# EDI Metadata Template (2020)[[1]](#footnote-1)

Data should be in csv text file. If starting with an Excel spreadsheet, please make sure it does not contain any formulas and comments on cells. If you need comments put them in their own column. If data were used in a database and major table linking is necessary to analyze, please de-normalize into a flat file, not just database table exports.

## Dataset Title

(be descriptive, more than 5 words):

## A unified dataset of co-located sewage pollution, periphyton, and benthic macroinvertebrate community and food web structure from Lake Baikal (Siberia)

## Abstract

(include what, why, where, when, and how)

## Sewage released from lakeside development can introduce nutrients and micropollutants that can restructure aquatic ecosystems. Lake Baikal, the world’s most ancient, biodiverse, and voluminous lake, has been experiencing localized sewage pollution from lakeside settlements.

## Increasing filamentous algal abundance suggests benthic communities are responding to this localized pollution. We surveyed 40-km of Lake Baikal’s southwestern shoreline 19-23 August 2015 for sewage indicators, including pharmaceuticals, personal care products, and microplastics with co-located periphyton, macroinvertebrate, stable isotope, and fatty acid sampling. Unique identifiers corresponding to sampling locations are retained throughout all data files to facilitate interoperability among the dataset’s 150+ variables. The data are structured in a tidy format (a tabular arrangement familiar to limnologists) to encourage future reuse. For Lake Baikal studies, these data can support continued monitoring and research efforts. For global studies of lakes, these data can help characterize sewage prevalence and ecological consequences of anthropogenic disturbance across spatial scales.

## Investigators

(list in order as for a paper with e-mail addresses, organization and preferably ORCID ID, if you don’t have one, get it, it’s easy and free: <http://orcid.org/>) add table rows as needed

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| First Name | Middle Initial | Last Name | Organization | e-mail address | ORCID ID (optional) |
| Michael | F | Meyer | School of the Environment, Washington State University | michael.f.meyer@wsu.edu |  |
| Ted |  | Ozersky | Large Lakes Observatory, University of Minnesota – Duluth | tozersky@d.umn.edu |  |
| Kara | H | Woo | Center for Environmental Research Education & Outreach, Washington State University | kara.woo@gmail.com |  |
| Kirill |  | Shchapov | Large Lakes Observatory, University of Minnesota – Duluth |  |  |
| Aaron | W. E. | Galloway | Oregon Institute of Marine Biology, University of Oregon |  |  |
| Julie | B | Schram | Oregon Institute of Marine Biology, University of Oregon |  |  |
| Daniel | D | Snow | School of Natural Resources, University of Nebraska-Lincoln |  |  |
| Maxim | A | Timofeyev | Biological Research Institute, Irkutsk State University |  |  |
| Dmitry | Yu. | Karnaukhov | Biological Research Institute, Irkutsk State University |  |  |
| Matthew | R | Brousil | Center for Environmental Research Education & Outreach, Washington State University | matthew.brousil@wsu.edu |  |
| Stephanie | E | Hampton | Center for Environmental Research Education & Outreach, Washington State University | s.hampton@wsu.edu |  |

## Other personnel names and roles

(dataset creators & contact, field crew, data entry etc. with e-mail addresses, organization and ORCID ID)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| First Name | Middle Initial | Last Name | Organization | e-mail address | ORCID ID (optional) | Role in project |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

## License

(Select a license for release of your data. We have 2 recommendations: [CCO – most accommodating of data reuse](https://creativecommons.org/publicdomain/zero/1.0/), & [CCBY – requires attribution](https://creativecommons.org/licenses/by/4.0/))

CCO

## Keywords

(List keywords and separate with commas. Using keywords from a controlled vocabulary (CV) will improve the future discovery and reuse of your data. The LTER CV is effective at describing ecological and environmental data. [Access the LTER CV here](http://vocab.lternet.edu/vocab/vocab/index.php). [Try this text mining service to extract LTER CV keywords from your abstract or methods](http://vocab.lternet.edu/keywordDistiller/). Additionally, please determine one or two keywords that best describe your lab, station, and/or project (e.g., Trout Lake Station, NTL LTER). This will help others discover your data by site/project).

Pharmaceuticals and Personal Care Products, Amphipoda, diatom, spirogyra, ulothrix, fatty acids, foodwebs

## Funding of this work:

Add rows to table if several grants were involved, list only the main PI, start with main grant first:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| PI First Name | PI Middle Initial | PI Last Name | PI ORCID ID (optional) | Title of Grant | Funding Agency | Funding Identification Number |
| Stephanie | E | Hampton |  | Dimensions of Biodiversity | NSF | NSF-DEB-1136637 |
| Michael | F | Meyer |  | Graduate Research Fellowship | NSF | NSF-DGE-1347973 |
| Michael | F | Meyer |  | Fulbright Fellowship | US State Department |  |
| Maxim | A | Timofeyev |  |  | Russian Ministry of Science and Education | N FZZE-2020-0026 |

