JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION
Vol. 52, No. 1 AMERICAN WATER RESOURCES ASSOCIATION February 2016

THE STREAM-CATCHMENT (STREAMCAT) DATASET: A DATABASE OF WATERSHED METRICS FOR THE CONTERMINOUS UNITED STATES[1]

Ryan A. Hill, Marc H. Weber, Scott G. Leibowitz, Anthony R. Olsen, and Darren J. Thornbrugh[2]

ABSTRACT: We developed an extensive database of landscape metrics for ~2.65 million stream segments, and their associated catchments, within the conterminous United States (U.S.): The Stream-Catchment (StreamCat) Dataset. These data are publically available (http://www2.epa.gov/national-aquatic-resource-surveys/streamcat) and greatly reduce the specialized geospatial expertise needed by researchers and managers to acquire landscape information for both catchments (i.e., the nearby landscape flowing directly into streams) and full upstream watersheds of specific stream reaches. When combined with an existing geospatial framework of the Nation's rivers and streams (National Hydrography Dataset Plus Version 2), the distribution of catchment and watershed characteristics can be visualized for the conterminous U.S. In this article, we document the development and main features of this dataset, including the suite of landscape features that were used to develop the data, scripts and algorithms used to accumulate and produce watershed summaries of landscape features, and the quality assurance procedures used to ensure data consistency. The StreamCat Dataset provides an important tool for stream researchers and managers to understand and characterize the Nation's rivers and streams.

(KEY TERMS: streams; catchments; watersheds; watershed metrics; National Hydrography Dataset Plus; database; conterminous United States.)

Hill, Ryan A., Marc H. Weber, Scott G. Leibowitz, Anthony R. Olsen, and Darren J. Thornbrugh, 2016. The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States. *Journal of the American Water Resources Association* (JAWRA) 52(1): 120-128. DOI: 10.1111/1752-1688.12372

INTRODUCTION

Stream environments reflect, in part, the hydrologic integration of upstream watershed characteristics (Allan, 2004). These characteristics can include both natural (e.g., climate, geology, and soils) and anthropogenic (e.g., agriculture and urbanization) features, and understanding their distribution and composition within watersheds is critical to river research and management. We present a new, publically available database of watershed features for several million streams within the conterminous United States (U.S.). This database provides an easily accessible suite of landscape metrics for scientists and managers to analyze, map, and understand the distribution of characteristics and conditions of the Nation's rivers and streams.

Despite advances in both computing power and tools for hydrologic analyses, it is a major challenge to accurately delineate upstream watershed boundaries and characterize landscape features within them without specialized geospatial expertise. These tasks are especially challenging if hundreds or thou-

[1]Paper No. JAWRA-15-0061-P of the *Journal of the American Water Resources Association* (JAWRA). Received May 6, 2015; accepted August 18, 2015. © 2015 American Water Resources Association. This article is a U.S. Government work and is in the public domain in the USA. **Discussions are open until six months from issue publication**.

[2]Oak Ridge Institute for Science and Education Post-doctoral Participant (Hill, Thornbrugh), Geographer (Weber), Research Ecologist (Leibowitz), and Environmental Statistician (Olsen), National Health and Environmental Effects Research Laboratory, Western Ecology Division, U.S. Environmental Protection Agency, 200 SW 35th St., Corvallis, Oregon 97333 (E-Mail/Hill: hill.ryan@epa.gov).

sands of watershed delineations are needed or if a study spans a large geographic extent. Furthermore, we need the ability to easily apply analytical results to new, unsampled streams, or to a complete network of streams for mapping. This ability would provide a powerful tool for visualizing and understanding analytical results in a spatially explicit way.

We developed an extensive database of natural and anthropogenic landscape metrics for ~2.65 million streams, and their associated catchments, within the conterminous U.S.: The Stream-Catchment (StreamCat) Dataset (http://www2.epa.gov/national-aquatic-resource-surveys/streamcat). These metrics are statistical summaries of GIS layers (e.g., climate data) for both catchments (i.e., the nearby landscape that flows directly into a stream segment) and full upstream watersheds (see section Definition of Terms for additional details). We developed the StreamCat Dataset using the National Hydrography Dataset (NHD) Plus Version 2 (NHDPlusV2) (McKay *et al.*, 2012). The NHDPlusV2 is a publically available geospatial framework that depicts the network of streams and rivers within the conterminous U.S. based on digitized lines of U.S. Geological Survey (USGS) topographic quadrangle maps. The StreamCat Dataset and NHDPlusV2 can be linked to rapidly provide spatially explicit watershed information for analysis and management applications. Similar datasets have been developed for NHDPlus Version 1 (e.g., Wang *et al.*, 2011), but NHDPlusV2 substantially improved the spatial representation of hydrologic features and replaces NHDPlus Version 1 (McKay *et al.*, 2012).

