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

**Instructions:**

Metadata provides enough structured information for other scientists to understand and use your data. To prepare your metadata, you will need to fill in the information in the tables below and take the followings steps:

1. Fill in the tables below for your dataset that you will be making available. If you have more than one dataset, then fill in information requested for Table 2 (the data dictionary) for each dataset.
2. Save this file in this RTF format and upload your metadata to the *L&O-Letters* website when you submit your manuscript.
3. Timing of depositing your data in a repository: You should submit your data to a repository at the time of submission, however, you do not need to provide the link to the data until the manuscript has received a decision of major or minor revision. During the review process, we will review your metadata. In some cases, reviewers may ask for the data during the review stage, at which point you need to make it available.

[PLEASE DELETE THESE INSTRUCTIONS ONCE YOU FILL THIS FORM IN]

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

|  |  |
| --- | --- |
| **Title of dataset** | *Matchups between Water Reflectance and Landsat data* |
| **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 dataset represents Remote Sensing reflectance (Rrs) data obtained from different providers worldwide matched with Landsat Collection 2 Level 2 product. The dataset was used in the study “*Validity of the Landsat surface reflectance archive for aquatic science: Implications for cloud-based analysis” to validate the Landsat product.   *Data were obtained after a matchup analysis described in Maciel et al. (2023) and, from an initial dataset of ~12000 points, 1100 matchups were obtained. It represents data filtered for:*   1. *.* |
| **Keywords** | Remote Sensing, Water Quality, Freshwaters, Trends, Time-series, Google Earth Engine. |
| **Lead author for the dataset** | *Daniel Andrade Maciel* |
| **Title and position of lead author** | *Ph.D candidate* |
| **Organization and address of lead author** | *National Institute for Space Research (INPE), São José dos Campos, São Paulo, Brazil.* |
| **Email address of lead author** | *daniel.maciel@inpe.br* |
| **Additional authors or contributors to the dataset** | *Nima Pahlevan; Claudio Clemente Faria Barbosa* |
| **Organization associated with the data** | *INPE/NASA GSFC /SSAI/USGS* |
| **Funding** | *List the principle investigator names, title of grant, funding agency, and funding identification number* |
| **License** | [***CCBY***](https://creativecommons.org/licenses/by/4.0/) |
| **Geographic location – verbal description** | *Global coverage* |
| **Geographic coverage bounding coordinates** | *-122.4366667, 177.404267, -45.87291, 58.79015 (xmin, xmax, ymin, ymax)* |
| **Time frame - Begin date** | *04-15-1994* |
| **Time frame - End date** | *12-21-2021* |
| **General study design** | *The aim of this study is to evaluate the validity of the Landsat Collection 2 Level 2 surface reflectance products for aquatic applications. For that, in-situ measured reflectance were matched with Landsat data and accuracy analysis was performed.* |
| **Methods description** | *This dataset is a compilation of satellite reflectance data and in-situ reflectance data.*  *In-situ reflectance data were matched with Landsat Collection 2 Level 2 based on several criteria described in Maciel et al. (2023). The in-situ data was measured by different groups worldwide and complete descriptions of these datasets are available in referenced papers:*  *Lehmann et al. (2023): GLORIA Dataset*  *Pahlevan et al. (2022): Augmented GLORIA Dataset*  *Maciel et al. (2021): LabISA dataset* |
| **Laboratory, field, or other analytical methods** | *For the in-situ measurement of remote sensing reflectance, the water leaving radiance, downwelling irradiance, and sometimes the sky radiance was measured and the Remote Sensing reflectance was calculated. For this measurement, different instruments (e.g., TriOS RAMSES, ASD FieldSpec) were used. The papers that published/used each dataset (GLORIA, Augmented GLORIA and LabISA) describes the instruments and protocols used for each measurement.*  *The satellite data was obtained from United States Geological Survey data hosted in Microsoft Planetary Computer. The product corresponds to the Collection 2 Level 2 surface reflectance data (i.e., corrected by atmospheric effects). For atmospheric correction LEDAPS algorithm is used for Landsat-5/TM and Landsat-7/ETM+ and LaSRC algorithm is used for Landsat-8/OLI and Landsat-9/OLI-2.* |
| **Taxonomic species or groups** | *Not applicable.* |
| **Quality control** | *The matchups were analyzed and removed if some filter criteria is match. For example:*   1. *Time difference between 2 days in inland waters* 2. *Time difference between 2 hours in coastal waters* 3. *Filter for only water pixels based on a 5x5 window and Landsat Pixel\_QA band.* 4. *Filter for removing negative values.* 5. *Filter for removing outliers by considering errors higher than 200% in green band as invalid* |
| **Additional information** | *Any additional information that may help future users of the data not included in the above rows, or in the table below.* |
|  |  |

**Table 2.** Data dictionary: description of the variables (i.e., columns) in EACH dataset. You must provide sufficient detail for another user to understand and use the data. If there are 10 variables (i.e., columns) in the dataset, then there should be 10 rows in this table that describe each column. Be sure to include all relevant information for your dataset, including the unique identifiers for your dataset or system, dates, replicate numbers, latitude and longitude of sampling locations, etc.

