Metadata template[[1]](#footnote-1) for datasets of *L&O-Letters* articles

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

|  |  |
| --- | --- |
| **Title of dataset** | *Data supporting “Synchronous variation of dissolved organic carbon in Adirondack lakes at multiple timescales”* |
| **URL of dataset** | *https://github.com/NatCoombs/ADK-DOC-Synch* |
| **Abstract** | *Dissolved organic carbon (DOC) is a key component of aquatic ecosystems with complex effects on ecosystem function. While long-term increases in DOC termed ‘brownification’ have received considerable attention, directional trends typically account for a minority of variance. DOC concentrations also fluctuate on seasonal to multi-annual timescales, but the causes of such variations are less understood. We used a wavelet-based approach to study timescale-specific, spatially synchronous fluctuations in DOC across 49 lakes in the Adirondacks, New York, USA. DOC varies synchronously among lakes at within-season, annual, and interannual timescales, but relationships with external drivers and internal processes indicated by lake chemistry differed across timescales. External drivers explained 78% of spatial DOC synchrony at the annual time scale. Beyond positive trends related to regional recovery from acidification, variability in DOC is a consequence of fluctuations at several timescales that are common among Adirondack lakes in precipitation, solar radiation, and internal chemical concentrations.*  *Data supporting this study consist of monthly time series of DOC for 49 lakes, several lake chemistry variables including major cations and anions, precipitation, solar radiation, acid deposition, and the North Atlantic Oscillation.* |
| **Keywords** | Lake, dissolved organic carbon, acid deposition, climate, aquatic chemistry, solar radiation, Adirondacks, New York |
| **Lead author for the dataset** | *Jonathan Walter* |
| **Title and position of lead author** | *Senior Researcher (PI)* |
| **Organization and address of lead author** | *Department of Environmental Sciences, University of Virginia, Charlottesville, VA, USA*  *Center for Watershed Sciences, University of California, Davis, CA, USA* |
| **Email address of lead author** | *Jaw3es@virginia.edu, jawalter@ucdavis.edu* |
| **Additional authors or contributors to the dataset** | *Michael L. Pace, Nat J. Coombs* |
| **Organization associated with the data** |  |
| **Funding** | *Michael L. Pace: National Science Foundation DEB-1754712.* |
| **License** | *CCBY* |
| **Geographic location – verbal description** | *Adirondack Park, NY* |
| **Geographic coverage bounding coordinates** | *Longitude min: -75.31831*  *Longitude max: -73.29350*  *Latitude min: 43.05239*  *Latitude max: 44.87769* |
| **Time frame - Begin date** | *1992-06-01* |
| **Time frame - End date** | *2017-09-01* |
| **General study design** | *This study is based on a compilation of publicly available datasets. The primary data set is the Adirondack Long Term Monitoring Program (ALTM) a field survey of monthly water quality measurements from lakes in Adirondack Park, NY. Protocols are described in Driscoll et al. (2003) doi: 10.1021/es020924h. This dataset was augmented with data on weather (PRISM:* [*https://prism.oregonstate.edu*](https://prism.oregonstate.edu)*), climate (NAO;* [*https://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml*](https://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml)*), acid deposition (NADP:* [*https://nadp.slh.wisc.edu*](https://nadp.slh.wisc.edu)*), and solar radiation (NSRDB:* [*https://nsrdb.nrel.gov/data-sets/spectral-on-demand-data*](https://nsrdb.nrel.gov/data-sets/spectral-on-demand-data) *).* |
| **Methods description** | *Data were made interoperable by matching the temporal resolution (monthly) and, to the extent possible, the spatial resolution (49 lakes) of the ALTM lake chemistry data. Data at sub-monthly resolutions (solar radiation) were aggregated to monthly by averaging. The NAO is a non-spatial variable representing a large-scale climate mode, so its values were taken to apply across all lakes. Sampling stations used by the NADP to monitor acid deposition did not match lake locations, so a single acid deposition time series representative of the study area was produced by averaging values from monitoring stations in the reason and the regional average taken to apply across lakes.* |
| **Laboratory, field, or other analytical methods** | *ALTM Protocols are described in Driscoll et al. (2003) doi: 10.1021/es020924h.* |
| **Taxonomic species or groups** | *NA* |
| **Quality control** | *Observational time series were plotted to assess the presence of outliers. 33 (0.2%) DOC observations were removed because they were implausible values likely resulting from laboratory anomalies.* |
| **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. You must provide sufficient detail for another user to understand and use the data. If there are 10 variables (i.e., columns) in the dataset, then there should be 10 rows in this table that describe each column. Be sure to include all relevant information for your dataset, including the unique identifiers for your dataset or system, dates, replicate numbers, latitude and longitude of sampling locations, etc.