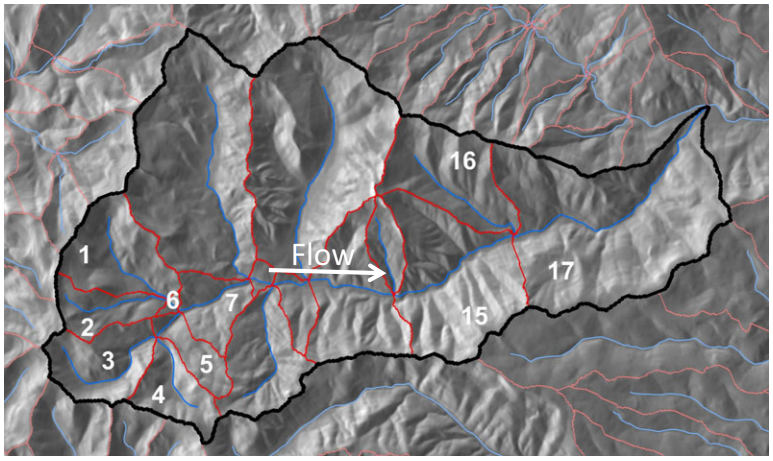
In this article, we describe the development and main features of the StreamCat Dataset. We begin with definitions of terms and then a brief description of the NHDPlusV2 framework, upon which the dataset was built. We next describe the suite of natural and anthropogenic landscape layers that are the basis of the watershed metrics. We then detail the algorithm and the process we used to intersect and accumulate these geospatial layers with the NHDPlusV2 to provide both catchment and full watershed-level metrics. Considerable effort was expended to document the quality and consistency of each dataset and processing step, and we provide a brief description of these efforts. In addition to this overview, the StreamCat website (accessible through: http://www2.epa.gov/national-aquatic-resource-surveys/streamcat) contains comprehensive metadata for each watershed metric. All Python (Lutz, 2006) and R (R Development Core Team 2013) scripts used in data development are also available online. The development of the StreamCat Dataset is part of a multiphase project. Future efforts will process additional landscape layers, providing watershed metrics that will be appended to the website and be available for download, including associated metadata. The StreamCat Dataset will provide scientists and managers with rapid access to information on upstream characteristics, such as land use and land cover, for specific stream reaches. Spatially explicit maps of this information could, for example, inform users of how land use influences water quality and quantity within a state or region and improve both ecological and environmental modeling. These data could provide a powerful tool to identify stream reaches that are at risk of impairment or, conversely, stream reaches that could be prioritized for protection to maintain water quality and habitat. In addition to the StreamCat Dataset, we are developing a similar dataset that will contain landscape characterizations of lakes and their associated catchments (LakeCat).

METHODS

Definition of Terms

The nested nature of stream networks, and their associated watersheds, can lead to confusion when discussing the various components of hydrographic data. In the U.S., the terms "catchments," "basins," and "watersheds" are often used interchangeably. However, we found it useful to define and adhere to a set of terms that helped to distinguish between the different components and scales of the NHDPlusV2. Here, we provide explicit definitions of these terms as used in this article and in the StreamCat Dataset.

Stream segments (blue lines in Figure 1, called flowlines in the NHD) are the basis of the NHD and are defined as sections of contiguous stream or river between upstream and downstream tributaries (as depicted in USGS topographic maps), except where the segment is a headwater or terminal stream (e.g., an outlet to the ocean). We avoided using the term "stream reach" because it is used in fluvial geomorphology and ecology and its definition differs from the one provided for stream segments here (Montgomery and Buffington, 1997). The average length of stream segments within the NHD is 1.8 km (SD = 1.9 km). *Catchments* represent the portion of the landscape where surface flow drains directly into an NHD stream segment, excluding any upstream contributions. For example, the red boundaries in Figure 1 depict the catchments of the stream segments within them. We used the term *watershed* to refer to the set of hydrologically connected catchments, consisting of all upstream catchments that contribute flow into any catchment (e.g., the thick



FROM	TO
0	1
0	2
0	3
0	4
1	6
2	6
3	5
4	5
.	.
.	.
.	.
15	17
16	17

FIGURE 1. An Example Watershed and Associated FROM-TO Flow Table. Background image is a landscape raster depicting relief (hill shade) based on a 30-m digital elevation model (DEM). Red boundaries represent catchments associated with each stream segment (blue lines), that is, the portion of the topography where precipitation would flow directly into each stream line, excluding any upstream contributions. All catchments within the black outline flow to the most downstream catchment, labeled catchment 17. These local flow connections are documented in the FROM-TO flow table. Since headwater catchments have no upstream contributors, these catchments have a zero in the FROM column.

black boundary in Figure 1 is the watershed for catchment 17). In this definition, we also include the catchment (e.g., catchment 17) as part of the watershed. However, each catchment has a full watershed associated with it. For headwater streams, catchments and watersheds are equivalent.

Additional terms used in this article include *landscape layer*, *landscape raster*, *catchment metric*, and *watershed metric*. We use landscape layer to refer generically to any geospatial layer from which we derived statistical summaries for both catchments and watersheds. These landscape layers were comprised of several GIS formats, including vector (points, lines, or polygons) and raster. Landscape rasters are gridded representations of particular features, such as elevation (e.g., shaded relief map in Figure 1), soil properties, or climate. We used *catchment metric* and *watershed metric* to refer, respectively, to catchment- and watershed-level statistical summaries of landscape layers or rasters that comprise the primary products of the StreamCat Dataset.

