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** | No lake left behind? Lake protection in the continental US |
| **URL of dataset** | <https://github.com/cont-limno/FreshwaterConservation>; this will be published in a permanent archive with a DOI (e.g., Zenodo) when the manuscript is accepted for publication. Some data files are currently too large to be stored on github, but they will be part of the final archival. |
| **Abstract** | In 2010, the Convention on Biological Diversity stated that by 2020 "at least 17 percent of terrestrial and inland water areas, and 10 percent of coastal and marine areas, especially areas of particular importance for biodiversity and ecosystem services, are conserved through effectively and equitably managed, ecologically representative and well connected systems of protected areas and other effective area-based conservation measures, and integrated into the wider landscapes and seascapes." (Aichi target 11). This repository contains data and analysis scripts to assess lake protection in the continental US with respect to these conservation targets. |
| **Keywords** | freshwater conservation, inland waters, LakeCat, PADUS, pond, reservoir, protected area |
| **Lead author for the dataset** | Ian M. McCullough |
| **Title and position of lead author** | Postdoctoral research associate |
| **Organization and address of lead author** | Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824, USA |
| **Email address of lead author** | immccull@gmail.com |
| **Additional authors or contributors to the dataset** | Nicholas K. Skaff, Patricia A. Soranno, Kendra Spence Cheruvelil |
| **Organization associated with the data** | Department of Fisheries and Wildlife, Michigan State University |
| **Funding** | This project was supported by the US National Science Foundation Macrosystems Biology program (EF #1638679 and #1638554). PAS was supported by the USDA National Institute of Food and Agriculture, Hatch Project 101354. Lead PI: P. Soranno. Co-PI: K. Cheruvelil, P. Tan, J. Zhou, E. Stanley, C. Gries, N. Lottig, T. Wagner, E. Hanks, E. Schliep. |
| **License** | CCO |
| **Geographic location – verbal description** | Full continental US |
| **Geographic coverage bounding coordinates** |  |
| **Time frame - Begin date** | 2019 |
| **Time frame - End date** | 2019 |
| **General study design** | Analysis of the extent to which US protected areas currently meet international conservation targets for lakes |
| **Methods description** | Protected lakes were identified using the US Protected Areas Database (PADUS v. 1.4). Lake protection was analyzed across US states ecoregions used by the US Environmental Protection Agency’s National Aquatic Resource Survey. Characteristics of protected lakes were quantified using catchment data from LakeCat (Hill et al. 2018; <https://doi.org/10.1086/697966>). |
| **Laboratory, field, or other analytical methods** | This analysis was based on existing data. See final table in this document for R code descriptions. Additional details are embedded within individual scripts. |
| **Taxonomic species or groups** |  |
| **Quality control** |  |
| **Additional information** |  |

**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: *LakeConn\_1ha\_LAGOS\_NHD\_Xwalk.csv*

Dataset description: *Crosswalk table between lakes in LAGOS and National Hydrography Dataset (NHD); used to identify LAGOS lake connectivity class in NHD lakes.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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.* |
| lagoslakei | Unique LAGOS lake ID |  |  | Numeric | NA |
| LakeConnec | Lake connectivity class |  | lakes with inflow streams and at least one upstream lake (drainage lake/stream), lakes with inflow streams (drainage stream), lakes at the headwaters of stream networks with at least one outflow stream (headwater) and lakes with no inflows or outflows (isolated) | Text | NA |
| nhdplusv2\_reachcode | attribute from NHD; used for locating waterbody |  |  | Numeric | NA |
| nhdplusv2\_comid | Unique NHD lake ID |  |  | Numeric | NA |