## Timeframe

* Begin date: August 19 2015
* End date August 23 2015
* Data collection ongoing/completed: completed

## Geographic location

* Verbal description: Southwestern shore of Lake Baikal between the town of Listvyanka and the village of Bolshoe Goloustnoe
* North bounding coordinate (decimal degree): 52.02693
* South bounding coordinate (decimal degree): 51.85530
* East bounding coordinate (decimal degree): 105.4724
* West bounding coordinate (decimal degree): 104.8148

## Taxonomic species or groups

## Methods

(please be specific, include instrument descriptions, or point to a protocol online, if this is a data compilation please specify datasets used, preferably their DOI or URL plus general citation information)

*Site description*

The vast majority of Lake Baikal’s 2,000-km shoreline lacks lakeside development (Moore et al. 2009; Timoshkin et al. 2016). Our study focused on a 40-km section of Baikal’s southwestern shoreline, which included three settlements of different size (Figure 1). From 19 through 23 August 2015, we sampled 14 littoral and 3 pelagic locations along our 40-km transect.

At both pelagic and littoral sites, samples were collected for nutrient, chlorophyll, microplastic, and pharmaceutical and personal care product (PPCP) analysis. Samples were collected by hand from 0.75 m depth for each littoral site and with a bucket from aboard the ship for pelagic sites. Each water sample collection procedure is described below.

*Inverse distance weighted (IDW) population calculation for each sampling location*

We recognized that sewage indicator concentrations at each sampling location may be related to a sampling location’s spatial position relative to both the size and proximity of neighboring developed sites. Therefore, we created the inverse distance weighted (IDW) population metric to compress, into a single metric, information about human population size, density, and location along the shoreline as well as distance between developed sites and sampling locations.

Our workflow for calculating IDW population required five steps. First, we traced polygons and shorelines from satellite imagery for each developed site in Google Earth. Second, polygon and line geometries were downloaded from Google Earth as a .kml file. Third, the .kml file was imported into the R statistical environment (R Core Team, 2019), where using the sf package (Pebesma, 2018) we calculated shoreline length, polygon area, and centroid location for each developed site. Fourth, we joined point locations of each sampling site with the spatial polygons to calculate the distance from each sampling location to each developed site’s centroid. Fifth, we calculated IDW population for each sampling location, using formula (1).

(1)

where I is the IDW population at sampling location j, P is the population at each of the three developed sites Listvyanka (LI), Bolshie Koty (BK), Bolshoe Goloustnoe (BGO), A is the area of a developed site in km2, L is the shoreline length at a developed site in km, and D is the distance from developed site j to each developed site’s centroid in km.

*Nutrients*

Water samples for nutrient analyses were collected in 150 mL glass jars that had been washed with phosphate-free soap and rinsed three times with water from the sampling location. Samples were collected in duplicates and immediately frozen at -20°C until processing at the A.P.Vinogradov Institute of Geochemistry (Siberian Branch of the Russian Academy of Sciences, Irkutsk). Samples were not filtered prior to freezing, meaning that nitrogen and ammonium concentrations may potentially include intracellular nitrogen and overestimate dissolved nitrogenous forms in the water column.

For each water sample, nitrate, ammonium, and total phosphorus concentrations were measured. For ammonium (2016a) and nitrate (2017) concentrations, samples were analyzed with a spectrophotometer following the addition of Nessler’s reagent and disulfuric acid respectively. Total phosphorus concentration was measured with a spectrophotometer following the addition of persulfate (2016b).

*Chlorophyll a*

Water samples were collected in 1.5 L plastic bottles from a depth of approximately 0.75 m. Within 12 h of collection, three subsamples (up to 150 mL each) were filtered through 25-mm diameter, 0.2 µm pore size nitrocellulose filters. Filters were then placed in a 35 mm petri dish and frozen in the dark until processing.

Chlorophyll samples were processed in a manner similar to that of Parson (1963) and Lorenzen (1967). Nitrocellulose filters were ground in 90% acetone, in which chlorophyll extraction was allowed to proceed overnight. Samples were then centrifuged for 15-20 minutes. After centrifugation, absorbance of the chlorophyll extract was measured in a spectrophotometer at 630, 645, 665, and 750 nm. Concentrations were calculated using the formula: ; where A is the absorbance value of a particular wavelength, V1 is the volume of the filtered water, and V2 is the volume of extract. Concentrations are reported as mg/L.

*Pharmaceuticals and Personal Care Products (PPCPs)*

Water samples for PPCP analysis were collected in 250 mL amber glass bottles that were rinsed with either methanol or acetone and then three times with sample water prior to collections. Following collection, samples were refrigerated and kept in the dark until solid phase extraction (SPE).

