al. 2013). In addition to commercial fishing, declining fish populations also threaten enjoyment of recreational fishing (Hoagland et al. 2002). When water quality as measured by water clarity improves, residents in an area might increase their frequency of swimming and fishing (Vesterinen et al. 2010). Harmful algal blooms result in lake and reservoir closures that inhibit boating, swimming, and fishing (Backer et al. 2010).

Furthermore, although nutrients are currently the main source of impairment of U.S. streams, nonpoint source pollution also increases several other potential stressors for ecosystems such as sediment, salinity, and pesticides. Excess sediment can cause direct mortality in invertebrates and fish, and increased turbidity inhibits primary production, depressing growth and reproduction up the food web (Henley et al. 2000). Elevated salinity in streams interferes with invertebrate osmoregulation, resulting in reduced decomposition of organic matter and potentially impairing food availability in higher trophic levels (Schäfer et al. 2012, Tyree et al. 2016). Pesticides in sub-lethal concentrations can alter periphyton community structure and suppress respiration (Dorigo et al. 2010), alter invertebrate community structure (Liess and von der Ohe 2005), and reduce decomposition of organic matter (Schäfer et al. 2012).

NATIONAL BIOASSESSMENT POLICY FOUNDATIONS

Identifying national trends in the conditions of U.S. waterbodies is vital for crafting effective regulation and using taxpayer dollars effectively. Yet, the ability to measure national aquatic conditions is relatively new. Although the CWA set ambitious goals for water quality with the overall purpose to “restore and maintain the chemical, physical, and biological integrity of the Nation’s waters,” the structure of the law made it difficult to acquire national-

level data to monitor progress toward this goal (Fowler 2014). Responsibility to monitor waterbodies and set water quality objectives, as outlined by section 305(b), falls primarily to the states, which then report their progress back to the EPA every two years. The EPA then compiles the states' data into a single report called the National Water Quality Inventory, which is delivered to Congress as required under section 303(d).

Flexibility afforded the states, however, historically resulted in datasets too fragmented and disparate to inform national water quality status. Data was measured at temporal and spatial scales that varied between states, and states chose to use different indicators and sampling methods (Braden et al. 2014). State data could not even consistently provide state-level trends because some states did not assess the same waterbodies each year, and although some states used randomized sampling designs, others purposefully targeted problematic waterbodies (Andreen 2013).

Just five years after the passing of the CWA, the National Research Council (NRC) issued a report critiquing the EPA with the comment, "Responsibilities for environmental monitoring are fragmented, and there is inadequate coordination among programs serving different purposes" (NRC 1977). In 1989, the EPA initiated a program called the Environmental Monitoring and Assessment Program (EMAP) to research and develop ecological indicators, standardized protocols, and study designs suited to measuring environmental health at the regional and national scales (Shapiro et al. 2008), but early EMAP efforts did not immediately culminate in a large-scale application (Bradley and Landy 2000).

In 1995, the White House Office of Science and Technology Policy asked the H. John Heinz III Center for Science, Economics, and the Environment to create a report on the state of

the nation's environment. The report, released in 2002, acknowledged data was insufficient to report national trends for many indicators of freshwater ecosystem health. For example, U.S. Geological Survey (USGS) stream gages provided adequate national data to examine temporal trends in streamflow, but other indicators, such as extent of stream miles, fish die-offs, and biological community integrity did not have enough data to establish national conditions. Other indicators, such as stream habitat quality, had not even been sufficiently developed to analyze existing data (H. John Heinz III Center 2002). The United States General Accounting Office (GAO) also released a report in 2000 at the request of Congress, which definitively stated the EPA's National Water Quality Inventory reports did not accurately portray nationwide conditions of water quality. In addition to urging the EPA to address problems with differing methodologies, standardization of definitions, and comparability of data, the GAO highlighted the need for sampling designs that enabled statistically valid extrapolation of conditions to unmonitored sites (GAO 2000).

These and similar reports underscored the need for new monitoring approaches that would provide nationally consistent datasets capable of accurately characterizing the conditions of U.S. water resources. Meanwhile, the EPA implemented its first regional pilot in the EMAP program called the Mid-Atlantic Integrated Assessment (MAIA), covering a region from southern New York to northern North Carolina (Bradley and Landy 2000). Success with MAIA resulted in a second pilot termed the EMAP-West that included 12 western states plus the coasts of Alaska and Hawaii (USEPA 2001). Both the MAIA and EMAP-West served as proofs-of-concept for the technical foundations developed in the EMAP, and the EPA next

turned to its most ambitious assessment to date: the first national survey of U.S. streams called the Wadeable Streams Assessment (WSA), conducted in 2004.

The EMAP pilots and the WSA represent a turning point in how the federal government monitors U.S. water resources. Standardization of methods and collaboration between states now enable assessment of national conditions and insight into how the nation's water resources are faring (Paulsen et al. 2008). National and regional water quality assessments have become integral within the EPA. The WSA was succeeded by the National Rivers and Streams Assessment, first conducted in 2008 and then again in 2013. Another NRSA is scheduled to begin collection in 2018.

STREAM QUALITY IN THE SOUTHEAST U.S.

Although the evolution of the EPA's national surveys was vital to bioassessment in establishing an understanding of the biological condition of US streams over a large geographical area, these surveys are based on a probabilistic sampling design in order to extrapolate results from monitored sites to unmonitored sites (Hughes and Peck 2008). The tradeoff of this probabilistic design is that it informs average values of stressors rather than encompassing their full range. EPA national surveys are thus best considered as complementary to targeted surveys, which better address specific questions of interest (Rehn and Ode 2009). Starting in 2013, the USGS began conducting a series of regional surveys of stream quality based on a targeted survey design throughout the country termed Regional Stream Quality Assessments, or RSQA (USGS 2017).

The second RSQA conducted by the USGS was the Southeast Stream Quality Assessment (SESQA). The goal of SESQA was to establish the relation of multiple stressors such as

contaminants, nutrients, sediment, and streamflow alteration to ecological conditions in streams throughout the Piedmont and southern Appalachian Mountains to inform policymakers and stakeholders in the region about the state of their water quality (USGS 2014). Sites were sampled along urban multi-stressor and hydrologic-alteration-stressor gradients (Journey et al. 2014). Sampling sites extended through Alabama, Georgia, North Carolina, South Carolina, Tennessee, and Virginia.

The Southeast has a long history of extreme changes in land use. Deforestation for agricultural expansion began in the 19th century, followed by conversion of agricultural land to coniferous forests by the forest industry in the 1930s-1970s (Trimble 1974). More recently, agricultural and forested lands are being converted to urban land as metropolitan areas in the region expand, with the Southeast experiencing the fastest net population growth in the nation (O'Driscoll et al. 2010). Still, 33% of land in the SESQA study area is operated as farmland (National Agricultural Statistics Service 2017). The result is an array of potential biochemical stressors such as elevated nutrients, pesticides, pharmaceuticals, and sediment (Scott et al. 2002, Gregory and Calhoun 2006). These biochemical stressors are worsened by hydrologic alterations throughout the region, with the highest density of dams nationwide (Graf 1999). Even in streams that have undergone restoration, impacts of historical land use remain evident today (Surasinghe and Baldwin 2014).

DIATOMS IN REGIONAL ASSESSMENTS

Both NRSA and RSQA rely on biological indictors as a measure of stream health, representing the CWA's valuing of aquatic biological integrity as well as utilizing bioindicators' ability to assimilate and represent their environmental conditions better than snapshot

chemical indicators alone (Karr and Chu 1999). In particular, diatoms have long been recognized as valuable indicators of stream health due to their ubiquity and responsiveness to environmental conditions (Patrick 1973). Short algal lifespan and rapid rate of reproduction make algae effective indicators of short-term environmental conditions (Smucker et al. 2013). Hering et al. (2006) found that diatoms reflected eutrophication better than macrophytes, macroinvertebrates, or fish, but were unresponsive to hydrologic impairment, highlighting the importance of algal indicator use in conjunction with other bioindicators. However, although use of diatoms as bioindicators has proved useful in measuring stream impairment at a local scale (Stevenson 1998, Nodine and Gaiser 2014), large regional assessments of water quality using diatoms have not yet seen the success of macroinvertebrate assessments (Potapova and Charles 2005, Stevenson et al. 2008, Potapova and Carlisle 2011), potentially due to two issues that this thesis seeks to address: inconsistency in taxonomy between analysts and the method with which diatoms have traditionally been counted.

