Diatom enumeration method influences biological assessments of southeastern USA streams

Meredith A. Tyree1,4, Daren M. Carlisle2,5, and Sarah A. Spaulding1,3,6

1Institute of Arctic and Alpine Research, University of Colorado, 4001 Discovery Drive, Boulder, Colorado 80303 USA
2United States Geological Survey, 1217 Biltmore Drive, Lawrence, Kansas 66049 USA
3United States Geological Survey, 4001 Discovery Drive, Boulder, Colorado 80303 USA

Abstract: Current fixed-count enumeration methods for benthic diatoms are likely inadequate for most research and monitoring objectives. These methods underestimate taxa richness and may fail to detect losses of species caused by human impacts. Consequently, the full potential of diatoms is not realized in current assessments of biological integrity or species diversity. In this study, we hypothesize that alternative enumeration methods differ in their ability to quantify species composition. Furthermore, we hypothesize that an alternative to the traditional fixed-count method will improve both performance of observed/expected (O/E) indices derived from River Invertebrate Prediction and Classification System models and the discrimination of reference-quality and human-modified sites by other standard metrics used in biological assessments. To test these hypotheses, we assessed 1) how well 3 counting methods characterized species richness in a subset of 15 samples of stream benthic diatoms and 2) how counting method affected the performance of O/E indices and metrics by comparing the traditional fixed-count method against the best-performing alternative method. These latter comparisons were based on samples collected from 68 reference-quality streams and 20 streams located along an urban disturbance gradient. We demonstrate that traditional fixed counts failed to detect >½ of species present in most of the 68 reference-quality sites. Instead, timed-presence data produced the O/E index with the best performance and a level of precision similar to published invertebrate O/E indices. Furthermore, the O/E index based on the timed-presence data allowed us to determine which species are most often lost with urbanization. We found that traditional fixed-count and alternative timed-presence data produce metrics that are nearly equally able to discriminate between reference and disturbed sites. This study demonstrates that alternative counting methods improve species detection and require up to ~30% less effort.

Key words: diatoms, RIVPACS, O/E indices, metrics, taxonomic completeness, species richness

The Clean Water Act (1972 Federal Water Pollution Control Act [33 U.S.C. §§1251–1387]) mandates the restoration and maintenance of the biological integrity of the United States' waters. A primary objective of biological assessments, therefore, is to characterize the composition and diversity of species to assess and monitor aquatic biological integrity. Furthermore, biological assessments should measure alteration and losses of species diversity caused by anthropogenic disturbances. However, current enumeration methods for benthic diatoms produce data that fall short on both counts because they fail to consistently detect low-abundance taxa. Consequently, diatom assemblages are not used to their full potential as indicators of biological integrity despite their ubiquity and responsiveness to environmental conditions (Patrick 1973).

Quantifying assemblage composition is a critical step when using diatoms to infer water quality conditions (Patrick and Strawbridge 1963, Patrick and Reimer 1966, Charles et al. 2002, Kelly et al. 2008, USEPA 2009). Species composition is typically determined from fixed counts, in which an analyst identifies and enumerates cells until a specified number of cells has been reached. Early compositional studies of diatoms were based on very high counts, often exceeding 10,000 valves, or half-cells (Patrick and Strawbridge 1963, Patrick and Reimer 1966). Modern assessments, however, are based on much lower counts, typically 600 valves (Pappas...
Across taxonomic groups, ecological communities almost universally follow a truncated lognormal curve, with several abundant species and many rare species (Magurran 2004). Therefore, data based on low fixed counts can poorly characterize the abundance of non-dominant taxa (Cao and Hawkins 2005, Gillett et al. 2011). In some cases, important taxa such as the nuisance species Didymosphenia geminata (Lyngbye) Mart.Schmidt may be completely missed (Spaulding and Elwell 2007). Discussions of the value of these poorly-characterized, rare taxa in assessment have increased (Hawkins et al. 2000, Cao et al. 2001, Clarke and Murphy 2006, Van Sickle et al. 2007). Some authors argue that when rare taxa are omitted there is a loss of information, potentially reducing assessment sensitivity (Cao et al. 2001). Other authors report that including rare taxa adds noise to index scores, thereby decreasing index precision, sensitivity, or both (Hawkins et al. 2000, Clarke and Murphy 2006, Van Sickle et al. 2007). Moreover, low fixed counts may limit our ability to detect the loss of species diversity caused by anthropogenic impacts. Neither North American nor European assessments address species loss (e.g., Pan et al. 1996, Kelly et al. 2008). The biological condition gradient approach (Davies and Jackson 2006) is directed at characterizing loss of sensitive native species across human-disturbance gradients, but particular species were not recognized as absent from degraded sites in a biological condition gradient study of diatoms (Hausmann et al. 2016).

It is possible that arguments both for and against the inclusion of rare taxa are partially correct if they address different subsets of rare taxa. Some taxa in samples appear rare because they are incidental, such as planktonic taxa that drift downstream from a reservoir (Stevenson and Peterson 1991). These incidental taxa are unlikely to be informative in bioassessments because they do not represent conditions at the site from which they were sampled and therefore introduce statistical noise. However, consistent detection of common but non-abundant species could provide an important signal to distinguish the effects of anthropogenic disturbance on species diversity if these species have distinct environmental preferences (Gillett et al. 2011). High fixed counts could detect these common, non-abundant species, but such counts are time- and cost-prohibitive. To date, comparisons of the efficacy of low fixed counts to alternatives have not been evaluated. We compare the low fixed-count method to 2 alternative methods to evaluate our ability to characterize species richness and detect the loss of key indicator taxa.

