nsink: An R package for flow path nitrogen removal estimation

2021-10-21

Summary

The `nsink` package estimates cumulative nitrogen (N) removal along a specified flow path and is based on methodologies outlined in Kellogg et al. (2010). For a user-specified watershed, `nsink` downloads all required datasets from public datasets in the United States, prepares data for use, summarizes N removal along a flow path and creates several static watershed maps. The results of an `nsink` analysis may be exported to standard geospatial files for use in other applications.

Statement of need

Excess N delivery via surface water to downstream aquatic resources contributes to impaired water quality and leads to several ecosystem impacts including harmful algal blooms (HABs) and hypoxia (Rabalais, Turner, and Scavia 2002). Identifying landscape N sinks (i.e., areas where N is effectively removed from the aquatic system) and analysing N delivery at the watershed scale is helpful to watershed managers, land use planners and conservation organizations. The theoretical underpinnings for identifying N sinks rely on decades of research and are explained in Kellogg et al. (2010).

The first implementation of this approach was done case-by-case. Data acquisition and manipulation were mostly manual and took weeks to months to complete for a single watershed. The effort required for the analysis limited it's application as scaling beyond a few pilot studies was not feasible. The goal of `nsink` was to address this limitation and provide an open source solution that could be run on a single small watershed in minutes to hours with minimal manual input.

The `nsink` package

Package Installation

The `nsink` package is available from https://github.com/usepa/nsink and may be installed in R with the following:

```r
# If not installed, install remotes
install.packages("remotes")

# Install nsink from GitHub
remotes::install_github("USEPA/nsink", build_vignettes = TRUE)
```

Package Details

The `nsink` package is designed around the major steps in running a N-Sink analysis and includes the following:

1. Prepare for analysis
   - Get data
   - Prepare data for analysis
   - Calculate relative N removal layer for hydric soils, lakes and streams.

2. Run an point-based analysis
    - Calculate a flow path
    - Summarize relative N removal along a flow path
3. Run a watershed-based analysis
    - Develop static maps
    - Generate output datasets

Required Data

The ability to run an `nsink` analysis relies on several national scale datasets for the United States. By limiting our approach to these national datasets we are ensuring scalability of `nsink` because the datasets will be available for most locations in the United States. The datasets that `nsink` uses are the National Hydrography Dataset Plus (NHDPlus), Soil Survey Geographic Database (SSURGO), the National Land Cover Dataset (NLCD) land cover, and the National Land Cover Dataset (NLCD) impervious surface [ADD CITATION FOR DATASETS]. These datasets are all available via either an Application Programming Interface (API) or via direct download.

Dependencies

The `nsink` package depends on several existing R packages to facilitate spatial data handling, data acquisition, data management, data analysis and data processing. These are detailed in Table 1.

Table 1. R Package Dependencies for the `nsink` package

Package	Task	Citation
`sf`	Spatial Data Handling and Analysis	E. Pebesma (2018); E. Pebesma (2021b)
`raster`	Spatial Data Handling and Analysis	Hijmans (2021)
`stars`	Spatial Data Handling and Analysis	E. Pebesma (2021c)
`fasterize`	Spatial Data Handling and Analysis	Ross (2020)
`lwgeom`	Spatial Data Handling and Analysis	E. Pebesma (2021a)
`gstat`	Spatial Data Handling and Analysis	E. J. Pebesma (2004); Gräler, Pebesma, and Heuvelink (2016); E. Pebesma and Graeler (2021)
`sp`	Spatial Data Handling and Analysis	E. J. Pebesma and Bivand (2005); Bivand, Pebesma, and Gomez-Rubio (2013); E. Pebesma and Bivand (2021)
`units`	Unit Transformations	E. Pebesma, Mailund, and Hiebert (2016); E. Pebesma et al. (2021)
`FedData`	Data Acquisition	Bocinsky (2020)
`httr`	Data Acquisition	Wickham (2020)
`dplyr`	Data Management and Analysis	Wickham et al. (2021)
`zoo`	Data Management and Analysis	Zeileis and Grothendieck (2005); Zeileis, Gorthendieck, and Ryan (2021)

Package	Task	Citation
`igraph`	Data Management and Analysis	Csardi and Nepusz (2006); Csardi et al. (2020)
`readr`	Data Management and Analysis	Wickham and Hester (2020)
`foreign`	Data Management and Analysis	R Core Team (2020)
`rlang`	Data Management and Analysis	Henry and Wickham (2021)
`furrr`	Parallel Processing	Vaughan and Dancho (2021)
`future`	Parallel Processing	Bengtsson (2021); Bengtsson (2020)

Functionality

Currently, `nsink` provides 10 exported functions to facilitate a flow path analysis of relative N removal.

