

Utah State University

DigitalCommons@USU

---

Ecology Center Publications

Ecology Center

---

2-13-2020

## Reduction of Taxonomic Bias in Diatom Species Data

Meredith A. Tyree

*University of Colorado, Boulder*

Ian W. Bishop

*University of Rhode Island*

Charles P. Hawkins

*Utah State University*

Richard Mitchell

*United States Environmental Protection Agency*

Sarah A. Spaulding

*University of Colorado, Boulder*

Follow this and additional works at: [https://digitalcommons.usu.edu/eco\\_pubs](https://digitalcommons.usu.edu/eco_pubs)

 Part of the [Ecology and Evolutionary Biology Commons](#)

---

### Recommended Citation

Tyree, M.A., Bishop, I.W., Hawkins, C.P., Mitchell, R. and Spaulding, S.A. (2020), Reduction of taxonomic bias in diatom species data. *Limnol Oceanogr Methods*. doi:10.1002/lom3.10350

This Article is brought to you for free and open access by the Ecology Center at DigitalCommons@USU. It has been accepted for inclusion in Ecology Center Publications by an authorized administrator of DigitalCommons@USU. For more information, please contact [digitalcommons@usu.edu](mailto:digitalcommons@usu.edu).



## Reduction of taxonomic bias in diatom species data

Meredith A. Tyree <sup>1,\*</sup> Ian W. Bishop,<sup>2</sup> Charles P. Hawkins,<sup>3</sup> Richard Mitchell,<sup>4</sup> Sarah A. Spaulding<sup>1,5</sup>

<sup>1</sup>Institute of Arctic and Alpine Research (INSTAAR), University of Colorado Boulder, Boulder, Colorado, USA

<sup>2</sup>Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island, USA

<sup>3</sup>Department of Watershed Sciences, Western Center for Monitoring and Assessment of Freshwater Ecosystems, The Ecology Center, Utah State University, Logan, Utah, USA

<sup>4</sup>United States Environmental Protection Agency, Office of Water, Office of Wetlands, Oceans, and Watersheds, Washington, District of Columbia, USA

<sup>5</sup>United States Geological Survey, University of Colorado Boulder, Boulder, Colorado, USA

### Abstract

Inconsistency in taxonomic identification and analyst bias impede the effective use of diatom data in regional and national stream and lake surveys. In this study, we evaluated the effect of existing protocols and a revised protocol on the precision of diatom species counts. The revised protocol adjusts four elements of sample preparation, taxon identification and enumeration, and quality control (QC). We used six independent data sets to assess the effect of the adjustments on analytical outcomes. The first data set was produced by three laboratories with a total of five analysts following established protocols (Charles et al., Protocols for the analysis of algal samples collected as part of the U.S. Geological Survey National Water-Quality Assessment, 2002) or their slight variations. The remaining data sets were produced by one to three laboratories with a total of two to three analysts following a revised protocol. The revised protocol included the following modifications: (1) development of coordinated precount voucher floras based on morphological operational taxonomic units, (2) random assignment of samples to analysts, (3) postcount identification and documentation of taxa (as opposed to an approach in which analysts assign names while they enumerate), and (4) increased use of QC samples. The revised protocol reduced taxonomic bias, as measured by reduction in analyst signal, and improved similarity among QC samples. Reduced taxonomic bias improves the performance of biological assessments, facilitates transparency across studies, and refines estimates of diatom species distributions.

Over the past 30 years, the use of diatoms as indicators of biotic condition has increased because of the value of diatom species composition as an important ecological endpoint (Stoermer and Smol 2010). In the United States, federal programs have adopted diatom indicators to complement assessments based on aquatic invertebrates and fish. Many federal surveys have a large geographic extent, covering regional and continental scales (Pan et al. 1996; Potapova and Charles 2007), and can include thousands of samples. Taxonomic expertise is often spread across several analysts and laboratories. Unfortunately, taxonomic data sets produced by different laboratories have not been taxonomically comparable with one another within studies (Lee et al. 2019), which compromises assessments (Cao et al. 2007). The lack of taxonomic consistency in

freshwater diatom data is a serious issue (Kelly et al. 2009; Kahlert et al. 2016; Werner et al. 2016), and the European Water Framework Directive has initiated cross-analyst comparisons to quantify the magnitude of the problem (Besse-Lototskaya et al. 2006, summarized in Kahlert et al. 2016). Here, we consider aspects commonly shared among protocols used in the United States in support of diatom assessments and recommend revisions to those protocols for ensuring taxonomically consistent and verifiable diatom species data.

Lack of consistency in taxonomic identification of diatoms stems, in part, from a reliance on European floras and the resulting misapplication of names to North American species (Kocielek and Spaulding 2000). Analysts in different laboratories often use different taxonomic references, which results in analysts arriving at different species names for the same or similar specimens. In an Idaho study, Cao et al. (2007) assembled diatom taxonomic data from 256 reference-quality sites to develop a state-wide diatom index. The species names used by the three laboratories, however, could not be reconciled or harmonized. As a result, data from only one laboratory could

\*Correspondence: meredithtyree@gmail.com

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

be used (149 samples), reducing the robustness of the final index. Following recognition of incompatibility in taxonomic identification in the Idaho study, other studies were found to suffer from significant “analyst bias” (Lee et al. 2019).

