The Journal of Open Source Software

gamut: A Geospatial R Package to Analyze Multisectoral Urban Teleconnections

Kristian D. Nelson[1], **Sean W. Turner**[1], **Chris R. Vernon**[1], **and Jennie S. Rice**[1]

1 Pacific Northwest National Laboratory, Richland, Washington, USA

DOI: 10.21105/joss.03383

Software

- Review
- Repository
- Archive

Editor: Kristina Riemer

Reviewers:

- @fsanchez-trigueros
- @jsta

Submitted: 05 May 2021
Published: 21 October 2021

Summary

Most cities in the United States withdraw surface water to meet public water supply needs. The lands on which this water is generated are often developed for human activities—such as agriculture, mining, and industry—that may compete for water resources or contaminate water supplies. Cities are thereby connected to other sectors through their water supply catchments. These connections are an example of a multisectoral urban teleconnection, or an interdependency to a geographically disparate region from a source region where events in one (e.g., land use changes) often impact the other (Seto et al., 2012). This term was brought about to bring greater understanding of the connections between urbanization and land use changes (Seto et al., 2012). The Geospatial Analytics for Multisectoral Urban Teleconnections (`gamut`) package provides national-scale information on these urban teleconnections for 235 cities by combining land use data with hydrological analysis to characterize and quantify urban source watershed human interactions across the conterminous United States (Figure 1).

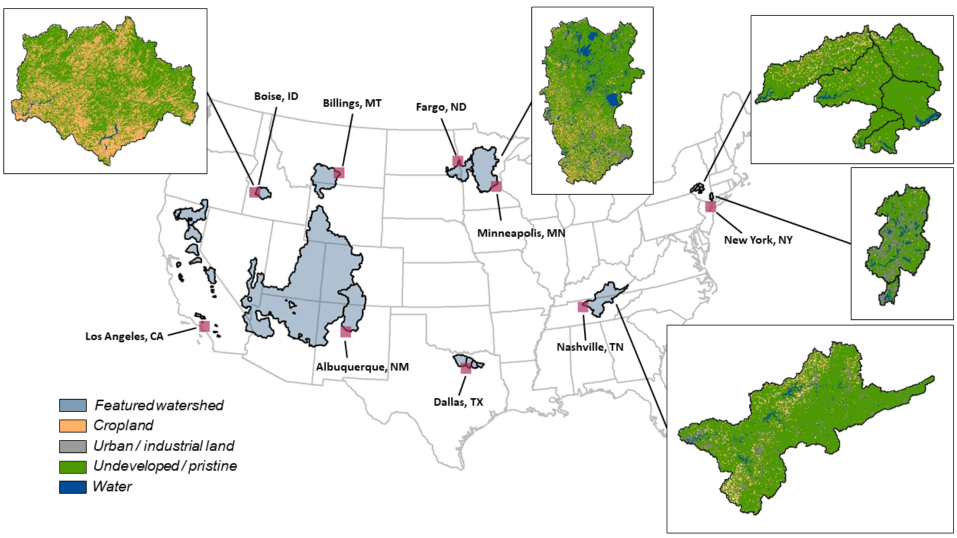

Figure 1: The `gamut` package analyzes urban cities and their watersheds across the conterminous U.S. As shown in the figure, it can look at characteristics like land use inside watershed boundaries.

The `gamut` package computes dozens of city-level metrics that inform on the geographical nature of surface water supply catchments and the presence, intensity, and impact of human activities in those catchments. The package cycles through a 3-step process. First, it connects cities to their drinking water resources by using the Urban Water Blueprint dataset (McDonald

& Shemie, 2014). This dataset is combined with an enhancment dataset of source contribution estimates, river flow statistics, and high resolution runoff data(Nelson, Turner, Vernon, Rice, & Kao., 2021). After linking the cities to the watersheds, the second step is using the watershed boundary as a mask for the input geospatial layers. Additionally, `gamut` relies heavily on the use of the `st_intersection` function from the sf package to intersect features and watersheds, which joins the information from the geospatial layers. These layers encompass a wide variety of watershed characteristics including land use, land cover, power generation, infrastructure, hydrological data (irrigation, stream flow, runoff), and population data. Similar to other multisector packages, `gamut` produces new data from a combination of multiple datasets, which are then used to create new statistics. Non-geospatial layers in the form of data tables are joined through city names and municipal IDs. The input layers used in this package have been combined into an open-source dataset and can be accessed here (Nelson, Turner, Vernon, & Rice, 2021). The last step in the process is the creation of output metrics, which is done through the calculation of statistics based on the masked input data for each city.