- `nsink_get_huc_id()`: A function for searching the name of a USGS Watershed Boundary Dataset Hydrologic Unit (https://www.usgs.gov/core-science-systems/ngp/national-hydrography/watershed-boundary-dataset) and retrieving its 12-digit Hydrologic Unit Code (HUC).
- `nsink_get_data()`: Using any acceptable HUC ID (e.g. 2-digit to 12-digit), this function downloads the NHDPlus, SSURGO, NLCD Land Cover, and the NLCD Impervious for that HUC.
- `nsink_prep_data()`: `nsink` needs data in a common coordinate reference system, from mutliple NHDPlus tables, and from different portions of SSURGO. This function completes these data preparation steps and outputs all data, clipped to the HUC boundary.
- `nsink_calc_removal()`: Quantifying relative N removal across a landscape is a key aspects of an `nsink` analysis. The `nsink_calc_removal()` function takes the object returned from `nsink_prep_data()` and calculates relative N removal for each landscape sink. See Kellogg et al (2010) for details on relative N removal estimation for each sink.
- `nsink_generate_flowpath()`: This function uses a combination of flow determined by topography, via a flow-direction raster, for the land-based portions of a flow path and of downstream flow along the NHDPlus stream network.
- `nsink_summarize_flowpath()`: Summarizing removal along a specified flow path requires relative N removal and a generated flow path. This function uses these and returns a summary of relative N removal along a flow path for each sink.
- `nsink_generate_static_maps()`: This function analyzes N removal at the watershed scale by summarizing the results of multiple flow paths. Four static maps are returned: 1)removal efficiency; 2)loading index; 3)transport index; 4)delivery index. Removal efficiency is a rasterized version of the `nsink_calc_removal()` output. Loading index is N sources based on NLCD categories. Transport index is a heat map with the cumulative relative N removal along flow paths originating from a grid of points, density set by the user, across a watershed, highlighting the gradient of downstream N retention. Delivery index is the result of multiplying the loading index and the transport index, and shows potential N delivery from different sources, taking into account the relative N removal as water moves downstream.
- `nsink_plot()`: A function that plots each raster in the list returned from `nsink_generate_static_maps()`.
- `nsink_build()`: One of the drivers behind the development of the `nsink` package was to provide `n-sink` analysis output that could be used more broadly (e.g. within a GIS). The `nsink_build()` runs a complete `nsink` analysis and outputs R objects, shapefiles and/or TIFFs.
- `nsink_load()`: Essentially the inverse of the `nsink_build()` function, this function takes a folder of files, likely created by `nsink_build()`, and reads them into R.

The R package documentation contains both a detailed description of each function and a vignette that outlines a typical workflow for running an N-Sink analysis with the `nsink` package. Upon install, the vignette is accessed in R with `vignette("intro", package = "nsink")`.

Acknowledgements

Many people have contributed in various ways to the development of the N-Sink concept. In particular, we would like to thank, Chet Arnold, Cary Chadwick, David Dickson, and Emily Wilson of the University of Connecticut's Center for Land Use Education and Research as well as Peter August, Chris Damon, and Art Gold of the University of Rhode Island's Department of Natural Resources Science. Both the UCONN and URI crews have contributed tremendously to the development of the N-Sink concept. Additionally, we are grateful to X X, X X, X X, Joe LiVolsi, Tim Gleason, and Wayne Munns from the US EPA, Atlantic Coastal Environmental Sciences Division for constructive reviews of this paper. The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. Any mention of trade names, products, or services does not imply an endorsement by the U.S. Government or the U.S. Environmental Protection Agency. The EPA does not endorse any commercial products, services, or enterprises. This contribution is identified by the tracking number ORD-XXXXX of the Atlantic Coastal Environmental Sciences Division, Office of Research and Development, Center for Environmental Measurement and Modeling, US Environmental Protection Agency.

References

Bengtsson, Henrik. 2020. *Future: Unified Parallel and Distributed Processing in r for Everyone*. https://CRAN.R-project.org/package=future.

———. 2021. "A Unifying Framework for Parallel and Distributed Processing in R using Futures." *The R Journal*. https://doi.org/10.32614/RJ-2021-048.

Bivand, Roger S., Edzer Pebesma, and Virgilio Gomez-Rubio. 2013. *Applied Spatial Data Analysis with R, Second Edition*. Springer, NY. https://doi.org/10.1007/978-1-4614-7618-4.

Bocinsky, R. Kyle. 2020. *FedData: Functions to Automate Downloading Geospatial Data Available from Several Federated Data Sources*. https://CRAN.R-project.org/package=FedData.

Csardi, Gabor, and Tamas Nepusz. 2006. "The Igraph Software Package for Complex Network Research." *InterJournal* Complex Systems: 1695. https://igraph.org.