One alternative means to enumerate diatoms is the stratified-count method (Stevenson et al. 2010). The purpose of a stratified count is to focus an appropriate amount of effort on the common species to adequately characterize their abundance, then to put additional effort into uncovering the presence and abundance of rare species. The stratified-count method involves enumerating all valves and tracking the area of the slide examined until a pre-specified number of diatom half-cells, or valves (e.g., 50), of 1 taxon is reached. Then, identification and enumeration of valves, except of that taxon, continues. Each time the pre-specified number of valves is reached for a taxon, enumeration of that taxon stops, but enumeration of all other taxa continues until a desired total count is reached. Valve counts of dominant taxa that exceeded the pre-specified number of valves are then estimated by scaling actual taxon counts by the total area of the slide that was examined, giving the effective counts of these dominant taxa. Effective counts estimate the actual number of valves that would have been counted if every valve of the dominant taxa had been recorded rather than estimated by the area of the slide examined (Stevenson et al. 2010). Stratified counts of the same number of total actual individuals should better detect rare taxa because fewer abundant taxa are directly counted, focusing more effort on characterizing rare species presence and abundance (Spaulding et al. 1997). However, this method has not been directly compared with the fixed-count method.

The second alternative enumeration method we consider is the timed-presence method, in which the analyst identifies and enumerates the first 100 valves on a slide, then continues to examine the slide for a pre-specified length of time and records (but does not enumerate) the observed taxa. This method is designed to maximize the detection of taxa richness while characterizing dominant taxa abundance. This method has not been used in any published studies.

One limitation to the application of diatoms in biological assessments is the dependence on a small number of skilled taxonomists. An experienced analyst requires an average of 2 to 3.5 h to complete a fixed-count analysis of 300 cells for each sample, and difficult samples can require up to 8 h to count. If an enumeration method is faster and can improve characterization of species richness, or both, it would offer an improvement over current methods.

We hypothesized that enumeration methods (fixed count, stratified count, and timed presence) would differ in their ability to quantify species composition and hence bioassessment inferences. We evaluated our hypothesis by comparing how the traditional fixed-count method and the best-performing alternative method affected 1) estimates of species richness in a subset of samples and 2) the performance of observed/expected (O/E) indices and estimates of assemblage-level metrics. These comparisons were based on a large number of reference sites and samples collected along an urban gradient.

**METHODS**

**Site selection**

We obtained permanent slides for 68 reference sites used across 3 regional assessments conducted in the southeastern
Field sampling and slide preparation

NAWQA and SESQA samples were collected following richest-targeted habitat protocols (Moulton et al. 2002). Where possible, periphyton was scraped from coarse substrate (cobble or wood) within riffle habitats with a stiff-bristled brush from a 12-cm²-delimited area at each of 11 transects. When coarse substrate was not available, the top 1 cm of sediment was removed with a syringe from the streambed within a 12-cm²-delimited area (Gilliom et al. 1995, Journey et al. 2015). NRSA samples were collected following multi-habitat transect-based sampling protocols (USEPA 2009). Prior to sampling, the initial transect at each site was randomly assigned a sampling station (left bank, right bank, or center), and each subsequent transect was systematically assigned a sampling station thereafter (in order from left, center, right) for a total of 11 transects. During sampling, periphyton was scraped from coarse substrate in the assigned sampling station with a toothbrush and an area delimiter. If no coarse substrate was available at the sampling station, the top 1 cm of sediment was removed from the streambed with a syringe within the delimited area (USEPA 2009). For all 3 regional studies, samples from all transects at a site were combined into a single composite sample, preserved in formalin, and transported to the laboratory on ice for processing. Periphyton material was digested to remove organic material to allow species-level identification of the silica valves, rinsed, then pipetted onto coverslips and mounted to permanent slides with Naphrax or Zrax mounting medium (Acker et al. 1999).

Voucher compilation and diatom enumeration with the fixed-count method

Previous work has highlighted problems with taxonomic consistency between datasets produced by different laboratories, which can limit the use of these datasets in bioassessment (Cao et al. 2007, Lee et al. 2019). To avoid analyst bias, all samples were identified and counted by a single analyst for this study.

Prior to analysis, we developed a pre-count voucher flora of images to ensure between-sample consistency in the morphological concept of each taxon (Bishop et al. 2017). We compiled a final voucher flora of 2780 images representing 647 taxa, which serves as a permanent record (https://instaar.colorado.edu/research/labs-groups/diatom-laboratory/research-detail/). Encyonema minutum (Hilse) D.G.Mann and Encyonema lineolatum Krammer could not be consistently distinguished during enumeration and were combined...
into a single entity during analysis with all 3 enumeration methods.

We then enumerated taxa on 68 reference slides and 20 test slides based on 600-valve fixed counts. The analyst examined slides along transects at 1000× magnification, and taxa were identified based on morphological operational taxonomic unit codes in the voucher flora. Valves were only counted if at least 60% of a valve was visible within a field of view.

