Metadata template for datasets of *L&O-Letters* articles

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

|  |  |
| --- | --- |
| **Title of dataset** | Carbon and environmental data from 253 northern circumpolar ponds |
| **URL of dataset** | <http://www.cen.ulaval.ca/nordicanad/> |
| **Abstract** | As permafrost thaws, new ponds emerge and vast stocks of carbon stored in frozen soils are released to aquatic ecosystems. By optical analyses of dissolved organic matter (DOM) in 253 ponds, we show the impact of thawing permafrost on DOM pools across the circumpolar North. Further, using stable isotopes, we quantified the contribution of terrestrial sources to the DOM pool in a subset of ponds in subarctic Quebec. Optical analyses underscored a higher proportion of terrestrial carbon, and a lesser algal compounds contribution to DOM in ponds impacted by thawing permafrost. In the selected subarctic ponds, DOM composition was dominated by terrestrial substances (93.2 ± 8.2%) in sites that were influenced by thawing permafrost compared to ponds with bedrock (36.2 ± 34.9%) or non-permafrost tundra soils (42.1 ± 20.3%) in the watershed. Our results demonstrate the strong terrestrial imprint on freshwater ecosystems in degrading permafrost catchments, and the likely shift towards dominance of land-derived organic carbon in high-latitude waters with ongoing permafrost thaw. |
| **Keywords** | permafrost thaw, northern circumpolar ponds, DOM, chlorophyll *a*, DOC, CDOM, PARAFAC, stable isotopes, d13C, d2H |
| **Dataset lead author**  | Maxime Wauthy |
| **Position of data author** | PhD student |
| **Address of data author** | Département des sciences fondamentalesUniversité du Québec à Chicoutimi (UQAC) 555, boulevard de l’Université Chicoutimi (QC) G7H2B1 Canada |
| **Email address of data author** | maximewauthy@hotmail.com |
| **Primary contact person for dataset** | Same as author |
| **Position of primary contact person** | n/a |
| **Address of primary contact person** | n/a |
| **Email address of primary contact person** | n/a |