NHDPlusV2 Framework

The NHDPlusV2 is a value-added product of the original NHD (McKay *et al.*, 2012). It was created, in part, by using digital stream networks (i.e., geospatial vector data) from the original USGS NHD (USGS, 2001) in combination with the ridgelines from the Watershed Boundary Dataset (USGS & USDA, 2013) to adjust 30-m digital elevation models (DEM) (Figure 1). Doing so ensured that depictions of flow across DEM surfaces (i.e., flow directions and accumulations) matched the hydrologic connections of the NHD (McKay *et al.*, 2012). These flow rasters were then used to delineate catchments (red boundaries in Figure 1) for NHD stream segments (blue lines in Figure 1).

Catchments are the main geospatial unit we used to derive the StreamCat Dataset. Critically, the NHDPlusV2 represents the hydrologic relationship between each stream segment, and hence their associated catchments, in topological flow tables (Figure 1). A topological flow table typically contains FROM and TO columns to identify upstream unit(s) where the reach comes from and downstream unit(s) where the reach flows to, respectively. Values within each column represent the unique IDs ("COMID" in NHDPlusV2) of associated catchments. If a stream segment contributes flow into a downstream segment, its COMID is listed in the FROM column and the COMID of the receiving stream segment is listed in the TO column. For example, stream segments 1-4 in Figure 1 are headwater streams and the values in the FROM column of the flow table are zero because there are no contributing upstream segments besides themselves. However, stream segments 1 and 2 flow into segment 6. Hence, the FROM column contains values 1 and 2 that correspond to value 6 in the TO column. Likewise, stream segments 3 and 4 flow into segment 5 (Figure 1). This sequence of FROM-TO relationships continues to stream segment 17, which receives flow directly from segments 15 and 16 (bottom of table in Figure 1). FROM-TO relationships are documented for ~2.65 million stream segments within the NHDPlusV2. We wrote scripts to link these rela-

tionships in a way that accumulates statistical summaries of landscape layers for each catchment (see Generating Local-Catchment and Watershed Summaries for details).

There are several exceptions to the general description of the NHDPlusV2 we provided here. For example, not all line segments within the NHDPlusV2 represent streams. Some represent canals and pipelines, while others are artificial connectors through lakes to maintain topological relationships. In most cases, these canals and pipelines are not included in the NHDPlusV2 flow tables. Thus, the NHDPlusV2 flow tables represent the natural hydrologic connections of river networks and exclude man-made networks. However, artificial line segments that connect lake inlets and outlets are generally included in flow tables, have catchment delineations, and are therefore included in the StreamCat Dataset. In addition, some stream segments were too short for a catchment to be delineated. In such cases, these records were included within an upstream or downstream catchment. Line segments within the NHDPlusV2 that are missing catchment delineations or are not contained within flow tables are excluded from the StreamCat Dataset. Last, some catchments represent sinks (e.g., internally draining basins in eastern Oregon) that contain no streams with defined flow direction in NHDPlusV2. Potential users of the StreamCat Dataset are encouraged to first familiarize themselves with the NHDPlusV2 (McKay *et al.*, 2012).

Landscape Layers

To develop the StreamCat Dataset, we derived watershed-scale summaries of landscape layers with consistent and complete coverage across the conterminous U.S. in two phases. The first phase was based on a literature review that identified watershed metrics that had been linked to instream biota or could be linked to habitat condition. For the second phase, we identified new landscape layers that we hypothesized could improve our characterization of upstream watershed conditions. Phase two of this project is in progress and new layers and spatial representations of these layers will be added as they become available (see the StreamCat Dataset for updates, available through http://www2.epa.gov/national-aquatic-resource-surveys/streamcat).

Phase One Layers. For the first phase of this project, we obtained layers that had been used previously in three papers: Carlisle *et al.* (2009), Falcone *et al.* (2010), and Wang *et al.* (2011) (see Appendix S1 and S2 for a complete list of phase-one landscape layers and watershed metrics, respectively). Carlisle *et al.* (2009) and Falcone *et al.* (2010) used empirical techniques to relate stream condition to a suite of landscape metrics. We selected several layers from these studies that were important for explaining differences in condition among sample sites. We used all metrics identified in Wang *et al.* (2011) because it was a similar dataset to the StreamCat Dataset, except that it was built on NHDPlus Version 1. Natural layers consisted of land cover (Homer *et al.*, 2007), soils (USDA, 2006), lithology (Cress *et al.*, 2010), runoff (McCabe and Wolock, 2011), and topography (i.e., DEM; USGS, 2006). Anthropogenic layers included roads (TIGER Lines; USCB, 2014), dams (USACE, 2009), mines (USGS, 2003), US Census data on population and housing unit densities (TIGER Lines; USCB, 2014), land use (urbanization and agriculture) (Homer *et al.*, 2007), imperviousness of man-made surfaces (Homer *et al.*, 2007), and EPA Facilities Registry Service locations (e.g., Superfund sites; USEPA, 2006). In addition to these raw layers, we manipulated or combined several layers to match metrics derived by one or more of these previous studies. For example, we combined DEM-derived slopes with the National Land Cover Database (NLCD) (Homer *et al.*, 2007) to identify agricultural land on slopes $\geq$10% based on a metric reported by Carlisle *et al.* (2009). In addition, we derived several metrics that constrained calculations to within 100-m buffers (Carlisle *et al.*, 2009) of the NHD stream lines (Appendix S2).