Determining the source of inconsistency in taxonomic identification can be confounded by the species-specific geographic distributions of diatoms. In past national surveys, samples were intentionally assigned to analysts on a geographic basis. For example, one laboratory might receive samples from the Northeast United States, whereas another laboratory received samples from another region. However, when analysts working in different regions subscribe to different species concepts and use different taxonomic guides, “analyst effect” can be confounded with diatom geographic distributions. Post hoc taxonomic harmonization has been advocated to correct taxonomic inconsistencies (Manoylov 2014), but the process is not only time-consuming, but flawed. Post hoc corrections result in reduced detection of species diversity because many species are combined into higher-level taxonomic groups in an attempt to reduce “analyst signal” (Lee et al. 2019). As a result, bias in the application of diatom names continues to compromise the use of diatom data in biological assessments, particularly at regional and national scales.

In this study, we examined six diatom data sets to quantify variation in analysts’ abilities to recognize species (as morphological operational taxonomic units [mOTUs]). Based on our results, we developed revisions to common aspects of processing protocols to produce consistent, verifiable diatom species data. The revised protocol reduced taxonomic bias (i.e., analyst signal) and improved similarity among enumeration of replicate samples.

## Materials and procedures

### Diatom taxonomic data sets

Diatom relative abundance data from six recent surveys of periphyton in streams and lakes were examined in this study. One survey was collected as a joint effort between the U.S. Geological Survey (USGS) and the U.S. Environmental Protection Agency (USEPA): the 2013 Midwest Stream Quality Assessment (MSQA) (Van Metre et al. 2012; Garrett et al. 2017). Four studies were collected as part of the USGS National Water

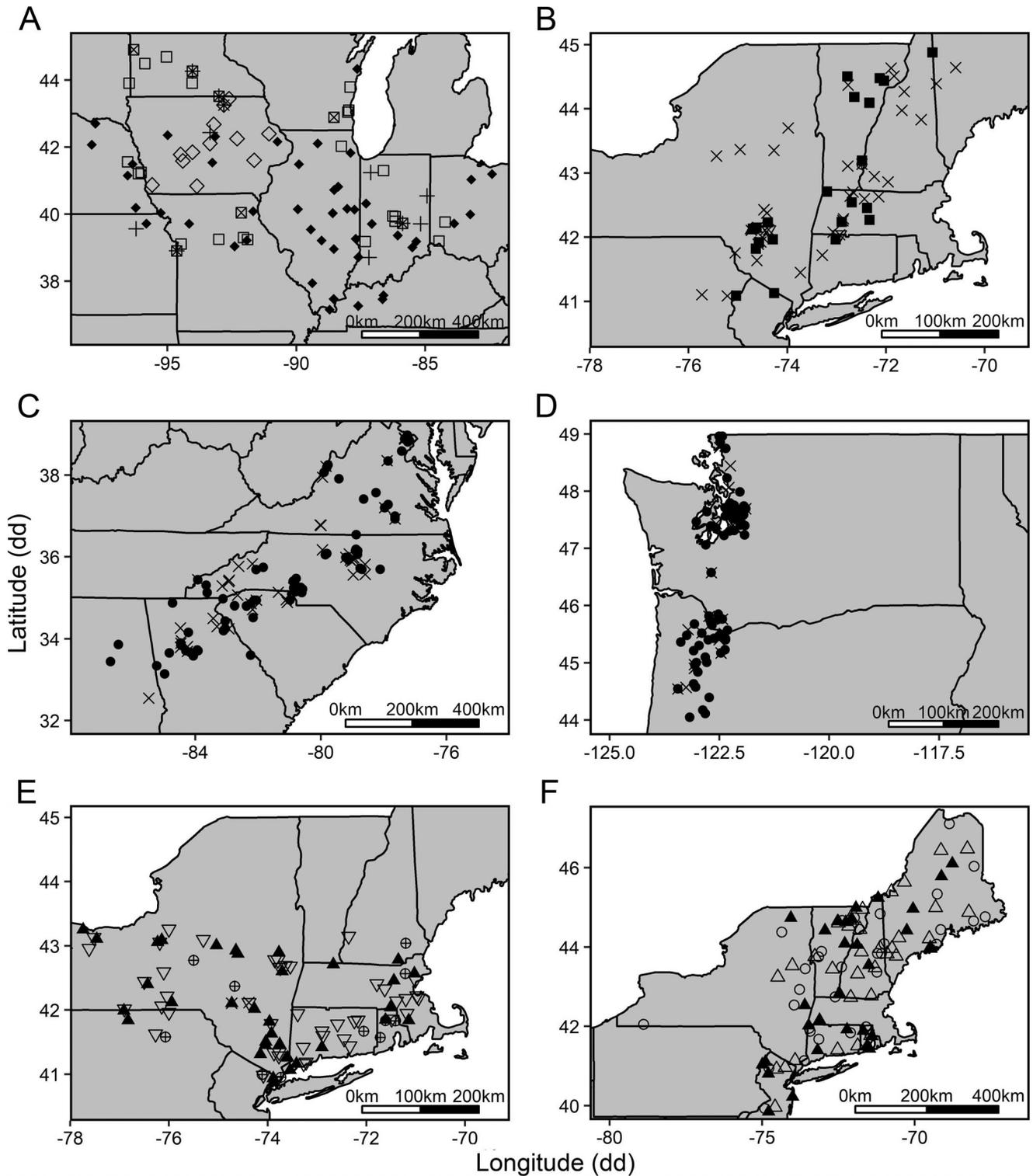
Quality Assessment program: the 2014 Northeast Stream Quality Assessment (NESQA 2014) (Van Metre et al. 2016; Coles et al. 2018), the 2014 Southeast Stream Quality Assessment (SESQA) (Van Metre and Journey 2014; Journey et al. 2015), the 2015 Pacific Northwest Stream Quality Assessment (PNSQA) (Van Metre et al. 2015; Sheibley et al. 2017), and the NESQA 2016 (Van Metre et al. 2016; Coles et al. 2018) (Table 1). Finally, a 2017 regional survey of lake sediments from northeast states, termed the 2017 Northeast Lake Sediment Diatom Voucher Flora project (NE Lakes), was examined (Fig. 1). Between 60 and 107 sites are represented in each study, and 80–140 taxonomic analyses of 600 valves were collected (Table 1). The geographic extent of each survey is roughly equivalent, extending across two to four Level III ecoregions (Omernik and Griffith 2014).