|  |  |
| --- | --- |
| **Organization associated with the data** | Centre for Northern Studies, Institut national de la recherche scientifique, Laval University, Université du Québec à Chicoutimi, University Centre in Svalbard, University of Copenhagen. |
| **Usage Rights** | Publicly available |
| **Geographic region** | Circumpolar Arctic, including regions in Alaska, Canada, Greenland, Finland and Russia |
| **Geographic coverage**  | Circumpolar arctic permafrost zone, from 149°36’W to 62°56’E (more details in Table S1) |
| **Temporal coverage - Begin date** | Summer 2002 |
| **Temporal coverage - End date** | Summer 2016 |
| **General study design** | In this study, we investigated the influence of permafrost thaw to the stocks and composition of dissolved organic material (DOM) in circumpolar North freshwater systems, using a suite of chemical (DOC, etc.), biological (chlorophyll *a*), optical (spectrophotometric indexes, EEMs, PARAFAC) and stable isotopic (d13C and d2H) indicators.  |
| **Methods description** | During the summer periods from 2002 to 2016, we sampled a total of 253 ponds distributed in 14 circumpolar regions, for a total of 356 samples, a fraction of ponds being sampled more than one time during these 15 years. The regions span over a very wide geographic area, covering ~200° in longitude (from Alaska to Russia) and ~30° in latitude (from Subarctic to High Arctic), and resulting in a large range of thawing permafrost influence.We divided the ponds into three categories according to their exposure to permafrost thaw: (1) bedrock ponds, characterized by a bedrock catchment and not directly affected by thawing permafrost; (2) tundra, not impacted by permafrost thaw, but characterized by forest, shrub or desert tundra watersheds, depending on the region; and (3) thaw ponds, directly affected by thawing permafrost.In order to investigate the impact of thawing permafrost on DOM pool, we used chemical, biological, optical and stable isotopic indicators. The different sampling methods and laboratory analyses are described in the next row for each dataset and each variable.To determine how the chemical, biological and optical properties were individually influenced by permafrost thaw, we performed Kruskal-Wallis by rank tests. Pair-wise comparisons were applied using a post hoc test (Bonferroni). We carried out PCAs and PERMANOVAs on log10(x+1)-transformed data to illustrate and test the influence of pond type on optical properties. The data were centered and standardized before applying the PCAs. Regarding the PERMANOVAs, we used Euclidean distance as dissimilarity index, and the number of permutations was fixed at 999. Pair-wise comparisons were performed using Bonferroni correction to identify differences among the types of ponds. The multivariate homogeneity of group dispersions was verified by performing a PERMDISP. All statistical analyses were performed on R v 3.3.2.Regarding the stable isotopes, we performed a dual Bayesian mixing model adapted from Wilkinson et al. (2014), using d13C and d2H as end-members. We considered three potential sources contributing to DOM in each pond type: soils, phytoplankton, and benthic bulk for bedrock and tundra ponds, or soils, phytoplankton, and macrophytes for thaw ponds. The model was run in R v 3.3.2. |
| **Laboratory, field, or other analytical methods** | **CDOM\_Database.csv*****Biological and chemical properties***We collected water samples from the surface to measure **pH**, total phosphorus (**TP**), total nitrogen (**TN**) and total dissolved iron (**Fe**) concentrations. We filtered surface water through pre-rinsed cellulose acetate filters (0.2 μm) in order to analyze dissolved organic carbon concentration (**DOC**) and perform optical analyses on the chromophoric DOM (CDOM) (see below). We stored samples in acid-washed and combusted glass vials, at 4°C in the dark, and DOC quantification was carried out using a carbon analyzer (TOC-5000A or TOC-VCPH, Shimadzu, Kyoto, Japan). We also filtered surface water onto GF/F filters to determine pelagic chlorophyll *a* (**Chla**) concentration following Nusch (1980).***Spectrophotometric and fluorometric indexes***We measured CDOM absorbance between 250 and 800 nm using a UV-visible Cary 100 (Agilent, Santa Clara, CA), Cary 300 (Agilent, Santa Clara, CA) or LAMBDA 650 (PerkinElmer, Waltham, MA) spectrophotometer, depending on the sample origin. After subtracting the blank spectrum, we applied a null-point adjustment, using the mean value from 750 to 800 nm.We calculated the specific ultraviolet absorbance at 254 nm (**SUVA254**) from DOC normalized A254 values. In order to remove the iron effect on SUVA254, we applied the following equation when the Fe concentration was higher than 2 mg L-1: $A\_{254 corrected}= A\_{254 measured} -\left(0.0653 × \left[Fe\right]\right)$We reported CDOM as the absorption coefficient at 320 (**a320**) and 440 nm (**a440**) according to the equation: $a\_{λ}=2.303 × A\_{λ}/ L ,$**where aλ is the absorption coefficient (m-1) at wavelength λ, Aλ the absorbance corrected at wavelength λ and L the path length of the cuvette (m).