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 for “Biotic and thermal drivers alter zooplankton phenology in western Lake Erie”* |
| **URL of dataset** | *Provide the URL of the data repository for the dataset. For the manuscript review stage, it is acceptable to say that this is forthcoming upon decision at the first review stage* |
| **Abstract** | *This study examined how environmental change influenced the seasonality and phenology of Daphnia retrocurva, Dressenid spp. Veligers, Skistodiaptomus oregonensis, and Mesocyclops spp in the western basin of Lake Erie. Time series data from eight sites in western Lake Erie during 1995-2022 are provided for the four focal taxa as well as the following predictors: water temperature, phytoplankton with and without grazer deterrents, phytoplankton < 10 µm, zooplankton biomass (excluding predacious cladocerans), Leptodora kindtii, Bythotrephes longimanus, and planktivorous fish.* |
| **Keywords** | Phenology; Zooplankton; Eutrophication; Cyanobacteria; Bythotrephes; Lake Erie, LEPAS, Aquatic Ecology Laboratory |
| **Lead author for the dataset** | *Jim Hood* |
| **Title and position of lead author** | *Associate Professor* |
| **Organization and address of lead author** | *Aquatic Ecology Laboratory*  *Dept. of Evolution, Ecology, and Organismal Biology*  *Translational Data Analytics Institute*  *The Ohio State University*  *1314 Kinnear Road*  *Columbus, Ohio USA 43212* |
| **Email address of lead author** | [*Hood.211@osu.edu*](mailto:Hood.211@osu.edu) |
| **Additional authors or contributors to the dataset** | *J. Bailey* |
| **Organization associated with the data** | *The Ohio State University* |
| **Funding** | Federal Aid in Sport Fish Restoration Program (F-69-P, Fish Management in Ohio) administered jointly by the United States Fish and Wildlife Service and the ODNR-DOW (project FADX09 to JMH and S. Ludsin).  United States Environmental Protection Agency through an assistance agreement with the Ohio Environmental Protection Agency (OSU-FDERIE19 and OSU-FDERIE22 to JMH) |
| **License** | [***CCBY***](https://creativecommons.org/licenses/by/4.0/) *– requires attribution* |
| **Geographic location – verbal description** | *Western basin of Lake Erie* |
| **Geographic coverage bounding coordinates** | *41.44-41.87 N and -83.3 – -82.6 W* |
| **Time frame - Begin date** | *May 1995* |
| **Time frame - End date** | *Sept. 2022* |
| **General study design** | Zooplankton, phytoplankton, and temperature data were provided by the Lake Erie Plankton Abundance Study (LEPAS), which has monitored plankton and physicochemical variables biweekly at eight sites in the western basin of Lake Erie during May–September since 1995.  For estimates of planktivore density (catch per hectare), we used Ohio Department of Natural Resources–Division of Wildlife (ODNR-DOW) trawl data for sites nearby the western basin Lake Erie LEPAS sites. Data provided by ODNR-DOW. |
| **Methods description** | *The methods for the temperature and plankton dataset are described in “00c\_LEPAS Methods\_20221118.pdf”, which is in the data repository.*  *The methods for the fish dataset are described in:* (ODNR-DOW) Ohio Department of Natural Resources - Division of Wildlife. 2023. Ohio’s Lake Erie Fisheries 2022 Annual data report. Federal Aid in Fish Restoration Project F-69-P. Page 88. Ohio Department of Natural Resources, Division of Wildlife, Lake Eries Fisheries Units, Fairport Harbor and Sandusky. |
| **Laboratory, field, or other analytical methods** | *The lab, field, or other processing methods for each variable included in the data table are described extensively in ““00c\_LEPAS Methods\_20221118.pdf” which is available in the data repository.* |
| **Taxonomic species or groups** | *Daphnia retrocurva, Dressenid spp. Veligers, Skistodiaptomus oregonensis, Mesocyclops spp, Leptodora kindtii, and Bythotrephes longimanus,* |
| **Quality control** | Plankton data are validated (i.e., checked for completeness, correctness, and errors) at several steps. 1) When samples are brought back to the laboratory, we check to ensure that the sample is preserved and the label is intact and accurate. Data entered into the sample inventory is also checked for completeness. 2) When the count data is transferred from a paper datasheet to an electronic datasheet, the information is checked for anomalous data or sample information, and for entry errors. 3) When data is entered into the LEPAS database and biomass and densities are calculated, the data is checked for completeness, duplicates, missing information, and to make sure it corresponds to previous data.  Five percent of samples are recounted for QC. Recounts are conducted by a different research assistant and involve recounting a subsample without looking at the data from the first sample..  Prior to analyses, all data were screen for reasonableness and outliers. |
| **Additional information** | *N/A* |
|  |  |