Strewn slides (Charles et al. 2002) were prepared for the MSQA study. Although slide preparation does not relate directly to taxonomic bias, Battarbee chambers were used to prepare slides for the remaining studies. These chambers were developed to produce quantitative, replicate slides of microfossils for analysis of lake sediments (Battarbee 1973). Evaporative settling chambers, such as Battarbee chambers, allow for even settling of cells resulting in random distribution of diatoms on coverslips. The method reduces edge effects, that is, differences in density and distribution of cells, especially near the margins of the coverslip. Specimens are distributed on the coverslip in a more random distribution than with strewn slides (Battarbee 1973). Furthermore, replicate slides prepared by a single person from Battarbee chambers were shown to have a lower coefficient of variation of cell number than replicate slides prepared by other methods (Wolfe 1997). Strewn preparations using square coverslips are vulnerable to uneven distribution of cells, both from edge effects and by cell sorting caused by vibration. Diatomists are familiar with the “X” pattern that valves form when coverslips are subject to even minimal amounts of vibration while drying.

For the MSQA survey, five analysts in three labs completed diatom identifications and enumeration following Charles et al. (2002), which did not include our proposed revisions. For the other studies, we made modifications to data collection and quality assurance/quality control (QA/QC). The modifications include (1) development of coordinated precount voucher floras with provisional names (Bishop et al. 2017), (2) random

**Table 1.** Descriptive attributes of data sets for each of six regional studies. Studies include the EPA-USGS MSQA, USGS NESQA 2014, USGS SESQA, USGS PNSQA, USGS NESQA 2016, and EPA NE Lakes. The table lists the study name and number of sites, analytical organism counts, replicate sample collections, replicate (self and cross) counts, analysts, and labs.

Study	Sites	Analytical organism counts	Replicate collections	Replicate counts	Analysts	Labs
MSQA	92	98	5	5	5	3
NESQA 2014	60	80	0	19	2	2
SESQA	107	134	5	19	2	1
PNSQA	104	112	8	21	2	1
NESQA 2016	119	125	8	26	3	2
NE Lakes	106	140	6	27	3	3



**Fig. 1.** Geographic distribution of samples for six regional surveys: MSQA (A), NESQA 2014 (B), SESQA (C), PNSQA (D), NESQA 2016 (E), and NE Lakes (F). Symbols represent analysts, carry across surveys, and correspond to symbols used in Fig. 2.

assignment of samples to analysts, (3) postcount identification of taxa (as opposed to most protocols, in which analysts assign names while they enumerate), and (4) implementation of a

“multi-party” QC process with increased collection and analysis of QC data. Samples were analyzed by two to three analysts in one to three laboratories (Table 1).

### Precount voucher floras

For the MSQA survey, diatoms were identified by generally following established protocols (Charles et al. 2002; USEPA 2009). Each analyst was instructed to provide the project leader with a digital image of a representative specimen for all taxa that made up 5% or more of an individual count. In general, these images were not shared among MSQA analysts during the enumeration process.

In contrast, the other five studies used precount voucher floras shared by all analysts during enumeration to coordinate species concepts and naming convention (Bishop et al. 2017). Multiple images were collected to document the size and morphological range of each taxon, which were then assigned a unique mOTU code. The images were assembled into image catalogs, or precount voucher floras, and made available to analysts. Analysts then had the opportunity to comment and reformulate mOTUs and their member images before the formal analysis began. Once analysis of microslides started, newly encountered taxa were photographed, assigned a code, added to the flora, and made available to project analysts for use throughout the analysis period. Analysts could use the voucher flora in either electronic, printed, or both forms during analysis. The SESQA flora was published (Bishop et al. 2017), and the additional floras are available for download (<https://diatoms.org/practitioners/what-is-a-voucher-flora>).

### Random assignment of samples

For the MSQA survey, samples were not randomly assigned to analysts, but rather, sample assignments generally adhered to state boundaries. In contrast, for the other five studies, samples were randomly assigned to analysts, accounting for different workloads assigned to each (Table 1).

### Formal scientific names

For the MSQA survey, analysts assigned scientific or provisional names to diatom valves as they were encountered during microscopic analysis using taxonomic references available within their given laboratory. At the end of the survey, names were finalized by a taxonomic coordinator. In contrast, analysts in the remaining five surveys enumerated taxa based on the mOTU codes from the shared voucher floras. When analyses were completed, mOTU codes were reconciled, and formal scientific names were assigned during collaborative workshops. During the workshops, project analysts, as well as one to three additional diatomists, assigned scientific names to mOTUs. For each scientific name assigned, analysts recorded their initials and the taxonomic concept reference used when identifying the taxon. Any mOTU that did not correspond with a validly published species was referenced by its project-specific code (e.g. "*Achnantheidium* sp. 1 SESQA"). Such names were entered into the U.S. Geological Survey BioData Taxonomic Database (U.S. Geological Survey 2019) as a bench name, with the voucher flora serving as the image representation of the name.

