J. M. Polanco-Martínez and F. J. Fernández-Macho, "*Package W2CWM2C: Description, Features, and Applications*," in **Computing in Science & Engineering,** vol. 16, no. 6, pp. 68-78, Nov.-Dec. 2014, doi: 10.1109/MCSE.2014.96.

Package W2CWM2C: description, features and applications

Josué M. Polanco-Martínez*(1) **F. Javier Fernández-Macho(2)**
(1) BC3 - Basque Centre for Climate Change
Alameda Urquijo 4, 4o 1a | 48008 Bilbao, Bizkaia (Spain)
(2) Dept. of Econometrics and Statistics and Institute for Public Economics
University of the Basque Country, Spain
*E-mail: `josue.m.polanco@gmail.com, josue.polanco@bc3research.org`

Abstract

This article describes new graphical tools that help with the interpretation of wavelet correlation and cross-correlation analysis, in both bivariate and multivariate cases. We have created an R package called **W2CWM2C** with the following main functions: Wavelet (Cross) Correlation and Wavelet Multiple (Cross) Correlation. To illustrate its strengths, we use data from several European Union stock market indices, comparing the new graphical output with the classical plots.

Keywords: maximal overlap discrete wavelet transform, wavelet correlation, wavalet cross-correlation, wavelet multiple correlation, wavelet multiple cross-correlation.

1. Introduction

Engineers and scientists often seek to identify relationships between variables. For this, one of the most widely used statistical tools is the estimation of correlation coefficients (e.g., Kendall's, Pearson's, Spearman's, etc.) together with the cross-correlation function that operates on the whole time span of the variables involved. However, many phenomena studied in several branches of engineering (environmental, electric, hydraulic, etc.) and scientific disciplines (physics, climatology, ecology, economics, finance, etc.) operate on several different time scales and horizons (multi-scale phenomena) and their variables, commonly represented as time series, are not stationary in the sense that their statistical properties (including the mean, variance, etc.) change with time [4, 8]. This presents a problem, as traditional correlation analysis techniques are designed to analyze stationary data and do not allow multi-scale analysis.

In this context, the use of the wavelet correlation (WC) and cross-correlation (WCC) to study the relationship between variables (either using discrete or continuous wavelet transforms) has been gaining popularity due to the fact that these statistical tools are able to handle multi-scale analysis and non-stationarity. In essence, WCs and WCCs analyze the association between time series in the combined time-and-scale domain performing a multi-scale analysis. Especially since the publication of a the seminal paper by Whitcher *et al.* [12] and of textbooks on the subject [see [4], [8], among others] as well as the release of the **waveslim** package [11] as free software, these types of correlations have been widely used in many different scientific fields including the atmospheric sciences, ecology, geophysics, etc., and more recently, economics and finance [4, 8, 12]. However, to our knowledge, few improvements have been made either to the wavelet correlation methodology or the wavelet graphical tools. One exception is the work of Fernández-Macho [2], who provided an extension from the bivariate to the multivariate case and also released a computational package **wavemulcor** [3] for applying this new methodology.

The **waveslim** and **wavemulcor** packages [3, 11] do help considerably with wavelet correlation analysis. But they have a key weakness: the graphical representation of the output from the wavelet correlation functions contained in the packages. For example, when the `wave.correlation` function [**waveslim** package; 11] is used to analyze several time series (see Figure 4), the number of wavelet correlation plots is usually large, equal to $C_{N,2} = N!/(N-2)!2!$ (that is, the number of combinations without repetition, where N is the number of time series under study), making the analysis of the results rather cumbersome in practice. What is more, when we are working with wavelet cross-correlations, the `spin.correlation` function [**waveslim** package; 11] (and to some extent `wave.multiple.cross.correlation` in the **wavemulcor** package; [3]) the number of plots generated increases even further, due the multiple pairwise comparisons (see Figures 6 and 8). Specifically, with `spin.correlation`, it depends on the number of wavelet scales J and the number of time series under study N (that is, $J \times C_{N,2}$ pairwise comparison plots), and while `wave.multiple.cross.correlation` reduces this somewhat, you still need as many plots as there are wavelet scales J.

To address these difficulties for visualizing the wavelet correlation outputs, and thereby facilitate the interpretation of wavelet correlation analysis, we have developed three novel graphical tools that we have called `WC` ("wavelet

Digital Object Indentifier 10.1109/MCSE.2014.96 1070-9924/$26.00 2014 IEEE

This article has been accepted for publication in Computing in Science and Engineering but has not yet been fully edited.
Some content may change prior to final publication.

correlation"), `WCC` ("wavelet cross-correlation") and `WMCC` ("wavelet multiple cross-correlation"), implemented as an `R` package named **W2CWM2C**. The graphs that we are proposing are based on the following functions: `wave.correlation`, `spin.correlation` (bivariate case) [**waveslim**; 11] and `wave.multiple.cross.correlation` (multivariate case) [**wavemulcor**; 3] respectively. We believe that they would be useful for scientists, engineers, economists, indeed researchers in any field, interested in applying wavelet correlation analysis to study correlations between several, possibly non-stationary, time series characterized by multi-scale phenomena.

The main idea behind these graphical tools is to summarize the key information contained in classical wavelet (cross)-correlation outputs in a single plot regardless of the dimensionality of the multivariate set involved, dramatically reducing the number of plots used to represent wavelet correlation analysis. With the plots produced, users can figure out at a glance which time series pairs or groups are correlated, as well as the strength of this correlation and at what wavelet scales the correlation is strongest or weakest.

That is, the new functions give two types of output. On the one hand, the `WC` function helps to analyze wavelet correlations by generating a single plot instead of $N!/[(N-2)!2!]$ plots and it provides a table of the wavelet correlation coefficients (see Figure 3). On the other hand, the `WCC` and `WMCC` functions greatly reduce the number of plots used to represent wavelet cross-correlations, and they also indicate graphically where in time, i.e., at what time lag, there are the strongest wavelet correlations (see Figure 5 and 7). In addition, note that the `WMCC` plot is based on a new wavelet correlation technique [2] that is able to handle multivariate time series. Wavelet multiple correlation analysis is a recently-develop subfield of wavelet analysis.