Phase Two Layers. We conducted an extensive search for publically available, conterminous U.S.-wide landscape layers that we hypothesized could improve our representation of natural and anthropogenic watershed features. To date, we have identified and are processing several landscape layers, including year-specific and summer season PRISM air temperature and precipitation (Daly *et al.*, 2008); forest cover change during 2000-2010, including fires (Rollins and Frame, 2006; Hansen *et al.*, 2013); atmospheric N deposition (NADP, 2007); several natural and human-related sources of N and P (Sobota *et al.*, 2013); locations and areal extents of protected lands (USGS, 2012); and barriers to fish passage (a modified dams layer) (Ostroff *et al.*, 2013). In addition to new layers, we are exploring several riparian and upstream distance-weighting schemes to better characterize the spatial arrangement of landscape features within catchments and watersheds (Van Sickle and Burch Johnson, 2008; Peterson *et al.*, 2011). As new tables become available, they will be added to the StreamCat website with their associated metadata.

Generating Local-Catchment and Watershed Summaries

Catchments. The first step to create watershed metrics was to generate statistical summaries of landscape layers for each catchment. We overlaid the catchment boundaries onto landscape layers and calculated a suite of summary statistics (ArcGIS Zonal Statistics tool implemented with a Python script, ESRI, 2014) (Figure 2a). For example, we might calculate the sum, mean, minimum, and maximum of pixel or point values as well as the count of pixels or points for continuous data types (e.g., soil permeability raster and reservoir locations and volumes). For categorical rasters (e.g., land cover type), we calculated the sum of pixels of each category within each catchment. These data were stored in tables with their associated catchment COMIDs (Figure 2b; only count and sum are shown). During processing, all rasters were resampled to a 30-m resolution to match the resolution of NHDPlusV2 flow rasters. Doing so standardized the area of each pixel to facilitate calculations and avoided small catchments not properly intersecting with large resolution rasters.

Accumulating Metrics. The NHDPlusV2 flow tables report the immediate, local flow relationships (e.g., table in Figure 1) and several approaches exist for navigating and routing information using these relationships (e.g., the NHDPlusV2 CA3TV2 Tool) (McKay *et al.*, 2012). Two common approaches are to use the flow tables to route flow, and thus information, downstream (e.g., Tsang *et al.*, 2014) or to create lists of all upstream catchments that flow into each catchment (e.g., Table 1 and Figure 1). As an example of the latter approach, a list of catchments (1-16)

