



## Original software publication

## NonParRolCor: An R package for estimating rolling correlation for two regular time series

Josué M. Polanco-Martínez<sup>a,b,\*</sup>, José L. López-Martínez<sup>c</sup><sup>a</sup> GECOS-IME, Universidad de Salamanca, Salamanca, Spain<sup>b</sup> Basque Centre for Climate Change (BC3), Leioa, Spain<sup>c</sup> Faculty of Mathematics, Universidad Autónoma de Yucatán (UADY), Mérida, Yucatán, Mexico

## ARTICLE INFO

## Article history:

Received 31 December 2022

Received in revised form 8 February 2023

Accepted 2 March 2023

## Keywords:

Heat map

Monte-Carlo simulation

Multiple testing problem

Non-parametric test

Rolling window correlation

## ABSTRACT

The R package NonParRolCor estimates rolling window correlations between two regular time series. The statistical significance estimated for the rolling correlation coefficients addresses effects due to multiple testing. This is done via Monte Carlo simulations by permuting one of the variables and keeping the other fixed. NonParRolCor uses parallel computing to improve computation time when statistical significance is estimated. NonParRolCor contains four functions for estimating and plotting the correlation coefficients for a single window-length or a band of window-lengths. NonParRolCor's functions are highly flexible, since they contain several parameters for controlling the estimation of correlation and the plot output. Some applications are presented to illustrate its use.

© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Code metadata

Current code version	v0.80
Permanent link to code/repository used for this code version	<a href="https://github.com/ElsevierSoftwareX/SOFTX-D-22-00402">https://github.com/ElsevierSoftwareX/SOFTX-D-22-00402</a>
Code Ocean compute capsule	Not available
Legal Code License	GPL ( $\geq 2$ ) License
Code versioning system used	None
Software code languages, tools, and services used	R
Compilation requirements, operating environments & services used	R ( $\geq 3.6$ ), gtools, pracma, colorspace, doParallel
If available Link to developer documentation/manual	<a href="https://cran.r-project.org/web/packages/NonParRolCor/NonParRolCor.pdf">https://cran.r-project.org/web/packages/NonParRolCor/NonParRolCor.pdf</a>
Support email for questions	<a href="mailto:josue.m.polanco@gmail.com">josue.m.polanco@gmail.com</a>

## 1. Motivation and significance

One of the most common tasks in scientific research is the search for relationships between two or more time series. This is mainly done by estimating the correlation coefficient between two time series using either the Pearson or Spearman estimators [1,2]. However, correlations estimated for the full study period act as an overall measure of association, and although this is a useful initial approach for analysing correlation between two time series, relationships between time series can vary over time [3,4].

Furthermore, time series could contain information on multiple time-scales and in many cases these series and the relationships between them may not be stationary [3,5–7]. A simple but powerful way to deal with these challenges is rolling, running or sliding window correlation. Rolling window correlation is also useful for assessing the evolution and stability of correlation over time. However, one collateral effect due to multiple testing when statistical significance is estimated for the rolling window correlation is the inflation of the Type I error. This is also commonly known as the multiple testing or comparison problem [8,9].

In the literature there are very few R packages available for estimating rolling window correlation and, even worse, these packages do not include statistical significance tests or do not address the multiple comparison problem. As far as we know,

\* Corresponding author at: GECOS-IME, Universidad de Salamanca, Salamanca, Spain.

E-mail address: [josue.polanco@usal.es](mailto:josue.polanco@usal.es) (Josué M. Polanco-Martínez).

there are three R packages that estimate rolling correlations but they do not estimate the statistical significance of rolling correlation coefficients. They are the following: (1) *zoocat*, whose function for carrying out this task is `rollcor()` [10]—although this package was removed from CRAN in October 2020; (2) *roll*, which is also able to estimate rolling correlation through its function `roll_cor()` [11]; and (3) *TTR*, which includes the function `runCor()` for estimating rolling correlation [12]. There is also an R package that includes the estimation of statistical significance for rolling correlation coefficients, but the test used to estimate significance does not address the risks that arise from repeating multiple significance tests simultaneously. This R package, named *dendroTools*, uses the `daily_response()` function to estimate rolling correlation [13]. To the best of our knowledge, there are only two R packages that address the multiple comparison problem: (1) *RollWinMulCor*, which uses a classical approach, i.e. correcting p-values to address the multiple comparison problem [9]; and (2) *NonParRollCor* [14], which is the R package presented in this software paper.

The R package *NonParRollCor* is based on a previous work [15] in which we presented a non-parametric test for estimating statistical significance for rolling window correlations addressing the multiple comparison problem. This previous work was partially inspired by the work by Telford (2013) [16] and Polanco-Martínez (2019) [4]. The non-parametric test used in *NonParRollCor* is based on Monte Carlo simulations by permuting one of the variables (e.g. the dependent) under analysis and keeping the other variable (e.g. the independent) fixed. We have improved this test via parallel computing to reduce the computation time: it is now possible to use as many CPU cores as available. We would like to highlight that, to the best of our knowledge, *NonParRollCor* is the only one R package that uses an “experimental” (Monte Carlo) statistical significance test to address the multiple comparison problem. That test also takes into account the window-lengths for which the rolling correlations are computed. In addition, *NonParRollCor* is able to estimate and plot rolling window correlation coefficients for only a single window-length and for a band of window-lengths as a heat map. Rolling window correlation coefficients can also be estimated through Pearson, Spearman and Kendall, and it is also possible to use different levels of statistical significance.

This software paper describes the main features and functionalities of *NonParRollCor* and provides some examples with real-life time series. *NonParRollCor* was originally developed to analyse environmental time series, but this software can be used with time series from different scientific fields (e.g. economics, finance, energy, geology, etc.). *NonParRollCor* is free, open source software (GPL license) and is available on The Comprehensive R Archive Network (CRAN), the official repository of the R project.

This paper is structured as follows. Section 2 briefly describes the non-parametric test and Section 3 describes the *NonParRollCor* software in detail. Section 4 provides some applications. Section 5 sets out the impact of the software and concludes the paper.