Dataset filename: *Matchups.csv*

Dataset description: *explain what is in this dataset*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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.* |
| ID | The unique ID of each stations | - | - | String | - |
| region | Region where the data was collected | - | US: United States  UK: United Kingdon | String | - |
| lat | Latitude of the station | Decimal degrees |  | EPSG:4326 | - |
| long | Longitude of the station | Decimal degrees |  | EPSG:4326 | - |
| Date | Date of the in situ measurement | - | - | MM/DD/YYYY |  |
| dif\_days | Difference (in days) between in situ and satellite overpass | Days | - | Days | - |
| tss | Concentration of total suspended sediments | (mgL-1) | - | - | NA |
| Chla | Concentration of Chlorophyll-a | (ugL-1) | - | - | NA |
| Secchi | Secchi Disk depth | Meters | - | - | NA |
| Sensor | Sensor correspondent to the matchup |  | TM: Landsat-5/TM  ETM+: Landsat-7/ETM+  OLI: Landsat-8/OLI  OLI2: Landsat-9/OLI-2 | - | - |
| Class | Optical Water Type Class | Class | 1: OWT1  2: OWT2  3: OWT3  4: OWT4  5: OWT5  6: OWT6  7: OWT7  NA: not classified |  | NA |
| CA\_insitu | In-situ measured Remote Sensing Reflectance simulated to the specific sensor (column ‘Sensor’ band) for CA band. | (sr-1) | - | Float | NA |
| Blue\_insitu | In-situ measured Remote Sensing Reflectance simulated to the specific sensor (column ‘Sensor’ band) for Blue band. | (sr-1) | - | Float | NA |
| Green\_insitu | In-situ measured Remote Sensing Reflectance simulated to the specific sensor (column ‘Sensor’ band) for Green band. | (sr-1) | - | Float | NA |
| Red\_insitu | In-situ measured Remote Sensing Reflectance simulated to the specific sensor (column ‘Sensor’ band) for Red band. | (sr-1) | - | Float | NA |
| NIR\_insitu | In-situ measured Remote Sensing Reflectance simulated to the specific sensor (column ‘Sensor’ band) for Near-infrared band. | (sr-1) | - | Float | NA |
| CA\_satellite | Satellite remote sensing reflectance extracted based on median values of a 5x5 pixel window for CA band of the specific sensor (column “Sensor”) | (sr-1) | - | Float | NA |
| Blue\_satellite | Satellite remote sensing reflectance extracted based on median values of a 5x5 pixel window for Blue band of the specific sensor (column “Sensor”) | (sr-1) | - | Float | NA |
| Green\_satellite | Satellite remote sensing reflectance extracted based on median values of a 5x5 pixel window for Green band of the specific sensor (column “Sensor”) | (sr-1) | - | Float | NA |
| Red\_satellite | Satellite remote sensing reflectance extracted based on median values of a 5x5 pixel window for Red band of the specific sensor (column “Sensor”) | (sr-1) | - | Float | NA |
| NIR\_satellite | Satellite remote sensing reflectance extracted based on median values of a 5x5 pixel window for Near-infrared band of the specific sensor (column “Sensor”) | (sr-1) | - | Float | 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)** |
| *GLORIA dataset* | *https://doi.pangaea.de/10.1594/PANGAEA.948492* | *Moritz Lehmann, Daniela Gurlin, Nima Pahlevan et al.* | [moritz.lehmann@gmail.com](mailto:moritz.lehmann@gmail.com) |
|  |  |  |  |

**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** |
| *00\_Planetary\_computer\_Extraction.R* | *Extract data from Microsoft Planetary Computer* | *R* |
| 01\_Generate\_correlations\_heatmap.R | Apply the validation used in the study | R |
| 02\_GEE\_Time\_Series.R | Create the plots used in the time-series generation | R |

**Notes and Comments:**

The full R project and Google Earth Engine codes are available on this paper GitHub: <https://github.com/dmaciel123/Landsat_Aquatic_Ref_Validation>

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