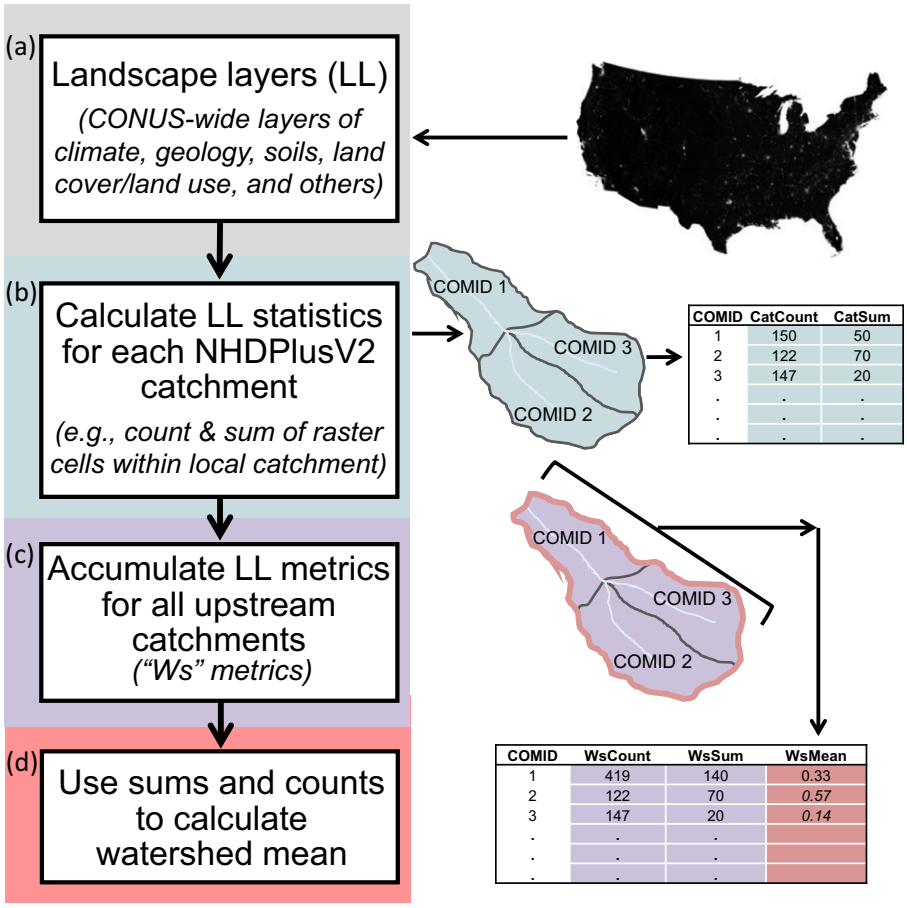


FIGURE 2. Process to Generate Catchment and Watershed-Level Metrics from (a) a Suite of Landscape Layers. Local-catchment summaries of landscape layers are produced and stored in a table with the associated catchment's unique ID (COMID) (b). An algorithm produces accumulations of all upstream catchments (c) that can then be used to calculate watershed-level summaries, such as a watershed mean (d). In this example, catchments 2 and 3 are headwater catchments and their catchment and watershed values are, therefore, equal (*cf.* output from b and d). In contrast, catchment 1 receives flow from catchments 2 and 3 and its watershed values (c and d) are the sum of all three outputs from (b). The watershed (Ws) mean (WsMean) is the sum of raster pixel values (WsSum) divided by the count of pixels (WsCount) within the watershed. Other abbreviations used in this figure: CatCount = count of raster pixels within a catchment, CatSum = sum of pixel values within a catchment.

TABLE 1. Example of Accumulation Lists for Catchments Depicted in Figure 1. The first record in each list is the focal catchment for which watershed metrics will be calculated. Headwater catchments (e.g., catchments 1-4) contain no additional catchments in their respective lists. In contrast, the list for catchment 17 contains all contributing catchments within the black border of Figure 1. Catchment lists were then used to query statistical summaries of landscape information that were made for each catchment to calculate full watershed-level metrics. Lists for catchments 8-16 are not shown in this example.

Focal Catchment COMID	List of Contributing Upstream Catchments						
1							
2							
3							
4							
5	4	3					
6	2	1					
7	6	5	4	3	2	1	
.							
.							
.							
17	16	15	14	.	.	.	1

within the black border of Figure 1 is associated with catchment 17 (Table 1). We used a Python script to implement this approach to calculate watershed-level metrics for all ~2.65 million catchments within the NHDPlusV2. Once these associations were made, we summarized catchment-level metrics up to watershed-level metrics. However, the specific algorithm depended on the type of data being summarized. For rasters of continuous data types, such as soil permeability, we calculated watershed averages as follows:

$$R_w = \sum_c x_c / \sum_c n_c \quad c \in w \tag{1}$$

where R_w is the raster average for watershed w, x_c is the sum of raster pixel values in catchment c, and n_c is the count of raster pixels with data in catchment c. Catchment c is a member of the list of catchments that compose watershed w (i.e., $c \in w$). For categorical rasters, such as land cover class, we calculated the percent of the watershed composed of each class with:

$$P_{w,i} = 100 \sum_c x_{c,i} / \sum_c n_c \quad c \in w \tag{2}$$

where $P_{w,i}$ is the percent of watershed w composed of class i and $x_{c,i}$ is the count of pixels of class i within catchment c (all other notation from Equation 1). Note that n_c is the count of pixels with data across all classes.

We did not use watershed areas to calculate the watershed metrics in Equations (1) and (2), because some catchments cross international borders whereas most landscape layers do not. That is, dividing the sum of pixel values by the accumulated watershed areas would artificially down-weight metrics in those watersheds. To account for international borders with vector point or line data (e.g., dams or roads), we calculated watershed densities and averages as follows:

$$V_w = \sum_c x_c / \sum_c \pi_c A_c \quad c \in w \tag{3}$$

in which V_w is the density or average value of vector point or line values in watershed w, x_c is the count of points or sum of point or line values in catchment c, π_c is the proportion of area of catchment c within conterminous U.S. borders, and A_c is the total area of catchment c, including areas that cross international borders. As with Equations (1) and (2), w includes all upstream catchments and the catchment of interest. It is important to note that in braided networks, our approach does not split flow, and thus information, between divergent stream segments. Instead, each side of the braid is associated with the full, upstream-contributing watershed. This approach is appropriate for relating upstream features (e.g., vurbanization) to water quality parameters. However, this approach may not be appropriate for modeling conservative water features, such as flow. Approaches exist for splitting flow across braided networks and these approaches should be considered depending on the required application (McKay *et al.*, 2012).

The accumulation operation was done for each landscape layer and for each of the ~2.65 million NHDPlusV2 catchments. In addition, because the NHDPlusV2 is divided and distributed as major river basins (Figure 3), the Python script passed data between these hydrologically connected regions. Specifically, accumulated data from regions 05, 06, 07, 10U, 10L, and 11 were passed from one to another and ultimately to region 08 (Figure 3), the outlet of the Mississippi Basin, to produce a hydrologically complete characterization of upstream watershed features. Likewise, the script passed data from the upper (region 14) to the lower (region 15) Colorado Basins (Figure 3).