### Quality control

For the MSQA survey, the QC procedure required that 10% of samples be reanalyzed by an analyst outside the primary lab (Charles et al. 2002). These analyses are referred to as "taxonomic harmonization counts." The first analyst used a diamond objective marker (i.e., Zeiss No. 46 29 60) to etch the microslide transects that were examined. Then, the second analyst located the scribed line and enumerated 600 valves following the same transects as the first analyst.

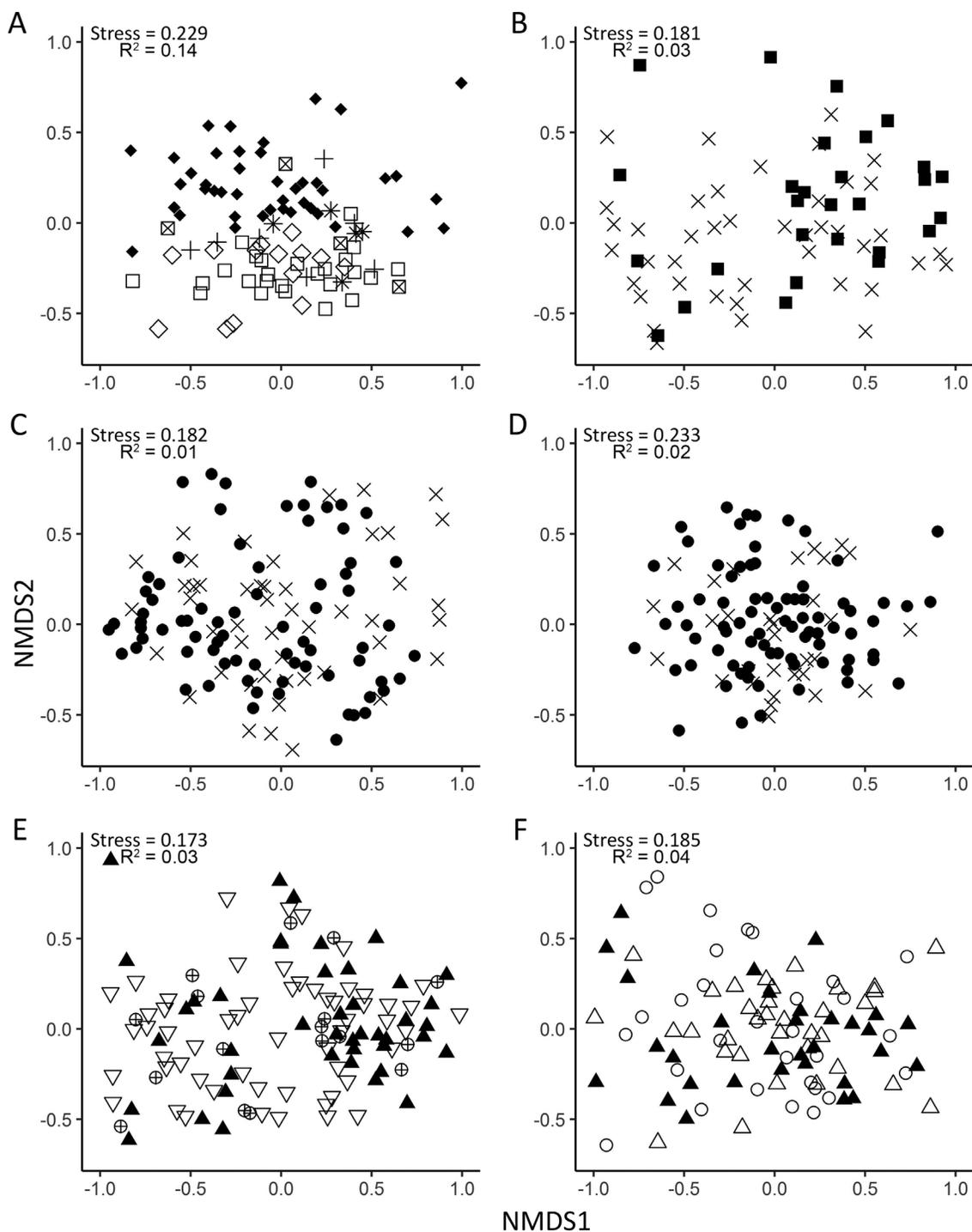
In contrast, for the other five studies, analyses of QC samples were divided among the analysts completing the primary analyses. Each analyst examined a random 10% of the samples that the other analyst(s) had completed. Additionally, analysts repeated analysis of 10% of their own samples (self-QC counts). Thus, a total of 20% of samples were recounted, with each analysis conducted on a replicate slide produced from the same Battarbee chamber.

### Statistical methods

Permutational multivariate ANOVA (PERMANOVAs) were used to test for analyst bias. Data were square root transformed and converted to Bray–Curtis similarity matrices prior to analysis. PERMANOVAs were based on a one-factor, fixed design and type III sums of squares. "Analyst" was the factor, and calculations used unrestricted permutation of raw data, with 9999 permutations. Percent variation explained by each analyst was calculated by dividing the sum of squares of the analyst factor by the total sum of squares to give the coefficient of determination ( $R^2$ ). Nonmetric multidimensional scaling (NMDS) was used to create visual representations of relative differences among analysts. Tests for similarity among replicate samples were based on Bray–Curtis similarity matrices in which Bray–Curtis values were arcsine transformed and evaluated with one-way ANOVA. The relation between sample diversity and QC sample similarity was examined with ranged major axis regression (RMA). RMA is designed to accommodate independent variables estimated with error (e.g., species richness) for data with no outliers (Legendre and Legendre 2012). Data preparation and graphic production was conducted in R (R Core Team 2017) and analyzed in R and Primer 6 (PRIMER-E).

