Software Impacts 16 (2023) 100514

Contents lists available at ScienceDirect

Software Impacts

journal homepage: www.journals.elsevier.com/software-impacts

Original software publication

VisualDom: An R package for estimating dominant variables in dynamical systems

Josué M. Polanco-Martínez *

GECOS-IME, University of Salamanca, Salamanca, Spain
Basque Centre for Climate Change (BC3), Leioa, Spain

A R T I C L E   I N F O

Keywords:
Dynamical systems
Dynamic wavelet correlation
Nonlinear dynamics
MODWT
Multi-scale phenomena
Wavelet multiple correlation

A B S T R A C T

The R package VisualDom estimates and plots the correlation coefficients obtained via the wavelet local multiple correlation and the variables that maximizes the wavelet multiple correlation through time and scale, i.e. the "dominant" variables of a dynamical system. The novel graphical tool to find out dominant variables that we are proposing is able to obtain knowledge from diverse types of dynamical systems, e.g. the climate system. The functions included in VisualDom are quite flexible because these contain multiple parameters for controlling the plot of the time series under analysis and the heat maps of the correlation coefficients and dominant variables.

Code metadata

Current code version	*v0.8*
Permanent link to code/repository used for this code version	https://github.com/SoftwareImpacts/SIMPAC-2023-161
Code Ocean compute capsule	https://codeocean.com/capsule/1853089/tree/v1
Legal Code License	*GPL (>= 2) License*
Code versioning system used	*None*
Software code languages, tools, and services used	*R*
Compilation requirements, operating environments & dependencies	*R (≥ 3.6), wavemulcor, waveslim, plot3D*
If available Link to developer documentation/manual	*https://github.com/jomopo/VisualDom/tree/main/docs*
Support email for questions	josue.m.polanco@gmail.com

1. Introduction to VisualDom

The wavelet local multiple correlation (WLMC) is a signal processing tool that was developed to estimate dynamic wavelet correlation for multivariate time series [1]. This tool is based on the concept of multiple regression and measures a non-stationary time-evolving correlation structure at different scales within a multivariate set of time series. The WLMC consists of one set of multiscale correlations over time, each of them calculated as the square root of the regression coefficient of determination in a linear combination of locally weighted wavelet coefficients for which such coefficient of determination is a maximum. The WLMC produces a set of multi-scale correlations over time and scale that are represented visually as a heat map [1–4]. The WLMC has also an associated R package: `wavemulcor` [5,6], which is available on The Comprehensive R Archive Network (CRAN).

The WLMC is an extension of the work by [7], published a decade ago. Although the statistical signal processing techniques introduced in this work, named wavelet multiple correlation (WMC) and wavelet multiple cross-correlation (WMCC) are not dynamic in time. That is, the WMC and WMCC can be used to estimate the correlation as a global measure at different wavelet scales for a set of multivariate time series, but these techniques cannot be used to study the evolution of correlation though time. The R package `wavemulcor` [5] was originally developed to include these tools, and the R package `W2CWM2C` [8,9] was developed to provide some graphical improvements of the `wavemulcor` [5]. The WLMC was originally developed to analyse financial time series [1], but this tool has been introduced to analyse other kind of data from different scientific disciplines, for example,

The code (and data) in this article has been certified as Reproducible by Code Ocean: (https://codeocean.com/). More information on the Reproducibility Badge Initiative is available at https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals.

* Correspondence to: GECOS-IME, University of Salamanca, Salamanca, Spain.
E-mail address: josue.polanco@usal.es.

https://doi.org/10.1016/j.simpa.2023.100514
Received 13 April 2023; Received in revised form 26 April 2023; Accepted 9 May 2023

energy research [3,10], climate sciences [4], chaotic and nonlinear dynamics [2], technology sector [11], among others.