The article is organized as follows. Section 2 provides a brief introduction to wavelet correlation analysis. Section 3 describes the **W2CWM2C** package. In Section 4, we briefly outline the main features of **W2CWM2C** through some examples (bivariate and multivariate cases) using stock market time series. We have used these series for the purposes of illustration since they are a good example of the type of data that can be treated with the package, being considered, in general, non-stationary and containing information at multiple scales -or frequencies- and they are freely available on the internet. We underline, however, that other types of data from different scientific disciplines and engineering branches can also be analyzed. Finally, in Section 5, we summarize what can be achieved by using this package.

2. Wavelet correlation analysis

Fourier analysis-based methods are among the most useful statistical ap-

Digital Object Indentifier 10.1109/MCSE.2014.96 1070-9924/$26.00 2014 IEEE

proaches for time series analysis. Specifically, the Fourier transform decomposes the variance of a time series into its components in the frequency domain and it is an invaluable aid for identifying periodic events and other data features. However, this technique is designed for the analysis of stationary data and many types of time series (climatic, environmental, financial, geophysical, medical, etc.) rarely satisfy this requirement. One extension of Fourier analysis to tackle the problem of nonstationarity is the windowed Fourier transform. However, the window width, and hence time resolution, is assumed constant for all decomposed frequencies, resulting in an over- and under-representation of high and low frequency components respectively. Consequently, windowed analysis is both inefficient and inaccurate as a method of time-frequency localization [6, 10].

Wavelet analysis is a statistical tool that improves on the windowed Fourier transform (and other similar techniques) overcoming the problem of non-stationarity in noisy time series. In contrast to Fourier analysis, wavelet analysis uses *wavelet* ("small wave") functions as a basis, instead of sines and cosines, and performs a localized time-scale (frequency) decomposition of a time series determining the dominant modes of variability and how those modes vary in time and scale [1, 4, 8, 10].

There are two approaches to performing wavelet analysis: on the one hand, the continuous wavelet transform (CWT) and, on the other, the discrete wavelet transform (DWT) [8, 10]. The CWT contains a large amount of redundant information (neighbouring scales and times) and has a high computational cost, whereas the DWT uses a limited number of time and scale decompositions avoiding redundant information [4].

In this article, we focus on a non-orthogonal version of the DWT, the maximal overlap discrete wavelet transform (MODWT). While the MODWT is similar to the classical one in many respects, the choice is not arbitrary as the MODWT offers certain advantages. To start with, the MODWT can handle any sample size N, whilst the DWT of level J restricts the sample size to a multiple of 2^J. Further, unlike for a DWT, wavelet and scaling coefficients from an MODWT are invariant to circularly shifting the time series under analysis and its multiresolution detail and smooth coefficients are associated with zero-phase filters. Finally, the wavelet variance estimator based on the MODWT is asymptotically more efficient than that based on the DWT, which in turn makes it more suitable when calculating wavelet correlations [2, 4, 7, 8].

In pioneering work, Whitcher *et al* [12] used the MODWT to extend bivariate wavelet analysis to study the degree of association between two time series. They provided estimators to calculate the wavelet covariance, correlation and cross-correlation, and their corresponding confidence intervals. The first application of this methodology was to analyze climatic time series. Since then, however, it has been applied to many kinds of non-stationary

Digital Object Indentifier 10.1109/MCSE.2014.96 1070-9924/$26.00 2014 IEEE

This article has been accepted for publication in Computing in Science and Engineering but has not yet been fully edited.
Some content may change prior to final publication.

time series (with [12] having been cited 66 and 170 times in peer-reviewed publications, according to *Web of Science* and *Google Scholar* respectively at the time of writing, and the trend is growing).

More recently, Fernández-Macho [2] developed a multivariate wavelet correlation technique based on MODWT to calculate wavelet multiple correlations and cross-correlations together with their corresponding confidence intervals. The wavelet multiple correlation consists of a single set of multi-scale correlations each of them calculated as the square root of the regression coefficient of determination corresponding to the linear combination of wavelet coefficients that maximizes this coefficient of determination. On the other hand, the wavelet multiple cross-correlation is obtained similarly but additionally it allows a certain number of lags between observed and fitted values from the same linear combination as before at each of the wavelet scales [2]. Since its publication this novel methodology has mainly been applied to financial time series. However, the potential field of applications is much broader, including many areas of science and engineering where there is a need to analyze a multivariate set of variables.

3. Description of the package

The **W2CWM2C** package [9] developed in R ver. 2.14.1 runs on all major operating systems and it is available from the CRAN repository (`http://cran.r-project.org/web/packages/W2CWM2C/index.html`) and from the corresponding author upon request. It has been programmed in R to be run from the command line. The **W2CWM2C** package contains four functions (namely, `WC`, `WCC`, `WMC` and `WMCC`) and their respective outputs can be displayed on the screen or be saved as PNG, JPG, PDF or EPS graphics files. **W2CWM2C** depends on the **waveslim** [11] and **wavemulcor** [3] packages to calculate the wavelet correlations, as well as on the **colorspace** package [5] to plot the heatmaps.

Figure 1 provides some limited information about usage statistics of the **W2CWM2C** package and direct dependents. It shows their weekly downloads (together with a `loess` fit line and 95% confidence limits) made by unique IP addresses from the RStudio "0-Cloud" CRAN mirror between October 2012 and March 2014. It emerges that the combined total number of "0-Cloud" downloads of **wavemulcor** + **W2CWM2C** packages during that period amounts to 3960 unique installations. This means a historical average of about 23 new installations per week for each of these two packages. In this respect, we note that these data do not include downloads from the primary CRAN mirror or any of the 90 other CRAN mirrors nor do they include downloads prior to the release of the first RStudio log file (Oct 2012) (nor, indeed, manual package installs from non CRAN downloads). Therefore, the actual number of total downloads differ by a substantial factor.

Digital Object Indentifier 10.1109/MCSE.2014.96 1070-9924/$26.00 2014 IEEE

This article has been accepted for publication in Computing in Science and Engineering but has not yet been fully edited.
Some content may change prior to final publication.

