Package 'NonParRolCor'

April 7, 2021

Type Package

Title a Non-Parametric Statistical Significance Test for Rolling Correlation

Version 0.6.0

Maintainer Josue M. Polanco-Martinez <josue.m.polanco@gmail.com>

Depends R (>= 3.5.0), gtools, pracma, colorspace, doParallel

Imports foreach, scales

Description Estimates and plots (as a heat map) the statistical significance of rolling window correlation coefficients, which is carried out through a non-parametric computing-intensive method. This method addresses the effects due to the multiple testing (inflation of the Type I error) when the statistical significance is estimated for the rolling window correlation coefficients. The method is based on Monte Carlo simulations by permuting one of the variables (dependent) under analysis and keeping fixed the other variable (independent). We improve the computational efficiency of this method to reduce the computation time through parallel computing. The 'NonParRolCor' package also provides examples with synthetic and real-life ecological time series to exemplify its use. Methods derived from R. Telford (2013) <https://quantpalaeo.wordpress.com/2013/01/04/> and J.M. Polanco-Martinez (2020) <doi:10.1016/j.ecoinf.2020.101163>.

License GPL (>= 2)

Repository CRAN

Encoding UTF-8

LazyData true

Date/Publication 2021-04-07 10:00:02 UTC

NeedsCompilation no

Author Josue M. Polanco-Martinez [aut, cph, cre] (<https://orcid.org/0000-0001-7164-0185>), Jose L. Lopez-Martinez [ctb] (<https://orcid.org/0000-0002-2489-7559>)

R topics documented:

`NonParRolCor-package` *Non-parametric statistical significance test for rolling correlation*

Description

'NonParRolCor' estimates and plots as a heat map the statistical significance of rolling window correlation coefficients between two regular (sampled on identical time points) time series. The statistical significance is computed through a non-parametric computing-intensive method. This method (test) addresses the effects due to the multiple testing problem (inflation of the Type I error) when the statistical significance is estimated for rolling window correlation coefficients. The method is based on Monte Carlo simulations by permuting one (dependent) of the variables under analysis and keeping fixed the other (independent) variable. We improve the computational efficiency of this method to reduce the computation time through parallel computing. It has been designed especially for ecological data although this can be applied to other kinds of data sets as well. 'NonParRolCor' contains two functions: (1) 'estimation_NonParRolCor' to perform the test and (2) 'heatmap_NonParRolCor' to plot the time series under analysis and to create a heat map of the rolling window correlation coefficients that are statistically significant. The functions contained in 'NonParRolCor' are highly flexible since these contains several parameters to control the estimation of correlation and the features of the plots of the time series, e.g. to remove potential linear trend contained in the time series under analysis or to personalise the plot of the time series under analysis. The 'NonParRolCor' package also provides examples with synthetic ('syntheticdata' data set) and real-life ecological ('ecodata' data set) time series to exemplify its use.

Details

Package:	NonParRolCor
Type:	Package
Version:	0.6
Date:	2020-04-06
License:	GPL (>= 2)
LazyLoad:	yes

NonParRolCor package contains two functions: (1) `estimation_NonParRolCor` that estimates the rolling window correlation coefficients and their respective statistical significance through a non-parametric computing-intensive method, and (2) `heatmap_NonParRolCor` that plots the time series under scrutiny and that creates a heat map of the rolling window correlation coefficients that are statistically significant. NonParRolCor also contains two data sets: (1) `syntheticdata` and (2) `ecodata` to exemplify the use of the aforementioned functions. The significance test is based on and inspired from Telford (2013) and Polanco-Martínez (2020) whereas the heat map plot is based

on Polanco-Martínez (2020). The implementation is described in detail in Polanco-Martínez and López-Martínez (2021).

Note

Dependencies: *stat*, *gtools*, *pracma*, *colorspace*, *scales*, *foreach*, *parallel*, *doParallel*.