Unlike small-scale assessments, in which taxonomic samples are few enough to be analyzed by a single individual or laboratory, samples from large regional assessments are typically contracted out to multiple laboratories. For example, macroinvertebrate samples in the WSA were contracted out to 25 taxonomists associated with 8 different laboratories. An assessment of taxonomic consistency in the WSA found the degree of taxonomic error varied among laboratories when compared to an independent quality control taxonomist, but error decreased among laboratories after coordination through a conference call and specific procedural corrections assigned to some of the laboratories (Stribling et al. 2008). Similarly, macroinvertebrate samples collected throughout streams in Germany as part of the European

Union Water Framework Directive were contracted out to 7 commercial laboratories. More than 30% of taxa differed in original counts compared to those conducted by independent auditors (Haase et al. 2010). In both studies, approximately 20% of sites changed ecological quality classification after taxonomic changes resulting from quality control efforts (Stribling et al. 2008, Haase et al. 2010).

Taxonomic consistency has been particularly problematic in algal bioassessment. The Academy of Natural Sciences in Philadelphia has hosted a series of annual diatom taxonomic workshops since 1999 to improve harmonization of names applied by different analysts to difficult species in the USGS Nation Water Quality Assessment program (Potapova et al. 2008). Still, taxonomic consistency between analysts continues to create problems in bioassessment. In the first RSQA, conducted in the Midwest in 2013, diatom analysis was contracted out to 5 analysts; as a result, clear patterns in taxonomic composition among sites were observed and appeared to be due to different analysts (Bishop et al., in prep). Similarly, in the EPA's 2008-2009 National Rivers and Streams Assessment, diatom samples were contracted out to 3 laboratories and 11 analysts for enumeration. The "analyst signal" in the diatom data was stronger than the environmental signal, rendering the data problematic for bioassessment in its original form (Lee et al., in prep). Although Lee et al. used *post hoc* approaches to retroactively adjust the taxonomic data to improve the environmental signal relative to the analyst signal, the process was time-consuming. A number of *a priori* approaches could prevention taxonomic inconsistency in the first place, saving time and money (Bishop et al. in prep).

In addition to issues with taxonomy, the enumeration method traditionally used to characterize diatom communities has been applied without question. The traditional count

method prescribes slides be enumerated in transects until a designated total number of valves is reached, usually between 300-600 valves (Charles et al. 2002). In samples dominated by a few taxa, this count is often not high enough to capture the species richness of a site. Low diversity in some sites (i.e. dominated by a few taxa), in particular, puts diatom communities at risk of mischaracterization, especially when rare species are excluded to reduce statistical noise (Cao et al. 2001). Depending on the desired project outcomes, two alternative methods of diatom enumeration methods could better capture sample richness for use in diatom bioassessment: stratified counting and timed presence. I discuss these methods in detail in Chapter 2.

CHAPTER 2

IMPROVING DIATOM ENUMERATION METHODS FOR USE IN PREDICTIVE BIOASSESSMENT MODELS

INTRODUCTION

Diatoms have been integral to aquatic bioassessment for the past forty years (Patrick 1973) and are now included in regional and national assessments of water quality (Potapova and Charles 2005, Kelly et al. 2008b, Stoermer and Smol 2010). Biological indicators of water quality typically rely on taxonomic data (i.e., counts of taxa at each sampling site) derived from laboratory processing with an analysis method called a "fixed count" method, in which cells are identified and enumerated until a specified number of cells is reached. However, the fixed count method may not be appropriate for assessments utilizing measures of species richness, such as predictive bioassessment models, because the fixed count method characterizes dominant taxa well but inconsistently characterizes non-dominant taxa. In this study, we evaluate the efficacy of three counting methods (fixed, stratified, and timed presence) to characterize richness of diatom assemblages.

Across environments and taxonomic groups, ecological communities are almost universally composed of few species that are very abundant, and many more species that are rare. Species abundance distributions thus follow a truncated lognormal curve, where the majority of species fall within the curve's long tail (Magurran 2004). Rare species are commonly removed during analysis of bioassessment data because rare species are perceived as adding noise, thereby obscuring relationships between environmental stress and biological communities (Poos and Jackson 2012). This noise, however, might result from improperly characterizing the tail of

species abundance distributions, rather than from rare species lacking a clear environmental response to environmental stress.

One commonly-used method for assessing the biological quality of waters is taxonomic completeness (Hawkins 2006), which considers native species diversity as an inherently valuable indicator of the health of a waterbody, directly addressing the Clean Water Act's goal of restoring and maintaining the "biological integrity" of US waters (USEPA 1972). Taxonomic completeness is quantified simply as the ratio of observed (O) taxonomic composition to what is expected (E) under minimal anthropogenic influence (Bailey et al. 2004). Although O/E indicators have been successfully developed for invertebrates (Armitage et al. 1987) and fish (Meador and Carlisle 2009), less success has been achieved for diatoms (Cao et al. 2007, Ritz 2010). It is possible that past diatom O/E model performance suffered because the enumeration method traditionally used to characterize diatom communities is not well-suited to O/E models.

Although diatoms are routinely analyzed in bioassessment, there is fundamentally no agreement on the acceptable minimum number of diatom half-cells (called valves) needed to characterize an assemblage. In early studies examining the effects of pollutants on rivers, fixed counts of 3000-8000 valves were recommended, depending on diversity and the type and degree of impairment (Patrick et al. 1954). Such large fixed counts are time-intensive and costly, and for the past three decades, an emphasis on characterizing total abundance or relative abundance of dominant taxa by counting a far lower number of valves, 300-600, has prevailed (Battarbee 1986, Prygiel et al. 2002). These less costly, abbreviated fixed counts have since become the norm in regional and national assessment, and are a component of routine

protocols. For example, assessment of rivers in the United Kingdom, following the European Union Water Framework Directive, relied on fixed counts of 300 valves (Kelly et al. 2008a). In the United States, the U.S. Environmental Protection Agency (EPA) Environmental Monitoring and Assessment Program (EMAP) used 500-valve counts (Pan et al. 1996). Current assessment protocols specify that 600 valves are counted in the EPA National Rivers and Streams Assessments (NRSA) and the U.S. Geological Survey (USGS) National Water Quality Assessment (NAWQA) program (Charles et al. 2002, USEPA 2009).

The 300-600 fixed-valve count method appears to have originated for the specific purpose of characterizing the dominant taxa in the community. For example, Battarbee (1986) calculated the change in the percent frequency of dominant taxa as the number of valves counted increased. Battarbee concluded that because the relative abundance of dominant taxa changed markedly between a count of 100 to 200 valves but changed only marginally between 400 to 500 valves, a fixed count of 300 to 600 valves could be recommended for most analyses. Most biological indicators of water quality, however, depend on accurate measures of species relative abundance and species richness (Cao et al. 2007, Potapova and Carlisle 2011, Kelly 2013). Due to dominance by one or two taxa, up to 70% of diatom species might be classified as rare and excluded from further analysis (Potapova and Charles 2002). It is likely, however, that many excluded species are not actually spatially rare (i.e. occur at few sites). Instead, many excluded species are widespread, but low in abundance. If these species are identified at sites consistently, taxa with low abundance and high occurrence could provide an important signal to distinguish impacted from reference sites (Gillett et al. 2011).

