

Package ‘reddPrec’

January 19, 2024

Title Reconstruction of Daily Data - Precipitation

Version 2.0.1

Description Applies quality control to daily precipitation observations; reconstructs the original series by estimating precipitation in missing values; and creates gridded datasets of daily precipitation.

License GPL-3

Encoding UTF-8

Imports terra, stats, foreach, doParallel, reshape, qmap

RoxygenNote 7.2.3

NeedsCompilation no

Author Roberto Serrano-Notivoli [aut, cre]
(<https://orcid.org/0000-0001-7663-1202>)

Maintainer Roberto Serrano-Notivoli <roberto.serrano@unizar.es>

Repository CRAN

Date/Publication 2024-01-19 12:50:03 UTC

R topics documented:

gapFilling	1
gridPcp	3
qcPrec	4

Index	7
--------------	----------

gapFilling	<i>Estimating new values in original missing values data series of daily precipitation</i>
------------	--

Description

This function uses the neighboring observations to estimate new precipitation values in those days and locations where no records exist.

Usage

```
gapFilling(
  prec,
  sts,
  dates,
  stmethod = NULL,
  thres = NA,
  neibs = 10,
  coords,
  crs,
  coords_as_preds = TRUE,
  window,
  ncpu = 2
)
```

Arguments

<code>prec</code>	matrix or data.frame containing the original (cleaned) precipitation data. Each column represents one station. The names of columns must coincide with the names of the stations.
<code>sts</code>	matrix or data.frame. A column "ID" (unique ID of stations) is required. The rest of the columns (all of them) will act as predictors of the model.
<code>dates</code>	vector of class "Date" with all days of observations (yyyy-mm-dd).
<code>stmethod</code>	standardization method. 'quant' or 'ratio', see details.
<code>thres</code>	numeric. Maximum radius (in km) where neighboring stations will be searched. NA value uses the whole spatial domain.
<code>neibs</code>	integer. Number of nearest neighbors to use.
<code>coords</code>	vector of two character elements. Names of the fields in "sts" containing longitude and latitude.
<code>crs</code>	character. Coordinates system in EPSG format (e.g.: "EPSG:4326").
<code>coords_as_preds</code>	logical. If TRUE (default), "coords" are also taken as predictors.
<code>window</code>	odd integer. Length of data considered for standardization
<code>ncpu</code>	number of processor cores used to parallel computing.

Details

After the gap filling, "stmethod" allows for an standardization of the predictions based on the observations. It only works for daily data. For other timescales (monthly, annual) use "stmethod=NULL". The "window" parameter is a daily-moving centered window from which data is collected for each year (i.e. a 15-day window on 16th January will take all predictions from 1st to 30th January of all years to standardize them with their corresponding observations. Only standardized prediction of 16th January is returned. Process is repeated for all days).

Examples

```
## Not run:
set.seed(123)
prec <- round(matrix(rnorm(30*50, mean = 1.2, sd = 6), 30, 50), 1)
prec[prec<0] <- 0
prec <- apply(prec, 2, FUN = function(x){x[sample(length(x),5)] <- NA; x})
colnames(prec) <- paste0('sts_',1:50)
sts <- data.frame(ID = paste0('sts_',1:50), lon = rnorm(50,0,1),
                 lat = rnorm(50,40,1), dcoast = rnorm(50,200,50))
filled <- gapFilling(prec, sts,
                   dates = seq.Date(as.Date('2023-04-01'),
                                     as.Date('2023-04-30'),by='day'),
                   stmethod = "ratio", thres = NA, coords = c('lon','lat'),
                   coords_as_preds = TRUE, crs = 'EPSG:4326', neibs = 10,
                   window = 11, ncpu = 2)

str(filled)
summary(filled)

## End(Not run)
```

gridPcp

Gridded dataset creation

Description

This function creates a gridded precipitation dataset from a station-based dataset.

Usage

```
gridPcp(
  prec,
  grid,
  sts,
  dates,
  ncpu,
  thres,
  neibs,
  coords,
  crs,
  coords_as_preds
)
```

Arguments

prec matrix or data.frame containing the original (cleaned) precipitation data. Each column represents one station. The names of columns must coincide with the names of the stations.

grid	SpatRaster. Collection of rasters representing each one of the predictors.
sts	matrix or data.frame. A column "ID" (unique ID of stations) is required. The rest of the columns (all of them) will act as predictors of the model.
dates	vector of class "Date" with all days of observations (yyyy-mm-dd).
ncpu	number of processor cores used to parallel computing.
thres	numeric. Maximum radius (in km) where neighboring stations will be searched. NA value uses the whole spatial domain.
neibs	integer. Number of nearest neighbors to use.
coords	vector of two character elements. Names of the fields in "sts" containing longitude and latitude.
crs	character. Coordinates system in EPSG format (e.g.: "EPSG:4326").
coords_as_preds	logical. If TRUE (default), "coords" are also taken as predictors.