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

***Data used to conduct these analyses are provided, however, The Lake Erie Plankton Abundance Study (LEPAS) database is not publicly available due to the sponsors restrictions, although data is available by request.***

Dataset filename: *01\_LEPASzpData.csv (in 01\_Data)*

Dataset description: Biomass data for four focal taxa at eight LEPAS sites in western Lake Erie during 1995-2022. *This dataset was generated from the LEPAS database, which is not publicly available.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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.* |
| Sample\_ID | Sample ID is a combination of date and site | - | - | text |  |
| Sample\_date | Sample date | - | - | month/day/2-digit year |  |
| Sample\_site | Sample site | - | - | text |  |
| Dretrocurva | *Daphnia retrocurva* biomass | (µg /L) | - | numeric | “NA” |
| Veliger | Dreissena spp. Veliger biomass | (µg /L) | - | numeric | “NA” |
| Mesocyclops | *Mesocyclops spp.* Biomass | (µg /L) | - | numeric | “NA” |
| Soregonensis | *Skistodiaptomus oregonensis* biomass | (µg /L) | - | numeric | “NA” |
| DOY | Day of year | - | - | numeric |  |
| Y | Four digit year | - | - | numeric |  |
| Chl\_a\_ugL\_L | Chlorophyll a | (µg /L) | - | numeric | “NA” |
| Secchi\_depth\_m | Secchi depth | m | - | numeric | “NA” |
| AvgTemp | Average of surface and bottom temperature | °C | - | numeric | “NA” |

Dataset filename: *22\_10\_18 Hood planktivore catch at LEPAS sites.xlsx*

Dataset description: *Planktivorous fish catch data from bottom trawls at LEPAS sites.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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.* |
| Crn | catch record number; | - | unique for each sample | Numeric | - |
| Year | Year | - | - | 4-digit | - |
| month | Month | - | - | Numeric | - |
| Day | Day | - | - | Numeric 1-31 | - |
| t\_mon | intended trawl month (occasionally, one trawl day has to be pushed into the next month due to weather; e.g. June trawl on July 1st). | - | - | Numeric 1-12 | - |
| Siteno | site number (older data may have same site in different locations, current sites are permanent) | - | - | text | - |
| Lat\_start | latitude at start of trawl | decimal degrees |  | Numeric |  |
| Lon\_start | longitude at start of trawl | decimal degrees |  | Numeric |  |
| Species | interagency species code |  | Codes are defined in the common\_name column | Text |  |
| Common\_name | Common names of species |  |  | Text |  |
| Agecode | age code |  | 1 = Age-0/YOY; 2 = Age-1/YAO; 3 = Age-1+; 2 = Age-2+; 5 = all ages | numeric |  |
| Ages | age code range |  |  | text |  |
| Fish\_per\_ha | |  | | --- | | Catches prior to 2002 standardized to new R/V Explorer using Fishing Power Corrections (Tyson et al. 2006). NOTE: some species and age classes may not have an associated FPC due to low catches during that study, resulting in catch numbers that be not be comparable before and after 2002. | |  | |  | |  | |  | |  | |  | |  |  | numeric |  |