Dataset filename: *altm\_data.csv*

Dataset description: *Monthly water quality data from the Adirondack Long Term Monitoring Program (ALTM).*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| ConfDate | Sample date conformed to a monthly time interval; observations were assigned to the first of the month they were taken. |  |  | YYYY.MM.DD |  |
| LabID | Sample unique identifier |  |  | 7 alphanumeric characters |  |
| SiteID | Lake unique identifier |  |  | 5 or 6 alphanumeric characters |  |
| Station | Station identifier |  |  | 1 or 2 digits |  |
| depth | Depth at which the water sample was taken | meters |  | numeric |  |
| SampleDate | Date the sample was taken |  |  | YYYY.MM.DD |  |
| SampleTime | Time the sample was taken |  |  | 3 or 4 digits; last two are minutes, first one or two are hours; hours are given using 24-hour time. |  |
| SO4 | Sulfate concentration | mg/L |  | numeric | Blank cell |
| NO3 | Nitrate concentration | mg/L |  | numeric | Blank cell |
| F | Fluoride concentration | mg/L |  | numeric | Blank cell |
| ANC | Acid neutralizing capacity | mol/kg |  | numeric | Blank cell |
| DIC | Dissolved inorganic carbon | mg/L |  | numeric | Blank cell |
| DOC | Dissolved organic carbon | mg/L |  | numeric | Blank cell |
| SiO2 | Silica concentration | mg/L |  | numeric | Blank cell |
| Ca | Calcium concentration | mg/L |  | numeric | Blank cell |
| Mg | Magnesium concentration | mg/L |  | numeric | Blank cell |
| Na | Sodium concentration | mg/L |  | numeric | Blank cell |
| K | Potassium concentration | mg/L |  | numeric | Blank cell |
| NH4 | Ammonium concentration | mg/L |  | numeric | Blank cell |
| AL\_TD | Total dissolved aluminum concentration | ug/L |  | numeric | Blank cell |
| AL\_TM | Total monomeric aluminum concentration | ug/L |  | numeric | Blank cell |
| AL\_OM | Organic monomeric aluminum concentration | ug/L |  | numeric | Blank cell |
| AL\_IM | Inorganic monomeric aluminum concentration | ug/L |  | numeric | Blank cell |
| LABPH | Laboratory measured pH concentration |  |  | numeric | Blank cell |
| AIREQPH | Air equilibrated pH concentration |  |  | numeric | Blank cell |
| TRUCOLOR | True color |  |  | numeric | Blank cell |
| SCONDUCT | Specific conductivity | us/cm |  | numeric | Blank cell |
| TP | Total phosphorus concentration | mg/L |  | numeric | Blank cell |
| Chlorophyll\_a | Chlorophyll-a concentration | mg/L |  | numeric | Blank cell |

Dataset filename: *precip.csv*

Dataset description: *Lake-specific time series of monthly total precipitation obtained from PRISM in “wide” format.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| lake | Lake unique identifying code |  |  | 5 or 6 alphanumeric characters |  |
| 1992.06 | Precipitation in June 1992 | mm |  | numeric |  |
| 1992.07 | Precipitation in July 1992 | mm |  | numeric |  |
| … | … | … |  | … |  |
| 2017.09 | Precipitation in September 2017 | mm |  | numeric |  |

Dataset filename: *solar\_radiation.csv*

Dataset description: *Regionally averaged solar radiation time series*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| date | date |  |  | YYYY.MM |  |
| par | Photosynthetically active radiation | watts/m^2 |  | numeric |  |
| uv | Ultraviolet radiation | watts/m^2 |  | numeric |  |

Dataset filename: *acid\_deposition.csv*

Dataset description: *Regionally averaged acid deposition time series*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| date | date |  |  | YYYY.MM |  |
| par | Photosynthetically active radiation | watts/m^2 |  | numeric |  |
| uv | Ultraviolet radiation | watts/m^2 |  | numeric |  |

Dataset filename: *nao.csv*

Dataset description: *Monthly time series of the NAO index*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| yyyy | Year |  |  | 4 digits |  |
| mm | Month |  |  | 1 or 2 digits |  |
| NAO | NAO index value |  |  | numeric |  |

Dataset filename: *lake\_coords.csv*

Dataset description: *Geographic coordinates of ALTM lakes*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| SiteID | Lake unique identifying code |  |  | 5 or 6 alphanumeric characters |  |
| LONGITUDE | Longitude |  |  | numeric |  |
| LATITUDE | Latitude |  |  | numeric |  |
| NAME | Lake name |  |  | Character |  |

**Table 3. Data provenance**

If you used data derived from other sources, provide the information here so future users know where the data came from.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset title** | **Dataset DOI or URL** | **Creator (name & email)** | **Contact (name & email)** |
| Adirondack Long Term Monitoring Program | http://www.adirondacklakessurvey.org |  |  |
| PRISM monthly precipitation | https://prism.oregonstate.edu |  |  |
| National Solar Radiation Database Spectral On-Demand Data | https://nsrdb.nrel.gov/data-sets/spectral-on-demand-data |  |  |
| National Acid Deposition Program National Trends Network | https://nadp.slh.wisc.edu/networks/national-trends-network/ |  |  |
| Monthly mean NAO index | https://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml |  |  |

**Scripts/code (software)**

It is recommended that you also provide your scripts along with your data, although it is not required at this time in our journal.

|  |  |  |
| --- | --- | --- |
| **File name** | **Description** | **Scripting language** |
| mainResults.R | Reproduces analyses reported in the main text and generates figures | R |
| supplement\_spatSync.R | Reproduces analyses reported in supplementary material concerning spatial patterns of synchrony in DOC | R |
| supplement\_Variation.R | Reproduces analyses reported in supplementary material concerning temporal variation at intra-annual and interannual scales. | R |

**Notes and Comments:**

Analyses were implemented in R version 4.2.1 using the RStudio graphical user interface version 2022.07.0. The primary analysis package is ‘wsyn’ version 1.0.4.

1. *This document liberally borrows from a similar document provided by the Environmental Data Initiative* [↑](#footnote-ref-1)