Creating the links between cities and watershed characteristics enables the `gamut` package to calculate numerous metrics that may be used for multiple types of city-level multisector dynamics research. Metrics reported by `gamut` fall into four main categories: geographical characteristics of watersheds (e.g., climate zones, land area, distance from city, hydrology), potential water contamination concentrations (nonpoint and point), withdrawal/consumption of water from other sectors, and presence/intensity of multisectoral land uses. Table 1 shows the metrics that are created and includes descriptions and units. An R vignette is provided to help users to get started with `gamut` and may be accessed here.

Table 1: Metrics reported in `gamut`

Metric Name	Description	Units
city_population	The population of the city being analyzed	people
n_watersheds	Number of watersheds that city uses to source drinking water	watersheds
n_other_cities	Number of other cities pulling off the same watersheds	cities
dependent_city_pop	Total population of people dependent on that city's watersheds	people
watershed_area_sqkm	Combined area of all the source watersheds of a city	square kilometers
storage_BCM	Combined storage capacity of all the city catchments	billion cubic meters
yield_BCM	Combined yield capacity of all the city catchments	billion cubic meters
irr_cons_BCM	Combined water consumption that is used for irrigation with the watersheds	billion cubic meters
n_climate_zones	Number of climate zones that the source watersheds cover	zones
n_hydro_plants	Number of hydroelectric power plants operating within the source watersheds	plants
n_thermal_plants	Number of thermal power plants operating within the source watersheds	plants
n_fac_agcrop	Number of agricultural crop facilities within the source watersheds	facilities
n_fac_aglivestock	Number of agicultural livestock facilities within the source watersheds	facilities
n_fac_cnsmnf	Number of construction and manufacturing facilities within the source watersheds	facilities
n_fac_mining	Number of mining facilities within the source watersheds	facilities

Metric Name	Description	Units
n_fac_oilgas	Number of oil and gas facilities within the source watersheds	facilities
n_fac_total	Total number of facilities operating within the source watersheds	facilities
hydro_gen_MWh	Combined hydroelectric generation from all the facilities within the source watersheds	megawatt-hours
thermal_gen_MWh	Combined thermal generation from all the facilities within the source watersheds	megawatt-hours
thermal_cons_BCM	Combined water consumption that is used for thermal generation	billion cubic meters
thermal_with_BCM	Combined water withdrawal for thermal generation	billion cubic meters
n_utilities	Number of electric utilities within the source watersheds	utilities
n_ba	Number of balancing authorities within the source watersheds	balancing authorities
n_crop_classes	Total number of different types of crops within the source watersheds	crops
cropland_fraction	Fraction of land that is used for crops	fraction
developed_fraction	Fraction of land that is developed	fraction
ag_runoff_max	Agricultural runoff as proportion of total runoff (worst-case watershed)	fraction
ag_runoff_av_exgw	Agricultural runoff as proportion of total runoff in supply (exc. groundwater)	fraction
ag_runoff_av	Agricultural runoff as proportion of total runoff in supply (inc. groundwater)	fraction
dev_runof_max	Urban runoff as proportion of total runoff (worst-case watershed)	fraction
dev_runof_av_exgw	Urban runoff as proportion of total runoff in supply (exc. groundwater)	fraction
dev_runof_av	Urban runoff as proportion of total runoff in supply (inc. groundwater)	fraction
np_runoff_max	Max amount of non-point source runoff within the source watersheds	fraction
np_runoff_av_exgw	Nonpoint Proportion of Potentially Contaminated Supply (PPCS) (exc. groundwater)	fraction
np_runoff_av_exgw_unweighted	Nonpoint supply contamination averaged across watersheds	fraction
np_runoff_av	Nonpoint Proportion of Potentially Contaminated Supply (PPCS)	fraction
n_economic_sectors	Total number of different economic sectors within the source watersheds	sectors
max_withdr_dis_km	Maximum distance between a city's intake points	kilometers
avg_withdr_dis_km	Average distance between a city's intake points	kilometers
n_treatment_plants	Total number of waste water treatment plants operating within the source watersheds	plants
watershed_pop	Total number of people living within the source watershed boundaries	people
pop_cons_m3sec	Combined water consumption from the source watersheds that is used for people	m3/sec
av_fl_sur_conc_pct	Point PPCS (surface water only, based on flow)	%
av_fl_sur_conc_pct_unweighted	Point PPCS (surface water only, based on flow, not weighted by source importance)	%

