StreamCatTools: An R package for working with StreamCat and LakeCat watershed data in R

Marc H. Weber[1], **Ryan A. Hill**[2], **Travis Hudson**[3], **Selia Markley**[4], **and Alan Brookes**[2]

1 Office of Water, United States Environmental Protection Agency **2** Office of Research and Development, United States Environmental Protection Agency **3** Oak Ridge Associated Universities Student Services Contractor c/o United States Environmental Protection Agency **4** Oak Ridge Institude for Science and Education Fellow c/o United States Environmental Protection Agency

Contents
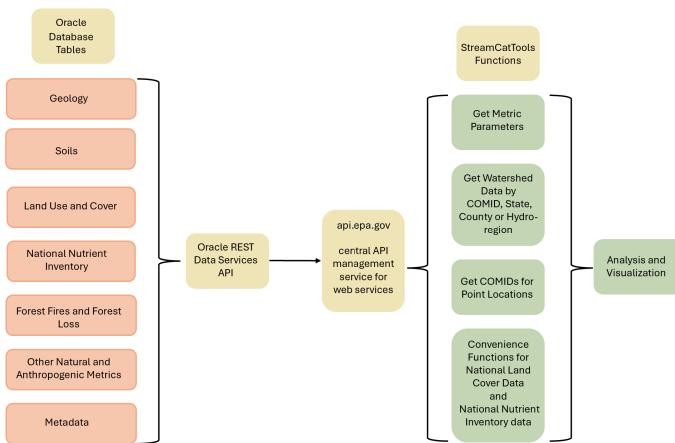
Summary

`StreamCatTools` provides functions for easily working with, visualizing and analyzing StreamCat (Hill, Weber, Leibowitz, Olsen, & Thornbrugh, 2016) and LakeCat (Hill, Weber, Debbout, Leibowitz, & Olsen, 2018) watershed metrics within **R**. The StreamCat and LakeCat datasets provide hundreds of landscape metrics for both the local catchment (e.g. landscape draining to a particular stream reach) and full watershed for every stream reach and lake depicted in the medium resolution National Hydrography Dataset Plus Version 2.1 (NHDPlusV21)(McKay et al., 2012) for the contiguous United States (CONUS). `StreamCatTools` functions wrap the application programming interface (API) for the StreamCat and LakeCat online database and facilitate simple, straightforward access and use of these datasets within R.

Statement of Need

Easily accessible, robust, and consistent watershed data is an underpinning of hydrology research, water quality monitoring programs, and predictive modelling applications, to name just a few examples. The StreamCat (Hill et al., 2016) and LakeCat (Hill et al., 2018) datasets fill this need by providing nationally consistent curated watershed data for CONUS that has had stringent quality control applied. The data encompasses hundreds of watershed metrics for every stream reach and lake feature represented in the NHDPlusV21 (McKay et al., 2012). StreamCatTools fills the need for easily accessible watershed metrics for CONUS by: (1) providing a simple interface in R to the StreamCat and LakeCat web services, (2) providing convenient functionality to find available StreamCat and LakeCat

Figure 1: Diagram of the StreamCat and StreamCatTools framework. A backend Oracle database and web service are exposed through api.gov to functions in StreamCatTools which simplify accesss and analysis of the data in R via the application programming interface.

metric names and information, (3) extracting StreamCat and LakeCat metrics by COMID (a unique identifier in the NHDPlusV2 framework), by state, by county, by NHD Hydroregion, or for all of CONUS. Providing this valuable watershed data via web services in R follows the FAIR principles laid out in (Wilkinson et al., 2016).

StreamCatTools Functionality

`StreamCatTools` provides a simple streamlined set of functions to easily query and ingest watershed landscape metrics into an R session. Figure 1 shows the overall framework of the StreamCat database, application programming interface, and functionality in the package that simplifies data access in R using web services for the StreamCat and LakeCat datasets.

