Consideration of species-specific diatom indicators of anthropogenic stress in the Great Lakes

Euan D. Reavie*, Meijun Cai

Natural Resources Research Institute, University of Minnesota Duluth, Duluth, Minnesota, United States of America

* ereavie@d.umn.edu

Abstract

Robust inferences of environmental condition come from bioindicators that have strong relationships with stressors and are minimally confounded by extraneous environmental variables. These indicator properties are generally assumed for assemblage-based indicators such as diatom transfer functions that use species abundance data to infer environmental variables. However, failure of assemblage approaches necessitates the interpretation of individual dominant taxa when making environmental inferences. To determine whether diatom species from Laurentian Great Lakes sediment cores have the potential to provide unambiguous inferences of anthropogenic stress, we evaluated fossil diatom abundance against a suite of historical environmental gradients: human population, agriculture, mining, atmospheric nutrient deposition, atmospheric temperature and ice cover. Several diatom species, such as *Stephanodiscus parvus*, had reliable relationships with anthropogenic stress such as human population. However, many species had little or no indicator value or had confusing relationships with multiple environmental variables, suggesting one should be careful when using those species to infer stress in the Great Lakes. Recommendations for future approaches to refining diatom indicators are discussed, including accounting for the effects of broad species geographic distributions to minimize region-specific responses that can weaken indicator power.

Introduction

In a report on 50 priority research questions in paleoecology [1], two of the priorities were as follows and are the subject of the current analysis: What methods can be used to develop more robust quantitative paleoenvironmental reconstructions? How can paleoecologists disentangle the separate and combined effects of multiple causal factors in paleoecological records?

Knowing indicator properties of species is critical for robust inferences of environmental condition in monitoring programs and paleolimnological applications. In figuring out these species characteristics, the present-day distributions of diatoms (or any of several indicator organisms) are typically calibrated across the gradient of a selected environmental parameter.
In Anthropocene studies these variables tend to be analytes related to stressors such as nutrient enrichment (e.g. phosphorus), acidification (e.g. pH) and warming (e.g. salinity) [2]. These tools are usually assemblage-based and use a species weighted-average approach and have consistently provided valuable tools to infer past conditions from sedimentary records [3]. For instance, Potapova & Charles [4] derived species conductivity optima from over 3000 benthic diatom samples. While valuable, water quality variables can have uncertain meaning, such as conductivity which can reflect pollution or a broad suite of natural factors (e.g. carbonate bedrock and soils). Monitoring programs often use biological indices such as the diatom-based Trophic Diatom Index (TDI) [5], which assigns a single environmental condition based on a weighted compilation of species characteristics for all taxa within a sample. Such distilled indices are desirable for aquatic management, and there is little question that strong relationships often exist between the abundance and structure of diatom communities and prevalent stressor variables. However, the limitations of aggregate community metrics were recognized by King & Baker [6], who emphasized that it is critically important to know which taxa are affected by a stressor of interest. This is especially important when aggregate metrics fail, such as recently occurred when applying a phosphorus transfer function to fossil diatom assemblages in Lake Huron [7].

It has long been known that diatoms are powerful indicators of environmental condition and change [8]. When we examine modern diatom assemblages and fossil profiles, there is a desire to have clearly-defined stressor responses for our indicator species so that we may make clear interpretations of past conditions. This is challenging for many species due to their simultaneous responses to many environmental variables. Take *Fragilaria crotonensis* as an example. This pennate, phytoplanktonic diatom is common in the Laurentian Great Lakes and tends be most common during summer months [9]. Based on Reavie et al. [9], this taxon has relatively high optimum for phosphorus and turbidity and a low optimum for nitrogen. It has been noted in paleoecological studies as increasing in relative abundance following artificial phosphorus enrichment resulting in cultural eutrophication [10]. In an experimental study, *F. crotonensis* increased in response to nitrogen (and not silica or phosphorus) enrichment [11], an observation backed by observations of natural assemblages from Lake Michigan [12]. It is apparently also positively supported by water column mixing that enables heavy colonies to remain suspended [13]. Clearly, the multi-variable context of an aquatic environment has a critical bearing on the abundance of *F. crotonensis* and what it indicates relative to anthropogenic stressors. Such dynamics make indicator interpretations challenging, and furthermore, uncertainties in explaining the presence of a species is subject to unexplained noise in the data, which is expected when one deals with highly complex biological data such as diatom assemblages. These complicated indicator relationships likely contribute to poor performance of diatom-based indicator models, something that has been highlighted relatively recently [14].

A useful index requires accurate environmental calibration of species. However, geographic constraints can limit the large sample size needed to adequately calibrate diatom indicators [15]. Training sets of samples used to calibrate diatom indicators are preferably selected within a narrow range of physicochemical conditions that at the same time maximize the gradient of the stressor variable. Such an ideal dataset is rarely possible, so it is typical to include deep and shallow lakes that, respectively, may favor phytoplanktonic or benthic species, which is fine if one hopes to reconstruct changing lake depth and not some chemical parameter. Variations in catchment soil and bedrock determine ionic properties of aquatic systems, which in turn naturally drive assemblage properties, a phenomenon that determines diatom assemblages throughout the Laurentian Great Lakes basin [9]. Hence, the natural context of water quality can limit how a species responds to stress, if it even occurs at all, under certain natural conditions. Such geographic considerations have been applied to biotic indicators on continental
scales in Europe [16], and are likely a major confounding factor in diatom indicator development. To our knowledge corrections for geographic context have not been applied in diatom-based transfer functions in North America.

Due to their significance as water resources, great lakes worldwide are becoming subjects of paleolimnological evaluations aimed at long-term management (e.g. [17,18]). Hence, robust metrics derived from fossil records will be increasingly used to support monitoring and remedial programs. Aggregate metrics such as transfer functions and biological indices will continue to be undermined by multivariate determinants of diatom occurrence, so it is likely that species-specific interpretations will continue to be required for useful inferences of condition. We aimed to determine whether diatom species exist that have clear relationships with stress. Investigation was based on a suite of fossil diatom data and historical stressor data on population growth, agriculture, deforestation and climate change in the Laurentian Great Lakes Basin. Consideration is given to the importance of spatial variation and whether it can be overcome in the pursuit of indicator development.

