



OPEN

## Dynamic wavelet correlation analysis for multivariate climate time series

Josué M. Polanco-Martínez<sup>1✉</sup>, Javier Fernández-Macho<sup>2</sup> & Martín Medina-Elizalde<sup>3</sup>

The wavelet local multiple correlation (WLMC) is introduced for the first time in the study of climate dynamics inferred from multivariate climate time series. To exemplify the use of WLMC with real climate data, we analyse Last Millennium (LM) relationships among several large-scale reconstructed climate variables characterizing North Atlantic: i.e. sea surface temperatures (SST) from the tropical cyclone main developmental region (MDR), the El Niño-Southern Oscillation (ENSO), the North Atlantic Multidecadal Oscillation (AMO), and tropical cyclone counts (TC). We examine the former three large-scale variables because they are known to influence North Atlantic tropical cyclone activity and because their underlying drivers are still under investigation. WLMC results obtained for these multivariate climate time series suggest that: (1) MDRSST and AMO show the highest correlation with each other and with respect to the TC record over the last millennium, and: (2) MDRSST is the dominant climate variable that explains TC temporal variability. WLMC results confirm that this method is able to capture the most fundamental information contained in multivariate climate time series and is suitable to investigate correlation among climate time series in a multivariate context.

The climate system is highly dynamic and involves interactions among five main Earth system components or subsystems: the atmosphere, the hydrosphere, the cryosphere, the lithosphere and the biosphere<sup>1,2</sup>. Processes taking place within these components and during their interactions occur at different spatial and temporal scales, from molecular to planetary levels and from seconds to millions of years<sup>1,3</sup>. These Earth system components interact through mass, energy and momentum exchanges creating feedback loops and chains (the cascade effect) despite the climate system itself being a closed system (the climate system is considered as a closed system, i.e. this system does not exchange mass outside of the system, although there is still exchange of energy<sup>3,4</sup>). The climate system is also under the influence of external forcings, such as aerosol emissions from volcanic eruptions, changes in Earth's orbital parameters (astronomical forcing), fluctuations of solar radiation and increases in the atmospheric composition of greenhouse gases due to human activities<sup>1,5</sup>. Each subsystem affects the response of another, ultimately determining a climate state. Interactions among the different parts of the climate system generate disproportionate relations between inputs and outputs, and therefore originating a complex, non-linear and non-stationary dynamic system<sup>4,6</sup>.

The complexity, nonlinearity and nonstationarity of the climate system is reflected by the resulting essential climate variables—ECV (variables that play a critical role in characterizing the Earth's climate). Such climate variables are represented as time series, for example, of atmospheric or ocean surface temperature, ocean salinity, precipitation, etc.<sup>4,7,8</sup>. In addition to the nonlinearity and nonstationarity nature of climate dynamics, time series of climate variables used to characterize it also have other properties irrespective of their source (i.e. archives of climate information come mainly from direct instrumental measurements or indirect evidence, from climate model simulations, and paleoclimate reconstructions). For example, typically, climate time series show strong autocorrelation (memory or persistence), are generally short, noisy, contain uncertainty, may be unevenly spaced, may contain periodic, quasi-periodic or transient signals, and are a composition of numerous packages of information in time-scales (i.e. multiscale phenomena)<sup>3,8–12</sup>.

A useful approach to extract information from climate time series is the application of diverse kinds of statistical methods, particularly time series analysis and signal processing techniques<sup>3,8,13,14</sup>. Statistical data analysis are traditionally used in climate research in support of scientific affirmations: in order to estimate and assign confidence intervals around observations and predictions, to detect non-negligible noise level in climate variables and to identify fundamental relationships among them<sup>3,10</sup>. Regarding the latter, there are a great number

<sup>1</sup>Basque Centre for Climate Change (BC3), 48940 Leioa, Spain. <sup>2</sup>Dept. of Quantitative Methods, University of the Basque Country, 48015 Bilbao, Spain. <sup>3</sup>Dept. of Geosciences, University of Massachusetts, Amherst, MA, USA. ✉email: josue.polanco@bc3research.org

of techniques to study the relationship between two climate time series (bivariate analysis): from the classical Pearson's and Spearman's correlation or the cross-correlation function (CCF)<sup>3,10</sup>, binned<sup>8,15</sup> and synchrony<sup>8</sup> correlation to most sophisticated techniques such as kernel methods<sup>16,17</sup>, cross-recurrence plots<sup>18,19</sup>, and wavelet correlation<sup>20,21</sup> and wavelet coherence<sup>22–25</sup>. In contrast, large number of techniques are not available to study the relationship among multiple variables.