Dataset filename: *LakeProtection\_byNARS\_perm.csv*

Dataset description: *output data file of protected lakes by National Aquatic Resource Survey (NARS) ecoregion. Lakes from the National Hydrography Dataset plus v2 (NHD). Contains only permanent lakes.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| Ecoregion | Ecoregion name, 3-letter abbreviation |  |  | Text | NA |
| nLakes | Number of lakes >= 1 ha |  |  | Numeric | NA |
| ProtectedLakes\_gap12\_ctr | Number of protected lakes based on lake centroids occurring in strictly protected areas |  |  | Numeric | NA |
| ProtectedLakes\_gap3\_ctr | Number of protected lakes based on lake centroids occurring in multi-use areas |  |  | Numeric | NA |
| ProtectedLakes\_gap12\_Cat80 | Number of protected lakes based on 80% catchment protection occurring in strictly protected areas |  |  | Numeric | NA |
| ProtectedLakes\_gap3\_Cat80 | Number of protected lakes based on 80% catchment protection occurring in multi-use areas |  |  | Numeric | NA |
| ProtectedLakes\_gap12\_Cat100 | Number of protected lakes based on full catchments occurring in strictly protected areas |  |  | Numeric | NA |
| ProtectedLakes\_gap3\_Cat100 | Number of protected lakes based on full catchments occurring in multi-use areas |  |  | Numeric | NA |
| unprotected\_lakes | Number of lakes with centroids occurring in neither strictly protected nor multi-use areas |  |  | Numeric | NA |
| PropProtected\_gap12\_ctr | Proportion of lakes protected based on lake centroids occurring in strictly protected areas |  |  | Numeric | NA |
| PropProtected\_gap3\_ctr | Proportion of lakes protected based on lake centroids occurring in multi-use areas |  |  | Numeric | NA |
| PropProtected\_gap12\_Cat80 | Proportion of lakes protected based on 80% catchment protection occurring in strictly protected areas |  |  | Numeric | NA |
| PropProtected\_gap3\_Cat80 | Proportion of lakes protected based on 80% catchment protection occurring in multi-use areas |  |  | Numeric | NA |
| PropProtected\_gap12\_Cat100 | Proportion of lakes protected based on full catchments occurring in strictly protected areas |  |  | Numeric | NA |
| PropProtected\_gap3\_Cat100 | Proportion of lakes protected based on full catchments occurring in multi-use areas |  |  | Numeric | NA |
| PropUnprotected | Proportion of lakes with centroids occurring in neither strictly protected nor multi-use areas |  |  | Numeric | NA |

Dataset filename: *LakeProtection\_byState\_perm.csv*

Dataset description: *Lakes from the National Hydrography Dataset plus v2 (NHD). Contains only permanent lakes.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| State | state name |  |  | Text | NA |
| nLakes | number of lakes >= 1 ha |  |  | Numeric | NA |
| Strict\_ctr | number and proportion (in parentheses) of protected lakes based on lake centroids occurring in strictly protected areas |  |  | Text | NA |
| Multi\_ctr | number and proportion (in parentheses) of protected lakes based on lake centroids occurring in multi-use areas |  |  | Text | NA |
| Stict\_Cat80 | number and proportion (in parentheses) of protected lakes based on 80% catchment protection occurring in strictly protected areas |  |  | Text | NA |
| Multi\_Cat80 | number and proportion (in parentheses) of protected lakes based on 80% catchment protection occurring in multi-use areas |  |  | Text | NA |
| Strict\_Cat100 | number and proportion (in parentheses) of protected lakes based on full catchments occurring in strictly protected areas |  |  | Text | NA |
| Multi\_Cat100 | number and proportion (in parentheses) of protected lakes based on full catchments occurring in multi-use areas |  |  | Text | NA |
| Unprotected | number and proportion (in parentheses) of unprotected lakes based on lake centroids occurring in neither strictly protected nor multi-use areas |  |  | Text | NA |

Dataset filename: *PADUS.csv*

Dataset description: *percent lake catchments and watersheds protected based on US Protected Areas Database (v 1.4) (PADUS). See dataset for details:* [*https://www.usgs.gov/core-science-systems/science-analytics-and-synthesis/gap/pad-us-data-manual*](https://www.usgs.gov/core-science-systems/science-analytics-and-synthesis/gap/pad-us-data-manual)*.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| COMID | unique NHD lake ID; used to link to lakes from the National Hydrography Dataset plus v2 (NHD). |  |  | Numeric | NA |
| CatAreaSqKm | catchment area | Square km |  | Numeric | NA |
| WsAreaSqKm | watershed area | Square km |  | Numeric | NA |
| PctNo\_GAP\_StatusCat | percent of catchment under any protection status (GAP 1-4) | % | GAP status 1: managed for biodiversity – disturbance events proceed or are mimicked (strict protection)  GAP status 2: managed for biodiversity – disturbance events suppressed (strict protection)  GAP status 3: managed for multiple uses – subject to extractive (e.g. mining or logging) or OHV use  GAP status 4: no known mandate for biodiversity protection (may include easements, land trusts, etc., but no permanent legal protection) | Numeric | NA |
| PctGAP\_Status1Cat | percent of catchment under GAP status 1 (strict protection) | % |  | Numeric | NA |
| PctGAP\_Status2Cat | percent of catchment under GAP status 2 (strict protection) | % |  | Numeric | NA |
| PctGAP\_Status3Cat | percent of catchment under GAP status 3 (multi-use) | % |  | Numeric | NA |
| PctGAP\_Status4Cat | percent of catchment under GAP status 4 (no permanent protection) | % |  | Numeric | NA |
| PctGAP\_Status1Ws | percent of watershed under GAP status 1 (strict protection) | % |  | Numeric | NA |
| PctGAP\_Status2Ws | percent of watershed under GAP status 2 (strict protection) | % |  | Numeric | NA |
| PctGAP\_Status3Ws | percent of watershed under GAP status 3 (multi-use) | % |  | Numeric | NA |
| PctGAP\_Status4Ws | percent of watershed under GAP status 4 (no permanent protection) | % |  | Numeric | NA |