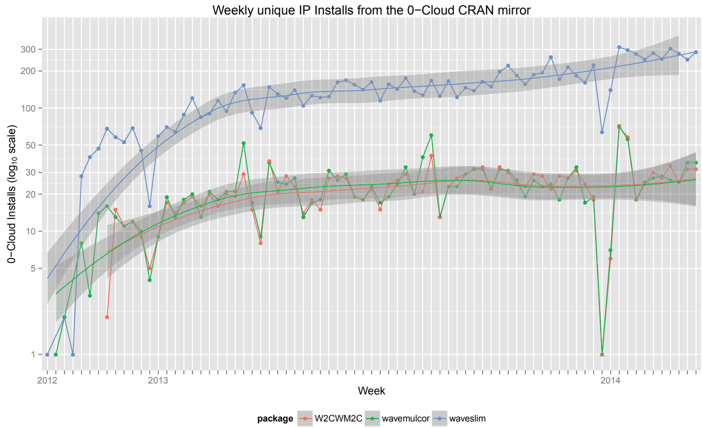

Figure 1: Usage statistics of the package **W2CWM2C** based on the RStudio "0-Cloud" CRAN mirror from October 2012 to March 2014.

However, assuming that all package sources have similar time behaviour, we estimate that the number of installs is increasing at a monthly growth rate of 7%. Finally, the map in Figure 2 shows the geographical distribution of the **W2CWM2C** package. It can be seen that USA, Japan, China, India, Korea and Germany are the countries where the package has generated the greatest interest so far.

The first function, called `WC`, calculates the wavelet correlation (bivariate case) by means of the routine `wave.correlation` from the **waveslim** package [11] and uses the `heat_hcl` function from the **colorspace** package [5] to plot the heatmap. The `WC` function can be applied to between two and

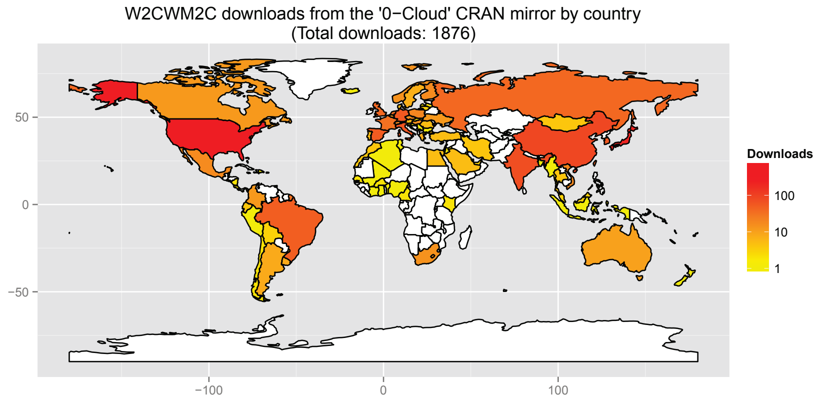

Figure 2: Geographical distribution of the **W2CWM2C** package based on the RStudio "0-Cloud" CRAN mirror from October 2012 to March 2014.

Digital Object Indentifier 10.1109/MCSE.2014.96 1070-9924/$26.00 2014 IEEE

This article has been accepted for publication in Computing in Science and Engineering but has not yet been fully edited.
Some content may change prior to final publication.

seven time series (if you need to analyze more than seven time series, we strongly recommend using the wavelet multiple correlation, implemented in the **W2CWM2C** package as the `WMC` function or the `wave.multiple.correlation` function implemented in the **wavemulcor** package [3]). It generates a plot of the $C_{N,2}$ pairwise comparisons, which are plotted in descending order with respect to the sum of all the wavelet correlation coefficients for each pairwise comparison. The WC plot (see Figure 3) also contains a table of the wavelet correlation coefficients (within a 95% confidence interval) from the pairwise comparisons. The `WC` function has the following syntax:

```
R> WC(inputDATA, Wname, J, device="screen", filename,
        Hfig, WFig, Hpdf, Wpdf)
```

where `inputDATA` is the input data, here an array of multivariate time series as a time series object; `Wname`, the wavelet function or filter to be used in the decomposition (by default the Daubechies least asymmetric wavelet filter of length 8 or “la8”); `J`, the wavelet the depth of the decomposition (the maximum decomposition level J is given by $log_2(N)$ [4]; `device`, the type of the output device (“screen” by default, the other options being “jpg”, “png”, “eps” and “pdf”); `filename`, the output filename; and `Hfig, WFig, Hpdf`, and `Wpdf` the height and width of the output for the jpg and png format and for eps and pdf formats respectively.

`WC` also generates a list of numerical values as output. On the one hand, `wavcor.modwtsDAT` is a $C_{N,2} \times J \times 3$ array; its first element, `wavcor.modwtsDAT`$[,,1]$, contains the wavelet correlation coefficients (which are used to make the WC plot) and the other two elements are the lower and upper bounds of the confidence interval, respectively. On the other hand, the array `to3DpL` contains only the wavelet correlation coefficients (`wavcor.modwtsDAT`$[,,1]$), but the it has been sequenced in ascending order with respect to the wavelet scale `J`.

The second function, called `WCC` (for the bivariate case), calculates the wavelet cross-correlation using the `spin.correlation` function from the **waveslim** package [11] and uses the `heat_hcl` function from the **colorspace** package [5] to plot the heatmap. The `WCC` function presents the wavelet cross-correlation (within a 95% confidence interval) in a novel way (see Figure 5) that reduces the number of plots compared to the classical function `spin.correlation`. The `WCC` contains the following elements:

```
R> WCC(inputDATA, Wname, J, lmax, device="screen", filename,
        Hfig, WFig, Hpdf, Wpdf)
```

where the arguments `inputDATA, Wname, J, device, filename, Hfig, WFig, Hpdf` and `Wpdf` are as described above, and the `lmax` indicates the

Digital Object Indentifier 10.1109/MCSE.2014.96 1070-9924/$26.00 2014 IEEE

maximum lag for which the wavelet cross-correlation is calculated (up to 30 by default, the value being variable depending on the length of the data under study). The numerical output of the `WCC` function is a multivariate time series `returns.cross.cor` of dimensions $(2\,\mathtt{lmax} + 1) \times J$, containing the wavelet correlation coefficients for each lag and level.