The core functions in `StreamCatTools` leverage the `httr2` library (Wickham, 2025) for a modern, pipeable method for working with APIs. Specifically, `StreamCatTools` simplifies the calls to the API for StreamCat and LakeCat data within the R programming language to allow a user to: (1) access details on available StreamCat and LakeCat metrics, (2) match users' sites to NHDPlusV21 waterbodies (streams or lakes) to access StreamCat and Lakecat data, and (3) retrieve StreamCat and LakeCat watershed metrics by COMID, by state, by county, by hydroregion, or for all of CONUS. Additionally, several convenience functions retrieve data from particular years for data with time series, such as the National Land Cover Database (NLCD) (USGS, 2024) and the National Nutrient Inventory (NNI) (Brehob et al., 2025).

You can install the most recent version of `StreamCatTools` from CRAN by running:

```
# install.packages("StreamCatTools")
```

You can install the most recent development version from GitHub by running:

```
# library(remotes)
# install_github("USEPA/StreamCatTools", build_vignettes=FALSE)
```

`StreamCatTools` is loaded into an **R** session:

```
library(StreamCatTools)
```

StreamCatTools includes several functions to facilitate accessing and working with the data. First, users can list metric names and find out more about StreamCat and LakeCat data available in StreamCat using using the `sc_get_params` function. This function also provides details about metrics, such as area of interest for which data are available (e.g., local catchment vs. watershed) and years of available data if applicable. `sc_get_params` returns a tibble which facilitates automated access and searches of metric characteristics. Additional functions provided for getting metadata about the underlying StreamCat and LakeCat data include the `sc_fullname` and `lc_fullname` functions and the `sc_get_params` and `lc_get_params` functions, respectively. Users can also filter metric names and information by the metric year(s), the indicator categories for metrics, the metric dataset names, or the areas of interest the metrics are available for using the `sc_get_metric_names` or `lc_get_metric_names` functions. More details on these functions can be found at the [package introduction page](#).

The primary package functionality is in the `sc_get_data` and `lc_get_data` functions which allow users to extract catchment or watershed metrics of interest by providing NHDPlusV21 COMIDs for streams or lakes within the database. Users can also request data for a given state(s), county(ies), hydroregion(s), or all of CONUS. Additionally, convenience functions are provided for accessing the NLCD and NNI datasets, respectively, using `sc_nlcd` and `lc_nlcd` and `sc_nni` and `lc_nni`.

The following example shows a request for two catchment and watershed metrics for three stream reaches:

```
df <- sc_get_data(metric='pcturbmd2019,damdens',
                  aoi='cat,ws',
                  comid='179,1337,1337420')
```

The first line requests percent of NLCD medium intensity developed land cover in 2019 and the density of dams. The second line specifies that the geographic area of interest (aoi) is the drainage to the local reach scale (i.e. 'cat', short for catchment) and the full watershed. Finally, the final line request this data for three NHDPlusV2 stream segments specified by their unique COMIDs.

