

User Guide to LAGOS-US NETWORKS v1: Data module of surface water networks characterizing connections among lakes, streams, and rivers in the conterminous U.S.



CITATION:

To cite this document, please use the following citation for the data package:

King, K.B.S., Wang, Q., Rodriguez, L.K., Haite, M., Danila, L., Tan, P.-N., Zhou, J., and Cheruvilil, K.S. 2021. LAGOS-US NETWORKS v1: Data module of surface water networks characterizing connections among lakes, streams, and rivers in the conterminous U.S. Environmental Data Initiative.

<https://doi.org/XXXXXX>. Dataset accessed XX/XX/2020.

Data creation and documentation contribution paragraph:

KBSK created the vision behind the database design and metrics for the NETWORKS module. She led the team in data creation, documentation, and validation processes for NETWORKS, and drafted text and figures. QW wrote the code and created the networks and metrics. LKR created the naming conventions for the module, wrote the data source and dictionary tables, drafted figures, assisted in metric validation, and drafted text. MH assisted in metric validation, wrote the EDI metadata, and drafted figures and tables. LD assisted in code writing and metric creation. PNT and JZ supervised the computer science students and provided feedback during data creation. KSC conceived of the data module and created the research team, provided the structure for the data and documentation, and supervised the project. All co-authors collectively framed project ideas and scope and reviewed the entire guide.

Acknowledgements:

We thank Dana Infante and Arthur Cooper for their dam database and expertise, Arika Hawkins for her hourly help with manual dam classification, Nicole Smith for GIS support, and Katherine Webster for her QA/QC support. We also thank the Continental Limnology Project team for contributions and discussions throughout the development of this data module.

Overview of the LAGOS-US research platform

LAGOS-US provides an extensible research-ready platform to study the 479,950 lakes and reservoirs larger than or equal to 1 ha in the conterminous U.S. at multiple scales of space and time and at broad spatial extents. Although lakes are the focal unit of study, studying land-water interactions requires not only *in situ* lake water quality measurements, but also descriptions of the lakes, their watersheds, and their landscape ecological context (i.e., the land use, geologic, climatic, and hydrologic setting of lakes). Each lake's ecological context can be characterized at a variety of spatial extents (e.g., ecoregions, watersheds), which we call spatial divisions. Some of these ecological context variables are relatively static through time and are therefore characterized for a single date, whereas others are dynamic through time and are characterized at multiple time steps. Whenever possible, we include data for all lakes which can be used as the 'census' population of lakes ≥ 1 ha in the conterminous U.S.

The LAGOS-US research platform includes three core data modules (Figure I):

1. LAGOS-US GEO for geospatial ecological context at multiple spatial and temporal scales for lakes and their watersheds,
2. LAGOS-US LOCUS for location, identifiers, and physical characteristics of lakes and their watersheds, and
3. LAGOS-US LIMNO for *in situ* lake surface-water physical, chemical, and biological measurements through time.

These LAGOS-US core modules were created using existing datasets from a variety of data sources, such as national-scale climate, land use/cover, and hydrology data, as well as government, tribal, and community science lake data. In building this research platform, we followed a set of three fundamental principles that are similar to those used to create LAGOS-NE, an earlier version of the database system for a subset of U.S. states (Soranno et al. 2015, 2017). The first principle is that LAGOS-US should be based on a foundation of open science by which we make our data publicly available when each module is completed, error-checked, and documented. This includes providing a permanent identifier and a versioning system (e.g. v1) to facilitate future reuse of the data. Second, we document and describe the original data sources, our methods for integrating data, possible errors that may exist in the data, and we provide code for such methods, when possible. Third, we preserve the provenance of the original data as much as possible.

The LAGOS-US research platform was designed to be modular, i.e., each data module is made of variables that were derived using similar methods or data sources. This modularity facilitates documentation of the entire database and makes the data tables of manageable size. In addition, LAGOS-US was designed to be easily extensible, i.e., to allow other users to build extension modules that can be easily integrated into the LAGOS-US research platform. Future extension modules will be able to connect to any core module of LAGOS-US through common identifiers.

The LAGOS-US research platform has three extension modules currently in development by members of our team and that connect to the LOCUS module via a unique lake identifier (Figure I):

4. LAGOS-US LAKE DEPTH for observed mean and maximum depths of lakes,
5. LAGOS-US RESERVOIR for classifying lakes as natural lakes or reservoirs, and
6. LAGOS-US NETWORKS for surface water networks characterizing connections among lakes, streams, and rivers, including dams.

The LAGOS-US Platform

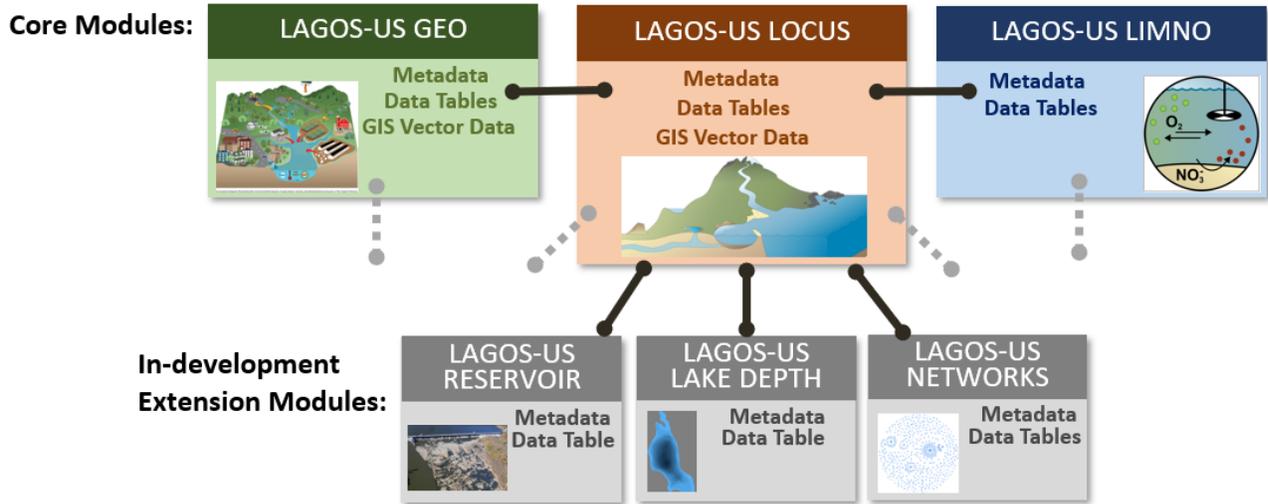


Figure I. The LAGOS-US Platform includes three core LAGOS-US modules: GEO, LOCUS, and LIMNO for geospatial, locational, and limnological data, respectively. Three in-development extension modules that will connect to LOCUS are shown in gray boxes along the bottom (RESERVOIR, LAKE DEPTH, and NETWORKS), with future extension modules connected to LOCUS and other core modules indicated by gray dashed lines. The major categories of tables within each module (each with many variables) are shown in bold. All core and in-development extension modules are connected via a unique lake identifier. Images for GEO, LOCUS, LIMNO, and RESERVOIR are from the Integration and Application Network, University of Maryland Center for Environmental Science (ian.umces.edu/imagelibrary/IAN image library). The image for LAKE DEPTH is courtesy of Joe Stachelek and for NETWORKS is courtesy of Laura Danila.

Description of LAGOS-US NETWORKS

This documentation is written as a guide for potential users of the LAGOS-US NETWORKS v1 module. We have structured the documentation into sections that start as an overview of the module and become more detailed (e.g., methods) as you move through the document. The numbered sections and subsections, as well as the table of contents can be used for navigation. Additional context for the motivation and purpose of this module are found in the associated data paper (King et al. under review) and additional methodological details are found in associated code (Wang & King 2020).

Knowing the degree of surface water connectivity among aquatic ecosystems can help scientists better understand and predict the movement of materials and biota across ecosystems. Connectivity metrics that incorporate both streams and lakes to describe surface water network structure at broad-spatial scales are needed to better understand the influence of connectivity on biotic and abiotic ecosystem properties. Additionally, barriers to movement, such as dams, can prohibit movement within a network. Methods to quantify surface water networks that include lake and stream connections at broad spatial scales are rare because it is difficult to balance accurate estimates of surface water connectivity and computational challenges. The NETWORKS module fills this data gap by quantifying surface connectivity metrics for lake networks, including dams, across the conterminous United States.

We applied a graph theory approach to identify lake networks for lakes ≥ 1 ha in the conterminous US using the medium resolution National Hydrography Dataset (NHDplusV2; 1-100,00-scale; abbreviated as “medres”) and the National Anthropogenic Barrier Dataset (NABD; Ostroff et al. 2013), and subsequently derived surface water connectivity metrics for individual lakes and entire lake networks. We define ‘lake networks’ as a set of lakes connected by ephemeral or permanent streams regardless of the directionality of those connections (e.g. upstream, downstream, or both), excluding connections through the Great Lakes, oceans, or estuaries. Networks do not stop at dams; therefore, these data are broadly applicable to both abiota and biota. Using this definition and our analytical approach, we created 898 networks that include 86,511 lakes, with the networks ranging in size from 2 to 32,811 lakes, the largest network being the Mississippi River basin (Figure 1).

The NETWORKS module includes one table with metrics for connected lakes, such as connections among lakes (both upstream and downstream), connections to dams, and network position (spatial orientation of the lake within its network). Although there are 479,950 total lakes ≥ 1 ha in the conterminous US, NETWORKS includes only those that are connected to other lakes and does not include isolated lakes or lakes only connected to streams. Two additional tables include a bidirectional and unidirectional distance between every pair of connected lakes. The fourth table is a flow table that describes the flow path direction between two flowlines (e.g., TO and FROM) that was used to create the networks and metrics in the other three tables. All metrics can be linked to individual lakes through the LAGOS-US LOCUS database via [lagoslakeids](#) (Smith et al. 2020).