Author(s)

Josué M. Polanco-Martínez (a.k.a. jomopo).
BC3 - Basque Centre for Climate Change, Bilbao, SPAIN.
Web1: https://scholar.google.es/citations?user=8djLIhcAAAAJ&hl=en.
Web2: https://www.researchgate.net/profile/Josue-Polanco-Martinez.
Email: <josue.m.polanco@gmail.com>
José L. López-Martínez.
Faculty of Mathematics, Universidad Autónoma de Yucatán (UADY), Tizimín, MEXICO.
Web1: https://scholar.google.es/citations?user=552PKVEAAAAJ&hl=es.
Web2: https://www.researchgate.net/profile/Jose-Lopez-Martinez-3.
Email: <jose.lopez@correo.uady.mx>.

Acknowledgement:
The first author acknowledges to the SEPE (Spanish Public Service of Employment) for its funding support.

References

Polanco-Martínez, J. M. and López-Martínez, J.M. (2021). A non-parametric method to test the statistical significance in rolling window correlations, and applications to ecological time series. Submitted to Ecological Informatics (2021).

Polanco-Martínez, J. M. (2020). NonParRolCor: an R package for estimating rolling window multiple correlation in ecological time series. Ecological Informatics, 60, 101163. <URL: doi: 10.1016/j.ecoinf.2020.101163>.

Polanco-Martínez, J. M. (2019). Dynamic relationship analysis between NAFTA stock markets using nonlinear, nonparametric, non-stationary methods. Nonlinear Dynamics, 97(1), 369-389. <URL: doi: 10.1007/s1107101904974y>.

Telford, R.: Running correlations – running into problems. (2013). <URL: https://quantpalaeo.wordpress.com/2013/01/04/>.

ecodata *Ecological data set to exemplify the use of the functions contained in* NonParRolCor

Description

The data set ecodata contains four columns, the first one (named "Years") is the time (years from 1989 to 2008, monthly resolution), the second (named "SST") are monthly anomalies of sea surface temperature (SST) of the south of Gran Canaria (28.5 N/16.5 W) (NOAA 2021a), the third column (named "NAO") are the monthly index of the North Atlantic Oscillation (NAO) (NOAA 2021b), and the last column (named "CPUE") are monthly catches of common octopus (measured as CPUE or Catch Per Unit of Effort) from an artisanal fisheries from the Southwest of Gran Canaria Islands (Caballero-Alfonso et al. 2010, Polanco et al. 2011, Polanco-Martínez 2012).

Usage

```
data(ecodata)
```

Format

One file in ASCII format containing four columns and 240 rows, columns are separated by spaces.

Source

Caballero-Alfonso, A, Ganzedo, U., Trujillo-Santana, A., Polanco, J., del Pino, A. S., Ibarra-Berastegi, G., Castro-Hernández, J. (2010). The role of climatic variability on the short-term fluctuations of octopus captures at the Canary Islands. Fisheries Research, 102(3), 258-265. <URL: doi: 10.1016/j.fishres.2009.12.006>.

NOAA Optimum Interpolation (OI) Sea Surface Temperature (SST) V2, <URL: https://psl.noaa.gov/data/gridded/data.noaa.oisst.v2.html>, accessed: 2021-02-28.

NAO index, <URL: https://psl.noaa.gov/data/correlation/nao.data>, accessed: 2021-02-28.

Polanco, J., Ganzedo, U., Sáenz, J., Caballero-Alfonso, A. M., & Castro-Hernández, J. J. (2011). Wavelet analysis of correlation among Canary Islands octopus captures per unit effort, sea-surface temperatures and the North Atlantic Oscillation. Fisheries Research, 107(1-3), 177-183. <URL: doi: 10.1016/j.fishres.2010.10.019>.

Polanco-Martínez, J.M. (2012). Aplicación de técnicas estadísticas en el estudio de fenómenos ambientales y ecosistémicos, Ph.D. thesis, University of Basque Country, Spain. <URL: https://addi.ehu.es/handle/10810/11295>.

`estimation_NonParRolCor` *Estimates the statistical significance of the rolling window correlation coefficients*

Description

The `estimation_NonParRolCor` function estimates the statistical significance of the rolling window correlation coefficients through a non-parametric computing-intensive method to be applied to two time series sampled on identical time points for all the possible window-lengths (time-scales) or for a band of window-lengths. To carry out the computational implementation we extend the works of Telford (2013), Polanco-Martínez (2019) and Polanco-Martínez (2020). The statistical and computational implementation is described in Polanco-Martínez and López-Martínez (2021). The `estimation_NonParRolCor` function is highly flexible since this contains several parameters to control the estimation of the correlation. A list of parameters are described in the following lines.

Usage

```
estimation_NonParRolCor(inputdata, CorMethod="pearson", typewidthwin="FULL",
                    widthwin_1=3, widthwin_N=dim(inputdata)[1], Align="center",
                        rmltrd=TRUE, Scale=TRUE, MCSim=1000, prob=0.95, Np=2)
```

Arguments

`inputdata`	A matrix of 3 columns: time (regular/evenly spaced), the first (independent) variable, and the second (dependent) variable.
`CorMethod`	The method used to estimate the correlations via "pearson", but other options ("spearman" and "kendall") are available (please look at: R>?cor.test).
`typewidthwin`	"FULL" is to estimate the windows from 2, 4, ..., to dim(inputdata)[1]) if *Align* is equal to "left" or "right", or from 3, 5,..., to dim(inputdata)[1]) if Align is "center". The other option is "PARTIAL", please you should take into account that *widthwin_1* and *widthwin_1* MUST be ODD if the *Align* option is "center".
`widthwin_1`	First value for the size (length) of the windows when the option *typewidthwin*="PARTIAL" is selected, the minimum value is 3 (the default value), but you must define this parameter (please note that *widthwin_1* < *widthwin_N*).
`widthwin_N`	Last value for the size (length) of the windows when the option *typewidthwin*="PARTIAL" is selected, by default is dim(inputdata)[1], but you must define this parameter (please note that *widthwin_1* < *widthwin_N*).
`Align`	To align the rolling object, RolWinMulCor uses three options: "left", "center", and "right" (please look at: R>?running). However, there are some restrictions, which have been described lines above. We recommend to use the "center" option to ensure that variations in the correlations are aligned with the variations in the relationships of the variables under study, rather than being shifted to left or to right (Polanco-Martínez 2019, 2020), but this imply that the window-lengths must be odd.

rmltrd	Remove (by default is "TRUE"; "FALSE" otherwise) the linear trend in the variables under analysis.
Scale	Scale (by default is "TRUE"; "FALSE" otherwise) is used to "normalize" or "standardize" the variables under analysis.
MCSim	Number of Monte-Carlo simulations to permute the second variable. It is advisable to use at least 1000 simulations.
prob	Numeric vector of probabilities with values in the interval [0,1], by default prob=0.95 (p=0.05), please look at R?quantile, Telford (2013), or Polanco-Martínez and López-Martínez (2021) for more information.
Np	Number of cores, by default is 2 (please verify the number of cores of your computer. WARNING: it is not advisable to use the maximum number of cores of your computer).

Details

The estimation_NonParRolCor function estimates the statistical significance of the rolling window correlation coefficients through a non-parametric computing-intensive method to be applied to two time series sampled on identical time points for all the possible window-lengths or for a band of window-lengths. The function estimation_NonParRolCor uses the functions *cor.test* (package:stats) and *running* (package:gtools) to estimate the correlation coefficients and compute the rolling window correlations, and also the functions *foreach* (package:foreach) and *makeCluster* (package:parallel) to parallelize the estimation of the rolling window correlations.

Value

Outputs: A list containing eight elements: *the_matrixCOR* and *CRITVAL* contain the rolling window correlation coefficients and its respective critical values, *nwin* and *Rwidthwin* contain the number of window-lengths (time-scales) and the window-lengths, *left_win* and *righ_win* are used to accommodate the times of the rolling window correlation coefficients, finally *MCSim* indicates the number of Monte-Carlo simulations and *prob* the significance level.

Author(s)

Josué M. Polanco-Martínez (a.k.a. jomopo).
BC3 - Basque Centre for Climate Change, Bilbao, SPAIN.
Web1: https://scholar.google.es/citations?user=8djLIhcAAAAJ&hl=en.
Web2: https://www.researchgate.net/profile/Josue-Polanco-Martinez.
Email: <josue.m.polanco@gmail.com>.
José L. López-Martínez.
Faculty of Mathematics, Universidad Autónoma de Yucatán (UADY), Tizimín, MEXICO.
Web1: https://scholar.google.es/citations?user=552PKVEAAAAJ&hl=es.
Web2: https://www.researchgate.net/profile/Jose-Lopez-Martinez-3.
Email: <jose.lopez@correo.uady.mx>.