Applications

Watershed metrics from `StreamCatTools` can be easily visualized with functions from `nhdplusTools` (Blodgett & Johnson, 2023) and `ggplot2` (Wickham, 2016). The example below plots the NLCD percent imperviousness for the the local drainage (catchment in NHDPlusV2 syntax) and displays the values mapped to each stream reach and to the overall basin boundary:

```
library(nhdplusTools)
library(ggplot2)
library(ggspatial)
library(StreamCatTools)

start_comid = 23763517
nldi_feature <- list(featureSource = "comid", featureID = start_comid)
flowline_nldi <- nhdplusTools::navigate_nldi(nldi_feature, mode = "UT",
↪   data_source = "flowlines", distance=5000)
df <- sc_get_data(metric='pctimp2019', aoi='cat',
↪   comid=flowline_nldi$UT_flowlines$nhdplus_comid)
flowline_nldi <- flowline_nldi$UT_flowlines
flowline_nldi$PCTIMP2019 <- df$pctimp2019cat[match(flowline_nldi$nhdplus_comid,
↪   df$comid)]
basin <- nhdplusTools::get_nldi_basin(nldi_feature = nldi_feature)
```
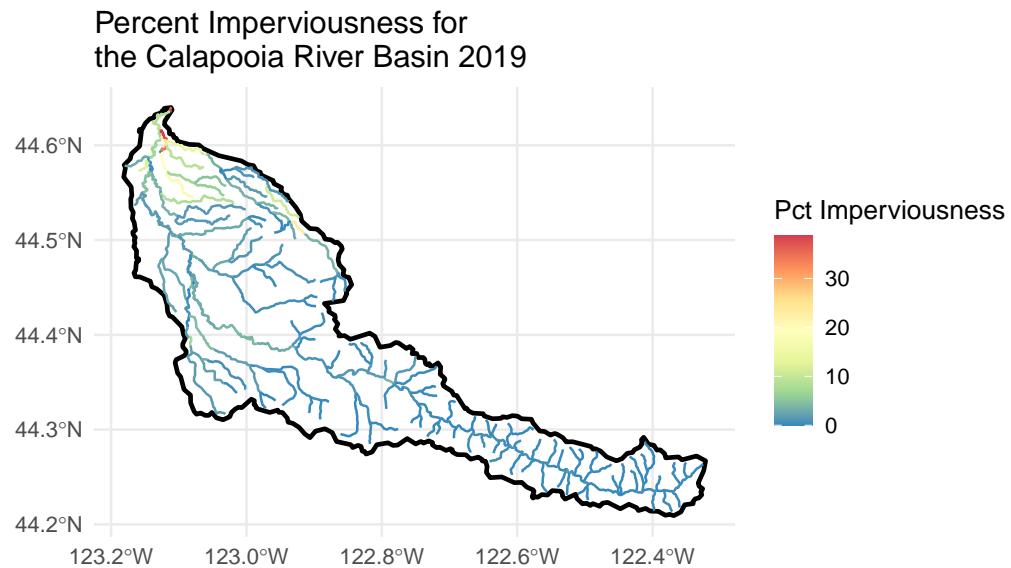
```r
calapooia <- ggplot() +
    geom_sf(data = basin,
            fill = NA,
            color = "black",
            linewidth = 1) +
    geom_sf(data = flowline_nldi,
            aes(colour = PCTIMP2019)) +
    scale_y_continuous() +
    scale_color_distiller(palette = "Spectral") +
    labs(color = "Pct Imperviousness") +
    theme_minimal(12) +
    ggtitle('Percent Imperviousness for \nthe Calapooia River Basin 2019')

plot(calapooia)
```


Also show: new functionality for accessing and plotting NNI data. Perhaps any other uses or applications we are aware of (such as CASTools R Shiny app).

Acknowledgements

References

Blodgett, D., & Johnson, M. (2023). *nhdplusTools: Tools for accessing and working with the NHDPlus.* Reston, VA: U.S. Geological Survey; U.S. Geological Survey. Retrieved from https://doi.org/10.5066/P97AS8JD

Brehob, M. M., Pennino, M. J., Compton, J. E., Zhang, Q., Weber, M. H., Hill, R. A., Markley, S., et al. (2025). The US EPA's national nutrient inventory: Critical shifts in US nutrient pollution sources from 1987 to 2017. *Environmental Science & Technology*, *59*(51), 27836–27852. doi:10.1021/acs.est.5c08196

Hill, R. A., Weber, M. H., Debbout, R. M., Leibowitz, S. G., & Olsen, A. R. (2018). The lake-catchment (LakeCat) dataset: Characterizing landscape features for lake basins within the conterminous USA. *Freshwater Science*, *37*(2), 208–221. doi:10.1086/697966

Hill, R. A., Weber, M. H., Leibowitz, S. G., Olsen, A. R., & Thornbrugh, D. J. (2016). The stream-catchment (StreamCat) dataset: A database of watershed metrics for the conterminous united states. *JAWRA Journal of the American Water Resources Association*, *52*(1), 120–128. doi:https://doi.org/10.1111/1752-1688.12372

McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., & Reah, A. (2012). *NHDPlus version 2: User guide.* Retrieved from http://www.horizon-systems.com/NHDPlus/NHDPlusV2_home.php

USGS. (2024). *Annual NLCD collection 1 science products: U.s. Geological survey data release.* U.S. Geological Survey; U.S. Geological Survey. Retrieved from %E2%80%AFhttps://doi.org/10.5066/P94UXNTS

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York. doi:10.1007/978-0-387-98141-3

Wickham, H. (2025). *httr2: Perform HTTP requests and process the responses.* Retrieved from https://httr2.r-lib.org

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific data*, *3*(1), 1–9. doi:10.1038/sdata.2016.18