Dataset filename: *06b\_PlanktonTempDat\_biWk.csv (in 03\_GeneratedData)*

Dataset description: *This dataset contains all of the temperature and plankton predictor data and was generated from the LEPAS database, which is not publicly available, with the script entitled “06b\_PredMunge\_Plankton\_BIWK”.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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.* |
|  | Row number |  |  |  |  |
| Sample\_ID | Sample ID is a combination of date and site | - | - | text |  |
| Sample\_date | Sample date | - | - | month/day/2-digit year |  |
| Sample\_site | Sample site | - | - | text |  |
| Response | Contains all predictor variable categories with the data in the column “Values” – meta data provided in the following columns |  | Codes in this column are explained in the next eight rows | text | No missing data |
| AvgTemp | Average of top and bottom temperature | °C |  |  |  |
| Byth | Bythotrphes longimanus biomass | µg/L |  |  |  |
| Chla | Chlorophyll a | µg/L |  |  |  |
| EdiblePP | Phytoplankton without grazer defenses (PPNGD) | mg/L |  |  |  |
| InEdiblePP | Phytoplankton with grazer defenses (PPGD) | mg/L |  |  |  |
| Lepto | Leptodora kindtii biomass | µg/L |  |  |  |
| MesoZPfood | Mesozooplankton biomass (excluding L. kindtii and B. lognimanus) | µg/L |  |  |  |
| PPLthen10 | Phytoplankton < 10 µm | mg/L |  |  |  |
| values | Values for the responses listed above | See units for each response |  | Numeric | No missing data |

**Scripts/code (software)**

A longer description of all scripts as well as the input and output files generated is provided in our “Read me” file entitled (“00a\_ReadMe”).

|  |  |  |
| --- | --- | --- |
| File name | Description | Scripting language |
| 02a\_Models\_Retrocurva.R | conducts the Bayesian GAMM hurdle models for D. retrocurva | *R* |
| 02b\_Models\_Oregonensis.R | conducts the Bayesian GAMM hurdle models for S. oregonensis | *R* |
| 02c\_Models\_Mesocyclops.R | conducts the Bayesian GAMM hurdle models for Mesocyclops spp. | *R* |
| 02d\_Models\_Veligers.R | conducts the Bayesian GAMM hurdle models for Dreissena spp. veligers | *R* |
| 03a\_GAMMpredictPeakBM\_Retrocurva.R | calculates DOY of peak biomass for D. retrocurva | *R* |
| 03b\_GAMMpredictPeakBM\_Oregonensis.R | calculates DOY of peak biomass for S. oregonensis | *R* |
| 03c\_GAMMpredictPeakBM\_Meso.R | calculates DOY of peak biomass for Mesocyclops | *R* |
| 03d\_GAMMpredictPeakBM\_Vel.R | calculates DOY of peak biomass for Dreissena spp. veligers | *R* |
| 04\_CummulativeBiomassAnalysis\_loess.R | Uses LOESS to predict the start, middle, and end of the season for four focal taxa | *R* |
| 05\_Fig2\_3\_S5\_SeasonalityPhenol\_loess.R | - Cleans up phenology metrics data and prepares for modeling  - Generates figures 2, 3, S5, S6, S7  - Models relationship between Biomass and start, middle, and end of season | *R* |
| 06a\_PredMunge\_Fish\_BIWK.R | prepares fish data | *R* |
| 06b\_PredMunge\_Plankton\_BIWK.R | – prepares Temperature, PPGD, PPNGD, PPL10, ZPBM, Bythotrephes, and Leptodora data.  - generates Fig. S6. | *R* |
| 07b\_PreparePhenologyPredictorsData\_BIWK.R | Combines phenology and predictor data | *R* |
| 08ab\_PredictPhenol\_Ret\_BiWk\_glmer.R | GLMM models predicting phenology for D. Retrocurva | *R* |
| 08bb\_PredictPhenol\_Oreg\_BiWk\_glmer.R | GLMM models predicting phenology for S. oregonensis | *R* |
| 08cb\_PredictPhenol\_Meso\_BiWk\_glmer.R | GLMM models predicting phenology for Mesocyclops | *R* |
| 08db\_PredictPhenol\_Vel\_BiWk\_glmer.R | GLMM models predicting phenology for Dreissena spp. veligers | *R* |
| 09b\_PhenologyPredictionPlots\_BiWk.R | Generates Fig. 5 | *R* |
| 10\_GAMS\_PhenologyTimeSeries\_GAMfits.R | Fits GAMMs to DOY15, DOY50 DOY85 time series | *R* |
| 10b\_GAMS\_PhenologyTimeSeries\_plots.R | - Generates Fig 4 | *R* |
| 11\_PredictorsInterannualPatterns\_MARSS.R | - Runs biased random walk models for all predictors  - Generates Fig. 1 & Fig. S4 | *R* |
| 12\_relBiomass\_FigS1 | Generates Fig. S1 | *R* |

**Notes and Comments:**

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