**Table 2.** Results of PERMANOVA analysis for all primary counts for each of the six regional studies.

Study	Pseudo $F$	$R^2$	$p$
MSQA	3.18	0.14	<0.001
NESQA 2014	2.69	0.03	0.006
SESQA	1.75	0.01	0.039
PNSQA	2.04	0.02	0.009
NESQA 2016	1.73	0.03	0.015
NE Lakes	2.13	0.04	<0.001

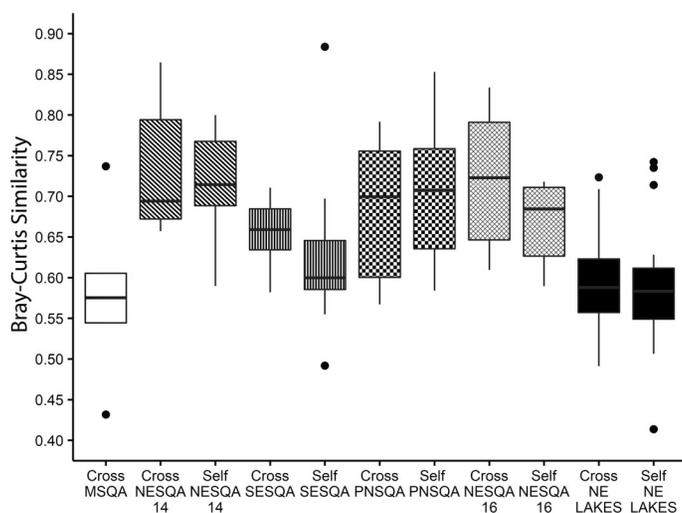


**Fig. 2.** Ordination of diatom species data in NMDS for sites in six regional surveys: MSQA (A), NESQA 2014 (B), SESQA (C), PNSQA (D), NESQA 2016 (E), and NE Lakes (F). Symbols represent analysts, carry across surveys, and correspond to symbols used in Fig. 1.  $R^2$  values represent the amount of variation in the data set attributable to analyst bias.

### Assessment

Analyst was a significant factor ( $p \leq 0.05$ ) in the MSQA study, with the analyst factor explaining 14% of the variation in the data ( $R^2 = 0.14$ ,  $p < 0.001$ ; Table 2). The results

can be visualized by NMDS (Fig. 2), in which sites are clustered based on analyst. Analyst was also a significant factor in the remaining five studies, but only accounted



**Fig. 3.** Boxplots comparing Bray–Curtis similarity between cross-counts and self-counts in six regional surveys. The MSQA study did not include analyst self-counts; only cross-count data are shown. Each study is shown in a different pattern.

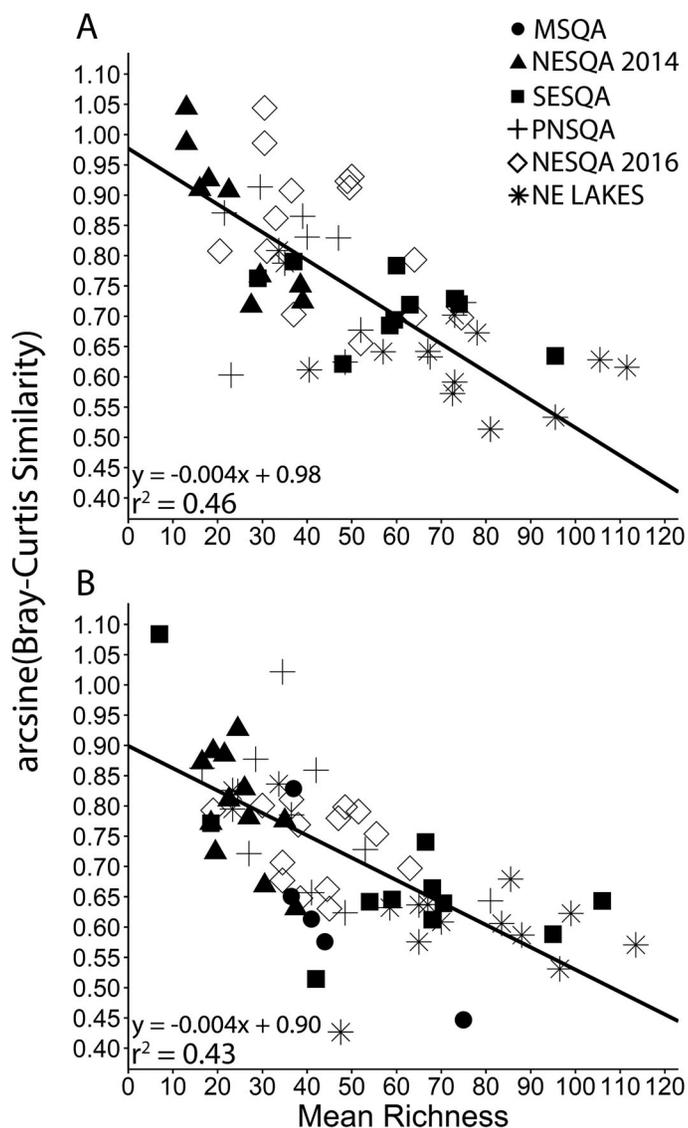
for 1–4% of variation (Table 2). NMDS plots of these studies do not demonstrate visual clustering by analyst (Fig. 2).

Average Bray–Curtis similarity between paired QC counts (cross-QC and self-QC counts) varied among studies (Fig. 3). For cross-QC counts, the median similarity was lowest in MSQA and NE Lakes and highest in NESQA 2014. Note that MSQA protocols did not include self-QC counts. For NESQA 2016, self-QC counts had lower average Bray–Curtis similarity than cross-QC counts, but the difference was small (0.039;  $p = 0.018$ ; Table 3). In the other four surveys, cross-QC and self-QC counts did not differ significantly from one another (Table 3).

Bray–Curtis similarity was strongly correlated with taxon richness (Fig. 4). Furthermore, slope and coefficient of determination were similar between self-QC counts ( $m = -0.004$ ;  $r^2 = 0.46$ ) and cross-QC counts ( $m = -0.004$ ;  $r^2 = 0.43$ ).

**Table 3.** ANOVA results comparing cross-QC and self-QC counts within studies. Note that MSQA protocols did not include self-QC counts.

Study	Median self-QC	Median cross-QC	Difference	$p$
NESQA 2014	0.694	0.714	-0.020	0.468
SESQA	0.660	0.600	0.060	0.583
PNSQA	0.699	0.707	-0.008	0.635
NESQA 2016	0.723	0.684	0.039	0.015
NE Lakes	0.588	0.583	0.005	0.958



**Fig. 4.** RMA linear regressions between mean taxa richness between replicate counts and arcsine-transformed Bray–Curtis similarity across six regional studies for (A) self-QC counts and (B) cross-QC counts. The MSQA study did not include analyst self-counts; only cross-count data are shown. Full linear equations and coefficients of determination ( $r^2$ ) are reported for each regression.

## Discussion

Taxonomic bias has been problematic for diatom data sets combined across analysts and laboratories (Cao et al. 2007; Kelly et al. 2009; Kahlert et al. 2016; Lee et al. 2019). While exact protocols differ among agencies and practitioners, no existing protocols include the four revisions presented here. Our methodological changes to common practice improve transparency in taxonomy, and substantially reduce bias in the following ways: (1) precount voucher floras provide shared species concepts to analysts and serve as permanent records of each study, (2) random assignment of samples prevents

confounding of analyst bias against diatom geographic distribution, (3) formal scientific names are assigned by analysts working collaboratively and each name is documented to a published source, and (4) improved QC distinguishes analyst bias between self-QC and cross-QC consistency. Implementation of these methods produced taxonomic data in which analyst bias was significantly reduced.

### Reduction of bias

The revised process outlined here resulted in reduced analyst bias for all five studies compared to the study (MSQA) that lacked the revisions. Although analyst was a significant factor for all six studies, the effect size was notably larger in MSQA ( $R^2 = 0.14$ ). The revised protocol resulted in <5% of the variance in the data being attributable to analyst bias ( $R^2 = 0.01$ – $0.04$ ).

### QC procedures

Many protocols for diatom analysis specify that the analyst marks the coverslip of each microslide with a diatom scribe, permanently etching the observed transect (Charles et al. 2002; USEPA 2012, 2018). If the slide is selected for one of the 10% repeat counts, a QC analyst returns to this same transect. An implicit assumption of this protocol is that the primary and secondary analyst both examine the same transects and specimens during a QC count. In practice, however, it is difficult for analysts to examine the same transect, with many analysts reporting they are unable to locate or follow previously etched transects (S. A. Spaulding, pers. observ.). Furthermore, the diamond objective marks frequently degrade the thin glass coverslips over time, destroying the integrity of the microslide (M. Potapova, Academy of Natural Sciences of Drexel University, pers. comm.). Many of the transects can no longer be observed in slides, defeating the purpose of long-term archives.

In addition, some protocols specify that the primary and secondary counts must be at least 60% similar based on Bray–Curtis similarity (USEPA 2019). Thus, two analysts examining the same specimens need only be 60% similar in their species identification. Analysts are therefore penalized for failing to match a standard of Bray–Curtis similarity when, instead, differences in identifications between the primary and secondary analyst result from inability to locate or follow etched transects. Species are distributed on slides in a heterogeneous manner such that the greater the species richness, the less likely analysts encounter the same taxa. Moreover, the greater the species richness, the more likely analysts are penalized, falsely, for being “inconsistent.” Not only does the revised protocol produce more consistent data, but it is a more fair assessment of analyst ability.

Some protocols require that QC samples be reanalyzed if they fall below a threshold of taxonomic similarity, irrespective of species richness (USEPA 2012, 2019). In contrast, inclusion of 10% self-QC counts and 10% cross-QC counts allows us to evaluate analyst self-consistency, analyst cross-consistency, and inherent sample heterogeneity as

independent factors. For example, NE Lakes had low median Bray–Curtis similarity compared with the other studies, but this survey also had the greatest species richness. Self-QC and cross-QC count similarities were statistically indistinguishable, indicating individual analysts were equally consistent compared with all analysts as they were recounting their own samples. In contrast, because MSQA did not include self-QC counts, we have no standard by which to judge the relatively low cross-QC count similarities. Inclusion of self-QC counts as a standard by which to evaluate cross-QC counts highlights the importance of ensuring analyst self-consistency, which is currently being addressed through a newly implemented program on diatom taxonomic certification (Lee et al. 2019; <https://diatoms.org/practitioners/diatom-taxonomic-certification>).

Our findings indicate QA/QC efforts must be considered within the context of sample diversity. In the United Kingdom and Ireland, accredited analysts are required to participate in ongoing ring tests, in which analysts report their identification of diatoms from a test slide. The results are compared to the identifications provided by a panel of experts of the same set of test slides. Participants must achieve a certain standard in terms of multi-metric index scores (Kelly 2013). To borrow concepts from this scheme, we believe it is conceivable to compute a “standard curve” characterizing the relationship between Bray–Curtis similarity of recounts and sample richness. Certified diatomists with verified self-consistency could serve as experts counting paired QC samples ranging from low to high diversity, for which Bray–Curtis similarity could be calculated, as in Fig. 4. In future analyses, analysts would be expected to fall within a predetermined range of the standard curve for a given taxon richness.

### Precount floras

The production and use of precount project floras allow transparency and data provenance. Because analysts can refer to mOTUs and their morphological “meaning” through images that are project-specific, data reproducibility and portability are improved. The images within the flora represent the “truth” of the entities encountered because they are derived from each study, rather than from a flora from some other geographic region. Moreover, understanding of the size diminution series, or “morphological space” throughout a species’ life history for each taxon, is shared by analysts. In addition, new taxa are incorporated and shared throughout the project timeline. Finally, project floras can be archived and made accessible so that future work to merge data is based on transparency of the species concepts used in each study.

There is a general perception that the development of an *a priori* flora requires a cost prohibitive effort. A significant amount of time and expertise is required to produce a precount flora, but the effort for subsequent identification and counting of diatom valves is equally reduced. For example, when an analyst has a customized flora for a project, none of

the analyst's time during the enumeration process is diverted to searching for literature references to apply names to taxa. The analyst's job becomes to pay attention to microscopy and specimen characters, not search for names or references. The use of bench reference codes improves data collection efficiency by compartmentalizing and streamlining workflow. Once a flora is produced for a region, it can be reused for subsequent surveys by providing a base document to which analysts add additional taxa as they are encountered.

### Random assignment of samples

The random assignment of samples to analysts does not, by itself, affect analyst bias, but random assignment allows bias to be detected. It is generally unappreciated that diatoms have biogeographic distributions (Kociolek and Spaulding 2000) and that species are regionally distributed across the broad expanse of North America (Potapova and Charles 2007). Random assignment of samples ensures that samples and analysts do not covary with environmental factors. In other words, random assignment of samples crosses potential geographic boundaries, so that bias can be revealed in the QA/QC process.

### Formal scientific names

In the process presented for assessed surveys (except MSQA), the assignment of formal scientific names was shifted to the final stage of analysis. The assignment of scientific names becomes the most speculative aspect of a project. If the image voucher represents the documentation of the truth—the range of specimens observed—then the assignment of formal scientific names is best interpreted as the proposal of a hypothesis (with acknowledgement to E. F. Stoermer). This assignment of taxonomic alignment might be better considered as provisional as more information is gained about diatom species and their meaning.

Pre-collection measures to prevent and reduce bias are easier and more efficient than *post hoc* taxonomic harmonization. However, most taxonomic coordination efforts are implemented after analyses have been completed (Besse-Lototskaya et al. 2006; Manoylov 2014). Back coordination of taxonomic data is time-consuming, sacrifices data resolution, and reduces analyst signal at the expense of also reducing the desired environmental signal (Lee et al. 2019).

### Comments and recommendations

We outline a process for analyzing diatom samples in national and regional surveys using multiple analysts. The process is also appropriate for smaller studies that need to ensure taxonomic consistency over time, or to place results in a larger, regional context of species' responses to environmental variables. For example, recent work to harmonize large diatom data sets shows that when analyst bias is removed, there is greater ability to observe species responses to phosphorus (Lee et al. 2019). The process includes four steps that, taken together, significantly decreased analyst bias in five recent regional surveys.

We highlight the importance of rigorous QA/QC methods, with consideration to self-QC as well as cross-QC counts. Furthermore, we suggest an approach to isolate analyst taxonomic consistency from sample heterogeneity. This process should be considered a central component of a modern, robust biological assessment of aquatic condition based on diatom data.

### References

- Besse-Lototskaya, A., P. F. M. Verdonchot, and J. A. Sinkeldam. 2006. Uncertainty in diatom assessment: Sampling, identification and counting variation. *Hydrobiologia* **566**: 247–260. doi:[10.1007/s10750-006-0092-5](https://doi.org/10.1007/s10750-006-0092-5)
- Bishop, I. W., R. M. Esposito, M. Tyree, and S. A. Spaulding. 2017. A diatom voucher flora from selected southeast rivers (USA). *Phytotaxa* **332**: 101–140. doi:[10.11646/phytotaxa.332.2.1](https://doi.org/10.11646/phytotaxa.332.2.1)
- 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. *J. North Am. Benthol. Soc.* **26**: 566–585. doi:[10.1899/06-078.1](https://doi.org/10.1899/06-078.1)
- 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. Patrick Center for Environmental Research Report No. 02-06.
- Coles, J. F., K. Riva-Murray, P. C. Van Metre, D. T. Button, A. H. Bell, S. L. Qi, C. A. Journey, and R. W. Sheibley. 2018. Design and methods of the U.S. Geological Survey Northeast Stream Quality Assessment (NESQA), 2016. U.S. Geological Survey Open-File Report No. 2018-1183.
- Garrett, J. D., J. W. Frey, P. C. Van Metre, C. A. Journey, N. Nakagaki, D. T. Button, and L. H. Nowell. 2017. Design and methods of the Midwest Stream Quality Assessment (MSQA), 2013. No. 2017-1073.
- 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. 2015. Design and Methods of the Southeast Stream Quality Assessment (SESQA), 2014. U.S. Geological Survey Open-File Report No. 2015-1095.
- Kahlert, M. E., and others. 2016. Quality assurance of diatom counts in Europe: towards harmonized datasets. *Hydrobiologia* **772**: 1–14. doi:[10.1007/s10750-016-2651-8](https://doi.org/10.1007/s10750-016-2651-8)
- Kelly, M. 2013. Building capacity for ecological assessment using diatoms in UK rivers. *J. Ecol. Environ.* **36**: 89–94. doi:[10.5141/ecoenv.2013.011](https://doi.org/10.5141/ecoenv.2013.011)
- Kelly, M., and others. 2009. A comparison of national approaches to setting ecological status boundaries in phytobenthos assessment for the European Water Framework Directive: results of an intercalibration exercise. *Hydrobiologia* **621**: 169–182. doi:[10.1007/s10750-008-9641-4](https://doi.org/10.1007/s10750-008-9641-4)
- Kociolek, J. P., and S. A. Spaulding. 2000. Freshwater diatom biogeography. *Nov. Hedwigia* **71**: 223–241.
- Lee, S. S., I. W. Bishop, S. A. Spaulding, R. M. Mitchell, and L. L. Yuan. 2019. Taxonomic harmonization may reveal a

- stronger association between diatom assemblages and total phosphorus in large datasets. *Ecol. Indic.* **102**: 166–174. doi:[10.1016/j.ecolind.2019.01.061](https://doi.org/10.1016/j.ecolind.2019.01.061)
- Legendre, P., and L. Legendre. 2012. Interpretation of ecological structures, p. 521–624. *In* Numerical ecology. Elsevier.
- Manoylov, K. M. 2014. Taxonomic identification of algae (morphological and molecular): species concepts, methodologies, and their implications for ecological bioassessment. *J. Phycol.* **50**: 409–424. doi:[10.1111/jpy.12183](https://doi.org/10.1111/jpy.12183)
- Omernik, J. M., and G. E. Griffith. 2014. Ecoregions of the conterminous United States: Evolution of a hierarchical spatial framework. *Environ. Manag.* **54**: 1249–1266. doi:[10.1007/s00267-014-0364-1](https://doi.org/10.1007/s00267-014-0364-1)
- Pan, Y., R. J. Stevenson, B. H. Hill, et al. 1996. Using diatoms as indicators of ecological conditions in lotic systems: A regional assessment. *J. North Am. Benthol. Soc.* **15**: 481–495.
- Potapova, M., and D. F. Charles. 2007. Diatom metrics for monitoring eutrophication in rivers of the United States. *Ecol. Indic.* **7**: 48–70. doi:[10.1016/j.ecolind.2005.10.001](https://doi.org/10.1016/j.ecolind.2005.10.001)
- Sheibley, R. W., J. L. Morace, C. A. Journey, P. C. Van Metre, A. H. Bell, N. Nakagaki, D. T. Button, and S. L. Qi. 2017. National water-quality assessment project design and methods of the Pacific northwest stream quality assessment (PNSQA), 2015. U.S. Geological Survey Open-File Report No. 2017-1103.
- Stoermer, E. F., and J. P. Smol. 2010. *The diatoms: applications for the environmental and earth sciences*, 2nd Edition. Cambridge Univ. Press.
- U.S. Environmental Protection Agency (USEPA). 2009. National rivers and streams assessment: Laboratory methods manual. EPA-841-B-07-010. USEPA.
- U.S. Environmental Protection Agency (USEPA). 2012. 2012 National lakes assessment. Laboratory operations manual. EPA-841-B-11-004. USEPA.
- U.S. Environmental Protection Agency (USEPA). 2018. National rivers and streams assessment 2018–19: Laboratory operations manual; EPA-841-B-17-004. USEPA.
- U.S. Environmental Protection Agency (USEPA). 2019. Standard operating procedure for phytoplankton analysis LG401, Revision 06. USEPA.
- U.S. Geological Survey. 2019. BioData—Aquatic bioassessment data for the nation available on the world wide web, accessed October 24, 2019. doi:[10.5066/F77W698B](https://doi.org/10.5066/F77W698B)
- Van Metre, P. C., J. W. Frey, and E. Tarquinio. 2012. The mid-west stream quality assessment. U.S. Geological Survey Fact Sheet 2012-3124.
- Van Metre, P. C., and C. A. Journey. 2014. The southeast stream quality assessment. U.S. Geological Survey Fact Sheet 2014-3023.
- Van Metre, P. C., J. L. Morace, and R. W. Sheibley. 2015. The Pacific northwest stream quality assessment. US Geological Survey Fact Sheet 3020.2.
- Van Metre, P. C., K. Riva-Murray, and J. F. Coles. 2016. The northeast stream quality assessment. U.S. Geological Survey Fact Sheet 2016-3021.
- Werner, P., S. Adler, and M. Dreßler. 2016. Effects of counting variances on water quality assessments: implications from four benthic diatom samples, each counted by 40 diatomists. *J. Appl. Phycol.* **28**: 2287–2297. doi:[10.1007/s10811-015-0760-9](https://doi.org/10.1007/s10811-015-0760-9)
- Wolfe, A. P. 1997. On diatom concentrations in lake sediments: results from an inter-laboratory comparison and other tests performed on a uniform sample. *J. Paleolimnol.* **18**: 261–268. doi:[10.1023/A:1007937300347](https://doi.org/10.1023/A:1007937300347)

#### Acknowledgments

We thank Paul Bliznik, Alec Camp and Leanna Jo Clarkson for assistance and contributions in the CWEST/INSTAAR lab. Sylvia Lee furthered discussion of the content of this manuscript. This material is based upon work supported by the U.S. Geological Survey under Cooperative Agreement #G15 AC00104. Any use of trade names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

*Submitted 28 October 2019*

*Revised 24 January 2020*

*Accepted 28 January 2020*

*Associate editor: John Smol*