Package 'CSHShydRology'

July 18, 2019

Type Package

Title Canadian hydrological analyses

Version 1.0.1

Date 2019-05-07

Description A collection of user submitted functions to aid in the analysis of hydrological data.

License AGPL-3 | file LICENSE

URL https://github.com/CSHS-hydRology/CSHShydRology

Depends R (>= 3.1)

Imports fields,
Kendall,
lubridate (>= 1.3),
plotrix,
timeDate,
stringr,
jsonlite,
curl,
gstat,
sp,
lmom,
lmomco,
lmomRFA,
numDeriv,
mnormt,
Matrix

Suggests knitr,
testthat,
rmarkdown

VignetteBuilder knitr

LazyData true

RoxygenNote 6.1.1

R topics documented:

`CSHShydRology-package` *Functions for Canadian hydrological analyses*

Description

CSHShydRology is intended for the use of hydrologists, particularly those in Canada. It will contain functions which focus on the use of Canadian data sets, such as those from Environment Canada. The package will also contain functions which are suited to Canadian hydrology, such as the important cold-region hydrological processes. **CSHShydRology** will also contain functions which work with Canadian hydrological models, such as Raven, CRHM, Watflood, and MESH.

This packages has been developed with the assistance of the Canadian Society for Hydrological Sciences (CSHS) `https://cwra.org/en/branches/affiliates/cshs-a` which is an affiliated society of the Canadian Water Resources Association (CWRA) `cwra.org`.

The **CSHShydRology** will contain functions grouped into several themes, including:

Statistical hydrology trend detection, data screening, frequency analysis, regionalization

Basic data manipulations input/conversion/adapter functions, missing data infilling

Visualization data visualization, standardized plotting functions

Spatial hydrology basin delineation, landscape data analysis, working with GIS

Streamflow measurement analysis rating curve analysis, velocity profiles, naturalization

Network design/analysis homogeneity assessment

Ecohydrology fisheries and ecological analysis

Wrappers/unwrappers between other packages and **CSHShydRology**

References

To cite **CSHShydRology** in publications, use the command `citation("CSHShydRology")` to get the current version of the citation.

Amax *Basic probability functions for distribution of annual maximums*

Description

Density, distribution function, quantile function and random generation for various distribution used in the modeling annual maximums.

Usage

```
dAmax(x, para, distr, log = FALSE)

pAmax(q, para, distr)

qAmax(p, para, distr)

rAmax(n, para, distr)

lAmax(l, distr, ...)
```

Arguments

x, q	Vector of quantiles.
para	Vector of parameters for the given distribution.
distr	Distribution family. See lmom2par.
log	Logical. If TRUE, probabilities p are given as log(p).
p	Vector of probabilities.
n	Number of observations.
l	L-moments.
...	Other parameters.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

Examples

```
u <- runif(5)
u
x <- qAmax(u, c(100,3,.001), 'gno')
pAmax(x, c(100,3,.001), 'gno')

x <- rAmax(100, c(100,30,0), 'gev')
sum(dAmax(x, c(100,30, 0), 'gev', log = TRUE))
```

```
lAmax(c(100, .3, .2), 'gev', lscale = FALSE)
```

`Basic_data_manipulation-functions` *Basic data manipulation functions*

Description

These functions read in or convert values among formats

read_ECDE_flows Reads a file of WSC daily flows from ECDataExplorer

get_wscstation Reads station information from a data file produced by ECDE

get_AHCCD_monthly Downloads monthly Adjusted and Homogenized Canadian Climate Data (AHCCD) values

read_AHCCD_daily Reads file of daily AHCCD values

read_AHCCD_monthly Reads file of monthly AHCCD values

`binned_MannWhitney` *Compares two time periods of data using Mann-Whitney test*

Description

It bins data based upon a bin size, extracting data for two time periods and tests for change between two such periods. Result can be passed to `polar_plot` for visualization

Usage

```
binned_MannWhitney(mdata, step, range1, range2, ptest = 0.05,
  station_ID = "", station_name = "", variable = "discharge")
```

Arguments

`mdata`	A data frame of hydrometric data. Must contain the variables `Date` and `Flow`.
`step`	An integer indicating the degree of smoothing eg. 1, 5, 11.
`range1`	The first and last year of first period, as `c(first,last)`
`range2`	The first and last year of second period as codec(first,last)
`ptest`	The significance level. The default is `0.05`.
`station_ID`	Optional ID of station.
`station_name`	Optional name of station.
`variable`	Name of variable. Default is 'discharge'

Value

Returns a list containing:

StationID ID of station

Station_lname Name of station

bin_width Smoothing time step

range1 range1 years

range2 range2 years

p_used p value used

fail TRUE if test failed due to missing values

bin_method method used for binning

test_method Mann-Whitney U

series a data frame containing:

period period numbers i.e. 1:365/step

period1 median values for each bin in period 1

period2 median values for each bin in period 2

mwu Mann Whitney U-statistic for each bin between the two periods

prob probability of U for each period

code significance codes for each bin

Author(s)

Paul Whitfield <paul.h.whitfield@gmail.com>

References

Whitfield, P.H., Cannon, A.J., 2000. Recent variations in climate and hydrology in Canada. Canadian Water Resources Journal 25: 19-65.

See Also

polar_plot polar_plot_prep

Examples

```
## Not run:
# fails due to missing data in both periods
range1 <- c(1960,1969)
range2 <- c(1990,1999)
b_MW <- binned_MannWhitney(W05AA008, step=5, range1, range2, ptest=0.05)
## End(Not run)

range1 <- c(1970,1979)
range2 <- c(1990,1999)
b_MW <- binned_MannWhitney(W05AA008, step = 5, range1, range2,
ptest = 0.05, station_ID = "05AA008", station_name= "Crowsnest River at Frank")
```

booth_plot *Create a Booth plot of peaks over a threshold*

Description

A Booth plot is a plot of peaks over threshold flood events with duration on the horizontal and either magnitude (default) or volume on the vertical axis

Usage

```
booth_plot(events, threshold, title, type = "mag", colour1 = 1,
  colour2 = 1)
```

Arguments

events	A data frame of POT events from the function get_peaks
threshold	The threshold used by get_peaks
title	Plot title
type	The plot type, either 'mag' (magnitude, the default) or 'vol' (volume)
colour1	A vector of length 12 with line colours of rings or symbols. Defaults to those used by Booth.
colour2	A vector of length 12 with fill colours of rings or symbols. Defaults to those used by Booth.

Value

No value is returned; a standard R graphic is created.

References

Booth, E.G., Mount, J.F., Viers, J.H. 2006. Hydrologic Variability of the Cosumnes River Floodplain. San Francisco Estuary & Watershed Science 4:21.

Whitfield, P.H., and J.W. Pomeroy. 2016. Changes to flood peaks of a mountain river: implications for analysis of the 2013 flood in the Upper Bow River, Canada. Hydrological Processes 30:4657-73. doi: 10.1002/hyp.10957.

See Also

get_peaks

Examples

```
threshold <- 0.1 * max(W05AA008$Flow)  # arbitrary threshold
peaks <- get_peaks(W05AA008, threshold)
events <- peaks$POTevents
booth_plot(events, threshold, title = "05AA008", type='mag')
booth_plot(events, threshold, title = "05AA008", type='vol')
```

CvRoi *Cross-validation using region of influence and kriging*

Description

Return a matrix of criteria evaluated by using region of influence (ROI) and kriging. It includes: Root mean square error (rmse), relative RMSE (rrmse), Nash-Sutcliffe (nsh), Mean absolute deviation (mad), relative MAD (rmad) and the skill score based on MAD (smad). The latter smad has the same form as nsh except that absolute error are taken instead of square error.

Usage

```
CvRoi(x, nk, phy, similarity, kriging = NULL, ker = TRUE,
  model = "Exp", fold = 5, verbose = TRUE)

## S3 method for class 'roicv'
head(x, crit = "mad", ...)

## S3 method for class 'roicv'
plot(x, crit = "mad", best.col = "red",
  best.pch = 16, best.cex = 1, ...)
```

Arguments

x	Data.
nk	List of neighborhood sizes to try.
phy	Formula defining the physical descriptors.
similarity	Formula defining the covariates used to evaluate the similarity between site, i.e. Euclidean distance to the target.
kriging	Formula defining the spatial covariates. Necessary for predicting residuals using spatial correlation.
ker	Should a (Epanechnikov) kernel be used in to weight local regression model. Otherwise uniform weight are used.
model	Variogram model. See vgm.
fold	Number of group used in the cross-validation scheme. Can also be a vector defining the group for each site.
verbose	Logical. Should a progress bar be displayed.
crit	Cross-validation criteria used to evaluate the best choice.
...	More arguments to pass to the plot function (par).
best.col, best.pch, best.cex	Argument for the point indication the best choice.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

References

Martin Durocher, Donald H. Burn, Shabnam Mostofi Zadeh & Fahim Ashkar (2019) Estimating flood quantiles at ungauged sites using nonparametric regression methods with spatial components, Hydrological Sciences Journal, 64:9, 1056-1070, https://doi.org/10.1080/02626667.2019.1620952

See Also

FitRoi

Examples

```
attach(flowUngauged)

## Using multidimensional scaling for projecting coordinates
coord <- cbind(lon,lat)
coord <- cmdscale(GeoDist(coord))
colnames(coord) <- c('lon','lat')

## Transform data if necessary
xdf <- data.frame(y      = log(l1),
                  area   = scale(log(area)),
                  wb     = scale(log(wb)),
                  stream = scale(log(stream)),
                  map    = scale(log(map)),
                  coord)

 ## select a validation and training set
 set.seed(9382)
 vid <- runif(nrow(xdf)) > .8
 tid <- !vid

 # formula of the relationship between flood quantile and descriptors
 fphy <- y ~ area + map + wb + stream
 fsimilarity <- ~ area + map

 ## Perform cross-validation.
 system.time(out <- CvRoi(x = xdf, nk = seq(20,150, 10), fold = 5,
                 phy = fphy,  similarity = fsimilarity, model = 'Exp'))

 head(out, 'nsh')
 plot(out, 'mad')
```

DataWide *Transpose dataset to a wide format*

Description

Return a matrix where columns correspond to data. Each row represent a specific time and missing values are added where need. The pivoted variable should be a numerical value.