References

Polanco-Martínez, J. M. and López-Martínez, J.M. (2021). A non-parametric method to test the statistical significance in rolling window correlations, and applications to ecological time series.

Submitted to Ecological Informatics (2021).

Polanco-Martínez, J. M. (2020). NonParRolCor: an R package for estimating rolling window multiple correlation in ecological time series. Ecological Informatics, 60, 101163. <URL: doi: 10.1016/j.ecoinf.2020.101163>.

Polanco-Martínez, J. M. (2019). Dynamic relationship analysis between NAFTA stock markets using nonlinear, nonparametric, non-stationary methods. Nonlinear Dynamics, 97(1), 369-389. <URL: doi: 10.1007/s1107101904974y>.

Telford, R.: Running correlations – running into problems. (2013). <URL: https://quantpalaeo.wordpress.com/2013/01/04/>.

Examples

```
 # Code to test the function "estimation_NonParRolCor"
# Example 4.1 "synthetic time series" (Figures 6 and 7) in
# Polanco-Martínez and López-Martínez (2021).
# Defining the 'NonParRolCor' parameters
TYPEWIDTHWIN="PARTIAL"
# Number of Monte-Carlo simulations (MCSim), please use at least 1000.
# WARNING: MCSim=5, it's just to test this example!
MCSim <- 2
Np    <- 2  # Number of cores
X_Y <- estimation_NonParRolCor(syntheticdata[1:350,], CorMethod="pearson",
                  typewidthwin=TYPEWIDTHWIN, widthwin_1=29,
                  widthwin_N=51, Align="center",
                  rmltrd=TRUE, Scale=TRUE, MCSim=MCSim, Np=Np)
```

---

heatmap_NonParRolCor *Plot the variables under analysis and a heat map of the rolling correlation coefficients that are statistically significant*

Description

The heatmap_NonParRolCor function plots the time series under study and create a heat map of the rolling window correlation coefficients that are statistically significant, which are generated by the estimation_NonParRolCor function.

Usage

```
heatmap_NonParRolCor(inputdata, corcoefs, CRITVAL, Rwidthwin="", typewidthwin="",
                     widthwin_1=3, widthwin_N=dim(inputdata)[1], varX="X",
                    varY="Y", coltsX="black", coltsY="blue", LWDtsX=1, LWDtsY=1,
                     CEXLAB=1.15, CEXAXIS=1.05)
```

Arguments

`inputdata`	The same data matrix (time, first and second variable) that was used with the `estimation_NonParRolCor` function.
`corcoefs`	Rolling correlation coefficients estimated with the `estimation_NonParRolCor` function.
`CRITVAL`	The critical values computed through the function `estimation_NonParRolCor`.
`Rwidthwin`	*Rwidthwin* is a vector that contain the window lengths, which come from the function `estimation_NonParRolCor`.
`typewidthwin`	Contains the type ("FULL" or "PARTIAL") of heat map that will be plotted, this information is provided by `estimation_NonParRolCor`. Please note that whether option *typewidthwin*="PARTIAL" is selected, and you will need to provide the parameters *widthwin_1* and *widthwin_N* to plot the heat map.
`widthwin_1`	First value for the size (length) of the windows when the option *typewidthwin*="PARTIAL" is selected, this value is provided by the `estimation_NonParRolCor` function.
`widthwin_N`	Last value for the size (length) of the windows when the option *typewidthwin*="PARTIAL" is selected, this value is provided by the `estimation_NonParRolCor` function.
`varX, varY`	Names of the first (e.g. X) and the second (e.g. Y) variables contained in *inputdata*. Please note that the names of these two variables should be provided (by default are X and Y) when these variables are plotted.
`coltsX, coltsY`	Colors to be used when the variables are plotted, by default are "black" for the first variable and "blue" for the second, but other colors can be used.
`LWDtsX, LWDtsY`	Line-widths for the first and the second variable when these are plotted, by default these have values of 1, but other values (widths) can be used.
`CEXLAB, CEXAXIS`	These parameters are used to plot the sizes of the X-axis and Y-axis labels and X- and Y-axis, by default these parameters have values of 1.15 and 1.05, respectively, but it is possible to use other values.