Furthermore, fixed counts are widely utilized by analysts with the presumption that analytical effort is standardized between samples by the number of valves counted. In reality, however, given species diversity varies widely among samples in many datasets, richness in samples with low diversity stands to be characterized much more accurately than in samples of high diversity using the fixed count method. Stevenson et al. (2010) suggest using a method called "stratified counting" for diatom assemblages in which one, or a few, taxa are dominant. Stratified counting involves enumerating all valves until a pre-specified number of valves of one taxon is reached and its relative abundance calculated. Then, enumeration of all valves except that taxon continues until the desired total count is reached. Valve counts of dominant taxa are estimated by scaling their actual counts by the area of the slide that was examined (Stevenson et al. 2010). However, although stratified counting has been employed to better detect rare taxa (Spaulding et al. 1997), it has not been directly compared to fixed count methods. Consequently, a means to standardize effort among samples using stratified counts has not been established. In addition, for applications in which abundance data is not required, such as O/E models, a low-effort counting method relying on presence/absence only and standardized by time could more accurately characterize site richness than fixed counts or stratified counts, and at significantly lower cost.

Our objectives were to characterize the nature of diatom communities in reference sites of varying diversity and use our results to develop two alternative enumeration methods that better consistently capture species richness for use in bioassessment: the stratified method and the timed presence method. We then compared the fixed enumeration method to our best-performing alternative method (timed presence), by developing O/E models using both the

traditional and timed presence enumeration methods and comparing model performance metrics.

METHODS

Sampling Design and Site Selection

Measuring stream impairment in bioassessment requires establishing a control to which test sites are compared by characterizing biotic assemblages expected at sites in the absence of impairment over the range of natural (e.g. climate/hydrology and stream gradient) environmental conditions expected at test sites. We therefore extracted diatom count data and sampling site information for 68 reference sites from three regional assessments described below. Each assessment individually identified reference sites as sites in least-disturbed condition (*sensu* Stoddard et al. 2006) based on land cover, in-stream chemical and physical conditions, and local expertise (Carlisle et al. 2008, Herlihy et al. 2008, Journey et al. 2014).

We used data collected at 21 sites from the USGS National Water-Quality Assessment (NAWQA) Program (1993-2000) in the southeastern US that were previously identified as reference quality (Carlisle et al. 2008). In addition, we used data from 20 reference sites sampled in 2014 as part of the NAWQA Program's Southeast Stream-Quality Assessment (SESQA; Journey et al. 2014). NAWQA samples were collected following richest-targeted habitat protocols. Periphyton was scraped from coarse substrate (cobble or wood) within riffle habitats using a stiff-bristled brush and an area delimiter, combined into a single composite sample, preserved with 10% buffered formalin, and transported to the laboratory on ice for processing (Moulton, S.R. et al. 2002). We also used data from an additional 27 reference sites sampled as part of the U.S. EPA NRSA program (2008/2009) using a probability-based sample design

(USEPA 2013, USEPA 2016). Samples were collected from the left, right, or center of each transect as designated at random. Where possible, periphyton was scraped from coarse substrate using an area delimiter. When coarse substrate was not available, the top 1 cm of sediment was vacuumed from the streambed within the delimited area using a syringe. Samples from all transects were combined into a single composite sample, preserved with 10% buffered formalin, and transported to the laboratory on ice for processing (Gilliom et al. 1995). In the laboratory, organic matter was removed from samples using nitric acid and a microwave digester (Acker et al. 1999a). Cleaned periphyton material was pipetted onto coverslips and mounted to permanent slides using Naphrax™ mounting medium (Acker et al. 1999b).

Diatom Analysis

First, a "working flora" of voucher images was developed to ensure consistency in the morphological concept of each taxon between samples and across analysts. The working flora consisted of a set of images of each taxon present in the slides from the complete group of reference sites from NAWQA, SESQA, and NRSA. Specimens were selected to demonstrate the morphological variation and size range of each taxon, resulting in 1-17 images for each. All 68 slides were examined under the light microscope (Olympus Vanox) using a 100x oil immersion objective (1.3 NA) and differential interference contrast (DIC). Images were collected using a Micropublisher 3.3 RTV QImaging digital camera, and assembled into groups by genus. During the counting process, additional taxa not encountered during flora-building were micrographed and added to the working flora as they were encountered. Thus, a voucher flora of 2780 images was compiled, grouped into species, and assigned provisional identification codes. The document serves as a permanent record of this study.

To understand the ramifications of characterizing diatom communities using traditional and alternative counting methods, we first analyzed each site using a fixed count of 600 valves using standard methods (Charles et al. 2002). We calculated Shannon Diversity for each site and used these values to assign sites to one of three diversity categories: low (0.45 – 0.59; 13 sites), medium (0.60 – 0.78; 28 sites), and high (0.79 – 0.97; 47 sites). We then randomly chose one site from each diversity category to serve as precursory "test sites" that would be analyzed 3 times using 3 different methods: 1) a fixed count, 2) a stratified count, and 3) a timed presence count. The test sites served as a preliminary comparison between counting methods, which we then used to choose the best alternative method (either stratified or timed presence) to be more thoroughly compared to the fixed method.

In all 3 analyses, slides were examined under the light microscope along transects at 1000× magnification. Identifications of taxa were based on following identification codes in the working flora. Valves were only counted if at least 60% of a valve was visible within the field of view. For the fixed and stratified methods, we evaluated how well each count captured the diatom assemblage by calculating the enumeration efficiency (Pappas and Stoermer 1996). This measure is an estimate of the probability that additional taxa will not be encountered with further enumeration (Eqn. 1).

$$Efficiency = 1 - \frac{number\ of\ species}{number\ of\ individuals}$$

In the stratified count method, valves were enumerated along a transect until 50 valves of a given taxon were encountered. At that point, we continued enumerating all valves *except* that taxon. As the analysis progressed, we repeated this pattern until a total 600 valves had been observed and identified. The dominant taxa counts were then scaled to their estimated

"effective counts" by multiplying their actual counts by the area of the slide examined. The effective counts thus estimate the number of valves of each taxon that would have been counted with the traditional method had we continued counting past 600 valves to the total effective count. We calculated the efficiency and total effective valve count for each site over the course of the stratified counts.

For a timed effort to uncover species presence, we recorded species presence over a fixed time frame. In this method, we enumerated the first 100 valves encountered along a transect to estimate the relative abundance of dominant taxa. We then scanned the slide in non-overlapping transects for a period of one hour, recording only newly encountered taxa.

To determine which alternative counting method was uncovering the greatest species richness, we calculated the number of species detected at each test site using the fixed, stratified, and timed presence methods. We used these richness values based on our 3 preliminary test sites to select the alternative method we would apply to all 68 sites for a more thorough comparison with the fixed method. We then explored whether new species detected in the best-performing method were truly spatially rare taxa versus widely distributed taxa of low abundance by determining which taxa were detected in each timed presence count that were not detected in the fixed count of the same site. Finally, we calculated how many of these taxa were spatially rare, defined as taxa occurring in less than 10% of sites.

Model construction

Our timed presence method was designed to maximize detection of species richness with minimal time and effort, but did so at the expense of collecting species abundance data. RIVPACs-type models (*sensu* Hawkins 2006), which do not require abundance data and have

been successfully applied using invertebrates, were therefore developed to evaluate the relative influence of each enumeration technique on one possible biological assessment outcome—taxonomic completeness as defined by O/E (Hawkins et al. 2000, Carlisle et al. 2008, Reynoldson et al. 2016).

The O/E index is defined as the ratio of observed taxa to expected taxa, where a high O/E score (close to 1) indicates an unimpaired site, and a low O/E score (considerably less than 1) indicates degradation of the biological condition of a site. To assess the performance of our models, we calculated the mean O/E predicted by each model for our reference-condition calibration sites to ensure mean O/E scores approximately equaled 1. In calculating O/E scores, we also calculated the precision of each model's ability to correctly predict the presence of taxa by calculating the standard deviation (SD) of O/E. High SD indicates low precision, whereas low SD indicates high precision in predicting community composition at reference sites.