Usage

```
DataWide(x, ...)

## S3 method for class 'data.frame'
DataWide(x, order.time = TRUE, order.site = FALSE,
  ...)

## S3 method for class 'formula'
DataWide(form, x, ...)

## S3 method for class 'matrix'
DataWide(x, ...)
```

Arguments

x	Dataset in long format. First column should be the value to transpose. The second and third columns are respectively site and time variable. If the time variable is omitted, the pivot will ignore it and all columns will start at the first position.
...	Other parameters.
order.time, order.site	Logical. Should the rows (time) or columns (site) of the output be sorted.
form	Formula that specifies the site and time variable. Must have the form: 'value ~ site + time'.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

Examples

```
data(flowAtlantic)
xd <- flowAtlantic$ams

## transpose by year
xd$year <- format(xd$date, '%Y')
DataWide(ams ~ id + year, xd)[60:70,1:5]

## Without time
DataWide(ams ~ id, xd)[1:6,1:5]
```

`DistSeason` *Distance in seasonal space*

Description

Return a matrix of distances between points in the seasonal space that characterizes timing and regularity. It is equivalent to Euclidean distance applied to regularity (radius) and timing (angle) separately.

Usage

```
DistSeason(x, ...)

## S3 method for class 'numeric'
DistSeason(x, a, w = 1/pi, ...)

## S3 method for class 'data.frame'
DistSeason(x, w = 1/pi, ...)

## S3 method for class 'formula'
DistSeason(form, x, w = 1/pi, ...)
```

Arguments

`x, a`	Coordinates in the seasonal space: radius ‘x‘ and angle ‘a‘.
`...`	Other parameters.
`w`	Weight to favor angle over radius. By default it is 1/pi, which bring angle in the interval [0,1].
`form`	Formula and dataset providing the coordinates of the seasonal space. Must be of the form ‘radius ~ angle‘.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

References

Durocher, M., Burn, D. H., & Ashkar, F. (2019). Comparison of estimation methods for a nonstationary index-flood model in flood frequency analysis using peaks over threshold. https://doi.org/10.31223/osf.io/rnepc

Examples

```
scoord <- data.frame(radius = runif(5),
                     angle = runif(5,0,2*pi))

DistSeason(radius ~ angle , scoord)
```

`ExtractAmax` *Extracts the annual maximums of a daily time series*

Description

Returns a dataset containing the annual maximums, the date and the number of observations during the year.

Usage

```
ExtractAmax(x, ...)

## S3 method for class 'formula'
ExtractAmax(form, x, tol = 0, ...)

## Default S3 method:
ExtractAmax(x, tol = 0, nlab = "n", ylab = "yy",
  ...)
```

Arguments

`x`	Data. If no formula is passed, the first column must be the value and the second the date.
`...`	Other parameters.
`form`	Formula of the form `value ~ date` that specifies the variable from which the annual maximums are extracted and a date variable.
`tol`	Filter the years having less than `tol` days.
`nlab, ylab`	Names for the added columns representing respectively the number of yearly observations and the year. If set to NULL the given column is not added.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

Examples

```
out <- ExtractAmax(flow ~ date, flowStJohn, tol = 350)
head(out)
```

`fdcurve` *Plot Flow Duration Curve*

Description

A flow duration curve is a plot of flow magnitude against exceedance probability. The plot may contain the Gustard Curves or they can be omitted. The default is for curves to be plotted against probability, but an option is to plot against the normalized exceedance probability. In that case, the x axis represents a normal distribution.

Usage

```
fdcurve(flow, title = "", normal = FALSE, gust = TRUE)
```

Arguments

`flow`	Vector containing daily flows
`title`	The plot title
`normal`	If `normal = TRUE` then exceedance probability is normalized. Default is `FALSE`
`gust`	If `TRUE` (the default), adds the curves from Gustard et al 1992

Value

Plots the flow durations and returns a data frame containing the exceedance probabilty and flow

Author(s)

Paul Whitfield <paul.h.whitfield@gmail.com>

References

Gustard, A., A. Bullock, and J.M. Dixon. 1992. Low flow estimation in the United Kingdom. Institute of Hydrology, 292. Wallingford: Institute of Hydrology.

Vogel, R.M., and N.M. Fennessy. 1994. Flow-duration curves. I: New Interpretation and confidence intervals. Journal of Water Resources Planning and Management ASCE 120:485-504.

Examples

```
flow <- W05AA008$Flow
# plot with Gustard 1992 curves
test <- fdcurve(flow, title="Station", normal=FALSE, gust=TRUE)

# plot with normalized exceedance probability
test <- fdcurve(flow, title="Station", normal=TRUE, gust=FALSE)
```

fgpa *Estimation of the Generalized Pareto distribution.*

Description

Low level functions for estimating of the generalized Pareto distribution(GPA) with two parameters. Can use either maximum likelihood or the method of L-moments. The algorithm of `fgpa2d` is using `optim` to directly optimize the log-likelihood (bivariate), while the algorithm of `fgpa1d` is using a transformation to use a univariate optimization routine. Additionally, `fgpa2d` constraint the shape parameter between -.5 and 1.

Usage

```
fgpa1d(x, sol = FALSE)

fgpa2d(x, sol = FALSE, par0 = NULL, ...)

fgpaLmom(x)
```

Arguments

`x`	Sample.
`sol`	Does solution from `optim` be returned. In case of `fgpa1d`, it returns the variance covariance matrix.
`par0`	Initial parameter.
`...`	aditional arguments to pass to `optim`

Reference

Davison AC, Smith RL. (1990) Models for Exceedances over High Thresholds. Journal of the Royal Statistical Society Series B (Methodological). 52(3):393-442.

Hosking JRM (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. Journal of the Royal Statistical Society Series B (Methodological). 52(1):105-24.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

Examples

```
x <- rgpa(1000, 1, -.2)
fgpa1d(x)
fgpa2d(x)
fgpaLmom(x)
```

`FindNearest` *Find the nearest sites*

Description

Return a dataset including only the nearest site to a target. The target has a zero distance and will be at the first column of the output.

Usage

```
FindNearest(x, distance, n, super.distance = NULL, super.n = NULL,
  row = FALSE)
```

Arguments

`x`	Dataset with sites in columns and time in row.
`distance`	Distances with the target.
`n`	Number of sites to keep.
`super.distance, super.n`	Distance and number of sites for the super region. This allows to pre-filter sites according to a complementary measure of similarity.
`row`	Logical. Should only the row of the nearest site be kept. Must be used with correlation or square matrix.

References

Mostofi Zadeh, S., Burn, D.H., 2019. A Super Region Approach to Improve Pooled Flood Frequency Analysis. Canadian Water Resources Journal 0, 1-14. https://doi.org/10.1080/07011784.2018.1548946

Durocher, M., Burn, D.H., Mostofi Zadeh, S., 2018. A nationwide regional flood frequency analysis at ungauged sites using ROI/GLS with copulas and super regions. Journal of Hydrology 567, 191-202. https://doi.org/10.1016/j.jhydrol.2018.10.011

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

See Also

DataWide

Examples

```
## Organize data in the proper format
attach(flowAtlantic)
ams$year <- format(ams$date, '%Y')
xmat <- DataWide(ams ~ id + year, ams)
```

```
dim(xmat)

## Find the nearest neirghbors using great-circle distance.
## Note that it is a good habit to make sure that the column names matches

h <- GeoDist(info[colnames(xmat), c('lon','lat')])
xmat0 <- FindNearest(xmat, h[1,], 5)
dim(xmat0)

## Distance based on catchment characteristics
charac <- scale(log(info[,c('area','map')]))
h.super <- as.matrix(dist(charac))

## Find among the 20 sites with the ones with the most similar
## characteristics and keep the 5 nearest neighbors of the target.
xmat0 <- FindNearest(xmat, h[1,], 5, h.super[1,], 20)
dim(xmat0)

## subsample a distance or correlation matrix
xmat0 <- FindNearest(h, h[1,], 5, row = TRUE)
dim(xmat0)
```

`FindThresh` *Find automatically a best threshold*

Description

The function `Findthresh` can be used to identify the best threshold according to a specific rules.

Usage

```
FindThresh(x, method = "sgn", tol.sgn = 0.25, tol.ppy = 2,
  qua = "q10", tol.qua = Inf, ppy = c(0, Inf))
```

Arguments

`x`	output form SearchThresh.
`method`	Method to identify the threshold. Must be one of `sgn`, `sgn-max`, `sgn-ppy`, `max` or `ppy`.
`tol.sgn`	Selection criteria associated with a given p-value of the Anderson-Darling test.
`tol.ppy`	Selection criteria associated with a given average of peaks per year.
`qua`	Flood quantiles. Must be one of `'q2'`, `'q5'`, `'q10'`, `'q20'`, `'q50'` or `'q100'`.
`tol.qua`	Selection criteria associated with a flood quantile.
`ppy`	Range of accepted averages of peaks per year.

Details

The method sgn find the first threshold that has a p-value over a given value (tol.sgn) of the Anderson-Darling test, while max find the maximum p-value. The ppy method find the threshold that has the nearest average of peaks per year to tol.ppy. The method sgn-max find the thresholds associated with both methods and return the lowest one. The method sgn-ppy try first to find the sgn threshold. If not threshold meets the required p-value, the ppy threshold is returned.

The criteria ppy, initially filter all the candidates thresholds that are not found in a given interval of average peaks per year. With the sgn method an additional condition associated to a flood quantile qua can be specified. It imposes that the relative discrepancies between the candidate threshold and a reference is respected in addition to the criteria with the p-value. This reference is the average of the flood quantiles of the 5 lowest candidate thresholds in the set of candidates. Normally, lower thresholds should have reached stability and their choice can be controlled by the argument ppy.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

References

Durocher, M., Zadeh, S. M., Burn, D. H., & Ashkar, F. (2018). Comparison of automatic procedures for selecting flood peaks over threshold based on goodness-of-fit tests. Hydrological Processes, 0(0). https://doi.org/10.1002/hyp.13223

Solari, S., Eguen, M., Polo, M. J., & Losada, M. A. (2017). Peaks Over Threshold (POT): A methodology for automatic threshold estimation using goodness of fit p-value. Water Resources Research, 53(4), 2833-2849. https://doi.org/10.1002/2016WR019426

Examples

```
## Create list of candidate threshold
lstu <- which.floodPeaks(flow~date, flowStJohn, u =500, r = 14)
lstu <- sort(unique(flowStJohn[lstu,'flow']))
lstu <- lstu[seq(1,length(lstu)-30,2)]

out <- SearchThresh(flow~date, flowStJohn, u = lstu, declust = 'wrc', r = 14)

cid <- c('u','ppy','ad')
FindThresh(out, method = 'sgn', tol.sgn = 0.25)[,cid]
FindThresh(out, method = 'ppy', tol.ppy = 2)[,cid]
FindThresh(out, method = 'max')[,cid]

## Make find the first threshold that have at-least p-value > 0.1
## and relative descripencies between Q100 of less than 20%.
## Otherwise the ppy threshold is chosen.
## Note here that sgn is chosen because the flood quantile stabilizes
## quickly.
FindThresh(out, method = 'sgn-ppy', tol.sgn = 0.25, tol.ppy = 1.5,
           ppy = c(1,3), qua = 'q10', tol.qua = 0.2)[,cid]
```

FitAmax *At-site frequency analysis using annual maximums*

Description

Return a fitting of a distribution, normally representing annual maximums. Both maximum likelihood and L-moments estimation methods are available. If the maximum likelihood is used and fails, the L-moments solution will be returned with a warning message is issued. When more than one distribution is passed. The best distribution is selected automatically according to the AIC criteria.