Within 12 h of collection, samples were filtered directly from the amber glass bottle using a single-stream 25-mm GF/F SPE cartridge setup (Waters Corporation, Milford, MA). Lab personnel wore gloves and face masks to minimize contamination. Prior to filtration, SPE cartridges were primed with at least 5 mL of either methanol or acetone and then washed with at least 5 mL of sample water. Rate of SPE occurred at approximately 1 drop per second. Extraction proceeded until water could no longer pass through the SPE cartridge or until all collected water was filtered. Cartridges were stored in Whirlpacks at -20°C until analysis for PPCPs following methods of Lee et al. (2016).

*Microplastics*

At each location, samples were collected in triplicate using 1.5 L clear plastic bottles that were washed thoroughly with sample water before each collection. Samples were collected by hand for each littoral site and with a metal bucket from aboard the ship for pelagic sites.

For processing, each sample was vacuum filtered on to a 47-mm diameter GF/F filter. During filtration, aluminum foil was used to cover the filtration funnel to prevent contamination from airborne microplastic particles. After filtration, filters were dried under vacuum pressure and then stored in 50-mm petri dishes. Following filtration of all three replicates, the filtrate was collected and then re-filtered through a GF/F filter as a control for contamination from the plastic vacuum funnel or potentially airborne microplastics.

Microplastic counting involved visual inspection of the entire GF/F according to Van Cauwenberghe et al. (2015). Visual enumeration was conducted under a stereo microscope with ~100x magnification, and microplastics were classified into one of three categories: fibers, fragments, or beads. For all categories, plastics were defined as observed objects with apparent artificial colors, to avoid enumerating plastics contributed from the sampling bottle itself. Fibers were defined as smooth, long plastics with consistent diameters. Fragments were defined as plastics with irregularly sharp or jagged edges. Beads were defined as spherical plastics. During enumeration, GF/Fs remained in the petri dish to minimize potential for contamination from the air.

*Periphyton abundance*

At each littoral site, we haphazardly selected three rocks representative of local substrate. A plastic stencil was used to define a surface area of each rock from which we scraped a standardized 14.5 cm2 patch of periphyton. Samples were preserved with Lugol’s solution and stored in plastic scintillation vials. Additional periphyton was collected in composite from each site for fatty acid and stable isotope analysis.

Periphyton taxonomic identification and enumeration was performed by subsampling 10 μL aliquots from each preserved sample. Cells, filaments, and colonies were counted for each taxonomic group until at least 300 cells were identified. For all 10 μL aliquots, the entire subsample was counted, even if 300 cells were counted prior to completing the aliquot. Taxa were classified into broad categories consistent with Baikal algal taxonomy (Izhboldina, 2007), using coarse groupings to capture general patterns in relative algal abundance. As a result, algal groups consisted of diatoms, Ulothrix, Spirogyra, and the green algal Order Tetrasporales.

*Benthic macroinvertebrate abundance*

Three kick-net samples were collected for assessment of benthic community composition and abundance. Using a D-net, we collected macroinvertebrates by flipping over 1-3 rocks, and then sweeping five times in a left-to-right motion across approximately 1 m. After the series of sweeps, the catch was rinsed into a plastic bucket. For each replicate, bucket contents were concentrated using a 64-μm mesh and placed in glass jars with 40% ethanol (vodka; the only preservative available to us at the time) for preservation and refrigerated at 4°C aboard the research vessel. The 40% ethanol preservative was replaced with ~80% ethanol upon return to the lab within 24 to 48 hours, and samples were stored at ~4°C.

Invertebrate taxonomic identification and enumeration were performed under a stereo microscope. All invertebrates were identified to species with the exception of juveniles (Taakhteev, 2015 for amphipods; Sitnikova, 2012 for mollusks; Table 2). All samples contained oligochaetes and polychaetes, but due to poor preservation, these taxa were not counted. Six samples of the 42 collected were not well-preserved and were excluded from further analyses, in order to reduce errors in identification. KD-1 and LI-1 were the only sites with 1 sample counted. BK-2 and KD-2 each had two samples counted.

*Stable Isotope Analysis*

Measurements of δ15N and δ13C were performed on an elemental analyzer-isotope ratio mass spectrometer (EA-IRMS; Finnigan DELTAplus XP, Thermo Scientific) at the Large Lakes Observatory, University of Minnesota Duluth.

*Fatty Acid Analysis*

Following freeze-drying, samples were transferred to 10 mL glass centrifuge vials, and 2 mL of 100% chloroform was added to each under nitrogen gas. Samples were allowed to sit in chloroform overnight at -80°C. Fatty acid extractions generally involved three phases: (1) 100% chloroform extraction, (2) chloroform-methanol extraction, and (3) fatty acid methylation. Fatty acid extraction methods were adapted from Schram et al. (2018).