Dataset filename: *state\_NARS\_COMID.csv*

Dataset description: *lookup table for state and ecoregion (National Aquatic Resources Survey; NARS) for each lake in the National Hydrography Dataset plus v2 (NHD)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| COMID | unique NHD lake ID |  |  | Numeric | NA |
| WSA9 | ecoregion name, 3-letter abbreviation |  |  | Text | NA |
| WSA9\_NAME | full ecoregion name |  |  | Text | NA |
| STUSPS | state name, 2-letter abbreviation (postal code) |  |  | Text | NA |

Dataset filename: *unprotected\_COMID.csv*

Dataset description: *table of unprotected lake IDs in the National Hydrography Dataset plus v2 (NHD)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| x (COMID) | unique NHD lake ID |  |  | Numeric | NA |

Dataset filename: *catchment\_table.csv*

Dataset description: *table of lake and catchment variables for analysis of protected lake characteristics from LakeCat See Hill et al. (2018) for detailed variable descriptions from LakeCat;* [*https://doi.org/10.1086/697966*](https://doi.org/10.1086/697966)*).*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| COMID | unique NHD lake ID |  |  | Numeric | NA |
| PctGAP\_Status12Cat | percent of catchment under strict protection (GAP status 1 or 2) | % | GAP status 1: managed for biodiversity – disturbance events proceed or are mimicked (strict protection)  GAP status 2: managed for biodiversity – disturbance events suppressed (strict protection)  GAP status 3: managed for multiple uses – subject to extractive (e.g. mining or logging) or OHV use | Numeric | NA |
| PctGAP\_Status3Cat | percent of catchment under multi-use (GAP status 3) | % |  | Numeric | NA |
| ProtectGAP12\_ctr | protected or unprotected based on lake centroid occurring in strictly protected area |  |  | Text | NA |
| ProtectGAP3\_ctr | protected or unprotected based on lake centroid occurring in multi-use area |  |  | Text | NA |
| ProtectGAP12Cat\_80 | protected or unprotected based on 80% catchment protection occurring in strictly protected area |  |  | Text | NA |
| ProtectGAP3Cat\_80 | protected or unprotected based on 80% catchment protection occurring in multi-use area |  |  | Text | NA |
| ProtectGAP12Cat\_100 | protected or unprotected based on full catchment occurring in strictly protected area |  |  | Text | NA |
| ProtectGAP3Cat\_100 | protected or unprotected based on full catchment occurring in multi-use area |  |  | Text | NA |
| Unprotected | protected or unprotected based on lake centroid occurring in strictly protected or multi-use area |  |  | Text | NA |
| AREASQKM | lake area (from NHD plus v2) | sq km |  | Numeric | NA |
| CatAreaSqKM | catchment area, square kilometers | sq km |  | Numeric | NA |
| DrainageRatio | lake area/watershed area ratio (approximation of drainage ratio) |  |  | Numeric | NA |
| ElevCat | mean catchment elevation (from NHD plus v2) | m |  | Numeric | NA |
| WetIndexCat | catchment topographic wetness index (from LakeCat WetIndx.csv) |  |  | Numeric | NA |
| PctTotalForest2011Cat | percent of catchment covered by forest based on 2011 NLCD (evergreen, mixed, deciduous) (from LakeCat NLCD2011.csv) | % |  | Numeric | NA |
| PctTotalAg2011Cat | percent of catchment covered by agriculture based on 2011 NLCD (cropland, hay) (from LakeCat NLCD2011.csv) | % |  | Numeric | NA |
| PctTotalWetland2011Cat | percent of catchment covered by wetlands based on 2011 NLCD (woody, herbaceous) (from LakeCat NLCD2011.csv) | % |  | Numeric | NA |
| PctConif2011Cat | percent of catchment covered by evergreen forest based on 2011 NLCD (from LakeCat NLCD2011.csv) | % |  | Numeric | NA |
| RdDensCat | catchment road density (2010) (from LakeCat RoadDensity.csv) | km/sq km |  | Numeric | NA |
| PctImp2011Cat | percent catchment covered by impervious surface (2011) (from LakeCat ImperviousSurfaces.csv) | % |  | Numeric | NA |
| RunoffCat | mean catchment runoff (1971-2010) (from LakeCat Runoff.csv) | mm |  | Numeric | NA |
| BFICat | catchment baseflow index (baseflow/total inflow) (2003) (%) (from LakeCat BFI.csv) | % |  | Numeric | NA |
| SN\_2008Cat | catchment sulfur and nitrogen deposition (2008) (from LakeCat NADP.csv) | kg/ha/yr |  | Numeric | NA |
| TotalPctFrstLossCat | percent catchment total forest loss (2002-2013) (from LakeCat ForestLossByYear0013.csv) | % |  | Numeric | NA |
| Precip8110Cat | catchment mean annual precipitation (1981-2010) (from LakeCat PRISM\_1981\_2010.csv) | mm |  | Numeric | NA |
| Tmean8110Cat | catchment mean annual air temperature (1981-2010) (from LakeCat PRISM\_1981\_2010.csv) | deg C |  | Numeric | NA |
| WSA9 | ecoregion name, 3-letter abbreviation |  |  | Text | NA |
| WSA9\_NAME | full ecoregion name |  |  | Text | NA |
| STUSPS | state name, 2-letter abbreviation (postal code) |  |  | Text | NA |