For each model, we also constructed a corresponding null model by omitting environmental clustering and calculating probabilities of capture using only the occurrence of each taxon in all reference sites. The difference between the null model SD and the model SD (Null SD – Model SD) thus measures the ability of each model to account for natural environmental variation in predicted assemblages, with high values indicating low model performance. We also calculated the replicate-sampling standard deviation (SD) of each model, which represents the variability expected from replicate samples taken from the same site on the same day. Together, the null model and the replicate-sampling SD represent the maximum and minimum SD that is theoretically possible for a given model, respectively (Van Sickle et al. 2005).

RIVPACs-type models were developed as described in USEPA (2013). To compare the fixed method with the timed presence method, all data was converted from abundance to binary values of presence or absence. Predicted assemblages were statistically determined by first clustering reference sites by biological similarity using cluster analysis, followed by random forest classification to assign reference sites to groups according to environmental variables (Hawkins et al. 2010). Site-specific probabilities of capture (P_c) for each taxon are computed by multiplying the probability of a site belonging to a particular cluster (based on random forest classification model) and the frequency of occurrence of the taxon across all reference sites that were assigned to that cluster (Wright et al. 1989). In addition to constructing models using species-resolution data for each counting method, we also constructed models at the genus level. Some studies have found comparable results between models using species and genus resolutions (Growns 1999, Hill et al. 2001), although results have been mixed (Rimet and Bouchez 2012).

RESULTS

The efficiency with which the traditional fixed counting method characterized the assemblages was related to the diversity of the community. When the efficiency was calculated over the course of a 600-valve fixed count for one site chosen at random from each diversity category, the low diversity site exceeded 95% efficiency by the end of its 600-valve fixed count, reaching a final 97% efficiency (Fig. 1). The medium diversity site fell slightly short of 95%, reaching a final 93% efficiency. The trendline estimated a total count of 830 valves was required to achieve 95% efficiency in this site. The high diversity site reached only 85% efficiency with a

fixed count. The trendline estimated this site required a total count of 1900 valves to achieve 95% efficiency.

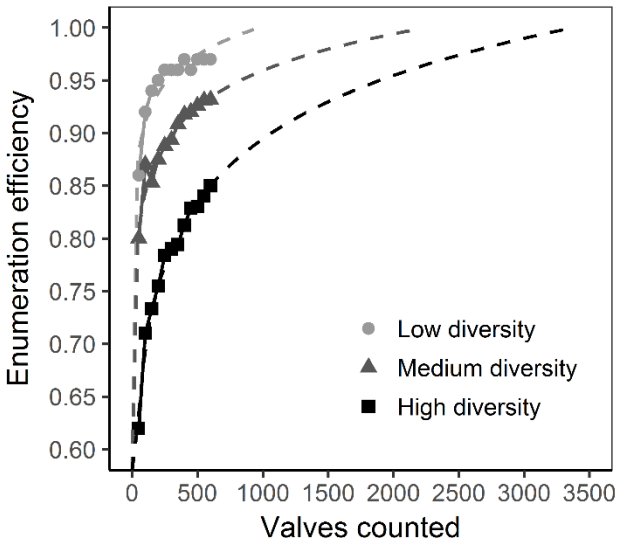

Figure 1. Plot showing the number of valves counted against enumeration efficiency for 3 samples. Enumeration efficiency was calculated every 50 valves over the course of fixed 600-valve counts (solid lines). The points are fit using logarithmic trend lines to estimate the number of valves that would need to be counted to achieve an efficiency of 1 (dashed lines).

Using a stratified method resulted in higher effective valve counts in all sites, but the magnitude differed substantially by diversity (Fig. 2a-c). The stratified method particularly increased the effective count of the low diversity site, where an actual count of 600 valves resulted in an effective count of 8348 valves (Fig. 2a). In the medium diversity site, an actual count of 600 valves gave an effective count of 2126 valves (Fig. 2b). In the high diversity site, the stratified count increased the effective count only marginally, with an actual count of 600 valves resulting in an effective count of 727 valves (Fig. 2c). Surprisingly, however, the stratified counting method had little effect on the final efficiency of all sites, increasing efficiency by 3% in the low diversity site and 4% in the medium and high diversity sites compared to traditional counts (Fig. 2d-f).

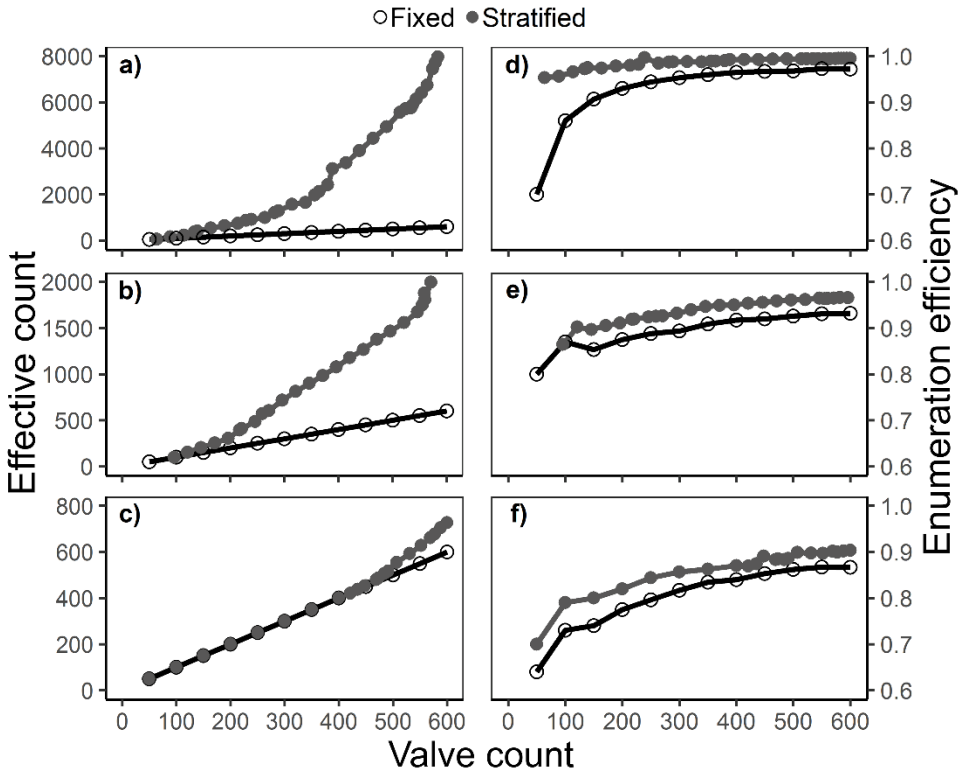

Figure 2. Total valve count plotted against effective count for fixed and stratified counts for a) low, b) medium, and c) high diversity sites. Total valve count plotted against enumeration efficiency for fixed and stratified counts for d) low, e) medium and f) high diversity sites.

The fixed count method detected the fewest species in all 3 of our test sites. The stratified method detected the most species in the low diversity site, and the timed presence method performed best overall, uncovering the most species in the medium and high diversity sites (Fig. 3). Based on the number of species uncovered and the shorter amount of time required to count samples using the timed presence method, we chose to recount all 68 sites using timed presence to compare the method more thoroughly with the fixed count method.