Methods
Diatom dataset

Sediment cores were collected from 10 locations throughout the Great Lakes, as detailed by Reavie et al. [19]. Cores were age dated using $^{210}$Pb, and details of those age models are provided by Reavie et al. [20] (in review). Diatom frustules in sediment samples were cleaned and prepared for light microscopy. Diatoms were identified and enumerated at 1000–1250 X magnification with oil immersion [21].

For every sediment sample analyzed for a diatom assemblage, we generated associated stressor data for human population, agriculture, forest cover and atmospheric variables (annual minimum temperature, water level, $\text{NH}_4^+$, $\text{NO}_3^-$, inorganic N and chloride). Historical data for a given sample were selected based on the $^{210}$Pb-inferred date in the middle of the sample interval. The long-term stressor data (S1 Table) were compiled several ways:

- Quantitative data for agriculture, forests and population were retrieved from a collection of historical data compiled by Reavie et al. [22]. These data were compiled from various Canadian and US historical records and extrapolated, if necessary, into yearly estimates of stressor values going back to 1780. For years with no data linear interpolations between adjacent years in a given watershed were used to fill data gaps. This stressor database is organized according to 60 watersheds comprising the Great Lakes watershed. Stressors for a given core location were associated according to adjacent sub-watersheds as depicted in Fig 1.

- Historical monthly minimum, maximum and mean air temperature data from mid 1800s to present were summarized for each of five Great Lakes by averaging the monthly minimum, maximum and mean data of meteorological stations assigned to each lake. Meteorological stations relevant to each lake were selected by Hunter et al. [23]. The monthly minimum, maximum and mean air temperature of each station was computed from daily data. Daily air temperature data were downloaded from the National Climatic Data Center (NCDC) [24] of the National Oceanic and Atmospheric Administration (NOAA) and from Canadian Weather [25]. The monthly minimum and maximum temperatures were calculated by taking the smallest and largest daily temperatures within that month. Because NCDC weather data did not give daily mean temperatures, the daily mean temperature was calculated by averaging daily minimum and maximum temperatures. Although we compiled these variations of temperature data, we ultimately selected analyses using annual minimum temperature (°C) due to its apparent dominant effect on phytoplankton communities in the Great
Lakes [19] and because, worldwide, it is the most rapidly increasing temperature parameter [26].

- The atmospheric deposition monitoring stations within 100 km distance from a given lake’s shoreline were selected to calculate loading relevant to that lake. Annual wet deposition loadings of ammonium, nitrate, inorganic nitrogen and chloride were downloaded from National Atmospheric Deposition Program (NADP) [27] and EPA Clean Air Status and Trends Network (CASTNET) [28] stations from 1979 to 2014. A total of 52 monitoring stations (34 currently active stations and 18 historical stations) were relevant to the 100-km

![Fig 1. Maps of the Laurentian Great Lakes catchment, indicating the sub-watersheds selected to relate stressors with the sedimentary record in each sediment core (circle symbols). Delineation of sub-watersheds and quantitative stressor data are available from Reavie et al. (2018).](https://doi.org/10.1371/journal.pone.0210927.g001)
Great Lakes buffer. Some stations were within 100 km of more than one lake. The loadings for a lake in kg/ha were calculated by using the following distance weighted-average formula:

$$\sum_{i=1}^{n} \frac{x_i}{D_i} / n$$

where $n$ is the number of stations within 100 km of the lake, $x_i$ is the loading at station $i$ and $D_i$ is the distance of station $i$ from the lake shoreline.

- Monthly ice cover concentration and annual ice duration of winter (December to May) maps from 1973 to 2013 were obtained from GLAHF [29]. GIS maps were computed in ArcMap using the zonation tool to get mean monthly ice cover area and annual duration for each lake. The monthly ice cover from December (of the previous year) through May were averaged to get mean % ice cover.

- Annual average water levels (water level anomaly in ft) from 1860 to 2014 were downloaded from EPA [30] archives.

**Data analysis**

To evaluate species-stressor relationships, sample relative abundances of individual diatom species (at least three non-zero occurrences) were related to historical stressor data using Pearson correlation (linear) and Gaussian modeling (quadratic). The latter method is based on the well-founded assumption that diatoms respond non-monotonically along environmental gradients. Using the R base package [31] the function `lm` was used to evaluate linear correlation followed by the use of the `glm` function (specifying the Gaussian’ family of distributions) to evaluate a quadratic fit. Based on significance of correlation ($P < 0.05$; linear $r$ or Gaussian $r^2$) the method with the lower $P$ was selected as the optimal model. For Gaussian relationships, a species’ optimum to a given variable was estimated as the maximum value of the Gaussian model. For models with Gaussian peaks that occurred outside the range of environmental measurements, minimum or maximum values (S1 Table) were used as appropriate to substitute for optima. This was first done using raw stressor data; then, for evaluations against temperature, atmospheric deposition and water level data, the analysis was performed again using standardized data. Standardization included transforming the stressor (x axis) and diatom relative abundance (y axis) data from a specific lake so that their standard deviations ranged from 0 to 1. After lake-specific standardization, the data were re-pooled for analysis. Standardization was deemed necessary due to major geographic differences in assemblage and stressor data. For instance, Lake Superior contains high relative abundances of cyclotelloid centric diatoms, more so in recent years, apparently due to recent warming [26]. However, because Lake Superior has always had high abundances of cyclotelloids, and because it is the coldest of the Great Lakes, regressions using raw data suggested that colder atmospheric temperatures favored higher relative abundances of Cyclotella. This was contrary to findings that increasing temperatures in a given lake tend to favor cyclotelloids. Hence, standardization by lake allowed correlation analyses to account for historical changes noted in paleolimnological records. The debatable appropriateness of this standardization is considered further in the Discussion section.