Examples

```
## Not run:
alt <- terra::rast(volcano, crs = 'EPSG:4326')
terra::ext(alt) <- c(-1,3,38,42)
lon <- terra::rast(cbind(terra::crds(alt), terra::crds(alt)[,1]), type='xyz', crs='EPSG:4326')
lat <- terra::rast(cbind(terra::crds(alt), terra::crds(alt)[,2]), type='xyz', crs='EPSG:4326')
dcoast <- terra::costDist(alt, target=min(terra::values(alt)))/1000
grid <- c(alt, lon, lat, dcoast)
names(grid) <- c('alt', 'lon', 'lat', 'dcoast')

prec <- round(matrix(rnorm(2*25, mean = 1.2, sd = 4), 2, 25), 1)+1
prec[prec<0] <- 0
colnames(prec) <- paste0('sts_', 1:25)
sts <- data.frame(ID = paste0('sts_', 1:25), as.data.frame(terra::spatSample(grid, 25)))
gridPcp(prec, grid, sts,
        dates = seq.Date(as.Date('2023-04-01'), as.Date('2023-04-02'), by='day'),
        thres = NA, coords = c('lon', 'lat'), coords_as_preds = TRUE,
        crs = 'EPSG:4326', neibs = 10, ncpu = 2)

r <- terra::rast(c('./pred/20230401.tif', './err/20230401.tif'))

## End(Not run)
```

qcPrec

Quality Control of daily precipitation observations

Description

This function apply several threshold-based criteria to filter original observations of daily precipitation.

Usage

```

qcPrec(
  prec,
  sts,
  crs,
  coords,
  coords_as_preds = TRUE,
  neibs = 10,
  thres = NA,
  qc = "all",
  qc3 = 10,
  qc4 = c(0.99, 5),
  qc5 = c(0.01, 0.1, 5),
  ncpu = 1
)

```

Arguments

prec	matrix or data.frame containing the original precipitation data. Each column represents one station. The names of columns have to be names of the stations.
sts	matrix or data.frame. A column "ID" (unique ID of stations) is required. The rest of the columns (all of them) will act as predictors of the model.
crs	character. Coordinates system in EPSG format (e.g.: "EPSG:4326").
coords	vector of two character elements. Names of the fields in "sts" containing longitude and latitude.
coords_as_preds	logical. If TRUE (default), "coords" are also taken as predictors.
neibs	integer. Number of nearest neighbors to use.
thres	numeric. Maximum radius (in km) where neighboring stations will be searched. NA value uses the whole spatial domain.
qc	vector of strings with the QC criteria to apply. Default is "all". See details.
qc3	numeric. Indicates the threshold (number of times higher or lower) from which a observation, in comparison with its estimate, should be deleted. Default is 10.
qc4	numeric vector of length 2. Thresholds of wet probability (0 to 1) and magnitude (in the units of input precipitation data) from which a observation of value zero, in comparison with its estimate, should be deleted. Default is c(0.99, 5).
qc5	numeric vector of length 2. Thresholds of dry probability (0 to 1) and magnitude (in the units of input precipitation data) from which a observation higher than a specific value (also in the original units), in comparison with its estimate, should be deleted. Default is c(0.01, 0.1, 5).
ncpu	number of processor cores used to parallel computing.

Details

Parameter "sts" must have an "ID" field containing unique identifiers of the stations.

"qc" can be "all" (all criteria are applied) or a vector of strings (e.g.: c("1","2","4")) indicating the QC criteria to apply to observations: "1" (suspect value): obs==0 & all(neibs>0); "2" (suspect zero): obs>0 & all(neibs==0); "3" (suspect outlier): obs is "qc3" times higher or lower than the estimate; "4" (suspect wet): obs==0 & wet probability > "qc4[1]" & estimate > "qc4[2]"; "5" (suspect dry): obs>"qc5[3]" & dry probability < "qc5[1]" & estimate < "qc5[2]"

Examples

```
## Not run:
set.seed(123)
prec <- round(matrix(rnorm(30*50, mean = 1.2, sd = 6), 30, 50), 1)
prec[prec<0] <- 0
colnames(prec) <- paste0('sts_',1:50)
sts <- data.frame(ID = paste0('sts_',1:50), lon = rnorm(50,0,1),
                 lat = rnorm(50,40,1), dcoast = rnorm(50,200,50))
qcdata <- qcPrec(prec, sts, crs = 'EPSG:4326', coords = c('lon','lat'),
               coords_as_preds = TRUE, neibs = 10, thres = NA,
               qc = 'all', qc3 = 10, qc4 = c(0.99, 5), qc5 = c(0.01, 0.1, 5),
               ncpu=2)

str(qcdata)

## End(Not run)
```

Index

gapFilling, 1

gridPcp, 3

qcPrec, 4