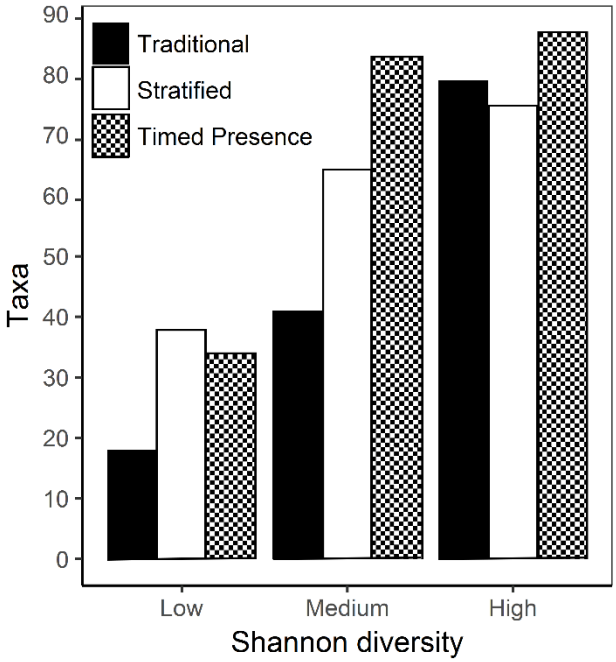

Figure 3. Shannon diversity in low, medium and high diversity sites and the number of taxa encountered using fixed, stratified, and timed presence counting methods.

When compared across all 68 sites, the timed presence method detected a total of 599 taxa not found in each fixed count for the same site. 289 of these taxa (48%) occurred at fewer than 10% of all sites and were considered spatially rare. The number of additional taxa considered rare differed by diversity. 12% of additional species found in low diversity sites were rare, 26% of additional species in medium diversity sites were rare, and 41% of additional species in high diversity sites were rare.

The timed presence method produced RIVPACS-type models with greater performance compared to the fixed count method using both species- and genus-level data, and the timed presence genus-level model performed best overall. Fixed count models at both the species- and genus-level had higher standard deviations than their corresponding null models (Table 1). The timed presence model at the species-level had a lower standard deviation than its

corresponding null model, but only marginally (Table 1). The timed presence genus-level model had a notably lower standard deviation than its null model (Table 1).

Table 1. Mean O/E scores, standard deviation (SD), and replicate standard deviation (Rep SD) for calibration sites. Difference between null model SD and model SD (Null SD – Model SD) measures the relative ability of the model to account for environmental variation in predicting the composition of diatom assemblages (values >0 indicate improvement relative to the Null model).

Model	Mean O/E	SD	Rep SD	Null Mean O/E	Null SD	Null SD - Model SD
Fixed: Species	0.975	0.280	0.193	1.000	0.257	-0.023
Fixed: Genus	1.010	0.206	0.144	1.000	0.202	-0.004
Timed P – Species	1.031	0.194	0.105	1.000	0.205	0.011
Timed P – Genus	1.010	0.147	0.095	1.000	0.172	0.025

DISCUSSION

The fixed counting method is widely used in diatom studies under the assumption it is standardized by the number of valves counted for each site (e.g. 600 valves). Our results indicated, however, that due to variation in the diversity of diatom assemblages, fixed counts do not represent a standardized counting effort among samples. Unequal characterization of sites based on diversity is particularly problematic for datasets set as ours, where diversity varied by 54% among sites and most reference sites in the region were highly diverse.

The traditional method might not suitably characterize even low diversity sites, however, depending on the goal of the study. The efficiency measure developed by Pappas and Stoermer (1996) was highly influenced by dominant taxa and largely unconcerned with less abundant, though not necessarily rare, taxa. Although the stratified method drastically increased the effective counts (Fig. 2a-c) and the number of species encountered (Fig. 3) in our test sites, improved representation of low abundance taxa was not reflected in the counting

efficiencies, likely because the efficiency equation was heavily influenced by abundance of rare taxa (Fig. 2d-e). In studies where species richness is potentially important, such as O/E models, quantifying the relative abundance of dominant taxa at the expense of recognizing the presence of less abundant taxa could negatively impact the quality of results, while the stratified method should be further explored for use in studies where relative abundance is important.

The stratified counting method significantly outperformed the traditional counting method in characterizing species richness for the low and medium diversity sites, detecting 53% more species in the low diversity site and 36% more species in the high diversity site (Fig. 3). The stratified method found 5% fewer species than the traditional method in the high diversity site, likely because this site was not dominated by any one taxon and thus did not suffer from mischaracterization by the traditional method. However, the timed presence method captured more species than the traditional method in all three test sites. When applied to all 68 sites, the timed presence method found 91 taxa not found in corresponding traditional counts. Only 48% of these additional taxa were spatially rare, indicating 52% of taxa uncovered with our timed presence method provided taxa signals that would be included in most models and provide potentially valuable information for bioassessment.

Our O/E models built using data collected with the fixed method compared to the timed presence method supported our conclusion that the timed presence method collected valuable information missed by the fixed method. Our O/E models built with traditionally counted data performed worse than their null models, indicating these models failed to account for influence of natural environmental variation on diatom community composition among reference sites.

Surprisingly, the genus-resolution model performed better than the species-resolution model using the timed presence method (Table 1), which should be considered with caution given species within a genus are known to many times have different environmental preferences (Kelly et al. 2014). When we examined which species composed the biological groups designated during model construction, we found that tolerant, widespread taxa dominated the groups with a high frequency of occurrence among sites within groups. In our most extreme example, *Achnanthidium minutissimum* occurred with 100% frequency in all samples among all groups. As a result, most groups were biologically distinguished from one another using only slight differences in the frequency with which these tolerant species occurred among sites within a group. It is possible more environmentally sensitive taxa could better distinguish groups if their signal was not lost in the signal of widespread, tolerant taxa. An important direction of future research is thus to test model performance after removing taxa that occur in most sites. Another possibility is the data was muddled by inconsistent species-level identifications. Others have documented the difficulties assigning individuals to cryptic species using light microscopy, with some species only distinguishable using scanning electron microscopy (Morales 2001, Morales et al. 2001, Morales and Hamilton 2002, Potapova and Hamilton 2007).

CONCLUSIONS

The traditional fixed count method characterized reference sites unequally due to varying diversity among sites as measured by both efficiency, which considers relative abundance of dominant taxa, and by the greater richness of species found using the stratified and timed presence methods. 69% of additional species found using timed presence were not

spatially rare and could provide valuable information for bioassessment, particularly O/E models.

We suggest an alternative counting method is needed in diatom assessment to better consistently identify the presence of less dominant species. In multimetric indices, which rely on both relative abundance and richness data, a stratified method could improve model accuracy. We concluded that efficiency is not a good metric by which to standardize stratified counts, so other types of standardization need to be explored. In predictive models that rely on species richness and do not require abundance data, such as O/E models, a timed presence method could better identify taxa presence than the traditional or stratified methods and with a considerably lower cost.

REFERENCES

ACKER, F., B. RUSSELL, AND E. HAGAN. 1999a. Cleaning by Nitric Acid Digestion with a Microwave Apparatus. Protocol P-13-42.

ACKER, F., B. RUSSELL, AND E. MORALES. 1999b. Preparation of Diatom Slides Using Naphrax ™ Mounting Medium. Protocol P-13-49.

ALBEMARLE-PAMLICO NATIONAL ESTUARY PARTNERSHIP. 2015. Fast Facts about the Albemarle-Pamlico Sounds and Watershed. http://www.apnep.org/web/apnep/fastfacts#1.

ANDREEN, W. L. 2013. Success and Backlash: The Remarkable (Continuing) Story of the Clean Water Act. George Washington Journal of Energy & Environmental Law 4:25–37.

BACKER, L. C., S. V MCNEEL, T. BARBER, B. KIRKPATRICK, C. WILLIAMS, M. IRVIN, Y. ZHOU, T. B. JOHNSON, K. NIERENBERG, M. AUBEL, R. LEPRELL, A. CHAPMAN, A. FOSS, S. CORUM, V. R. HILL, S. M. KIESZAK, AND Y. CHENG. 2010. Recreational exposure to microcystins during algal blooms in two California lakes. Toxicon 55:909–921.

BAILEY, R. C., R. H. NORRIS, AND T. B. REYNOLDSON. 2004. Bioassessment of freshwater ecosystems: using the reference condition approach. Kluwer Academic Publishers, Norwell, Massachusetts.