After overnight chloroform extraction, samples underwent a chloroform-methanol extraction three times. To each sample, we added 1 mL cooled 100% methanol, 1 mL chloroform:methanol solution (2:1), and 0.8 mL 0.9% NaCl solution. Samples were inverted three times and sonicated on ice for 10 minutes. Next, samples were vortexed for 1 minute, and centrifuged for 5 minutes (3,000 rpm) at 4°C. Using a double pipette technique, the lower organic layer was removed and kept under nitrogen. After the third extraction, samples were evaporated under nitrogen flow, and resuspended in 1.5 mL chloroform and stored at -20°C overnight.

Once resuspended in chloroform, 1 mL of chloroform extract was transferred to a glass centrifuge tube with a glass syringe as well as an internal standard of 4 μL of 19-carbon fatty acid. Samples were then evaporated under nitrogen, and then 1 mL of toluene and 2 mL of 1% sulfuric acid-methanol was added. The vial was closed under nitrogen gas and then incubated in 50°C water bath for 16 hours. After incubation, samples were removed from the bath, allowed to reach room temperature and stored on ice. Next, we performed a potassium carbonate-hexane extraction twice. To each sample, we added 2 mL of 2% potassium bicarbonate and 5 mL of 100% hexane, inverting the capped vial so as to mix the solution. Samples were centrifuged for 3 minutes (1,500 rpm) at 4°C. The upper hexane layer was then removed and placed in a vial to evaporate under nitrogen flow. Once almost evaporated, 1 mL of 100% hexane was added and stored in a glass amber autosampler vial for GC/MS quantification. GC/MS quantification was performed with a Shimadzu QP2020 GC/MS following Schram et al. (2018).

## Data Table

* Column name: exactly as it appears in the dataset. Please avoid special characters, dashes and spaces.
* Description: please be specific, it can be lengthy
* Unit: please avoid special characters and describe units in this pattern: e.g. microSiemenPerCentimeter, microgramsPerLiter, absoptionPerMolePerCentimeter
* Code explanation: if you use codes in your column, please explain in this way: e.g. LR=Little Rock Lake, A=Sample suspect, J=Nonstandard routine followed
* Date format: please tell us exactly how the date and time is formatted: e.g. mm/dd/yyyy hh:mm:ss plus the time zone and whether or not daylight savings was observed.
* Missing value code: If a code for ‘no data’ is used, please specify: e.g. -99999

Please add rows as needed

**Table name:** Chlorophyll data (chlorophylla.csv)

**Table description:** This .csv contains chlorophyll a data for each littoral and pelagic sampling location along Lake Baikal’s southwestern shoreline.

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Missing value code |
| site | Locational identifier for each sampling site |  |  |
| replicate | Replicate number for a given sampling site |  |  |
| filtered\_volume\_ml | Volume filtered for a given chlorophyll replicate | milliliters |  |
| sample\_volume\_ml | Volume of sample analyte during the chlorophyll extraction process | milliliters |  |
| raw\_fluo | Raw, uncorrected fluorometric reading for chlorophyll analysis |  |  |
| adjusted\_raw | Corrected fluorometric reading for chlorophyll analysis |  |  |
| chl\_conc | Chlorophyll a concentration | milligramsPerLiter |  |

**Table name:** Inverse-distance-weighted population (distance\_weighted\_population\_metrics.csv)

**Table description:** This .csv contains population data for each of the sampled locations. Although the majority of sites do not contain adjacent developments, we calculated inverse-distance-weighted population for each location based on neighboring settlements. A full description of the methods can be found in the associated manuscript (Meyer et al., XXXX).

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Missing value code |
| site | Locational identifier for each sampling site |  |  |
| distance\_weighted\_population | Inverse distance weighted population | NumberOfPeople |  |

**Table name:** Fatty Acid profiles (fatty\_acid.csv)

**Table description:** This file contains fatty acid concentrations for various benthic macroinvertebrate genera, periphyton, and endemic Draparnaldia spp. benthic algae collected from the 14 littoral sampling locations. A full description of the methods can be found in the associated manuscript (Meyer et al., XXXX).