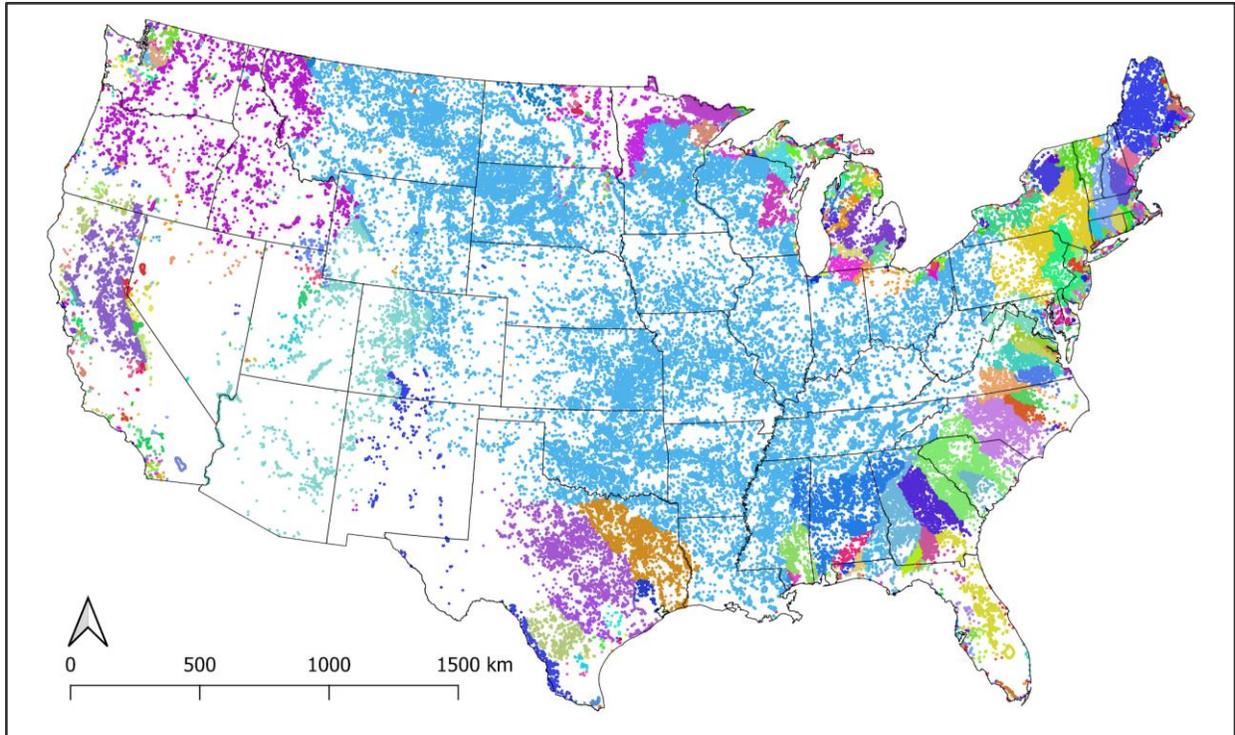


Figure 1. Lakes (n=86,511) in the LAGOS-US NETWORKS module, colored according to their network membership (n=898 networks). Although there are 479,950 total lakes ≥ 1 ha in the conterminous US, NETWORKS includes only those that are connected to other lakes.

1. Database Design for LAGOS-US NETWORKS

1.1. Definition of terms relating to entities

1.1.1. Study area

Although LAGOS-US LOCUS includes the 48 states in the conterminous U.S., ecologically relevant boundaries were delineated rather than using political boundaries. This approach resulted in the inclusion of some lakes that cross the border into either Canada or Mexico (Smith et al. 2020). However, we have excluded the Great Lakes and the oceans, meaning connections through these water bodies are not included in the NETWORKS module.

1.1.2. Lake

A 'lake' in LAGOS-US is a perennial body of relatively still water. LAGOS includes lakes and reservoirs that range from being completely natural to highly modified: lake basins can be entirely natural, modified natural (i.e., a water control structure on a natural lake), or a fully impounded stream or river (i.e., a reservoir). However, LAGOS explicitly excludes: sewage treatment ponds, aquaculture ponds, or other such retention ponds that are known to contain basins that are entirely artificial and were built for high-intensity human use. This definition of 'lake' has been developed for the purpose of the LAGOS-US database and its applications (e.g., to study lakes at macroscales) and is based on geographic representations of lakes from the high-resolution U.S. National Hydrography Dataset snapshot from 2016, which has an operational minimum lake size of 1 ha. For LAGOS-US NETWORKS, the medium

resolution (NHDplus V2; 1-100,000-scale) data was used, therefore, lakes are a subset of the entire LAGOS-US population.

1.1.3. Network

We define a ‘lake network’ or ‘network’ as a set of lakes connected by ephemeral or permanent streams regardless of the directionality of those connections (e.g. upstream, downstream, or both), excluding connections through the Great Lakes, oceans, or estuaries (Figure 2a). Lake networks were made using graph networks, consisting of nodes (lakes and streams) as well as stream lengths (Figure 2b), which were obtained from the NHDplusV2 flow table.

To make the networks, we used a bidirectional graph, where the graph traverses both downstream and upstream, including all lakes connected through streams. For example, in Figure 2, lake A, B, and C would not be connected to lake D and E in a unidirectional graph, but a bidirectional graph traverses downstream and upstream to connect Lake C to Lake D. Note that since we are characterizing *lake* networks, streams were not included in the graph when they do not connect to another lake (Figure 2b).

1.1.4. Entity Identifiers

The lakes are identified using a unique identifier ([lagoslakeid](#)) created for the LAGOS-US framework that was created from a snapshot of the lakes present in the NHD high resolution as of spring 2016 (Smith et al. 2020). However, the NETWORKS module was created using the medium resolution NHDplusV2 data; therefore, we also include the NHDplusV2 identifiers, which were matched to [lagoslakeids](#) using the [lake_link](#) table from the LAGOS-US LOCUS module (Smith et al. 2020).

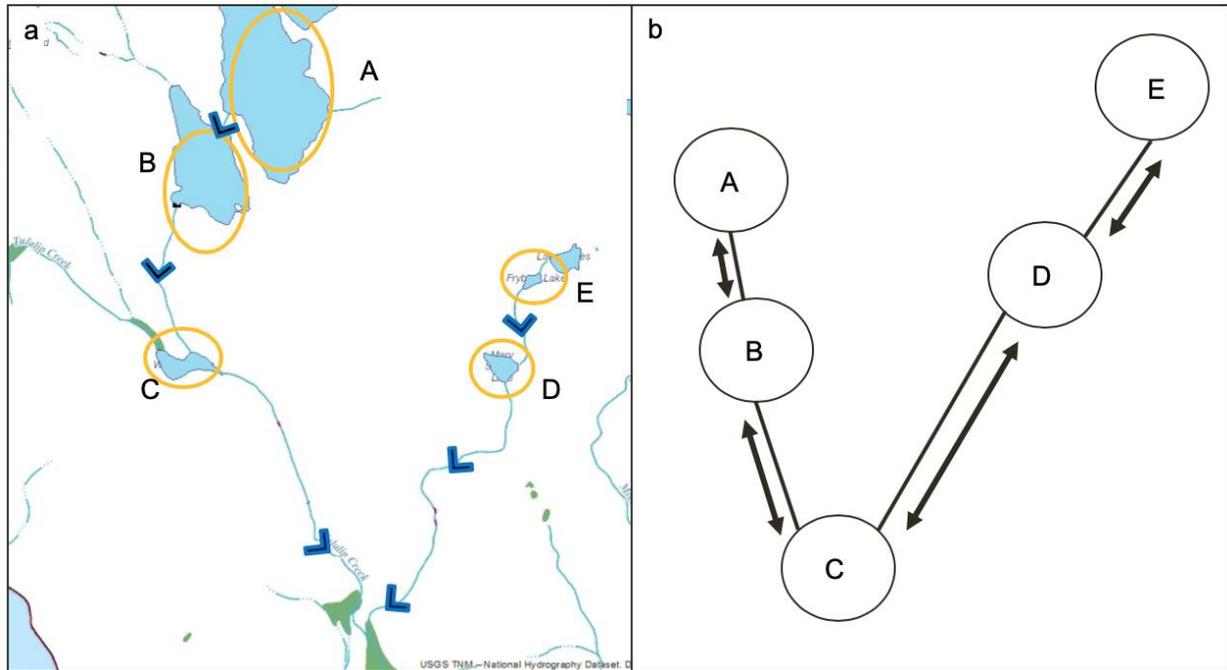


Figure 2. An example of (a) a lake network and (b) the bidirectional graph used to create the network, to illustrate how networks were defined. Because bidirectional graphs traverse downstream and upstream, lake C and lake D are connected. Note that since we are characterizing lake networks, streams were not included in the graph when they do not connect to another lake (example: Tulalip Creek flowing into lake C in (a) is not included in the graph in (b)).

1.2. Overview of data tables and variables

The variables in the NETWORKS module are organized into four machine-readable comma-separated values (CSV) data tables.

- 1.2.1. *nets_networkmetrics_medres* - This table includes connectivity metrics for each lake, as well as network-level metrics. We include lake identifier information, upstream and downstream connectivity metrics, upstream and downstream dams, and network metrics for the NHDplusV2 medium resolution (abbreviated medres) data.
- 1.2.2. *nets_binetworkdistance_medres* - This table includes distances between all pairs of connected lakes using a bidirectional graph. We include lake identifier information as well as upstream and downstream distances between pairs of connected lakes for the NHDplusV2 medium resolution data.
- 1.2.3. *nets_uninetworkdistance_medres* - This table includes distances between all pairs of connected lakes using a unidirectional graph. We include lake identifier information as well as the downstream distances between pairs of connected lakes for the NHDplusV2 medium resolution data.
- 1.2.4. *nets_flow_medres* - This table characterizes the downstream flow between every ‘flowline’, whereby flowlines include NHDplusV2 streams/rivers and artificial flowlines that show the flow path through a lake. We include the NHDplusV2 common identifiers for NHDFlowlines that describe the flow path direction between two flowlines (e.g., TO and FROM) that were used to create the networks and associated metrics in the other three tables.