Metric Name	Description	Units
av_ro_sur_conc_pct	Point PPCS (surface water only, based on runoff)	%
av_fl_all_conc_pct	Point PPCS (based on flow)	%
av_ro_all_conc_pct	Point PPCS (based on runoff)	%
av_fl_max_conc_pct	Point PPCS (based on flow, worst-case catchment only)	%
av_ro_max_conc_pct	Point PPCS (based on runoff, worst-case catchment only)	%
surface_contribution_pct	Proportion of total average supply made up from surface water	%
importance_of_worst_watershed_pct	Proportion of total average supply made up from most heavily contamined watershed	%

Statement of Need

Multisector Dynamics (MSD) research is the study of the co-evolution of human and natural systems. This research requires infrastructure expansion and land use scenarios, resource demand projections, and multisectoral modeling to capture the impacts of trends and shocks on human systems. The `gamut` package offers new data that meet a number of MSD needs. The package may be used to infer possible water resources expansion strategies for major cities in the United States. For example, cities found to be heavily exposed to potential contamination may be more likely to seek alternative means of supply (e.g., water transfers) or invest in water reuse facilities. In a study by Rice et al. (2013) which looked at de facto wastewater reuse across the US, it was found that there had been an increase in wastewater concentrations in drinking water treatment plants from a 1980 EPA report, especially at low flow conditions. The `gamut` package has the ability to look at wastewater discharge and average flow to find these concentrations at a much larger scale, showing that this package could be useful in studies like this in the future.

In addition to water contamination analysis, the `gamut` package has the ability to reveal which source watersheds are heavily protected by receiving cites. This information can inform land use and energy expansion scenarios applied in MSD research, for example by preventing significant expansion of human developments in protected source watersheds. `gamut` may also be used in large-scale hydrological modeling to correctly assign urban water demands to specific intakes. Whether research is being done on water scarcity, water pollution, or urbanization effects, the `gamut` package provides useful data that can brings greater understanding of anthropogenic impacts on urban source watersheds.

The `gamut` package is open source and may be downloaded using the devtools package with the code below (Wickham, Hester, et al., 2020). Further instructions on package download can be found in the documentation.

```
install.packages("devtools")
library(devtools)
devtools::install_github('IMMM-SFA/gamut')
library(gamut)
```

Dependencies

`gamut` relies on functionality from the following R packages: clisymbols (Csárdi & Sorhus, 2017), crayon (Csárdi, 2017), dplyr (Wickham, François, et al., 2020), dams (Goteti & Stachelek, 2020), exactextractr (Baston, 2020), foreign (R Core Team, 2020), geosphere

(Hijmans, 2019), ggplot2 (Wickham, 2016), lwgeom (Pebesma, 2020), magrittr (Bache & Wickham, 2014), purrr (Henry & Wickham, 2020), raster (Hijmans, 2020), readxl (Wickham & Bryan, 2019), reservoir (Turner & Galelli, 2016), rgdal (R. Bivand et al., 2020), rgeos (R. Bivand & Rundel, 2020), sf (Pebesma, 2018), sp (R. S. Bivand et al., 2013), spex (Sumner, 2020), stringr (Wickham, 2019), tibble (Müller & Wickham, 2020), tidyr (Wickham & Henry, 2020), vroom (Hester & Wickham, 2021), testthat (Wickham, 2011), rmarkdown (Xie et al., 2018), knitr (Xie, 2018).

Acknowledgements

This research was supported by the U.S. Department of Energy, Office of Science, as part of research in MultisectorDynamics, Earth and Environmental System Modeling Program.