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Missing value code |
| site | Locational identifier for each sampling site |  |  |
| Genus | Genus of the analyzed organism. Amphipods were identified to species, whereas snails and periphyton were kept as generalized groupings. Drapa spp. is an endemic alga and was likewise not identified to species. |  |  |
| Species | Amphipod species analyzed |  | NA |
| c12\_0 | Concentration of 12:0 fatty acid | microgramsPerMilligramOfTissue |  |
| i\_14\_0 | Concentration of i-14:0 fatty acid | microgramsPerMilligramOfTissue |  |
| c14\_0 | Concentration of 14:0 fatty acid | microgramsPerMilligramOfTissue |  |
| c14\_4n5 | Concentration of 14:4n-5 fatty acid | microgramsPerMilligramOfTissue |  |
| i\_15\_0 | Concentration of i-15:0 fatty acid | microgramsPerMilligramOfTissue |  |
| a\_15\_0 | Concentration of a-15:0 fatty acid | microgramsPerMilligramOfTissue |  |
| c15\_1w7 | Concentration of 15:1ω7 fatty acid | microgramsPerMilligramOfTissue |  |
| i\_16\_0 | Concentration of i-16:0 fatty acid | microgramsPerMilligramOfTissue |  |
| c16\_0 | Concentration of 16:0 fatty acid | microgramsPerMilligramOfTissue |  |
| c16\_1w9 | Concentration of 16:1ω9 fatty acid | microgramsPerMilligramOfTissue |  |
| c16\_1w8 | Concentration of 16:1ω8 fatty acid | microgramsPerMilligramOfTissue |  |
| c16\_1w7 | Concentration of 16:1ω7 fatty acid | microgramsPerMilligramOfTissue |  |
| c16\_1w6 | Concentration of 16:1ω6 fatty acid | microgramsPerMilligramOfTissue |  |
| c16\_1w5 | Concentration of 16:1ω5 fatty acid | microgramsPerMilligramOfTissue |  |
| i\_17\_0 | Concentration of i-17:0 fatty acid | microgramsPerMilligramOfTissue |  |
| a\_17\_0 | Concentration of a-17:0 fatty acid | microgramsPerMilligramOfTissue |  |
| c17\_0 | Concentration of 17:0 fatty acid | microgramsPerMilligramOfTissue |  |
| c17\_1w7 | Concentration of 17:1n-7 fatty acid | microgramsPerMilligramOfTissue |  |
| c16\_2w7 | Concentration of 16:2ω7 fatty acid | microgramsPerMilligramOfTissue |  |
| c16\_2w6 | Concentration of 16:2ω6 fatty acid | microgramsPerMilligramOfTissue |  |
| c16\_2w4 | Concentration of 16:2ω4 fatty acid | microgramsPerMilligramOfTissue |  |
| c16\_3w6 | Concentration of 16:3ω6 fatty acid | microgramsPerMilligramOfTissue |  |
| c16\_3w4 | Concentration of 16:3ω4 fatty acid | microgramsPerMilligramOfTissue |  |
| c16\_3w3 | Concentration of 16:3ω3 fatty acid | microgramsPerMilligramOfTissue |  |
| c16\_4w3 | Concentration of 16:4ω3 fatty acid | microgramsPerMilligramOfTissue |  |
| c16\_4w1 | Concentration of 16:4ω1 fatty acid | microgramsPerMilligramOfTissue |  |
| c18\_0 | Concentration of 18:0 fatty acid | microgramsPerMilligramOfTissue |  |
| c18\_1w9 | Concentration of 18:1ω9 fatty acid | microgramsPerMilligramOfTissue |  |
| c18\_1w7 | Concentration of 18:1ω7 fatty acid | microgramsPerMilligramOfTissue |  |
| c18\_2w6t | Concentration of 18:2ω6t fatty acid | microgramsPerMilligramOfTissue |  |
| c18\_2w6 | Concentration of 18:2ω6 fatty acid | microgramsPerMilligramOfTissue |  |
| c18\_3w6 | Concentration of 18:3ω6 fatty acid | microgramsPerMilligramOfTissue |  |
| c18\_3w3 | Concentration of 18:3ω3 fatty acid | microgramsPerMilligramOfTissue |  |
| c18\_4w4 | Concentration of 18:4ω4 fatty acid | microgramsPerMilligramOfTissue |  |
| c18\_4w3 | Concentration of 18:4ω3 fatty acid | microgramsPerMilligramOfTissue |  |
| c18\_5w3 | Concentration of 18:5ω3 fatty acid | microgramsPerMilligramOfTissue |  |
| c20\_0 | Concentration of 20:0 fatty acid | microgramsPerMilligramOfTissue |  |
| c20\_1w9 | Concentration of 20:1ω9 fatty acid | microgramsPerMilligramOfTissue |  |
| c20\_1w7 | Concentration of 20:1ω7 fatty acid | microgramsPerMilligramOfTissue |  |
| c20\_2w5\_11 | Concentration of 20:2 ω5(11) fatty acid | microgramsPerMilligramOfTissue |  |
| c20\_2w5\_13 | Concentration of 20: 2 ω5(13) fatty acid | microgramsPerMilligramOfTissue |  |
| c20\_2w6 | Concentration of 20:2ω6 fatty acid | microgramsPerMilligramOfTissue |  |
| c20\_3w6 | Concentration of 20:3ω6 fatty acid | microgramsPerMilligramOfTissue |  |
| c20\_4w6 | Concentration of 20:4ω6 fatty acid | microgramsPerMilligramOfTissue |  |
| c20\_3w3 | Concentration of 20:3ω3 fatty acid | microgramsPerMilligramOfTissue |  |
| c20\_4w3 | Concentration of 20:4ω3 fatty acid | microgramsPerMilligramOfTissue |  |
| c20\_5w3 | Concentration of 20:5ω3 fatty acid | microgramsPerMilligramOfTissue |  |
| c22\_0 | Concentration of 22:0 fatty acid | microgramsPerMilligramOfTissue |  |
| c22\_1w9 | Concentration of 22:1ω9 fatty acid | microgramsPerMilligramOfTissue |  |
| c22\_1w7 | Concentration of 22:1ω7 fatty acid | microgramsPerMilligramOfTissue |  |
| c22\_2w6 | Concentration of 22:2ω6 fatty acid | microgramsPerMilligramOfTissue |  |
| c22\_4w6 | Concentration of 22:4ω6 fatty acid | microgramsPerMilligramOfTissue |  |
| c22\_5w6 | Concentration of 22:5ω6 fatty acid | microgramsPerMilligramOfTissue |  |
| c22\_3w3 | Concentration of 22:3ω3 fatty acid | microgramsPerMilligramOfTissue |  |
| c22\_4w3 | Concentration of 22:4ω3 fatty acid | microgramsPerMilligramOfTissue |  |
| c22\_5w3 | Concentration of 22:5ω3 fatty acid | microgramsPerMilligramOfTissue |  |
| c22\_6w3 | Concentration of 22:6ω3 fatty acid | microgramsPerMilligramOfTissue |  |
| c24\_0 | Concentration of 24:0 fatty acid | microgramsPerMilligramOfTissue |  |
| comments | Data quality flag |  |  |