Among the correlation techniques, the wavelet correlation either via the discrete wavelet transform (DWT) or especially via the continuous wavelet transform (CWT), is highly used by climatologists and paleoclimatologist and by other environmental research communities<sup>11,20,22–28</sup>. The wavelet correlation via the wavelet transform (WT) can be seen as an “improved version” of the combination of Fourier transform and partial and sliding correlations in different periods (known as windowed or short-time Fourier transform—WFT<sup>29</sup>) although methodologically are different. The WFT represents an inaccurate and inefficient method of time–frequency localization, since it imposes a scale or “response interval” into the analysis, whereas the WT is a method of time–frequency localization that is scale independent<sup>22,30</sup>. The wavelet (uni and bivariate) analysis is an adequate and versatile mathematical tool to tackle several characteristics of climate time series, such as nonstationarity, in search of (quasi) periodical or oscillatory signals, and to examine multiscale phenomena. Bivariate wavelet analysis also permits to study the relationship between two climate time series and is particularly appropriate for tracking dual change in forcing by exogenous variables<sup>20,22,24,25,28,31</sup>. However, the use of wavelet correlation has been limited since its inception to the bivariate case (some exceptions are<sup>32</sup> and<sup>33</sup>), and the study of climate dynamics usually involves more than two variables interacting with each other simultaneously<sup>1,3</sup>. For this reason, a multivariate version of the wavelet correlation has been required.

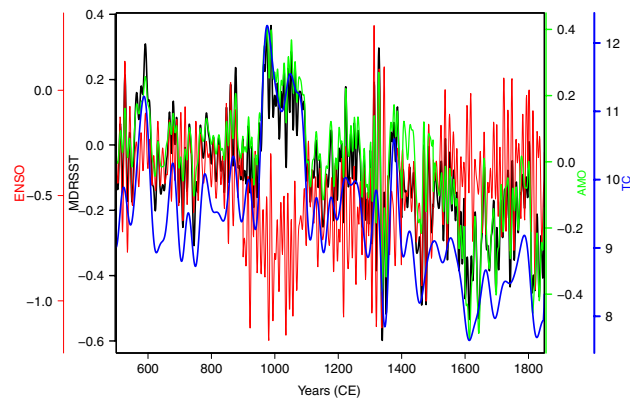
In line with this interest, the aim of this paper is to introduce for the first time to the climate community, and to those of related fields, the application of wavelet local multiple correlation (WLMC)<sup>34</sup> to analyze “dynamically” (i.e. through time) over different time-scales, multivariate climate time series. Particularly, to examine relationships among climate variables known to be physically related to each other. This method improves our ability to understand the underlying mechanisms driving climate change on different timescales. Climate variables, and the nature of the relationship among them, change over time and thus cannot be accurately understood using conventional statistical methods that do not take into account their time evolution. The best approach is therefore to use correlation methods and computational tools that address non-stationary relationships among multiple climate variables, such as the wavelet local multiple correlation (WLMC)<sup>34</sup>.