References

Bache, S. M., & Wickham, H. (2014). *magrittr: A Forward-Pipe Operator for R*. https://CRAN.R-project.org/package=magrittr

Baston, D. (2020). *exactextractr: Fast Extraction from Raster Datasets using Polygons*. https://CRAN.R-project.org/package=exactextractr

Bivand, R., Keitt, T., & Rowlingson, B. (2020). *rgdal: Bindings for the 'Geospatial' Data Abstraction Library*. https://CRAN.R-project.org/package=rgdal

Bivand, R., & Rundel, C. (2020). *rgeos: Interface to Geometry Engine - Open Source ('GEOS')*. https://CRAN.R-project.org/package=rgeos

Bivand, R. S., Pebesma, E., & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R, Second edition*. Springer, NY. https://asdar-book.org/

Csárdi, G. (2017). *crayon: Colored Terminal Output*. https://CRAN.R-project.org/package=crayon

Csárdi, G., & Sorhus, S. (2017). *clisymbols: Unicode Symbols at the R Prompt*. https://CRAN.R-project.org/package=clisymbols

Goteti, G., & Stachelek, J. (2020). *dams: Dams in the United States from the National Inventory of Dams (NID)*. https://CRAN.R-project.org/package=dams

Henry, L., & Wickham, H. (2020). *purrr: Functional Programming Tools*. https://CRAN.R-project.org/package=purrr

Hester, J., & Wickham, H. (2021). *vroom: Read and Write Rectangular Text Data Quickly*. https://CRAN.R-project.org/package=vroom

Hijmans, R. J. (2019). *geosphere: Spherical Trigonometry*. https://CRAN.R-project.org/package=geosphere

Hijmans, R. J. (2020). *raster: Geographic Data Analysis and Modeling*. https://CRAN.R-project.org/package=raster

McDonald, R. I., & Shemie, D. (2014). *Urban Water Blueprint: Mapping conservation solutions to the global water challenge*. The Nature Conservancy. https://www.nature.org/content/dam/tnc/nature/en/documents/Urban_Water_Blueprint.pdf

Müller, K., & Wickham, H. (2020). *tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble

Nelson, K., Turner, S., Vernon, C., & Rice, J. (2021). *Geospatial Input Datasets for Urban Teleconnections Analysis*. Zenodo. http://doi.org/10.5281/zenodo.4662993

Nelson, K., Turner, S., Vernon, C., Rice, J., & Kao., S.-C. (2021). *UWSCatCH: Urban Water Supply Catchment Contributions and Hydrological Statistics for large cities of the conterminous United States*. Zenodo. http://doi.org/10.5281/zenodo.4315195

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, *10*(1), 439–446. https://doi.org/10.32614/RJ-2018-009

Pebesma, E. (2020). *lwgeom: Bindings to Selected 'liblwgeom' Functions for Simple Features*. https://CRAN.R-project.org/package=lwgeom

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

Rice, J., Wutich, A., & Westerhoff, P. (2013). Assessment of De Facto Wastewater Reuse across the U.S.: Trends between 1980 and 2008. *Environmental Science & Technology*, *47*(19), 11099–11105. https://doi.org/10.1021/es402792s

Seto, K., Reenberg, A., Boone, C., Fragkias, M., Haase, D., Langanke, T., Marcotullio, P., Munroe, D., Olah, B., & Simon, D. (2012). Urban Land Teleconnections and Sustainability. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 7687–7692. https://doi.org/10.1073/pnas.1117622109

Sumner, M. D. (2020). *spex: Spatial Extent Tools*. https://CRAN.R-project.org/package=spex

Turner, S. W. D., & Galelli, S. (2016). Water supply sensitivity to climate change: An R package for implementing reservoir storage analysis in global and regional impact studies. *Environmental Modelling & Software*, *76*, 13–19. https://doi.org/10.1016/j.envsoft.2015.11.007

Wickham, H. (2011). testthat: Get Started with Testing. *The R Journal*, *3*(1), 5–10. https://doi.org/10.32614/RJ-2011-002

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4

Wickham, H. (2019). *stringr: Simple, Consistent Wrappers for Common String Operations*. https://CRAN.R-project.org/package=stringr

Wickham, H., & Bryan, J. (2019). *readxl: Read Excel Files*. https://CRAN.R-project.org/package=readxl

Wickham, H., François, R., Henry, L., & Müller, K. (2020). *dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2020). *tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr

Wickham, H., Hester, J., & Chang, W. (2020). *devtools: Tools to Make Developing R Packages Easier*. https://CRAN.R-project.org/package=devtools

Xie, Y. (2018). knitr: a comprehensive tool for reproducible research in R. In *Implementing reproducible research* (pp. 3–31). Chapman; Hall/CRC. https://doi.org/10.1201/9781315373461-1

Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R Markdown: The Definitive Guide*. Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown