Metadata template[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	*Station_Details.csv*
URL of dataset	*https://github.com/spencer-tassone/RiverineHeatwaves/blob/main/Station_Details.csv*
Abstract	*Heatwaves are increasing in frequency, duration, and intensity in ocean, coastal, and lake ecosystems. While positive water temperature trends have been documented in many rivers, heatwaves have not been analyzed. This study examined heatwaves in rivers throughout the United States between 1996-2021. Riverine heatwaves increased in frequency over the study period, with significant increases in summer, in mid to high order streams, and at free-flowing sites and sites above a reservoir. The increase in heatwave frequency was accompanied by an increase in moderate strength heatwaves as well as an increasing trend in the average number of heatwave days each year. Riverine heatwaves were most likely to occur during periods of normal or below-normal discharge conditions and at sites with a mean annual discharge ≤ 250 m3 s-1. These results provide the first assessment of heatwaves in rivers for a large geographic area in the United States.* *The dataset described here provides USGS station specific details regarding location, reservoir position, and stahler stream order.*
Keywords	*heatwave, USGS, river, reservoir, stream order*
Lead author for the dataset	*Spencer J. Tassone*
Title and position of lead author	*Graduate Student*
Organization and address of lead author	*Department of Environmental Sciences* *University of Virginia* *Clark Hall* *291 McCormick Rd* *P.O. Box 400123* *Charlottesville, VA 22904*
Email address of lead author	*sjt7jc@virginia.edu*
Additional authors or contributors to the dataset	*NA*
Organization associated with the data	*United States Geological Survey (USGS), United States Army Corps of Engineers (USACE)*
Funding	*NA*
License	*CCO*
Geographic location – verbal description	*United States of America*

Geographic coverage bounding coordinates	*Alabama, Alaska, Arizona, Arkansas, California, Colorado, Georgia, Michigan, Montana, Nevada, New Jersey, New York, Oregon, Pennsylvania, South Carolina, Texas, Washington*
Time frame - Begin date	*01-01-1996*
Time frame - End date	*12-31-2021*
General study design	*This study analyzed publicly, and freely available USGS derived water temperature data to examine riverine heatwaves throughout the U.S. Two of the variables examined for trends in riverine heatwaves were site position relative to a reservoir, and stahler stream order. The USACE maintains the National Inventory of Dams database while the USGS maintains the NHDPlus High Resolution database which includes stahler stream orders. Both of the databases are publicly, and freely available. Both databases were accessed using ESRI ArcMap version 10.8 to extract information regarding each sites position relative to a reservoir and each sites stream order.*
Methods description	*Stahler stream order was determined for each site position using the [USGS NHDPlus High Resolution geospatial database](#) using ESRI ArcMap version 10.8. Similarly, site position relative to a reservoir was determined using the [USACE National Inventory of Dams (NID)](#) using ESRI ArcMap version 10.8. Categorical assignment of site position relative to a reservoir (i.e., above, below, none) was further determined from aerial photographic visual inspection of each site's location relative to the USACE NID.*
Laboratory, field, or other analytical methods	*The USGS conducts high-frequency surface water monitoring throughout the U.S. as part of its national water information system. All sites with daily mean water temperature records available for the 26-year period of 1996-2021 were identified using the R package 'dataRetrieval' version 2.7. Tidally influenced and lake sites were removed, as were all flagged data other than those 'Approved', 'Approved Revised', 'Approved Edited', or 'Provisional'. Sites with < 90% of their daily records were also excluded. Linear interpolation was applied to water temperature gaps ≤ 2 days. For larger gaps, multiple linear regression models were developed using 1 km^2 resolution, daily climate data using the R package 'daymetr' version 1.6. Only those sites with regressions where $R^2 ≥ 0.80$ were used in this analysis (mean ± SD R^2 = 0.91 ± 0.04). Seventy long-term water temperature sites were identified, resulting in a total of 1,820 station years of water temperature data available for analysis.* *Site specific variables included region, Stahler stream order, and position in landscape relative to a reservoir. Regional assignment was classified according to [historically climatic consistent regions of the U.S.](#) Stahler stream order was determined using the USGS NHDPlus High Resolution geospatial database using ESRI ArcMap version 10.8. Categorical assignment of site position relative to a reservoir (i.e., above, below, none) was determined from aerial photographic visual inspection of each site's location relative to the USACE NID in ArcMap. Dams are listed in the NID if 1) their failure would cause loss of human life, economic loss, environmental damage, or disruption of lifeline facilities, 2) are ≥ 25 feet in height and ≥ 15 acre-feet in storage, or 3) ≥ 50-acre feet storage and ≥ 6 feet in height. Visual inspection of sites was warranted as the NID may exclude low head dams or historical mill dams.*