The third function, `WMC` (for the multivariate case), calculates the wavelet multiple correlation by means of the function `wave.multiple.correlation` from the **wavemulcor** package [3], but it also provides a way to handle multivariate time series easily as a list of N time series (the function `wave.multiple.correlation` requiring a list of N time series as input data). Furthermore, the `WMC` function is also able to save the output plot as jpg, png, eps or pdf graphics formats. The `WMC` function has the following syntax:

```
R> WMC(inputDATA, Wname, J, device="screen", filename,
       Hfig, WFig, Hpdf, Wpdf)
```

all have which have already been described. The numerical output of the function `WMC` is the list `LS`, which contains two objects, `xy.mulcor` and `YmaxR` (the same output values as the function `wave.multiple.correlation` from the **wavemulcor** package; [3]). The former contains a $J \times 3$ matrix, in which the first column corresponds to the wavelet multiple correlation coefficients, while the other two columns are the lower and upper bounds of the confidence interval. The latter is a numeric vector giving, at each wavelet level, the index number of the variable that gives the maximum correlation against a linear combination of the rest of the variables under study [2, 3].

Finally, the fourth function `WMCC` (multivariate case), calculates the wavelet multiple cross-correlation by means of the function `wave.multiple.cross.correlation` from the **wavemulcor** package [3] and it also uses the `heat_hcl` function from the **colorspace** package [5] to build the heatmap. We have added some improvements, namely a reduction in the number of plots used with respect to the typical plot from the function `wave.multiple.cross.correlation` (see Figure 7 compared to 8) and the functionality added to tackle multivariate time series easily as a list of N time series (the function `wave.multiple.cross.correlation` also needs a list of N time series as input data). The `WMCC` function has the following syntax:

```
R> WMCC(inputDATA, Wname, J, lmax, device="screen", filename,
       Hfig, WFig, Hpdf, Wpdf)
```

all these elements having been defined above. Besides the WMCC plot, function `WMCC` generates as output a list of two elements (like the function `wave.multiple.correlation`), but `xy.mulcor` includes one more dimension to take into account the lags.

Digital Object Indentifier 10.1109/MCSE.2014.96 1070-9924/$26.00 2014 IEEE

The **W2CWM2C** package also includes a set of multivariate time series of several European stock market indices (daily closing prices) FTSE MIB30 (Italy), IBEX35 (Spain), DAX30 (Germany), CAC40 (France), AEX25 (Netherlands), ATX20 (Austria) and NBEL20 (Belgium) spanning from 2 January 2004 to 29 June 2012. In order to cope with the different official holidays, we have adjusted the raw market indices data, carrying forward the daily closing price from the last working day before each of these holidays. The data set is called `dataexample` and is loaded using the instruction `load(dataexample)` (to visualize the name of each variable use the instruction `names(dataexample)`) in the command line of R. The data were obtained from ***Yahoo*** and the dimensions of this array are 2216×7 (number of elements $\times$ columns or variables).

4. Features and applications

4.1. The bivariate case

The first example in this paper is the application of the wavelet correlation pairwise comparisons between time series. We have used five time series, and hence, without repetition, there are 10 combinations. Figure 3 shows the wavelet correlation pairwise comparisons (C#-C#) of the daily log-returns of the indices of five European stock markets (IBEX35, FTSE MIB30, CAC40, DAX30 and AEX25), numbered 1 to 5 respectively, and was built using the `WC` function. The colour scale used in the figure indicates the strength of the wavelet correlation coefficients, from weak correlations in "blue" through to strong ones in "pink". At the same time, the wavelet correlation coefficients that are significantly different from zero (within a 95% confidence interval) are reported in a table superimposed on the plot.

Remarkable differences can be noted between our WC plot Figure 3 and the WC classical plots (Figure 4, using the `wave.correlation` function). Firstly, when we are working with multiscale statistical techniques, the number of wavelet correlation plots and pairwise comparisons tends to be large. In the present case, Figure 4 packs all the information contained in $C_{N,2} = 10$ plots into a single plot. Secondly, in Figure 3 the WC coefficients themselves are printed, unlike in Figure 4. Thirdly, Figure 3 makes it possible to appreciate the differences and similarities between wavelet scales, and this facilitates the comparison between pairs of wavelet correlations from the set of time series under study. On the other hand, our plot does not show the WC coefficients as a curve, although this information is reflected in the colours and the WC coefficients given as a table in our plot.

Digital Object Indentifier 10.1109/MCSE.2014.96 1070-9924/$26.00 2014 IEEE

This article has been accepted for publication in Computing in Science and Engineering but has not yet been fully edited.
Some content may change prior to final publication.

Figure 3: Wavelet correlation (new plot) between five EU market returns, MIB30 (C1), IBEX35 (C2), DAX30 (C3), CAC40 (C4), and AEX25 (C5).

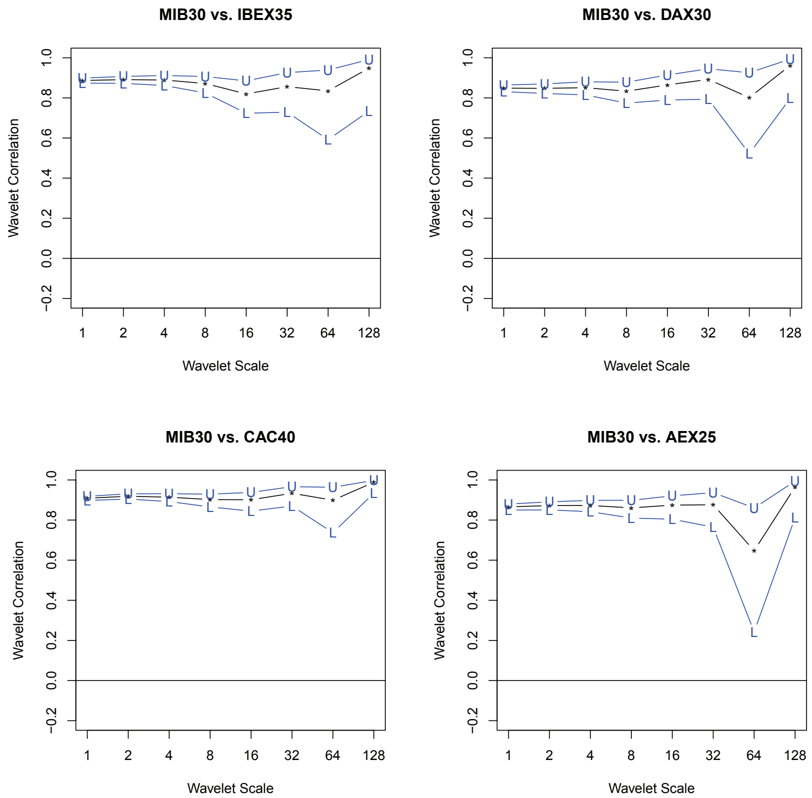

Figure 4: (Part 1) Wavelet correlation (classical plot) between five EU market returns, MIB30, IBEX35, CAC40, DAX30 and AEX25. The blue lines correspond to the upper and the lower bounds of the 95% confidence interval.