**Table name:** Macroinvertebrate species counts (invertebrates.csv)

**Table description:** This .csv contains benthic macroinvertebrate abundance data. A full description of the methods can be found in the associated manuscript (Meyer et al., XXXX).

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Missing value code |
| site | Locational identifier for each sampling site |  |  |
| replicate | Invertebrate replicate number for a given sampling site |  |  |
| Acroloxidae | Mollusk genus | NumerOfIndividuals |  |
| Asellidae | Endemic isopod genus | NumerOfIndividuals |  |
| Baicaliidae | Mollusk genus, most of which are endemic | NumerOfIndividuals |  |
| Benedictidae | Mollusk genus, most of which are endemic | NumerOfIndividuals |  |
| Brandtia\_latissima | Endemic amphipod species | NumerOfIndividuals |  |
| Brandtia\_parasitica\_parasitica | Endemic amphipod species | NumerOfIndividuals |  |
| Caddisflies | General grouping; were not identified to species. | NumerOfIndividuals |  |
| Cryptoropus\_inflatus | Endemic amphipod species | NumerOfIndividuals |  |
| Cryptoropus\_pachytus | Endemic amphipod species | NumerOfIndividuals |  |
| Cryptoropus\_rugosus | Endemic amphipod species | NumerOfIndividuals |  |
| Eulimnogammarus\_capreolus | Endemic amphipod species | NumerOfIndividuals |  |
| Eulimnogammarus\_cruentes | Endemic amphipod species | NumerOfIndividuals |  |
| Eulimnogammarus\_cyaneus | Endemic amphipod species | NumerOfIndividuals |  |
| Eulimnogammarus\_grandimanus | Endemic amphipod species | NumerOfIndividuals |  |
| Eulimnogammarus\_juveniles | Endemic amphipod genus. Identification kept at genus level so as to prevent misclassification. | NumerOfIndividuals |  |
| Eulimnogammarus\_maackii | Endemic amphipod species | NumerOfIndividuals |  |
| Eulimnogammarus\_marituji | Endemic amphipod species | NumerOfIndividuals |  |
| Eulimnogammarus\_verucossus | Endemic amphipod species | NumerOfIndividuals |  |
| Eulimnogammarus\_viridis\_viridis | Endemic amphipod species | NumerOfIndividuals |  |
| Eulimnogammarus\_vittatus | Endemic amphipod species | NumerOfIndividuals |  |
| Flatworms | Not identified beyond order. | NumerOfIndividuals |  |
| Leeches | Not identified beyond order, although 12 endemic species do exist. | NumerOfIndividuals |  |
| Maackia | Mollusk genus, most of which are endemic | NumerOfIndividuals |  |
| Pallasea\_brandtia\_brandita | Endemic amphipod species | NumerOfIndividuals |  |
| Pallasea\_brandtii\_tenera | Endemic amphipod species | NumerOfIndividuals |  |
| Pallasea\_cancelloides | Endemic amphipod species | NumerOfIndividuals |  |
| Pallasea\_cancellus | Endemic amphipod species | NumerOfIndividuals |  |
| Pallasea\_viridis | Endemic amphipod species | NumerOfIndividuals |  |
| Planorbidae | Mollusk genus, most of which are endemic | NumerOfIndividuals |  |
| Poekilogammarus\_crassimus | Endemic amphipod species | NumerOfIndividuals |  |
| Poekilogammarus\_ephippiatus | Endemic amphipod species | NumerOfIndividuals |  |
| Poekilogammarus\_juveniles | Endemic amphipod genus Identifying to species introduced risk of misclassification. | NumerOfIndividuals |  |
| Poekilogammarus\_megonychus\_perpolitus | Endemic amphipod species | NumerOfIndividuals |  |
| Poekilogammarus\_pictus | Endemic amphipod species | NumerOfIndividuals |  |
| Valvatidae | Mollusk genus, most of which are endemic | NumerOfIndividuals |  |

**Table name:** site-associated metadata (metadata.csv)

**Table description:** This .csv contains metadata for each of the sampled locations. A full description of the methods can be found in the associated manuscript (Meyer et al., XXXX).