1.3. Module data schema

The NETWORKS module consists of data and metadata located in two metadata tables and four data tables (Figure 3).

- 1.3.1. **Data Dictionary** (metadata) - Provides a definition for each variable name or ‘column’ of every table in the module, and includes other useful information such as units.
- 1.3.2. **Source Table** (metadata) - Includes a description of the sources used to create NETWORKS .
- 1.3.3. **Data tables** (data) - Contains the observations of the variables. This module contains four data tables that can be linked with other LAGOS modules through the common LAGOS-US-LOCUS identifier [lagoslakeid](#) (Figure 3).

LAGOS-US_{NETS}

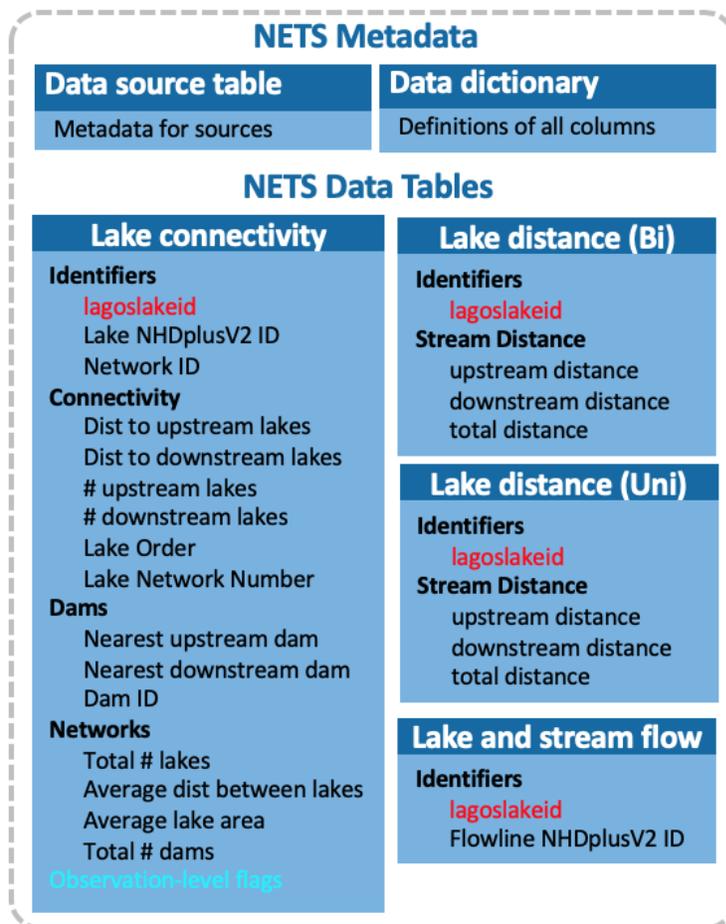


Figure 3. The LAGOS-US NETWORKS schema. NETWORKS includes metadata in the form of a source table and a data dictionary and four data tables (*nets_networkmetrics_medres*, *nets_binetworkdistance_medres*, *nets_uninetworkdistance_medres*, and *nets_flow_medres*). The tables are connected to each other and other LAGOS-US modules via *lagoslakeid*, depicted with red text. The lake connectivity data table also includes observation-level flags (section 2.6.), depicted with blue text. The variables in black text are grouped according to themes and are representative examples of variables included in the four data tables. See Table 1-4 for more details.

2. Data in LAGOS-US NETWORKS

The NETWORKS module contains information on all lakes in the conterminous U.S. ≥ 1 ha that are a part of a lake network as well as stream course distances (i.e. distance along a stream) derived from the medium resolution NHDplusV2. Here we describe the data used to create NETWORKS, as well as our approach for naming and organizing the data. Metrics for LAGOS-US NETWORKS were derived using data from three external sources, described below. These metrics were divided into four data tables, each with a unique naming convention (section 2.4). Furthermore, some observations of the data were flagged in order for users to be aware of unconventional results in the data table that may be of interest to users.

2.1. Data sources

2.1.1. National Hydrography Dataset, Medium Resolution (NHDplusV2)

The NETWORKS data product was derived from the lake and stream flow tables from the medium resolution U.S. National Hydrography Dataset (NHDplusV2) downloaded August 5, 2019. NHDplusV2 is a U.S. national geospatial surface water dataset that integrates information from the NHD, National Elevation Dataset (NED), and the Watershed Boundary Dataset (WBD) at a 1:100,000-scale. This data product was used to construct and validate the surface water connections for LAGOS-US NETWORKS.

Citation: U.S. Geological Survey. 2019. National Hydrography Dataset (NHDPlus Version 2.1 for Hydrologic Unit (HU) 4 - 2001 (published 20191002)). <https://www.epa.gov/waterdata/nhdplus-national-data>. Accessed August 5, 2019.

2.1.2. LAGOS-US LOCUS v1

Lakes, their identifiers, and surface area used in the creation of NETWORKS were from the LAGOS-US LOCUS module v1. This module includes lakes that were sources from the high resolution National Hydrography Dataset (NHD), a geospatial vector dataset used to map the nation's surface waters and hydrologic drainage areas.

Citation: Smith, N.J., K.E. Webster, L. Rodriguez, K.S. Cheruvilil, and P.A. Soranno. 2020. LAGOS-US LOCUS v1: Module of location, identifiers, and physical characteristics of lakes and their watersheds in the conterminous U.S. Environmental Data Initiative. Accessed May 13, 2019.

2.1.3. National Anthropogenic Barrier Dataset (NABD)

The NABD was used to establish the population of dams ($n=49,525$) that reside on streams or lakes and calculate metrics for all lakes and networks within the LAGOS-US NETWORKS database (Figure 4). The NABD is a dataset of large, anthropogenic barriers that are spatially linked to the NHDPlusV1 data product to facilitate analyses based on the NHD and National Inventory of Dams (NID). Cooper et al. (2017) augmented this database with 170 additional dams from the USFWS Fish Passage Decision Support Tool. In addition, the dams were linked to the NHDPlusV2 and dams that were identified as having been removed since the NABD was published were excluded using the 2018 American Rivers dam removal database.

Citations:

Ostroff, A., Wiefelich, D., Cooper, A., Infante, D., and USGS Aquatic GAP Program. 2013. 2012 National Anthropogenic Barrier Dataset (NABD): U.S. Geological Survey - Aquatic GAP Program: Denver, CO.

Cooper, A.R., Infante, D.M., Daniel, W.M., Wehrly, K.E., Wang, L., Brenden, T.O. 2017. Assessment of dam effects for streams and fish assemblages of the conterminous USA. *Science of the Total Environment* doi:10.1016/j.scitotenv.2017.02.067

Rivers American. American Rivers Dam Removal Database. In: Database: Figshare. 2019. Available. <https://doi.org/10.6084/m9.figshare.5234068.v6>

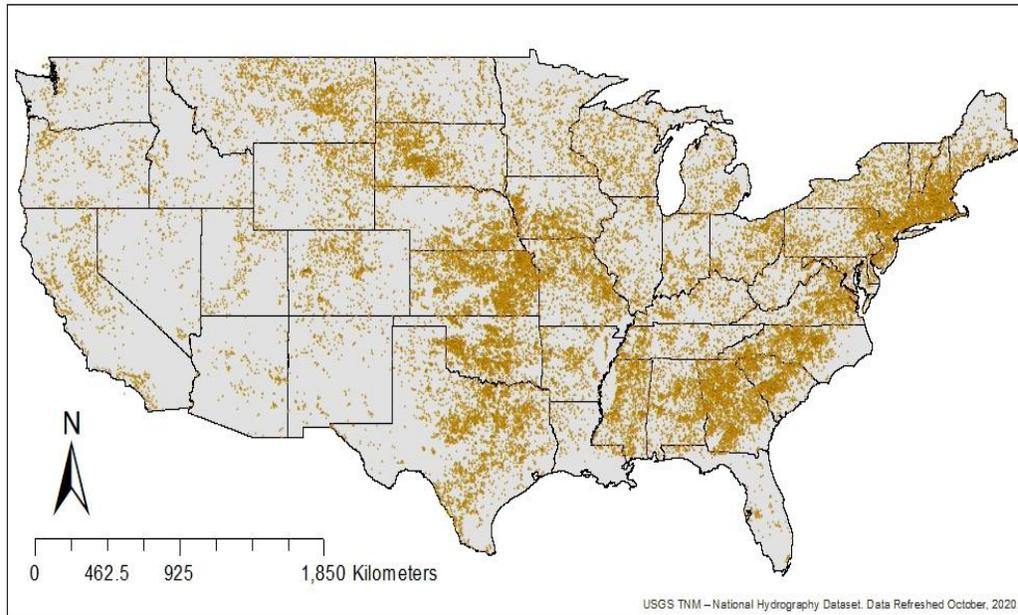


Figure 4. All dams included in the NETWORKS module (n=49,525). These dams are from the NABD, with modifications of additional dams from the USFWS Fish Passage Decision Support Tool and removal of dams that have been removed since the database was published using the 2018 American Rivers dam removal database.

2.2. Metadata tables

This section provides a description of variables in the two metadata tables contained in NETWORKS .

2.2.1. *data_dictionary_nets*

The *data_dictionary_nets* table can be used for the following purposes:

- Learn more information about the definition, units, and missingness of each variable. (table_name, variable_name, variable_description, units, missing_values)
- Understand the data provenance of a variable by identifying the code and data sources used to create it. (variable_source_code1, variable_source_code2, methods_tool)
- Explore variables by variable taxonomy groupings, or automate data analysis for all variables in a grouping. (taxonomy_*, variable_name_group)
- Specify the import of data into a new database or software program. (data_type, precision, domain, column_index)

2.2.2. *source_table_nets*

After joining *data_dictionary_nets* to *source_table_nets* on the variable_source_code1 and/or variable_source_code2 values, some additional uses are possible:

- Learn more information about the name and description of an input data source. (source_name, source_description, source_provider)
- Cite the input data sources for a variable. (source_citation)

- Locate data sources used to create NETWORKS for further analysis. (source_url)
- Answer questions about the purpose, contents, spatial resolution or timepoint of a source dataset. (source_metadata_url, source_spatial_resolution, source_access_date).