In addition to characterizing catchment and watershed features, we documented the percent area of each catchment and watershed that overlapped with each landscape layer because most layers stopped at the international borders or contained missing values in some locations. These percent-completeness metrics were retained and are available for download so that users can evaluate the appropriateness of these data for analyses.

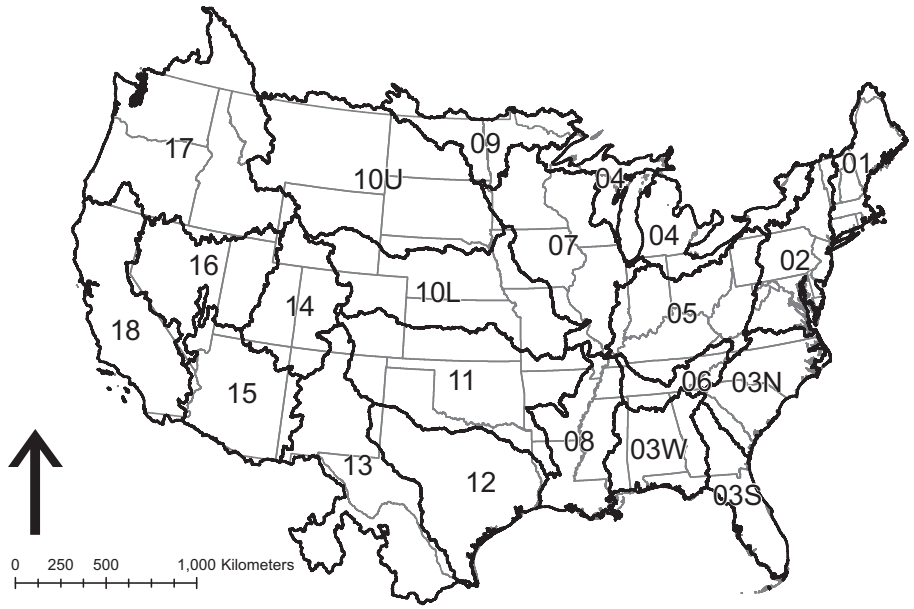


FIGURE 3. Major Basins of the National Hydrography Dataset Plus Version 2 (NHDPlusV2) for the Conterminous United States. The NHDPlusV2 divides and distributes data based on these 21 basin boundaries. Within each basin, topological relationships between catchments are included. However, NHDPlusV2 does not contain topologic relationships that cross basins.

DATA

StreamCat Dataset Validation

We developed and executed a quality assurance protocol for each step of the process to create the StreamCat Dataset, including checks of landscape layers (see StreamCat website for metadata and complete documentation of quality assurance procedures), catchment summaries, and full watershed metrics. Landscape layers comprised several data types (continuous, proportional, and categorical), resolutions, GIS formats (rasters and vector points, lines, and polygons), and spatial projections. Our quality assurance protocol documented these characteristics, the date of layer acquisition, and data units if applicable. All layers were opened within a GIS and visually inspected to identify any spatial gaps within the data and to verify that all datasets covered the extent of the conterminous U.S. After documenting the original characteristics of each dataset, we then transformed that data to standardize the spatial projection (USGS Albers Equal Area Conic projection), units (SI), and missing-data values across layers. Polygon layers were converted into rasters to facilitate processing. All data manipulations were executed in Python and R scripts that are available for download from the StreamCat website. In addition, the StreamCat website provides complete documentation of each layer's original characteristics and any manipulations that we executed to standardize the data.

We verified our accumulation process by comparing our results with the available set of watershed-level metrics that are distributed with NHDPlusV2: PRISM precipitation, the 2001 NLCD, and full watershed areas (McKay *et al.*, 2012). This comparison assumed that by matching accumulation values with NHDPlusV2, our algorithm correctly calculated both catchment- and watershed-level metrics. Our comparisons showed exact matches.

StreamCat Dataset Distribution and Format

The StreamCat Dataset is available as comma-delimited (.csv) text files from http://www2.epa.gov/national-aquatic-resource-surveys/streamcat, and is organized by U.S. State or NHDPlusV2 major basins (Figure 3). The data are distributed based on the GIS layers used to produce them. For example, catchment and watershed summaries of the NLCD data are distributed as a single table that contains the area of each catchment and watershed and the percent of the catchment and watershed composed of each land cover class. In addition, tables contain ancillary information regarding the percent of each catchment and watershed that overlapped with the GIS data layer, which accounts for missing data (e.g., due to international borders). Finally, each table contains a column of COMIDs that can be linked to the NHDPlusV2 stream lines or catchment polygons in a GIS. Furthermore, the data can be queried and accessed with a non-spatial database or software