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Missing value code |
| site | Locational identifier for each sampling site |  |  |
| year | Year sampling occurred | Year |  |
| month | Month sampling occurred | Month |  |
| day | Day sampling occurred | Day |  |
| time | Time sampling occurred | Hours:Minutes | NA |
| lat | Latitude of sampling location | DecimalDegrees |  |
| long | Longitude of sampling location | DecimalDegrees |  |
| site\_description | Researcher description of sampling location |  |  |
| distance\_to\_shore\_m | Distance from in situ sampled location to the shoreline | Meters | NA |
| depth\_m | Depth at in situ sampling location | Meters |  |
| air\_temp\_celsius | Temperature of air at sampling location | Celsius |  |
| surface\_temp\_celsius | Temperature of water’s surface at sampling location | Celsius | NA |
| mid\_temp\_celsius | Temperature of water midway between surface and bottom at sampling location | Celsius | NA |
| bottom\_temp\_celsius | Temperature of water near sediment at sampling location | Celsius | NA |
| comments | Notes in the field describing sampling conditions |  |  |
| shore\_photo | Whether or not photos of the shoreline were taken | YesOrNo |  |
| substrate\_photo | Whether or not photos of the substrate were taken | YesOrNo |  |
| sponges | Whether or not sponges were present at a sampling location | YesOrNo |  |
| brandtia | Whether or not brandtia (endemic amphipod) was present at a sampling location | YesOrNo |  |

**Table name:** Microplastics (microplastics.csv)

**Table description:** This file contains microplastics counts for each of the pelagic and littoral sampling locations. A full description of the methods can be found in the associated manuscript (Meyer et al., XXXX).

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Missing value code |
| site | Locational identifier for each sampling site |  |  |
| replicate | Replicate for a given location | Numerical values represent a replicate; the letter “C” represents a control. |  |
| fragments | Number of microplastic fragments observed | Number |  |
| fibers | Number of microplastic fibers observed | Number |  |
| beads | Number of microplastic beads observed | Number |  |
| comments | Observer comments while enumerating microplastics |  |  |
| volume\_filtered\_ml | Volume for a given replicate filtered | Milliliters |  |

**Table name:** Nutrient data (nutrients.csv)

**Table description:** This .csv contains nutrient data for each of the associated sampling locations. A full description of the methods can be found in the associated manuscript (Meyer et al., XXXX).

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Missing value code |
| site | Locational identifier for each sampling site |  |  |
| replicate | Replicate for a given location |  |  |
| nh4\_mg\_dm3 | Ammonium concentration | MilligramsPerCubicDecimeter |  |
| no3\_mg\_dm3 | Nitrate concentration | MilligramsPerCubicDecimeter |  |
| tp\_mg\_dm3 | Total phosphorus concentration | MilligramsPerCubicDecimeter |  |
| tpo43\_mg\_dm3 | Total phosphate concentration | MilligramsPerCubicDecimeter |  |

**Table name:** Periphyton abundance data (periphyton.csv)

**Table description:** This .csv contains periphyton abundance data for each of the sampled littoral sites. A full description of the methods can be found in the associated manuscript (Meyer et al., XXXX).

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Missing value code |
| site | Locational identifier for each sampling site |  |  |
| replicate | Replicate number for a given sampling site |  |  |
| subsamples\_counted | Number of 10 microliter subsamples counted for a given replicate |  | NA |
| diatom | Number of diatom cells counted for a given replicate | NumberOfCells | NA |
| spirogyra | Number spirogyra cells counted for a given replicate | NumberOfCells | NA |
| spirogyra\_filaments | Number of spirogyra filaments counted for a given replicate | NumberOfFilaments | NA |
| ulothrix | Number of ulothrix cells counted for a given replicate | NumberOfCells | NA |
| ulothrix\_filaments | Number of ulothrix filaments counted for a given replicate | NumberOfFilaments | NA |
| tetrasporales | Number of tetrasporales cells counted for a given replicate | NumberOfCells | NA |
| pediastrum | Number of pediastrum cells counted for a given replicate | NumberOfCells | NA |
| desmidales | Number of desmidales cells counted for a given replicate | NumberOfCells | NA |
| comments | Notes from the obsverer |  | NA |

**Table name:** Pharmaceutical and Personal Care Product (PPCP) data (ppcp.csv)

**Table description:** This .csv contains PPCP concentrations for each of the associated sampling locations. A full description of the methods can be found in the associated manuscript (Meyer et al., XXXX).