**Results**

Correlations with stressor variables identified several species likely to be good indicators of stress. For the strictly anthropogenic stressors (population, agriculture, deforestation, mining), Discostella pseudostelligera (CYCPSEU) is ranked the highest, having the strongest relationship (lowest $P$) with low agricultural intensity around the Great Lakes (Table 1). This is followed by
Table 1. Pearson correlations ($r$) and Gaussian squared correlation coefficients (Gauss. $r^2$(adj)) of diatom relative abundance versus anthropogenic stressors. Gaussian optima are marked as to whether they are lower than the minimum (-) or higher than the maximum (+) stressor values for Great Lakes catchments (i.e. actual optima are lower or higher, respectively). The top 25 taxon-stressor relationships are presented, sorted by smallest $P$ value. $N$ = number of observations in samples. A full list of the significant taxon-stressor relationships is in S2 Table.

| Diatom code | Full name | Variable | Best model | Linear $r$ | Gauss. $r^2$(adj) | Gauss. optimum | $P$       | $N$ |
|-------------|-----------|----------|------------|-----------|--------------------|----------------|----------|-----|
| CYCPSEU     | Discostella pseudostelligera (Hust.) Houk & Klee | Agriculture | Gauss. | 0.61 | 194.4 (-) | 3.3E-85 | 249 |
| STEPSBTR    | Stephanodiscus subtransylvanicus Gasse | Agriculture | Gauss. | 0.54 | 194.4 (-) | 1.7E-63 | 249 |
| SYNFLIE     | Syndra filiformis var. exilis A. Cl. | Agriculture | Gauss. | 0.49 | 194.4 (-) | 3.7E-52 | 249 |
| AULGRAN     | Aulacoseira granulata (Ehr.) Simonsen | Population | Gauss. | 0.48 | 9884672 (+) | 5.2E-52 | 252 |
| STEALPI     | Stephanodiscus alpinus Hust. | Population | Gauss. | 0.48 | 7616075 | 2.2E-51 | 252 |
| SYNOSTE     | Syndra ostenfeldii (Krieg.) A. Cl. | Agriculture | Gauss. | 0.48 | 194.4 (-) | 1.4E-50 | 249 |
| CYCOCEL     | Lindavia ocellata (Pant.) Nakov, Guilloy, M.L. Julius, E.C. Ther. and A.J. Alverosn | Population | Gauss. | 0.43 | 52.2 (-) | 1.2E-42 | 252 |
| ACHMINUS    | Achnanthes minutissima var. scotica (Carter) Lange-Bertalot | Agriculture | Gauss. | 0.43 | 194.4 (-) | 4.0E-41 | 249 |
| FRABREV     | Pseudostaurosira brevistriata (Grunowe) D.M. Williams & Round | Population | Gauss. | 0.41 | 9884672 (+) | 7.6E-39 | 252 |
| STEPARV     | Stephanodiscus parvus Stoerm. & Hak. | Population | linear | 0.70 | 8.0E-39 | 252 |
| DIATENU     | Diatoma tenuis C.Agardh | Population | Gauss. | 0.39 | 9884672 (+) | 2.7E-36 | 252 |
| CYCPSEU     | Discostella pseudostelligera (Hust.) Houk & Klee | Population | Gauss. | 0.38 | 52.2 (-) | 2.0E-34 | 252 |
| CYCCOMRC    | Lindavia delicatula (Hust.) Nakov, Guilloy, M.L. Julius, E.C. Ther. & A.J. Alverosn | Agriculture | Gauss. | 0.38 | 194.4 (-) | 1.0E-33 | 249 |
| RHERIE      | Urosolenia eriensis (H.L. Sm.) Round & R.M. Crawford | Agriculture | Gauss. | 0.35 | 194.4 (-) | 5.8E-30 | 249 |
| DIPSPE      | Diploneis sp. | New mines | Gauss. | 0.34 | 10.0 (+) | 1.9E-29 | 254 |
| STEPARV     | Stephanodiscus parvus Stoerm. & Hak. | Forestry | Gauss. | 0.34 | 722974 (-) | 7.9E-29 | 253 |
| SURMINU     | Surirella minuta Bréb. ex Kütz. | Population | Gauss. | 0.33 | 9884672 (+) | 1.4E-28 | 252 |
| CYCOCEL     | Lindavia ocellata (Pant.) Nakov, Guilloy, M.L. Julius, E.C. Ther. and A.J. Alverosn | Agriculture | Gauss. | 0.34 | 194.4 (-) | 1.6E-28 | 249 |
| STEPSBTR    | Stephanodiscus subtransylvanicus Gasse | Population | Gauss. | 0.33 | 52.2 (-) | 1.1E-27 | 252 |
| ACHMINUS    | Achnanthes minutissima var. scotica (Carter) Lange-Bertalot | Population | Gauss. | 0.32 | 52.2 (-) | 2.3E-27 | 252 |
| FRAINTEF    | Fragilaria intermedia var. fallax (Grun.) Stoerm. & Yang | Forestry | Gauss. | 0.32 | 19784219 (+) | 2.7E-27 | 253 |
| FRACAPR     | Fragilaria capucina var. rumpens (Kützing) Lange-Bertalot ex Bukht. | New mines | Gauss. | 0.32 | 10.0 (+) | 3.1E-27 | 254 |
| SYNOSTE     | Syndra ostenfeldii (Krieg.) A. Cl. | Population | Gauss. | 0.32 | 52.2 (-) | 4.7E-27 | 252 |
| CYCATOM     | Cyclotella atomus Hust. | Agriculture | Gauss. | 0.32 | 194.4 (-) | 1.2E-26 | 249 |
| RHIGRAC     | Rhizosolenia gracilis H.L. Smith | Agriculture | Gauss. | 0.32 | 194.4 (-) | 1.3E-26 | 249 |

https://doi.org/10.1371/journal.pone.0210927.t001
Stephanodiscus subtransylvanicus and Synedra filiformis var. exilis, also associated with lower agriculture. Due to their high relative abundance in the earliest years represented in our sediment cores these taxa had negative relationships with population and agriculture and a positive relationship with forestry (S2 Table). Many taxa indicated stress tolerance as suggested by having positive relationships with population (Aulacoseira granulata, Pseudostaurosira brevistriata, Stephanodiscus parvus). A total of 270 significant (with Bonferroni correction) species-stressor relationships occurred, 69 for agriculture, 54 for deforestation, 45 for mining stress, 39 for new mines and 63 for population. Using an uncorrected P (0.05), 306 significant relationships occurred. 91.3% (37.9% with Bonferroni correction) of the 335 taxa with sufficient occurrences for analysis 127 had significant relationships with at least one stressor.

For the stressor variables with natural components (chemical deposition and atmospheric temperatures), Discostella pseudostelligera had the strongest, negative relationship with minimum annual temperature (Table 2). The next highest ranked were Lindavia delicatula (with a chloride deposition optimum of 0.27 kg/ha), three species that were associated with low minimum annual temperature (Stephanodiscus subtransylvanicus, Synedra ostenfeldii, Lindavia ocellata) and Aulacoseira islandica which was associated with the highest nitrate deposition. A total of 709 significant (206 with Bonferroni correction) relationships occurred, 95 for chloride, 99 for inorganic N, 94 for water level, 45 for ice extent, 91 for ammonium, 97 for nitrates and 188 for minimum annual temperature.