One of the less exploited features of the WLMC is that this is able to estimate and identify the variables that maximizes the multiple correlation through time and scale for a set of multivariate time series under study that come from a dynamical system [2,4]. This characteristic could be interpreted to lesser extent – although these are not the same – as the phase-difference or coherence phase when the wavelet coherence is estimated via the continuous wavelet transform [2,12]. However, despite the R package `wavemulcor` includes a function to estimate the dominant variables, to the best of our knowledge, there is no exist a graphical representation for the dominant variables, much less a computational software to do it. For this reason, in 2020 we introduced this graphical representation of the WLMC to analyse multivariate climate time series [4], later on in 2022 we showed that this graphical tool is very useful to obtain knowledge from nonlinear and chaotic dynamical systems [2]. However, these works were focused on methodological aspects and science rather than on software applications. Following this necessity, at the end of 2022, we created a set of R functions to implement this quantitative and graphical tool and an R package to contain these functions, which is named `VisualDom`:a tool to visualize dominant variables in dynamic wavelet multiple correlation analysis [13].

`VisualDom` unlike `wavemulcor` that contains a great number of functions only contains three, thereby considerably simplifying the estimation and visualization of the wavelet local multiple correlation. In `VisualDom` one function was created to estimate the WLMC and other two were developed to plot as a heat map the wavelet correlation coefficients and the dominant variables that are statistically significant (within the 95% confidence interval) of the dynamical system under analysis. However, the most important contribution of `VisualDom` is the function used to plot the dominant variables of a dynamical system, which is not included in `wavemulcor` or `W2CWM2C`. Another feature of `VisualDom` is that the graphical functions include a multi-plot format, that is, these functions plot at the same graphic the time series under study and a heat map for the wavelet correlation coefficients or for the dominant variables. We would like to point out that this multi-plot feature contained in `VisualDom` is neither included in `wavemulcor` and `W2CWM2C`. This simple graphical feature helps to visualize at the same time the time series under scrutiny and the dynamic wavelet multiple correlation analysis of these series. The functions contained in `VisualDom` are quite flexible because these contains multiple parameters to personalize the visualization of the time series under study and the heat maps of the wavelet correlation coefficients and dominant variables. Furthermore, we have included in `VisualDom` two data sets named `rdata_climate` and `rdata_Lorenz` to illustrate its use.

This software paper is structured as follows. Section 2 describes the `VisualDom` software in detail. Section 3 provides an illustrative example. Finally, Section 4 presents the impact of this software and proposes some future work.

2. Software description

The R package `VisualDom` is free software (GPL license) and is available on CRAN [13], the main worldwide repository of R packages [14]. Since `VisualDom` is available on CRAN, installing it can be made from R by executing the instruction `install.packages(''VisualDom'')`. `VisualDom` contains three functions in two separate entities. One function is for estimation and the other two are for graphing, as explained visually in Fig. 1. `VisualDom` does not require compilation or others external computational languages and can be installed and used in the main current operating systems (Windows, GNU/Linux and MacOS), due to R is multiplatform, i.e., R runs in multiple computing platforms. The package `VisualDom` depends on the following three R packages: `waveslim` [15], `wavemulcor` [6], and `plot3D` [16], all of which are on CRAN. The first package is used to decompose through the MODWT (Maximal Overlap Discrete Wavelet Transform) the time series under study, the second one is used to estimate the wavelet local multiple correlation, whereas the third package is used to built the heat maps. In addition to these packages, several pieces of code (around 500 lines) in R have been written to estimate and visualize the multivariate time series under scrutiny, the wavelet correlation coefficients and the dominant variables.

2.1. Software functionalities

The R package `VisualDom` includes three functions: (1) `estim_WLMC()` for estimating the wavelet local multiple correlation, which is based on the `wave.local.multiple.correlation()` function contained in the R package `wavemulcor` [5,6]; (2) `plot_estim_WLMC()` for plotting the time series under analysis and the wavelet correlation coefficients as a heat map and uses the output of `estim_WLMC()`; and (3) `plot_dominant_WLMC()` for plotting the time series under study and the dominant variables as a heat maps. `plot_dominant_WLMC()` also uses the output of the `estim_WLMC()` function. A detailed description of their syntaxes is presented in the following lines:

```
(1) estim_WLMC(inputdata, wf="la8", J, window, M, Ymaxr=NULL)
```