## 2. Estimating statistical significance

*NonParRollCor* includes a non-parametric test for estimating statistical significance for rolling window correlation coefficients which takes into account the repeating of multiple significance tests simultaneously. This test is based on Monte Carlo simulations by permuting one of the variables (e.g. the dependent) under analysis and keeping the other variable (e.g. the independent) fixed. It computes a critical value “CRITVAL” for the rolling correlation coefficients for each window-length, enabling the correlation coefficients of each window-length to be contrasted with

their corresponding critical values [9,16]. The description of this test is based on a previous work [15], so here we provide only a brief description.

1. Input: one independent variable (e.g.  $X$ ) and one dependent variable (e.g.  $Y$ ) with  $N$  elements.
2. For each possible window-length or `widthwin` (window-lengths range from  $3/4$  for odd/even window-lengths to `maxwin`, the maximum window-length, i.e.  $N$  or  $N - 1$  depending on whether  $N$  is an odd or even number), a Monte-Carlo simulation is carried out to determine its corresponding CRITVALs, that is:
  - (a) Start the Monte-Carlo simulations to determine the null distribution: for  $i = 1$  to `MCsims` replicates (it is highly advisable to use at least 1000):
    - i. Permuting all the time segments of the dependent variable.
    - ii. Estimating the maximum or highest correlation for the set of absolute values of rolling window correlation coefficients (estimated via Pearson, Spearman or Kendall) obtained between the independent and the permuted dependent variables.
  - (b) Close the Monte-Carlo simulation loop.
  - (c) Estimate the critical value (“CRITVAL”, i.e. the quantile 0.95 ( $p=0.05$ , two-sided test; other significance levels can be used) of the `MCsims` maximum absolute correlation coefficients obtained in the ensemble (step (a)).
3. Output: determine the statistical significance of the correlation coefficients for each window-length, if an `abs(correlation coefficient) > “CRITVAL”` this correlation coefficient is statistically significant with a 95% confidence level.

The CRITVALs take values from 0 to 1 and depend on the number of elements in the time series under analysis and on the window-lengths under inspection [15,16]. Smaller window-lengths produce CRITVALs close to one and larger ones produce CRITVALs that tend to zero.

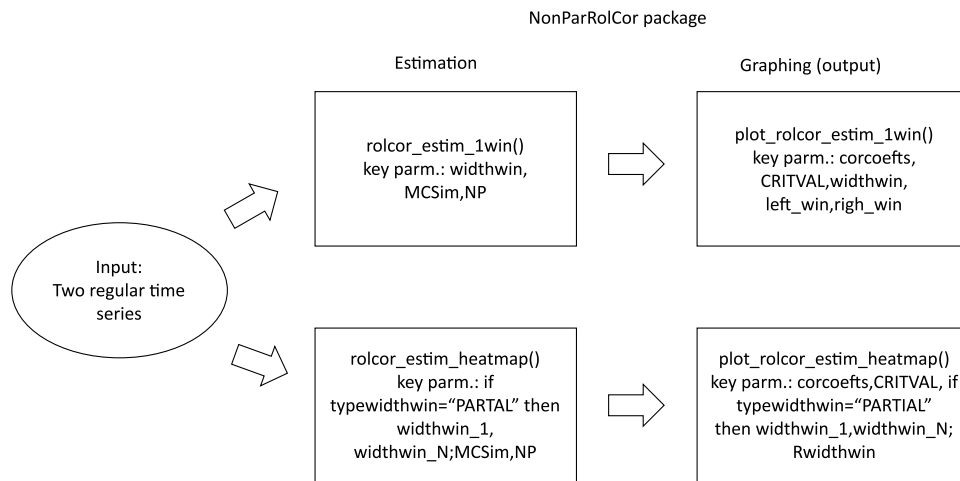
## 3. Software description

In the following subsections we provide details of the *NonParRollCor* system architecture (Section 3.1) and outline the software functionality (Section 3.3).

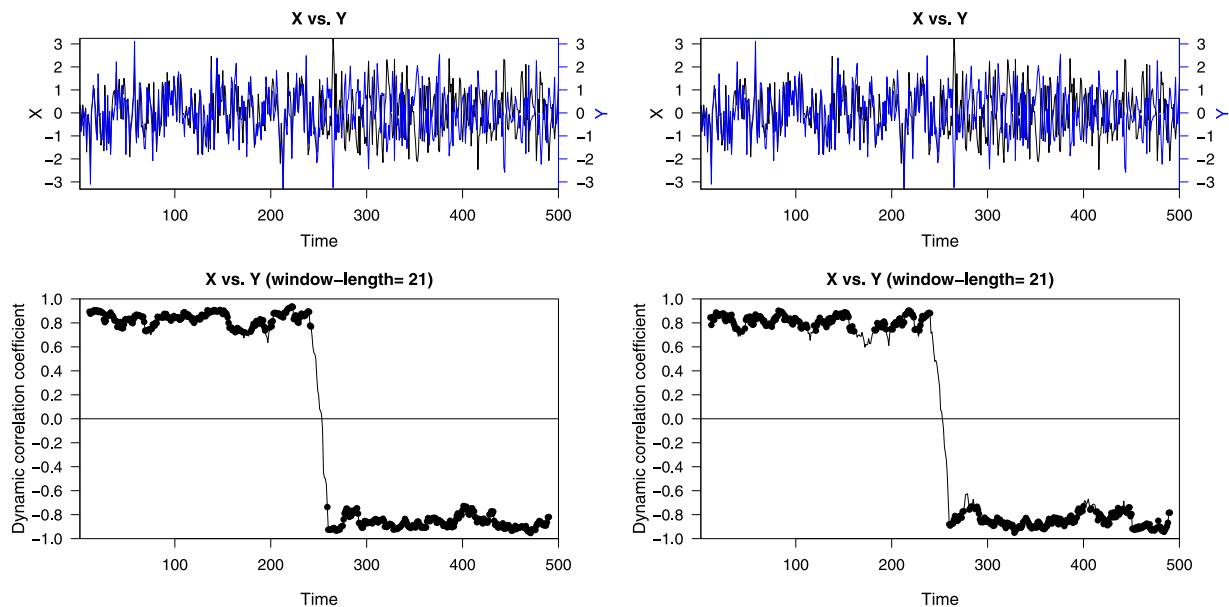
### 3.1. Software architecture

*NonParRollCor* contains four functions in two separate entities. Two functions are for estimations and the other two are for graphing, as explained visually in Fig. 1. There are two functions for graphing, but it is also possible to plot the outputs of the estimation functions through other internal or external R graphical functions.