Lake-based normalization of fossil and historical monitoring data resulted in a shift in the indicator properties of several diatom species and resulted in fewer significant relationships with stressors (Table 3), indicating geographical specifics can influence interpretations of condition. With Bonferroni correction applied, only 22 taxa had significant correlations, most with minimum annual temperature (S2 Table). As anticipated, several positive relationships were observed between minimum annual temperature and taxa belonging to Cyclotella sensu lato. Examining the regressions on two select taxa (Fig 2) indicates the effect of within-lake data standardization. The relationship between Lindavia laurentiana and temperature became stronger after standardization, whereas the relationship for Lindavia delicatula reversed, becoming positive and corresponding with that observed in paleorecords (i.e. increasing in relative abundance with higher atmospheric temperatures) [26].

Discussion

We identified several diatom taxa that individually reflect stress, or lack of stress, in the Great Lakes. In general, taxon-stressor relationships with the lower P and higher r values (Tables 1–3) can be considered better indicators. While there are too many taxa to discuss, high-ranking species like S. parvus and D. pseudostelligera should be considered good indicators of stress associated with population growth, agriculture and deforestation in the catchment of an aquatic system. S. parvus is a relatively consistent indicator of poorer conditions while D. pseudostelligera appears to reflect good or pre-impact conditions (lower population, agriculture and atmospheric temperatures) in the Great Lakes. Such a finding is not necessarily novel for a high-nutrient indicator such as S. parvus (which we select as a case example because of its prevalence in freshwater paleolimnology literature), but our results suggest this indicator status is minimally confounded by environmental variability throughout the Great Lakes. As one steps down the ranked lists of species, greater caution should be applied when associating a taxon with a particular stressor.

Diatoms have existed for millions of years, well before freshwater assemblages began to shift due to human drivers. Hence, unlike synthetic fossil indicators (such as PCBs), a diatom species cannot explicitly indicate anthropogenic stress. However, it is possible that pre-impact
Table 2. Pearson correlations ($r$) and Gaussian squared correlation coefficients (Gauss. $r^2$(adj)) of diatom relative abundance versus measured environmental variables considered to represent a mixture of natural and anthropogenic components. Gaussian optima are marked as to whether they are lower than the minimum (-) or higher than the maximum (+) stressor values for Great Lakes catchments (i.e. actual optima are lower or higher, respectively). The top 25 taxon-stressor relationships are presented, sorted by smallest P value. N = number of observations in samples. A full list of the significant taxon-stressor relationships is in S2 Table.

| Diatom code | Full name | Variable | Best model | Linear $r$ | Gauss. $r^2$(adj) | Gauss. optimum | P       | N   |
|-------------|-----------|----------|------------|-------------|-------------------|----------------|--------|-----|
| DISPSEU     | Discostella pseudostelligera (Hustedt) Houk & Klee | Min. annual temp. | Gauss. | 0.56 | -15.75 (-) | 2.99E-65 | 236 |
| CYCCOMRC    | Lindavia delicatula (Hust.) Nakov, Guillory, M.L.Julius, E.C. Ther. & A.J.Alerson | CI deposition | Gauss. | 0.65 | 0.27 (-) | 3.75E-50 | 122 |
| STEPSBTR    | Stephanodiscus subtransylvanicus Gasse | Min. annual temp. | Gauss. | 0.46 | -15.75 (-) | 2.51E-45 | 236 |
| AULISLA     | Aulacoseira islandica (O. Mull.) Simonsen | Nitrate deposition | Gauss. | 0.63 | 28.97 (+) | 2.78E-45 | 122 |
| SYNOSTE     | Synedra ostenfeldii (Krieg.) A. Cl. | Min. annual temp. | Gauss. | 0.43 | -15.75 (-) | 4.44E-39 | 236 |
| CYCCOMRC    | Lindavia delicatula (Hust.) Nakov, Guillory, M.L.Julius, E.C. Ther. & A.J.Alerson | Inorganic N deposition | Gauss. | 0.55 | 1.57 (-) | 5.04E-33 | 122 |
| CYCCOMRC    | Lindavia delicatula (Hust.) Nakov, Guillory, M.L.Julius, E.C. Ther. & A.J.Alerson | Nitrate deposition | Gauss. | 0.53 | 4.16 (-) | 3.67E-31 | 122 |
| ACHMINUS    | Achnanthes minutissima var. scotica (Carter) Lange-Bertalot | Min. annual temp. | Gauss. | 0.36 | -15.75 (-) | 2.07E-29 | 236 |
| AULISLA     | Aulacoseira islandica (O. Mull.) Simonsen | Inorganic N deposition | Gauss. | 0.50 | 10.36 (+) | 6.94E-28 | 122 |
| STEPARV     | Stephanodiscus parvus Stoerm. & Hak. | Min. annual temp. | Gauss. | 0.34 | -2.76 | 6.17E-27 | 236 |
| RHIGRAC     | Rhizosolenia gracilis H.L. Smith | Min. annual temp. | Gauss. | 0.32 | -15.75 (-) | 3.00E-25 | 236 |
| STEALP1     | Stephanodiscus alpinus Type I | CI deposition | Gauss. | 0.46 | 1.99 | 6.15E-24 | 122 |
| RHIERIE     | Urosolenia eriensis (H.L.Sm.) Round & R.M.Crawford | Min. annual temp. | Gauss. | 0.29 | -15.75 (-) | 4.31E-22 | 236 |
| TABFLOC3    | Tabellaria flocculosa str. 3p | Min. annual temp. | Gauss. | 0.29 | -15.75 (-) | 1.15E-21 | 236 |
| TABFENE     | Tabellaria fenestrata (Lyngh.) Kutz. | Min. annual temp. | Gauss. | 0.28 | -15.75 (-) | 2.30E-21 | 236 |
| AULGRAN     | Aulacoseira granulata (Ehr.) Simonsen | Min. annual temp. | Gauss. | 0.27 | -2.76 (+) | 2.73E-20 | 236 |
| SYNFILIE    | Synedra filiformis var. exilis A. Cl. | Min. annual temp. | linear | -0.55 | | 9.34E-20 | 236 |
| STEALP1     | Stephanodiscus alpinus Type I | Nitrate deposition | linear | 0.70 | 2.50E-19 | 122 |
| DISPSEU     | Discostella pseudostelligera (Hustedt) Houk & Klee | Inorganic N deposition | Gauss. | 0.40 | 1.57 (-) | 2.10E-18 | 122 |
| CYCCOMRC    | Lindavia delicatula (Hust.) Nakov, Guillory, M.L.Julius, E.C. Ther. & A.J.Alerson | Ammonium deposition | Gauss. | 0.39 | 0.81 (-) | 2.77E-18 | 122 |
| STEALP1     | Stephanodiscus alpinus Type I | Inorganic N deposition | linear | 0.68 | | 5.74E-18 | 122 |
| DISPSEU     | Discostella pseudostelligera (Hustedt) Houk & Klee | Ammonium deposition | Gauss. | 0.39 | 0.81 (-) | 1.31E-17 | 122 |
| STECONSP    | Stephanodiscus conspicuousporus Stoermer, Hakansson & Theriot | Inorganic N deposition | Gauss. | 0.38 | 1.57 (-) | 2.31E-17 | 122 |
| SYNOSTE     | Synedra ostenfeldii (Krieg.) A. Cl. | Inorganic N deposition | Gauss. | 0.37 | 1.57 (-) | 1.10E-16 | 122 |