Digital Object Indentifier 10.1109/MCSE.2014.96 1070-9924/$26.00 2014 IEEE

This article has been accepted for publication in Computing in Science and Engineering but has not yet been fully edited.
Some content may change prior to final publication.

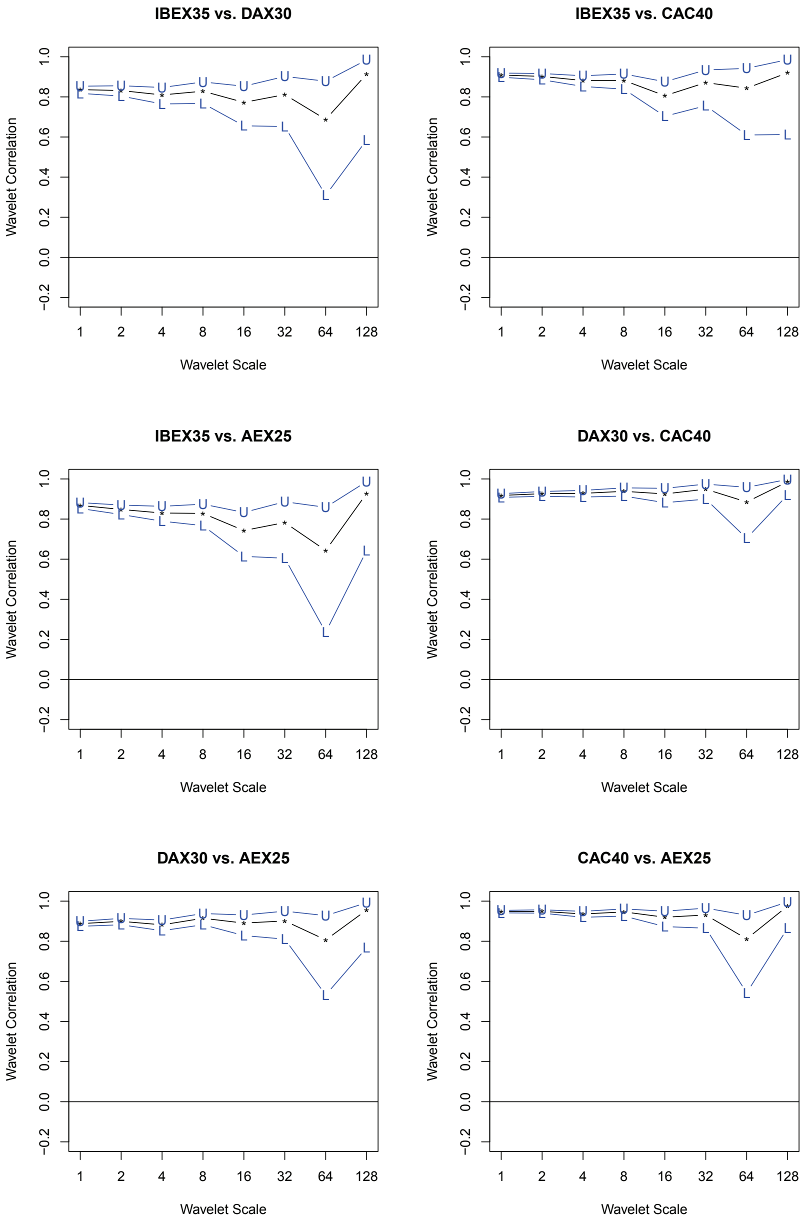

Figure 4: (Part 2) Wavelet correlation (classical plot) between five EU market returns, MIB30, IBEX35, CAC40, DAX30 and AEX25. The blue lines correspond to the upper and the lower bounds of the 95% confidence interval.

Digital Object Indentifier 10.1109/MCSE.2014.96 1070-9924/$26.00 2014 IEEE

This article has been accepted for publication in Computing in Science and Engineering but has not yet been fully edited.
Some content may change prior to final publication.

The second plot created by the `WCC` function is shown in Figure 5. It shows the result of applying the wavelet cross-correlation between the CAC40 and DAX30 daily log-returns. Figure 5 may be compared to the classical plot in Figure 6 constructed with the *spin.correlation* function. The main differences are the following. Again, the first advantage of our approach over the classical one is the reduction in the number of plots. As we can see, the classical approach needs J plots, in this example, eight, and ours only one. In fact, if we were analyzing N time series we would need to build $J\,C_{N,2}$ plots. The second advantage is that our approach offers the possibility of identifying graphically where in time, i.e., at what time lag, the strongest wavelet correlation coefficients are localized, these being indicated by long-dashed vertical lines. The third feature is that our plot allows us to distinguish the lags where the confidence interval spans zero (in Figure 5 the corresponding zones are shown in white). Overall, we believe that the plot presented in Figure 5, offering several advantages over the WCC classical plot, is a considerably more versatile graphical tool.

