

# Package ‘GSODR’

January 19, 2019

**Type** Package

**Title** Global Surface Summary of the Day ('GSOD') Weather Data Client

**Version** 1.3.2

**URL** <https://github.com/ropensci/GSODR>,  
<https://ropensci.github.io/GSODR/>

**BugReports** <https://github.com/ropensci/GSODR/issues>

**Description** Provides automated downloading, parsing, cleaning, unit conversion and formatting of Global Surface Summary of the Day ('GSOD') weather data from the from the USA National Centers for Environmental Information ('NCEI') for use in R. Units are converted from from United States Customary System ('USCS') units to International System of Units ('SI'). Stations may be individually checked for number of missing days defined by the user, where stations with too many missing observations are omitted. Only stations with valid reported latitude and longitude values are permitted in the final data. Additional useful elements, saturation vapour pressure ('es'), actual vapour pressure ('ea') and relative humidity are calculated from the original data and included in the final data set. The resulting data include station identification information, state, country, latitude, longitude, elevation, weather observations and associated flags. Additional data are included with this R package: a list of elevation values for stations between -60 and 60 degrees latitude derived from the Shuttle Radar Topography Measuring Mission ('SRTM'). For information on the 'GSOD' data from 'NCEI', please see the 'GSOD' 'readme.txt' file available from, <<http://www1.ncdc.noaa.gov/pub/data/g sod/readme.txt>>.

**Depends** R (>= 3.3.0)

**License** MIT + file LICENSE

**Imports** curl, dplyr (>= 0.7.0), future.apply, magrittr, purrr (>= 0.2.0), R.utils, readr, rlang, stats, tibble, utils

**Suggests** covr, future, ggplot2, ggthemes, gridExtra, knitr, lubridate, mapproj, maps, plotKML, raster, reshape2, rgdal (>= 1.1-9), rgeos, rmarkdown, roxygen2 (>= 6.1.0), sf, sp, spacetime, testthat, tidy

**RoxygenNote** 6.1.1

**Encoding** UTF-8

**NeedsCompilation** no

**Repository** CRAN

**LazyData** FALSE

**LazyLoad** FALSE

**ByteCompile** TRUE

**VignetteBuilder** knitr

**X-schema.org-applicationCategory** Tools

**X-schema.org-keywords** US-NCEI, meteorological-data, global-weather, weather, weather-data, meteorology, station-data, surface-weather, data-access, US-NCDC

**X-schema.org-isPartOf** <https://ropensci.org>

**Author** Adam Sparks [aut, cre] (<<https://orcid.org/0000-0002-0061-8359>>), Tomislav Hengl [aut] (<<https://orcid.org/0000-0002-9921-5129>>), Andrew Nelson [aut] (<<https://orcid.org/0000-0002-7249-3778>>), Hugh Parsonage [cph, ctb] (<<https://orcid.org/0000-0003-4055-0835>>), Bob Rudis [cph, ctb] (<<https://orcid.org/0000-0001-5670-2640>>), Gwenael Giboire [ctb] (Several bug reports in early versions and testing feedback), Łukasz Pawlik [ctb] (Reported bug in windspeed conversion calculation), Ross Darnell [ctb] (Reported bug in 'Windows OS' versions causing 'GSOD' data untarring to fail)

**Maintainer** Adam Sparks <[adamspark@gmail.com](mailto:adamspark@gmail.com)>

**Date/Publication** 2019-01-18 23:10:43 UTC

## R topics documented:

get_GSOD . . . . .	3
get_inventory . . . . .	5
GSODR . . . . .	6
nearest_stations . . . . .	7
print.GSODR.Info . . . . .	8
reformat_GSOD . . . . .	9
update_station_list . . . . .	10

<b>Index</b>	<b>12</b>
--------------	-----------

---

`get_GSOD`*Download and Return a Tidy Data Frame of GSOD Weather Data*

---

### Description

This function automates downloading, cleaning, reformatting of data from the Global Surface Summary of the Day (GSOD) data provided by the [US National Centers for Environmental Information \(NCEI\)](#). Three additional useful elements: saturation vapour pressure (es), actual vapour pressure (ea) and relative humidity (RH) are calculated and returned in the final data frame.

Parallel processing can be enabled using [plan](#) to set up a parallel backend of your choice, e.g., `future::plan("multisession")`. See examples for more.