BATTARBEE, R. W. 1986. Diatom Analysis. Pages 527–570*in* B. E. Berglund (editor).Handbook of Holocene Palaeoecology and Palaeohydrology. John Wiley & Sons Ltd., New York, NY.

BRADEN, J. B., M. C. JOLEJOLE-FOREMAN, AND D. W. SCHNEIDER. 2014. Humans and the water environment: The need for coordinated data collection. Water (Switzerland) 6:1–16.

BRADLEY, P. M., AND R. B. LANDY. 2000. The Mid-Atlantic Integrated Assessment (MAIA). Environmental Monitoring and Assessment 63:1–13.

CAO, Y., C. P. HAWKINS, J. OLSON, AND M. A. KOSTERMAN. 2007. Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. Journal of the North American Benthological Society 26:566–585.

CAO, Y., D. P. LARSEN, AND R. S.-J. THORNE. 2001. Rare species in multivariate analysis for bioassessment: some considerations. Journal of the North American Benthological Society 20:144–153.

CARLISLE, D. M., C. P. HAWKINS, M. R. MEADOR, M. POTAPOVA, AND J. FALCONE. 2008. Biological assessments of Appalachian streams based on predictive models for fish, macroinvertebrate, and diatom assemblages. Journal of the North American Benthological Society 27:16–37.

CHARLES, D. F., C. KNOWLES, AND R. S. DAVIS. 2002. Protocols for the analysis of algal samples collected as part of the U.S. Geological Survey National Water-Quality Assessment Program. Page Report No. 02-06. Philadelphia, PA.

CHESAPEAKE BAY FOUNDATION. 2012. The Economic Argument for Cleaning Up the Chesapeake Bay and its Rivers.

DORIGO, U., A. BERARD, F. RIMET, A. BOUCHEZ, AND B. MONTUELLE. 2010. In situ assessment of periphyton recovery in a river contaminated by pesticides. Aquatic Toxicology 98:396–406.

DUNCK, B., E. LIMA-FERNANDES, F. CASSIO, A. CUNHA, L. RODRIGUES, AND C. PASCOAL. 2015. Responses of primary production, leaf litter decomposition and associated communities to stream eutrophication. Environmental Pollution 202:32–40.

FALCONER, I. R. 2001. Toxic cyanobacterial bloom problems in Australian waters: Risks and impacts on human health. Phycologia 40:228–233.

FOWLER, L. 2014. Assessing the Framework of Policy Outcomes: The Case of the U.S. Clean Air Act and Clean Water Act. Journal of Environmental Assessment Policy and Management 16:1–20.

GILLETT, N. D., Y. PAN, K. M. MANOYLOV, R. STANCHEVA, AND C. L. WEILHOEFER. 2011. The potential indicator value of rare taxa richness in diatom-based stream bioassessment. Journal of Phycology 47:471–482.

GILLIOM, R. J., W. M. ALLEY, AND M. E. GURTZ. 1995. Design of the national water-quality assessment program: occurrence and distribution of water-quality conditions. U.S. Geological Survey circular 1112.

GRAF, W. L. 1999. Dam nation: A geographic census of american dams and their large-scale hydrologic impacts. Water Resources Research 35:1305–1311.

GREGORY, M., AND D. CALHOUN. 2006. Physical, Chemical, and Biological Responses of Streams to Increasing Watershed Urbanization in the Piedmont Ecoregion of Georgia and Alabama, 2003. Page U.S. Geological Survey Scientific Investigations Report 2006-5101-B.

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.

GULIS, V., V. FERREIRA, AND M. A. S. GRAÇA. 2006. Stimulation of leaf litter decomposition and associated fungi and invertebrates by moderate eutrophication: implications for stream assessment. Freshwater Biological Association 51:1655–1669.

H. JOHN HEINZ III CENTER FOR SCIENCE ECONOMICS AND THE ENVIRONMENT. 2002. The state of the nation's ecosystems: measuring the lands, waters, and living resources of the United States. Cambridge University Press.

HAASE, P., S. U. PAULS, K. SCHINDEHÜTTE, AND A. SUNDERMANN. 2010. First audit of macroinvertebrate samples from an EU Water Framework Directive monitoring program : human error greatly lowers precision of assessment results. Journal of the North American Benthological Society 29:1279–1291.

HADLEY, J., AND C. WIEGAND. 2014. A Social and Economic Analysis of Commercial Fisheries in North Carolina: Albemarle and Pamlico Sounds.

HAMILTON, D. P., S. A. WOOD, D. R. DIETRICH, AND J. PUDDICK. 2014. Costs of harmful blooms of freshwater cyanobacteria. Pages 247–256*in* N. K. Sharma, A. K. Rai, and L. J. Stal (editors). Cyanobacteria: An Economic Perspective. Wiley Blackwell, West Sussex, UK.

HAWKINS, C. P., Y. CAO, AND B. ROPER. 2010. Method of predicting reference condition biota affects the performance and interpretation of ecological indices. Freshwater Biology 55:1066–1085.

HAWKINS, C. P., R. H. NORRIS, J. N. HOGUE, AND J. W. FEMINELLA. 2000. Development and Evaluation of Predictive Models for Measuring the Biological Integrity of Streams. Ecological Applications 10:1456–1477.

HENLEY, W. F., M. A. PATTERSON, R. J. NEVES, AND A. D. LEMLY. 2000. Effects of Sedimentation and Turbidity on Lotic Food Webs: A Concise Review for Natural Resource Managers. Reviews in Fisheries Science 8:125–139.

HERING, D., R. K. JOHNSON, S. KRAMM, S. SCHMUTZ, K. SZOSZKIEWICZ, AND P. F. M. VERDONSCHOT. 2006. Assessment of European streams with diatoms, macrophytes, macroinvertebrates and fish: A comparative metric-based analysis of organism response to stress. Freshwater Biology 51:1757–1785.

HILL, B. H., R. J. STEVENSON, Y. PAN, A. T. HERLIHY, P. R. KAUFMANN, AND C. B. JOHNSON. 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.

HILLEBRAND, H., AND U. SOMMER. 2000. Diversity of benthic microalgae in response to colonization time and eutrophication. Aquatic Botany 67:221–236.

HOAGLAND, A. P., D. M. ANDERSON, Y. KAORU, AND A. W. WHITE. 2002. The Economic Effects of Harmful Algal Blooms in the United States : Estimates , Assessment Issues , and Information Needs Source : Estuaries , Vol . 25 , No . 4 , Part B : Dedicated Issue : Nutrient

Over-Enrichment in Coastal Waters : Global Patterns of C. Estuaries 25:819–837.

HUGHES, R. M., AND D. V. PECK. 2008. Acquiring data for large aquatic resource surveys: the art of compromise among science, logistics, and reality. Journal of the North American Benthological Society 27:837–859.

IHO, A., M. RIBAUDO, AND K. HYYTIÄINEN. 2015. Water protection in the Baltic Sea and the Chesapeake Bay: Institutions, policies and efficiency. Marine Pollution Bulletin 93:81–93.

JOURNEY, C. A., P. C. VAN METRE, A. H. BELL, J. D. GARRETT, D. T. BUTTON, N. NAKAGAKI, S. L. QI, AND P. M. BRADLEY. 2014. Design and Methods of the Southeast Stream Quality Assessment (SESQA).

KARR, J. R., AND E. W. CHU. 1999. Restoring life in running waters: better biological monitoring. Island Press, Washington, D.C.

KELLY, M. 2013. Data rich, information poor? Phytobenthos assessment and the Water Framework Directive. European Journal of Phycology 48:437–450.

KELLY, M. G., R. TROBAJO, L. ROVIRA, AND D. G. MANN. 2014. Characterizing the niches of two very similar Nitzschia species and implications for ecological assessment. Diatom Research:1–7.