Usage

```
FitAmax(x, distr = c("gev", "gno", "pe3", "glo"), method = "lmom",
  varcov = TRUE, nsim = 1000, ..., tol.gev = 0)
```

Arguments

x	Data.
distr	Distribution to fit. See Amax for the list of available distribution.
method	Estimation method. Either maximum likelihood ('mle') or L-moments ('lmom').
varcov	Should the variance-covariance matrix of the parameters be computed. For mle the covariance matrix is derived from the hessian matrix. For L-moments, non-parametric bootstrap is used.
nsim	Number of simulations used to evaluate the covariance matrix when using L-moment estimator.
...	Other parameters.
tol.gev	Accepted difference between the AIC of the GEV and the best best distribution. If the difference is inferior to tol.gev, the GEV distribution is prefered.

Value

data	Data Values.
lmom	L-moments.
para	Parameter estimates.
varcov	Covariance matrix of the parameter
llik	Value of the log-likelihood

References

Coles, S. (2001). An introduction to statistical modeling of extreme values. Springer Verlag.

Hosking, J. R. M., & Wallis, J. R. (1997). Regional frequency analysis: an approach based on L-moments. Cambridge Univ Pr.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

See Also

predict.amax, GofTest, plot.amax.

Examples

```
## Extract a time series of annual maxima
x <- ExtractAmax(flow~date, flowStJohn)$flow

## Fitting of GEV distribution using L-moments

fit <- FitAmax(x,'gev')
print(fit)
coef(fit)
AIC(fit)
fit$lmom

## The evaluation of the variance-covariance matrix can be turn down
fit <- FitAmax(x,'gev', varcov = FALSE)

## Using Maximum likelihood
fit <- FitAmax(x,'gev', method ='mle')
print(fit)
vcov(fit)

## Standard deviation of the parameter
sqrt(diag(vcov(fit)))

## Chose the best distribution according to AIC
FitAmax(x, distr = c('gev','glo','gno','pe3'), method = 'mle')
```

FitGev — *Fit Generalized Extreme Value (GEV) distribution using Generalized maximum likelihood*

Description

Fit a GEV distribution on annual maxima using the generalized maximum likelihood method with Beta prior. The output is of the class amax. See FitAmax. Asymptotic result are computed like the maximum likelihood approach.

Usage

```
FitGev(x, varcov = TRUE, mu = -0.1, sig2 = 0.015,
  method.optim = "BFGS", ...)
```

Arguments

x	Data.
varcov	Logical. Should the covariance matrix be returned.
mu, sig2	Mean and variance of the Beta prior.
method.optim	Optimisation method used by optim
...	Other parameter pass to optim.

References

Martins, E.S., Stedinger, J.R., 2000. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. Water Resour. Res. 36, 737-744. https://doi.org/10.1029/1999WR900330

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

Examples

```
x <- ExtractAmax(flow~date, flowStJohn)$flow

## Using the default physiographic prior.
fit <- FitGev(x)
print(fit)
coef(fit)
vcov(fit)
predict(fit, ci = 'delta')

## A uniform prior on interlval -0.5 to 0.5 can be used to approximate
## the maximum likelihood estimate.

AIC(FitAmax(x, distr = 'gev', method ='mle'))
AIC(FitGev(x, mu = 0, sig2 = 1/12))

## A regional study can be performed by using an empirical prior
## Here 20 sites with the same GEV distribution are simulated
## without priors.
## Then a regional estimate is obtained using an empirical prior

xmat <- replicate(20, rAmax(20, c(100,3, -.1), 'gev') )
flist <- apply(xmat,2, FitAmax, distr = 'gev', varcov = FALSE)
pmat <- sapply(flist, getElement,'para')

kap0 <- pmin(.5,pmax(-.5,pmat[3,]))

FitGev(xmat[,1], mu = mean(kap0), sig2 = var(kap0))
FitAmax(xmat[,1], distr = 'gev', method = 'mle')
```

FitPot *Peak over threshold (POT)*

Description

Fit the parameters of a thresholding model using Generalized Pareto distribution (GPA). Include declustering techniques.

Usage

```
FitPot(x, ...)

## S3 method for class 'data.frame'
FitPot(x, ...)

## S3 method for class 'matrix'
FitPot(x, ...)

## S3 method for class 'formula'
FitPot(form, x, ...)

## S3 method for class 'numeric'
FitPot(x, dt = NULL, u = 0, method = "mle",
  declust = "none", r = 1, rlow = 0.75, nsim = 1000,
  varcov = TRUE, unit = 365.25, ...)

## S3 method for class 'fpot'
coef(object, rate = FALSE, ci = FALSE, alpha = 0.05,
  ...)

## S3 method for class 'fpot'
vcov(object, rate = FALSE, ...)

## S3 method for class 'fpot'
print(x, ...)

## S3 method for class 'fpot'
logLik(object, ...)

## S3 method for class 'fpot'
AIC(object, k = 2, ...)
```

Arguments

x, form	Dataset. If a matrix or data.frame is passed as argument the first column must be the time and the second the values. A formula can be used to specify which variables of a data.frame to use. In this case, it must have the form value~time.

...	Other parameters.
dt	Date or time of observation.
u	Threshold.
method	Estimation method. Either 'lmom', 'mle' or 'mle2'.
declust	If necessary, declustering method. Either 'run' or 'wrc'.
r	Lag parameter for declustering. Either the running length between clusters or the minimum separating time between two flood Peaks. The scale must coincide with the observation date dt.
rlow	For WRC, recession level between two flood peaks in percentages.
nsim	Number of bootstrap samples.
varcov	Logical. Should the covariance matrix of the estimated parameters be evaluated.
unit	Length of cycle. Data are normaly years and unit = 365.25. Can be change to 12 or 52 for daily and weekly data.
object	Output of FitPot.
rate	Logical. Should the estimated probability of exeedance must be included in the covariance matrix.
ci	For 'coef' should confidence interval be returned.
alpha	Probability outside the confidence interval.
k	The penalty per parameter to be used.

Details

The access functions coef and vcov return respectively the parameters and the variance-covariance matrix of the POT model. For the L-moment method the covariance matrix is using bootstraps. The access function predict evaluates flood quantiles. If dt is a Date the return period is computed in years using the range of observation.

References

Coles S. (2001) An introduction to statistical modeling of extreme values. Springer Verlag.

Davison AC, Smith RL. (1990) Models for Exceedances over High Thresholds. Journal of the Royal Statistical Society Series B (Methodological). 52(3), 393-442.

See Also

which.floodPeaks, which.clusters, PlotMrl.

Examples

```
xd <- rgpa(100, 1, -.2)
fit <- FitPot(xd, u = 0)

print(fit)
vcov(fit)
```

```
predict(fit)
coef(fit, ci = TRUE)

fit <- FitPot(flow~date, flowStJohn, u = 1000,
              declust = 'wrc', r = 14)

print(fit)
plot(flow~date,flowStJohn, type = 'l')
points(fit$time,fit$excess+fit$u, col = 2, pch = 16)
abline(h=1000, col = 3, lwd = 2)

predict(fit, se = TRUE, ci = 'delta')
```

FitRegLmom *Fitting an index-flood model using regional L-moments*

Description

Returns the at-site L-moments, regional L-moments and the parameters of the regional growth curve for an index-flood model. If required, the homogeneity and goodness-of-fit measures are included. The function can be used for analyzing both annual maximums and peaks over threshold.

Usage

```
FitRegLmom(x, distr = NULL, type = "amax", diagnostic = FALSE,
  diagnostic.nsim = 1000, nmom = 4)

## S3 method for class 'reglmom'
coef(object, distr = NULL, ...)

## S3 method for class 'reglmom'
print(x, ...)

## S3 method for class 'reglmom'
plot(x, ...)

sitenames(x)
```

Arguments

x	Dataset with sites in columns and time in rows.
distr	Regional distribution. See Amax. If NULL an automatic selection is performed.
type	Type of input data. Either annual maximums ('amax') or peaks over threshold ('pot').
diagnostic	Should the homogeneity (H) and goodness of fit (Z-score) be evaluated.

`diagnostic.nsim`
: Number of simulations used to evaluate the diagnostic statistics.

`nmom`
: Number of L-moments to evaluate.

`object`
: Output from `FitRegLmom`.

`...`
: Other parameters.

Details

The function `coef` return the parameter of the at-site distributions estimated L-moments. The function `plot` will present a L-moment ratio diagram.

Value

type Type of input data

distr Regional distribution

para Parameters of the regional distribution

lmom At-site L-moments.

rlmom Regional L-moments.

nrec Record lengths.

stat Homogenous criteria and Z-score for goodness-of-fit.

discord Discordance measures.

References

Hosking, J. R. M., & Wallis, J. R. (1997). Regional frequency analysis: an approach based on L-moments. Cambridge Univ Pr.

Mostofi Zadeh, S., Burn, D.H. (2019). A Super Region Approach to Improve Pooled Flood Frequency Analysis. Canadian Water Resources Journal 0, 1-14. https://doi.org/10.1080/07011784.2018.1548946

Mostofi Zadeh, S., Durocher, M., Burn, D.H., Ashkar, F., 2019. Pooled flood frequency analysis: a comparison based on peaks-over-threshold and annual maximum series. Hydrological Sciences Journal 0, null. https://doi.org/10.1080/02626667.2019.1577556

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

See Also

predict.reglmom, PoolRemove.

Examples

```
data(flowAtlantic)
xd <- cbind(flowAtlantic$ams,
            year = format(flowAtlantic$ams$date, '%Y'))

xmat <- DataWide(ams ~ id + year, xd)

fit <- FitRegLmom(xmat)
print(fit)
plot(fit)
coef(fit, 'pe3')
```

FitRoi *Prediction at ungauged sites using region of influence and kriging*

Description

Return the prediction of a local regression model using region of influence (ROI) where the residuals are further predicted by kriging.