https://doi.org/10.1371/journal.pone.0210927.t002
Table 3. Pearson correlations (r) and Gaussian quadratic squared correlation coefficients (Gauss. r^2(adj)) of diatom relative abundance versus measured physical parameters and chemical atmospheric deposition. For this analysis diatom and environmental data were normalized by lake so that SDs ranged from 0 to 1. The top 25 taxon-stressor relationships are presented, sorted by smallest P value. N = number of observations in samples. A full list of the significant taxon-stressor relationships is in S2 Table.

| Diatom code | Full name | Variable | Best model | Linear r Gauss. r^2(adj) | Gauss. optimum | P         | N   |
|-------------|-----------|----------|------------|--------------------------|----------------|-----------|-----|
| CYCCOMES    | Lindavia comensis (Grunow) Nakov, Guillory, M.L.Julius, E.C. Ther. & A.J.Alverson | Min. annual temp. | linear | 0.53 | 4.01E-18 | 236 |
| CYCATOMF    | Cyclotella atomus “fine form” | Min. annual temp. | linear | 0.46 | 9.72E-14 | 236 |
| STECONSP    | Stephanodiscus conspicuopus Stoermer, Håkansson & Theriot | Min. annual temp. | linear | -0.47 | 1.79E-11 | 180 |
| NAVRHYN     | Navicula rhynchotella Lange-Bert. | Min. annual temp. | quadratic | 0.34 | 2.99 | 7.03E-11 | 89  |
| CYCOMMEI    | Lindavia laurentiana Alexson, Wellard Kelly, Estepp & Reavie | Min. annual temp. | linear | 0.38 | 1.33E-09 | 236 |
| CYCOMMRC    | Lindavia delicatula (Hust.) Nakov, Guillory, M.L.Julius, E.C. Ther. & A.J.Alverson | Min. annual temp. | linear | 0.41 | 2.26E-09 | 193 |
| TABFLOC     | Tabellaria flocculosa (Roth) Knud. | Min. annual temp. | linear | -0.39 | 4.40E-09 | 207 |
| COCPEDI     | Cocconeis pediculus Ehr. | Min. annual temp. | quadratic | 0.22 | 3.19 | 2.53E-08 | 118 |
| FRACAPUM    | Fragilaria mesolepta Rabenh. | Ammonium deposition | quadratic | 0.39 | -2.54 | 3.53E-08 | 52  |
| UNICENT     | Unidentified Centrales | Min. annual temp. | quadratic | 0.18 | -2.41 | 4.70E-08 | 147 |
| CYCAUXOS    | Cyclotella sp. auxospore | Min. annual temp. | quadratic | 0.34 | 3.19 | 1.23E-07 | 60  |
| CYCRADII    | Cyclotella aff. radiosa JCU | Water level | quadratic | 0.34 | -3.00 | 1.46E-07 | 57  |
| FRAVAUC     | Fragilaria vaucheriae (Kutz.) Peters. | CI deposition | quadratic | 0.27 | 2.92 | 7.71E-07 | 70  |
| FRACAPUM    | Fragilaria mesolepta Rabenh. | Inorganic N deposition | quadratic | 0.33 | -2.92 | 1.46E-06 | 52  |
| CYCBODA     | Lindavia bodanica (Eidenst. Ex Grunow) Nakov, Guillory, M.L. Julius, E.C. Ther. and A.J.Alverson | Min. annual temp. | linear | -0.44 | 1.54E-06 | 109 |
| STE10JCU    | Stephanodiscus sp. 10 JCU | Nitrate deposition | quadratic | 0.58 | 1.14 | 2.46E-06 | 18  |
| SYNFFL      | Synedra filiformis Grun. | Min. annual temp. | linear | 0.30 | 3.52E-06 | 236 |
| NITBACI     | Nitzschia bacillum Hust. | Water level | quadratic | 0.29 | -3.00 | 7.48E-06 | 54  |
| STESP10     | Stephanodiscus sp. #10 in house taxon | Min. annual temp. | linear | -0.28 | 3.43E-05 | 211 |
| PRAELLI     | Pseudoaestuastra elliptica (Schum.) Edlund, E.Morales and S. A.Spaull. | Min. annual temp. | linear | 0.53 | 3.48E-05 | 54  |
| CYCCOMES    | Lindavia comensis (Grunow) Nakov, Guillory, M.L.Julius, E.C. Ther. & A.J.Alverson | Max. annual ice extent | linear | -0.35 | 3.55E-05 | 132 |
| SYNFFL      | Synedra filiformis var. exilis A. Cl. | Min. annual temp. | linear | 0.38 | 3.60E-05 | 113 |
| STE10JCU    | Stephanodiscus sp. 10 JCU | Water level | quadratic | 0.42 | 1.69 | 5.47E-05 | 25  |
| PRAELLI     | Pseudoaestuastra elliptica (Schum.) Edlund, E.Morales and S. A.Spaull. | Water level | linear | -0.52 | 6.07E-05 | 54  |
| NITANGU     | Tryblionella angustata W.Sm. | Ammonium deposition | quadratic | 0.23 | 1.91 | 6.60E-05 | 59  |

