Environmental Research Letters

Metadata for

**Thermal structure and oxygen depletion provide mechanism for cyanoHABS in remote lakes**

Lienne R. Sethna1\*, Amelia E. Wilson-Jackson1, Nicole D. Wagner2, Alaina, A. Fedie1, Amber J. White1, Mark B. Edlund1, Adam J. Heathcote1

1 St. Croix Watershed Research Station, Science Museum of Minnesota, Marine on St. Croix, Minnesota, USA

2 Department of Biological Sciences, Oakland University, Rochester, Michigan, USA

\* Corresponding author: lsethna@smm.org

**Table 1.** Description of the fields needed to describe the creation of your dataset.

|  |  |
| --- | --- |
| **Title of dataset** | Surface water quality monitoring data and high-resolution buoy measurements for eight lakes within the Superior National Forest, Minnesota, U.S.A. |
| **URL of dataset** | Data will be made accessible via Zenodo. |
| **Abstract** | Harmful blooms of cyanobacteria (cyanoHABs) are increasingly reported in remote and low-nutrient lakes, prompting a paradigm shift in our understanding of the drivers of cyanoHABs and a need to study factors beyond watershed nutrient inputs. In this study, we examined eight lakes within the Superior National Forest (Minnesota, USA) to assess the control of lake thermal structure and oxygen depletion on nutrient and cyanobacteria abundance. This dataset includes monthly water quality information including nutrient concentrations (dissolved inorganic nitrogen, total phosphorus, dissolved silicon), chlorophyll-*a*, and phycocyanin as well as 30-minute readings of water column temperature and dissolved oxygen from buoy strings installed at the deepest point of each sample lake. The dataset ranges from 06-27-2022 through 10-26-2022 |
| **--Keywords** | Harmful algal blooms, cyanobacteria, climate change, internal nutrient loading, lake stratification, remote lakes |
| **Lead author for the dataset** | Lienne Sethna |
| **Title and position of lead author** | Assistant Scientist |
| **Organization and address of lead author** | St. Croix Watershed Research Station  16910 152nd St. NW  Marine on St. Croix, MN 55047 |
| **Email address of lead author** | lsethna@smm.org |
| **Additional authors or contributors to the dataset** | Amelia Wilson-Jackson, Nicole Wagner, Alaina Fedie, Amber White, Mark Edlund, Adam Heathcote |
| **Organization associated with the data** | St. Croix Watershed Research Station |
| **Funding** | Minnesota Environmental and Natural Resources Trust Fund: “Unprecedented Change Threatens Minnesota’s Pristine Lakes”; PI: Mark Edlund; Project 070-B, M.L. 2021, First Special Session, Chp. 6, Art. 5, Sec. 2, Subd. 20a1 |
| **License** | [**CCBY**](https://creativecommons.org/licenses/by/4.0/) – requires attribution |
| **Geographic location – verbal description** | Sampling sites were located in the Superior National Forest, Minnesota, U.S.A. |
| **Geographic coverage bounding coordinates** | 47.5°N - 48.0°N  -92°W - -90.0°W |
| **Time frame - Begin date** | June 27, 2022 |
| **Time frame - End date** | October 26, 2022 |
| **General study design** | We conducted lake monitoring between June 27 and October 26 2022, sampling over the deepest point in each of eight study lakes. Monthly sampling included the collection of nutrient samples (nitrate/nitrite, total phosphorus, ammonia, and dissolved silicon), chlorophyll-*a*, and phycocyanin. We also deployed monitoring buoys at the sampling location that collected water column temperature at meter intervals and hypolimnetic dissolved oxygen concentrations at 30-minute intervals throughout the monitoring period. |
| **Methods description** | We conducted monthly lake monitoring from June through October 2022, sampling over the deepest point in each lake. We collected water samples from the surface of the lake using a 60-mL polypropylene syringe that was rinsed with sample water before collection. Separate samples were collected for the analysis of dissolved inorganic nitrogen (nitrate/nitrite and ammonia), dissolved silica, total phosphorus, and chlorophyll-*a*. All dissolved constituents were field-filtered using a 0.45 µm nitrocellulose membrane filter and stored in 60-mL polypropylene bottles. All samples were transported on ice and frozen before analysis.  We deployed buoys at the deepest point in each lake, outfitted with a dissolved oxygen (DO) sensor (HOBO DO Logger U26-001; Onset Computer Co.; Bourne, MA) approximately 1 m off the bottom as well as temperature loggers (HOBO Temperature Pendants UA-002-64) every 1 m from the bottom DO sensor to 2m below the surface (Figure S1). These sensors monitored dissolved oxygen and temperature every thirty minutes between late June and late October 2022. The DO sensors were calibrated using a two-point calibration before deployment; recorded concentrations were corrected for temperature. |