Usage

```
FitRoi(x, xnew, nk, phy, similarity, kriging = NULL, model = "Exp",
  ker = TRUE, se = FALSE, fitted = FALSE)

## S3 method for class 'roi'
print(x, ...)

## S3 method for class 'roi'
predict(object, x, fold = 5, ...)

## S3 method for class 'roi'
residuals(object, x, fold = 5, ...)
```

Arguments

x	Data for training the model.
xnew	Data at new locations (Validation set).
nk	Number of sites in the neighborhoods. If nk is a list, the first element represents a vector of the neighborhood size for all sites used for training. The second element .
phy	Formula defining the physical descriptors.
similarity	Formula defining the covariates used to evaluate the similarity between site, i.e. Euclidean distance to the target.

kriging	Formula defining the spatial covariates. Necessary for predicting residuals using spatial correlation.
model	Variogram model. See vgm
ker	Should a (Epanechnikov) kernel be used in to weight local regression model. Otherwise uniform weights are used.
se	Logical. Should the standard error be returned.
fitted	Logical. Should the fitted value be returned. Useful to obtain fitted values at gauged sites from a ROI model when kriging is not used.
...	Other parameters.
object	Output from FitRoi.
fold	Number of group or group used to perform cross-validation.

Value

pred	Prediction at new sites.
pred.se	Standard deviation at new sites.
phy	Part of the prediction attributed to physical descriptor.
phy.se	Standard deviation associated with the physical descriptor.
fitted	Fitted values (training sites).
fitted.se	Standard deviation of the fitted values.
vgm	Sample variogram.
model	Fitted variogram model.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

References

Martin Durocher, Donald H. Burn, Shabnam Mostofi Zadeh & Fahim Ashkar (2019) Estimating flood quantiles at ungauged sites using nonparametric regression methods with spatial components, Hydrological Sciences Journal, 64:9, 1056-1070, https://doi.org/10.1080/02626667.2019.1620952

See Also

CvRoi, krige, lm

Examples

```
data(flowUngauged)

## Using multidimensional scaling for projecting coordinates
coord <- flowUngauged[,c('lon','lat')]
coord <- cmdscale(GeoDist(coord))
colnames(coord) <- c('lon','lat')
```

```
## Transform data if necessary
xdf <- with(flowUngauged,
         data.frame(y      = l1,
                    area   = scale(log(area)),
                    wb     = scale(log(wb)),
                    stream = scale(log(stream)),
                    map    = scale(log(map)),
                    coord))

 ## select a validation and training set
 set.seed(9382)
 vid <- runif(nrow(xdf)) > .8
 tid <- !vid

 # formula of the relationship between flood quantile and descriptors
 fphy <- log(y) ~ area + map + poly(wb,3) + poly(stream,3)
 fsimilarity <- ~ area + map

 ## Fit a local regression model
 fit <- FitRoi(x = xdf[tid,], xnew = xdf[vid,], nk = 60,
               phy = fphy, similarity = fsimilarity)
 print(fit)
 response <- log(xdf[vid,'y'])
 sd(response - fit$pred)

 ## Refit the model and perform the kriging of the residuals
 fitk <- FitRoi(x = xdf[tid,],
               xnew = xdf[vid, ],
               nk = 60,
               phy = fphy,
               similarity = fsimilarity ,
               model = 'Exp',
               kriging = ~ lon + lat)

 print(fitk)
 sd(response - fitk$phy)
 sd(response - fitk$pred)
```

`flowAtlantic` *Annual maximums from sites in the Atlantic region of Canada*

Description

Contains the annual maximums of 45 hydrometric stations found in the region '01' of Water Survery of Canada. Additionaly to the annual maximums, the output list includes catchment descriptors (longitude, latitude, drainage area, mean annual precipitation) and the geographical distance between each station.

Usage

```
flowAtlantic
```

Format

An object of class `list` of length 2.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

Source

https://wateroffice.ec.gc.ca/

`flowStJohn` *Streamflow data*

Description

Daily river discharge for the station 01AD002 on St-John River at Fort Kent, New Brunswick. Data range from 1926 to 2014 and have a drainage area of 14700 sq km.

Usage

```
flowStJohn
```

Format

An object of class `data.frame` with 32234 rows and 2 columns.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

Source

https://wateroffice.ec.gc.ca/

`flowUngauged` *L-moments and catchment descriptors of hydrometric stations in Canada*

Description

The L-moments of the annual maximums at 562 stations were extracted from Water Survery of Canada (HYDAT). Catchment descriptors are available for each station was provided by Environment and Climate Change Canada (ECCC). The best at-site distributions according to the AIC criterion are included.

Usage

```
flowUngauged
```

Format

An object of class `data.frame` with 562 rows and 11 columns.

Details

site Identification number of the station in HYDAT.

area Drainage area (sq km).

map Mean annual precipitation for the catchment (mm).

wb Drainage area covered by waterbodies (pct).

stream Stream density (km of streams per catchment sq km).

lon,lat Longitude and latitude of the catchment center.

l1, lcv, lsk Sample mean, L-coefficient of variation and L-coefficient of skewness of the annual maximum discharges.

dstr Distribution selected for the annual maximum discharge based on the Akaike Information Criterion (AIC).

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

Source

https://wateroffice.ec.gc.ca/

`flow_raster` *Raster plot of streamflows*

Description

Produces a raster plot: years x day of year, showing magnitude of flows. This produces a plot showing the flow data in colours, showing different context than in a hydrograph. High flows are in warm colours

Usage

```
flow_raster(dframe, title = "", rastercolours = c("lightblue", "cyan",
  "blue", "slateblue", "orange", "red"))
```

Arguments

`dframe`	A data frame of hydrometric data. Must contain the variables `Date` and `Flow`.
`title`	The (optional) title for the plot
`rastercolours`	A vector of colours used for the raster plot. The default is `c("lightblue","cyan","blue","slateblue'`

Value

No value is returned; a standard R graphic is created.

Author(s)

Paul Whitfield <paul.h.whitfield@gmail.com>

See Also

`flow_raster_trend` `flow_raster_qa`

Examples

```
flow_raster(W05AA008)
```

flow_raster_qa *Raster plot of streamflows with WSC quality flags*

Description

Produces a raster plot of years x day of year showing the flow data in grayscale overlain by the Water Survey of Canada quality flags. Colours are consistent with ECDataExplorer. Raster layout lets the use see the flags in a different context than in a hydrograph. The data flags are:

A (Partial) green

B (Below Ice) blue

D (Dry) yellow

E (Estimated) red

Usage

```
flow_raster_qa(dframe, title = "")
```

Arguments

dframe	A data frame of WSC hydrometric data. Must contain the variables Date, Flow and SYM which is the WSC data flag symbol.
title	The (optional) title for the plot

Value

No value is returned; a standard R graphic is created.

Author(s)

Paul Whitfield <paul.h.whitfield@gmail.com>

See Also

flow_raster_trend flow_raster

Examples

```
flow_raster_qa(W05AA008, "Station W05AA008")
```

`flow_raster_trend` *Raster plot and simple trends of observed streamflows*

Description

Creates a raster plot plus trend plots for day of year, and over time which may be binned by a number of days

Usage

```
flow_raster_trend(date, flow, step = 5, stationID = "", title = "",
  missing = FALSE, colours = c("lightblue", "cyan", "blue",
  "slateblue", "darkblue", "red"))
```

Arguments

`date`	A numeric vector of the date as an R date. Must be same length as the `flow`
`flow`	A numeric vector of daily streamflows
`step`	An integer indicating the degree of smoothing eg. 1, 5, 11.
`stationID`	Station ID number, e.g. "05BB001". This value is optional, but is included in the output to help you identify the results.
`title`	Title of the plot
`missing`	If `FALSE` years with missing data are excluded. If `TRUE` partial years are included.
`colours`	A vector of colours used for the raster plot. The default is `c("lightblue","cyan","blue","slateblue'`

Details

The plot contains four panels based upon binned data:

1. The annual maximum, minimum, and median dlow with a trend test number for each period: red arrows indicate decreases, blue arrows indicate increases.
2. The scale bar for the colours used in the raster plot,
3. The rasterplot with a colour for each period and each year where data exist, and
4. A time series plot of the minimum, median, and maximum annual bin values. If there is no trend ($p > 0.05$) the points are black. Decreasing trend are in red, increasing trends are in blue..

Value

a list containing:

stationID Station ID eg 05BB001

missing How missing values were used FALSE=used, TRUE=removed

step number of days in a bin

periods number of periods in a year
period period numbers i.e. 1:365/step
bins values for each period in each year
med_period median for each period
max_period maximum for each period
min_period minimum for each period
tau_period Kendall's Tau for each period
prob_period probability of Tau for each period
year years spanning the data
median_year median bin for each year
max_year maximum bin for each year
min_year minimum bin for each year
tau_median_year value of tau and probability for median_year
tau_maximum_year value of tau and probability for max_year
tau_minimum_year value of tau and probability for min_year

Author(s)

Paul Whitfield <paul.h.whitfield@gmail.com>

See Also

flow_raster

Examples

```
mdoy <- doys(W05AA008$Date)
mplot <- flow_raster_trend(W05AA008$Date, W05AA008$Flow, step=5, station="05AA008")
```

GeoDist *Great-circle distance*

Description

Returns the distance matrix of the great-circle distance between all pairs of sites.

Usage

```
GeoDist(x, ...)

## Default S3 method:
GeoDist(x, earth.radius = 6371, ...)

## S3 method for class 'formula'
GeoDist(form, x, ...)
```

Arguments

x	Dateset of coordinates (lon,lat).
...	Other parameters.
earth.radius	Radius of the earth. Default 6371 km.
form	Formula to identify coordinates in a dataset. Must be of the form : lon~lat.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

Examples

```
attach(flowAtlantic)

GeoDist(lon~lat, flowAtlantic$info)[1:5,1:5]
```

---

get_AHCCD_monthly *Retrieve AHCCD data from EC datamart*

Description

Retrieve AHCCD data from EC datamart

Usage

```
get_AHCCD_monthly(station, province, variable,
  url = "http://dd.weather.gc.ca/climate/ahccd/geojson/historical/monthly/")
```

Arguments

station	Required. The station number - either as numeric or as a string.
province	Required. Name of province/territory. Must one of AB, BC, MB, NB, NL, NS, NT, NU, ON, PE, QC, SK, YT.
variable	Required. Must be one of **variable** meaning **PCP** total precipitation **RA** rainfall **SN** snowfall **TMAX** max air temp **TMEAN** mean air temp **TMIN** max air temp **PSFC** surface air pressure

SFCWND surface wind speed

SLP sea level pressure

`url` Required. The default url currently works to access the data on the Environment Canada server. The url can be changed in case the site is moved.

Value

Returns a data frame with the monthly values and associated variables.

Note

Not all variables are available at all stations. Attempting to retrieve a non-existent variable will result in an error being returned.

