

# Package ‘GSODR’

August 13, 2022

**Type** Package

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

**Version** 3.1.6

**URL** <https://docs.ropensci.org/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'). 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 ('RH') are calculated from the original data using the improved August-Roche-Magnus approximation (Alduchov & Eskridge 1996) and included in the final data set. The resulting metadata include station identification information, country, state, latitude, longitude, elevation, weather observations and associated flags. For information on the 'GSOD' data from 'NCEI', please see the 'GSOD' 'readme.txt' file available from, <https://www1.ncdc.noaa.gov/pub/data/gsod/readme.txt>.

**Depends** R (>= 3.5.0)

**License** MIT + file LICENSE

**Imports** countrycode, curl, data.table, httr, R.utils, stats, utils

**Suggests** dplyr, spelling, ggplot2, gridExtra, knitr, mapproj, maps, remotes, rmarkdown, roxygen2, testthat, tidy

**RoxygenNote** 7.2.1

**Encoding** UTF-8

**Language** en-US

**NeedsCompilation** no

**Repository** CRAN

**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 H. 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>>), Taras Kaduk [ctb] (Suggestion for handling bulk station downloads more efficiently), 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, <<https://orcid.org/0000-0002-7973-6322>>), Tyler Widdison [ctb] (Reported bug where `nearest\_stations()` did not return stations in order of nearest to farthest)

**Maintainer** Adam H. Sparks <adamhsparks@gmail.com>

**Date/Publication** 2022-08-13 06:50:02 UTC

## R topics documented:

get_GSOD . . . . .	2
get_inventory . . . . .	4
nearest_stations . . . . .	5
print.GSODR.Info . . . . .	6
reformat_GSOD . . . . .	7
update_station_list . . . . .	8

<b>Index</b>	<b>10</b>
--------------	-----------

---

get_GSOD	<i>Download and return a data.table object of GSOD weather data</i>
----------	---

---

### 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 using the improved August-Roche-Magnus approximation (Alduchov and Eskridge 1996).

## Usage

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

## Arguments

years	Year(s) of weather data to download.
station	Optional. Specify a station or multiple stations for which to retrieve, check and clean weather data using <i>STATION</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 for download 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.
country	Optional. Specify a country for which to retrieve weather data; full name, 2 or 3 letter ISO or 2 letter FIPS codes can be used. All stations within the specified country will be returned.
max_missing	Optional. The maximum number of days allowed to be missing from a station's data before it is excluded from final file output.
agroclimatology	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

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

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, <https://www.ncei.noaa.gov/data/global-summary-of-the-day/doc/readme.txt>.

## Value

A data frame as a `data.table` object of GSOD weather data.

## References

Alduchov, O.A. and Eskridge, R.E., 1996. Improved Magnus form approximation of saturation vapor pressure. *Journal of Applied Meteorology and Climatology*, 35(4), pp.601-609. DOI: <10.1175

## Note

**GSODR** attempts to validate year and station combination requests, however, in certain cases the start and end date may encompass years where no data is available. In these cases no data will be returned. It is suggested that the user check the latest data availability for the station(s) desired using [get\\_inventory](#) as this list is frequently updated by the NCEI and is not shipped with **GSODR**.

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. A log of IP addresses accessing these data and products will be maintained and may be made available to data providers.”

## Author(s)

Adam H. Sparks, <adamhsparks@gmail.com>

## See Also

[reformat\\_GSOD](#)

## Examples

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

# Download weather data for the year 1929
w_1929 <- get_GSOD(years = 1929)

# Download weather data for the year 1929 for Ireland
ie_1929 <- get_GSOD(years = 1929, country = "Ireland")
```

---

get\_inventory

*Download and return a data.table object GSOD weather station data inventories*

---

**Description**

The NCEI maintains a document, <https://www1.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.table` object of station inventories

**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. A log of IP addresses accessing these data and products will be maintained and may be made available to data providers.”

**Author(s)**

Adam H. Sparks, <adamhsparks@gmail.com>

**Examples**

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

---

nearest_stations	<i>Find nearest GSOD stations to a specified latitude and longitude</i>
------------------	---

---

**Description**

Given latitude and longitude values entered as decimal degrees (DD), this function returns a list (as an atomic vector) of station ID 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**

By default a class `character vector` object of station identification numbers. in order from nearest to farthest in increasing order. If `return_full` is TRUE, a `data.table` with full station metadata including the distance from the user specified coordinates is returned.

**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

---

reformat_GSOD	<i>Tidy and return a data.table object of GSOD weather from local storage</i>
---------------	---

---

## Description

This function automates cleaning and reformatting of GSOD station files in “YEAR.tar.gz”, provided that they have been untarred or “STATION.csv” format that have been downloaded from the United States National Center for Environmental Information’s (NCEI) download page. 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 using the improved August-Roche-Magnus approximation (Alduchov and Eskridge 1996). All units are converted to International System of Units (SI), *e.g.*, Fahrenheit to Celsius and inches to millimetres.

## Usage

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

## Arguments

dsn	User supplied full file path to location of data files on local disk for tidying.
file_list	User supplied list of file paths to individual files of data on local disk for tidying. Ignored if dsn is set. Use if there are other files in the dsn that you do not wish to reformat.

## Details

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, a single station is 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 “csv” file format are supported by this function. If you have downloaded the full annual (“YYYY.tar.gz”) file you will need to extract the individual station files from the tar file first to use this function.

Note that reformat\_GSOD() will attempt to reformat any “.csv” files found in the dsn that you provide. If there are non- GSOD files present this will lead to errors.

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 [data.table](#) object of GSOD data.

## References

Alduchov, O.A. and Eskridge, R.E., 1996. Improved Magnus form approximation of saturation vapor pressure. Journal of Applied Meteorology and Climatology, 35(4), pp.601-609. DOI: <10.1175

**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. A log of IP addresses accessing these data and products will be maintained and may be made available to data providers.”

**Author(s)**

Adam H. Sparks, <adamhsparks@gmail.com>

**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.

**Examples**

```
# Download data to 'tempdir()'
download.file(
  url =
    "https://www.ncei.noaa.gov/data/global-summary-of-the-day/access/2010/95551099999.csv",
  destfile = file.path(tempdir(), "95551099999.csv"),
  mode = "wb"
)

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

tbar
```

---

update\_station\_list    *Download latest station list metadata and update internal database*

---

**Description**

This function downloads the latest station list (isd-history.csv) from the NCEI server and updates the data distributed with **GSODR** to the latest stations available. These data provide unique identifiers, country, state (if in U.S.) and when weather observations begin and end.

**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 isd\_history.csv file that is not available in the self-contained database distributed with **GSODR**.

To directly access these data, use:

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

**Author(s)**

Adam H. Sparks, <adamhsparks@gmail.com>

**Examples**

```
update_station_list()
```

# Index

character, [6](#)

data.table, [3](#), [5–7](#)

get\_GSOD, [2](#), [5](#), [8](#)

get\_inventory, [4](#), [4](#)

nearest\_stations, [5](#)

print.GSODR.Info, [6](#)

reformat\_GSOD, [4](#), [7](#)

update\_station\_list, [8](#)

vector, [6](#)