| **Laboratory, field, or other analytical methods** | Nutrient samples were analyzed colorimetrically following standard methods; samples for nitrate/nitrite were analyzed using the cadmium reduction method (4500-NO3- F; Bridgewater et al. 2017) and ammonia using the phenol-hypochlorite method (4500-NH3 H; Solórzano 1969) on a SEAL AQ400 Discrete Analyzer (SEAL Analytical; Mequon, WI). Samples for dissolved silica were analyzed using the heteropoly blue method (4500-Si-F; Sultan 2014) on a SmartChem 170 Discrete Analyzer (KPM Analytics, Westborough, MA). Samples for total phosphorus were determined colorimetrically on a SEAL AQ400 analyzer following a persulfate digestion (4500-P J; Bridgewater et al. 2017). Sample concentrations for each analyte were verified using external check standards and method detection limits (MDLs) were routinely calculated following EPA procedure 821-R-16-006. The MDLs for ammonia, nitrate/nitrite, dissolved silica, and total phosphorus were 8 µg N L-1, 5 µg N L-1, 0.3 mg SiO2 L-1, and 3.5 µg P L-1, respectively. Samples that fell below the MDLs were assigned half the MDL concentration (37% of ammonia, 82% of nitrate/nitrite, and 10% of dissolved silica).  We assessed the abundance of cyanobacteria using a Total Algae sensor attached to a YSI EXO2 multiparameter sonde (Yellow Springs Instruments; Yellow Springs, OH). This multi-band sensor simultaneously measures phycocyanin, a pigment specific to cyanobacteria, and chlorophyll-a (chl-a) concentrations. The sensor was calibrated before each sampling using rhodamine dye standards (Bowling et al., 2016). The sensor was lowered through the water column and pigment concentrations were integrated to account for average algal density within the water column. Phycocyanin:chl-a was calculated as a proxy of relative cyanobacterial abundance. We collected grab samples during observed cyanoHABs and used an inverted light microscope to identify cyanobacteria to the genus level.  Chlorophyll-a samples were collected by filtering 1-L of surface water onto a 1.2 µm glass-fiber filter (Whatman GF/C), which was stored in the dark and frozen until analysis. Chl-a was extracted from the filters using a 90% acetone extraction (EPA 445.0, Revision 1.2, Arar, et. al, 1997) for 18 hours. The extractant was analyzed using the Chl-a Non-Acidification Module (Model 7200-046) on a Trilogy fluorometer (Model 7200-200; Turner Designs; San Jose, CA). Concentrations were determined from absorbance values by accounting for the volume filtered.  We used the *rLakeAnalyzer* package in R (version 1.11.4.1, Winslow et al. 2019) to quantify lake thermal structure, including thermocline depths with the *thermo.depth* function. We defined mixing events by periods of time in which the sample lake transitioned from having a defined thermocline to no thermocline for a sustained period of at least 24 hours before re-stratifying.  We calculated the Schmidt stability, a metric for the resistance to lake mixing due to the energy stored in the thermal stratification of the water column (Hutchinson and Löffler 1956), of each lake using the *schmidt.stability* function in *rLakeAnalyzer*. Inputs to the function included lake bathymetry (U.S. Forest Service) and the 30-minute temperature profiles collected from the HOBO temperature loggers. |
| **Statistical methods** | Nutrient concentrations were related to phycocyanin and chl-a using repeated measures correlation with the rmcorr package (version 0.6.0; Bakdash & Marusich, 2017) in R (version 4.3.0; R Core Team, 2023). Repeated measures correlation accounts for the non-independence between samples taken from the same site through time by statistically adjusting for the variability between sites, calculates the best correlation for each site, and evaluates the overall, intra-site relationship between two variables. We calculated correlation coefficients for each paired measure using the rmcorr\_mat function, with confidence intervals set to 0.95. Data were visualized using the ggplot2 package (version 3.4.2; Wickham, 2016).  We quantified the synchrony between Schmidt stability and bottom water DO concentrations in each lake with Spearman rank correlations using the *cor.test* function in R (version 4.3.0; R Core Team). |
| **Taxonomic species or groups** | *Dolichospermum, Microcystis, and Woronichinia* |
| **Quality control** | Replicate samples were taken in the field at least once per month for each analyte to ensure repeatability in our field methods. Duplicate samples were analyzed in the laboratory using sample splits and instrumentation duplicates to verify laboratory and instrumentation methods. We also ran certified standards to verify instrumentation calibration curves. Method detection limits were established using EPA method 821-R-16-006 Samples with measured concentrations below the instrumentation detection limit were assigned a value of half the detection limit. |
| **Additional information** | *Any additional information that may help future users of the data not included in the above rows, or in the table below.* |
|  |  |