2.3. Data tables and variables

This section provides a description of variables contained in the four NETWORKS data tables. The tables all contain the key variable `lagoslakeid`, the unique lake identifier assigned by LAGOS-US LOCUS, that serves as the link across tables and modules in LAGOS-US. Variables in these data tables are characterized as four types: key, information, flags, or derived. Key variables consist of identifying information that is unique for a given lake. Information variables consist of identifying information for the dams. Observation-level flag variables indicate something about the observation that may be of interest to users (see Section 2.3.6). Finally, derived variables are quantitative or qualitative attributes of a lake or network that were generated through the methods described in section 3.

2.3.1. `nets_networkmetrics_medres` table

Definition: Table of variables describing lake connectivity to other lakes or dams via streams/ivers and network-scale metrics.

The variables in this table reflect freshwater connections between at least two lakes based on lake and stream (permanent and intermittent) connections in the NHDPlusV2. Therefore, these metrics do not include isolated lakes or lakes that are only connected to streams. Three types of metrics, and a total of eleven individual metrics, are included for each lake. The first type, connectivity between lakes, includes the distance (km) along the stream or river (i.e. stream course), the distance of a lake to the nearest lake, as well as the number of directly connected lakes both upstream and downstream. The second type, position within the network, includes lake network number and lake order (Riera et al. 2000, Martin and Soranno 2006). The third type, dam metrics, includes the distance to the nearest dam as well as the total number of dams both upstream and downstream, which were calculated for every lake. A set of four lake connectivity metrics are included at the network scale. These include the total numbers of dams and lakes in a network, the average lake area in a network, and the average distance between lakes in a network. Finally, this table includes observation-level flags for dams.

2.3.2. `nets_flow_medres` table

Definition: Table of stream and lake identifiers characterizing the downstream flow between surface water bodies.

This data characterizes the downstream flow between every flowline (i.e. streams/ivers and artificial flowlines that go through lakes). The FROM column and TO column contain waterbody identifiers denoting a direction of flow from one NHDPlusV2 flowline to the other. Our updated version of the NHDPlusV2 flow table, including where artificial flowlines are matched to lakes from the LAGOS-US database, can be found in this data table.

2.3.3 `nets_uninetworkdistance_medres` table

Definition: Table of stream course distance (in kilometers) between every pair of lakes, where stream traversal is in one direction (i.e., distances downstream).

This table includes all of the stream course distances between two lakes using a unidirectional graph. Because a unidirectional graph traverses the network only downstream, this table includes only a downstream distance. If there were multiple paths connecting the same two lakes, we include the path with the shortest length.

2.3.4. *nets_binetworkdistance_medres* table

Definition: Table of stream course distance (in kilometers) between every pair of lakes, regardless of direction (i.e., this distance includes the combination of upstream and downstream courses).

This table includes all of the stream course distances between two lakes using a bi-directional graph. A bi-directional graph traverses the network both up and downstream, therefore, this table includes many more observations than the *nets_uninetworkdistance_medres* table. This table includes a combination of the downstream distance *and* the upstream distance as well as the total distance between two lakes. If there were multiple paths connecting the same two lakes, we include the path with the shortest length.

2.4. Variable naming conventions

We created a naming convention for variables in the NETWORKS data module that is generally consistent with that used for all of LAGOS-US and that provides enough information for users to understand the meaning of variable names (CUAHSI ODM 2015). Some variables derive information from external data sources, as described in section 2.1.

2.4.1. Variables in *nets_networkmetrics_medres* tables

Variable names for this table were constructed following a set of rules that provide a consistent naming convention. Below, we describe those rules.

- a. Variable names are all lowercase.
- b. Parts of names are separated by an underscore ('_').
- c. Except for *lagoslakeid* and *nhdplusv2_comid*, names in each table must consist of at least one main feature. Depending on the variable, units may not be included in the name.
- d. Names consist of a maximum of three parts in the same order of i-iii below.
 - i. Spatial division, which is 'lake' or the 'net'-level observations.
 - ii. When appropriate, a shortened version of the module name, "nets", was included as a secondary identifier to avoid confusion with other similar metrics in the LAGOS data platform.
 - iii. Feature group is always present and reflects specifics about the data being provided. The main feature group can be a shorthand for the feature data source.
 - iv. The final part of the name is either units for a quantitative variable or 'flag,' which indicates the variable is an observation flag. Where a derived variable lacks units, the values are dimensionless, and thus the name includes no units. The unit abbreviations are standardized across LAGOS-US.

2.4.2. Variables in *nets_uninetworkdistance_medres* and *nets_binetworkdistance_medres*

Variable names for these tables were constructed following a set of rules that provide a consistent naming convention different from the ones described previously. Below, we describe those rules.

- a. Variable names are all lowercase.
- b. Parts of names are separated by an underscore ('_').
- c. Except for *lagoslakeid* names in each table must consist of at least one main feature.
- d. Names consist of a maximum of three parts in the same order of i-iii below.

- i. Feature group is always present and reflects specifics about the data being provided.
- ii. Direction from the lake the distance was calculated ('up', 'down', or 'total' for upstream, downstream, or total distance, respectively).
- iii. The final part of the name is units for a quantitative variable or if a derived variable lacks units, the values are dimensionless, and thus the name includes no units. The unit abbreviations are standardized across LAGOS-US.

2.4.3. Variables in *nets_flow_medres*

Variable names for this table were constructed following a set of rules that provide a consistent naming convention different from the ones described previously. Below, we describe those rules.

- a. Variable names are all lowercase.
- b. Parts of names are separated by an underscore ('_').
- c. The first part of the name is the flow direction (to or from).
- d. The final part of the name provides the identifier.

2.5. Data dictionary of variables

Below, we briefly summarize the variables in each data table within the NETWORKS module. Variables in these data tables are characterized by taxonomy as four types: key, information, flags, or derived. Key variables consist of identifying information that is unique for a given lake. Information variables consist of identifying information for the dams. Observation-level flag variables indicate something about the observation that may be of interest to users (see Section 2.3.6). Finally, derived variables are quantitative or qualitative attributes of a lake (or network) that were generated through the methods described in section 3.

Table 1. The NETWORKS module data dictionary for the *nets_networkmetrics_medres* table (n=86,511 lakes; file size ~11.1 MB).

variable_name	variable_description	taxonomy_type	missing_values	Units
lagoslakeid	Unique lake identifier developed by LAGOS-US.	key	N	NULL
nhdplusv2_comid	Unique lake identifier from the nhd for the medium resolution NHDplusV2.	key	N	NULL
lake_nets_upstreamlake_km	Distance to nearest upstream lake using a unidirectional graph.	derived	Y	kilometers
lake_nets_downstreamlake_km	Distance to nearest downstream lake using a unidirectional graph.	derived	Y	kilometers
lake_nets_bidirectionallake_km	Distance to the nearest lake upstream or downstream using a bi-directional graph.	derived	N	kilometers
lake_nets_upstreamlake_n	The number of upstream lakes directly connected through streams to a lake.	derived	N	number
lake_nets_downstreamlake_n	The number of downstream lakes directly connected through streams to a lake.	derived	N	number
lake_nets_lakeorder	Lake order follows the Strahler stream order of the stream that flows from it (outflowing), where the higher order stream is chosen if more than one outlet occurs (Riera et al. 2000, Martin and Soranno 2006). The exceptions are that headwater lakes are 0 and terminal lakes receive the order of the highest inflowing stream.	derived	Y	NULL
lake_nets_lnn	Lake network number (LNN) is the position of a lake	derived	N	NULL

	within the network in reference to other lakes. The lake at the top of a network (i.e. no upstream lakes) will be 1, the next lake downstream will be 2, etc. If a lake has more than one lake upstream it will take the higher LNN.			
lake_nets_nearestdamup_id	The NABD dam ID for the nearest upstream dam.	information	Y	NULL
lake_nets_nearestdamup_km	Distance to nearest upstream dam.	derived	Y	kilometers
lake_nets_nearestdamdown_id	The NABD dam ID for the nearest downstream dam.	information	Y	NULL
lake_nets_nearestdamdown_km	Distance to nearest downstream dam.	derived	Y	kilometers
lake_nets_totaldamup_n	The total number of upstream dams.	derived	N	number
lake_nets_totaldamdown_n	The total number of downstream dams.	derived	N	number
net_id	The unique identifier assigned by LAGOS-NETWORKS for each network.	derived	N	NULL
net_lakes_n	The total number of lakes in the lake network.	derived	N	number
net_averagelakedistance_km	Average distance between lakes in a network.	derived	N	kilometers
net_averagelakearea_ha	Average lake area in a network.	derived	N	hectares
net_dams_n	The total number of dams on a network.	derived	N	number
lake_nets_damonlake_flag	A value of 'Y' indicates that there is at least one dam on this lake. This means that the dam point falls onto one of the artificial flowlines that flows through a lake and is therefore associated with the lake and not a stream reach. An "N" indicates no flag.	flag	Y	NULL
lake_nets_multidam_flag	A value of 'Y' indicates that there are multiple dams on a lake. An "N" indicates no flag.	flag	Y	NULL

Table 2. The NETWORKS module data dictionary for the *nets_uninetworkdistance_medres* table (n = 124,251 lake pairs; file size ~2.6 MB).

variable_name	variable_description	taxonomy_type	missing_values	Units
laogslakeid	Identifier of lake 1 (lagoslakeid) that is connected to lake 2 using a unidirectional graph.	key	N	null
to_lagoslakeid	Identifier of lake 2 (lagoslakeid) connected to lake 1 using a unidirectional graph.	key	N	null
streamlength_down_km	Distance downstream from lake 1 to lake 2 (as indicated by lagoslakeid) using a unidirectional graph.	derived	Y	kilometers

Table 3. The NETWORKS module data dictionary for the *nets_binetworkdistance_medres* table (n = 39,498,506 lake pairs; file size ~1.9 GB).