KELLY, M., S. JUGGINS, R. GUTHRIE, S. PRITCHARD, J. JAMIESON, B. RIPPEY, H. HIRST, AND M. YALLOP. 2008a. Assessment of ecological status in U.K. rivers using diatoms. Freshwater Biology 53:403–422.

KELLY, M., S. JUGGINS, R. GUTHRIE, S. PRITCHARD, J. JAMIESON, B. RIPPEY, H. HIRST, AND M. YALLP. 2008b. Assessment of ecological status in U.K. rivers using diatoms. Freshwater Biology 53:403–422.

LIESS, M., AND P. C. VON DER OHE. 2005. Analyzing effects of pesticides on invertebrate communities in streams. Environmental Toxicology and Chemistry 24:954–965.

MAGURRAN, A. E. 2004. The commonness, and rarity, of species. Pages 18–71Measuring Biological Diversity. Blackwell Publishing.

MORALES, E. A. 2001. Morphological Studies in Selected Fragilarioid Diatoms (Bacillariophyceae) from Connecticut Waters (U.S.A). Proceedings of the Academy of Natural Sciences of Philadelphia2 151:105–120.

MORALES, E. A., P. A. SIVER, AND F. R. TRAINOR. 2001. Identification of Diatoms (Bacillariophyceae) during Ecological Assessments: Comparison between Light Microscopy and Scanning Electron Microscopy. Proceedings of the National Academy of Sciences of the United States of America 151:95–103.

MORALES, E., AND P. B. HAMILTON. 2002. Seventh NAWQA Workshop on Harmonization of Algal Taxonomy. Page Report No. 02-10. Philadelphia, PA.

MOULTON, S.R., I., J. G. KENNEN, R. M. GOLDSTEIN, AND J. A. HAMBROOK. 2002. Revised Protocols for Sampling Algal , Invertebrate , and Fish Communities as Part of the National Water-Quality Assessment Program. U.S. Geological Survey Open-File Report 02-150.

NATIONAL AGRICULTURAL STATISTICS SERVICE. 2017. Statistics by state. https://www.nass.usda.gov/Statistics_by_State/.

NATIONAL RESEARCH COUNCIL. 1977. Environmental Monitoring. Washington, DC: The National Academies Press.

NODINE, E. R., AND E. E. GAISER. 2014. Distribution of Diatoms Along Environmental Gradients in the Charlotte Harbor, Florida (USA), Estuary and Its Watershed: Implications for Bioassessment of Salinity and Nutrient Concentrations. Estuaries and Coasts 37:864–879.

O'DRISCOLL, M., S. CLINTON, A. JEFFERSON, A. MANDA, AND S. MCMILLAN. 2010. Urbanization Effects on Watershed Hydrology and In-Stream Processes in the Southern United States. Water 2:605–648.

PAERL, H. W., R. S. FULTON, P. H. MOISANDER, AND J. DYBLE. 2001. Harmful freshwater algal blooms, with an emphasis on cyanobacteria. The Scientific World 1:76–113.

PAN, Y. D., R. J. A N. STEVENSON, B. H. HILL, A. T. HERLIHY, AND G. B. COLLINS. 1996. Using diatoms as indicators of ecological conditions in lotic systems: A regional assessment. Journal of the North American Benthological Society 15:481–495.

PAPPAS, J. L., AND E. F. STOERMER. 1996. Quantitative method for determining a representative algal sample count. Journal of Phycology 32:693–696.

PATRICK, R. 1973. Use of Algae, Especially Diatoms, in the Assessment of Water Quality. Pages 76–95Biological Methods for the Assessment of Water Quality. American Society for Testing and Materials.

PATRICK, R., M. H. HOHN, AND J. H. WALLACE. 1954. A new method for determining the pattern of the diatom flora. Notulae Naturae 259:1–12.

PAULSEN, S. G., A. MAYIO, D. V. PECK, J. L. STODDARD, E. TARQUINIO, S. M. HOLDSWORTH, J. VAN SICKLE, L. L. YUAN, C. P. HAWKINS, A. T. HERLIHY, P. R. KAUFMANN, M. T. BARBOUR, D. P. LARSEN, AND A. R. OLSEN. 2008. Condition of stream ecosystems in the US: an overview of the first national assessment. Journal of the North American Benthological Society 27:812–821.

POOS, M. S., AND D. A. JACKSON. 2012. Addressing the removal of rare species in multivariate

bioassessments : The impact of methodological choices. Ecological Indicators 18:82–90.

POTAPOVA, B. M., AND D. M. CARLISLE. 2011. Development and application of indices to assess the condition of benthic algal communities in U.S. streams and rivers. Page U.S. Geological Survey Open-File Report 2011-1126, 40 p.

POTAPOVA, B. M. G., R. J. BIXBY, D. F. CHARLES, M. B. EDLUND, M. E. ENACHE, P. B. HAMILTON, R. L. LOWE, K. M. MANOYLOV, K. C. PONADER, L. REN, P. A. SIVER, S. A. SPAULDING, AND J. ZALACK. 2008. Eighteenth NAWQA Workshop on Harmonization of Algal Taxonomy Representatives of the genus Aulacoseira Thwaites in NAWQA samples. Philadelphia, PA.

POTAPOVA, M., AND D. F. CHARLES. 2005. Diatom metrics for monitoring eutrophication in rivers of the United States. Ecological Indicators 7:48–70.

POTAPOVA, M. G., AND D. F. CHARLES. 2002. Benthic Diatoms in USA Rivers: Distributions along Spatial and Environmental Gradients. Journal of Biogeography 29:167–187.

POTAPOVA, M., AND P. B. HAMILTON. 2007. Morphological and ecological variation within the Achnanthidium minutissimum (Bacillariophyceae) species complex. Journal of Phycology 43:561–575.

PRYGIEL, J., P. CARPENTIER, S. ALMEIDA, M. COSTE, J. DRUART, L. ECTOR, D. GUILLARD, M. HONORÉ, R. ISERENTANT, P. LEDEGANCK, C. LALANNE-CASSOU, C. LESNIAK, I. MERCIER, P. MONCAUT, M. NAZART, N. NOUCHET, F. PERES, V. PEETERS, AND F. RIMET. 2002. Determination of the biological diatom index (IBD NF T90 – 354): results of an intercomparison exercise. Journal of Applied Phycology 14:27–39.

REHN, A. C., AND P. R. ODE. 2009. Synthesis Report: Integrating probability and targeted survey designs in reigonal stream condition assessments with examples from southern coastal California.

REYNOLDSON, T. B., R. H. NORRIS, V. H. RESH, K. E. DAY, D. M. ROSENBERG, S. JOURNAL, N. AMERICAN, B. SOCIETY, AND N. DEC. 2016. The Reference Condition : A Comparison of Multimetric and Multivariate Approaches to Assess Water-Quality Impairment Using Benthic Macroinvertebrates. Journal of the North American Benthological Society 16:833–852.

RIMET, F., AND A. BOUCHEZ. 2012. Biomonitoring river diatoms: Implications of taxonomic resolution. Ecological Indicators 15:92–99.

RITZ, C. 2010. Capstone Projects and Theses Effectiveness of RIVPACS predictive models to evaluate diatom response to nutrient stress in coastal California streams. California State University, Monterey Bay.

SATZ, D., R. K. GOULD, K. M. A. CHAN, A. GUERRY, B. NORTON, T. SATTERFIELD, B. S. HALPERN, J. LEVINE, U. WOODSIDE, N. HANNAHS, X. BASURTO, AND S. KLAIN. 2013. The Challenges of Incorporating

Cultural Ecosystem Services into Environmental Assessment. Ambio 42:675–684.

SCHÄFER, R. B., M. BUNDSCHUH, D. A. ROUCH, E. SZÖCS, P. C. VON DER OHE, V. PETTIGROVE, R. SCHULZ, D. NUGEGODA, AND B. J. KEFFORD. 2012. Effects of pesticide toxicity, salinity and other environmental variables on selected ecosystem functions in streams and the relevance for ecosystem services. The Science of the Total Environment 415:69–78.