Author(s)

Kevin Shook

References

Use of the data must cite *Mekis, E and L.A. Vincent, 2011: An overview of the second generation adjusted daily temperature and precipitation dataset for trend analysis in Canada. Atmosphere-Ocean, 49 (2), 163-177.*

See Also

`read_AHCCD_daily` `read_AHCCD_monthly`

Examples

```
stoon_monthly_precip <- get_AHCCD_monthly("4057120", "SK", "PCP")
```

---

get_peaks *Extracts peak flows over a threshold*

Description

This function is development code being shared as is. It is expected that the user will be interested in the dataframe returned for POT analysis and for plotting (i.e. Booth_plots).

Usage

```
get_peaks(dataframe, threshold)
```

Arguments

`dataframe`	a data frame of streamflow data containing columns named 'Date' and 'Flow'
`threshold`	a value for the threshold. Values above the threshold are tested for peaks.

Details

This function retrieves peaks greater than the prescribed threshold. It returns a dataframe of peak characteristics suitable for subsequent analysis.

The portion under development is the It also returns a list of the flows during an event with the values of the three preceeding dates and three subsequnt dates.

Value

a list containing:

POTevents a dataframe contining details of the events

events a vector with the value 0 when the flow is below the threshold and 1 when above.

event_num a vector with the value 0 when the flow is below a threshold or the index of the events when the threshold was exceeded. i.e. 1,2,3, etc

st_date start date of events

case a list of the flows in each individual event (see details for more information)

The `POTevents` dataframe contains five columns: st_date (starting date), max_date (date of maximum in the event), max (maximum discharge), volume (volume of the event), and duration (in days).

The `case` list contains the flows during an event and also for three preceeding and subsequent days. The lists range from seven to n days in length.

References

Burn, D.H., Whitfield, P.H., Sharif, M., 2016. Identification of changes in floods and flood regimes in Canada using a peaks over threshold approach. Hydrological Processes, 39: 3303-3314. DOI:10.1002/hyp.10861

Whitfield, P.H., and J.W. Pomeroy. 2016. Changes to flood peaks of a mountain river: implications for analysis of the 2013 flood in the Upper Bow River, Canada. Hydrological Processes 30:4657-73. doi: 10.1002/hyp.10957.

See Also

`booth_plot`

Examples

```
threshold <- 0.9*max(W05AA008$Flow)  # arbitrary threshold
my_peaks <- get_peaks(W05AA008, threshold)
str(my_peaks)
```

get_wscstation *Reads station information from a data file produced by ECDE Retrieves station information for an individual Water Survey of Canada site, adds a text string at position 21 that combines key elements for a title.*

Description

Reads station information from a data file produced by ECDE Retrieves station information for an individual Water Survey of Canada site, adds a text string at position 21 that combines key elements for a title.

Usage

```
get_wscstation(stnID, stn)
```

Arguments

stnID	A Water Survey of Canada station number
stn	a data frame of station information from ECDataExplorer. The data frame 'HYDAT_list' is supplied with this package.

Value

Returns a data frame with 21 variables

- Station StationID
- StationName Station Name
- HYDStatus Active or Discontinued
- Prov Province
- Latitude
- Longitude
- DrainageArea km2
- Years # of years with data
- From Start Year
- To End Year
- Reg. Regulated
- Flow if TRUE/Yes
- Level if TRUE/Yes
- Sed if TRUE/Yes
- OperSched Continuous or Seasonal
- RealTime if TRUE/Yes

- RHBN if TRUE/Yes is in the reference hydrologic basin network
- Region if TRUE/Yes is in the reference hydrologic basin network
- Datum if TRUE/Yes is in the reference hydrologic basin network
- Operator if TRUE/Yes is in the reference hydrologic basin network
- Station_lname Added field combines ID,Name,Province and if RHBN an * is added

Author(s)

Paul Whitfield <paul.h.whitfield@gmail.com>

Examples

```
df <- HYDAT_list
s_info <- get_wscstation("05BB001", df)
title <- s_info[21]
print(title)
```

`GofTest` *Goodness-of-fit test*

Description

Return the statistic and p-value of a goodness-of-fit test on AMAX and POT models. The null hypothesis (composite) is that the data were generated by the fitted distribution.

Usage

```
GofTest(object, ...)

## S3 method for class 'amax'
GofTest(object, method = "ad", nsim = 1000, ...)

## S3 method for class 'fpot'
GofTest(object, method = "adtab", nsim = 1000, ...)
```

Arguments

`object`	Output from `FitAmax` or `FitPot`.
`...`	Other parameters.
`method`	Test to be performed. Either Anderson-Darling ad, or modified Shapiro-Wilk ('shapiro').
`nsim`	Number of simulations.

Details

For a POT model, the method `adtab` perform the Anderson-Darling test and interpolates the p-value from a table.

References

Choulakian, V., Stephens, M.A., 2001. Goodness-of-Fit Tests for the Generalized Pareto Distribution. Technometrics 43, 478-484. https://doi.org/10.2307/1270819

Heo, J.-H., Shin, H., Nam, W., Om, J., Jeong, C., 2013. Approximation of modified Anderson-Darling test statistics for extreme value distributions with unknown shape parameter. Journal of Hydrology 499, 41-49. https://doi.org/10.1016/j.jhydrol.2013.06.008

Ba, I., Ashkar, F., 2017. Discrimination between a group of three-parameter distributions for hydrometeorological frequency modeling. Can. J. Civ. Eng. 45, 351-365. https://doi.org/10.1139/cjce-2017-0416

Fahim Ashkar & Ba, I. (2017) Selection between the generalized Pareto and kappa distributions in peaks-over-threshold hydrological frequency modelling, Hydrological Sciences Journal, 62:7, 1167-1180, http://dx.doi.org/10.1080/02626667.2017.1302089

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

Examples

```
## The same simulated distribution should be accepted p-value >.05
x <- rAmax(100, c(100,30,-.1), 'gev')

fit <- FitAmax(x, 'gev', varcov = FALSE)
GofTest(fit, nsim = 50) ## nsim is small for speeding-up the demonstration

## The same apply to POT model
fit <- FitPot(flow~date, flowStJohn, u = 1000,
               declust = 'flood', r = 14)

## By default a table is used with GPA using MLE
GofTest(fit)
```

GPA *Generalized Pareto distribution (GPA)*

Description

Distribution, density, quantile and random function for the Generalized pareto distriution.

Usage

```
pgpa(q, alpha = 1, kap = 0, lower.tail = TRUE)

rgpa(n, alpha = 1, kap = 0)

dgpa(x, alpha = 1, kap = 0, log = FALSE)

qgpa(p, alpha = 1, kap = 0, lower.tail = TRUE)
```

Arguments

q, x	Vector of quantiles.
alpha	Scale parameter of the GPA
kap	Shape parameter of the GPA
lower.tail	Logical. Should the propability of the lower tail be returned
n	Number of simulations.
log	Should the log-density be returned
p	Vector of probabilities.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

References

Davison, A. C., & Smith, R. L. (1990). Models for Exceedances over High Thresholds. Journal of the Royal Statistical Society. Series B (Methodological), 52(3), 393-442. http://www.jstor.org/stable/2345667

Examples

```
kap <- -.2
a <- 1
xd1 <- rgpa(1e4, a, kap)
xd2 <- qgpa(runif(1e4), a, kap)

qqplot(xd1, xd2)


tt <- seq(0.001,6, len = 100)

hist(xd1[xd1<6], main = 'GPA distribution',
  freq = FALSE, ylim =c(0,1), xlim = c(0,6))

lines(tt, dgpa(tt,a,kap))
lines(tt,pgpa(sort(tt), a, kap), col = 2, lty = 2)
```

HYDAT_list *HYDAT_list*

Description

A dataframe of station information, as extracted from the EC Data Explorer

Usage

```
HYDAT_list
```

Format

A dateframe with 7791 rows and 20 columns.

Source

Water Survey of Canada

Variables:

- Station StationID
- StationName Station Name
- HYDStatus Active or Discontinued
- Prov Province
- Latitude
- Longitude
- DrainageArea km2
- Years # of years with data
- From Start Year
- To End Year
- Reg. Regulated
- Flow if TRUE/Yes
- Level if TRUE/Yes
- Sed if TRUE/Yes
- OperSched Continuous or Seasonal
- RealTime if TRUE/Yes
- RHBN if TRUE/Yes is in the reference hydrologic basin network
- Region if TRUE/Yes is in the reference hydrologic basin network
- Datum if TRUE/Yes is in the reference hydrologic basin network
- Operator if TRUE/Yes is in the reference hydrologic basin network
- Station_lname Added field combines ID,Name,Province and if RHBN an * is added

`hydrograph_plot` *Plot hydrographs*

Description

Creates a hydrograph plot for simulated, observed, and inflow hydrograph series, including precipitation if provided. The secondary y axis will be used to plot the precip time series. The function assumes that the supplied time series have the same length and duration in time. If this is not true, then the defined period or period calculated from the first available flow series will be used to determine the plotting limits in time. If the data is take from output from the **Raven** model, this is not a concern. The supplied time series should be in **xts** format, which can be obtained directly by using the `hyd.extract` function in the package **RavenR** . Note that a plot title is purposely omitted in order to allow the automatic generation of plot titles.

Usage

```
hydrograph_plot(flows = NULL, precip = NULL, prd = NULL,
  winter_shading = FALSE, range_mult_flow = NULL,
  range_mult_precip = 1.5, flow_labels = NULL,
  ylabel = "Flow [m3/s]", precip_label = "Precipitation [mm]",
  leg_pos = NULL, leg_box = NULL, zero_axis = T,
  plot_mode = "base")
```

Arguments

`flows`	data frame of flows to plot
`precip`	data frame of precipitation values to plot
`prd`	period to use in plotting
`winter_shading`	optionally adds a transparent cyan shading for the December 1st to March 31st period in each year that is plotted. Default is `FALSE`.
`range_mult_flow`	range multiplier for max value in hydrograph. This is useful in preventing overlap if precip is also plotted. This value should not be less than 1.0, otherwise the values will be cutoff in the plot.
`range_mult_precip`	range multiplier for max value in precipitation plot (default 1.5)
`flow_labels`	string vector of labels for flow values
`ylabel`	text label for y-axis of the plot (default 'Flow [m3/s]')
`precip_label`	text label for precipitation y-axis (default 'Precipitation [mm]')
`leg_pos`	string specifying legend placement on plot e.g. '`topleft`', '`right`', etc., and is consistent with the legend function options. If `NULL`, the function will place the legend left, if precip added, on the topleft otherwise).
`leg_box`	boolean on whether to put legend in an opaque white box or not. If `NULL` (the default), the function will automatically not use a white box and leave the background of the legend transparent.

zero_axis	fixes the y axis to start exactly at zero (default TRUE). By default, R will plot the values with a small buffer for presentation. Be warned that if this option is set to TRUE, the minimum value is set to zero without checking if any flow values are less than zero. This option should not be used for reservoir stage plotting, since most reservoir stage is typically reported as an elevation.
plot_mode	plot mode as 'base' or 'ggplot'. Currently only 'base' plot type is supported, 'ggplot' is under construction.