variable_name	variable_description	taxonomy_type	missing_values	Units
lagoslakeid	Identifier of lake 1 (lagoslakeid) that is connected to lake 2 using a bidirectional graph.	key	N	null
to_lagoslakeid	Identifier of lake 2 (lagoslakeid) connected to lake 1 using a bidirectional graph.	key	N	null
streamlength_total_km	Total stream distance from lake 1 to lake 2 (as indicated by lagoslakeid) using a bidirectional graph.	derived	N	kilometers
streamlength_up_km	Distance upstream from lake 1 to lake 2 (as indicated by lagoslakeid) using a bidirectional graph.	derived	Y	kilometers
streamlength_down_km	Distance downstream from lake 1 to lake 2 (as indicated by lagoslakeid) using a bidirectional graph.	derived	Y	kilometers

Table 4. The NETWORKS module data dictionary for the *nets_flow_medres* table (n=2,722,347 rows of flowlines, streams and artificial flowlines through lakes; file size ~ 76.2 MB). Note: these variables are unitless.

variable_name	variable_description	taxonomy_type	missing_values
nets_from_comid	Common identifier of the upstream NHDFlowline feature.	key	N
nets_to_comid	Common identifier of the downstream NHDFlowline feature.	key	N
nets_from_lagoslakeid	Identifier of the upstream lake as indicated by lagoslakeid.	key	Y
nets_to_lagoslakeid	Identifier of the downstream lake as indicated by lagoslakeid.	key	Y

2.6. Data flags

During construction of the NETWORKS module, we created a series of data flags for the *nets_networkmetrics_medres* table that convey something about a data observation that may be of interest to users. These flags are *informational* flags of general relevance to the data user; there are *no* cautionary flags that indicate potential concerns for inclusion of particular data observations in analysis (Table 5).

2.6.1. Informational flags

Table 5. Description and occurrence of informational data flags in *net_networkmetrics_medres* (number of lakes connected to other lakes = 86,511).

Flag	Value	Description	User Relevance	Number of occurrences	Percent of data
lake_nets_damonlake_flag	Y	A value of ‘Y’ indicates that there is at least one dam on this lake. This means that the dam point falls onto one of the artificial flowlines that flows through a lake and is therefore associated with the lake and not a stream reach. An “N” indicates no flag.	This flag primarily serves to alert the user of the presence of a dam directly on a lake as opposed to on a connecting stream reach.	12,630	14.6
lake_nets_multidam_flag	Y	A value of ‘Y’ indicates that there are multiple dams on a lake. An “N” indicates no flag.	This flag identifies lakes that have multiple dams. There may be a dam at multiple inlets or outlets or a dam at both locations.	132	0.15

3. Methods for LAGOS-US NETWORKS

This section outlines the methods used for creating the NETWORKS module. We explain how we derived lake networks and their associated connectivity metrics and how NABD dam data was linked to our networks. For further technical detail on this process, or to reproduce this effort, users can consult the published scripts (Wang & King 2020).

3.1. Software used for NETWORKS creation

We used a combination of Python 2.7.8 (Van Rossum & Drake 1998), ArcGIS 10.3 Desktop (ESRI 2014), and R version 3.6 to create LAGOS-US NETWORKS. The majority of the methods are associated with Python scripts which can be found at <http://doi.org/10.5281/zenodo.4383172>. Those scripts can be used to understand, reproduce, or adapt our methods. ArcGIS was used for some of the dam classifications, mapping, and verification of some metrics. We used the “nhdR” (Stachelek 2019) package to download NHDplusV2 data and the “hydrolinks” package (Winslow et al. 2018) to verify metrics in R.

3.2. Methods for creating the lake networks

Lake networks across the conterminous U.S. were created using the flow table from the NHDplusV2 database (USGS 2019). This flow table consists of every flowline (streams and artificial flowlines that go through lakes; Figure 5) either in the FROM column or TO column, denoting a direction of flow from one line to the other, as well as the distance for each connection between two flow lines. Prior to creating a graph, we removed coastline connections (Fcode 56600; McKay et al. 2012) so that the

connectivity networks would not connect through the ocean, estuaries, or the Great Lakes, as well as IDs associated with the Great Lakes water bodies. Artificial flowlines (Figure 5) were linked to water bodies (*nhdplusv2_comid*) and these water bodies were linked to *lagoslakeids* using the *lake_link* table from the LAGOS_US_LOCUS module (Smith et al. 2020). Our modified version of the NHDPlusV2 flow table including where artificial flowlines are matched to lakes from the LAGOS-US database can be found as the *nets_flow_medres* data table.

We applied a graph theory framework to create lake networks from this flow table. Graphs are mathematical structures used to model pairwise relations between objects, or nodes. In our case, we are interested in modeling the pairs of lakes that are connected by streams. Two types of graphs can be used to model connections: unidirectional graphs consider *either* downstream or upstream connections and bidirectional graphs consider *both* downstream and upstream connections. We created lake networks using bidirectional graphs with both lakes and streams as nodes (Figure 6). We used Dijkstra's algorithm (Cormen et al. 2001) to traverse the graph both up and downstream starting at a given lake. During the traversal, if a node was a stream, we continued traversing the graph until the node was a lake. We saved the distance from the given lake to this lake and stopped traversing. If there were multiple paths to connect the same two lakes, the algorithm chose and saved the path with the shortest length. This process outputs all the connections of the given lake to its neighbor lakes. This process was repeated for every lake until the connections and stream course distances between all lakes were known.

All lakes that are connected to another lake, up or downstream, are considered part of one network. We assigned each of these networks a unique identification number (*net_id*). All of the stream course distances between pairs of lakes can be found in the *nets_binetworkdistance_medres*. The artificial flowline distances through lakes were not included in these distances. This table includes upstream, downstream, and total distance between two lakes. The total distance may be smaller than the sum of the upstream and downstream columns because the graph does not have information on where the stream reaches intersect each other, therefore, an intersecting stream reach is only counted once for the total distance, but may be included in both the downstream and upstream distance columns (Figure 7).



Figure 5. An example of “artificial flowlines” in purple that show the connection of streams (teal) to lakes (blue) and are used to denote flow direction.

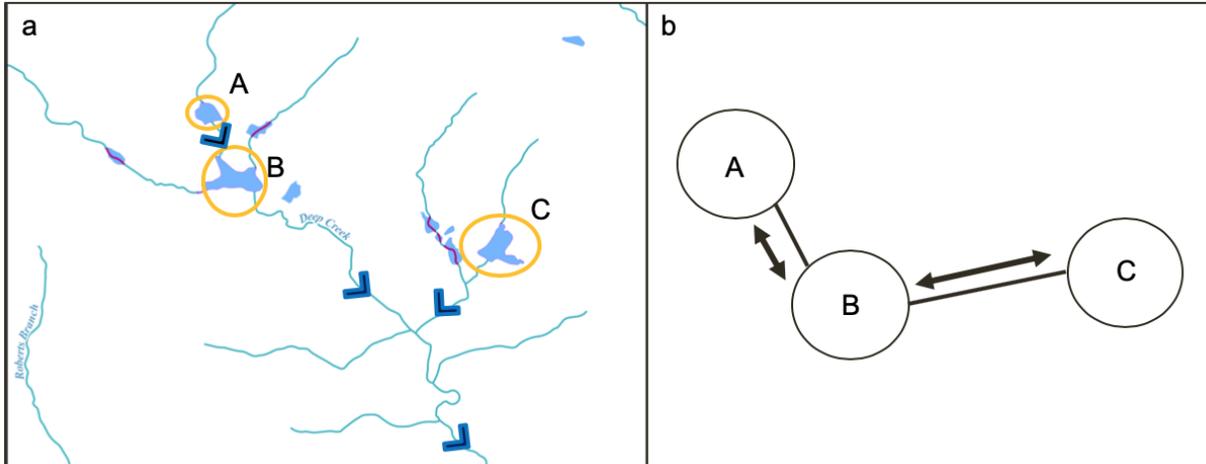


Figure 6. An example of (a) connected pairs of lakes and (b) a bidirectional graph depicting those same connections, to illustrate how networks were created and how upstream or downstream distances were defined. The distance between lake B and lake C includes traversing the network downstream and upstream.

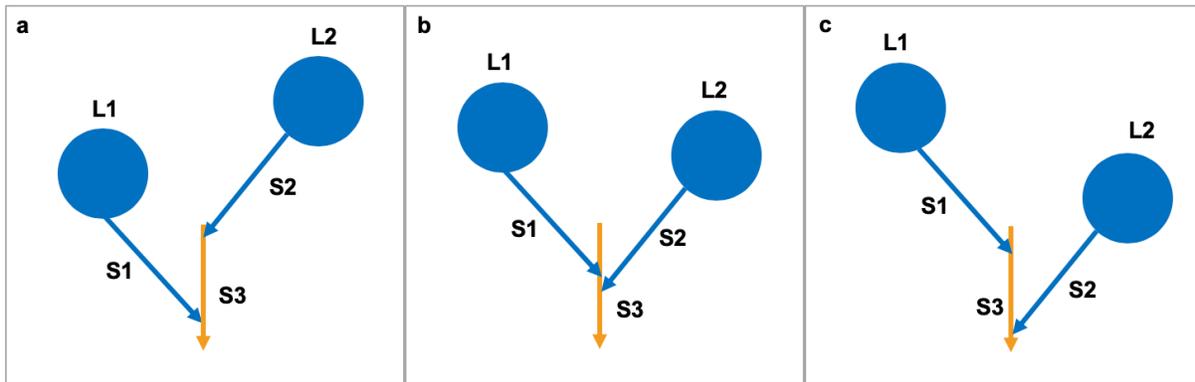


Figure 7. Different scenarios of connections between lake 1 (L1) and lake 2 (L2) using a bidirectional graph, where stream reach S1 and stream reach S2 might be connected to stream S3 at different points along the reach. In all three scenarios (a,b,c), S1 and S3 are included in downstream distance, S3 and S2 are included in the upstream distance, and S1, S2, and S3 are included in the total distance.