SCOTT, M. C., G. S. HELFMAN, M. E. MCTAMMANY, E. F. BENFIELD, AND P. V BOLSTAD. 2002. Multiscale Influences on Physical and Chemical Stream Conditions Across Blue Ridge Landscapes. Journal Of The American Water Resources Association 38:1379–1392.

SHAPIRO, M. H., S. M. HOLDSWORTH, AND S. G. PAULSEN. 2008. The need to assess the condition of aquatic resources in the US. Journal of the North American Benthological Society 27:808–811.

VAN SICKLE, J., C. P. HAWKINS, D. P. LARSEN, AND A. T. HERLIHY. 2005. A null model for the expected macroinvertebrate assemblage in streams. Journal of the North American Benthological Society 24:178–191.

SKLENAR, K., J. WESTRICK, AND D. SZLAG. 2016. Managing Cyanotoxins in Drinking Water: A Technical Guidance Manual for Drinking Water Professionals Managing Cyanotoxins in Drinking Water.

SMUCKER, N. J., M. BECKER, N. E. DETENBECK, AND A. C. MORRISON. 2013. Using algal metrics and biomass to evaluate multiple ways of defining concentration-based nutrient criteria in streams and their ecological relevance. Ecological Indicators 32:51–61.

SPAULDING, S. A., D. M. MCKNIGHT, E. F. STOERMER, AND P. T. DORAN. 1997. Diatoms in sediments of perennially ice-covered Lake Hoare, and implications for interpreting lake history in the McMurdo Dry Valleys of Antarctica. Journal of Paleolimnology 17:403–420.

STEVENSON, J. 1998. Diatom indicators of stream and wetland stressors in a risk management framework. Environmental Monitoring and Assessment 51:107–118.

STEVENSON, J. R., Y. PAN, AND H. VAN DAM. 2010. Assessing environmental conditions in rivers and streams with diatoms. Pages 57–85*in* J. P. Smol and E. F. Stoermer (editors).The Diatoms: Applications for the Environmental and Earth Sciences, 2nd edition. Cambridge University Press.

STEVENSON, R. J., Y. PAN, K. M. MANOYLOV, C. A PARKER, D. P. LARSEN, AND A. T. HERLIHY. 2008. Development of diatom indicators of ecological conditions for streams of the western US. Journal of the North American Benthological Society 27:1000–1016.

STODDARD, J. L., D. P. LARSEN, C. P. HAWKINS, R. K. JOHNSON, AND R. H. NORRIS. 2006. Setting

expectations for the ecological condition of streams: The concept of reference condition. Ecological Applications 16:1267–1276.

STOERMER, E. F., AND J. P. SMOL. 2010. The diatoms: applications for the environmental and earth sciences, 2nd edition. Cambridge University Press, Cambridge, U.K.

STRIBLING, J. B., K. L. PAVLIK, S. M. HOLDSWORTH, AND E. W. LEPPO. 2008. Data quality , performance , and uncertainty in taxonomic identification for biological assessments Published by : The University of Chicago Press on behalf of the Society for Freshwater Science Stable URL : http://www.jstor.org/stable/10.1899/07-175.1 Da. Journal of the North American Benthological Society 27:906–919.

SURASINGHE, T., AND R. F. BALDWIN. 2014. Ghost of land-use past in the context of current land cover: evidence from salamander communities in streams of Blue Ridge and Piedmont ecoregions. Canadian Journal of Zoology 92:527–536.

TRIMBLE, S. W. 1974. Man-induced soil erosion on the southern Piedmont, 1700–1970. Soil and Water Conservation Society, Ankeny, IA.

TYREE, M., N. CLAY, S. POLASKEY, AND S. ENTREKIN. 2016. Salt in our streams: even small sodium additions can have negative effects on detritivores. Hydrobiologia 775:109-122.

UNITED STATES GENERAL ACCOUNTING OFFICE. 2000. Water Quality: Key EPA and State Decisions Limited by Inconsistent and Incomplete Data.

USEPA (UNITED STATES ENVIRONMENTAL PROTECTION AGENCY). 1972. Clean Water Act.

USEPA (UNITED STATES ENVIRONMENTAL PROTECTION AGENCY). 2001. Environmental Monitoring and Assessment Program: West - Research Strategy.

USEPA (U.S. ENVIRONMENTAL PROTECTION AGENCY). 2009. National rivers and streams assessment: laboratory methods manual. EPA-841-B-07-010

.

USEPA (UNITED STATES ENVIRONMENTAL PROTECTION AGENCY). 2013. National Rivers and Streams Assessment 2008-2009: A Collaborative Survey.

USEPA (U.S. ENVIRONMENTAL PROTECTION AGENCY). 2016. National Rivers and Streams Assessment 2008 – 2009 Technical Report (EPA/841/R-16/008). Washington, D.C.

USGS (UNITED STATES GEOLOGICAL SURVEY). 2014. Fact Sheet 2014-3023, The Southeast Stream Quality Assessment.

USGS (UNITED STATES GEOLOGICAL SURVEY). 2017. Regional Stream Quality Assessments. https://txpub.usgs.gov/RSQA/Default.aspx.

VESTERINEN, J., E. POUTA, A. HUHTALA, AND M. NEUVONEN. 2010. Impacts of changes in water quality recreation behavior and benefits in Finland. Journal of Environmental Management 91:984–994.

WEIJTERS, M. J., J. H. JANSE, R. ALKEMADE, AND J. T. . . VERHOEVEN. 2009. Quantifying the effect of catchment land use and water nutrient concentrations on freshwater river and stream biodiversity. Aquatic Conservation: Marine and Freshwater Ecosystems 19:104–112.

WHITEHEAD, J. C., T. C. HAAB, AND J. C. HUANG. 2000. Measuring recreation benefits of quality improvements with revealed and stated behavior data. Resource and Energy Economics 22:339–354.

WRIGHT, J. F., P. D. ARMITAGE, M. T. FURSE, AND D. MOSS. 1989. Prediction of invertebrate communities using stream measurements. Regulated Rivers: Research and Management 4:147–155.

APPENDIX

Mean O/E scores, standard deviation (SD), and replicate standard deviation (Rep SD) for calibration sites using probability of capture (P_c) thresholds of 0, 0.25, 0.5, and 0.7. Difference between null model SD and model SD (Null SD – Model SD) measures the relative ability of the model to account for environmental variation.

	Mean O/E	SD	Rep SD	Null Mean O/E	Null SD	Null SD - Model SD
P_c = 0						
Fixed: Species	1.011	0.446	0.138	1.000	0.460	0.014
Fixed: Genus	1.010	0.385	0.159	1.000	0.347	-0.038
Timed P – Species	1.007	0.269	0.088	1.000	0.279	0.010
Timed P – Genus	1.011	0.242	0.109	1.000	0.226	-0.016
P_c = 0.25						
Fixed: Species	0.977	0.305	0.165	1.000	0.360	0.055
Fixed: Genus	0.987	0.246	0.155	1.000	0.266	0.020
Timed P – Species	1.030	0.226	0.094	1.000	0.238	0.012
Timed P – Genus	1.004	0.187	0.105	1.000	0.205	0.018
P_c = 0.5						
Fixed: Species	0.975	0.280	0.193	1.000	0.257	-0.023
Fixed: Genus	1.010	0.206	0.144	1.000	0.202	-0.004
Timed P – Species	1.031	0.194	0.105	1.000	0.205	0.011
Timed P – Genus	1.010	0.147	0.095	1.000	0.172	0.025
P_c = 0.7						
Fixed: Species	1.002	0.236	0.196	1.000	0.269	0.033
Fixed: Genus	1.018	0.153	0.127	1.000	0.173	0.020
Timed P: Species	1.024	0.167	0.111	1.000	0.164	-0.003
Timed P: Genus	1.017	0.109	0.081	1.000	0.110	0.001