Package ‘hddtools’

February 20, 2015

Type Package
Version 0.2.4
Date 2014-11-21
Title Hydrological Data Discovery Tools
Description Facilitates discovery and handling of hydrological data, non-programmatic access to catalogues and databases.
Author Claudia Vitolo [aut, cre], Simon Moulds [aut]
Maintainer Claudia Vitolo <cvitolodev@gmail.com>
URL http://cvitolo.github.io/r_hddtools/
BugReports https://github.com/cvitolo/r_hddtools/issues
Depends R (>= 3.0.2)
Imports sp, rgdal, raster, RCurl, XML, zoo
License GPL-3
LazyData true
Repository CRAN
Keywords hydrology, hydrological modelling, hydrologic modeling, time series, environmental data, web technologies and services

NeedsCompilation no
Date/Publication 2015-01-18 06:44:12

R topics documented:

getContent ................................................................. 2
GRDCCatalogue ........................................................... 2
GRDCMonthlyTS .......................................................... 3
Hydrological Data Discovery Tools ................................. 5
KGI ClimateClass ........................................................ 5
mopexCatalogue ......................................................... 6
mopexTS ................................................................. 7
TRMM ................................................................. 8

Index 10
getcontent

Extracts links from ftp page

Description
This function extracts all the links in a ftp page

Usage
getcontent(dirs)

Arguments
dirs is the url from which links should be extracted

Value
vector containing all the links in the page

Author(s)
Claudia Vitolo

Examples
# Retrieve mopex daily catalogue
getContent(url)

GRDCCatalogue

Interface for the Global Runoff Data Centre database catalogue

Description
This function interfaces the Global Runoff Data Centre database which provides river discharge data for about 9000 sites over 157 countries.

Usage
GRDCCatalogue(bbox = NULL, stationID = NULL, metadataColumn = NULL, entryValue = NULL, mdDescription = FALSE)
GRDCMonthlyTS

Arguments

bbox  bounding box, a list made of 4 elements: minimum longitude (lonMin), minimum latitude (latMin), maximum longitude (lonMax), maximum latitude (latMax)
stationID  Station ID number, it should be in the range [1104150,6990700]
metadataColumn  name of the column to filter
entryValue  value to look for in the column named metadataColumn
mdDescription  boolean value. Default is FALSE (no description is printed)

Value

list of stations within the bounding box

Author(s)

Claudia Vitolo

Examples

# Retrieve the whole catalogue
# GRDCCatalogue()

# Define a bounding box
# bbox <- list(lonMin=-3.82, latMin=52.41, lonMax=-3.63, latMax=52.52)

# Filter the catalogue
# GRDCCatalogue(bbox)

GRDCMonthlyTS  Interface for the Global Runoff Data Centre database of Monthly Time Series

Description

This function interfaces the Global Runoff Data Centre monthly mean daily discharges database.

Usage

GRDCMonthlyTS(stationID, plotOption = FALSE)

Arguments

stationID  7 character number that identifies a station, GRDC station number is called "grdc no" in the catalogue.
plotOption  boolean to define whether to plot the results. By default this is set to TRUE.
Details

Please note that not all the GRDC stations listed in the catalogue have monthly data available.

Value

The function returns a list of 3 tables:

**mddPerYear** This is a table containing mean daily discharges for each single year (n records, one per year). It is made of 7 columns which description is as follows:

- LQ: lowest monthly discharge of the given year
- month: associated month of occurrence
- MQ: mean discharge of all monthly discharges in the given year
- HQ: highest monthly discharge of the given year
- month: associated month of occurrence
- n: number of available values used for MQ calculation

**mddAllPeriod** This is a table containing mean daily discharges for the entire period (Calculated only from years with less than 2 months missing). It is made of 6 columns which description is as follows:

- LQ: lowest monthly discharge from the entire period
- MQ_1: mean discharge of all monthly discharges in the period [m3/s]
- MQ_2: mean discharge volume per year of all monthly discharges in the period [km3/a]
- MQ_3: mean runoff per year of all monthly discharges in the period [mm/a]
- HQ: highest monthly discharge from the entire period
- n: number of available months used for MQ calculation

**mddPerMonth** This is a table containing mean daily discharges for each month over the entire period (12 records covering max. n years. Calculated only for months with less then or equal to 10 missing days). It is made of 7 columns which description is as follows:

- LQ: lowest monthly discharge of the given month in the entire period
- year: associated year of occurrence (only the first occurrence is listed)
- MQ: mean discharge from all monthly discharges of the given month in the entire period
- HQ: highest monthly discharge of the given month in the entire period
- year: associated year of occurrence (only the first occurrence is listed)
- std: standard deviation of all monthly discharges of the given month in the entire period
- n: number of available daily values used for computation

Author(s)

Claudia Vitolo

Examples

```r
# x <- GRDCMonthlyTS(stationID=1107700)
```
Facilitates discovery and handling of hydrological data, non-programmatic access to catalogues and databases.