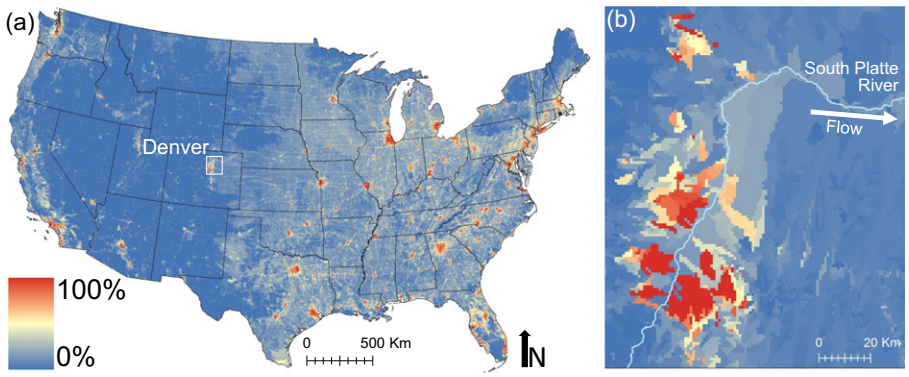


FIGURE 4. Map Showing the Percent of Each Watershed Composed of Urban Land Use for (a) the Conterminous United States and (b) Denver, Colorado. Note that although each value represents the full, upstream watershed metric, these values are mapped for each catchment. Thus, the map depicts the incremental watershed summaries in the downstream direction.

(e.g., MS Excel or R Statistical Software) if the COMIDs of study streams are known beforehand.

DISCUSSION AND CONCLUSION

There are a number of ways that the StreamCat Dataset could be used to improve our understanding, management, and protection of riverine systems. First, the StreamCat Dataset, in conjunction with the NHDPlusV2, can quickly and easily generate information for analyses. Simple queries can link field site locations to stream networks and, as a result, to a large variety of catchment and watershed metrics from the StreamCat Dataset. By providing national-level summaries of watershed metrics, the StreamCat Dataset greatly reduces the specialized technical expertise needed to access these types of data for researchers and managers. Second, analytical results based on a sample of catchments can be extrapolated to the StreamCat Dataset to provide results at new, unsampled locations. When combined with the NHDPlusV2 network and catchments, analytical results can be visualized with maps to better understand the behavior of models. Maps showing how natural and anthropogenic features vary spatially at the catchment and watershed levels can provide insight into how they may influence water quantity and quality. For example, Figure 4 depicts the percent of each watershed that is composed of urban land use (NLCD) for the conterminous U.S. (Figure 4a) and Denver, Colorado (Figure 4b). Note that percent urbanization was calculated for complete watersheds. However, we have depicted the full upstream values at each catchment because it is not possible to display the overlapping nature of incrementally nested watersheds. This visualization shows that watershed urbanization varies greatly within the Denver metropolitan area. Some watersheds are composed almost entirely of urban land use. In contrast, catchments of the South Platte River have relatively low amounts of urban influence because their upstream watersheds flow from the Southern Rocky Mountains before arriving to Denver. When combined with modeling results, such maps could provide important insight into model behavior to improve successive analyses. The StreamCat Dataset can provide a more comprehensive understanding of the distribution of characteristics of the Nation's rivers and streams.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article as outlined here: (1) Appendix S1 - Descriptions of phase-one landscape rasters used to create the StreamCat Dataset; and (2) Appendix S2 - Descriptions of phase-one catchment and watershed-level metrics included in the StreamCat Dataset.

ACKNOWLEDGMENTS

We thank James Falcone and David Wolock of the USGS for providing several landscape layers used to create the StreamCat Dataset. We also thank Cindy McKay of Horizon Systems Corporation (under contract to the EPA Office of Water) for her consultation and advice on use of NHDPlusV2. Comments by Cindy McKay, Richard Moore, Alan Rea, Tommy Dewald, James Markweise, Daren Carlisle, and three anonymous reviewers greatly improved the manuscript. The information in this document has been funded entirely by the U.S. Environmental Protection Agency, in part by an appointment to the Internship/Research Participation Program at the Office of Research and Development, U.S. Environmental Protection Agency, administered by the Oak Ridge Institute for

Science and Education through an interagency agreement between the U.S. Department of Energy and EPA. This manuscript has been subjected to Agency review and has been approved for publication. The views expressed in this journal article are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

LITERATURE CITED

Allan, J.D., 2004. Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. Annual Review of Ecology Evolution and Systematics 35:257-284.

Carlisle, D., J. Falcone, and M. Meador, 2009. Predicting the Biological Condition of Streams: Use of Geospatial Indicators of Natural and Anthropogenic Characteristics of Watersheds. Environmental Monitoring and Assessment 151:143-160.

Cress, J., D. Soller, R. Sayre, P. Comer, and W. Harumi, 2010. Terrestrial Ecosystems—Surficial Lithology of the Conterminous United States. Scientific Investigations Map 3126, U.S. Geological Survey, Denver, Colorado.

Daly, C., M. Halbleib, J.I. Smith, W.P. Gibson, M.K. Doggett, G.H. Taylor, J. Curtis, and P.P. Pasteris, 2008. Physiographically Sensitive Mapping of Climatological Temperature and Precipitation across the Conterminous United States. International Journal of Climatology 28:2031-2064.

ESRI, 2014. ArcGIS Desktop: Release 10.2.2. Environmental Systems Research Institute, Redlands, California.