****We also determined spectral slopes (S) following Loiselle et al. (2009) for the intervals 279-299 (S289),** 275-295 (S285) and 350-400 nm (S375). We performed the regression calculations on SciLab v 5.5.2. The slope ratio (**SR**) is S285 divided by S375.For a subsample of 100 ponds, we recorded fluorescence intensity on Cary Eclipse (Agilent, Santa Clara, CA) across excitation wavelengths of 250-450 nm (10 nm increments) and emission wavelengths of 300-560 nm (2 nm increments) in order to construct excitation-emission matrices (EEMs). We calculated the fluorescence index (**FI**) as the ratio of fluorescence emission intensities at 450 nm and 500 nm for an excitation at 370 nm.***PARAFAC Results***To identify and quantify the main DOM components, we ran a parallel factor (PARAFAC) model on 129 samples from 95 ponds in MATLAB v R2013a (MathWorks, Natick, MA) as in Murphy et al. (2013). We corrected EEMs for Raman and Rayleigh scattering and inner filter effects, and standardized fluorescence to Raman units using the FDOMcorr 1.4 toolbox (Murphy et al. 2010). The model was performed on corrected EEMs and validated by split-half analysis using the drEEM toolbox from Murphy et al. (2013). For each sample, we summed the maximum fluorescence [**Cx**] of the different components x to determine the total fluorescence (FT) and calculated the relative abundance (**%Cx**) of any component x, according to the following equation:$\%Cx = ([Cx]/ F\_{T}) × 100 $ (3)**Spectro\_Scans\_Raw\_800-0.4-250.4.csv, Spectro\_Scans\_Raw\_800-1-250.csv and Spectro\_Scans\_Raw\_800-2-250.csv**These three datasets include the **raw absorbance data** obtained by spectrophotometric scans of DOM samples (CDOM absorbance between 250 and 800 nm, see above). You are thus able to calculate the different spectrophotometric indexes with the original raw data. Note that we already have subtracted the blank spectrum. The raw scans are divided in three *.csv* files, depending of the increments used (0.4, 1 and 2 nm, respectively).**SIA\_Data.csv**To quantify the relative contribution of terrestrial carbon between pond types, we carried out stable isotope analyses (**SIA**) on surface DOM and its potential sources in a subsample of 10 ponds in the vicinity of Kuujjuarapik, subarctic Quebec. In addition to collecting soils surrounding ponds, DOM, benthic bulk material and seston were sampled in bedrock and tundra water bodies for **d13C** and **d2H** analyses. In thaw ponds, the same materials were sampled, except the benthic bulk which was replaced by macrophytes.For DOM samples, we filtered surface water through pre-rinsed cellulose acetate filters (0.2 μm), as for DOC and CDOM analyses. Soil samples were collected from the top layer (0-5 cm) close to the ponds. In order to remove the carbonate, we applied an acid fumigation to the soil samples during 96h prior the d13C analyses, as described in Ramnarine et al. (2011). Decaying submerged macrophytes (*Carex* sp.) were sampled from the edge of thaw ponds. In bedrock and tundra ponds, we sampled the surface of submerged rocks by scraping with a spatula to collect the epibenthic material. DOM samples were analyzed for d13C using an Aurora 1030W TOC Analyzer (O.I. Corporation, College Station, TX) coupled to a Finnigan DELTA plus Advantage MS (Thermo Fisher Scientific, Waltham, MA) in the G.G. Hatch Stable Isotope Laboratory (University of Ottawa, ON). The d13C analyses on soil and macrophyte samples were carried out using a FLASH 2000 OEA interfaced with a Delta V Plus MS (Thermo Fisher Scientific, Waltham, MA) in the RIVE Research Center (Université du Québec à Trois-Rivières, QC). All d2H analyses were performed in the Colorado Plateau Stable Isotope Laboratory (Northern Arizona University, AZ) as in Doucett et al. (2007), using a CONFLO II coupled to a Delta Plus XL MS (Thermo Fisher Scientific, Waltham, MA). To determine the phytoplankton d13C signature, we used specific algal fatty acids (FAs) extracted from bulk seston as a proxy, performing SIA on 16:1n7, 18:2n6, 18:3n3 and 20:5n3 fractions. The FAs were transmethylated according to a protocol adapted from Lepage and Roy (1984). The d13C analyses on FAs were carried out in the Stable Isotope Laboratory of Memorial University (Memorial University of Newfoundland, NL), using a 6890N GS (Agilent, Santa Clara, CA) linked to a Delta V Plus MS (Thermo Fisher Scientific, Waltham, MA). We estimated the phytoplankton d2H signature as in Grosbois et al. (2017). Benthic samples had not been acid fumigated and we therefore considered the benthic bulk d13C signatures to not be reliable. Consequently, as for the phytoplankton, we extracted the FAs from the benthic bulk material and performed d13C analyses on the 14:0, 16:0 and 18:0 fractions. |
| **Quality control** | Data were double check for accuracy with original data files, both during processing and prior submission for publication |
| **Additional information** | We also give access to a MATLAB file ***mydata\_Fluorometric\_Data.mat***, which contains the excitation-emission matrices (EEMs) for all the samples run in our PARAFAC model. These EEMs have been corrected for inner filter effects. We also standardized fluorescence to Raman units. With this file, you are able to run the PARAFAC model on MATLAB as we did (see Murphy et al. (2013) for the tutorial).Note that in those datasets, we give you rounded values. Therefore, it is possible that you have slightly different values if you redo some calculations. |

