Package: gumboot (via r-universe)

September 12, 2024

Type Package

Title Bootstrap Analyses of Sampling Uncertainty in Goodness-of-Fit Statistics

Version 1.0.1

Author Martyn Clark [aut], Kevin Shook [aut, trl, cre]

Maintainer Kevin Shook <kevin.shook@usask.ca>

Description Uses jackknife and bootstrap methods to quantify the sampling uncertainty in goodness-of-fit statistics. Full details are in Clark et al. (2021), ``The abuse of popular performance metrics in hydrologic modeling", Water Resources Research, <doi:10.1029/2020WR029001>.

License GPL-3

Encoding UTF-8

Depends R (>= 4.0)

Imports stats, dplyr, ggplot2, lubridate, stringr, ncdf4, reshape2

LazyData true

RoxygenNote 7.2.3

Suggests testthat, knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Date/Publication 2023-10-18 16:20:05 UTC

Repository https://kevinshook.r-universe.dev

RemoteUrl https://github.com/cran/gumboot

RemoteRef HEAD

RemoteSha fc5db8ab1a4d943ee22bc2a839f0517ab8bf254d

Contents

`gumboot-package` *Bootstrap Analyses of Hydrological Model Error*

Description

Does jackknife after bootstrap analyses of the error in hydrological models by estimating the empirical probability distributions of NSE (Nash-Sutcliffe efficiency) and KGE (Kling-Gupta efficiency) estimators.

Funding

The package was partly funded by the Global institute for Water Security (GIWS; https://water.usask.ca/) and the Global Water Futures (GWF; https://gwf.usask.ca/) program.

Author(s)

Coded by: Martyn Clark and Kevin Shook

Maintained by: Kevin Shook <kevin.shook@usask.ca>

References

The package code is described in:
Clark et al. (2021), "The abuse of popular performance metrics in hydrologic modeling", Water Resources Research, <doi:10.1029/2020WR029001>.

bootjack *Bootstrap-jacknife of flow calibration statistics*

Description

Bootstrap-jacknife of flow calibration statistics

Usage

```
bootjack(
  flows,
  GOF_stat = c("NSE", "KGE"),
  nSample = 1000,
  waterYearMonth = 10,
  startYear = NULL,
  endYear = NULL,
  minDays = 100,
  minYears = 10,
  returnSamples = FALSE,
  seed = NULL,
  bootYearFile = NULL
)
```

Arguments

flows	Required. Data frame containing the date, observed and simulated flows. The variable names must be 'date', 'obs', and 'sim', respectively. The date must be a standard R date.
GOF_stat	Required. Name(s) of simulation goodness of fit statistic(s) to be calculated. Currently both NSE and KGE are supported.
nSample	Required. Number of samples for bootstrapping.
waterYearMonth	Required. Month of beginning of water year. Default is 10 (October). If the calendar year is required, set waterYearMonth = 13.
startYear	Optional. First year of data to be used. If NULL then not used.
endYear	Optional. Last year of data to be used. If NULL then not used.
minDays	Required. Minimum number of days per year with valid (i.e. greater than 0) flows. Default is 100.
minYears	Required. Minimum number years to be used. Default is 10.
returnSamples	Optional. Default is FALSE. If TRUE, then sample statistics are returned. This is primarily used for debugging/testing.
seed	Optional. If NULL (the default) then no seed is specified for the random number generator used for the bootstrapping. If a value is specified then the bootstrapping will always use the same set of pseudo-random numbers.

`bootYearFile` Optional. If `NULL` (the default) the years used for the bootstrapping are neither output nor input. If a file is specified, and it it does not already exist, then the bootstrap years will be written to a .csv file as a table with the dimensions of years x nSample. If a file is specified, and it _does_ exist, then the years will be read in, and used for the bootstrapping.

Value

Returns a data frame containing the goodness of fit statistic name (i.e. 'NSE' and/or 'KGE'), and `seJack` = standard error of jacknife, `seBoot` = standard error of bootstrap, `p05, p50, p95`, the 5th, 50th and 95th percentiles of the estimates, `score` = jackknife score, `biasJack` = bias of jackknife, `biasBoot` = bias of bootstap, `seJab` = standard error of jackknife after bootstrap.

Author(s)

Martyn Clark and Kevin Shook

See Also

`read_CAMELS`

Examples

```
NSE_stats <- bootjack(flows_1030500, "NSE")
```

CAMELS_bootjack *Jackknife after bootstrap for all CAMELS sites*

Description

Hydrologic model simulations can be produced using input-response data from the 671 catchments in the CAMELS dataset (Catchment Attributes and MEteorology for Large-sample Studies). Newman et al. (2015) and Addor et al. (2017) provide details on the hydrometeorological and physiographical characteristics of the CAMELS catchments. The CAMELS catchments are those with minimal human disturbance (i.e., minimal land use changes or disturbances, minimal water withdrawals), and are hence almost exclusively smaller, headwater-type catchments (median basin size of 336 km^2^). The CAMELS data used for the large-domain model simulations are publicly available at the National Center for Atmospheric Research at https://ral.ucar.edu/solutions/products/camels.