### Usage

```
get_GSOD(years, station = NULL, country = NULL, max_missing = NULL,  
          agroclimatology = FALSE)
```

### Arguments

<code>years</code>	Year(s) of weather data to download.
<code>station</code>	Optional. Specify a station or multiple stations for which to retrieve, check and clean weather data using <i>STNID</i> . The NCEI reports years for which the data are available. This function checks against these years. However, not all cases are properly documented and in some cases files may not exist on the FTP server even though it is indicated that data was recorded for the station for a particular year. If a station is specified that does not have an existing file on the server, this function will silently fail and move on to existing files for download and cleaning from the FTP server.
<code>country</code>	Optional. Specify a country for which to retrieve weather data; full name or ISO codes can be used.
<code>max_missing</code>	Optional. The maximum number of days allowed to be missing from a station's data before it is excluded from final file output.
<code>agroclimatology</code>	Optional. Logical. Only clean data for stations between latitudes 60 and -60 for agroclimatology work, defaults to FALSE. Set to TRUE to include only stations within the confines of these latitudes.

### Details

Stations reporting a latitude of  $< -90$  or  $> 90$  or longitude of  $< -180$  or  $> 180$  are removed. Stations may be individually checked for number of missing days to assure data quality and omitting stations with too many missing observations.

All units are converted to International System of Units (SI), e.g. Fahrenheit to Celsius and inches to millimetres.

Alternative elevation measurements are supplied for missing values or values found to be questionable based on the Consultative Group for International Agricultural Research's Consortium for Spatial Information group's (CGIAR-CSI) Shuttle Radar Topography Mission 90 metre (SRTM 90m) digital elevation data based on NASA's original SRTM 90m data. Further information on these data and methods can be found on **GSODR's GitHub repository**.

Data summarise each year by station, which include vapour pressure and relative humidity elements calculated from existing data in GSOD.

All missing values in resulting files are represented as NA regardless of which field they occur in.

For a complete list of the fields and description of the contents and units, please refer to Appendix 1 in the **GSODR** vignette, `vignette("GSODR", package = "GSODR")`.

For more information see the description of the data provided by NCEI, [http://www7.ncdc.noaa.gov/CDO/GSOD\\_DESC.txt](http://www7.ncdc.noaa.gov/CDO/GSOD_DESC.txt).

## Value

A data frame as a `tibble` object of weather data.

## Note

While **GSODR** does not distribute GSOD weather data, users of the data should note the conditions that the U.S. NCEI places upon the GSOD data. "The following data and products may have conditions placed on their international commercial use. They can be used within the U.S. or for non-commercial international activities without restriction. The non-U.S. data cannot be redistributed for commercial purposes. Re-distribution of these data by others must provide this same notification."

## Author(s)

Adam H Sparks, <adamhsparks@gmail.com>

## References

Jarvis, A., Reuter, H. I, Nelson, A., Guevara, E. (2008) Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database <http://srtm.csi.cgiar.org>

## See Also

[reformat\\_GSOD](#)

## Examples

```
# Download weather station for Toowoomba, Queensland for 2010
tbar <- get_GSOD(years = 2010, station = "955510-99999")

tbar

# Download data for Australia from 2010 to 2015
AUS <- get_GSOD(years = 2010:2015, country = "Australia")
```

```
AUS

# Download agroclimatology data for 2015 using parallel processing
future::plan("multisession")
ag <- get_GSOD(years = 2015, agroclimatology = TRUE)

ag

# Download global data for 2010 to 2015 with a maximum allowed 5 missing days
# of data using parallel processing

future::plan("multisession")
global <- get_GSOD(years = 2010:2015, max_missing = 5)

global
```

---

get_inventory	<i>Download and Return a Tidy Data Frame of GSOD Weather Station Data Inventories</i>
---------------	---

---

## Description

The NCEI maintains a document, <ftp://ftp.ncdc.noaa.gov/pub/data/noaa/isd-inventory.txt>, which lists the number of weather observations by station-year-month from the beginning of the stations' records. This function retrieves that document and prints an information header displaying the last update time with a data frame of the inventory information for each station-year-month.

## Usage

```
get_inventory()
```

## Value

A data frame as a [tibble](#) object of station inventories

## Note

The GSOD data, which are downloaded and manipulated by **GSODR**, stipulate that the following notice should be given. "The following data and products may have conditions placed on their international commercial use. They can be used within the U.S. or for non-commercial international activities without restriction. The non-U.S. data cannot be redistributed for commercial purposes. Re-distribution of these data by others must provide this same notification."

The download process can take quite some time to complete.