Selecting the alternative enumeration method

We determined which alternative counting method to apply to the entire set of samples by randomly selecting a subset of 15 reference sites and enumerating diatoms at each of those sites with all 3 methods.

For the stratified-count method, we recorded microscope stage coordinates and then enumerated valves along a transect until 50 valves of the most common taxon had been encountered. At that point, we recorded the stage coordinates and continued enumerating all valves except that taxon. This enumeration process for examining transects and successively omitting species was repeated for each taxon for which 50 valves was counted until we reached a total actual count of 600 valves. We then calculated the effective count of each dominant taxon that exceeded 50 valves by multiplying the total number of each taxon actually counted (i.e., 50) by the area of the slide examined and then dividing by the area of the slide required to encounter the first 50 valves of that taxon. The effective count thus estimates the number of valves of each taxon that would have been observed with the fixed-count method had we continued counting beyond 600 valves to the total effective count (total effective counts ranged from 643 to 23,722 valves). Effective counts produced by the stratified-counting method thus estimated large fixed counts with a fraction of the required effort.

For the timed-presence method, we identified and enumerated the first 100 observed valves to characterize the abundance of the most dominant taxa. Relative abundance data is not required for O/E indices or the metrics we used, but it was collected for potential qualitative evaluations. After the first 100 valves, we observed and identified additional species present on the slide for a period of 1 h. Abundances of these additional taxa were not measured or recorded.

We recorded taxonomic richness in each of the 15 sites based on each enumeration method. Based on the taxon richness detected with each enumeration method, we chose to apply the timed-presence method to samples collected at all 68 reference sites and 20 test sites. We then confirmed the results from the subset of sites by comparing richness estimated with the fixed-count method with richness estimated with the timed-presence method at all sites.

Bioassessment application: O/E indices

We used O/E indices derived from River Invertebrate Prediction and Classification System (RIVPACS)-type models (sensu Hawkins 2006) to evaluate the relative influence of each enumeration technique on O/E index performance. The O/E index, or taxonomic completeness, represents the proportion of taxa observed (O) in a sample to the taxa expected (E) under reference conditions. RIVPACS-type models require accurate and precise estimates of taxonomic composition to predict taxa occurrences. We constructed both genus- and species-level models with composition data from the fixed-count and timed-presence methods to produce 4 O/E indices.

Detailed methods on RIVPACS model and O/E index development are provided elsewhere (Moss et al. 1987, Hawkins 2006), and we provide only a brief overview here. We first clustered reference sites into biologically-similar groups with the unweighted pair group method with arithmetic mean based on pairwise Bray–Curtis similarities calculated from presence/absence data. This procedure produced a dendrogram that we used to identify groups of sites that had the maximum within-group similarity, consistent between-group similarity, and ≥3 sites (Carlisle et al. 2008). We then used random-forest models (Cutler et al. 2007) to predict the probabilities that sites belong to different biological clusters based on their natural environmental setting (Hawkins et al. 2010; see Table 1 for the predictors we used in these models). Next, we computed the predicted probability of capturing each species at a given site (Pc) by weighting the frequency of each species within the biological groups by the probabilities of group membership for each site, as estimated by the random-forest model. In building the models, we also evaluated multiple levels of Pc (ranging from 0–0.7) for the timed-presence species-level model following Van Sickle et al. (2007). We evaluated multiple levels of Pc because the appropriate Pc level for diatom assemblages has not been investigated (Table 2). We evaluated multiple levels of Pc for just a single model to confirm that O/E indices built with diatom assemblage data conform to trends observed for invertebrate data. We subsequently chose Pc > 0.5 to allow direct comparisons with invertebrate O/E indices (Van Sickle et al. 2007). Finally, we calculated the expected richness of predicted taxa (E) at each site by summing the Pc of all species at that site with Pc > 0.5. O/E values were then calculated as the number of taxa predicted to occur with Pc > 0.5 that were observed (O) divided by E.

The ability of RIVPACS-derived O/E indices to detect alteration depends on the model’s capacity to accurately and precisely predict how assemblage composition varies across naturally-occurring environmental gradients. Thus, we compared accuracy, precision, and sensitivity of O/E indices developed with data from the fixed-count method with O/E indices developed with data from the timed-presence method. We quantified accuracy by evaluating mean O/E
We quantified precision as the standard deviation of reference site O/E values. Observed values were compared with theoretical upper and lower bounds. We estimated the lower bounds of model precision for each full model by calculating a null model in which environmental information was ignored and E was the same at all sites (Van Sickle et al. 2005). We estimated the upper bounds of model precision by approximating replicate-sampling variability following Van Sickle et al. (2005). Finally, we evaluated the sensitivity of the O/E indices to detect biological response along gradients of known anthropogenic perturbation. For each model, we computed O/E for each of the 20 test sites, then correlated (Pearson’s) O/E with % impervious cover in each watershed. We used imperviousness as a measure of disturbance because it is a continuous variable that is often correlated with stream degradation, alteration of biodiversity, and species loss (Schueler et al. 2009).

RIVPACS models predict which taxa should have occurred in test sites prior to human-caused degradation. Knowledge of the taxa that are expected but not detected can aid in identifying the environmental stressors that drive shifts in assemblage composition (Carlisle and Hawkins 2008). Moreover, loss of diatom biodiversity as a result of anthropogenic disturbance has received little attention as a component of bioassessment. We estimated the taxa that were lost from test sites based on each enumeration method by recording how many species predicted to occur (\( \text{Pc} \geq 0.5 \)) at test sites were not detected.