Value

Returns TRUE if the function is executed properly.

Author(s)

Robert Chlumsky <rchlumsk@gmail.com>

Examples

```
# example with synthetic random data
dd <- seq.Date(as.Date("2010-10-01"), as.Date("2013-09-30"),by = 1)
x <- abs(rnorm(length(dd)))
y <- abs(rnorm(length(dd))) * x
df <- data.frame("Date" = dd, x, y)
myprd <- "2011-10-01/2012-09-30"

precip <- data.frame("Date" = dd," precip" = abs(rnorm(length(dd))) * 10)

# basic hydrograph plot
hydrograph_plot(flows = df, winter_shading = FALSE)

# with different labels
hydrograph_plot(flows = df, winter_shading = FALSE, flow_labels = c("simulated", "observed"))

# with a few more options turned on
hydrograph_plot(flows = df, precip = precip)

# increase the plot ranges to separate flows and precip; add a legend box
hydrograph_plot(flows = df, precip = precip, range_mult_flow = 1.7,
range_mult_precip = 2, leg_box = TRUE)
```

Intersite *Estimating intersite correlation*

Description

Return a matrix of intersite correlation between paired observations. By default the empirical matrix is estimated and corrected to be positive definite. Can also be estimated by a power exponential model where the weighted least squares approach is used with weights proportional to record lengths.

Usage

```
Intersite(x, method = "emp", distance = NULL, nmin = 0,
  na.sub = "avg", defpos = TRUE, smooth = 1, distance.max = Inf,
  start = NULL)

## S3 method for class 'isite'
print(x, ...)

IntersiteMdl(para, distance)
```

Arguments

x	Dataset in the wide format
method	Estimation method. Can be either 'emp' for estimating the empirical matrix or 'exp' for fitting power exponential model.
distance	Matrix of distances. Necessary when using 'exp'.
nmin	Minimal number of pairs necessary to compute pairwise correlations.
na.sub	Value for imputing missing pairwise correlation with 'emp'.
defpos	Logical. Should a correction be applied to the empirical matrix to ensure positive definiteness.
smooth	Smooth parameter of the power exponential model.
distance.max	Maximal distance to consider for paired observations in the fitting of the power exponential model.
start	Initial parameter for optimization with the exp method. It must have the form (nugget,range).
...	Other parameters.
para	Paramter of the power exponential model.

Value

method Estimation method.

para Parameter of the fitted model. For 'emp' it is the average.

corr Raw estimation of the correlations.

model Final estimate of the correlations.

rmse For 'exp', root mean square errors of the fitted model.

References

Durocher, M., Burn, D.H., Mostofi Zadeh, S., 2018. A nationwide regional flood frequency analysis at ungauged sites using ROI/GLS with copulas and super regions. Journal of Hydrology 567, 191-202. https://doi.org/10.1016/j.jhydrol.2018.10.011

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

Examples

```
data(flowAtlantic)

## Organize annual maximums
ams <- flowAtlantic$ams
ams$year <- format(ams$date, '%Y')
xmat <- DataWide(ams ~ id + year, ams)

## Compute distance between sites
coord <- flowAtlantic$info[, c('lon','lat')]
rownames(coord) <- flowAtlantic$info$id
h <- GeoDist(coord)

## make sure that columns match
h <- h[colnames(xmat), colnames(xmat)]

## estimate intersite correlation using a model
isite <- Intersite(xmat, distance = h, method = 'exp')
print(isite)
plot(isite, xmat, h)

## Evaluate emprirical correlation matrix
isite <- Intersite(xmat[,1:5])
print(isite)
round(isite$model,2)
```

`JulianPlot` *Plotting Julian date*

Description

Create axis for plotting circular statistics in a unitary circle.

Usage

```
JulianPlot(rose.col = "gray40", rose.lwd = 1.5, rose.cex = 1.5,
  rose.radius = seq(0.25, 1, 0.25), ...)
```

Arguments

`rose.col, rose.lwd, rose.cex`	Property of the polar axes.
`rose.radius`	Vector of the position of the circular axis.
`...`	Other parameter passed to points.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

See Also

SeasonStat.

Examples

```
data(flowAtlantic)

ss <- SeasonStat(date ~ id, flowAtlantic$ams)

JulianPlot()
points(y ~ x, ss, pch = 16, col = cut(ss[,'radius'], c(0,.5,.75,1)))
```

---

plot.amax *Return level plot*

Description

Create a plot of the estimated versus sample flood quantiles, where the x-axis is expressed in terms of return period.

Usage

```
## S3 method for class 'amax'
plot(x, main = "Return level plot",
  xlab = "Return period (year)", ylab = "Flood quantiles",
  ci = FALSE, col.ci = "red", lty.ci = 2, lwd.ci = 1, ...)
```

Arguments

x	Output from FitAmax.
main, ylab, xlab	Graphical parameters. See par.
ci	Logical. Should confidence intervals be displayed. See predict.amax with argument Delta method.
col.ci, lty.ci, lwd.ci	Graphical parameters determining the confidence intervals.
...	Other graphical parameters. See par.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

Examples

```
data(flowStJohn)

x <- ExtractAmax(flow~date, flowStJohn)$flow

fit <- FitAmax(x, distr = 'gev', method = 'mle')

plot(fit, ci = TRUE)
```

`plot.isite` *Plot of the pairwise intersite correlation.*

Description

Produce a graphic of the pairwise intersite correlation coefficient with respect to the distance.

Usage

```
## S3 method for class 'isite'
plot(x, xmat, distance, xlab = NULL, ylab = NULL, ...)
```

Arguments

`x`	Output from `Intersite`.
`xmat`	Dataset in wide format.
`distance`	Matrix of distances.
`xlab, ylab`	Label of the xy-axis.
`...`	Other parameters.

See Also

Intersite, DataWide.

Examples

```
#' Data(flowAtlantic)

## Organize annual maximums
ams <- flowAtlantic$ams
ams$year <- format(ams$date, '%Y')
xmat <- DataWide(ams ~ id + year, ams)

## Compute distance between sites
coord <- flowAtlantic$info[, c('lon','lat')]
rownames(coord) <- flowAtlantic$info$id
```

```
h <- GeoDist(coord)

## make sure that columns match
h <- h[colnames(xmat), colnames(xmat)]

## estimate intersite correlation using a model
isite <- Intersite(xmat, distance = h, method = 'exp')
print(isite)
plot(isite, xmat, h)
```

`PlotMrl` *Mean residual life plot*

Description

Produce a mean residual Life plot to help selecting to help selecting a proper threshold.

Usage

```
PlotMrl(form, x, u, declust = NULL, r = 1, rlow = 0.75,
  alpha = 0.05, ylab = "Mean Residual Life", xlab = "Threshold",
  col = "black", lty = 1, lwd = 1, col.ci = "black", lty.ci = 3,
  lwd.ci = 1, ylim = NULL, display = TRUE, ...)
```

Arguments

`form, x`	Formula and sample. The must be of the form.
`u`	Series of candidate thresholds.
`declust, r, rlow`	Parameter for declustering. See FitPot.
`alpha`	Confidence interal with level `1-alpha/2`.
`col.ci, lty.ci, lwd.ci`	Parameter for controling the confidence intervals.
`ylim, ylab, xlab, col, lty, lwd`	Parameters for controling the graphic.
`display`	Logical. Should the graph be display.
`...`	Other arameters for controling the graphic.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

See Also

FitPot, which.floodPeaks.

Examples

```
## Find list of candidate thresholds
lstu <- seq(500,2500, len = 50)

PlotMrl(flow~date, flowStJohn, u = lstu, declust = 'wrc', r = 14)

x <- PlotMrl(flow~date, flowStJohn, u = lstu, declust = 'wrc', r = 14)

head(x)
```

`PlotThresh` *Visual diagnostic for peaks over threshold*

Description

Create various graphics to assess the selection of a threshold. Includes a plot with respect oh the threshold for the p-values for the Anderson-Darling test, estimated parameter and flood quantiles. Vertical lines are shown for thresholds associated to 1 and 1.5 peaks per years.

Usage

```
PlotThresh(obj, type = "ad", ppy = NULL, ...)
```

Arguments

`obj`	Output from SearchThresh.
`type`	Type of plot to display. Must be one or more of `'ad'`,`'mrl'`,`'alpha'`,`'kappa'`.
`ppy`	Range of peaks per year to display.
`...`	Other parameters.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

See Also

SearchThresh, FindThresh.

Examples

```
lstu <- seq(500,2000,20)
out <- SearchThresh(flow~date, flowStJohn, u = lstu, declust = 'wrc', r = 14)

PlotThresh(out)
```

```
PlotThresh(out, type = 'q50')

PlotThresh(out, type = paste0('q',c(2,5,10,20,50,100)))

par(mfrow = c(2,2))
PlotThresh(out, type = c('ad','kappa', 'mrl','q10'))
```

`polar_plot` *Polar plot of daily streamflows*

Description

Produces a polar plot similar to that used in *Whitfield and Cannon, 2000.* It uses output from the function `binned_MannWhitney` or a data structure created using the function `polar_plot_prep`.

Usage

```
polar_plot(bmw, lcol1 = c("black", "gray50"), lcol2 = c("black",
  "gray50"), lfill = c("yellow", "green"), lsig = c("red", "blue"))
```

Arguments

`bmw`	output from binned_MannWhitney
`lcol1`	line colour, default is c("black","gray50")
`lcol2`	point colour, default is c("black","gray50")
`lfill`	fill colour, default is c("yellow","green")
`lsig`	significance symbol colour, default is ("red","blue")

Value

No value is returned; a standard R graphic is created.

Author(s)

Paul Whitfield <paul.h.whitfield@gmail.com>

References

Whitfield, P.H. and A.J. Cannon. 2000. Polar plotting of seasonal hydrologic and climatic data. Northwest Science 74: 76-80.

Whitfield, P.H., Cannon, A.J., 2000. Recent variations in climate and hydrology in Canada. Canadian Water Resources Journal 25: 19-65.

See Also

`binned_MannWhitney` `polar_plot_prep`

Examples

```
range1 <- c(1970,1979)
range2 <- c(1990,1999)
b_MW <- binned_MannWhitney(W05AA008, step = 5, range1, range2,
ptest = 0.05, station_ID = "05AA008", station_name = "Crowsnest River at Frank")
polar_plot(b_MW)
```

---

polar_plot_prep *Creates a data structure to be passed to* polar_plot.

Description

Could be used to move data from a different type of analysis different to the binned_MannWhitney function which uses flows. The two series need to be of the same length and their length is related to the step size. For examples 73 periods links to 5 day periods.