**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)** |
| LakeCat | <https://doi.org/10.1086/697966>) | Ryan Hill ([hill.ryan@epa.gov](mailto:hill.ryan@epa.gov)) | Ryan Hill ([hill.ryan@epa.gov](mailto:hill.ryan@epa.gov)) |
| LAGOS | <https://lagoslakes.org/products/data-products/> | Nicole Smith  ([nicole.j.smith@gmail.com](mailto:nicole.j.smith@gmail.com)) | Nicole Smith  ([nicole.j.smith@gmail.com](mailto:nicole.j.smith@gmail.com)) |
| NHD plus v2 | <https://www.epa.gov/waterdata/get-data> | US Geological Survey | US Geological Survey |
| PADUS v 1.4 | <https://www.usgs.gov/core-science-systems/science-analytics-and-synthesis/gap/science/pad-us-data-overview?qt-science_center_objects=0#qt-science_center_objects> | US Geological Survey | US Geological Survey |

**Scripts/code (software) –** *OPTIONAL*

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** |
| Rcode/LakeCat\_LAGOS\_connectivity\_crosswalk.R | Produces crosswalk table to match lake connectivity classes from LAGOS to NHD | R |
| **Rcode/LakeCat\_prep\_data\_4Nick.R** | Wrangle LakeCat and NHD data into single table for logistic regression analysis | R |
| **Rcode/LakeProtectionByStateEcoregion.R** | Executes custom functions to calculate lake protection across US states and ecoregions | R |
| **Rcode/LakeProtectionByStateEcoregion\_80pctProtect.R** | Same as previous script, but adjusted for the 80% catchment protection threshold | R |
| **Rcode/Lakes\_in\_ProtectedAreas\_2019.R** | Main script; calculates lake protection across different definitions of protection, produces figures | R |
| **Rcode/functions/protected\_lakes\_by\_NARS.R** | Custom function to calculate lake protection across ecoregions (National Aquatic Resource Survey) | R |
| **Rcode/functions/protected\_lakes\_by\_state.R** | Custom function to calculate lake protection across US states | R |
| **Logistic regression.R** | Logistic regression for characteristics of protected lakes across ecoregions | R |

**Notes and Comments:**

LakeCat files were downloaded from Hill et al. (2018) (cited above). Ryan Hill and Marc Weber of the US Environmental Protection Agency calculated percent catchment and watershed protection using PADUS data (v 1.4, cited above). Large files were compressed prior to uploading for archival.

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