**Table 2.** Data dictionary: description of the variables (i.e., columns) in each dataset

Dataset filename: WL\_LabChem.csv

Dataset description: monthly grab samples for nutrient concentrations and chlorophyll-*a*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| *The name of the variable in the dataset; avoid special characters, dashes and spaces* | *A detailed description of the variable* | *Units the variable is measured in* | *If you use codes in your column, please explain each code, such as: LR = Little Rock Lake; A=sample; etc.* | *State exactly how the data are stored; for dates, state how it is formatted, including time zone, etc.* | *If data are missing, indicate how they are stored, such as NULL, NA, blank cell, etc.* |
| Lake | Lake name where samples were collected |  |  | chr | NA |
| CollectionDate | Date of sample collection |  |  | MM/DD/YYYY | NA |
| TP | Total Phosphorus | µg P/L |  | dbl | NA |
| TN | Total Nitrogen | mg N/L |  | dbl | NA |
| NH4 | Ammonia | mg N/L |  | dbl | NA |
| NO3\_NO2 | Nitrate and Nitrite | mg N/L |  | dbl | NA |
| SRP | Soluble Reactive Phosphorus, | µg P/L |  | dbl | NA |
| DIC | Dissolved Inorganic Carbon | mg C/L |  | dbl | NA |
| DOC | Dissolved Organic Carbon | mg C/L |  | dbl | NA |
| DSi | Dissolved Silica | mg SiO2/L |  | dbl | NA |
| NH4\_detect | Ammonia samples where samples below the detection limit were assigned half of the MDL | mg N/L |  | dbl | NA |
| NO3\_NO2\_detect | Nitrate and Nitrite samples where samples below the detection limit were assigned half of the MDL | mg N/L |  | dbl | NA |
| DIN | Dissolved Inorganic Nitrogen | mg N/L |  | dbl | NA |
| Chla\_ugL | Chlorophyll-a concentration | µg/L |  | dbl | NA |

Dataset filename: WL\_algaepigment\_ysi.csv

Dataset description: Monthly chlorophyll-*a* and phycocyanin pigment concentrations measured by YSI EXO sonde

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| *The name of the variable in the dataset; avoid special characters, dashes and spaces* | *A detailed description of the variable* | *Units the variable is measured in* | *If you use codes in your column, please explain each code, such as: LR = Little Rock Lake; A=sample; etc.* | *State exactly how the data are stored; for dates, state how it is formatted, including time zone, etc.* | *If data are missing, indicate how they are stored, such as NULL, NA, blank cell, etc.* |
| site | Lake site where samples were collected |  |  | chr | NA |
| date | Date of sample collection |  |  | MM/DD/YYYY (GMT-5), chr | NA |
| bga\_pc\_ugL\_avg | Blue green algae (as phycocyanin) Averaged throughout the water column | µg/L |  | dbl | NA |
| chla\_ugL\_avg | Chlorophyll-a Averaged throughout the water column | µg/L |  | dbl | NA |
| pc\_chla\_ysi | Phycocyanin to Chlorophyll-a ratio as recorded by the YSI |  |  | dbl | NA |

Dataset filename: WL\_summer23DO\_bot.csv

Dataset description: Bottom water dissolved oxygen concentrations collected at 10-minute intervals from 1 m off the sediment at the deepest point of each study lake

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| *The name of the variable in the dataset; avoid special characters, dashes and spaces* | *A detailed description of the variable* | *Units the variable is measured in* | *If you use codes in your column, please explain each code, such as: LR = Little Rock Lake; A=sample; etc.* | *State exactly how the data are stored; for dates, state how it is formatted, including time zone, etc.* | *If data are missing, indicate how they are stored, such as NULL, NA, blank cell, etc.* |
| datetime | Date and time of measurement recorded |  |  | MM/DD/YYYY HH:MM:SS %p | NA |
| site | Lake site where buoys were deployed |  | W.Twin= West Twin  E.Twin= East Twin | chr | NA |
| bottom\_DO\_conc | Bottom water dissolved oxygen concentration | mg/L |  | dbl | NA |
| bottom\_temp | Bottom water temperature | °C |  | dbl | NA |
| bottom\_DO\_adj\_conc | Bottom water dissolved oxygen concentration adjusted for temperature | mg/L |  | dbl | NA |
| bottom\_DO\_perc\_sat | Bottom water dissolved oxygen percent saturation | mg/L |  | dbl | NA |

Dataset filename: WL\_allbuoy\_30min\_temp.csv

Dataset description: Water column temperature measurements collected at 30-minute intervals between 2 m and z-1 m from each study lake.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| *The name of the variable in the dataset; avoid special characters, dashes and spaces* | *A detailed description of the variable* | *Units the variable is measured in* | *If you use codes in your column, please explain each code, such as: LR = Little Rock Lake; A=sample; etc.* | *State exactly how the data are stored; for dates, state how it is formatted, including time zone, etc.* | *If data are missing, indicate how they are stored, such as NULL, NA, blank cell, etc.* |
| datetime | Date and time of measurement recorded |  |  | MM/DD/YYYY HH:DD:SS %p | NA |
| site | Lake site where buoys were deployed |  | W.Twin= West Twin  E.Twin= East Twin | chr | NA |
| depth | Depth temperature sensors were located | meters |  | int | NA |
| temp\_C | Temperature | °C |  | dbl | NA |