https://doi.org/10.1371/journal.pone.0210927.t003
paleorecords may be used to strengthen species indicator characteristics. For instance, did *S. parvus* exist in high abundance prior to the Anthropocene? A lack of such a record would strengthen the characterization of *S. parvus* as an indicator of stress. In Great Lakes containing *S. parvus* in modern records, the relative abundance of *S. parvus* was consistently low in fossil records prior to the 20th century [32,33]. While not a complete assessment of the paleolimnology literature, a review of several manuscripts where the species was observed [34–37] indicates that *S. parvus* was not present in significant numbers prior to human population expansion and associated nutrient enrichment of aquatic systems. Even in a Swiss lake with a long record of human impacts, *S. parvus* did not achieve notable abundance until catchment pasturing as early as the 14th century [38]. Indeed, all of the Great Lakes now contain dominant species that were not previously observed, or existed in trace amounts, prior to the Anthropocene. Several species in the *Cyclotella sensu lato* complex, especially *Lindavia delicatula* and *Lindavia comensis*, are clearly dominant due to human changes to Great Lakes conditions [26]. While outside the scope of this paper, such scanning of available literature may be a worthwhile method to qualify the indicator properties of species.

Although we provide some clarifications, confirming an independent response of a species to stress remains challenging. A species’ consistent response to stressors such as population, agriculture and deforestation is helpful, but most taxa reveal relationships with variables with

---

**Fig 2.** Temperature-diatom relationships for two cyclotelloid species in the Great Lakes. The upper plots use the original temperature and percent abundance data, and lower plots use data that have been standardized by lake and recombined.

https://doi.org/10.1371/journal.pone.0210927.g002
counterintuitive meaning. For example, based on correlations, *Achnanthidium minutissimum* is associated with a mixture of stressors: low temperatures, agriculture, population and nitrogen deposition, and seemingly in conflict, high mining influence. This may be a result of intra-taxon complexity, as *A. minutissimum* contains a complex of varieties that are difficult to discern using light microscopy and so it may contain many taxa with their own unique environmental optima. Bradshaw and Anderson [39] noted that the performance of a diatom-based transfer function for phosphorus was strongly driven by one species, *Cyclostephanos dubius*, which has a high phosphorus optimum. Subsequent analyses identified a strong relationship between *C. dubius* abundance and nitrate in five Danish lakes. In a case like this, inferences of productivity may be sufficient since these two variables tend to co-vary. But what if the variables were phosphorus and pH? Such a finding would complicate interpretations of condition because those variables typically indicate different stress types. Saros [40] summarizes several examples of experiments indicating the confounding effects of multivariate control over species abundance has on the responses of those species to an environmental variable one may wish to reconstruct. Further analysis of the independence of species relationships with a given stressor may help further refine these indicators, but a more detailed set of measured variables than available to us would be needed (e.g. long-term records of phosphorus, conductivity, grazer and filter feeder abundance, water clarity).

Where do we go from here? As recommended by Saros [40], additional indicators (whether or not they employ diatoms) should be considered. We have focused on diatom taxonomic composition of samples, but we reiterate what other researchers have said [41]: several possible complementary indicators exist to provide a more refined or diverse interpretation of condition from paleorecords. Additional biological (e.g. chironomids, chrysophytes), chemical (e.g. geochemistry, isotopes) and physical (e.g. grain sizes) indicators are at our disposal, should funds allow.

To expand the utility of these indicators beyond the Great Lakes system may require input from experts, a method that had success on a diatom-based indicator of stream condition in the northeast US [42]. Briefly, this approach employed diatom experts to vote on the indicator properties of taxa based on available data and their experience as diatom researchers. While it is difficult to quantify such associations between diatom species and stress, expert input is likely to apply historical knowledge from outside the immediate study area and so is likely to provide a broader context on a diatom species’ indicator status.

Could indices and transfer functions be improved by eliminating mundane or highly confounded species? This has been explored by Racca et al. [43] on a set of arctic lakes using a neural network approach that pruned species from a training set according to their unique influence by the variable of interest. This appeared to improve transfer function performance. Unfortunately, since publication of that study we can find no similar applications, though other methods for removing non-responsive taxa [44,45] and sample locations having extreme values of a confounding variable (e.g. elevation) [46] from transfer functions appear to be promising. Perhaps future applications should consider a permanent method for determination of confounded species prior to transfer function use.

Isolating species cultures and evaluating against stressor surrogates (e.g. mesocosm or “bottle” studies of phosphorus concentrations) have been applied to better understand species responses (e.g. Saros et al.) [47]. Such studies are impractical for the thousands of diatom taxa that have been observed in modern and fossil records in the Great Lakes, but assessments of the most common taxa could be helpful. Further, it has long been recognized that species act differently under culture due to the confining of environmental condition that prevents accurate characterization of in-situ species and assemblage behavior [48].
Though examination of fossil diatom records and development of diatom-based indicator models typically focus on relative abundance data, additional data considerations beyond the scope of this evaluation may be useful. With $^{210}\text{Pb}$-dated records there is the option to calculate diatom accumulation rates, which may be a better reflection of in-lake primary production. Use of algal biovolume and cell densities has been the focus of other indicator approaches in Great Lakes programs [9], and whether there are useful applications to long-term records is yet to be determined.

In previous indicator development across geographically expansive regions [49] indicators have been scored using typology; ecosystem type and ecosystem services. If we assume that pre-standardization of data by geographic location is appropriate, then it follows that developing indicator tools such as transfer functions that rely on water quality will be negatively affected if geographic correction is not applied. Clearly, a species like *Lindavia delicatula* (Fig 2) will have very different indicator coefficients using data with and without spatial standardization. Further, it is highly likely such geographic artefacts have had important negative influences on transfer functions, hence the recommendation to ensure narrow environmental variation outside the reconstructed variable of interest [50]. Certainly, applying a transfer function using data calibrated using all of the Great Lakes [9] should be performed carefully, employing validation that fossil assemblages indicate assemblage responses to the variable of interest. Geography is one consideration, but other confounding variables include water quality variables that may not be measured during the species calibration phase.