3.3. Methods for linking LAGOS-US NETWORKS with NABD dams

The NABD is a dataset of large, anthropogenic barriers that are spatially linked to the NHDPlusV1 data product to facilitate analyses based on the NHD and National Inventory of Dams (NID) (Ostroff et al. 2013). Cooper et al. (2017) augmented this database with 170 additional dams from the USFWS Fish Passage Decision Support Tool and excluded ~250 dams that were identified as having been removed since the NABD was published (Rivers 2019). The dams were linked to the NHDPlusV2 flowlines and were incorporated into networks. Dams were assigned to a *lagoslakeid* if they were less than 50 m from a lake (Polus et al. *in review*). Dams that fall directly on a lake could not be considered as up- or downstream because they were on the node and therefore, did not have a direction in reference to that node. Therefore, these dams were assigned as upstream or downstream from a lake using two methods:

1) Using ArcGIS, lake inlets and outlets were identified using the start and end vertices associated with the artificial flowlines and extracted as points representing inlets and outlets. When multiple artificial

paths were present, the uppermost artificial flowline was identified for inlet locations and the downstream-most artificial flowline was identified for outlet locations. For each dam point location, the nearest three inlets or outlets (combined) were identified using euclidean distance in the ArcGIS GenerateNear tool. If the nearest inlet was less than 250m away, and no outlets or other lakes were also nearby, the dam was automatically designated as upstream of the associated lake. An equivalent, symmetrical rule was applied for nearby outlets. If both inlets and outlets for the same lake were very near each other or an inlet or outlet for another lake was very near, the dam position was assigned for manual review. "Very near" was defined as follows: if the second closest junction is within 50m of the closest junction or if the second closest junction is within 100m of the closest one *and* the closest junction is within 25m of the dam. Methods are available as Python code within the LAGOS GIS Toolbox (http://github.com/cont-limno/LAGOS_GIS_Toolbox; national_outlets_inlets.py, dams_link_lake_junctions.py). There were 11,551 dams that were assigned upstream or downstream of a lake using this method.

2) The remaining dams (n=1,079) that could not be identified by the automated process described above in (1) were then manually classified by visual inspection of the dam location in comparison to the NHD polygons and flowlines and manually assigning them as either on the upstream or downstream side of a lake.

Two data flags were created during the process of linking dams to lakes and streams/rivers. These flags are for cases when dams fall onto an artificial flowline contained within a lake or when multiple dams fall on the same lake (Figure 8; Table 5).

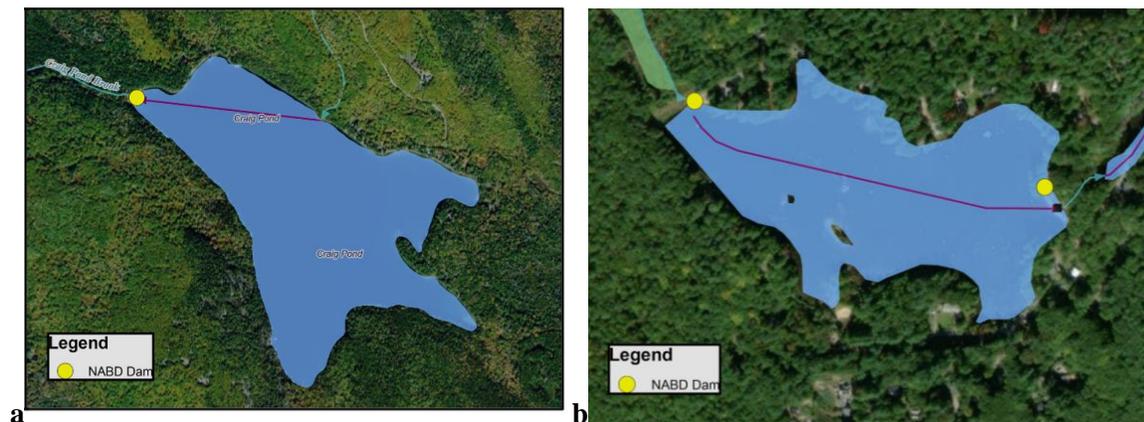


Figure 8. Examples of informational data flags, where (a) is a dam on a lake (`lake_nets_damonlake_flag`) and (b) is multiple dams on a lake (`lake_nets_multidam_flag`)

3.4. Methods for connectivity metrics

After creating the connectivity networks, several metrics were created at the lake scale using a unidirectional graph. Unidirectional graphs consider only *downstream* connections. For example, in Figure 9 there is a downstream distance between lake A and lake B that is the same distance upstream from lake B to lake A. The connection between lake B and lake C is not included because the unidirectional graph does not traverse both down and upstream. We used Dijkstra's algorithm (Cormen et al. 2001) to traverse the graph downstream only starting at a given lake. During the traversal, if a node was a stream, we continued traversing the graph until the node was a lake. We saved the distance from the

given lake to this lake and stopped traversing. If there were multiple paths to connect the same two lakes, the algorithm chose and saved the path with the shortest length. This outputs all the connections of the given lake to its neighbor lakes. This process was repeated for every lake until the connections and stream course distances between all lakes were known. These stream course distances between two lakes using a unidirectional graph can be found in the [nets_uninetworkdistance_medres](#) table.

The nearest lake distance was determined by comparing the distance between each lake and all of its neighboring lakes and choosing the nearest distance upstream (Figure 10a) and the nearest distance downstream (Figure 10b) from the unidirectional graph. Note that not all lakes have both an upstream and downstream lake. The number of directly connected lakes upstream was computed as the indegree of a lake, i.e. the number of lakes upstream only connected through streams flowing into the lake. Similarly, the number of directly connected downstream lakes was calculated using the outdegree of a lake, i.e. lakes directly connected through streams flowing out of a lake. There are instances when a lake does not have any directly connected upstream or directly connected downstream lakes because the lake is only connected through the bidirectional graph to the lake network (e.g. Figure 9, lake C; n=7,617). Therefore, we also included the nearest bidirectional distance (Figure 10c). This distance is often the same as the nearest downstream or nearest upstream value, however, it can be different if the nearest lake is connected through a bidirectional graph (Figure 10c).

Two metrics that describe the position of a lake within the network and landscape were derived from the unidirectional graph: Lake Network Number (LNN; Figure 10d) and Lake Order (LO; Figure 10e) (Riera et al. 2000; Martin and Soranno 2006). LNN was computed by starting at the first lake in a network (e.g. no upstream lakes) and assigning that lake a “1”, then moving downstream in the network to another lake and assigning that lake a “2”, and so on. Therefore, multiple lakes in a network could be assigned a “1” if they did not have any upstream lakes. Lakes with multiple upstream lakes were assigned the larger sequential number (Martin and Soranno 2006). LO was assigned using the Strahler stream order from the NHDplusV2 attributes. LO follows the Strahler stream order of the outflowing stream, where the higher order stream is chosen if more than one outlet is present (Riera et al. 2000, Martin and Soranno 2006). There were two exceptions to this: headwater lakes were assigned a “0” and terminal lakes received the order of the inflowing stream (Riera et al. 2000; Martin and Soranno 2006). To differentiate between headwater lakes and lakes that had inflowing streams but not upstream lakes, we considered inflowing streams for LO calculation. There were instances when a loop between two lakes occurred (0.02% of all connections), for example lake A flows to lake B and lake B flows back to lake A. In these instances, we randomly removed one connection.

Several dam metrics were derived to characterize connectivity barriers. The Depth First Search (DFS; Cormen et al. 2001) algorithm was used to traverse the lake-stream network to find all the upstream dams and downstream dams. The DFS algorithm is a common computer science technique that is used for traversing graphs by starting at one node and exploring every branch of the graph. Dijkstra’s algorithm was used to compute the distance to the nearest upstream and downstream dams (Cormen 2001). Because we used a graph to create the network, the algorithm did not have the exact location of the dam on the stream reach, just the flowline it is located on. Therefore, when deriving the metrics for the nearest dam, the entire stream reach that the dam is located on was included in the distance calculation. Thus, there were instances when two or more dams fell on the same stream flowline (8.7 % occurrence). In these instances, all dams were considered as the nearest up- or downstream dams, they have the same distance from the lake, and all of the dam identifiers were included and separated by a semicolon. Similarly, if multiple dams were on a lake, all the dams were considered the nearest dam, all dam identifiers were included, and dams located on a lake were assigned the distance of 0 km.

The completed lake networks were traversed using the DFS algorithm that counts total on-network lakes, the average distances between lakes in a network, and the total number of dams in each lake network. The average area of the on-network lakes in was calculated using the area from LAGOS-US LOCUS v1 polygons (Smith et al. 2020), grouping lakes by networks, and then using the Calculate Geometry tool in ArcGIS.

Lake networks were created for NETWORKS based on the medium resolution NHDplusV2 flow data; therefore, connectivity may differ from connectivity metrics in LAGOS-US LOCUS that were created based on the NHD high resolution (Smith et al., 2020). Metrics were only included for lakes connected to other lakes, and therefore do not include isolated lakes or lakes that are only connected to streams.

