

Package: reumetsat (via r-universe)

October 22, 2024

Type Package

Title EUMETSAT Offline Data Products

Version 0.1.0

Date 2024-10-19

Description Functions for reading files downloaded from AC SAF EUMETSAT server containing data for offline products. Initially only Surface UV data supported.

License GPL (>= 2)

VignetteBuilder knitr

Depends R (>= 4.0.0)

Imports utils, rhdf5 (>= 2.45.0)

Suggests knitr (>= 1.43), rmarkdown (>= 2.23), spelling, testthat (>= 3.1.10)

LazyLoad yes

ByteCompile true

Encoding UTF-8

URL <https://docs.r4photobiology.info/reumetsat/>

BugReports <https://github.com/aphalo/reumetsat/issues/>

RoxygenNote 7.3.2

Language en-US

Config/testthat/edition 3

Repository <https://aphalo.r-universe.dev>

RemoteUrl <https://github.com/aphalo/reumetsat>

RemoteRef HEAD

RemoteSha 681bb432e7d0b1adfc87370c6b40a87e3a6bc2fa

Contents

read_AC_SAF_hdf5	2
read_AC_SAF_txt	4

read_AC_SAF_hdf5 *Offline AC SAF Surface UV*

Description

The dates are decoded from the file names, expecting these to be those set by the data provider. The grid is expected to be identical in all files that are imported in a call to `read_AC_SAF_hdf5()`, and grid subsetting is currently not supported. If not all the files named in the argument to `'files'` are accessible, an error is triggered early. If the files differ in the grid, an error is triggered when reading the first mismatching file. Missing variables named in `'vars.to.read'` if detected when reading the first file, are filled with the `'fill'` value, otherwise they trigger an error when an attempt is made to read them.

Usage

```
read_AC_SAF_hdf5(
  files,
  data.product = NULL,
  group.name = "GRID_PRODUCT",
  vars.to.read = NULL,
  fill = NA_real_,
  verbose = interactive()
)

vars_AC_SAF_hdf5(files, data.product = NULL, group.name = "GRID_PRODUCT")

grid_AC_SAF_hdf5(files, expand = FALSE)
```

Arguments

<code>files</code>	character A vector of file names, no other limitation in length than available memory to hold the data.
<code>data.product</code>	character Currently only "Surface UV" supported.
<code>group.name</code>	character The name of the 'group' in the HDF5 files.
<code>vars.to.read</code>	character A vector of variable names. If <code>'NULL'</code> all the variables present in the first file are read.
<code>fill</code>	numeric The R value used to replace the fill value used in the file, which is retrieved from the file metadata, and also used to fill missing variables.
<code>verbose</code>	logical Flag indicating if progress, and time and size of the returned object should be printed.
<code>expand</code>	logical Flag indicating whether to return ranges or a full grid.

Details

Import gridded daily "offline products" data files from EUMETSAT AC SAF (Atmospheric Composition Monitoring) in HDF5 format. Currently only the "Surface UV" data product files as downloaded from the FMI server are supported.

Value

A data frame with columns named "Date", "Longitude", "Latitude", the data variables with their original names, and "QualityFlags". The data variables have their metadata stored as R attributes.

Note

The constraint on the consistency among all files to be read allows very fast reading into a single data frame. If the files differ in the grid or set of variables, this function can be used to read the files individually into separate data frames. These data frames can later be row-bound together.

Variable 'QualityFlags' is encoded as 64 bit integers in the HDF5 file and read as a double. R package 'bit64' can be used to print these values as 64 bit integers.

When requesting the data from the EUMETSAT AC SAF FMI server at <https://acsaf.org/> it is possible to select the range of latitudes and longitudes and the variables to be included in the file. This is more efficient than doing the selection after importing the data into R. The data are returned as a .zip compressed file containing one HDF5 file for each day in the range of dates selected. For world coverage each of these files can be as large as 10 MB in size depending on how many variables they contain. These files in HDF5 format are binary files so the size in RAM of a 'data.frame' containing one-year of data can be a few 10's of GB.

This function's performance is quite fast as long as there is enough RAM available to hold the data frame and the files are read from a reasonably fast SSD. It has been tested by importing one-year's worth of data with worldwide coverage on a PC with 64GB RAM. The example data included in the package are only for Spain and five summer days.

References

Kujanpää, J. (2019) _PRODUCT USER MANUAL Offline UV Products v2 (IDs: O3M-450 - O3M-464) and Data Record R1 (IDs: O3M-138 - O3M-152)_. Ref. SAF/AC/FMI/PUM/001. 18 pp. EUMETSAT AC SAF.