Many governmental bodies and institutions are currently committed to publish open data as the result of a trend of increasing transparency, based on which a wide variety of information produced at public expense is now becoming open and freely available to improve public involvement in the process of decision and policy making. Discovery, access and retrieval of information is, however, not always a simple task. Especially when programmatic access to data resources is not allowed, downloading metadata catalogue, select the information needed, request datasets, de-compression, conversion, manual filtering and parsing can become rather tedious. The R package hddtools is an open source project, designed to make all the above operations more efficient by means of reusable functions. The package facilitate non programmatic access to various online data sources such as the Global Runoff Data Centre, NASA’s TRMM mission, the Data60UK database amongst others. This package complements R’s growing functionality in environmental web technologies to bridge the gap between data providers and data consumers and it is designed to be the starting building block of scientific workflows for linking data and models in a seamless fashion.

hddtools is a collaborative effort of Claudia Vitolo and Simon Moulds.


kgclimateclass

Function to identify the updated Koppen-Greiger climate zone (on a 0.1 x 0.1 degrees resolution map).

Given a bounding box, the function identifies the overlapping climate zones.

KGClimateClass(bbox, updatedBy = "Peel", verbose = FALSE)
mopexCatalogue

Arguments

bbox bounding box, a list made of 4 elements: minimum longitude (lonMin), minimum latitude (latMin), maximum longitude (lonMax), maximum latitude (latMax)

updatedBy this can either be Kottek

verbose if TRUE more info are printed on the screen

Value

List of overlapping climate zones.

Author(s)

Claudia Vitolo

References


Examples

# Define a bounding box
# bbox <- list(lonMin=-3.82, latMin=52.41, lonMax=-3.63, latMax=52.52)

# Get climate classes
# KGClimateClass(bbox)

mopexCatalogue Interface for the MOPEX database catalogue

Description

This function interfaces the MOPEX database catalogue (available from ftp://hydrology.nws.noaa.gov/pub/gcip/mopex/US_Data/) containing 438 daily datasets.

Usage

mopexCatalogue(bbox = NULL, metadataColumn = NULL, entryValue = NULL, verbose = FALSE)

Arguments

bbox bounding box, a list made of 4 elements: minimum longitude (lonMin), minimum latitude (latMin), maximum longitude (lonMax), maximum latitude (latMax)

metadataColumn name of the column to filter

entryValue value to look for in the column named metadataColumn

verbose if TRUE it returns whether the data is coming from live or cached data sources
Value

This function returns a data frame made of 5 columns: "id" (hydrometric reference number), "name", "location", "Latitude" and "Longitude".

Author(s)
Claudia Vitolo

Examples

# Retrieve the whole catalogue
# mopexCatalogue()

---

mopexTS

*Interface for the MOPEX database of Daily Time Series*

Description

This function extract the dataset containing daily rainfall and streamflow discharge at one of the MOPEX locations.

Usage

mopexTS(hydroRefNumber, plotOption = FALSE, timeExtent = NULL)

Arguments

- `hydroRefNumber`: hydrometric reference number
- `plotOption`: boolean to define whether to plot the results. By default this is set to TRUE.
- `timeExtent`: is a vector of dates and times for which the data should be retrieved

Value

The function returns a data frame containing 2 time series (as zoo objects): "P" (precipitation) and "Q" (discharge).

Author(s)
Claudia Vitolo

Examples

# mopexTS("14359000")
Description

The TRMM dataset provide global historical rainfall estimation in a gridded format.

Usage

```r
TRMM(fileLocation = "-/", url = "ftp://disc2.nascom.nasa.gov/data/TRMM/Gridded/", product = "3B43", version = 7, year = 2012, type = "precipitation.accum", bbox = NULL, timeExtent = NULL)
```

Arguments

- `fileLocation` file path where to save the GeoTiff
- `url` url where data is stored (e.g. "ftp://disc2.nascom.nasa.gov/data/TRMM/Gridded/3B43_V7/2012/")
- `product` this is the code that identifies a product, default is "3B43"
- `version` this is the version number, default is 7
- `year` year of interest, default is 2012
- `type` this is the type of information needed, default is "precipitation.accum". Other types could be "gaugeRelativeWeighting.bin" and "relativeError.bin"
- `bbox` bounding box, a list made of 4 elements: minimum longitude (lonMin), minimum latitude (latMin), maximum longitude (lonMax), maximum latitude (latMax)
- `timeExtent` is a vector of dates and times for which the data should be retrieve

Details

This code is based upon Martin Brandt’s blog post: http://matinbrandt.wordpress.com/2013/09/04/automatically-downloading-and-processing-trmm-rainfall-data/ and on the TRMM FAQ: http://disc.sci.gsfc.nasa.gov/additional/faq/precipitation_faq.shtml

Value

Data is loaded as rasterbrick, then converted to a multilayer Geotiff that can

Author(s)

Claudia Vitolo
Examples

# Define a bounding box
# bbox <- list(lonMin=-3.82, latMin=52.41, lonMax=-3.63, latMax=52.52)

# TRMM(product="3B43",
#     version=7,
#     year=2012,
#     bbox)
Index

getcontent, 2
GRDCCatalogue, 2
GRDCMonthlyTS, 3

Hydrological Data Discovery Tools, 5
Hydrological Data Discovery Tools-package (Hydrological Data Discovery Tools), 5

KGClimateClass, 5
mopexCatalogue, 6
mopexTS, 7

TRMM, 8