Falcone, J.A., D.M. Carlisle, and L.C. Weber, 2010. Quantifying Human Disturbance in Watersheds: Variable Selection and Performance of a GIS-Based Disturbance Index for Predicting the Biological Condition of Perennial Streams. Ecological Indicators 10:264-273.

Hansen, M.C., P.V. Potapov, R. Moore, M. Hancher, S.A. Turubanova, A. Tyukavina, D. Thau, S.V. Stehman, S.J. Goetz, T.R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C.O. Justice, and J.R.G. Townshend, 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science 342:850-853.

Homer, C., J. Dewitz, J. Fry, M. Coan, N. Hossain, C. Larson, N. Herold, A. McKerrow, J.N. VanDriel, and J. Wickham, 2007. Completion of the 2001 National Land Cover Database for the Conterminous United States. Photogrammetric Engineering and Remote Sensing 73:337-341.

Lutz, M., 2006. Programming Python. O'Reilly Media Inc., Sebastopol, California.

McCabe, G.J. and D.M. Wolock, 2011. Independent Effects of Temperature and Precipitation on Modeled Runoff in the Conterminous United States. Water Resources Research 47:W11522.

McKay, L., T. Bondelid, T. Dewald, J. Johnston, R. Moore, and A. Reah, 2012. NHDPlus Version 2: User Guide. ftp://ftp.horizon-systems.com/NHDPlus/NHDPlusV21/Documentation/NHDPlusV2_User_Guide.pdf, *accessed* February 2015.

Montgomery, D.R. and J.M. Buffington, 1997. Channel-Reach Morphology in Mountain Drainage Basins. Geological Society of America Bulletin 109:596-611.

NADP (National Atmospheric Deposition Program), 2007. Illinois State Water Survey. NADP Program Office, Champaign, Illinois.

Ostroff, A., D. Wieferich, A. Cooper, and D. Infante, 2013. 2012 National Anthropogenic Barrier Dataset (NABD). National Fish Habitat Partnership Data System. https://www.sciencebase.gov/catalog/item/512cf142e4b0855fde669828, *accessed* May 2015.

Peterson, E.E., F. Sheldon, R. Darnell, S.E. Bunn, and B.D. Harch, 2011. A Comparison of Spatially Explicit Landscape Representation Methods and Their Relationship to Stream Condition. Freshwater Biology 56:590-610.

R Development Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rollins, M.G. and C.K. Frame, 2006. The LANDFIRE Prototype Project: Nationally Consistent and Locally Relevant Geospatial Data for Wildland Fire Management. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado.

Sobota, D.J., J.E. Compton, and J.A. Harrison, 2013. Reactive Nitrogen Inputs to US Lands and Waterways: How Certain Are We about Sources and Fluxes? Frontiers in Ecology and the Environment 11:82-90.

Tsang, Y.-P., D. Wieferich, K. Fung, D. Infante, and A. Cooper, 2014. An Approach for Aggregating Upstream Catchment Information to Support Research and Management of Fluvial Systems Across Large Landscapes. SpringerPlus 3:589.

USACE (U.S. Army Corps of Engineers), 2009. National Inventory of Dams. U.S. Army Corps of Engineers, Washington, D.C.

USCB (U.S. Census Bureau), 2014. TIGER/Line Shapefiles (machine-readable data files). http://www.census.gov/geo/maps-data/data/tiger.html, *accessed* March 2014.

USDA (U.S. Department of Agriculture), 2006. US General Soil Map (STATSGO)|NRCS NCGC. http://www.ncgc.nrcs.usda.gov/products/datasets/statsgo/, *accessed* January 2014.

USEPA (U.S. Environmental Protection Agency), 2006. National Facility Registry Service (FRS). http://www.epa.gov/enviro/html/fii/index.html, *accessed* February 2014.

USGS (U.S. Geological Survey), 2001. National Hydrography Dataset (NHD). Reston, Virginia.

USGS (U.S. Geological Survey), 2003. Active Mines and Mineral Processing Plants. http://tin.er.usgs.gov/metadata/mineplant.faq.html, *accessed* January 2014.

USGS (U.S. Geological Survey), 2006. National Elevation Dataset (NED). http://ned.usgs.gov, *accessed* February 2014.

USGS (U.S. Geological Survey), 2012. Protected Areas Database of the United States (PADUS) Version 1.3. USGS National Gap Analysis Program. http://gapanalysis.usgs.gov/padus/, *accessed* April 2015.

USGS (U.S. Geological Survey) & USDA (U.S. Department of Agriculture), 2013. Federal Standards and Procedures for the National Watershed Boundary Dataset (WBD) (Fourth Edition). U.S. Geological Survey Techniques and Methods 11-A3, 63 pp.

Van Sickle, J. and C. Burch Johnson, 2008. Parametric Distance Weighting of Landscape Influence on Streams. Landscape Ecology 23:427-438.

Wang, L., D. Infante, P. Esselman, A. Cooper, D. Wu, W. Taylor, D. Beard, G. Whelan, and A. Ostroff, 2011. A Hierarchical Spatial Framework and Database for the National River Fish Habitat Condition Assessment. Fisheries 36:436-449.