- `inputdata`: it is a matrix of N columns (variables or dynamic systems' components under study) by P rows (number of elements of the variables under study). The first column is the time that must be regular or evenly spaced and the other $N-1$ columns are the variables under analysis.
- `wf`: this is the name of the wavelet filter used in the MODWT decomposition of the time series under analysis. There are available several wavelet filters to be used, but we use by default the Daubechies orthonormal compactly supported wavelet of length L = 8, that is, LA(8) or "la8". A relatively long wavelet filter (e.g. LA(4) or LA(8)) is adequate to analyse non-stationary time series but also correlation structures that are not stationary [1,4].
- `J`: this parameter defines the maximum level of the MODWT decomposition and this must be an integer number. There are some works that recommend to use `round`$(log_2(\mathrm{N})-3)$ [1,4]. N is the number elements (rows) of `inputdata`.
- `window`: this is the weight or window function. The default function is the Gaussian window ("gaussian"), but other five window functions can be used: "uniform", "Bartlett's triangular", "Cleveland's tricube", "Wendland's truncated power" and "Epanechnikov's parabolic". For more information look at the function `wave.local.multiple.correlation()` contained in the R package `wavemulcor` [5,6].
- `M`: this is the length of the weight or window function and must be an integer number. It is recommended to use `round`$(\mathrm{N}/8)$ [1,4].
- `Ymaxr`: this parameter is used to indicate which variable will be used to maximize the multiple correlation for each wavelet scale, by default is NULL. That is, we do not define a priori an specific variable but instead let the WLMC select one [1,4].

The output of `estim_WLMC()` is a list named `LISTvals` that contains four elements:

- `CORCOEF`: this element contains the dynamic wavelet correlation coefficients.
- `CIlo` and `CIup`: these two elements are the lower and upper 95% confidence intervals (CIs) of the dynamic wavelet correlation coefficients.
- `YmaxR`: this contains the indices (numbers from one to the number of variables) of the corresponding variables whose correlation is calculated against a linear combination of the rest.

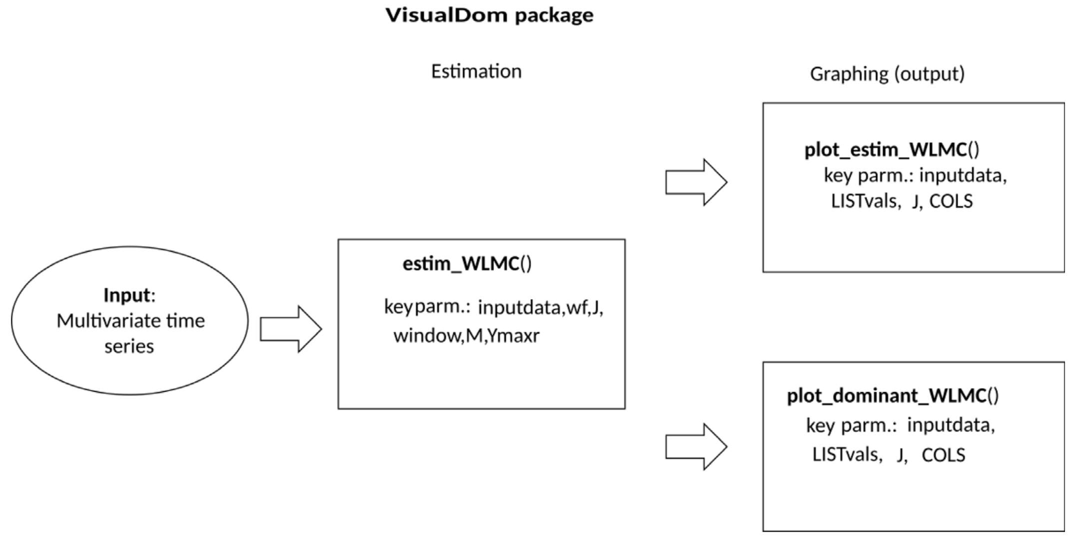

Fig. 1. Software architecture of the R package VisualDom [13]. Given a set of multivariate time series that come from a dynamical system, VisualDom estimates and plots the wavelet local multiple correlation (WLMC) coefficients and the dominant variables or components that are statistically significant (within the 95% confidence interval).

```
(2) plot_estim_WLMC(inputdata, LISTvals, J, fac=1, FLAG=TRUE, FLAGNA=1,
     COLS=c(1:5), LTY=c(rep(1,5)), LWD=c(rep(1.2,5)), DIST=c(seq(0,10,2)))
```