Csardi, Gabor, Tamas Nepusz, Szabolcs Horvat, Vincent Traag, and Fabio Zanini. 2020. *Network Analysis and Visualization*. https://CRAN.R-project.org/package=zoo.

Gräler, Benedikt, Edzer Pebesma, and Gerard Heuvelink. 2016. "Spatio-Temporal Interpolation Using Gstat." *The R Journal* 8: 204–18. https://doi.org/10.32614/RJ-2016-014.

Henry, Lionel, and Hadley Wickham. 2021. *Rlang: Functions for Base Types and Core r and 'Tidyverse' Features*. https://CRAN.R-project.org/package=rlang.

Hijmans, Robert J. 2021. *Raster: Geographic Data Analysis and Modeling*. https://CRAN.R-project.org/package=raster.

Kellogg, DQ, Arthur J Gold, Suzanne Cox, Kelly Addy, and Peter V August. 2010. "A Geospatial Approach for Assessing Denitrification Sinks Within Lower-Order Catchments." *Ecological Engineering* 36 (11): 1596–606. https://doi.org/10.1016/j.ecoleng.2010.02.006.

Pebesma, Edzer. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal* 10 (1): 439–46. https://doi.org/10.32614/RJ-2018-009.

———. 2021a. *Lwgeom: Bindings to Selected 'Liblwgeom' Functions for Simple Features*. https://CRAN.R-project.org/package=lwgeom.

———. 2021b. *Simple Features for r.* https://CRAN.R-project.org/package=sf.

———. 2021c. *Stars: Spatiotemporal Arrays, Raster and Vector Data Cubes.* https://CRAN.R-project.org/package=stars.

Pebesma, Edzer J. 2004. "Multivariable Geostatistics in S: The Gstat Package." *Computers & Geosciences* 30: 683–91. https://doi.org/10.1016/j.cageo.2004.03.012.

Pebesma, Edzer J., and Roger S. Bivand. 2005. "Classes and Methods for Spatial Data in R." *R News* 5 (2): 9–13. https://CRAN.R-project.org/doc/Rnews/.

Pebesma, Edzer, and Roger Bivand. 2021. *Sp: Classes and Methods for Spatial Data.* https://CRAN.R-project.org/package=sp.

Pebesma, Edzer, and Benedikt Graeler. 2021. *Gstat: Spatial and Spatio-Temporal Geostatistical Modelling, Prediction and Simulation.* https://CRAN.R-project.org/package=gstat.

Pebesma, Edzer, Thomas Mailund, and James Hiebert. 2016. "Measurement Units in R." *R Journal* 8 (2): 486–94. https://doi.org/10.32614/RJ-2016-061.

Pebesma, Edzer, Thomas Mailund, Tomasz Kalinowski, and Iñaki Ucar. 2021. *Units: Spatiotemporal Arrays, Raster and Vector Data Cubes.* https://CRAN.R-project.org/package=units.

R Core Team. 2020. *Foreign: Read Data Stored by 'Minitab', 's', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ...* https://CRAN.R-project.org/package=foreign.

Rabalais, Nancy N, R Eugene Turner, and Donald Scavia. 2002. "Beyond Science into Policy: Gulf of Mexico Hypoxia and the Mississippi River: Nutrient Policy Development for the Mississippi River Watershed Reflects the Accumulated Scientific Evidence That the Increase in Nitrogen Loading Is the Primary Factor in the Worsening of Hypoxia in the Northern Gulf of Mexico." *BioScience* 52 (2): 129–42. https://doi.org/10.1641/0006-3568(2002)052%5B0129:BSIPGO%5D2.0.CO;2.

Ross, Noam. 2020. *Fasterize: Fast Polygon to Raster Conversion.* https://CRAN.R-project.org/package=fasterize.

Vaughan, Davis, and Matt Dancho. 2021. *Furrr: Apply Mapping Functions in Parallel Using Futures.* https://CRAN.R-project.org/package=furrr.

Wickham, Hadley. 2020. *Httr: Tools for Working with URLs and HTTP.* https://CRAN.R-project.org/package=httr.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, and Jim Hester. 2020. *Readr: Read Rectangular Text Data.* https://CRAN.R-project.org/package=readr.

Zeileis, Achim, Gabor Gorthendieck, and Jeffrey A. Ryan. 2021. *Zoo: S3 Infrastructure for Regular and Irregular Time Series (Z's Ordered Observations).* https://CRAN.R-project.org/package=zoo.

Zeileis, Achim, and Gabor Grothendieck. 2005. "Zoo: S3 Infrastructure for Regular and Irregular Time Series." *Journal of Statistical Software* 14 (6): 1–27. https://doi.org/10.18637/jss.v014.i06.