**Table 2.** Description of the variables (i.e., columns) in EACH dataset in 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 column that describe each column.

Dataset filename: **CDOM\_Database.csv**

|  |  |  |
| --- | --- | --- |
| **Column name** | **Definition** | **Units** |
| Sample N° | Number of the sample | n/a |
| Sample name | Name of the sample | Text |
| Region | Sample origin (region) | Text |
| Site | Sample origin (site inside the region, if applicable) | Text |
| Ponds name | Name of the pond | Text |
| Type of ponds | The type of pond (*Thaw*, *Tundra* or *Bedrock*) | Text |
| Chla | Pelagic chlorophyll *a* concentration | µg L-1 |
| pH | Potential of hydrogen | n/a |
| Fe | Total dissolved iron concentration. For ponds without Fe measures, we used mean values from the literature, if they exist | mg L-1 |
| TP | Total phosphorus concentration | µg L-1 |
| TN | Total nitrogen concentration | mg L-1 |
| DOC | Dissolved organic carbon concentration | mg L-1 |
| Cuvette size | Size of the cuvette used for the absorbance scans on the spectrophotometer (2 sizes: 1 or 10 cm) | cm |
| SUVA254 | Specific ultraviolet absorbance at 254 nm from DOC normalized A254 | L mgC-1 m-1 |
| a320 | Absorption coefficient at 320 nm wavelength | m-1 |
| a440 | Absorption coefficient at 440 nm wavelength | m-1 |
| SR | Slope ratio, which is the ratio S285/S375(S285 is the spectral slope for the interval 275-295 nm ; (S375 is the spectral slope for the interval 350-400 nm) | n/a |
| S289 | Spectral slope for the interval 279-299 nm | nm-1 |
| FI | Fluorescence index, which is the ratio of fluorescence emission intensities at 450 nm and 500 nm for an excitation at 370 nm | n/a |
| [C1] | Maximum fluorescence of the component C1 identified by our PARAFAC model | Raman units (R.U.) |
| [C2] | Maximum fluorescence of the component C2 identified by our PARAFAC model | Raman units (R.U.) |
| [C3] | Maximum fluorescence of the component C3 identified by our PARAFAC model | Raman units (R.U.) |
| [C4] | Maximum fluorescence of the component C4 identified by our PARAFAC model | Raman units (R.U.) |
| [C5] | Maximum fluorescence of the component C5 identified by our PARAFAC model | Raman units (R.U.) |
| [Total] | Sum of the maximum fluorescence of the 5 components identified by our PARAFAC model for the corresponding sample | Raman units (R.U.) |
| %C1 | Relative abundance of the component C1 identified by our PARAFAC model | % |
| %C2 | Relative abundance of the component C2 identified by our PARAFAC model | % |
| %C3 | Relative abundance of the component C3 identified by our PARAFAC model | % |
| %C4 | Relative abundance of the component C4 identified by our PARAFAC model | % |
| %C5 | Relative abundance of the component C5 identified by our PARAFAC model | % |
| %Total | Sum of the relative abundance of the 5 components identified by our PARAFAC model for the corresponding sample. Must equal 100% | % |

Dataset filename: **Spectro\_Scans\_Raw\_800-0.4-250.4.csv**

|  |  |  |
| --- | --- | --- |
| **Column name** | **Definition** | **Units** |
| Wavelength (nm) | Wavelengths, from 800 to 250.4 nm (0.4 nm increment) | nm |
| BI09\_BYL1 | Absorbance measured at the specific wavelength for the sample BI09\_BYL1 | Absorbance units (A.U.) |
| BI09\_BYL22 | Absorbance measured at the specific wavelength for the sample BI09\_BYL22 | Absorbance units (A.U.) |
| … | … | Absorbance units (A.U.) |
| KW09\_KWK38 | Absorbance measured at the specific wavelength for the sample KW09\_KWK38 | Absorbance units (A.U.) |