Taxonomic species or groups	*NA*
Quality control	*Visual inspection of sites was warranted as the NID may exclude low head dams or historical mill dams.*
Additional information	*Many USGS site numbers begin with the number zero. Noted in table 2, all site numbers in this analysis < 10301500 (< line 52) begin with a zero.*

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

Dataset filename: *Station_Details.csv*
Dataset description: *This dataset provides USGS station specific details regarding location, stahler stream order, and position relative to a reservoir.*

Column name	Description	Units	Code explanation	Data format	Missing data code
station_nm	*Station name issued by the USGS*	*NA*	*NA*	*Character*	*NA*
site_no	*Site number issued by the USGS. All site numbers < 10301500 (< line 52) start with a zero which gets dropped in csv file format*	*NA*	*NA*	*Character*	*NA*
lat	*Site latitude provided by the USGS*	*Decimal degrees*	*NA*	*Numeric*	*NA*
long	*Site longitude provided by the USGS*	*Decimal degrees*	*NA*	*Numeric*	*NA*
state	*United States - state name where the site is located*	*NA*	*NA*	*Character*	*NA*
STUSAB	*United States - two letter state name abbreviation*	*NA*	*NA*	*Character*	*NA*
StreamOrder	*Stahler stream order for each site*	*NA*	*1 = stahler stream order 1, 2 = stahler stream order 2, etc.*	*Integer*	*NA*
Reservoir	*Site position relative to a reservoir.*	*NA*	*Above = site is above a reservoir, Below = site is below a reservoir, None = site is free-flowing*	*Character*	*NA*

Table 3. Data provenance

Dataset title	Dataset DOI or URL	Creator (name & email)	Contact (name & email)
United States Geological Survey – NHDPlus	*https://www.usgs.gov/national-hydrography/national-hydrography-dataset*	*United States Geological Survey, https://answers.usgs.gov/*	*United States Geological Survey, https://answers.usgs.gov/*
United States Army Corps of Engineers - National Inventory of Dams	*https://nid.usace.army.mil*	*United States Army Corps of Engineers, NID@USACE.ARMY.MIL*	*United States Army Corps of Engineers, NID@USACE.ARMY.MIL*

Table 4. Scripts/code (software)

File name	Description	Scripting language
DataPull.R	*This file pulls the daily mean water temperature and discharge (Q) data from USGS using the 'dataRetrieval' R package. Similarly, the data associated with station_details.csv gets appended to those sites that eventually meet the data standards of the analysis.*	*R*

Notes and Comments:

- The Riverine Heatwave GitHub page where all the code to download and analyze the data for the riverine heatwave project is here: https://github.com/spencer-tassone/RiverineHeatwaves

- The specific code and GitHub page where the station_details.csv was used is here: https://github.com/spencer-tassone/RiverineHeatwaves/blob/main/DataPull.R

- The GitHub link to the csv file used for the daymetr batch download is here: https://github.com/spencer-tassone/RiverineHeatwaves/blob/main/Station_Details.csv

- This metadata for the csv file above is also linked to the Riverine Heatwave GitHub here: https://github.com/spencer-tassone/RiverineHeatwaves/blob/main/Station_Details_METADATA.pdf