- `inputdata`: this is the same data set used in `estim_estim()`.
- `LISTvals`: it is the list generated by `estim_estim()`.
- `J`: this is the maximum level of the MODWT decomposition and was previously defined.
- `fac`: this parameter is used to scale the wavelet time-scales or "periods" (if the phenomenon studied occurs almost periodically) when the time scale is not the unit, by the default is 1.
- `FLAG`: this flag is used to plot the Y axis of the multivariate time series if the number of these series is less than four, by default is TRUE.
- `FLAGNA`: this is used to plot (by the default is 1) or not (please use 0) the correlation coefficients that are (not) statistically significant (within the 95% confidence interval).
- `COLS`, `LTY`, and `LWD`: these parameters are the colours, the type of lines, and the tick sizes used to plot the multivariate time series under analysis.
- `DIST`: this parameter is used to define the distances among the Y axis when the multivariate time series are plotted.

The output of `plot_estim_WLMC()` is a multi-plot displayed via screen containing the multivariate time series under study and a heat map of the dynamic wavelet correlation coefficients statistically significant of these series. We would like to highlight that the way to address and show graphically the statistical significance is a remarkable feature of `VisualDom`, which is not included in `wavemulcor` [5,6]. This multi-plot can be saved in your preferred format (e.g. JPG, PNG, PDF, etc.).

```
(3) plot_dominant_WLMC(inputdata, LISTvals, J, fac=1, FLAG=TRUE, FLAGNA=1,
     COLS=c(1:5), LTY=c(rep(1,5)), LWD=c(rep(1.2,5)), DIST=c(seq(0,10,2)))
```

All of which have already been described.

The output of `plot_dominant_WLMC()` is also a multi-plot showed via screen, which contain a plot of the time series under scrutiny and a heat map of the dominant variables that are statistically significant (within the 95% confidence interval). This graphical representation of the dominant variables is the main and flagship function of `VisualDom`.

2.2. Auxiliary data sets

`VisualDom` includes two data sets to exemplify its use. The first one `rdata_climate` contains three annual reconstructed climatic indices (anomalies in °C) spanning the years from 500 to 1850 CE, i.e.: (1) sea Surface Temperatures anomalies from the Main Developed Region (SSTMDR) for tropical cyclones [17]; (2) the ENSO (El Niño-Southern Oscillation) SST anomalies (el Niño 3 region) [18]; and (3) the AMO (The North Atlantic Multidecadal Oscillation) SST anomalies [18]. The second data set is named `rdata_Lorenz`, which contains the three components (X, Y and Z) of the Lorenz system [2,13]. The Lorenz system is without a doubt one of the most famous nonlinear dynamical mathematical models and consists of three coupled first-order ordinary differential equations. This dynamical system generates a multivariate set of time series that is ideal to be analysed with the functions contained in `VisualDom`, as was showed recently in [2].

3. Illustrative example

The illustrative example presents a dynamic wavelet correlation analysis for multivariate paleoclimate time series. Specifically, we analyse Last Millennium (LM) relationships among three large-scale annual reconstructed climate variables characterizing the North Atlantic climate system, that is: (1) the SSTMDR for tropical cyclones [17]; (2) the ENSO SST anomalies [18]; and (3) the AMO SST anomalies [18]. These climate variables are included in `rdata_climate`, which is part of `VisualDom`. This data set is loaded automatically when `VisualDom` is called. We examine these large-scale variables since they are known to influence North Atlantic tropical cyclone activity and because their underlying drivers are not well understood [4,19]. We show with this illustrative example that the functions contained in `VisulDom` are useful to shade some light for the study of this complex system, as it was demonstrated few years ago by [4], where a preliminary version of `VisualDom` was used. Although this study was more focused on methodological and scientific aspects rather than software applications. On the other hand, more recently, [9] applied the R package W2CWM2C, in particular, the WMCC function – a different but related approach from a statistical point of view – to the same paleoclimate data set, but in this work it was not possible to analyse the evolution of correlation and dominant variables through time. However, despite of this, the results were quite similar.