Finally, we tested the robustness of the timed-presence method to variability in how quickly different analysts might work and thus the number of taxa different analysts might encounter and identify during the 1 h of scanning employed by the timed-presence method. To test the resilience of O/E indices to the number of species detected, we constructed subsets of species-level, timed-presence data. We removed 1, 5, 10, or 20% of species from the dataset in this resiliency analysis. Species were removed mostly at random, but we did not remove any species found in the first 100 diatom

| \( P_c \) | Mean modeled O/E | Modeled O/E SD | Rep SD | Null O/E SD | Null SD – Modeled SD |
|---------|-----------------|----------------|--------|-------------|----------------------|
| 0       | 1.01            | 0.27           | 0.09   | 0.28        | 0.01                 |
| 0.35    | 1.04            | 0.22           | 0.10   | 0.24        | 0.02                 |
| 0.40    | 1.04            | 0.21           | 0.10   | 0.25        | 0.05                 |
| 0.45    | 1.04            | 0.19           | 0.10   | 0.23        | 0.05                 |
| 0.50    | 1.04            | 0.19           | 0.11   | 0.21        | 0.02                 |
| 0.55    | 1.04            | 0.19           | 0.11   | 0.20        | 0.01                 |
| 0.60    | 1.04            | 0.18           | 0.11   | 0.19        | 0.00                 |
| 0.65    | 1.03            | 0.18           | 0.11   | 0.16        | 0.01                 |
| 0.70    | 1.02            | 0.16           | 0.11   | 0.16        | 0.01                 |

Table 1. Predictors for all 4 RIVPACS-type models.

| Model                        | Predictors                                                                 |
|-----------------------------|-----------------------------------------------------------------------------|
| Fixed count: species        | Surficial geology class 21 Mean % MgO content Mean % CaO content % of catchment that contains the level 3 ecoregion Piedmont Mean uniaxial compressive strength Elevation range |
| Timed presence: species     | Mean % CaO content Mean % FeO content Average value of clay content Base flow index Elevation range |
| Fixed count: genus          | Mean % SiO content Mean % MgO content Mean % CaO content % of total precipitation as snow Number of wet days in October Maximum watershed elevation (m) Elevation range |
| Timed presence: genus       | Surficial geology class 21 Soil material < 5 mm Soil material < 0.07 mm Topographic wetness index Mean % MgO content Mean % CaO content |

Table 2. Mean modeled observed/expected (O/E) values (Mean modeled O/E), modeled O/E standard deviations (Modeled O/E SD), estimated replicate sample standard deviations (Rep SD), null model-based O/E standard deviations (Null O/E SD) based on 68 reference sites, the difference between null and model-based O/E standard deviations (Null SD – Modeled SD), and probabilities of capture (\( \text{Pc} \)) thresholds for the timed-presence species-level model. The differences between Modeled O/E SD, Null model SD, and Rep SD quantify the relative ability of the RIVPACS-type models to account for environmental variation (values >0 indicate improvement relative to the null model).
valves enumerated at each site because we assumed that abundant taxa would be detected consistently among analysts. We developed O/E indices with each of these reduced datasets and evaluated each index’s accuracy, precision, and sensitivity with the methods described above.

**Bioassessment application: Metric evaluation**

Multi-metric indices (MMIs) are widely used for biological assessment of diatoms, particularly in the European Union (Kelly and Whitton 1995, Kelly et al. 2009, Almeida et al. 2014). MMIs are developed by systematically evaluating each of a large set of candidate biological metrics (e.g., pollution tolerance, habit, etc.) for responsiveness to anthropogenic influences (Stoddard et al. 2008). Individual metrics that statistically discriminate between reference and disturbed sites are considered for inclusion in the MMI. We explored the effects of each alternative enumeration method on this step of MMI development. We used data from the fixed-count and timed-presence methods to compute the number of taxa possessing 61 different specific traits (i.e., metrics; Table S1) at reference and disturbed sites. These metrics were compiled from a number of sources (Lange-Bertalot 1979, Bahls 1993, Van Dam et al. 1994, Potapova and Charles 2007, Larras and Montuelle 2014, Tang et al. 2016, Spaulding et al. 2019).

We used Wilcoxon–Mann–Whitney tests, non-parametric tests that determine whether 2 independent samples belong to separate populations, to assess if each metric differed between reference and disturbed sites.

**RESULTS**

The counting methods differed in their estimates of mean species richness across the 15 sites in which we used all 3 methods (Fig. 2). The fixed-count method detected the lowest species richness compared with other methods for all 15 sites, but the relative difference in the number of species detected varied depending on total richness. For example, at the 5 sites with the lowest species richness (as determined by all 3 methods), the fixed-count method detected only ~10 to 30% of the taxa detected by the other 2 methods. At the 5 sites with the highest species richness, the fixed-count method detected ~63 to 78% of the taxa of the other 2 methods. The stratified-count method detected the greatest species richness at sites with a low number of taxa, but performed similarly to the fixed-count method at sites with a high number of taxa. However, the stratified-count method also required the longest time to count. For most sites, the timed-presence method detected the greatest number of taxa (10 of 15 sites).