The WLMC measures a non-stationary time-evolving correlation structure at different scales within a multivariate set of data, and consists of one single set of multiscale correlations along time, each of them calculated as the square root of the regression coefficient of determination in that linear combination of locally weighted wavelet coefficients for which such coefficient of determination is a maximum<sup>34,35</sup>. The WLMC is a powerful statistical and computational tool that was originally developed to estimate correlation among multivariate, non-stationary, financial time series<sup>34,35</sup>. However, as we demonstrate in this study, the WLMC has a strong potential to be used with multivariate, stationary and non-stationary, climate and environmental time series as well. It is important to introduce this tool to the climate community since the economic and financial communities have poor connection with the climate community, despite some international networks and initiatives to bring them together (e.g. Climate Econometrics, <https://www.climateeconometrics.org/>). In this study we also improve graphical outputs of the WLMC and propose a didactic and useful way to visualize the “dominant” variable(s) that maximizes multiple correlation through time for a set of climate time series. This is the first time that this kind of graphical representation for the WLMC is presented (please look at Fig. 3 right).

This study exemplifies the use of WLMC by analyzing multivariate paleoclimate time series. Specifically, we examine relationships among various large-scale climate proxy reconstructions spanning the Last Millennium (LM) that are known to influence tropical Atlantic hurricane frequency and intensity. These variables come from<sup>36,37</sup> and represent: (1) sea surface temperatures (SST) anomalies in the main developmental region (MDR) for tropical cyclones (MDRSST); (2) the El Niño–Southern Oscillation (ENSO) sea surface temperature anomalies (ENSO SSTs, or ENSO hereafter); (3) the Atlantic Multidecadal Oscillation (AMO) sea surface anomalies (AMOSSTs, or AMO hereafter); and (4) tropical cyclone counts (TCC, or TC hereafter) as a proxy of cyclone activity. We examine ENSO, AMO and MDR SSTs because they are known to influence North Atlantic tropical cyclone activity and their underlying processes are still under investigation. These climate variables have been invoked to explain climate dynamics inferred from paleoclimate records on previous studies; e.g. Ref.<sup>26</sup>.

## Results

**North Atlantic tropical cyclone activity over the past 1500 years.** Figure 1 shows the climate time series examined in this study. Because MDRSST, AMO and TC have similar variability patterns through time, it is expected for them to have a relative high degree of correlation at “global” level, which can be corroborated by estimating the degree of correlation with each other. The correlation matrix obtained via Spearman's rank correlation (Table 1) shows that the highest correlation ( $r = 0.910$ ) corresponds to the pair MDRSST–AMO, followed by the pairs AMO–TC ( $r = 0.843$ ) and MDRSST–TC ( $r = 0.774$ ). In contrast, ENSO shows an inverse relationship with AMO and TC (except relative to MDRSST which is direct), reflected by the negative correlation coefficients. This is confirmed by the pairwise comparison between ENSO and TC ( $r = -0.296$ ) and between ENSO and AMO ( $r = -0.132$ ). The negative correlation between ENSO and TC is expected because tropical cyclones can be suppressed when El Niño increases wind shear in the tropical Atlantic and warm upper tropospheric temperature anomalies<sup>38–40</sup>. On the other hand, the negative or inverse correlation between ENSO and AMO can be explained straightforwardly by the fact that during the negative AMO phase, both El Niño and La Niña events tend to be stronger than during a positive AMO phase<sup>41</sup>. Moreover, this inverse correlation between ENSO and AMO can be explained by the so-called atmospheric “bridge–thermocline feedback”<sup>41,42</sup>.



