GSODR: Global Summary Daily Weather Data in R

Adam H Sparks[1]**, Tomislav Hengl**[2]**, and Andrew Nelson**[3]

1 Centre for Crop Health, University of Southern Queensland, Toowoomba Queensland 4350, Australia **2** ISRIC - World Soil Information, P.O. Box 353, 6700 AJ Wageningen, The Netherlands **3** Faculty of Geo-Information and Earth Observation (ITC), University of Twente, Enschede 7500 AE, The Netherlands

Summary

The GSODR package (Sparks, Hengl, and Nelson 2017) is an R package (R Core Team 2016) providing automated downloading, parsing and cleaning of Global Surface Summary of the Day (GSOD) (United States National Oceanic and Atmospheric Administration National Climatic Data Center 2016) weather data for use in R or saving as local files in either a Comma Separated Values (CSV) or GeoPackage (GPKG) (Open Geospatial Consortium 2014) file. It builds on or complements several other scripts and packages. We take advantage of modern techniques in R to make more efficient use of available computing resources to complete the process, e.g., data.table (Dowle et al. 2015), plyr (Wickham 2011) and readr (Wickham, Hester, and Francois 2016), which allow the data cleaning, conversions and disk input/output processes to function quickly and efficiently. The rnoaa (Chamberlain 2016) package already offers an excellent suite of tools for interacting with and downloading weather data from the United States National Oceanic and Atmospheric Administration, but lacks options for GSOD data retrieval. Several other APIs and R packages exist to access weather data, but most are region or continent specific, whereas GSOD is global. This package was developed to provide:

- two functions that simplify downloading GSOD data and formatting it to easily be used in research; and

- a function to help identify stations within a given radius of a point of interest.

Alternative elevation data based on a 200 meter buffer of elevation values derived from the CGIAR-CSI SRTM 90m Database (Jarvis et al. 2008) are included. These data are useful to help address possible inaccuracies and in many cases, fill in for missing elevation values in the reported station elevations.

When using this package, GSOD stations are checked for inaccurate longitude and latitude values and any stations that have missing or have incorrect values are omitted from the final data set. Users may set a threshold for station files with too many missing observations for omission from the final output to help ensure data quality. All units are converted from the United States Customary System (USCS) to the International System of Units (SI), e.g., inches to millimetres and Fahrenheit to Celsius. Wind speed is also converted from knots to metres per second. Additional useful values, actual vapour pressure, saturated water vapour pressure, and relative humidity are calculated and included in the final output. Station metadata are merged with weather data for the final data set.

References

Chamberlain, Scott. 2016. *Rnoaa: 'NOAA' Weather Data from R*. https://CRAN.R-project.org/package=rnoaa.

Dowle, M, A Srinivasan, T Short, S Lianoglou with contributions from R Saporta, and E Antonyan. 2015. *Data.table: Extension of Data.frame*. https://CRAN.R-project.org/package=data.table.

Jarvis, Andy, Hannes I Reuter, Andy Nelson, and Edward Guevara. 2008. "Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database." http://srtm.csi.cgiar.org.

Open Geospatial Consortium. 2014. "GeoPackage Encoding Standard." http://www.opengeospatial.org/standards/geopackage.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Sparks, Adam, Tomislav Hengl, and Andrew Nelson. 2017. *GSODR: Global Summary Daily Weather Data in R*. http://ropensci.github.io/GSODR/.

United States National Oceanic and Atmospheric Administration National Climatic Data Center. 2016. "Global Surface Summary of Day (GSOD)." https://data.noaa.gov/dataset/global-surface-summary-of-the-day-gsod.

Wickham, Hadley. 2011. "The Split-Apply-Combine Strategy for Data Analysis." *Journal of Statistical Software* 40 (1): 1–29. http://www.jstatsoft.org/v40/i01/.

Wickham, Hadley, Jim Hester, and Romain Francois. 2016. *Readr: Read Tabular Data*. https://CRAN.R-project.org/package=readr.