In the initial 15 sites, fixed counts required ~2 to 3.5 h/slide, with a maximum of 8 h. In contrast, timed-presence counts required 1.5 to 2 h/slide, with a maximum of just over 2 h. Stratified counts required ~4 to 8 h/slide in low-diversity sites and 2 to 3.5 h/slide in high-diversity sites.

The timed-presence method better estimated species richness and required less time to analyze samples than other methods, so we analyzed all 68 reference sites and 20 test sites with the timed-presence method. We then compared these results with the fixed-count method. Calculations of species richness for all reference sites confirmed the results from our initial set of 15 sites: the timed-presence method detected significantly more species in reference sites than the fixed-count method (t = 21.5, df = 67, p < 0.001). The fixed-count method detected <½ the number of species as the timed-presence method in all but 1 reference site (Fig. 3).

O/E precision was lowest at a Pc threshold of 0 (Table 2). A Pc threshold of 0.45 gave the greatest precision relative to the precision of the null model, but improvement was marginal compared with Pc = 0.5. We therefore used the Pc threshold of 0.5 in the final models to align with conventions used in many macroinvertebrate assessments (Van Sickle et al. 2007, Ostertiller and Hawkins 2009).

O/E indices based on either the species- or genus-level taxonomic data and timed-presence methods were equally accurate and more precise and responsive than indices based on either of the fixed-count models (Table 3). The genus timed-presence model had the lowest standard deviation (SD = 0.16) and was marginally more precise than the null model (SD = 0.17). However, it was not responsive to watershed impervious cover. The species timed-presence index was the next most precise (model SD = 0.19, null model SD = 0.21), and it was also the most responsive because it was most strongly correlated with imperviousness (r = −0.43; Fig. 4A–D). The species fixed-count model (SD = 0.32) and the genus fixed-count model (SD = 0.21) were both slightly less precise than the null models (SD = 0.26 and 0.20; Table 3) and less responsive to stress because of
their low correlations with imperviousness (species-level $r = 0.01$; genus-level $r = -0.11$).

Analysis of resilience to missed low-abundance species showed that O/E indices based on the timed-presence method were resilient to some degree of variation in species detected. When between 1 and 20% of low-abundance species were randomly removed from the data, the precision of the O/E indices changed little (SD = 0.19 to 0.20) relative to the original index (Table 4). However, responsiveness to anthropogenic disturbance varied based on the % of species removed. Correlations between test-site O/E values and the urbanization gradient varied little when 1 or 5% of species were removed, but dropped 7 percentage points when 10% of species were removed ($r = -0.36$) and 21 percentage

Figure 3. Number of taxa detected with fixed-count (gray bars) and timed-presence (black bars) methods in 68 reference sites, ordered from the site with the highest number of taxa to the lowest number of taxa based on the timed-presence method. Each horizontal bar marks 1 of the 68 sites. Site information is provided in Table S3.

Table 3. Mean modeled observed/expected (O/E) values (Mean modeled O/E), modeled O/E standard deviations (Modeled O/E SD), replicate sample O/E standard deviations (Rep SD), null model-based O/E standard deviations (Null O/E SD), the difference between null and model-based O/E standard deviations (Null SD – Modeled SD), and Pearson’s $r$ correlations between O/E values and % impervious surface for 20 test sites. Null SD – Modeled 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). $r$ values measure the responsiveness of the O/E indices to a human-disturbance gradient.

| Model               | Mean modeled O/E | Modeled O/E SD | Rep SD | Null O/E SD | Null SD – Modeled SD | $r$ with % impervious surface |
|---------------------|------------------|----------------|--------|-------------|----------------------|-------------------------------|
| Fixed count: species| 0.97             | 0.32           | 0.20   | 0.26        | -0.06                | 0.02                          |
| Timed presence: species | 1.03             | 0.19           | 0.11   | 0.21        | 0.02                 | -0.43                         |
| Fixed count: genus  | 0.97             | 0.21           | 0.14   | 0.20        | -0.01                | -0.04                         |
| Timed presence: genus | 0.97             | 0.16           | 0.09   | 0.17        | 0.01                 | -0.15                         |
points when 20% of species were removed ($r = -0.22$). Therefore, we expect that analysts could differ up to 10% in the number of species that they observe and identify without affecting the responsiveness of the O/E index. Differences $>20\%$ in the number of species recorded, however, will lead to a reduction in responsiveness.

The identity of species predicted but not detected at urban sites differed between RIVPACS models based on different enumeration methods. Output from the timed-presence model indicated *Platessa stewartii* (R.M.Patrick) Potapova was most frequently absent in urban-influenced basins (15 of 20 test sites; Table S2), followed by *Psammothidium subatomoides* (Hust.) Bukht., *Encyonema minutum* var. *pseudogracilis* (Cholnoky 1968) Czarn., *Synedra pulchella* var. *flexella* Boyer, and *Gomphonema louisiananum* Kalinsky (13, 12, 12, and 11 sites, respectively). These diatom species are known to be sensitive to watershed disturbance (Table S2). Furthermore, these taxa were absent from a large percentage of urban sites (55–75%). In contrast, the models based on fixed counts estimated that taxa were absent from $\leq 40\%$ of the urban-influenced basins where they were predicted to occur. Thus, the fixed-count models appeared to be less able to estimate species loss than other models. Moreover, those that were not detected are not known to be sensitive to watershed disturbance (Table S2).

Of the 61 diatom traits evaluated, the frequencies of occurrence of 16 differed statistically between reference and disturbed sites. Overall, individual metrics based on the fixed-count method and the timed-presence method were equally responsive to urbanization (Table 5, Fig. 5). Specifically, 10 of the 16 metrics were responsive to urbanization regardless of the enumeration method used. In contrast, 4 metrics

| % species removed at random | Mean modeled O/E | Modeled O/E SD | Rep SD | Null O/E SD | Null SD – Model SD | $r$ with % impervious surface |
|-----------------------------|------------------|----------------|--------|-------------|-------------------|------------------------------|
| 0                           | 1.02             | 0.19           | 0.11   | 0.21        | 0.02              | −0.43                        |
| 1                           | 1.05             | 0.19           | 0.11   | 0.24        | 0.05              | −0.41                        |
| 5                           | 1.03             | 0.20           | 0.12   | 0.25        | 0.05              | −0.44                        |
| 10                          | 1.06             | 0.19           | 0.12   | 0.21        | 0.02              | −0.36                        |
| 20                          | 1.04             | 0.19           | 0.15   | 0.25        | 0.06              | −0.22                        |

Table 4. Mean modeled observed/expected (O/E) values (Mean modeled O/E) for reference sites, modeled O/E standard deviations (Modeled O/E SD), estimated replicate sample standard deviations (Rep SD), null model-based O/E standard deviations (Null O/E SD), the difference between null and modeled O/E standard deviations (Null SD – Modeled SD), and Pearson’s $r$ correlations with % impervious surface for indices constructed with timed-presence species-level data in a sensitivity analysis of rare species removal. The % of species that are removed at random represents how indices would be expected to differ with differing analysts working for a 1-h analysis. 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).
(Trophic 5, Salinity 2, BC 5, Very small) were responsive only when using the fixed-count method, whereas 2 metrics (N fixer, Small) were responsive only when using the timed-presence method (Fig. 5).

**DISCUSSION**

The fixed-count method is used widely in diatom studies, ostensibly because it standardizes the enumeration process—although whether its use achieves standardization has not been examined. However, our results indicate that low fixed counts are inferior to stratified and timed-presence counts for calculating O/E indices, comparing metrics, and detecting species loss for at least 3 reasons. First, low fixed counts fail to adequately characterize richness. Relative to the fixed-count method, stratified-count and timed-presence methods markedly improved species detection in sites with low to medium species richness. This result demonstrates that abundant taxa that dominate a sample can obscure the presence of less abundant taxa when low fixed counts are used, supporting the findings of Cao and Hawkins (2005) for invertebrates. Second, some taxa that are not detected by the fixed-count method appear to be sensitive to anthropogenic disturbances. O/E indices based on fixed-count data were not as responsive to a gradient of watershed urbanization as indices based on timed-presence methods. Third, the timed-presence method requires a fraction of analyst time compared with the fixed-count method and produced metrics just as responsive to anthropogenic disturbance as those based on fixed counts.

**Performance detecting richness**

Overall, in the 15 sites in which we used all 3 methods, the timed-presence method detected the greatest richness in sites and required, on average, the least time for analysis. However, the stratified-count method detected the greatest richness in the lowest richness sites, albeit with more average analysis time required compared to fixed and timed presence counts. The stratified-count method should, therefore, be considered a viable alternative to the fixed-count method, particularly in sites with low diversity or when obtaining species relative abundance data is also necessary.

**Evaluation of taxonomic completeness**

Our results indicate that exclusion of locally-rare taxa improves index performance, as observed in other studies (e.g., Van Sickle et al. 2007). Specifically, the use of intermediate $P_c$ values (e.g., 0.5) produced an index in which taxa that are moderately to very common were included and locally-rare taxa were omitted. However, we also observed that the timed-presence method resulted in a better performing index than one based on fixed counts. Therefore, inclusion of rare taxa that are widespread but low in abundance appears to further differentiate reference sites from...
impaired sites by increasing index precision and better detecting the loss of sensitive rare species, as proposed by Cao et al. (2007).

Furthermore, the timed-presence species O/E index performed better than previously-published diatom O/E indices, particularly in terms of responsiveness to anthropogenic stress. A diatom O/E index developed by Ritz (2010) was less precise (SD = 0.22) and showed essentially no responsiveness to environmental degradation. Cao et al. (2007) reported a slightly more precise diatom index than ours (SD = 0.17), but they noted that the index failed to show a response to general stress. Finally, a diatom index reported by Carlisle et al. (2008) also had a slightly higher precision than ours (SD = 0.16), but it also suffered from low responsiveness. That index primarily predicted tolerant, unresponsive taxa in reference sites, many of which were not correspondingly lost in test sites. To be clear, responsiveness of an index varies across studies due to variation in reference and test site quality, so it is difficult to fairly compare across indices. Our results comparing O/E indices built using the traditional method of producing diatom data (i.e., low fixed counts) to the timed-presence method suggest that more consistent detection of rare species using timed presence better characterizes sensitive species loss in non-reference sites. In addition, the timed-presence method using species-level data produced a diatom O/E index nearly as precise as invertebrate indices (Hawkins 2006). To place the importance of precision in context, Van Sickle et al. (2007) found that relatively small increases in precision (reduction in index SD) can result in large increases in index sensitivity for some test sites (a decrease in 0.02 O/E units resulted in a sensitivity increase of up to 25%), but no increase in sensitivity in others. Although our timed-presence species index is slightly less precise than some existing invertebrate indices, its responsiveness to environmental stressors (i.e., % imperviousness) is relatively high, indicating potentially high utility.