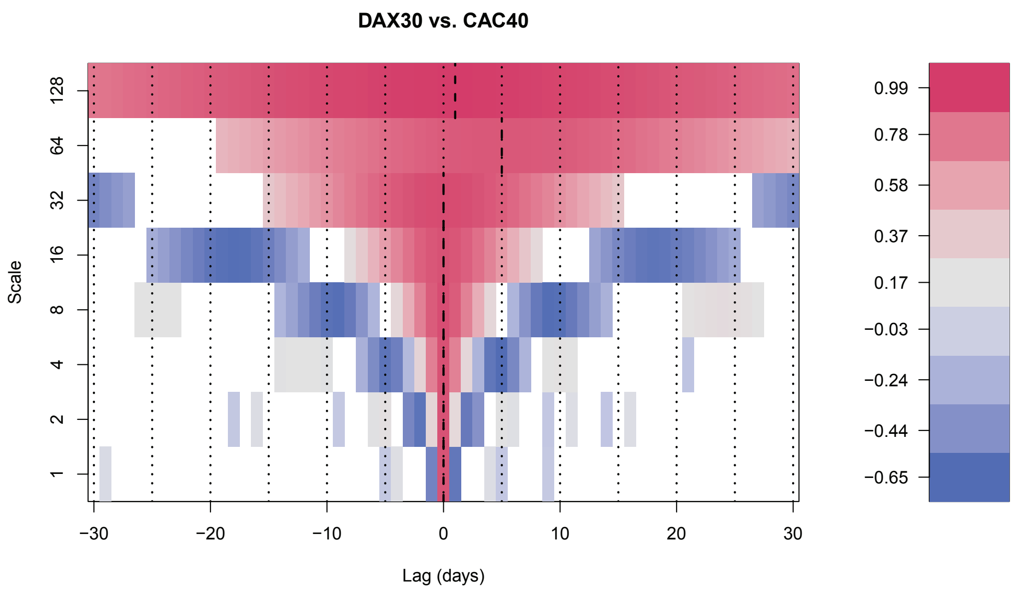

Figure 5: Wavelet cross-correlation (new plot) between the DAX30 and CAC40. The wavelet coefficients are within the 95% confidence interval for each wavelet cross-correlation. The white zones indicate that the zero is included in the 95% confidence interval. The long-dashed vertical lines indicates where in time the strongest wavelet correlation values are localized.

Digital Object Indentifier 10.1109/MCSE.2014.96 1070-9924/$26.00 2014 IEEE

This article has been accepted for publication in Computing in Science and Engineering but has not yet been fully edited.
Some content may change prior to final publication.

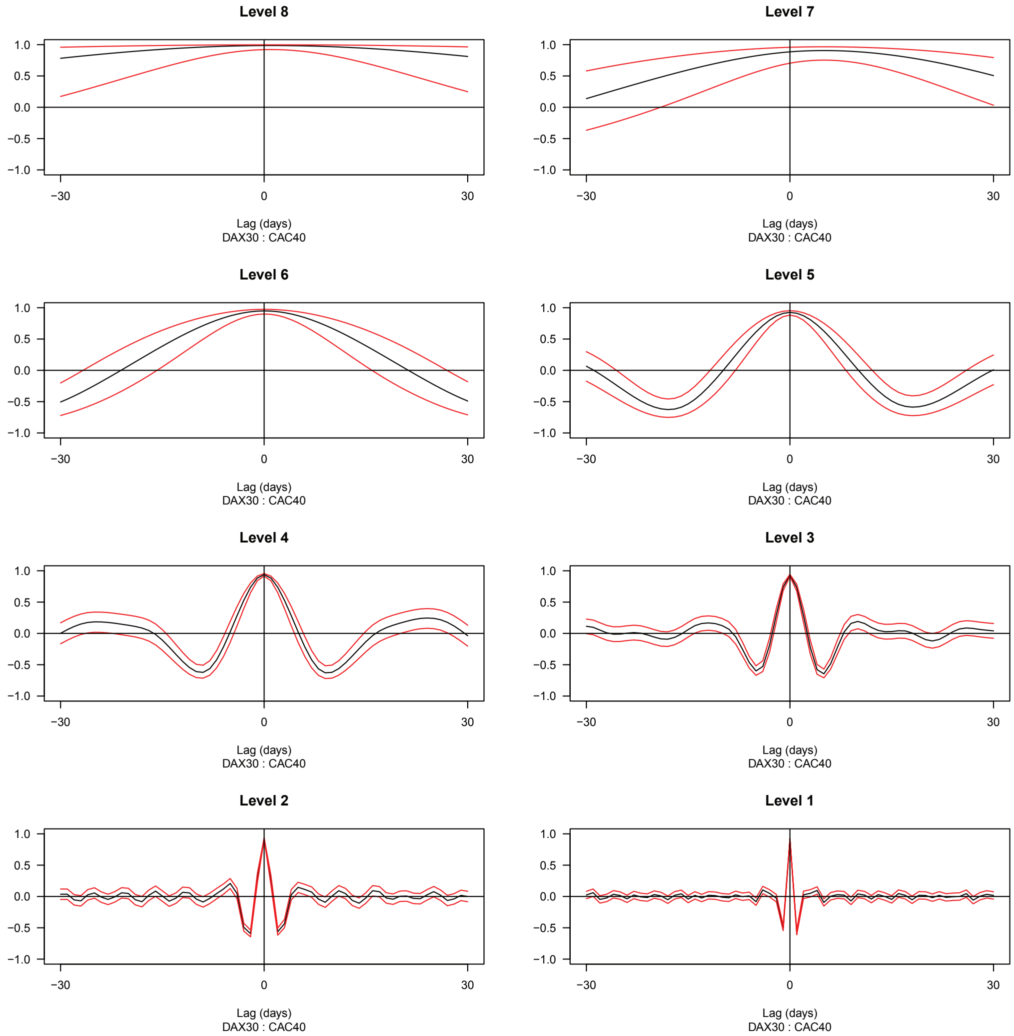

Figure 6: Wavelet cross-correlation (classical plot) between DAX30 and CAC40. The red lines correspond to the upper and the lower bounds of the 95% confidence interval.

Digital Object Indentifier 10.1109/MCSE.2014.96 1070-9924/$26.00 2014 IEEE

This article has been accepted for publication in Computing in Science and Engineering but has not yet been fully edited.
Some content may change prior to final publication.

4.2. The Multivariate case

Figure 7 presents the wavelet multiple cross-correlation created by applying the `WMCC` function to the daily log-returns of the indices of seven European stock markets (MIB30, IBEX35, CAC40, DAX30, AEX25, ATX20 and BEL20). As we can see, this approach offers various advantages over the the WMCC classical plot obtained using the `wave.multiple.cross.correlation` that is shown in Figure 8. Firstly, as in the bivariate case, our new plot combines the J plots (eight in this example) into one single plot, which facilitates the interpretation of the results. Secondly, it is easier to accurately identify at what time lag the strongest wavelet correlation values are localized (indicated by the long-dashed vertical lines). Moreover, the name of the variable that maximizes the multiple cross-correlation against a linear combination of the rest of the variables is stated explicitly on the right side of each wavelet scale [2]. Thirdly, it is now easier to identify whether the confidence interval spans zero (especially when the values are very close to zero) than when using plots of the type shown in Figure 8. Considering these useful features, our plot produced with the `WMCC` function has evident advantages.