Details

The `heatmap_NonParRolCor` function plots the variables (time series) under analysis and a heat map of the rolling correlation coefficients that are statistically significant and that are estimated through a non-parametric computing-intensive method. The `heatmap_NonParRolCor` function uses the outputs of the `estimation_NonParRolCor` function. To implement this method we extend the works of Telford (2013), Polanco-Martínez (2019) and Polanco-Martínez (2020), and to implement the heat map we follow to Polanco-Martínez (2020). A detailed description of this method can be found in Polanco-Martínez and López-Martínez (2021). `heatmap_NonParRolCor` uses the functions *diverge_hcl* (package:colorspace) and *alpha* (package:scales) to create the palette of colors.

Value

Outputs: A plot of the time series under analysis and a heat map (a multi-plot via screen) of the rolling window correlation coefficients statistically significant. This multi-plot can be saved in your preferred format.

Author(s)

Josué M. Polanco-Martínez (a.k.a. jomopo).
BC3 - Basque Centre for Climate Change, Bilbao, SPAIN.
Web1: https://scholar.google.es/citations?user=8djLIhcAAAAJ&hl=en.
Web2: https://www.researchgate.net/profile/Josue-Polanco-Martinez.
Email: <josue.m.polanco@gmail.com>.

References

Polanco-Martínez, J. M. and López-Martínez, J.M. (2021). A non-parametric method to test the statistical significance in rolling window correlations, and applications to ecological time series. Submitted to Ecological Informatics (2021).

Polanco-Martínez, J. M. (2020). NonParRolCor: an R package for estimating rolling window multiple correlation in ecological time series. Ecological Informatics, 60, 101163. <URL: doi: 10.1016/j.ecoinf.2020.101163>.

Examples

```
# Code to test the function "heatmap_NonParRolCor"
# Defining NonParRolCor parameters
# Example 4.1 "synthetic time series" (Figures 6 and 7) in
# Polanco-Martínez and López-Martínez (2021).
TYPEWIDTHWIN="PARTIAL"
# Number of Monte-Carlo simulations (MCSim), please use at least 1000.
# WARNING: MCSim=2, it's just to test this example!
MCSim <- 2
Np    <- 2  # Number of cores
X_Y <- estimation_NonParRolCor(syntheticdata[1:350,], CorMethod="pearson",
                  typewidthwin=TYPEWIDTHWIN, widthwin_1=29,
                  widthwin_N=51, Align="center", rmltrd=TRUE,
                  Scale=TRUE, MCSim=MCSim, Np=Np)
heatmap_NonParRolCor(syntheticdata[1:350,], X_Y$matcor, X_Y$CRITVAL,
                  Rwidthwin=X_Y$Windows, typewidthwin=TYPEWIDTHWIN,
                  widthwin_1=29, widthwin_N=51)
```

`syntheticdata` *Synthetic data set to exemplify the use of the functions contained in* NonParRolCor

---

Description

The data set `syntheticdata` contains three columns: the first one are the "times" (from 1 to 500) (named "Times"), the second (named "X") and the third (named "Y") columns were generated by a bi-variate AR1 process with similar autocorrelation coefficients of 0.25. We generate two correlated

bi-variate AR1 time series with positive (direct) correlation (0.85) for the first 250 elements and with negative (inverse) correlation (-0.85) for the last 250 elements (Polanco-Martínez and López-Martínez 2021).

Usage

```
data(syntheticdata)
```

Format

One file in ASCII format containing three columns and 500 rows, columns are separated by spaces.

Source

Author's own production (Josué M. Polanco-Martínez) based on: mpiktas (<URL: https://stats.stackexchange.com/users/2116/mpiktas>). How to simulate two correlated AR(1) time series?, Cross Validated. (2013). <URL: https://stats.stackexchange.com/q/71831> (version: 2013-10-03).

Polanco-Martínez, J. M. and López-Martínez, J.M. (2021). A non-parametric method to test the statistical significance in rolling window correlations, and applications to ecological time series. Submitted to Ecological Informatics (2021).

Index