Assessments based on metrics

The timed-presence method detects a greater number of species, but higher richness did not always translate to an increase in metric discrimination between reference and test sites. The timed-presence method improved performance
of metrics that rely on taxa that are often low in abundance, such as N-fixing taxa and taxa that prefer low total phosphorous and nitrogen. However, increased detection of taxa that are widespread but low in abundance did not provide an advantage in discriminating between metrics reliant on abundant taxa (taxa that prefer high nutrients, taxa that are indifferent to nutrients, taxa that prefer moderate oxygen saturation, and fresh-brackish water taxa). We think it is significant that most of the metrics we tested were derived from European studies (Lange-Bertalot 1979, Bahls 1993, Van Dam et al. 1994), but the metrics that best distinguished reference from test sites were derived from North American studies (Potapova and Charles 2007, Spaulding et al. 2019). We conclude from these results that both methods are similar in their ability to discriminate reference and disturbed sites in terms of metrics.

Conclusions

Relative or absolute abundance of diatom species has, for many decades, been a standard in biological assessment. This study, though limited to rivers in the southeast US, demonstrates that failure to achieve accurate estimates of diatom biodiversity is related to the widespread use of a low fixed-count protocol. We propose that improved measures of species richness will allow diatom O/E indices to perform as well as invertebrate O/E indices. We urge future work to test whether these results are applicable to rivers and lakes in other regions.

ACKNOWLEDGEMENTS

Author contributions: MAT led the enumeration of samples, identification of taxa, final data analysis, and much of the early manuscript writing. DMC led construction of RIVPACS models, performed initial metrics analysis, and contributed significantly to manuscript refinement. SAP contributed significantly to metric analysis and manuscript writing as well as provided much of the contextual knowledge upon which this study was founded.

Thank you to Ian Bishop and Nicholas Schulte for valuable discussions as well as for assistance identifying taxa. We also thank Alec Camp and Paul Bliznik for laboratory assistance at the Institute of Arctic and Alpine Research, University of Colorado. Finally, thank you to Chuck Hawkins for providing early feedback on our RIVPACS model performance. This material is supported by the United States Geological Survey under Cooperative Agreement #G15AC00104 with the University of Colorado.

LITERATURE CITED

Acker, F., B. Russell, and E. Morales. 1999. Preparation of diatom slides using Naphrax™ mounting medium. Protocol P-13–49. Academy of Natural Sciences of Philadelphia, Patrick Center for Environmental Research, Philadelphia, Pennsylvania.

Almeida, S. F. P., C. Elias, J. Ferreira, E. Tornés, C. Puccinelli, F. Delmas, G. Dör, G. Urbani, S. Marcheggiani, J. Rosebery, L. Mancini, and S. Sabater. 2014. Water quality assessment of rivers using diatom metrics across Mediterranean Europe: A methods intercalibration exercise. Science of the Total Environment 476:768–776.

Bahls, L. L. 1993. Periphyton bioassessment methods for Montana streams. Water Quality Bureau, Department of Health and Environmental Sciences, Helena, Montana.

Bishop, I. W., R. M. Esposito, M. Tyrree, and S. A. Spaulding. 2017. A diatom voucher flora from selected southeast rivers (USA). Phytotaxa 332:101–140.

Cao, Y., and C. P. Hawkins. 2005. Simulating biological impairment to evaluate the accuracy of ecological indicators. Journal of Applied Ecology 42:954–965.

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., and C. P. Hawkins. 2008. Land use and the structure of western US stream invertebrate assemblages: Predictive models and ecological traits. Journal of the North American Benthological Society 27:986–999.

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. Report No. 02-06. Academy of Natural Sciences of Philadelphia, Patrick Center for Environmental Research, Philadelphia, Pennsylvania.

Cholnoky, B. J. 1968. Die ökologie der Diatomeen in Binnengewässern. National Institute for Water Research Council for Scientific and Industrial Research, Pretoria, South Africa.

Clarke, R. T., and J. F. Murphy. 2006. Effects of locally rare taxa on the precision and sensitivity of RIVPACS bioassessment of freshwaters. Freshwater Biology 51:1924–1940.

Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. Ecology 88:2783–2792.

Davies, S., and S. Jackson. 2006. The biological condition gradient: A descriptive model for interpreting change in aquatic ecosystems. Ecological Applications 16:1251–1266.

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. US Geological Survey Circular No. 1112. Sacramento, California.

Hausmann, S., D. F. Charles, J. Gerritsen, and T. J. Belton. 2016. A diatom-based biological condition gradient (BCG) approach for assessing impairment and developing nutrient criteria for streams. Science of the Total Environment 562:914–927.
Hawkins, C. P. 2006. Quantifying biological integrity by taxonomic completeness: Its utility in regional and global assessments. Ecological Applications 16:1277–1294.

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.

Herlihy, A. T., S. G. Pauslen, J. Van Sickle, J. L. Stoddard, C. P. Hawkins, and L. L. Yuan. 2008. Striving for consistency in a national assessment: The challenges of applying a reference-condition approach at a continental scale. Journal of the North American Benthological Society 27:860–877.

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. No. 2015–1095. Reston, Virginia.

Kelly, M., C. Bennett, M. Coste, C. Delgado, F. Delmas, L. Denys, L. Ector, C. Fauville, M. Ferreol, M. Golub, A. Jarlman, M. Kahler, J. Lucey, B. N. Chathain, I. Pardo, P. Pfister, P.-F. Joanna, J. Rosebery, C. Schranz, J. Schaumburg, H. Van Dam, and S. Vilbaste. 2009. A comparison of national approaches to setting ecological status boundaries in phytothemosis assessment for the European Water Framework Directive: Results of an intercalibration exercise. Hydrobiologia 621:169–182.

Kelly, M., S. Juggins, G. Guthrie, S. Pritchard, J. Jamieson, B. Rippey, H. Hirst, and M. Yallop. 2008. Assessment of ecological status in UK rivers using diatoms. Freshwater Biology 53:403–422.

Kelly, M. G., and B. A. Whitton. 1995. The trophic diatom index: A new index for monitoring eutrophication in rivers. Journal of Applied Phycology 7:433–444.

Lange-Bertalot, H. 1979. Pollution tolerance of diatoms as a criterion for water quality estimation. Nova Hedwigia, Beihefte 64:285–304.

Larras, F., and B. Montuelle. 2014. Linking diatom sensitivity to herbicides to phylogeny: A step forward for biomonitoring? Environmental Science & Technology 48:1921–1930.

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. Ecological Indicators 102:166–174.

Maguran, A. E. 2004. The commonness, and rarity, of species. Pages 18–71 in Measuring biological diversity. Blackwell Publishing, Carleton, Victoria, Australia.

Moss, D., M. T. Furse, J. F. Wright, and P. D. Armitage. 1987. The prediction of the macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. Freshwater Biology 17:41–52.

Moulton II, S. R., 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. US Geological Survey Open-File Report 02-150. Reston, Virginia.

Ostermiller, J., and C. Hawkins. 2009. Effects of sampling error on bioassessments of stream ecosystems: Application to RIVPACS-type models. Journal of the North American Benthological Society 23:363–382.

Pan, Y., R. J. 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 76–95 in J. Cairns and K. Dickson (editors). Biological methods for the assessment of water quality. American Society for Testing and Materials International, West Conshohocken, Pennsylvania.

Patrick, R., and D. Strawbridge. 1963. Methods of studying diatom populations. Journal of the Water Pollution Control Federation 35:151–161.

Patrick, R. M., and C. W. Reimer. 1966. The diatoms of the United States (exclusive of Alaska and Hawaii). Monographs of the Academy of Natural Sciences of Philadelphia 13.

Potapova, M., and D. F. Charles. 2007. Diatom metrics for monitoring eutrophication in rivers of the United States. Ecological Indicators 7:47–70.

Ritz, C. 2010. Effectiveness of RIVPACS predictive models to evaluate diatom response to nutrient stress in coastal California streams. MS thesis. California State University, Monterey Bay, California.

Schueler, T. R., L. Fraley-McNeal, and K. Cappiella. 2009. Is imperious cover still important? Review of recent research. Journal of Hydrologic Engineering 14:309–315.

Spaulding, S. A., I. W. Bishop, M. B. Edlund, S. Lee, and M. Potapova. 2019. Diatoms of North America. (Available from: http://diatoms.org)

Spaulding, S. A., and E. Elwell. 2007. Increase in nuisance blooms and geographic expansion of the freshwater diatom *Didymosphenia geminata*. US Geological Survey open-file report no. 2007–1425. US Departement of the Interior, US Geological Survey, Reston, Virginia. (Available from: https://pubs.usgs.gov/of/2007/1425/report.pdf)

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, R. J., 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, Cambridge, United Kingdom.

Stevenson, R. J., and C. G. Peterson. 1991. Emigration and immigration can be important determinants of benthic diatom assemblages in streams. Freshwater Biology 26:279–294.

Stoddard, J. L., A. T. Herlihy, D. V. Peck, R. M. Hughes, T. R. Whittier, and E. Tarquinio. 2008. A process for creating multimetric indices for large-scale aquatic surveys. Journal of the North American Benthological Society 27:878–891.

Tang, T., R. J. Stevenson, and D. M. Infante. 2016. Accounting for regional variation in both natural environment and human
disturbance to improve performance of multimetric indices of lotic benthic diatoms. Science of the Total Environment 568:1124–1134.

USEPA (United States Environmental Protection Agency). 2009. National rivers and streams assessment: Field operations manual. EPA-841-B-07-009. Washington DC. (Available from: https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100AVOF.txt)

Van Dam, H., A. Mertens, and J. Sinkeldam. 1994. A coded checklist and ecological indicator values of freshwater diatoms from The Netherlands. Netherlands Journal of Aquatic Ecology 28:117–133.

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.

Van Sickle, J., D. P. Larsen, and C. P. Hawkins. 2007. Exclusion of rare taxa affects performance of the O/E index in bioassessments. Journal of the North American Benthological Society 26:319–331.