This study has several outcomes. (1) Geographic considerations are extremely important when calibrating diatom indicators. While constraining a training set geographically can reduce sample size, there are methods such as spatial normalization that may be useful to account for these confounding spatial factors. (2) There are several Great Lakes diatom taxa that are strong indicators of environmental stress and should be citable as such. (3) Many (most?) diatom taxa are apparently poor indicators of these stressors due to mundane responses or confounding by multiple variables, so caution should be used when making inferences based on these species.

Supporting information

**S1 Table.** Compilation of long-term stressor and measured data relative to each sediment core.
(XLSX)

**S2 Table.** Pearson correlations (r) and Gaussian quadratic squared correlation coefficients (Gauss. $r^2(\text{adj})$) of diatom relative abundance versus anthropogenic stressors (columns A–I) and other measured variables (columns K–S). Columns U–AC contain relationships with measured variables following Normalized standardization by lake. Taxa with a significant correlation (P < 0.05) are presented, sorted by smallest P value. N = number of observations in samples. The "spp list" spreadsheet provides the full species names for each diatom code.
(XLSX)

**S3 Table.** Diatom relative abundance data used for all species-environmental relationship analyses.
(XLSX)

Acknowledgments

We thank K. Kennedy, A.R. Kireta, R.W. Sterner, C.A. Stow and the Research Vessel Lake Guardian and Blue Heron field crews for their help collecting core samples. Sediment dating
was supported by D.R. Engstrom and personnel at the St. Croix Watershed Research Station. We thank Sonja Hausmann for input on a previous draft of this manuscript.

Author Contributions
Conceptualization: Euan D. Reavie.
Data curation: Euan D. Reavie.
Formal analysis: Meijun Cai.
Funding acquisition: Euan D. Reavie.
Investigation: Euan D. Reavie.
Methodology: Euan D. Reavie, Meijun Cai.
Project administration: Euan D. Reavie.
Resources: Euan D. Reavie.
Supervision: Euan D. Reavie.
Validation: Euan D. Reavie.
Visualization: Euan D. Reavie, Meijun Cai.
Writing – original draft: Euan D. Reavie, Meijun Cai.
Writing – review & editing: Euan D. Reavie, Meijun Cai.

References
1. Seddon AW, Mackay AW, Baker AG, Birks HJ, Breman E, Buck CE, et al. Looking forward through the past: identification of 50 priority research questions in palaeoecology. J Ecol. 2014; 102: 256–267. https://doi.org/10.1111/1365-2745.12195
2. Juggins S, Birks HJB. Quantitative environmental reconstructions from biological data. In: Birks HJB, Lotter AF, Juggins S, Smol JP, editors. Tracking environmental change using lake sediments. Dordrecht: Springer; 2012. pp. 431–494.
3. Birks HJB. Numerical methods for the analysis of diatom assemblage data. In: Smol JP, Stoermer EF, editors. The diatoms: applications for the environmental and earth sciences, 2nd edn. Cambridge: Cambridge University Press; 2010. pp. 23–54.
4. Potapova M, Charles DF. Distribution of benthic diatoms in US rivers in relation to conductivity and ionic composition. Freshw Biol. 2003; 48:1311–1328.
5. Kelly MG, Whitton BA. The trophic diatom index: a new index for monitoring eutrophication in rivers. J Appl Phycol. 1995; 7: 433–444.
6. King RS, Baker ME. Considerations for analyzing ecological community thresholds in response to anthropogenic environmental gradients. J N Am Benthol Soc 2010; 29: 998–1008.
7. Sgro GV, Reavie ED. Fossil diatoms, geochemistry, and the Anthropocene paleolimnology of Lake Huron. J Great Lakes Res. 2018; 44: 765–778.
8. Dixit SS, Smol JP, Kingston JC, Charles DF. Diatoms: powerful indicators of environmental change. Environ Sci Technol. 1992; 26:22–33.
9. Reavie ED, Heathcote AJ, Shaw Chraibi VL. Laurentian Great Lakes phytoplankton and their water quality characteristics, including a diatom-based model for paleoconstruction of phosphorus. PLoS ONE. 2014; 9: https://doi.org/10.1371/journal.pone.0104705 PMID: 25105416
10. Ekdahl EJ, Teranes JL, Guilderson TP, Turton CL, McAndrews JH, Wittkop CA, Stoermer EF. 2004. Prehistorical record of cultural eutrophication from Crawford Lake, Canada. Geology. 2004; 32: 745–748.
11. Saros JE, Michel TJ, Interlandi SJ, Wolfe AP. Resource requirements of Asterionella formosa and Fragilaria crotonensis in oligotrophic alpine lakes: implications for recent phytoplankton community reorganizations. Can J Fish Aquat Sci. 2005; 62: 1681–1689.
12. Stoermer EF, Ladewski BG, Schelske CL. Population responses of Lake Michigan phytoplankton to nitrogen and phosphorus. Hydrobiologia. 1978; 57: 249–65.

13. Bailey-Watts AE. The ecology of planktonic diatoms, especially *Fragilaria crotonensis*, associated with artificial mixing of a small Scottish loch in summer. Diatom Res. 1986; 1: 153–168.

14. Juggins S. 2013. Quantitative reconstructions in palaeolimnology: new paradigm or sick science? Quaternary Sci Rev. 2013; 64: 20–32.

15. Reavie ED, Juggins S. Exploration of sample size and diatom-based indicator performance in three North American phosphorus training sets. Aquat Ecol. 2011; 45: 529–538.

16. Pont D, Hugueny B, Beier U, Goffaux D, Melcher A, Noble R, Rogers C, Roset N, Schmutz S. Assessing river biotic condition at a continental scale: a European approach using functional metrics and fish assemblages. J Appl Ecol. 2006; 43:70–80.

17. Stager JC, Hecky RE, Grzesik D, Cumming BF, Kling H. Diatom evidence for the timing and causes of eutrophication in Lake Victoria, East Africa. Hydrobiologia 2009; 636:463–78.

18. Sgro GV, Reavie ED. Lake Erie’s ecological history reconstructed from the sedimentary record. J Great Lakes Res. 2018; 44: 54–69.

19. Reavie ED, Sgro GV, Estepp LR, Bramburger AJ, Pillsbury VL, et al. Climate warming and changes in *Cyclotella sensu lato* in the Laurentian Great Lakes. Limnol Oceanogr. 2017; 62: 788–793.

20. Reavie ED, Cai M, Werne JP, Meyer-Jacob C, Smol JP. Long-term productivity trends in the Laurentian Great Lakes: a comparison of geochemical methods. Limnol Oceanogr. Forthcoming 2019.

21. Shaw Chraibi VL, Kireta AR, Reavie ED, Brown TN, Cai M. A palaeolimnological assessment of human impacts on Lake Superior. J Great Lakes Res. 2014; 40: 886–897.

22. Reavie ED, Cai M, Brown TN. Historical watershed stressors for the Laurentian Great Lakes. Geosci Data J. Forthcoming 2018.

23. Hunter TS, Clites AH, Campbell KB, Gronewold AD. Development and application of a North American Great Lakes hydrometeorological database—Part i: Precipitation, evaporation, runoff, and air temperature. J Great Lakes Res. 2015; 41: 65–77.

24. NCDC. 2016. https://www.ncdc.noaa.gov/cdo-web/search, http://www.ncdc.noaa.gov/has/HAS.FileAppRouter?datasetname=3505v2&subqueryby=STATION&apppname=&outdest=FILE. Data accessed in February 2016.

25. Canada Weather. 2016. http://climate.weather.gc.ca/advanceSearch/searchHistoricData_e.html. Data accessed in February 2016.

26. Easterling DR, Meehl GA, Parmesan C, Changnon SA, Karl TR, Mearns LO. Climate extremes: observations, modeling, and impacts. Science. 2000; 289: 2068–2074. PMID: 11000103

27. NADP. http://nadp.sws.uiuc.edu/. Data accessed in November 2015.

28. CASTNET. http://www.epa.gov/castnet. Data accessed in November 2015.

29. GLAHF. http://glahf.org/data/. Data accessed in February 2016.

30. US Environmental Protection Agency. Climate Change Indicators in the United States—Ecosystems. Data accessed in November 2015. https://www.epa.gov/climate-indicators/ecosystems

31. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2018. https://www.R-project.org/.

32. Stoermer EF, Emmert G, Julius ML, Schelske CL. Paleolimnologic evidence of rapid recent change in Lake Erie’s trophic status. Can J Fish Aquat Sci 1996; 53: 1451–1458.

33. Stoermer EF, Wolin JA, Schelske CL, Conley DJ. An assessment of ecological changes during the recent history of Lake Ontario based on siliceous algal microfossils preserved in the sediments 1, 2. J Phycol. 1985; 21: 257–276.

34. Bradbury JP. A climatic-limnologic model of diatom succession for paleolimnological interpretation of varved sediments at Elk Lake, Minnesota. J Paleolimnol. 1988; 1: 115–131.

35. Lotter AF. Subfossil and modern diatom plankton and the paleolimnology of Rotsee (Switzerland) since 1850. Aquat Sci. 1989; 51: 338–350.

36. Heinsalu A, Alliksaar T, Leebeen A, Noges T. Sediment diatom assemblages and composition of pore-water dissolved organic matter reflect recent eutrophication history of Lake Peipsi (Estonia/Russia). In: Shallow Lakes in a Changing World. Dordrecht: Springer; 2007. pp. 133–143.

37. Eilers JM, Kann J, Cornell J, Moser K, Amand AS. Paleolimnological evidence of change in a shallow, hypereutrophic lake: Upper Klamath Lake, Oregon, USA. Hydrobiologia. 2004; 520: 7–18.
38. Hausmann S, Lotter AF, Van Leeuwen JFN, Ohlendorf C, Lemcke G, Grönlund E, et al. Interactions of climate and land use documented in the varved sediments of Seebergsee in the Swiss Alps. The Holocene 2002; 12: 279–289.

39. Bradshaw EG, Anderson NJ. Environmental factors that control the abundance of *Cyclostephanos dubius* (Bacillariophyceae) in Danish lakes, from seasonal to century scale. European Journal of Phycology. 2003; 38: 265–276.

40. Saros JE. Integrating neo-and paleolimnological approaches to refine interpretations of environmental change. J Paleolimnol. 2009; 41: 243–252.

41. Bennion H, Davidson TA, Sayer CD, Simpson GL, Rose NL, Sadler JP. Harnessing the potential of the multi-indicator palaeoecological approach: an assessment of the nature and causes of ecological change in a eutrophic shallow lake. Freshwater Biol. 2015; 60: 1423–1442.

42. Hausmann S, Charles DF, Gerritsen J, Belton TJ. A diatom-based biological condition gradient (BCG) approach for assessing impairment and developing nutrient criteria for streams. Sci Total Environ. 2016; 562: 914–927. https://doi.org/10.1016/j.scitotenv.2016.03.173 PMID: 27128024

43. Racca JM, Gregory-Eaves I, Pienitz R, Prairie YT. Tailoring palaeolimnological diatom-based transfer functions. Can J Fish Aquat Sci. 2004; 61: 2440–2454.

44. Denys L. Calibration of littoral diatoms to water chemistry in standing fresh waters (Flanders, lower Belgium): inference models for historical sediment assemblages. J Paleolimnol. 2006; 35: 763–787.

45. Racca JM, Racca R, Pienitz R, Prairie YT. PaleoNet: new software for building, evaluating and applying neural network-based transfer functions in paleoecology. J Paleolimnol. 2007; 38: 467–472.

46. Hausmann S, Kienast F. A diatom-inference model for nutrients screened to reduce the influence of background variables: application to varved sediments of Greifensee and evaluation with measured data. Palaeogeogr Palaeoclim. 2006; 233: 96–112.

47. Saros JE, Stone JR, Pederson GT, Slemmons KE, Spanbauer T, Schliep A, et al. Climate-induced changes in lake ecosystem structure inferred from coupled neo-and paleoecological approaches. Ecology. 2012; 93: 2155–2164. PMID: 23185877

48. Schindler DW. Whole-ecosystem experiments: replication versus realism: the need for ecosystem-scale experiments. Ecosystems. 1998; 1: 323–334.

49. Maes J, Lique C, Teller A, Erhard M, Paracchini ML, Barredo JL, et al. An indicator framework for assessing ecosystem services in support of the EU Biodiversity Strategy to 2020. Ecosystem Services. 2016; 17:14–23.

50. Smol JP. Pollution of lakes and rivers: a paleoenvironmental perspective. 2nd ed. Oxford: Blackwell Publishing; 2008.