**Author(s)**

Adam H Sparks, <adamhsparks@gmail.com>

**Examples**

```
inventory <- get_inventory()
inventory
```

---

GSODR

*Global Surface Summary of the Day (GSOD) Weather Data Client*

---

**Description**

Provides automated downloading, parsing, cleaning, unit conversion and formatting of Global Surface Summary of the Day (GSOD) weather data from the from the USA's National Centre for Environmental Information (NCEI) for use in **R**. All units are converted from United States Customary System (USCS) units to International System of Units (SI). Stations may be individually checked for number of missing days defined by the user, where stations with too many missing observations are omitted. Only stations with valid reported latitude and longitude values are permitted in the final data. Additional useful elements, saturation vapour pressure ('es'), actual vapour pressure ('ea') and relative humidity (RH) are calculated from the original data and included in the final data set. The resulting data include station identification information, state, country, latitude, longitude, elevation, weather observations and associated flags.

**Details**

Additional data are included: a list of elevation values for stations between -60 and 60 degrees latitude derived from the Shuttle Radar Topography Measuring Mission (SRTM).

For information on the GSOD data from NCEI, please see the GSOD readme.txt file available from, <http://www1.ncdc.noaa.gov/pub/data/ggod/readme.txt>.

For climate data that have been formatted specifically for use with the **GSODR** package, please see the **GSODRdata** package (Sparks *et al.*) available from GitHub: <https://adamhsparks.github.io/GSODRdata/>. Four data frames of climate data are provided from various sources for GSOD station locations.

**Author(s)**

Adam Sparks, Tomislav Hengl and Andrew Nelson

**Source**

<https://data.noaa.gov/dataset/dataset/global-surface-summary-of-the-day-gsod/>

## References

- Karger, D. N., Conrad, O., Bohner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., *et al.* (2016) Climatologies at high resolution for the Earth land surface areas. *arXiv preprint arXiv:1607.00217*
- New, M., Lister, D., Hulme, M., Makin, I., (2002) A high-resolution data set of surface climate over global land areas. *Climate Research* **21**:1–25
- Sparks A., Hengl T., Nelson A. (2018) GSODRdata: Extra Climate Data for the GSODR Package. R package version 0.2.10, <https://adamhsparks.github.io/GSODRdata/index.html>.
- Wilson A. M., Jetz W. (2016) Remotely Sensed High-Resolution Global Cloud Dynamics for Predicting Ecosystem and Biodiversity Distributions. *PLoS Biol* **14**(3): e1002415. doi:10.1371/journal.pbio.1002415

## See Also

### GSODR functions:

[get\\_GSOD](#) Download, Clean, Reformat Generate New Elements and Return a Tidy Data Frame of GSOD Weather Data

[reformat\\_GSOD](#) Clean, Reformat Generate New Elements and Return a Tidy Data Frame of GSOD Weather Data from Local Disk

[nearest\\_stations](#) Find Nearest GSOD Stations to Specified a Latitude and Longitude

[update\\_station\\_list](#) Download the Latest Station List Information and Update GSODR's Internal Database

[get\\_inventory](#) Download and return a tidy data frame of GSOD weather station data inventories

### Useful links:

- Static documentation at <https://ropensci.github.io/GSODR/>
- Development repository at <https://github.com/ropensci/GSODR>
- Report bugs at <https://github.com/ropensci/GSODR/issues>

---

nearest\_stations

*Find Nearest GSOD Stations to a Specified Latitude and Longitude*

---

## Description

Given latitude and longitude values entered as decimal degrees (DD), this function returns a list (atomic vector) of STNID values, which can be used in [get\\_GSOD](#) to query for specific stations as an argument in the `station` parameter of that function.

## Usage

```
nearest_stations(LAT, LON, distance)
```

**Arguments**

LAT	Latitude expressed as decimal degrees (DD) (WGS84)
LON	Longitude expressed as decimal degrees (DD) (WGS84)
distance	Distance in kilometres from point for which stations are to be returned.

**Value**

`vector` object of station identification numbers

**Note**

The GSOD data, which are downloaded and manipulated by **GSODR** stipulate that the following notice should be given. “The following data and products may have conditions placed on their international commercial use. They can be used within the U.S. or for non- commercial international activities without restriction. The non-U.S. data cannot be redistributed for commercial purposes. Re-distribution of these data by others must provide this same notification.”

**Author(s)**

Adam H Sparks, <adamhsparks@gmail.com>

**Examples**

```
# Find stations within a 100km radius of Toowoomba, QLD, AUS
n <- nearest_stations(LAT = -27.5598, LON = 151.9507, distance = 100)
n
```

---

```
print.GSODR.Info      Prints GSODR.info object.
```

---

**Description**

Prints GSODR.info object.