Examples

```
# find location of one example file
one.file.name <-
  system.file("extdata", "O3MOUV_L3_20240621_v02p02.HDF5",
             package = "reumetsat", mustWork = TRUE)

# available variables
vars_AC_SAF_hdf5(one.file.name)

# available grid
grid_AC_SAF_hdf5(one.file.name)
```

```

# read all variables
midsummer_spain.tb <- read_AC_SAF_hdf5(one.file.name)
dim(midsummer_spain.tb)
summary(midsummer_spain.tb)

# read two variables
midsummer_spain_daily.tb <-
  read_AC_SAF_hdf5(one.file.name,
                  vars.to.read = c("DailyDoseUva", "DailyDoseUvb"))
dim(midsummer_spain_daily.tb)
summary(midsummer_spain_daily.tb)

# find location of three example files
three.file.names <-
  system.file("extdata",
             c("03MOUV_L3_20240621_v02p02.HDF5",
               "03MOUV_L3_20240622_v02p02.HDF5",
               "03MOUV_L3_20240623_v02p02.HDF5"),
             package = "reumetsat", mustWork = TRUE)

summer_3days_spain.tb <- read_AC_SAF_hdf5(three.file.names)
dim(summer_3days_spain.tb)
summary(summer_3days_spain.tb)

```

read_AC_SAF_txt

Offline AC SAF Surface UV time series

Description

All information is in the files, including dates, and no information is decoded from file names, that users will most likely want to rename. Each file corresponds to a single geographic location. If not all the files named in the argument to ‘files’ are accessible, an error is triggered early. If the files differ in the coordinates, an error is triggered when reading the first mismatching file if coordinates are not being added to the data frame. Missing variables named in ‘vars.to.read’ are currently ignored.

Data from multiple files are concatenated. By default, the geographic coordinates are added in such a case.

Usage

```

read_AC_SAF_txt(
  files,
  vars.to.read = NULL,
  add.geo = length(files) > 1,
  keep.QC = TRUE,
  verbose = interactive()
)

```

```
vars_AC_SAF_txt(files, keep.QC = TRUE)
```

```
grid_AC_SAF_txt(files)
```

Arguments

files	character A vector of file names, no other limitation in length than available memory to hold the data.
vars.to.read	character A vector of variable names. If 'NULL' all the variables present in the first file are read.
add.geo	logical Add columns 'Longitude' and 'Latitude' to returned data frame.
keep.QC	logical Add to the returned data frame the quality control variables, always present in the files.
verbose	logical Flag indicating if progress, and time and size of the returned object should be printed.

Details

Import time series "offline products" data files from EUMETSAT AC SAF (Atmospheric Composition Monitoring) in text format. Currently only the "Surface UV" data product files as downloaded from the FMI server are supported.

Value

A data frame with columns named "Date", "Longitude", "Latitude", the data variables with their original names (with no units). The data variables have no metadata stored as R attributes. When reading multiple files, by default the format is similar to that from function 'read_AC_SAF_hdf5()'. Column names are the same by column order can differ. File headers are saved as a list in R attribute 'file.headers'.

Note

When requesting the data from the EUMETSAT AC SAF FMI server at <https://acsaf.org/> it is possible to select the variables to be included in the file, the period and the geographic coordinates of a single location. The data are returned as a .zip compressed file containing one text file with one row for each day in the range of dates selected. These files are fairly small.

This function's performance is not optimized for speed as these single location files are rather small. The example time series data included in the package are for one summer in Helsinki, Finland.

References

Kujanpää, J. (2019) _PRODUCT USER MANUAL Offline UV Products v2 (IDs: O3M-450 - O3M-464) and Data Record R1 (IDs: O3M-138 - O3M-152)_. Ref. SAF/AC/FMI/PUM/001. 18 pp. EUMETSAT AC SAFT.

Examples

```
# find location of one example file
one.file.name <-
  system.file("extdata", "AC_SAF_ts_Viikki.txt",
             package = "reumetsat", mustWork = TRUE)

# Available variables
vars_AC_SAF_txt(one.file.name)

# Grid point coordinates
grid_AC_SAF_txt(one.file.name)

# read all variables
summer_viikki.tb <-
  read_AC_SAF_txt(one.file.name)
dim(summer_viikki.tb)
colnames(summer_viikki.tb)
str(sapply(summer_viikki.tb, class))
summary(summer_viikki.tb)
attr(summer_viikki.tb, "file.headers")

# read all data variables
summer_viikki_QCf.tb <-
  read_AC_SAF_txt(one.file.name, keep.QC = FALSE)
dim(summer_viikki_QCf.tb)
summary(summer_viikki_QCf.tb)

# read all data variables including geographic coordinates
summer_viikki_geo.tb <-
  read_AC_SAF_txt(one.file.name, keep.QC = FALSE, add.geo = TRUE)
dim(summer_viikki_geo.tb)
summary(summer_viikki_geo.tb)

# read two variables
summer_viikki_2.tb <-
  read_AC_SAF_txt(one.file.name,
                 vars.to.read = c("DailyDoseUva", "DailyDoseUvb"))
dim(summer_viikki_2.tb)
summary(summer_viikki_2.tb)
```

Index

grid_AC_SAF_hdf5 (read_AC_SAF_hdf5), 2
grid_AC_SAF_txt (read_AC_SAF_txt), 4

read_AC_SAF_hdf5, 2
read_AC_SAF_txt, 4

vars_AC_SAF_hdf5 (read_AC_SAF_hdf5), 2
vars_AC_SAF_txt (read_AC_SAF_txt), 4