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Missing value code |
| site | Locational identifier for each sampling site |  |  |
| paraxanthine | Concentration of paraxanthine, also known as 1,7-dimethylxanthine. Paraxanthine is the main human metabolite of caffeine | MicrogramsPerLiter |  |
| acetaminophen | Concentration of acetaminophen, also known as paracetamol. | MicrogramsPerLiter |  |
| amphetamine | Concentration of amphetamine. | MicrogramsPerLiter |  |
| caffeine | Concentration of caffeine | MicrogramsPerLiter |  |
| carbamazepine | Concentration of carbamezapine | MicrogramsPerLiter |  |
| cimetidine | Concentration of cimetidine | MicrogramsPerLiter |  |
| cotinine | Concentration of cotinine, which is the main human metabolite of nicotine | MicrogramsPerLiter |  |
| diphenhydramine | Concentration of diphenhydramine | MicrogramsPerLiter |  |
| mda | Concentration of mda | MicrogramsPerLiter |  |
| mdma | Concentration of mdma | MicrogramsPerLiter |  |
| methamphetamine | Concentration of methamphetamine | MicrogramsPerLiter |  |
| morphine | Concentration of morphine | MicrogramsPerLiter |  |
| phenazone | Concentration of phenazone | MicrogramsPerLiter |  |
| sulfachloropyridazine | Concentration of sulfachloropyridazine | MicrogramsPerLiter |  |
| sulfamethazine | Concentration of *sulfamethazine* in micrograms per liter. | MicrogramsPerLiter |  |
| sulfamethoxazole | Concentration of sulfamethoxazole | MicrogramsPerLiter |  |
| thiabendazole | Concentration of thiabendazole | MicrogramsPerLiter |  |
| trimethoprim | Concentration of trimethoprim | MicrogramsPerLiter |  |
| collection\_year | Year sample was collected in the field | Year |  |
| collection\_month | Month sample was collected in the field | Month |  |
| collection\_day | Day sample was collected in the field | Day |  |
| analysis\_year | Year sample was analyzed | Year |  |
| analysis\_month | Month sample was analyzed | Month |  |
| analysis\_day | Day sample was analyzed | Day |  |

**Table name:** Stable Isotopes data (stable\_isotopes.csv)

**Table description:** This .csv contains carbon and nitrogen stable isotope data within periphyton and macroinvertebrate tissue for each of the associated sampling locations. A full description of the methods can be found in the associated manuscript (Meyer et al., XXXX).

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Missing value code |
| site | Locational identifier for each sampling site |  |  |
| C13 | Carbon 13 values |  |  |
| N15 | Nitrogen 15 values |  |  |
| Genus | Amphipod genus or Periphyton |  |  |
| Species | Amphipod species |  | NA |

**Table name:** Total Lipid data (total\_lipid.csv)

**Table description:** This .csv contains lipid gravimetry data for periphyton and benthic macroinvertebrate tissue for each of the associated sampling locations. A full description of the methods can be found in the associated manuscript (Meyer et al., XXXX).

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Missing value code |
| site | Locational identifier for each sampling site |  |  |
| Genus | Genus of the analyzed organism. |  |  |
| Species | Amphipod species |  | NA |
| total\_lipid\_mg\_per\_g | Total amount of lipids in a sample | MilligramsOfLipidPerGramOfTissue |  |
| deviation | Samples were weighed three times and deviation in measurement was calculated. | MilligramsOfLipidPerGramOfTissue |  |

## Scripts/code (software)

(List any software scripts/code you would like to archive along with your data. These may include processing scripts you wrote to create, clean, or analyze the data.)

|  |  |  |
| --- | --- | --- |
| File name | Description | Scripting language |
| 01\_data\_cleaning.R | This script aggregates data among replicates for a given sampling location. It outputs aggregated CSVs that are used for successive analytics scripts. | R |
| 02\_sewage\_indicator\_analysis.R | This script takes outputs from script 01 ti relate each sewage indicator with inverse distance weighted population. It outputs a plot with aggregated model results and plotted data for each sewage indicator. | R |
| 03\_community\_composition\_analysis.R | This script takes outputs from script 01 to relate sewage indicators and inverse distance weighted population with changes in benthic periphyton and macroinvertebrate community composition. | R |
| 04\_fatty\_acid\_analysis.R | This script takes outputs from script 01 to analyze benthic algae and macroinvertebrate fatty acid compositions, and then relates those fatty acid profiles with sewage indicators. | R |
| 05\_table\_formatting.R | This script formats descriptive tables using the outputs from script 01. | R |
| 06\_map\_making.R | This script generates a map of the sampled locations. | R |
| panel\_cor\_function.R | This script is sourced to analyze cross-correlation among variables. | R |

1. This document liberally borrows from similar documents at SBC and GCE [↑](#footnote-ref-1)