Usage

```
polar_plot_prep(station, plot_title, step, x0, x1, stat, prob, test_s,
  variable = "discharge", bin_method = "unstated",
  test_method = "unstated", lline1 = "Period 1", lline2 = "Period 2",
  pvalue = 0.05)
```

Arguments

station	Typically a station number
plot_title	Polar plot title - usually a station name
step	The number of days binned
x0	Time series of length n for a single seasonal cycle
x1	Time series of length n for a single seasonal cycle
stat	Time series of length n for statistical test value for each bin
prob	Time series of length n of probability of test value
test_s	Vector with values of -1, 0, 1 for significance, -1 negative, 1 positive, 0 not significant
variable	Name of variable plotted. Default is "discharge"
bin_method	Default is "unstated"
test_method	Default is "unstated"
lline1	Names of first period, default is "Period 1"
lline2	Names of second period, default is "Period 2"
pvalue	Value of p used. Default is 0.05

Value

Returns a list containing:

StationID ID of station

Station_lname Name of station

variable Name of variable

bin_width Smoothing time step

range1 range1 years

range2 range2 years

p_used p value used

fail TRUE if test failed due to missing values

bin_method method used for binning

test_method Mann-Whitney U

series a data frame containing:

period period numbers i.e. 1:365/step

period1 median values for each bin in period 1

period2 median values for each bin in period 2

mwu Mann Whitney U-statistic for each bin between the two periods

prob probability of U for each period

code significance codes for each bin

Author(s)

Paul Whitfield <paul.h.whitfield@gmail.com>

References

Whitfield, P.H. and A.J. Cannon. 2000. Polar plotting of seasonal hydrologic and climatic data. Northwest Science 74: 76-80.

Whitfield, P.H., Cannon, A.J., 2000. Recent variations in climate and hydrology in Canada. Canadian Water Resources Journal 25: 19-65.

See Also

`binned_MannWhitney` `polar_plot`

PoolRemove *Remove heterogenous sites from a pooling group*

Description

Return a regional model (`'FitRegLmom'`) where heterogenous sites are removed in a stepwise manner. At each step, the removed site that best improves the homogeneity statistic is permanently removed until a stopping criterion. The first site that was passed in a dataset is considered as the target and it cannot be removed.

Usage

```
PoolRemove(obj, method = "H1", tol = 2, nmin = 15, ntot.min = 0,
  nsim = 1000, distr.fix = FALSE, verbose = TRUE)
```

Arguments

`obj`	An output from FitRegLmom.
`method, tol`	Which heterogenity statistics used in the procedure. The choices are `'H1'`, `'H2'` and `'H3'`. The algorithm stop when the heterogenity goes below the value `tol`.
`nmin, ntot.min`	Additional stopping criteria. Respectively the minimal number of sites or station-year to be included in the region.
`nsim`	Number of simulations used to evaluate the heterogenous statistic
`distr.fix`	Logical, should the selection of the distribution be re-evaluated after removing the site.
`verbose`	Logical. Should a trace of the removed sites be display.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

Examples

```
data(flowAtlantic)

## Organize data
ams <- flowAtlantic$ams
ams$year <- format(ams$date, '%Y')
xmat <- DataWide(ams ~ id + year, ams)

## Compute distance
coord <- flowAtlantic$info[,c('lon','lat')]
rownames(coord) <- flowAtlantic$info$id
h <- GeoDist(coord)

## Fit a index flood model inside a pooling group
```

```
xmat0 <- FindNearest(xmat, h[1, colnames(xmat)], 25)
fit <- FitRegLmom(xmat)

## Remove site until homogeneity is reached
fit.new <- PoolRemove(fit, tol = 2)

## Get the data from the pooling group
xmat1 <- xmat[,sitenames(fit.new)]
```

`predict.amax` *Predict return levels*

Description

Return the flood quantile of annual maximum distribution and Confident intervals are provided by bootstrap.

Usage

```
## S3 method for class 'amax'
predict(object, q = c(0.5, 0.8, 0.9, 0.95, 0.98, 0.99),
  se = FALSE, ci = "none", alpha = 0.05, nsim = 1000,
  out.matrix = FALSE, ...)
```

Arguments

`object`	Output from `FitAmax`
`q`	Probabilities associated to the return level. For example, a 100 years return period is equivalent to `q = 0.99`.
`se`	Return the standard deviation of the return level using the delta method. The fitted model must
`ci`	Method to compute the confident interval. One of `'delta'` for the delta method, `'boot'` for parametric boostrap and `'norm'` for Monte-Carlo approximation assuming normality of the parameters.
`alpha`	Probability outside the confident interval.
`nsim`	Number of simulation use for resampling.
`out.matrix`	Logical. Should the resampling be returned. If true, a list is returned containing the prediction table (`pred`), the parameters (`para`) and the return levels (`qua`).
`...`	Other parameters.

Examples

```
#' ## Extract an time series of annual maxima
x <- ExtractAmax(flow~date, flowStJohn)$flow

## Fitting of GEV distribution using L-moments
fit <- FitAmax(x,'gev', method = 'mle')

## Get the estimated quantile of 10 and 100 years return period
rp <- 1-1/c(10,100)
predict(fit, rp)
predict(fit, se = TRUE, ci = 'delta')

## The bootstrap sample used for CI are returned
fit <- FitAmax(x,'gev', varcov = FALSE)
boot <- predict(fit, rp, se = FALSE, ci = 'boot',
                nsim = 500, out.matrix = TRUE)
```

`predict.fpot` *Prediction of flood quantiles of a given return period.*

Description

Return a vector or matrix of the flood quantiles, its standard deviation and confidence intervals.

Usage

```
## S3 method for class 'fpot'
predict(object, rt = c(2, 5, 10, 20, 50, 100),
  se = FALSE, ci = "none", alpha = 0.05, nsim = 1000, ...)
```

Arguments

`object`	Output of FitPot.
`rt`	Return period.
`se`	Logical. Should the standard deviation be returned.
`ci, alpha`	Method for evaluation the confidence intervals with probability `1-alpha`. Available methods are : Delta method (`'delta'`), profile likelihood (`'profile'`) and nonparametric bootstrap (`'boot'`)
`nsim`	Number of bootstrap sample.
`...`	Other parameters.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

References

Coles S. (2001) An introduction to statistical modeling of extreme values. Springer Verlag.

See Also

FitPot

Examples

```
data(flowStJohn)

fit <- FitPot(flow~date, flowStJohn, u = 1000,
              declust = 'wrc', r = 14)

predict(fit, se = TRUE, ci = 'delta')
```

`predict.reglmom` *Flood quantiles estimates*

Description

Predict the flood quantile of index-flood model for a specific scale factor. By default the flood quantile of the first site (target) is returned.

Usage

```
## S3 method for class 'reglmom'
predict(object, q = c(0.5, 0.8, 0.9, 0.95, 0.98, 0.99),
  ci = FALSE, corr = 0, nsim = 1000, alpha = 0.05, ...)
```

Arguments

`object`	An output from FitRegLmom.
`q`	Probability associated to the flood quantiles.
`ci`	Logical. Should the confident intervals and the standard deviation be evaluated?
`corr`	Intersite correlation. Either a matrix or a constant coefficient for all pairs
`nsim`	Number of simulations used for approximating the confident intervals.
`alpha`	Probability outside the confidence intervals.
`...`	Other parameters.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

References

Hosking, J. R. M., & Wallis, J. R. (1997). Regional frequency analysis: an approach based on L-moments. Cambridge Univ Pr.

See Also

FitRegLmom, Intersite.

Examples

```
data(flowAtlantic)

h <- GeoDist(flowAtlantic$info[,c('lon','lat')])
ams <- flowAtlantic$ams
ams$year <- format(ams$date, '%Y')
xd <- DataWide(ams ~ id + year, ams)

xd <- FindNearest(xd, h[1,], 25)
h  <- FindNearest(h , h[1,], 25, row = TRUE)

## Fit the regional model
fit <- FitRegLmom(xd)

## estimate flood quantiles
predict(fit, c(.3,.7))

## Evaluate intersite correlation
isite <- Intersite(xd)

predict(fit, ci = TRUE, corr = isite$model)
```

---

read_AHCCD_daily *Reads AHCCD daily file*

Description

This program reads an Adjusted and Homogenized Canadian Climate Data (AHCCD) of daily precipitation or temperatures. The values are arranged as month x day, which makes them difficult to read using standard R functions.

Usage

```
read_AHCCD_daily(daily_file)
```

Arguments

daily_file	Required. Name of the file to be read.

Value

If successful, returns the values in a dataframe, consisting of the date, the value and the data code. If unsuccessful, returns the value FALSE.

Author(s)

Kevin Shook

References

Monthly AHCCD data are available from http://www.ec.gc.ca/dccha-ahccd. Daily values must be requested. Any use of the data must cite *Mekis, E and L.A. Vincent, 2011: An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. Atmosphere-Ocean, 49 (2), 163-177.*

See Also

read_AHCCD_monthly get_AHCCD_monthly

Examples

```
## Not run:
stoon_daily_tmax <- read_AHCCD_daily("dx40657120.txt")
## End(Not run)
```

---

read_AHCCD_monthly *Reads AHCCD monthly file*

Description

This program reads an Adjusted and Homogenized Canadian Climate Data (AHCCD) data of precipitation or temperatures. The values are arranged as year x month, which makes them difficult to read using standard R functions.

Usage

```
read_AHCCD_monthly(monthly_file = NULL)
```

Arguments

monthly_file	Required. Name of the file to be read.

Value

If successful, returns the values in a dataframe, consisting of the year, the month, the value and the data code. The meanings of the codes can be found in the

Author(s)

Kevin Shook

References

Monthly AHCCD data are available from https://www.canada.ca/en/environment-climate-change/services/climate-change/science-research-data/climate-trends-variability/adjusted-homogenized-canad surface-air-temperature-access.html. Any use of the data must cite *Mekis, E and L.A. Vincent, 2011: An overview of the second generation adjusted daily temperature and precipitation dataset for trend analysis in Canada. Atmosphere-Ocean, 49 (2), 163-177.*

See Also

read_AHCCD_daily get_AHCCD_monthly

Examples

```
## Not run:
Stoon_monthly_precip <- read_AHCCD_monthly("mt4057120.txt")
NB_monthly_tmean <- read_AHCCD_monthly("mm4045695.txt")
## End(Not run)
```

---

read_ECDE_flows — *Reads a file of WSC daily flows from ECDataExplorer Reads in a file WSC daily flows as returned from the program ECDataExplorer, and omits the last 3 lines as these contain the data disclaimer.*

Description

Reads a file of WSC daily flows from ECDataExplorer Reads in a file WSC daily flows as returned from the program ECDataExplorer, and omits the last 3 lines as these contain the data disclaimer.

