

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

Marc H. Weber¹, Ryan A. Hill², Travis Hudson¹, Selia Markley¹, and Alan Brookes¹

¹ Office of Water, United States Environmental Protection Agency ² Office of Research and Development, United States Environmental Protection Agency

DOI:

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted:

Published:

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Contents

1	Summary	1
2	Statement of Need	1
3	StreamCatTools Functionality	2
4	Applications	3
5	Acknowledgements	4
	References	4

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 data 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 metric names and information, (3) extracting StreamCat and LakeCat metrics by COMID (a unique identifier in the NHDPlusV2 framework), by state,

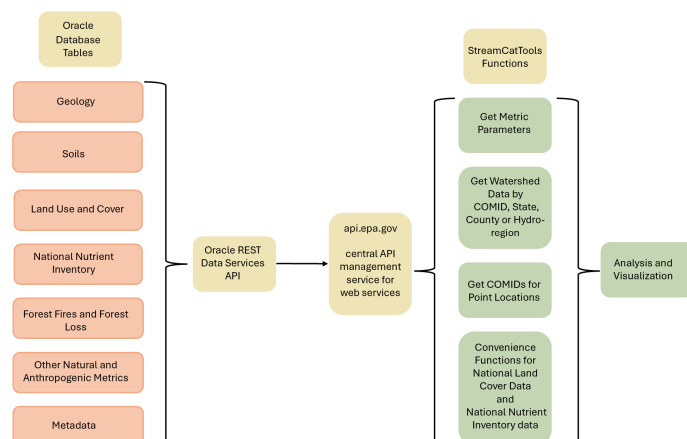


Figure 1: The StreamCat framework: a backend Oracle database, web services exposed through api.gov, and functions in StreamCatTools which simplify access and analyzing the data in R via the API.

by county, by NHD Hydro-region, 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 functionality of **StreamCatTools** simplifies the calls to the API for StreamCat and LakeCat data within the R programming language to allow a user to: 1) ask for details on available StreamCat and LakeCat metrics 2) Reference sites with location information to NHDPlusV2 catchments for stream and lake features in order to ask for StreamCat and LakeCat metrics and 3) Get StreamCat and LakeCat watershed metrics by COMID, by state, by county, by hydroregion, or for all of CONUS. Additional functionality is provided in convenience functions get data all data for particular years for particular data sources such as the National Land Cover Database (NLCD) ((USGS), 2024) and the National Nutrient Inventory (NNI) (Brehob et al., 2025). The core functions in **StreamCatTools** leverage the **httr2** library **httr2**(Wickham, 2025) for a modern, pipeable API for working with web APIs.

Installing StreamCatTools

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

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

```
library(StreamCatTools)
```

As we detail in the [package introduction page](#) we can list metric parameters and find out more about StreamCat and LakeCat data available in **StreamCatTools** using using the **sc_get_params** function. We can also see what metrics are available for what areas of interest and what years using the **sc_get_params** function (which returns a tibble of information about StreamCat metrics). Additional functions provided for getting metadata on

the underlying StreamCat and LakeCat data include the `sc_fullname` and `lc_fullname` functions and the `sc_get_params` and `lc_get_params` functions. We can also filter metric names and information by the metric year(s), the indicator categories for metrics, the metric data set names, or the Areas of Interest the metrics are available for using the `sc_get_metric_names` or `lc_get_metric_names` functions.

The primary package functionality is in the `sc_get_data` and `lc_get_data` functions which allow users to extract whatever suite of metrics they are interested in from StreamCat or LakeCat using a provided set of unique identifiers of COMIDs for NHDPlusV2, or request a set of metrics 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 using `sc_nlcd` and `lc_nlcd` and `sc_nni` and `lc_nni`.

An example of a simple request for several watershed metrics for a particular stream reach in **StreamCatTools** looks like the following:

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

Where we are asking in the first line for the percent of NLCD 1019 medium intensity developed land and the number of dams and in the second line specifying at the drainage to the local reach scale (i.e. 'cat', short for catchment) and the full watershed scale as our areas of interest, and requesting this data just for three NHDPlusV2 stream segments specified by their three COMID values.

Applications

We can easily visualize derived watershed metrics in **StreamCatTools** by using functions from **nhdplusTools** and **ggplot2**, as in this example where plot the NLCD percent imperiousness for the the local drainage (catchment in NHDPlusV2 syntax) and display the values mapped to each stream reach as well as displaying 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 = "nldi")
# df <- sc_get_data(metric='pctimp2019', aoι='cat', comid=flowline_nldi$UT_flowline_comid)
# flowline_nldi <- flowline_nldi$UT_flowlines
# flowline_nldi$PCTIMP2019 <- df$pctimp2019cat[match(flowline_nldi$nhdplus_comid,
#                                                    flowline_nldi$UT_flowline_comid)]
# basin <- nhdplusTools::get_nldi_basin(nldi_feature = nldi_feature)
#
# 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 Imperiousness") +
```

```
# 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

Examples of using StreamCat and LakeCat make extensive use of `nhdplusTools` (Blodgett & Johnson, 2023) and the functions for accessing the API are facilitated through use of `httr2` (Wickham, 2025). Figures were created using `ggplot2` (Wickham, 2016).

We would like to sincerely thank the editor and reviewers for all of their helpful feedback which greatly improved both the software and the manuscript.

The United States Environmental Protection Agency (EPA) GitHub project code is provided on an “as is” basis and the user assumes responsibility for its use. EPA has relinquished control of the information and no longer has responsibility to protect the integrity, confidentiality, or availability of the information. Any reference to specific commercial products, processes, or services by service mark, trademark, manufacturer, or otherwise, does not constitute or imply their endorsement, recommendation or favoring by EPA. The EPA seal and logo shall not be used in any manner to imply endorsement of any commercial product or activity by EPA or the United States Government.

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](https://doi.org/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](https://doi.org/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), U. S. G. S. (2024). *Annual NLCD collection 1 science products: U.s. Geological survey data release*. U.S. Geological Survey; U.S. Geological Survey. Retrieved from <https://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](https://doi.org/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](https://doi.org/10.1038/sdata.2016.18)