Fig. 2 shows the results after applying the functions `estim_WLMC()` and `plot_estim_WLMC()`. The R code used to produce this figure is

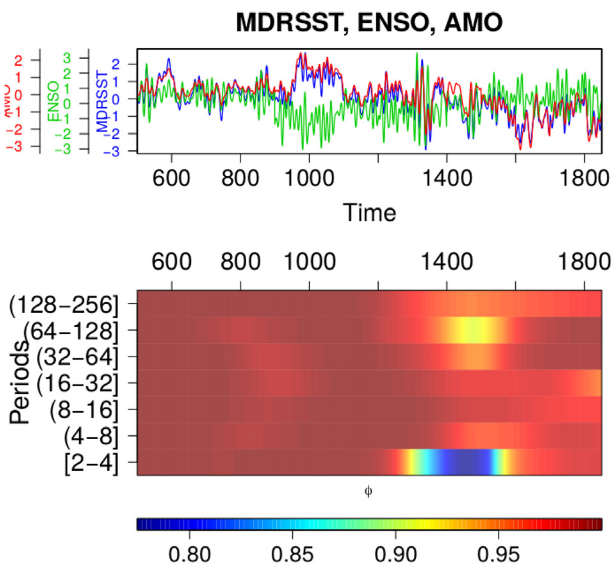

Fig. 2. Multivariate paleoclimate time series and the heat map of the wavelet correlation coefficients statistically significant (95% confidence interval) obtained via the WLMC, which was estimated and plotted through the `estim_WLMC()` and `plot_estim_WLMC()` functions from the R package `VisualDom` [13]. Time and Periods are in years. ϕ indicates the wavelet correlation coefficients.

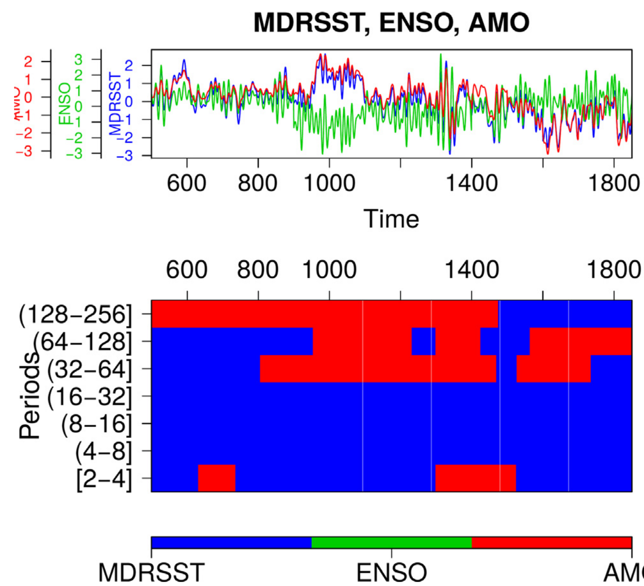

Fig. 3. Multivariate paleoclimate time series and the heat map of the dominant variables statistically significant (95% confidence interval) obtained via the WLMC, which was estimated and plotted through the `estim_WLMC()` and `plot_dominant_WLMC()` functions from the R package `VisualDom` [13]. Time and Periods are in years.

```
library("VisualDom")
inputdata  <- rdata_climate
N          <- nrow(inputdata)
wf         <- "la8"
window     <- "gaussian"
J          <- 7
M          <- 168

LISTvals  <- estim_WLMC(inputdata, wf=wf, J=J, window=
    window, M=M, Ymaxr=NULL)

pdf("WLMC_climate_data.pdf")
plot_estim_WLMC(inputdata, LISTvals=LISTvals, J=J, fac
    =1, FLAG=TRUE, FLAGNA=1,
   COLS=c(4,3,2), LTY=c(rep(1,5)), LWD=c(rep(1.2,5)),
    DIST=c(seq(0,10,2.75)))
dev.off()
```

Listing 1: Code used to produce Fig. 2.

```
LISTvals   <- estim_WLMC(inputdata, wf=wf, J=J, window=
    window, M=M, Ymaxr=NULL)

pdf("dominant_variables_climate_data.pdf")
plot_dominant_WLMC(inputdata, LISTvals=LISTvals, J=J,
    fac=1, FLAG=TRUE, FLAGNA=1,
   COLS=c(4,3,2), LTY=c(rep(1,5)), LWD=c(rep(1.2,5)),
    DIST=c(seq(0,10,2.75)))
dev.off()
```

Listing 2: Code used to produce Fig. 3.

in Listing 1. The fist salient observation is the high level of correlation (from ca. 0.8 to ca. 0.99) for practically the full length of the time series under study, though this correlation is not homogeneous. For instance, there is a short interval approximately from 1300 to 1550 for the shortest wavelet scale where the correlation coefficients take the lowest values. The second salient observation is that the dynamic wavelet correlation analysis can be done observing at the same time the time series under study. This last feature is exclusive to `VisualDom`. As can be seen on Fig. 2, this graphical tool is a useful initial approach for analysing dynamic wavelet correlation for a set of multivariate time series, but it is not enough to discern which are the dominant variables of the phenomenon of study, and a knowledge of this is essential for studying complex dynamical systems as the climate system.

Fig. 3 shows the result after applying the `estim_WLMC()` and `plot_dominant_WLMC()` functions to the multivariate paleoclimate time series. The corresponding R code used to produce this figure is in Listing 2. We do not define a priori an specific climate variable that would maximize the multiple correlation for each wavelet scale (parameter `ymaxr=NULL`) but instead let the WLMC select one. The reason for this choice is that although these climate variables are expected to be correlated among them, their causal relationship remains a matter of research [4,19]. On the one hand, it is clear that the dominant variable for the shortest and medium wavelet scales (from 2 to 32 years) is the SSTMDR, with some exceptions for the shortest scale (2–4 years) and for the time intervals around 600–700 years and 1300–1500 years, where the dominant variable is AMO. On the other hand, the dominant variable for the largest wavelet scales (from 32 to 256 years) is AMO, as expected. This result is similar to the one obtained by [9] via `W2CWM2C` (please look at Fig. 1 right) although the statistical technique used is different (that is, the WMCC) and does not take into account the time dimension. It is noticeable to observe that the time interval 1300–1500 years is precisely the same interval that in Fig. 2 has the lowest wavelet correlation coefficients. It is also remarkable to note that this does not make much sense that AMO is a dominant variable in this short wavelet scale, since AMO has its main spectral signature multidecadally [18,20]. Even still, it is possible that AMO might be played an important role in the interannual variability during these time intervals. Although it is beyond the aim of this paper to address this issue.

4. Impact and future work

The computational R package `VisualDom` is designed for scientists and researchers from different fields of physical sciences who would like to conduct applied research on multivariate time series analysis that come from dynamical systems. VisualDom is easy to use, is well documented, is free software (GPL license) and is available on CRAN [13], the official package repository of the R project. The main contribution included in VisualDom is the function plot_dominant_WLMC(), which display graphically the dominant variables that maximize the multiple correlation for a set of multivariate time series. This tool is used to find which variables play an important role in different time-scales and temporal intervals in a group of multivariate time series under study. This is especially important in physics research for the study of nonstationary and nonlinear phenomena that take place at different time scales, frequencies and temporal periods, such as the climate system. This novel graphical tool has an enormous potential to be used with different types of multivariate time series, see e.g., [2–4]. For this reason, we are sure that VisualDom will be also useful for a wide range of scientific and engineering disciplines.

VisualDom contains three functions: (1) estim_WLMC(), (2) plot_estim_WLMC() and (3) plot_dominant_WLMC(); and two data sets: (1) rdata_climate and (2) rdata_Lorenz) to illustrate the use of these functions. The functions included in VisualDom are highly flexible, since these functions contain multiple parameters for controlling the plot of the time series under analysis and the heat maps of the wavelet correlation coefficients and dominant variables. We would like highlight the flagship feature of the plot_dominant_WLMC() function, i.e. the graphical representation of dominant variables, which is not included in the R package wavemulcor [5,6], and to the best of our knowledge, this has not been implemented in any other computational package.

Future plans include the implementation of some functions to be added to VisualDom. For instance, a graphical tool to estimate and to represent the dynamical wavelet local multiple cross-correlation (WLMCC) [1,5]. Other interesting functionalities can be the implementation of some functions that allow the use of irregular time series, which address irregular (unevenly spaced) time series that are common in climatology, oceanography, geophysics, among others. Furthermore, more sophisticated future directions of research include the computational implementation of causality tests to be added to VisualDom, which will be used to infer causation between multiple variables of dynamical systems.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Josué M. Polanco Martınez declares no potential conflict of interests.