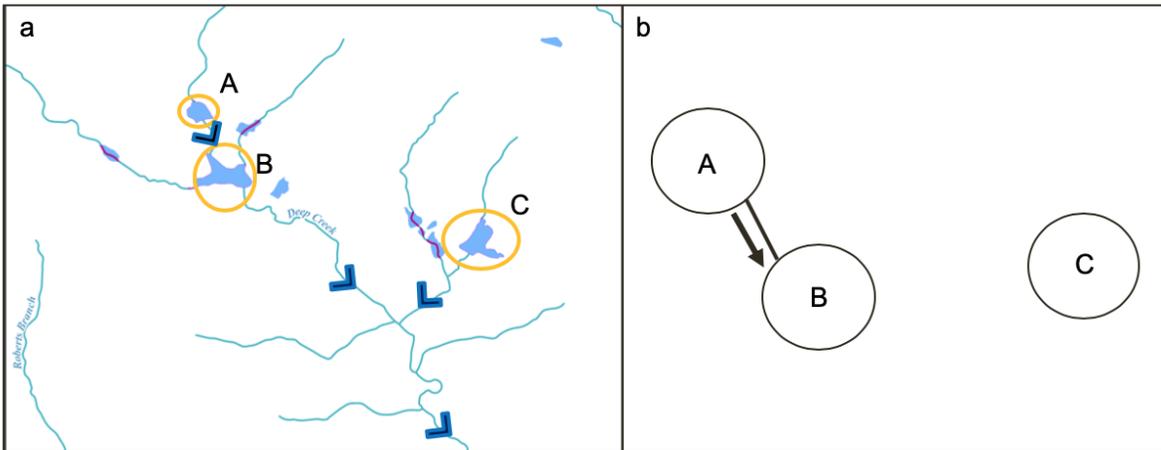


Figure 9. An example of (a) connected pairs of lakes compared to (b) a unidirectional graph to illustrate how upstream or downstream distances were defined for some connectivity metrics and for the unidirectional table.

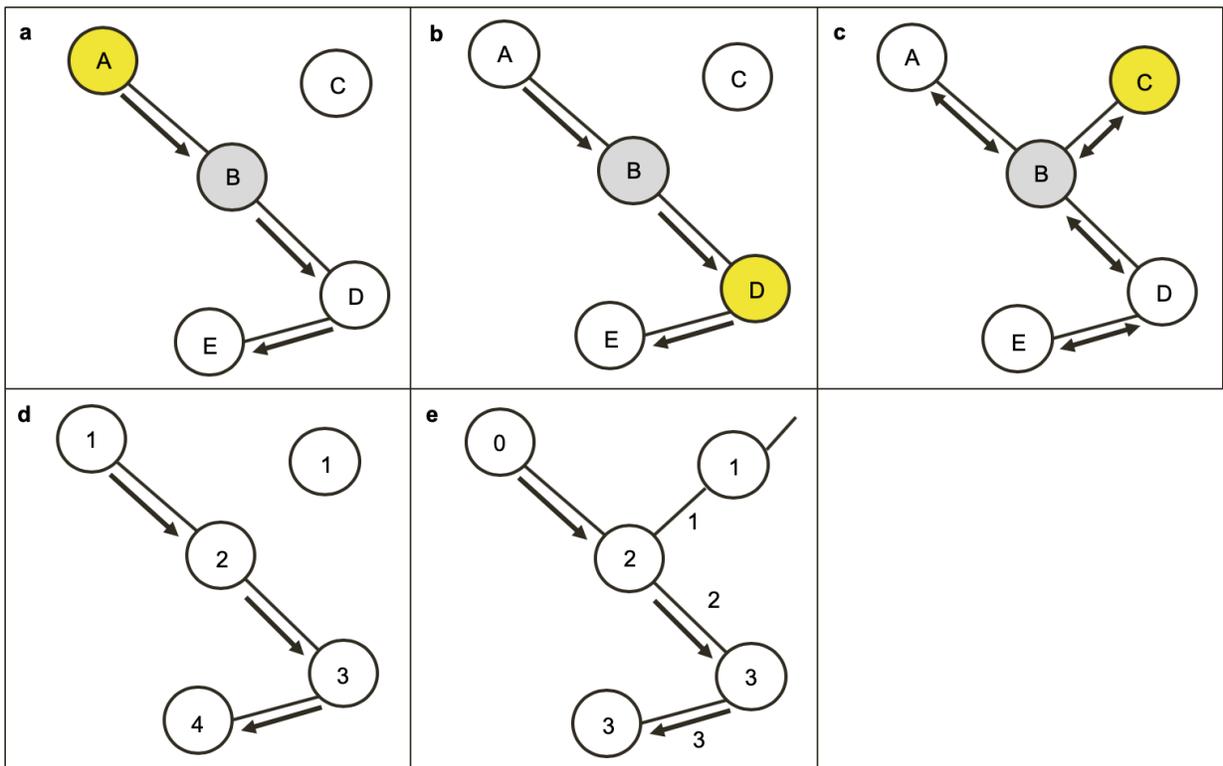


Figure 10. Examples of some of the connectivity metrics in NETWORKS. (a) The nearest upstream lake (A; yellow) to lake B using a unidirectional graph, (b) the nearest downstream lake (D; yellow) to lake B using a unidirectional graph, (c) the nearest bidirectional lake (C; yellow) to lake B using a bidirectional graph. Both (d) Lake Network Number and (e) Lake Order use a unidirectional graph and indicate network position. In panel (d) and (e), numbers within lakes (circles) are the LNN or LO, respectively. In panel (e), numbers beside lines are the Strahler stream order assigned to stream reaches, which are used to assign LO to a lake.

4. Quality Control/Quality Assurance (QA/QC) for LAGOS-US: NETWORKS

The validation and Quality Control/Quality Assurance (QAQC) process is intended to ensure that the procedures used to create the values for NETWORKS variables resulted in the intended outcomes. We used two methods for validation and QAQC.

First, during the initial creation of the metrics, a simulation graph was created to validate the code. This simulation graph included paths that were unidirectional as well as bidirectional, multiple connections between lakes, lakes that were directly connected to other lakes without streams, and a simulated Great Lake. From this simulation graph, we checked that the distance between pairs of lakes was correct for downstream, upstream, and bi-directional connections. Then, we ensured that the code selected the shorter distance if there were multiple connections between lakes for both the unidirectional and bidirectional connections. For lakes that do not have a stream connection between them, we ensured the code output for the downstream and upstream distance was 0 km. Finally, we tested that the code ignored connections to the Great Lakes. The NETWORKS authors manually examined resulting networks and associated metrics using either GIS or the “hydrolinks” package in R (Winslow et al. 2018), which downloads and traverses paths for the NHD medium resolution data to identify potential issues with either the input data or code. All solvable issues were reconciled and the networks or metrics were regenerated and retested until no further issues were found.

Second, after metrics were quantified, the NETWORKS metrics data table ([nets_networkmetrics_medres](#)) was queried to: 1) identify potential data or geoprocessing issues and 2) verify that data values were sensible (e.g., are within expected ranges and expected completeness of data). These checks of individual feature values assessed that the workflow generating data accurately reflects both the source data and the lake-specific values.

For this process, the [nets_networkmetrics_medres](#) data table, in csv (comma-separated values) format, was imported by semi-automated R scripts into a companion R markdown script that summarizes the data table, ensures comparability with the source GIS layer and data dictionary, summarizes and maps values for each variable, and automatically generates scores for three main evaluation criteria. A QAQC summary report is provided in html format. The third-party R packages we used in this process include tidyverse, sf, summarytools, and Rmarkdown.

The R scripts generating the QAQC report automatically evaluate five core QAQC checks as either **Pass**, **Warn**, **Fail**, or **N/A** (e.g., not applicable). A **Warn** score indicates that although the data have failed to fully meet a criterion, this failure was due to a known issue that is not solvable (most commonly this occurs where the data layer extent is smaller than the zonal extent generating missing data values). This serves as an indication to users that although the data are considered usable, certain applications may not be appropriate or may need adjustments.

The following describes the three evaluation criteria applicable to the NETWORKS module and the actions followed to rectify potential problems. Note that actions are iterative; the QAQC review feeds back into the data creation process, which then re-exports the data table and then re-runs the entire QAQC process.

- 1) **Match with GIS data:** This check compares the list of [lagoslakeid](#) in the data table with those in the corresponding LAGOS-US LOCUS v1 reference shapefiles maintained in an ArcGIS

geodatabase (GIS_LOCUS; Smith et al. 2020). A list of non-matching lagoslakeids either in the feature table or the geodatabase is provided in the QAQC report and a **Fail** warning is generated.

Action: where a **Fail** warning is generated, non-matching lagoslakeids are manually investigated to identify the source of the mismatch between the datatable and the reference GIS datalayer.

- 2) **Match with metadata:** Variable names in the datatable are compared with the master list of variable names maintained in the metadata / data dictionary. Where there is no match, due to missing or incorrect names in either the data dictionary or the datatable, a **Fail** warning is generated and the mismatches are listed in a table in the QAQC report.

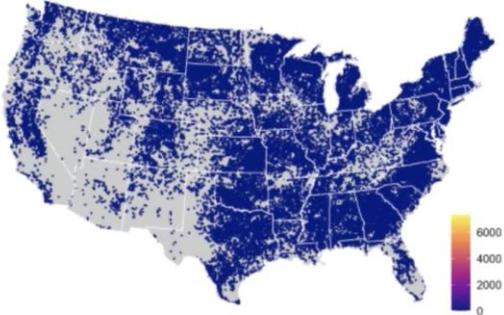
Action: where a **Fail** warning is generated, the data dictionary and datatable variable names are examined and the name(s) in error are fixed as necessary either in the data dictionary or in the python stream which exports the data from GIS.

- 3) **Missing value check:** This check counts the number of zones with missing values, lists them, and produces maps of their location. A **Warn** evaluation is created for this criterion as, like the zonal completeness check, it is not unexpected to have zones with missing values.

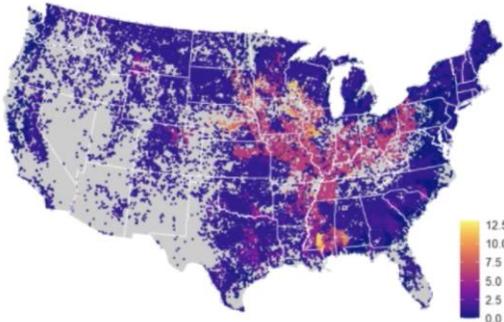
Action: A **Warn** value for this criteria is not unexpected for zones near borders or coasts that extend beyond the extent of the input data layer. If zones with missing values are not located at the edges of feature data layers, they are inspected to make sure there are no gaps in the input data or some other geoprocessing issue.

In addition to responding to a **Fail** or **Warn** check from the above evaluation criteria, we manually examined other output in the QAQC report to identify potential issues with either the input data or geoprocessing. The reviewer then discussed any discrepancies or questions with the QAQC report creator and all solvable issues were reconciled and the data files were regenerated and retested until no further issues were found. Below we show select variable outputs from the QAQC report (Figure 11, 12, and 13).