Dataset filename: **Spectro\_Scans\_Raw\_800-1-250.csv**

|  |  |  |
| --- | --- | --- |
| **Column name** | **Definition** | **Units** |
| Wavelength (nm) | Wavelengths, from 800 to 250 nm (1 nm increment) | nm |
| KW14\_R1 | Absorbance measured at the specific wavelength for the sample KW14\_R1 | Absorbance units (A.U.) |
| KW14\_R2 | Absorbance measured at the specific wavelength for the sample KW14\_R2 | Absorbance units (A.U.) |
| … | … | Absorbance units (A.U.) |
| WH15\_WardHunt | Absorbance measured at the specific wavelength for the sample WH15\_WardHunt | Absorbance units (A.U.) |

Dataset filename: **Spectro\_Scans\_Raw\_800-2-250.csv**

|  |  |  |
| --- | --- | --- |
| **Column name** | **Definition** | **Units** |
| Wavelength (nm) | Wavelengths, from 800 to 250 nm (2 nm increments) | nm |
| RB02\_R1 | Absorbance measured at the specific wavelength for the sample RB02\_R1 | Absorbance units (A.U.) |
| RB02\_R3 | Absorbance measured at the specific wavelength for the sample RB02\_R3 | Absorbance units (A.U.) |
| … | … | Absorbance units (A.U.) |
| KW02\_T10 | Absorbance measured at the specific wavelength for the sample KW02\_T10 | Absorbance units (A.U.) |

Dataset filename: **SIA\_Data.csv**

|  |  |  |
| --- | --- | --- |
| **Column name** | **Definition** | **Units** |
| Type of samples | The type of sample analyzed (*DOM*, *Phytoplankton*, *Terrestrial*, *Macrophytes*, or *Benthic* *bulk*) | Text |
| Type of ponds | The type of pond (*Thaw*, *Tundra* or *Bedrock*) | Text |
| d13C Mean | Mean of the d13C signature for the type of sample and pond corresponding | ‰ |
| d13C SD | Standard deviation of the d13C signature for the type of sample and pond corresponding | ‰ |
| d2H Mean | Mean of the d2H signature for the type of sample and pond corresponding | ‰ |
| d2H SD | Standard deviation of the d2H signature for the type of sample and pond corresponding | ‰ |

**References**

Doucett, R. R., J. C. Marks, D. W. Blinn, M. Caron, and B. A. Hungate. 2007. Measuring terrestrial subsidies to aquatic food webs using stable isotopes of hydrogen. Ecology **88**:1587-1592.

Grosbois, G., P. A. del Giorgio, and M. Rautio. 2017. Zooplankton allochthony is spatially heterogeneous in a boreal lake. Freshwater Biology **62**:474-490.

Lepage, G., and C. C. Roy. 1984. Improved recovery of fatty acid through direct transesterification without prior extraction or purification. J Lipid Res **25**:1391-1396.

Loiselle, S. A., L. Bracchini, A. Cózar, A. M. Dattilo, A. Tognazzi, and C. Rossi. 2009. Variability in photobleaching yields and their related impacts on optical conditions in subtropical lakes. Journal of Photochemistry and Photobiology B: Biology **95**:129-137.

Murphy, K. R., K. D. Butler, R. G. M. Spencer, C. A. Stedmon, J. R. Boehme, and G. R. Aiken. 2010. Measurement of dissolved organic matter fluorescence in aquatic environments: An interlaboratory comparison. Environmental Science & Technology **44**:9405-9412.

Murphy, K. R., C. A. Stedmon, D. Graeber, and R. Bro. 2013. Fluorescence spectroscopy and multi-way techniques. PARAFAC. Analytical Methods **5**:6557-6566.

Nusch, E. A. 1980. Comparison of different methods for chlorophyll and phaeopigments determination. Archiv für Hydrobiologia **14**:14-35.

Ramnarine, R., R. P. Voroney, C. Wagner-Riddle, and K. E. Dunfield. 2011. Carbonate removal by acid fumigation for measuring the δ13C of soil organic carbon. Canadian Journal of Soil Science **91**:247-250.

Wilkinson, G. M., S. R. Carpenter, J. J. Cole, and M. L. Pace. 2014. Use of deep autochthonous resources by zooplankton: Results of a metalimnetic addition of 13C to a small lake. Limnology and Oceanography **59**:986-996.