Since the *NonParRollCor* R package is available on The Comprehensive R Archive Network (CRAN), installing it can be made from R by executing `> install.packages(“NonParRollCor”)`. The source code is also available on CRAN. *NonParRollCor* can be installed and used in the main operating systems (Windows, GNU/Linux and MacOS) since R is multiplatform.



**Fig. 1.** Software architecture of NonParRolCor. Given two regular time series, NonParRolCor estimates rolling window correlation coefficients and their statistical significance, addressing the multiple comparison problem. NonParRolCor also enables the outputs of the estimation functions to be visualised.



**Fig. 2.** Visual features of outputs of the `rolcor_estim_1win` and `plot_rolcor_estim_1win` functions via Pearson (left) and Spearman (right). Correlation coefficients that are statistically significant (95% confidence level) are in black dots. X and Y are two AR1 time series contained in `syntheticdata` [14].

### 3.1.1. Estimation

We provide the syntax of the two functions for estimation:

1. `rolcor_estim_1win`: this estimates rolling correlation coefficients for a single window-length for two regular time series and their statistical significance. The syntax is as follows:  
`rolcor_estim_1win(inputdata, CorMethod="pearson", widthwin=3, Align="center", rmltrd=TRUE, Scale=TRUE, MCSim=1000, Np=2, prob=0.95)`
2. `rolcor_estim_heatmap`: this estimates rolling correlation coefficients as the function `rolcor_estim_1win`, but the estimation is performed for a band of window-lengths or for all possible window-lengths, from three to the number of elements in the time series under analysis (i.e. `dim(inputdata)` [5]). This function has the following syntax:  
`rolcor_estim_heatmap(inputdata, CorMethod="pearson", typewidthwin="FULL", widthwin_1=3, widthwin_N=dim(inputdata)[1], Align="center",`

`rmltrd=TRUE, Scale=TRUE, MCSim=1000, Np=2, prob=0.95)`

The parameters are described below:

- `inputdata`: a matrix of three columns where the first column must be the times and the second (the “independent”) and third (the “dependent”) columns must be the time series under study.
- `CorMethod`: the default method for estimating correlations is “pearson”, but it is also possible to use “spearman” and “kendall”.
- `widthwin`: the window-length or size of the window used when the rolling window correlation coefficients are estimated. It can take values from 3 to the number of elements in the time series under study.
- `Align`: this is used to align the rolling object. There are three options: “left”, “center” and “right”, but the “center” option is used by default 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 the left or right [4,9]. However, this implies that the window-lengths must be odd.

- **typewidthwin**: this parameter has two modalities: (1) `typewidthwin= 'PARTIAL'` to estimate rolling correlations for a band of window-lengths (from `widthwin_1` to `widthwin_N`); and (2) `typewidthwin= 'FULL'` to estimate rolling correlations from all possible window-lengths (from 3 to `dim(inputdata)` [5]).
- **rmltrd**: this is used to remove the linear trend contained in `inputdata`.
- **Scale**: this is used to normalise or standardise the variables under analysis (the default option is "TRUE"; otherwise it is "FALSE").
- **MCSim**: number of Monte-Carlo simulations. It is advisable to use at least 1000 replicates.
- **Np**: number of CPU cores. It is not advisable to use the maximum number of cores of your computer (by default `Np` is set to two cores).
- **prob**: this defines the level of statistical significance (a typical value used is `prob = 0.95`, but other values can be used).

The output of `rolcor_estim_1win` is a list that contains six elements:

- **Correlation\_coefficients**: this element is an array that contains times (for the default case `Align = "center"` they are central times) and rolling window correlation coefficients.
- **CRITVAL**: this contains the critical values used to determine the statistical significance of the rolling window correlation coefficients.
- **CorMethod**: this is the correlation method (estimator) used to estimate the rolling correlations (e.g. "pearson").
- **widthwin**: this element contains the window-length.
- **left\_win** and **right\_win**: these two auxiliary elements are used to accommodate the times of the rolling window correlation coefficients.

The output of `rolcor_estim_heatmap` is a list that contains eight elements:

- **matcor**: this element contains an array with the rolling window correlation coefficients.
- **CRITVAL**: this is an array that contains the critical values for the window-lengths.
- **NoWindows**: this is the number of windows.
- **Windows**: this element contains the window-lengths.
- **left\_win** and **right\_win**: these two auxiliary elements are as described above.
- **MCSim** and **prob**: these two elements are also as defined above.

### 3.1.2. Graphing

A description of the syntax of the two functions used for graphing follows:

1. **plot\_rolcor\_estim\_1win**: this plots the output of the `rolcor_estim_1win` function as a single plot indicating the correlation coefficients which are statistically significant. This function also plots the time series under study. The syntax is as follows:  

```
plot_rolcor_estim_1win(inputdata, corcoefs, CRITVAL, widthwin, left_win, right_win, varX="X", varY="Y", coltsX="black", coltsY="blue", rmltrd=TRUE, Scale=TRUE)
```

2. **plot\_rolcor\_estim\_heatmap**: this plots the output of the `rolcor_estim_heatmap` function as a heat map and only plots those coefficients which are statistically significant, i.e. non-significant coefficients are left blank. This function also plots the data under analysis. This function has the following syntax:

```
plot_rolcor_estim_heatmap(inputdata, corcoefs, CRITVAL, Rwidthwin="", typewidthwin="", widthwin_1=3, widthwin_N=dim(inputdata)[1], varX="X", varY="Y", coltsX="black", coltsY="blue")
```

We define only the parameters that have not already been described.

- **varX** and **varY**: these parameters are used to define the names for the time series under study, e.g. X for the first series and Y for the second.
- **coltsX** and **coltsY**: these two parameters are used to define the colours for the plots of the time series under study.

There are other parameters (e.g. `HeigWin1`, `HeigWin2`, `colCOEF`, `CEXLAB`, `LWDtsX`, etc.) that serve to personalise the plots of the series under study and the rolling correlation coefficients, but we do not describe them here because they are not essential for the use of `NonParRolCor`. They are, however, described in the documentation for `NonParRolCor` [14].

### 3.2. Auxiliary data sets

`NonParRolCor` contains three auxiliary data sets: (1) `syntheticdata`, which contains three columns: the first is "time" (from 1 to 500), and the second (named X) and third (named Y) are generated by a bi-variate AR1 (autoregressive process of order 1) process with similar autocorrelation ( $\rho = 0.25$ ), with positive correlation ( $r = 0.85$ ) for the first 250 elements and negative correlation ( $r = -0.85$ ) for the last 250, though the correlation for the whole time interval is zero [15]; (2) `ecodata`, which contains four columns: the first is `Years`, which shows the time (years from 1989 to 2008 with monthly resolution), the second (`SST`) contains monthly anomalies of sea surface temperature (SST) to the south of Grand Canary (28.5 N/16.5 W)<sup>1</sup>, the third (`NAO`) shows the monthly index of the North Atlantic Oscillation (NAO)<sup>2</sup>, and the last column (`CPUE`) gives monthly catches of common octopus (measured as CPUE or Catch Per Unit of Effort) from an artisanal fishery to the Southwest of the island of Grand Canary [17]; and (3) `ecodata2`, which has three columns: the first shows times (years from 1700 to 1936), the second and third columns show reconstructions of total solar irradiance (TSI) [18] and the first principal component (PC1) of the reconstructed Atlantic Bluefin Tuna (BFT) captures [19,20].

### 3.3. Software functionalities with examples

This section contains two subsections: the first describes and exemplifies the main features of the estimation of rolling window correlation for a single window-length and the second does the same but for a band of window-lengths.

<sup>1</sup> SST data can be retrieved from <https://psl.noaa.gov/data/gridded/data.noaa.oisst.v2.html>

<sup>2</sup> NOA data can be retrieved from <https://psl.noaa.gov/data/correlation/nao.data>



Listing 1: Code used to produce Fig. 2.

```
> library("NonParRolCor")
> XY_P <- rolcor_estim_1win(syntheticdata, CorMethod="pearson",
  widthwin=21, MCSim=1000, Np=6)

> plot_rolcor_estim_1win(syntheticdata,
  corcoefs=XY_P$Correlation_coefficients,
  CRITVAL=XY_P$CRITVAL, widthwin=XY_P$widthwin,
  left_win=XY_P$left_win, right_win=XY_P$right_win)
```

Listing 2: Code used to produce Fig. 3.

```
> XY_P <- rolcor_estim_heatmap(syntheticdata[150:350,],
  CorMethod="pearson", typewidthwin="PARTIAL",
  widthwin_1=11, widthwin_N=31, MCSim=1000, Np=6)

> plot_rolcor_estim_heatmap(syntheticdata[150:350,],
  corcoefs=XY_P$matcor, CRITVAL=XY_P$CRITVAL,
  Rwidthwin=XY_P$Windows, typewidthwin="PARTIAL",
  widthwin_1=11, widthwin_N=31, varX="X2", varY="Y2",
  coltsX="brown", coltsY="green")
```

### 3.3.1. Rolling correlation for only one window-length

One of the most important functionalities of NonParRolCor is its ability to estimate the rolling correlation for several different correlation estimators. Fig. 2 shows a comparison between rolling correlation coefficients estimated for a window-length of 21 elements for Pearson and Spearman (two of the most widely used estimators for correlation) using the syntheticdata data set included in NonParRolCor. It is immediately apparent that there is lack of correlation around time 250, as expected. Correlation coefficients have similar values but they vary roughly for each estimator in a such way that coefficients estimated via Spearman take slightly lower values than those via Pearson. The R code used to produce these figures is in Listing 1 (please note that to generate rolling correlations via Spearman you must change the parameter CorMethod to “spearman”).

### 3.3.2. Rolling correlation for a band of window-lengths

To show more of the main functionalities of NonParRolCor, we use the functions rolcor\_estim\_heatmap and plot\_rolcor\_estim\_heatmap to analyse and plot a subset (from 150 to 350) of the variables contained in syntheticdata [14]. As can be seen on the heat maps (Fig. 3) for the rolling correlation coefficients, there are three visual features to help users to analyse their own data when NonParRolCor is used. That is, our R package enables users realise that: (1) there are no statistically significant correlation coefficients for the shortest window-lengths, especially for the heat map correlation estimated via Spearman; (2) correlations around 250 are zero for all window-lengths (from 11 to 31); and (3) Pearson correlation gives better estimations than Spearman, as suggested in Section 3.3.1 for a single window-length.

Estimating rolling correlation for multiple window-lengths may take a little time depending on the number and speed of the CPU cores used. The R code required to produce Fig. 3 is in Listing 2.

## 4. Illustrative examples and comparison with RolWilMulCor

As illustrative examples we provide two snippets containing working R code.

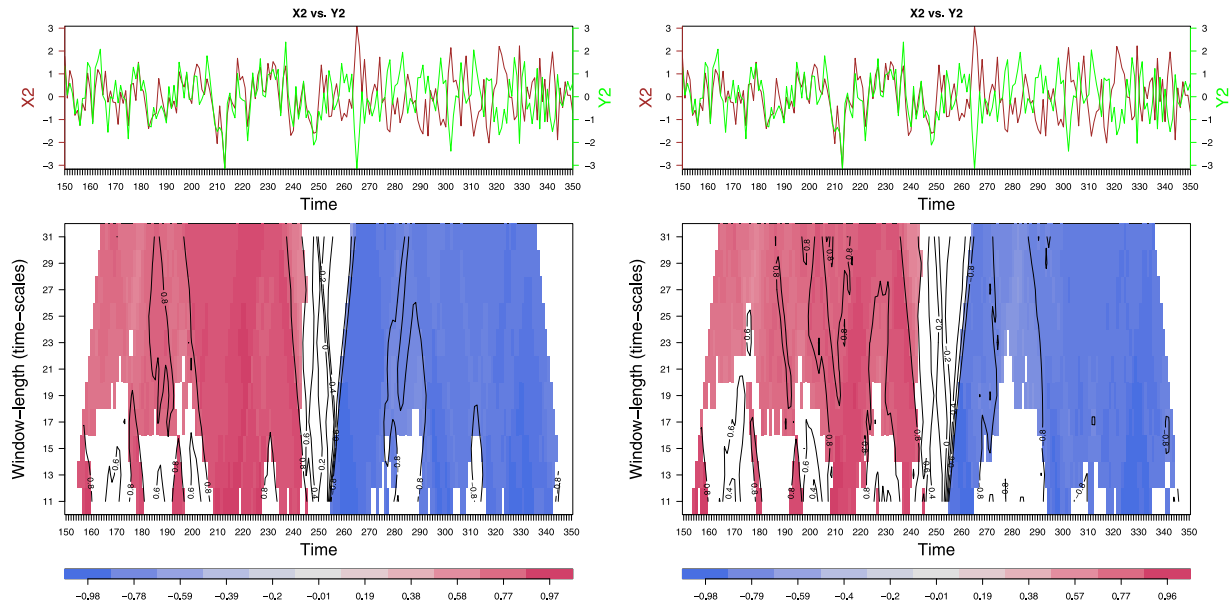
The first snippet (Listing 3) required to produce the first illustrative example (Figs. 4 and 5) shows the R instructions

for estimating and plotting rolling correlations between Total Solar Irradiance (TSI) [18] and the PC1 (principal component 1) of reconstructed Atlantic Bluefin Tuna (BFT) catches [19,20] for window-lengths of 11 and 81 years. We use these time series because in previous studies we found a strong correlation between these two variables [9,19]. These window-lengths were chosen since it is well known that solar irradiance varies at different time scales, e.g. each 11 (Schwabe cycle) or 81 (Gleissberg cycle) years, and it is ideal for showing the features of NonParRolCor.

Fig. 4 (left) shows that there is no statistically significant correlation for the window-length of 11 years for any correlation coefficients except three isolated coefficients, but those significant correlation coefficients could be expected to occur by chance. Fig. 4 (right) was produced without taking into account the multiple comparison problem. As shown, there are several correlation coefficients with p-values of less than 0.05, i.e. these coefficients are statistically significant. However, it must be realised that these rolling correlations do not address the problem of multiple comparison and their significance may be due to chance, but our NonParRolCor R package is able to provide us with information about this problem. The lack of correlation can be explained by a combination of two factors: (1) the time frequency of data analysed (yearly); and (2) the shortness of this window-length (11 data points).

On the other hand, there is a relative high correlation that is statistically significant for the window-length of 81 years (Fig. 5 left), the correlation coefficients of which cover the time interval approximately from 1775 to 1840. We have ruled out the potential effects of multiple comparisons, so these correlations are more likely to represent real associations between the two variables analysed. Fig. 5 (right) shows the rolling correlation estimated via RolWilMulCor [14] for a window-length of 81 years, but these rolling correlations have not undergone any p-value correction method or our non-parametric test. However, the results are practically similar with the exception of a few coefficients which are significant. This indicates that there could be a “cause and effect” relationship between the TSI and the PC1 of tuna captures in the long-term. This result is supported by previous studies [9,19], where the same time series but different statistical techniques were used.

The second code snippet (Listing 4) is used to create Fig. 6. This illustrative example shows the features of the functions



**Fig. 3.** Visual features of the outputs of the `rolcor_estim_heatmap` and `plot_rolcor_estim_heatmap` functions via Pearson (left) and Spearman (right). Rolling correlation coefficients that are not statistically significant (95% confidence level) are blank. Line contours indicate similar values of rolling correlation coefficients. X2 and Y2 are subsets of the two AR1 time series contained in `syntheticdata` [14].

Listing 3: Code used to produce example 1 (Fig. 4).

```
> XYP <- rolcor_estim_1win(ecodata2, CorMethod="pearson",
  widthwin=11, MCSim=1000, Np=6)
> XYS <- rolcor_estim_1win(ecodata2, CorMethod="spearman",
  widthwin=11, MCSim=1000, Np=6)

> plot_rolcor_estim_1win(ecodata2,
  corcoefs=XYP$Correlation_coefficients, CRITVAL=XYP$CRITVAL,
  widthwin=XYP$widthwin, left_win=XYP$left_win,
  right_win=XYP$right_win, varX="TSI", varY="PC1",
  coltsX="orange", coltsY="black", LWDtsX=2, LWDtsY=2)
> plot_rolcor_estim_1win(ecodata2,
  corcoefs=XYS$Correlation_coefficients, CRITVAL=XYS$CRITVAL,
  widthwin=XYS$widthwin, left_win=XYS$left_win,
  right_win=XYS$right_win, varX="TSI", varY="PC1",
  coltsX="orange", coltsY="black", LWDtsX=2, LWDtsY=2)
```

Listing 4: Code used to produce example 2 (Fig. 6).

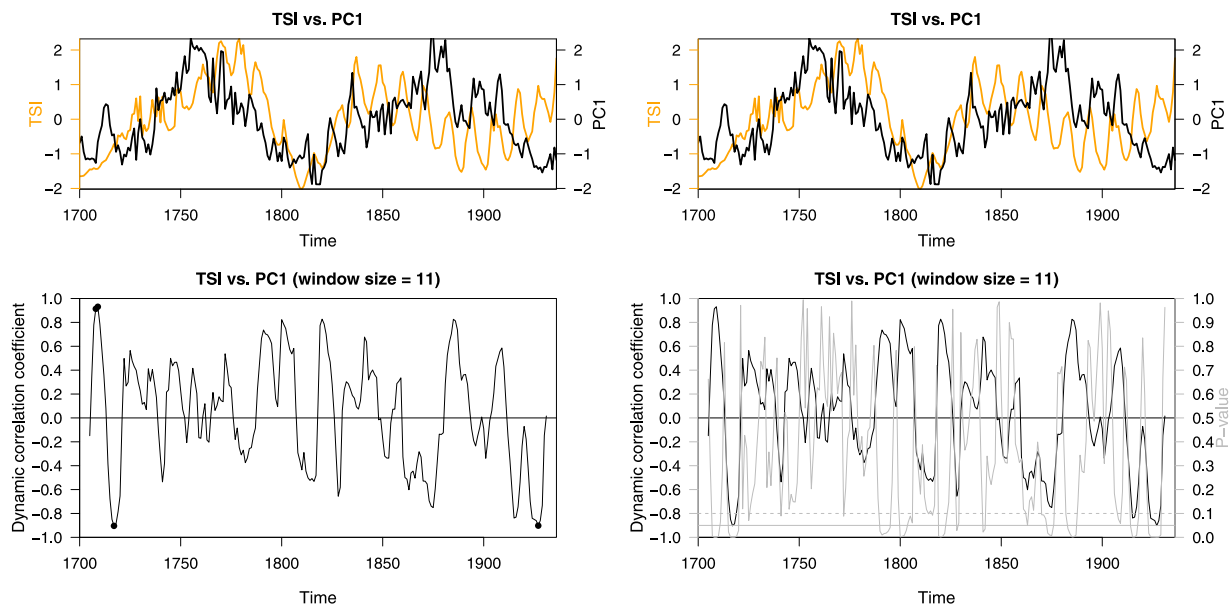
```
> XY_P <- rolcor_estim_heatmap(ecodata2, CorMethod="pearson",
  typewidthwin="FULL", MCSim=1000, Np=5)

> plot_rolcor_estim_heatmap(ecodata2, corcoefs=XY_P$matcor,
  CRITVAL=XY_P$CRITVAL, Rwidthwin=XY_P$Windows,
  typewidthwin="FULL", varX="TSI", varY="PC1",
  coltsX="orange", coltsY="black", LWDtsX=2, LWDtsY=2)
```

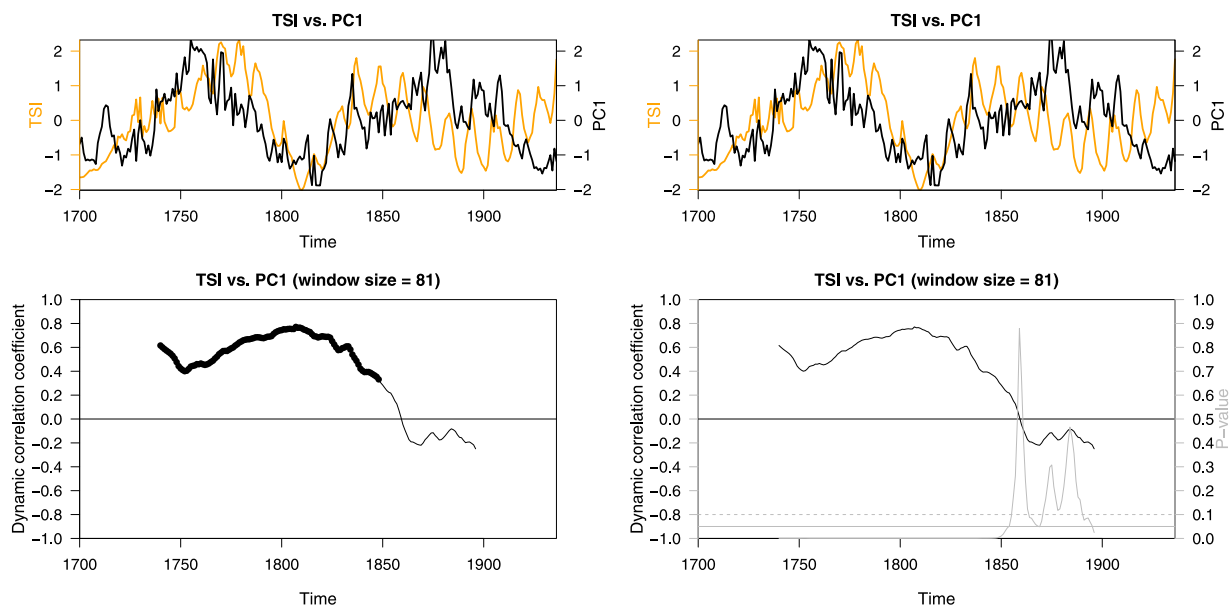
`rolcor_estim_heatmap` and `plot_rolcor_estim_heatmap`. It also compares the statistical significance estimated with our statistical significance test, included in `NonParRolCor`, with the  $p$ -value correction methods of Bonferroni and Benjamini and Hochberg (BH), included in `RolWinMulCor` [14]. The main feature is that `NonParRolCor` works in a way roughly similar to these two  $p$ -value correction methods, but our statistical significance test is a little more restrictive than BH and a little more permissive than Bonferroni. This result is consistent with a previous study where this test was introduced and validated [15].

## 5. Impact and conclusions

One of the most common tasks performed in scientific research is the search for relationships between two or more time series. This is mainly done by estimating the correlation coefficient via Pearson or Spearman. However, correlations estimated for the full study period act as an overall measure of association and relationships between two time series can also vary over time [3,9]. The rolling window correlation functions from the `NonParRolCor` R package presented here help to analyse regular



**Fig. 4.** Visual features of the outputs of `rolcor_estim_1win` and `plot_rolcor_estim_1win` functions from the `NonParRolCor` R package for a window-length of 11 years (left). The second plot (right) was created using the `RolWinMulCor` R package [9]. Rolling correlation coefficients that are statistically significant (95% confidence level) are in black dots (left) and p-values of 0.05 or less are on or below the continuous thick horizontal gray line (right).

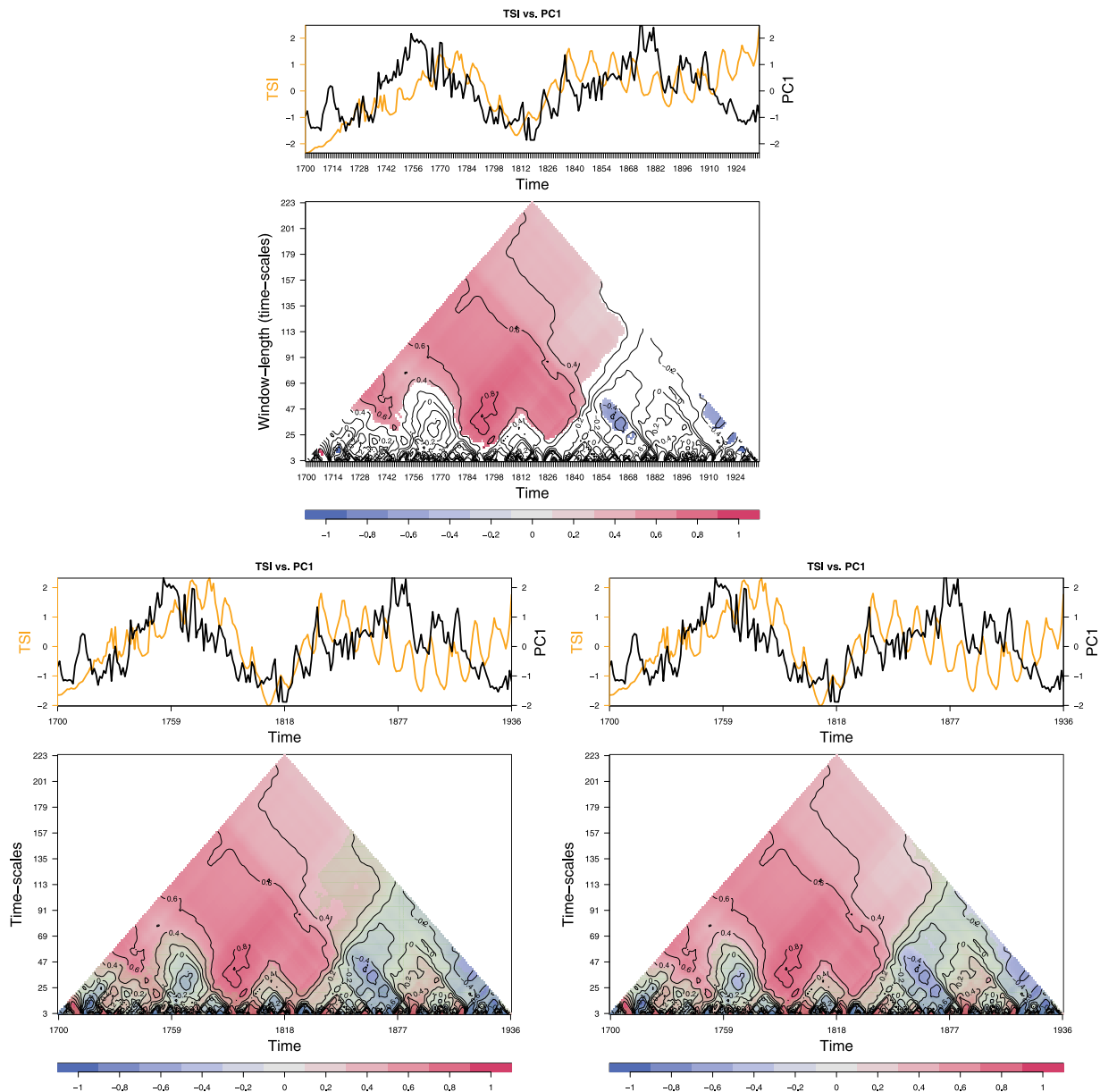


**Fig. 5.** Visual features of the outputs of the `rolcor_estim_1win` and `plot_rolcor_estim_1win` functions from the `NonParRolCor` R package for a window-length of 81 years (left). The second plot (right) was created using the `RolWinMulCor` R package [9]. Rolling correlation coefficients that are statistically significant (95% confidence level) are in black dots (left) and p-values of 0.05 or less are on or below the continuous thick horizontal gray line (right).

time series. In particular, they are useful for finding correlations between two time series for certain subintervals of the series even if overall correlation between the two is weak or non-existent. This makes `NonParRolCor` especially useful for analysing non-stationary time series or when the relationship between two series is not stationary.

There are significant gaps in the literature in regard to R packages for estimating rolling window correlation. Some exceptions are `zoocat` [10], `dendroTools` [13], `TTR` [12] and `roll` [11]. However, these packages do not include an estimation of statistical significance for rolling correlation coefficients and, even worse, none of them includes a way to get around inflation of the Type I error that is produced by repeating multiple significance tests simultaneously. To date there is only one R package,

named `RolWilMulCor`, that includes a estimation of statistical significance that address the multiple comparison problem [9]. Nevertheless, `RolWilMulCor` uses a classical approach that corrects p-values to tackle the multiple comparison problem. The package that we present in this paper, `NonParRolCor`, includes a non-parametric statistical significance test based on Monte Carlo simulations to estimate the significance of rolling correlation coefficients addressing the multiple comparison problem. That test also takes into account the window-lengths for which the rolling correlations are computed. To the best of our knowledge, `NonParRolCor` is the first R package and probably the only software to include this novel methodology for estimating the significance of rolling correlation coefficients. Moreover, rolling window correlation coefficients can be estimated using Pearson,



**Fig. 6.** Heat maps (top centre) of the rolling correlations between TSI and PCI of tuna captures obtained via the `rolcor_estim_heatmap` and `plot_rolcor_estim_heatmap` functions. The other two heat maps are the rolling correlations for the same variables but using the  $p$ -value correction methods of Bonferroni (left) and Benjamini and Hochberg (right). A Pearson estimator is used for all three heat maps. Rolling correlation coefficients that are not statistically significant (95%) are in blank for `NonParRolCor` and in green for `RolWinMulCor` [9]. Line contours indicate similar rolling correlation coefficients. .