Acknowledgements

JMPM acknowledges to the Excellence Unit GECOS (reference number CLU-2019-03), Universidad de Salamanca for funding support.

References

[1] J. Fernández-Macho, Time-localized wavelet multiple regression and correlation, Phys. A Stat. Mech. Appl. 492 (2018) 1226–1238.

[2] J. Polanco-Martínez, A computational and graphical approach to analyze the dynamic wavelet correlation among components of a nonlinear dynamical system, J. Appl. Nonlinear Dyn. (2023) 1–12, (Forthcoming).

[3] J. Polanco-Martínez, L. Abadie, J. Fernández-Macho, A multi-resolution and multivariate analysis of the dynamic relationships between crude oil and petroleum-product prices, Appl. Energy 228 (2018) 1550–1560.

[4] J. Polanco-Martínez, J. Fernández-Macho, M. Medina-Elizalde, Dynamic wavelet correlation analysis for multivariate climate time series, Sci. Rep. 10 (1) (2020) 1–11.

[5] J. Fernandez-Macho, Package wavemulcor: Wavelet multiple regression and correlation in R, Comput. Sci. Eng. 21 (2019) 63–73.

[6] J. Fernández-Macho, Wavemulcor: Wavelet routines for global and local multiple correlation, 2021, URL https://CRAN.R-project.org/package=wavemulcor. R package version 3.1.2.

[7] J. Fernández-Macho, Wavelet multiple correlation and cross-correlation: A multiscale analysis of Eurozone stock markets, Phys. A Stat. Mech. Appl. 391 (4) (2012) 1097–1104.

[8] J.M. Polanco-Martínez, F.J. Fernández-Macho, Package W2CWM2C: description, features, and applications, Comput. Sci. Eng. 16 (6) (2014) 68–78.

[9] J.M. Polanco-Martínez, W2CWM2C reloaded: Ten years later, Softw. Impacts 16 (2023) 100495.

[10] K. Raath, K. Ensor, Time-varying wavelet-based applications for evaluating the Water-Energy Nexus, Front. Energy Res. 8 (2020) 118.

[11] Z. Umar, N. Trabelsi, F. Alqahtani, Connectedness between cryptocurrency and technology sectors: International evidence, Int. Rev. Econ. Finance 71 (2021) 910–922.

[12] J. Polanco, U. Ganzedo, J. Sáenz, A. Caballero-Alfonso, J. Castro-Hernández, Wavelet analysis of correlation among Canary Islands octopus captures per unit effort, sea-surface temperatures and the North Atlantic Oscillation, Fish. Res. 107 (1–3) (2011) 177–183.

[13] J. Polanco-Martínez, VisualDom: Visualize dominant variables in wavelet multiple correlation, 2022, URL https://CRAN.R-project.org/package=VisualDom. R package version 0.8.0.

[14] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2020, URL https://www.R-project.org/.

[15] B. Whitcher, The waveslim Package, 2022, URL https://CRAN.R-project.org/package=waveslim. R package version 1.8.4.

[16] K. Soetaert, plot3D: Plotting multi-dimensional data, 2021, URL https://CRAN.R-project.org/package=plot3D. R package version 1.4.

[17] M. Mann, J. Woodruff, J.P. Donnelly, Z. Zhang, Atlantic hurricanes and climate over the past 1,500 years, Nature 460 (7257) (2009) 880–883.

[18] M. Mann, Z. Zhang, S. Rutherford, R. Bradley, M. Hughes, D. Shindell, C. Ammann, G. Faluvegi, F. Ni, Global signatures and dynamical origins of the Little Ice Age and Medieval Climate Anomaly, Science 326 (5957) (2009) 1256–1260.

[19] M. Medina-Elizalde, J.M. Polanco-Martínez, F. Lases-Hernández, R. Bradley, S. Burns, Testing the "tropical storm" hypothesis of Yucatan Peninsula climate variability during the Maya Terminal Classic Period, Quat. Res. 86 (2) (2016) 111–119.

[20] C.M. Patricola, R. Saravanan, P. Chang, The impact of the El Niño–Southern Oscillation and Atlantic meridional mode on seasonal Atlantic tropical cyclone activity, J. Clim. 27 (14) (2014) 5311–5328.