Usage

```
read_ECDE_flows(filename)
```

Arguments

filename	Datafile retrieved from ECDataExplorer

Value

Returns a dataframe with the last three rows removed and the Date as Date

Author(s)

Paul Whitfield <paul.h.whitfield@gmail.com>

Examples

```
mfile <- system.file("extdata", "04JD005_Daily_Flow_ts.csv", package = "CSHShydRology")
mdata <- read_ECDE_flows(mfile)
```

---

regime_plot *Plots the regime of daily streamflows*

Description

Produces a regime hydrograph similar to that in the reference. It shows the flow quantiles for each day of the year and the maximum and minimum. Parameters can be set to change colours and fix the y scale to allow plots of same scale to be produced.

Usage

```
regime_plot(date, flow, title = "", wyear = 1, colour = TRUE,
  mx = 1)
```

Arguments

date	Vector of dates
flow	Vector of daily streamflows. Must be the same length as date
title	Text to be used as the graph title
wyear	Beginning month of water year. USe wyear = 10 for October water year, wyear = 1 for calendar year
colour	Logical. If TRUE plot is in colour, if FALSE plot is grayscale
mx	The maximum y value; if mx = 1 then maximum value of the flows is used to set the maximum y-axis value. The value of mx can be specified to produce a series of plots with the same scale.

Value

No value is returned; a standard R graphic is created.

Author(s)

Paul Whitfield

References

MacCulloch, G. and P. H. Whitfield (2012). Towards a Stream Classification System for the Canadian Prairie Provinces. Canadian Water Resources Journal 37: 311-332.

Examples

```
regime_plot(W05AA008$Date, W05AA008$Flow, title = "05AA008", colour = TRUE, wyear = 10)
```

RegSim *Simulation of a regional dataset with intersite correlation.*

Description

Returns a dataset containing multiple time series in the form of a matrix where the sites are in columns. Different record lengths can be specified for each site and missing values are filled accordingly at the beginning. The rows (time) are independent and the intersite correlation model is based on a multivariate Normal distribution.

Usage

```
RegSim(x, ...)

## S3 method for class 'matrix'
RegSim(x, distr, nrec, corr = 0, corr.sqrt = FALSE,
  lmom = TRUE, lscale = FALSE, long = FALSE, ...)

## S3 method for class 'reglmom'
RegSim(x, n = 1, corr = 0, margin = "atsite", ...)
```

Arguments

x	Matrix (in rows) of parameters or L-moments for all sites to simulate. Can also be an output form FitRegLmom.
...	Other parameters.
distr	Marginal distribution of each site. If only one value is passed as argument, the same is used for all sites.
nrec	Record lengths of the sites. If only one value is passed in argument, the same is used for all sites.
corr	Correlation matrix for the dependence between site. If only one value is passed, the correlation is assumed the for every pair of sites.
corr.sqrt	Squared correlation matrix. Can be passed to speed up multiple calls.
lmom	Logical. Is the argument 'x' a matrix of L-moments or distribution parameters
lscale	Logical. Is the second L-moments the scale ('TRUE') or the LCV ('FALSE').
long	Logical. Should the output be returned in a long format.
n	Number of simulations.
margin	Which marginal distribution should be used. Either based on the at-site (`'atsite'`) or regional (`'reg'`) distribution.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

Examples

```
## Extract data
data(flowUngauged)
lmom <- flowUngauged[1:5, c('l1','lcv','lsk')]

## Simulate data base on at-site L-moments
sim <- RegSim(lmom, distr = 'gev', nrec = 11:15, corr = .4)

head(sim)

sim <- RegSim(lmom,
              distr = c(rep('gev',4),'gno'),
              nrec = 10:15,
              corr = .4, long = TRUE)
head(sim)
```

SearchThresh *Fit POT models for a list of candidate thresholds*

Description

The function SearchThresh returns a data frame containing key value for the automatic selection of a threshold and where each row corresponds to a candidate threshold. The result is extracted from several calls to the function FitPot.

Usage

```
SearchThresh(form, x, u, nmin = 20, verbose = TRUE, ...)
```

Arguments

x, form	Dataset and formula passed to FitPot.
u	Vector of candidate thresholds.
nmin	Stopping condition verifying that a minimal number of peaks are extracted.
verbose	Logical. Should a progress bar be displayed.
...	Other arguments passed to FitPot.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

Examples

```
# Create a list of candidate threshold
lstu <- which.floodPeaks(flow~date, flowStJohn, u =500, r = 14)
lstu <- sort(unique(flowStJohn[lstu,'flow']))
lstu <- lstu[seq(1,length(lstu)-30,2)]

out <- SearchThresh(flow~date, flowStJohn, u = lstu, declust = 'wrc', r = 14)
head(out)

FindThresh(out, method = 'sgn', tol.sgn = 0.1)
FindThresh(out, method = 'ppy', tol.ppy = 2)
FindThresh(out, method = 'max')
```

SeasonStat *Seasonal statistics for flood peaks*

Description

Return the circular or seasonal statistics of flood peaks. The angle represent the average timing of the floods and the radius its regularity. For instance a radius of one represent perfect regularity. Can perform the analysis of multiple sites.

Usage

```
SeasonStat(x, ...)

## S3 method for class 'data.frame'
SeasonStat(x, ...)

## S3 method for class 'formula'
SeasonStat(form, x, ...)
```

Arguments

x	Data. If data.frame with two columns, they must be respectively the date and a site variable.
...	Other parameters.
form	Formula that specifies the date and site variable. Must be of the form `date ~ site`.

Author(s)

Martin Durocher <mduroche@uwaterloo.ca>

References

Burn, D.H. (1997). Catchment similarity for regional flood frequency analysis using seasonality measures. Journal of Hydrology 202, 212-230. https://doi.org/10.1016/S0022-1694(97)00068-1

Examples

```
dt <- ExtractAmax(flow~date, flowStJohn)$date

SeasonStat(dt)

## Illustration of the analysis of multiple sites

F0 <- function(ii) data.frame(site = ii, dt = sample(dt, replace = TRUE))
x <- lapply(1:10, F0)
x <- do.call(rbind, x)

st <- SeasonStat(dt ~ site, x)

JulianPlot()
points(y ~ x, st, col = 2, pch = 16)
```

StatisticalHydrology-functions

Statistical analysis functions

Description

These functions perform statistical analyses

binned_MannWhitney Compares two time periods of data using Mann-Whitney test

fdcurve Finds flow exceedence probabilities

get_peaks Finds peak flows over a specified threshold

Visualization-functions

Visualization functions

Description

These functions are primarily intended for graphing, although some analyses may also be done.

booth_plot Plot of peaks over a threshold

flow_raster Raster plot of streamflows

flow_raster_qa Raster plot of streamflows with WSC quality flags

flow_raster_trend Raster plot and simple trends of observed streamflows

hydrograph_plot Plots hydrographs and/or precipitation

polar_plot Polar plot of daily streamflows

regime_plot Plots the regime of daily streamflows

`W05AA008` *W05AA008*

Description

A dataframe of Water Survey of Canada (WSC) daily flows for station W05AA008, CROWSNEST RIVER AT FRANK Alberta.

Usage

```
W05AA008
```

Format

A dateframe with 25252 rows and 5 columns spanning the period 1910-2013.

Source

Water Survey of Canada

Variables:

- ID StationID
- PARAMParameter 1=Flow, 2=Level
- DateR date
- FlowDaily flow in m^3/s
- SYMWater Survey FLags A, B, D, E

`which.floodPeaks` *Extracting peaks*

Description

Returns the indices of the peaks above a threshold according to the declustering method put in place by the Water Resources Council or simple run declustering. See Lang et al. (1999) for more details.

Usage

```
which.floodPeaks(x, ...)

## S3 method for class 'numeric'
which.floodPeaks(x, dt = NULL, u, r = 1,
  rlow = 0.75, ini = "wrc", ...)

## S3 method for class 'formula'
which.floodPeaks(form, x, u, ...)

## S3 method for class 'data.frame'
which.floodPeaks(x, u, ...)

which.clusters(x, ...)

## S3 method for class 'formula'
which.clusters(form, x, u, r = 1, ...)

## S3 method for class 'numeric'
which.clusters(x, dt = NULL, u, r = 1, ...)
```

Arguments

`x, form`	If numeric, `x` is a vector of data. If a formula is passed
`...`	Other parameters.
`dt`	Date or time of observations. If not provided, regular step are asumed.
`u`	Threshold.
`r, rlow, ini`	Declustering parameters. See details.

Details

Two conditions are required for peaks to not be rejected. First, two peaks `Q1` and `Q2` must be separated by a period of at least `r` days, where one recommendation is

$$4days + log(A)$$

and A is the drainage area in squared kilometers. The second conditions is

$$Xmin > rlow * min(Q1, Q2).$$

where Xmin is lowest point between Q1 and Q2. By defautlt, rlow = 0.75. When one of the two conditions is not statisfied the lowest of the two peaks is discarded. The 2 conditions are verify sequentially, from an initial set of peaks. If ini = 'run', a run declustering method with one lag is first used to filter the initial set of peaks to the maximums of the respective clusters. If ini = 'lmax' the 2 condition are verified accross all local maximums. If ini = 'wrc' (default), the second condition id verified next the first condition is verify on the previously extracted peaks. The two version are very similar and differ only on few cases where the modified version is more conservative and reject peaks that are kept in the initial version.

The function which.clusters is returning the indices of the peaks identified by the run declustering method where clusters are separated by a period of r consecutive values under the threshold.

References

Lang M, Ouarda TBMJ, Bobee B. (1999) Towards operational guidelines for over-threshold modeling. Journal of Hydrology. Dec 6;225(3):103-17.

Examples

```
# Declustering using the flood recommendation.
cid <- which.floodPeaks(flow~date, flowStJohn, u = 1000, r = 14, rlow = .75,
                        ini = 'wrc')

plot(flowStJohn, type = 'l')
points(flowStJohn[cid,], col = 'red', pch = 16)

## Simpler run declustering
cid <- which.clusters(flowStJohn$flow, u = 1000, r = 14)

plot(flowStJohn, type = 'l')
points(flowStJohn[cid,], col = 'red', pch = 16)
```

---

wtr_yr *Designation of the water year*

Description

Display water year

Usage

```
wtr_yr(dates, start_month = 10)
```

Arguments

dates	A vector of dates with actual year
start_month	Month in which the year starts (defaults to October)

Value

Year starting in start_month

Source

http://stackoverflow.com/questions/27626533/r-create-function-to-add-water-year-column

Examples

```
date <- seq(as.Date("1910/1/1"), as.Date("1912/1/1"), "days")
wtr_yr_date <- wtr_yr(dates=date, start_month=10)
data.frame(wtr_yr_date,date)
```

Index