**Usage**

```
## S3 method for class 'GSODR.Info'
print(x, ...)
```

**Arguments**

x	GSODR.info object
...	ignored



## Description

This function automates cleaning and reformatting of GSOD station files in "WMO-WBAN-YYYY.op.gz" format that have been downloaded from the United States National Center for Environmental Information's (NCEI) FTP server. Three new elements; saturation vapour pressure (es), actual vapour pressure (ea) and relative humidity are calculated and returned in the final data as well. All units are converted to International System of Units (SI), e.g. Fahrenheit to Celsius and inches to millimetres. Alternative elevation measurements are supplied for missing values or values found to be questionable based on the Consultative Group for International Agricultural Research's Consortium for Spatial Information group's (CGIAR-CSI) Shuttle Radar Topography Mission 90 metre (SRTM 90m) digital elevation data based on NASA's original SRTM 90m data.

## Usage

```
reformat_GSOD(dsn = NULL, file_list = NULL)
```

## Arguments

<code>dsn</code>	User supplied file path to location of data files on local disk for tidying.
<code>file_list</code>	User supplied list of files of data on local disk for tidying.

## Details

Parallel processing can be enabled using [plan](#) to set up a parallel backend of your choice, e.g., `future::plan("multisession")`. See examples for more.

If multiple stations are given, data are summarised for each year by station, which include vapour pressure and relative humidity elements calculated from existing data in GSOD. Else, single stations are tidied and a data frame is returned.

All missing values in resulting files are represented as NA regardless of which field they occur in.

Only station files in the original ".op.gz" file format are supported by this function. If you have downloaded the full annual "gsod\_YYYY.tar" file you will need to extract the individual station files from the tar file first to use this function.

For a complete list of the fields and description of the contents and units, please refer to Appendix 1 in the **GSODR** vignette, `vignette("GSODR", package = "GSODR")`.

## Value

A data frame as a [tibble](#) object of weather data.

**Note**

While **GSODR** does not distribute GSOD weather data, users of the data should note the conditions that the U.S. NCEI places upon the GSOD data. “The following data and products may have conditions placed on their international commercial use. They can be used within the U.S. or for non-commercial international activities without restriction. The non-U.S. data cannot be redistributed for commercial purposes. Re-distribution of these data by others must provide this same notification.”

**Author(s)**

Adam H Sparks, <adamhsparks@gmail.com>

**References**

Jarvis, A., Reuter, H.I, Nelson, A., Guevara, E. (2008) Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database <http://srtm.csi.cgiar.org>

**See Also**

For automated downloading and tidying see the [get\\_GSOD](#) function which provides expanded functionality for automatically downloading and expanding annual GSOD files and cleaning station files.

[get\\_GSOD](#)

**Examples**

```
# Download data to 'tempdir()'
download.file(url =
  "ftp://ftp.ncdc.noaa.gov/pub/data/gsod/2010/955510-99999-2010.op.gz",
  destfile = file.path(tempdir(), "955510-99999-2010.op.gz"),
  mode = "wb")

# Reformat station data files in R's tempdir() directory
tbar <- reformat_GSOD(dsn = tempdir())

tbar
```

**Description**

This function downloads the latest station list (isd-history.csv) from the NCEI FTP server and updates the data distributed with **GSODR** to the latest stations available. These data provide unique identifiers, country, state (if in U.S.), latitude, longitude, elevation and when weather observations begin and end. Stations with invalid latitude and longitude values will not be included.

**Usage**

```
update_station_list()
```

**Details**

Care should be taken when using this function if reproducibility is necessary as different machines with the same version of **GSODR** can end up with different versions of the isd\_history.csv file internally.

There is no need to use this unless you know that a station exists in the **GSODR** data that is not available in the self-contained database.

To directly access these data, use:

```
load(system.file("extdata", "isd_history.rda", package = "GSODR"))
```

**Author(s)**

Adam H Sparks, <adamhsparks@gmail.com>

**Examples**

```
## Not run:  
update_station_list()  
  
## End(Not run)
```

# Index

`get_GSOD`, [3](#), [7](#), [10](#)  
`get_inventory`, [5](#), [7](#)  
`GSODR`, [6](#)  
`GSODR-package (GSODR)`, [6](#)  
  
`nearest_stations`, [7](#), [7](#)  
  
`plan`, [3](#), [9](#)  
`print.GSODR.Info`, [8](#)  
  
`reformat_GSOD`, [4](#), [7](#), [9](#)  
  
`tibble`, [4](#), [5](#), [9](#)  
  
`update_station_list`, [7](#), [10](#)  
  
`vector`, [8](#)