Spearman or Kendall, and it is also possible to use different levels of statistical significance.

Having software for estimating rolling window correlation and its statistical significance, especially software freely available on the Internet and accompanied by good documentation as is the case with the `NonParRolCor` R package, enables users to easily analyse their own data by themselves. It is true that `NonParRolCor` was originally developed for environmental research, but it can also be used in other fields, e.g. industry, financial services, etc. In terms of impact and use, CRAN download statistics indicate that `NonParRolCor` was downloaded and installed 5974 times between 31-03-2021 (when the package was published on CRAN) and 31-01-2023. Its sole “competitor” `RolWinMulCor` was downloaded and installed 6139 times during the same period. As can be seen, the level of interest in the two packages is similar despite `NonParRolCor` not having its corresponding software paper having been published on CRAN more recently. For this

reason, we are sure that a software paper for `NonParRolCor` that shows its main features and functionalities, along with adequate documentation, will increase the potential use of this computational package.

`NonParRolCor` is free software (GPL license), can be used by all the main operating systems (Windows, GNU/Linux and MacOS) and is available on CRAN. The non-parametric statistical significance test included in the functions of `NonParRolCor` is a little more restrictive than BH and a little more permissive than Bonferroni. Our package uses parallel computing (via CPU cores) to improve computation time when statistical significance is estimated for rolling window correlation coefficients. `NonParRolCor` contains four functions for estimating and plotting correlation coefficients and their respective statistical significances. The functions available on `NonParRolCor` are highly flexible, and include several parameters for controlling the estimation of correlation and the features of the plot output. The package



also contains three auxiliary data sets to exemplify its use. Finally, we provide some examples that show the usefulness and effectiveness of NonParRolCor.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data availability

Data sets are contained in the R package on CRAN.

### Acknowledgements

JMPM acknowledges to the Excellence Unit GECOS, Spain (reference number CLU-2019-03) at the Universidad de Salamanca for funding support. We would like to show our gratitude to the anonymous reviewers and the Editor Prof. Sobie whose comments and suggestions have been very helpful in improving this paper.

### References

- [1] Pearson K. VII. Mathematical contributions to the theory of evolution.-III. Regression, heredity, and panmixia. *Philos Trans R Soc London Ser A Contain Pap Math Phys Character* 1896;(187):253–318.
- [2] Spearman C. The proof and measurement of association between two things. *J Psychol* 1904;15(1):72–101.
- [3] Hereş A-M, Polanco-Martínez JM, Petritan IC, Petritan AM, Yuste JC. The stationary and non-stationary character of the silver fir, black pine and Scots pine tree-growth-climate relationships. *Agricult Forest Meteorol* 2022;325:109146.
- [4] Polanco-Martínez JM. Dynamic relationship analysis between NAFTA stock markets using nonlinear, nonparametric, non-stationary methods. *Nonlinear Dynam* 2019;97:369–89.
- [5] Cazelles B, Chavez M, Berteaux D, Ménard F, Vik JO, Jenouvrier S, et al. Wavelet analysis of ecological time series. *Oecologia* 2008;156(2):287–304.
- [6] Polanco-Martínez JM, Fernández-Macho J, Medina-Elizalde M. Dynamic wavelet correlation analysis for multivariate climate time series. *Sci Rep* 2020;10(1):1–11.
- [7] Yuan N, Xoplaki E, Zhu C, Luterbacher J. A novel way to detect correlations on multi-time scales, with temporal evolution and for multi-variables. *Sci Rep* 2016;6:27707.
- [8] Lehmann EL, Romano JP. Testing statistical hypotheses. Springer Science & Business Media; 2006.
- [9] Polanco-Martínez J. RolWinMulCor: An R package for estimating rolling window multiple correlation in ecological time series. *Ecol Inform* 2020;60:101163.
- [10] He R-R. Zoocat: 'Zoo' objects with column attributes. 2016, URL <https://CRAN.R-project.org/package=zoocat> R package version 0.0.2.
- [11] Foster J. Roll: Rolling and expanding statistics. 2020, URL <https://CRAN.R-project.org/package=roll> R package version 1.1.6.
- [12] Ulrich J. TTR: Technical trading rules. 2020, URL <https://CRAN.R-project.org/package=TTR> R package version 0.24.2.
- [13] Jevšenak J, Levanič T. dendroTools: R package for studying linear and nonlinear responses between tree-rings and daily environmental data. *Dendrochronologia* 2018;48:32–9.
- [14] Polanco-Martínez JM, López Martínez J. NonParRolCor: A non-parametric statistical significance test for rolling window correlation. 2022, URL <https://CRAN.R-project.org/package=NonParRolCor> R package version 0.8.0.
- [15] Polanco-Martínez JM, López-Martínez JL. A non-parametric method to test the statistical significance in rolling window correlations, and applications to ecological time series. *Ecol Inform* 2021;64:101379.
- [16] Telford R. Running correlations - running into problems. 2013, URL <https://quantpalaeo.wordpress.com/2013/01/04/running-correlations-running-into-problems/>.
- [17] Polanco J, Ganzedo U, Sáenz J, Caballero-Alfonso A, Castro-Hernández J. Wavelet analysis of correlation among Canary Islands octopus captures per unit effort, sea-surface temperatures and the North Atlantic Oscillation. *Fish Res* 2011;107(1–3):177–83.
- [18] Lean J. Evolution of the Sun's spectral irradiance since the maunder minimum. *Geophys Res Lett* 2000;27(16):2425–8.
- [19] Ganzedo U, Polanco-Martínez JM, Caballero-Alfonso A, Faria SH, Li J, Castro-Hernández JJ. Climate effects on historic bluefin tuna captures in the Gibraltar strait and western Mediterranean. *J Mar Syst* 2016;158:84–92.
- [20] Polanco-Martínez JM, Caballero-Alfonso AM, Ganzedo U, Castro-Hernández JJ. A reconstructed database of historic bluefin tuna captures in the Gibraltar strait and western Mediterranean. *Data Brief* 2018;16:206–10.