Usage

```
CAMELS_bootjack(
  CAMELS_sites = NULL,
  NetCDF_file = NULL,
  sim_var = "kge",
  GOF_stat = c("NSE", "KGE"),
  nSample = 1000,
```

```
  waterYearMonth = 10,
  startYear = NULL,
  endYear = NULL,
  minDays = 100,
  minYears = 10,
  seed = NULL,
  bootYearFile = NULL,
  quiet = FALSE
)
```

Arguments

CAMELS_sites	Required. Data frame of CAMELS sites. Must contain a field called 'hcdn_site'. The data frame hcdn_conus_sites will work. You can subset this data frame if you want to use fewer sites.
NetCDF_file	Required. NetCDF file containing CAMELS modelled and gauged flows.
sim_var	Required. Name of variable containing simulated flows in NetCDF.
GOF_stat	Required. Name(s) of simulation goodness of fit statistic(s) to be calculated. Currently both NSE and KGE are supported.
nSample	Required. Number of samples for bootstrapping.
waterYearMonth	Required. Month of beginning of water year. Default is 10 (October). If the calendar year is required, set waterYearMonth = 13.
startYear	Optional. First year of data to be used. If NULL then not used.
endYear	Optional. Last year of data to be used. If NULL then not used.
minDays	Required. Minimum number of days per year with valid (i.e. greater than 0) flows. Default is 100.
minYears	Required. Minimum number years to be used. Default is 10.
seed	Optional. If NULL (the default) then no seed is specified for the random number generator used for the bootstrapping. If a value is specified then the bootstrapping will always use the same set of pseudo-random numbers.
bootYearFile	Optional. If NULL (the default) the years used for the bootstrapping are neither output nor input. If a file is specified, and it it does not already exist, then the bootstrap years will be written to a .csv file as a table with the dimensions of years x nSample. If a file is specified, and it _does_ exist, then the years will be read in, and used for the bootstrapping.
quiet	Optional. If FALSE (the default) a progress bar is displayed. If TRUE, it is not.

Value

Returns a data frame containing the following variables:

CAMELS_site CAMELS site number

lat CAMELS site latitude

lon CAMELS site longitude

GOF_stat Goodness of fit statistics (i.e. NSE or KGE)

`seJack` standard error of jacknife

`seBoot` standard error of bootstrap

`p05, p50, p95` the 5th, 50th and 95th percentiles of the estimates

`score` the jackknife score

`biasJack` the bias of the jackknife

`biasBoot` the bias of the bootstrap

`seJab` the standard error of the jackknife after bootstrap

Author(s)

Martyn Clark and Kevin Shook

References

N. Addor, A. Newman, M. Mizukami, and M. P. Clark, 2017. Catchment attributes for large-sample studies. Boulder, CO: UCAR/NCAR. doi: 10.5065/D6G73C3Q

Addor, N., Newman, A. J., Mizukami, N. and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21, 5293–5313, doi: 10.5194/hess2152932017, 2017.

See Also

`read_CAMELS`

Examples

```
## Not run:
camels <- CAMELS_bootjack(CAMELS_sites = sites, NetCDF_file = "CAMELS_flow.nc")

## End(Not run)
```

---

flows_1030500 *Observed and simulated flows for a single location*

Description

A data frame containing observed and simulated flows for USGS site 1030500

Usage

```
flows_1030500
```

Format

A data frame with 6940 rows and 3 variables:

date Date of flows

obs observed flows (m^3)/s)

sim simulated flows (m^3)/s)

`ggplot_estimate_uncertainties` *Plots uncertainties in model error estimates*

Description

Plots uncertainties in model error estimates

Usage

```
ggplot_estimate_uncertainties(JAB_stats, fill_colour = NULL)
```

Arguments

`JAB_stats`	Required. Data frame of jackknife after boot statistics for a large number of model runs, as produced by `CAMELS_bootjack`.
`fill_colour`	Optional. If `NULL` (the default), then all data series are plotted as lines. If specified, e.g.`fill_colour = "orange"`, the plot of 2 x the Jackknife estimate of the standard error will be filled with the specified colour.

Value

Returns a `ggplot2` object of the plots, faceted by goodness of fit statistic, i.e. NSE/KGE. The confidence interval (difference between the 95^th^ and 5^th^ quantiles, and the value of 2 x the Bootstrap estimate of the standard error are plotted as lines. The values of 2 x the Jackknife estimate of the standard error are plotted as filled)

Author(s)

Martyn Clark and Kevin Shook

See Also

`CAMELS_bootjack`

Examples

```
## Not run:  p <- ggplot_estimate_uncertainties(all_stats, "orange")
```

hcdn_conus_sites *Locations of HCDN sites in CONUS*

Description

A data frame containing the locations of the USGS Hydro-Climatic Data Network site for the continental US (CONUS). These are the same sites used by CAMELS (Catchment Attributes and METeorology for Large-sample Studies).

Usage

```
hcdn_conus_sites
```

Format

A data frame with 670 rows and 3 variables:

hcdn_site HCDN site number (integer)

lat Site latitude (decimal degrees)

lon Site longitude (decimal degrees)

Source

This data set is described in Lins, H. F. (2012). USGS Hydro-climatic data network 2009 (HCDN-2009). U.S. Geological Survey Fact Sheet 2012-3047. Retrieved from https://pubs.usgs.gov/fs/2012/3047/. The data can be downloaded at doi: 10.5066/P9HP0WFJ.

read_CAMELS *Reads simulated and observed values from CAMELS netcdf file for a single location*

Description

Reads simulated and observed values from CAMELS netcdf file for a single location

Usage

```
read_CAMELS(nc_file, site, obsName = "obs", simName = "kge")
```

Arguments

nc_file	Required. netCDF file to read CAMELS data from.
site	Required. Site number to extract data.
obsName	Required. Name for variable containing observations. Default is "obs".
simName	Required. Name for variable containing simulations. Default is "kge".

Value

Returns a data frame containing the date, observed and simulated flows. The name of the observed flow variable is `obs`, the name of the simulated flow variable is `sim`.

Author(s)

Martyn Clark and Kevin Shook

See Also

CAMELS_bootjack

Examples

```
## Not run:
flows <- read_CAMELS(nc_file = "CAMELS_flow.nc", site = 1030500)

## End(Not run)
```

Index