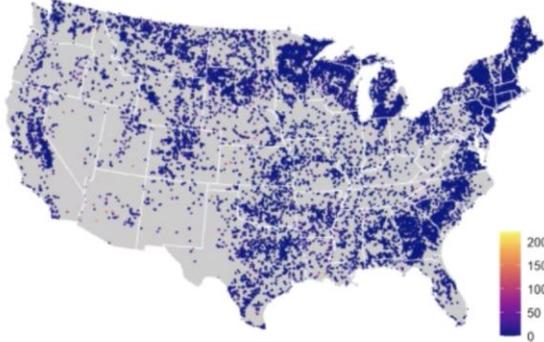
lake_net_upstreamlake_n



lake_net_downstreamlake_n



lake_net_upstreamlake_km



lake_net_downstreamlake_km

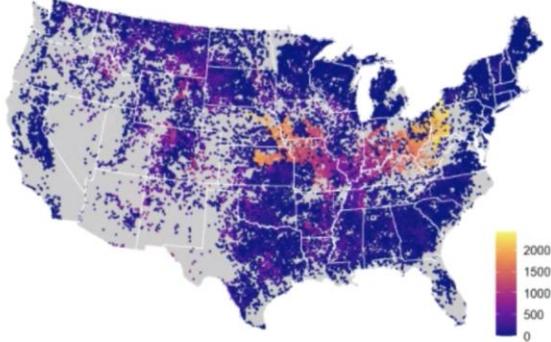
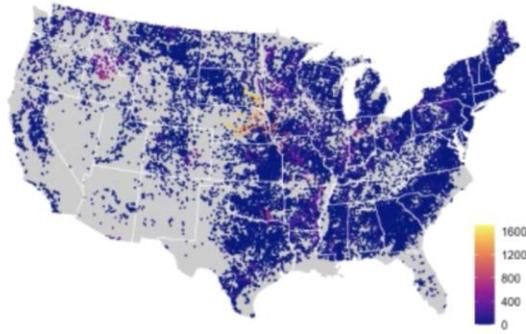
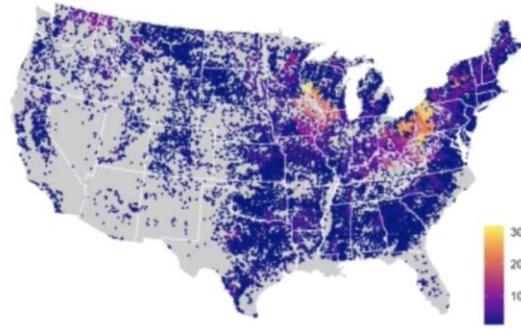


Figure 11. Spatial distribution of upstream and downstream observations in NETWORKS.

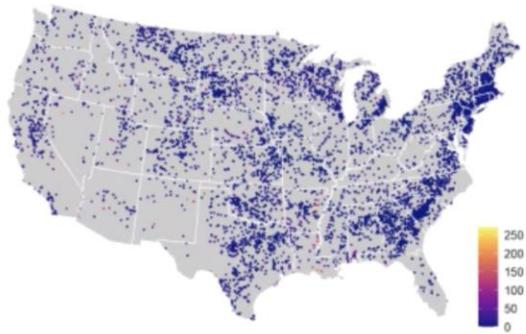
lake_net_nearestdamdown_km



lake_net_totaldowndam_n



lake_net_nearestdamup_km

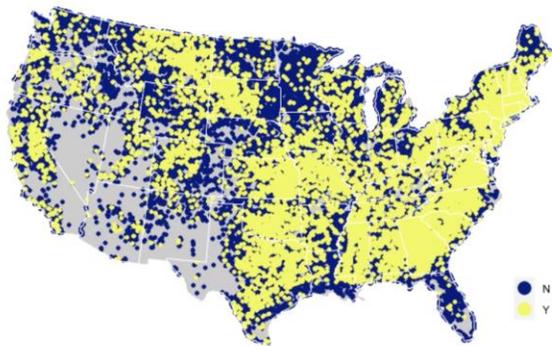


lake_net_totaldamup_n



Figure 12. Spatial distribution of dam observations in NETWORKS.

lake_net_damonlake_flag



lake_net_multidam_flag



Figure 13. Spatial distribution of data flag occurrences in NETWORKS.

5. Accessing, Using, and Citing LAGOS-US NETWORKS

5.1 Data access and use

LAGOS-US NETWORKS v1 is made up of multiple files of various types: data and metadata tables that are csv files, as well as this documentation guide all of which are available for public download via the EDI repository. However, for ease of use, NETWORKS data can be downloaded through the LAGOS-US R package and linked to other LAGOS-US core and extension modules.

LAGOS-US R package: Stachelek J. 2020. LAGOSUS: Interface to the Lake Multi-scaled Geospatial and Temporal Database. R package version 0.0.1.

5.2 Data citation

When data from the NETWORKS module are included in analyses, users should cite them. We have also written a data paper that describes the motivation and context for creating the NETWORKS module, which may be cited by users.

- 1) **NETWORKS v1 data and documentation:** King, K., Wang, Q., Rodriguez, L.K., Haite, M., Danila, L., Pang-Ning, T., Zhou, J., and Cheruvellil, K.S. 2020. LAGOS-US NETWORKS v1: Data module of surface water networks characterizing connections among lakes, streams, and rivers in the conterminous U.S. Environmental Data Initiative. <https://doi.org/XXXXXX>. Dataset accessed XX/XX/XXXX.
- 2) **Data paper:** King, K., Wang, Q., Rodriguez, L.K., and Cheruvellil, K.S. *under review*. Lake networks for the conterminous U.S. (LAGOS-US NETWORKS). *Limnology and Oceanography Letters*. Submitted for peer review December 2020.

5.3 Reproducible code

There is code available to reproduce, extend, or adapt our networks.

Wang, Q. & King, K. (2020). Code for LAGOS-US NETWORKS v1.0 (Version v1.0.0) *Zenodo*. <http://doi.org/10.5281/zenodo.4383172>

6. References

- Consortium of Universities for the Advancement of Hydrologic Science, Inc. CUAHSI ODM 2015. <https://www.cuahsi.org/ODMControlledVocabulary>. Accessed 2 December 2011.
- Cooper, A.R., Infante, D.M., Daniel, W.M., Wehrly, K.E., Wang, L., Brenden, T.O. 2017. Assessment of dam effects for streams and fish assemblages of the conterminous USA. *Science of the Total Environment* doi:10.1016/j.scitotenv.2017.02.067
- Cormen, T.H., Leiserson, C.E., Rivest, R.L., and Stein, C. 2001. *Introduction to Algorithms*, Second Edition. MIT Press and McGraw-Hill. ISBN 0-262-03293-7. Section 22.3: Depth-first search and Section 24.3: Dijkstra's algorithm. pp. 540–549, 595-601.
- ESRI. 2014. ArcGIS Desktop: Release 10.3. Redlands, CA: Environmental Systems Research Institute.
- Martin, S. L., & Soranno, P. A. 2006. Lake landscape position: Relationships to hydrological connectivity and landscape features. *Limnology & Oceanography*, 51(2), 801–814.
- McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., and Rea, A. 2012. NHDPlus Version 2: User Guide. U.S. Geological Survey.
- Ostroff, A., Wieferich, D., Cooper, A., Infante, D., and USGS Aquatic GAP Program. 2013. 2012 National Anthropogenic Barrier Dataset (NABD): U.S. Geological Survey - Aquatic GAP Program: Denver, CO.
- Polus, S.M., L.K. Rodriguez, Q. Wang, J. Díaz Vázquez, K.E. Webster, P.-N. Tan, J. Zhou, L. Danila, P.A. Soranno and K.S. Cheruvilil. LAGOS-US RESERVOIR: Data module classifying conterminous U.S. lakes 4 hectares and larger as natural lakes or reservoirs. Environmental Data Initiative. *in review*.
- Riera, J. L., Magnuson, J. J., Kratz, T. K., & Webster, K. E. 2000. A geomorphic template for the analysis of lake districts applied to the Northern Highland Lake District, Wisconsin, U.S.A. *Freshwater Biology*, 43(3), 301–318. <https://doi.org/10.1046/j.1365-2427.2000.00567.x>
- Rivers American. American Rivers Dam Removal Database. In: Database: Figshare. 2019. Available <https://doi.org/10.6084/m9.figshare.5234068.v6>
- R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Smith, N.J., K.E. Webster, L. Rodriguez, K.S. Cheruvilil, and P.A. Soranno. 2020. LAGOS-US LOCUS v1: Module of location, identifiers, and physical characteristics of lakes and their watersheds in the conterminous U.S. Environmental Data Initiative. Accessed May 13, 2019.
- Stachelek, J. 2019. Tools for working with the National Hydrography Dataset. Version 0.5.3. <https://cran.r-project.org/package=nhdR>.
- Stachelek J. 2020. LAGOSUS: Interface to the Lake Multi-scaled Geospatial and Temporal Database. R package version 0.0.1.

- U.S. Geological Survey. 2019. National Hydrography Dataset (NHDPlus Version 2.1 for Hydrologic Unit (HU) 4 - 2001 (published 20191002)).
<https://www.epa.gov/waterdata/nhdplus-national-data>. Accessed August 5, 2019.
- Van Rossum, G. & Drake Jr, F.L., 1995. *Python reference manual*, Centrum voor Wiskunde en Informatica Amsterdam.
- Wang, Q. & King, K. 2020. Code for LAGOS-US NETWORKS v1.0 (Version v1.0.0) *Zenodo*.
<http://doi.org/10.5281/zenodo.4383172>
- Winslow, LA, TH Hahn, SD Princiotta, TH Leach, KC Rose. 2018. hydrolinks: A new tool to link macroscale to inland water bodies. Version 0.10.0. <https://cran.r-project.org/web/packages/hydrolinks/index.html>