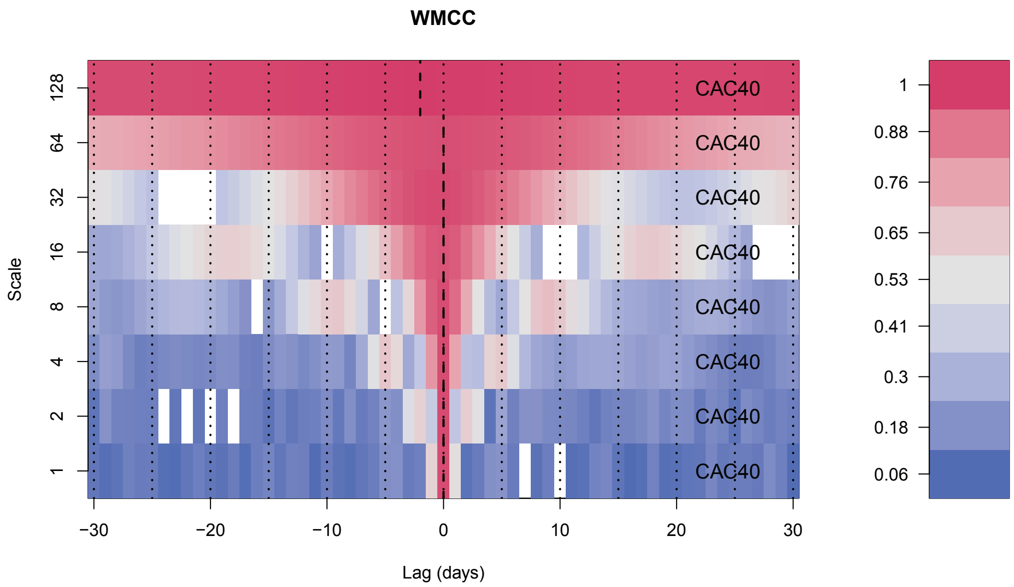

Figure 7: Wavelet multiple cross-correlation (new plot) between seven European market returns, MIB30, IBEX35, CAC40, DAX30, AEX25, ATX20 and BEL20. The wavelet coefficients are within of the 95% confidence interval for each wavelet correlation. Zones in which the 95% interval spans zero are indicated in white. The long-dashed vertical lines indicates where in time the strongest wavelet correlation values are localized.

Digital Object Indentifier 10.1109/MCSE.2014.96 1070-9924/$26.00 2014 IEEE

This article has been accepted for publication in Computing in Science and Engineering but has not yet been fully edited.
Some content may change prior to final publication.

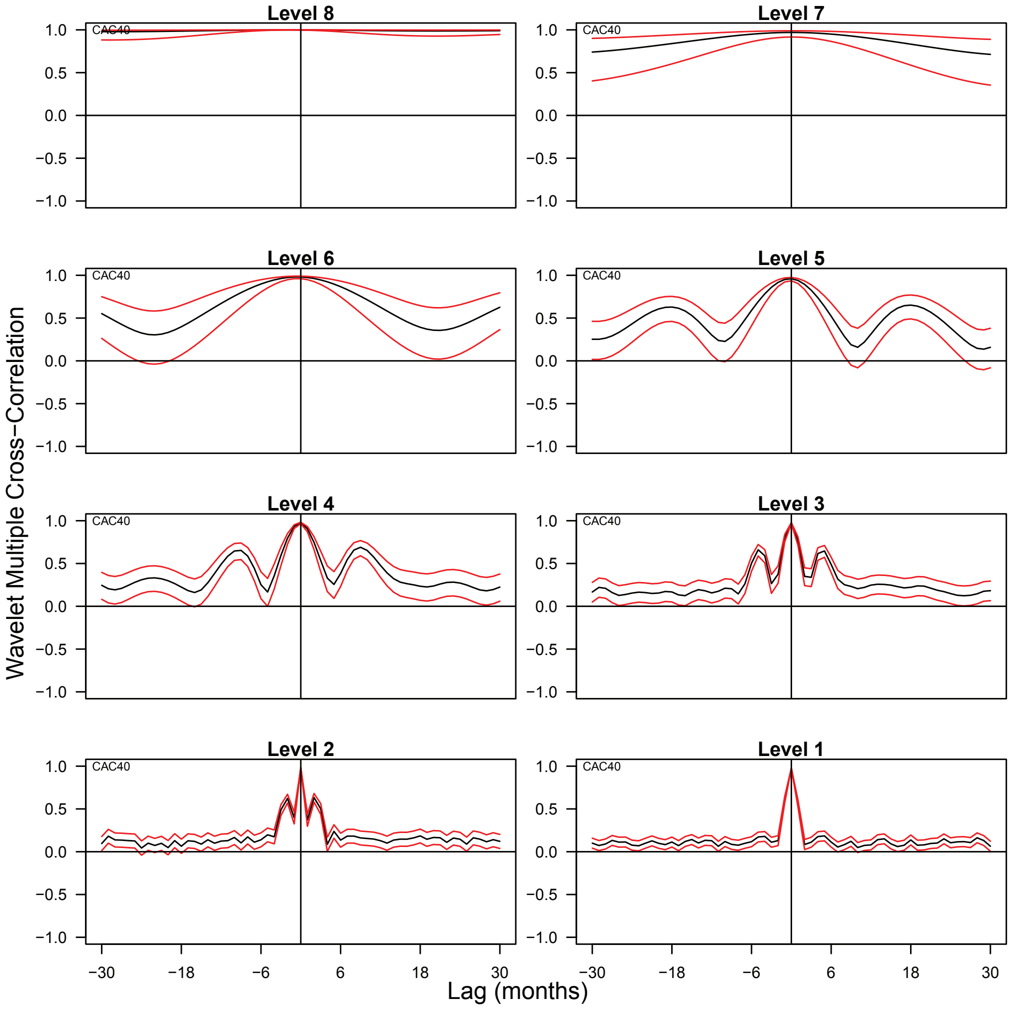

Figure 8: Wavelet multiple cross-correlation (classical plot) between seven European stock market returns, MIB30, IBEX35, CAC40, DAX30, AEX25, ATX20 and BEL20. The red lines correspond to the upper and the lower bounds of the 95% confidence interval.

5. Conclusions

The **W2CWM2C** package provides versatile graphical tools that can be employed to analyze multivariate data sets, in particular with a view to multiscale wavelet correlation analysis. A data set from stock market indices has been included in the package and we have used this to illustrate the power of the tools. The **W2CWM2C** package has been programmed in R using existing R package routines **waveslim** [11] and **wavemulcor** [2, 3]. In addition, we have used the R package **colorspace** [5] to build the heatmaps. This computational tool has been designed to be run from the command-line

Digital Object Indentifier 10.1109/MCSE.2014.96 1070-9924/$26.00 2014 IEEE

interface. The main contributions of the **W2CWM2C** package are: 1) the `WC` function helps to analyze pairwise wavelet correlations by generating a single plot instead of $C_{N,2} = N!/(N-2)!2!$ plots (where N is the number of time series under study), while 2) the `WCC` and `WMCC` functions summarize all the information from $J\,C_{N,2}$ and J classical plots, respectively, into a single plot and 3) they also provide a graphical way to identify at what time lag the strongest wavelet correlation coefficients are localized. In brief, the R package **W2CWM2C** addresses key shortcomings in the types of graphical tools previously available from the **waveslim** [11] and **wavemulcor** [3] packages to represent wavelet correlation and wavelet cross-correlation analysis of bivariate and multivariate time series.

Acknowledgement

Financial support from the *Econometrics Research Group* at the University of the Basque Country (UPV/EHU), Basque Government grants (IT-334-07 and IT-642-13), and the FP7 BASE project (Ref. No. 308337) is gratefully acknowledged. We are thankful to Jan de Leeuw and Achim Zeileis for suggesting the use of the R package **colorspace** and two anonymous `CiSE` reviewers for their constructive suggestions for improving this manuscript. *Ideas Need Communicating Language Services* S. L. (Julia) helped refining the English. We also thank **Yahoo! Finance** for making stock market indices available, RStudio for their recent release of "0-Cloud" CRAN log files and Felix Schönbrodt and Boris Hejblum for the R scripts to analyze the log data.

References

[1] I. Daubechies. *Ten lectures on wavelets*. Society for Industrial Mathematics, 1992.

[2] J. Fernández-Macho. Wavelet multiple correlation and cross-correlation: A multiscale analysis of Eurozone stock markets. *Physica A: Statistical Mechanics and its Applications*, 391(4):1097–1104, 2012a.

[3] J. Fernández-Macho. ***wavemulcor**: wavelet routine for multiple correlation*, 2012b. R package version 1.2. http://CRAN.R-project.org/package=wavemulcor.

[4] R. Gencay, F. Selcuk, and B. Whitcher. *An introduction to wavelets and other filtering methods in finance and economics*. Academic Press, San Diego., 2002.

Digital Object Indentifier 10.1109/MCSE.2014.96 1070-9924/$26.00 2014 IEEE

[5] R Ihaka, P Murrell, K Hornik, JC Fisher, and A Zeileis. ***colorspace**: Color Space Manipulation*, 2012. R package version 1.2-0. http://CRAN.R-project.org/package=colorspace.

[6] K. M. Lau and H. Weng. Climate signal detection using wavelet transform: How to make a time series sing. *Bulletin of the American Meteorological society*, 76(12):2391–2402, 1995.

[7] D.B. Percival and H.O. Mofjeld. Analysis of Subtidal Coastal Sea Level Fluctuations Using Wavelets. *Journal of the American Statistical Association*, 92(439):868–880, 1997.

[8] D.B. Percival and A.T. Walden. *Wavelet methods for time series analysis*. Cambridge Univ Press, Cambridge., 2000.

[9] J Polanco-Martínez. ***W2CWM2C**: set of functions to produce new graphical tools for wavelet correlation (bivariate and multivariate cases)*, 2012. R package version 1.0. http://CRAN.R-project.org/package=W2CWM2C.

[10] C. Torrence and G. P. Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*, 79(1):61–78, 1998.

[11] B. Whitcher. ***waveslim**: Basic wavelet routines for one-, two- and three-dimensional signal processing*, 2012. R package version 1.7.1. http://CRAN.R-project.org/package=waveslim.

[12] B. Whitcher, P. Guttorp, and D.B. Percival. Wavelet analysis of covariance with application to atmospheric time series. *Journal of Geophysical Research-Atmosheres*, 105(D11):941–962, 2000.

Short author bios

Josué M. Polanco M. is a postdoctoral researcher at the *Basque Centre for Climate Change*. He holds a *PhD* (2012) in Engineering Physics at the Dpt. of Physics Applied (supported by the *University of the Basque Country* UPV/EHU) and also worked as a research officer (2009-2012) at the Dpt. of Econometrics and Statistics (Applied Economics III) and the Public Economics Institute (UPV/EHU). His research interests include environmental and financial time series analysis, climatology and statistics applied, scientific software development, and more recently interdisciplinary research on climate change. In 2012, he obtained the ANECA (National Agency for the Quality Assessment and Accreditation, Spain) accreditation as an Assistant Professor. Contact him at josue.m.polanco@gmail.com or josue.polanco@bc3researcher.org.

Digital Object Indentifier 10.1109/MCSE.2014.96 1070-9924/$26.00 2014 IEEE

Complete contact information:

Josué M. Polanco M.
BC3 - Basque Centre for Climate Change
Alameda Urquijo 4, 4 1a | 48008 BILBAO, Bizkaia (Spain)
Phone: +34 94 401 4690 ext. 136
Email: `josue.m.polanco@gmail.com`, `josue.polanco@bc3researcher.org`.

Javier Fernández-Macho is Full Professor of Econometrics at the Department of Econometrics and Statistics (Applied Economics III) of the University of the Basque Country (UPV/EHU) where he lectures on Econometrics and Time Series Analysis. His main research topics include theoretical contributions on state-space and frequency domain time series econometrics and spatial stochastic processes. He has also co-authored several books on the theory and practice of Econometrics. Fernández-Macho has a *PhD* in Economics and an *MSc* in Statistics from the *London School of Economics* (Department of Statistics and Mathematical Science). Contact him at `javier.fernandezmacho@ehu.es`.

Complete contact information:

postal address:
F. Javier Fernández-Macho,
Dpt of Econometrics and Statistics (EA3),
University of the Basque Country,
Lehendakari Agirre 83,
48015 BILBAO (Spain)

Phone: +34946013655,
Fax: +34946013754,
Email: `javier.fernandezmacho@ehu.es`

Digital Object Indentifier 10.1109/MCSE.2014.96 1070-9924/$26.00 2014 IEEE