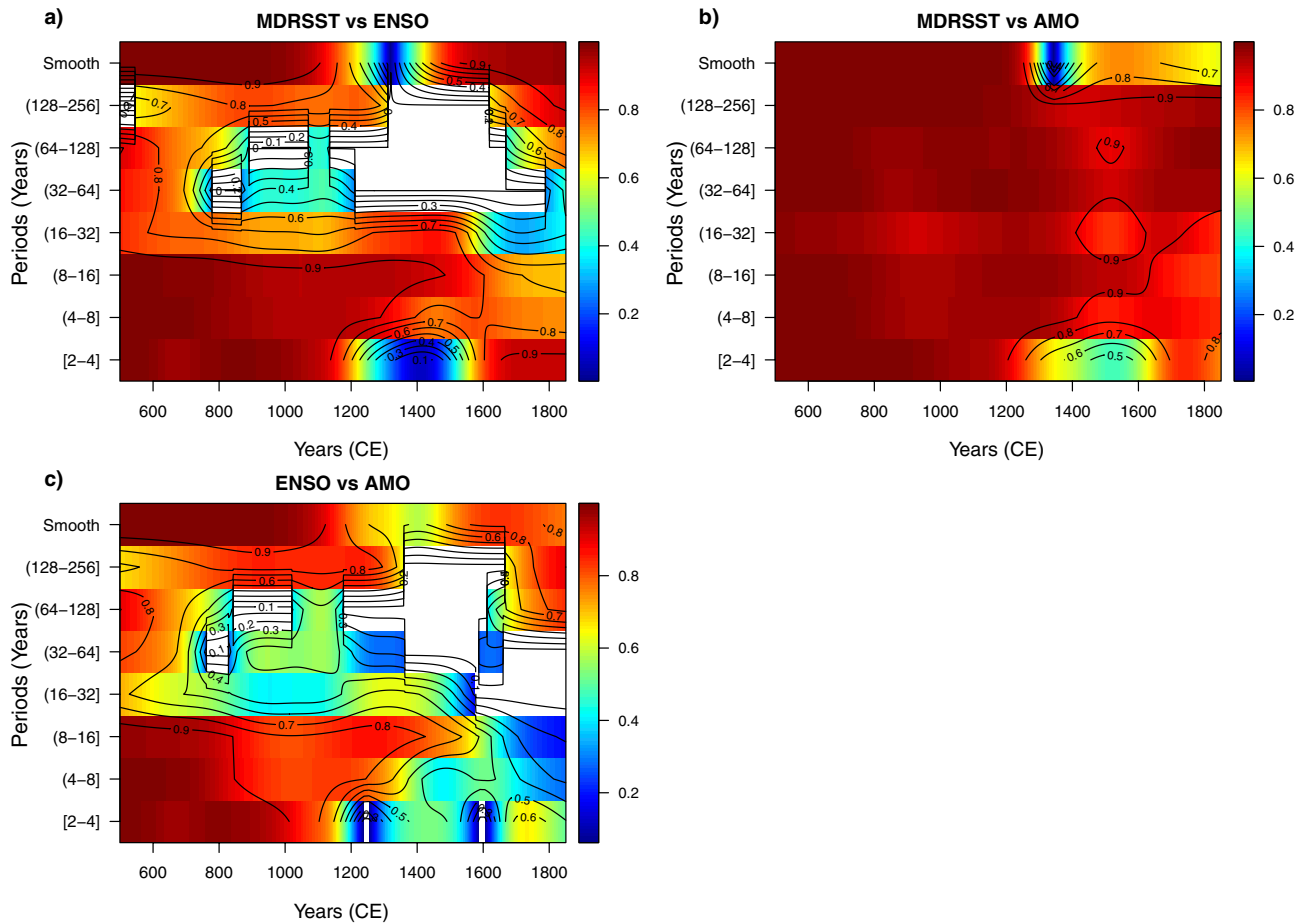
**Figure 1.** Sea surface temperature (SST) for the main development region (MDR) for Atlantic tropical cyclones<sup>36</sup> (black); El Niño–Southern Oscillation (ENSO) SST (based on El Niño 3 region)<sup>37</sup> (red); North Atlantic Atlantic Multidecadal Oscillation (AMO) SST averaged over the North Atlantic ocean<sup>37</sup> (green); Long-term Atlantic tropical cyclone counts (TC)<sup>36</sup> (blue). Climate variables are in anomalies ( $^{\circ}\text{C}$ ), cover the time interval CE 500–1850, and the number of elements is 1350.

|        | MDRSST       | ENSO         | AMO          | TC           |
|--------|--------------|--------------|--------------|--------------|
| MDRSST | 1.000        | <b>0.155</b> | <b>0.910</b> | <b>0.774</b> |
| ENSO   | <b>0.155</b> | 1.000        | –0.132       | –0.296       |
| AMO    | <b>0.910</b> | –0.132       | 1.000        | <b>0.843</b> |
| TC     | <b>0.774</b> | –0.296       | <b>0.843</b> | 1.000        |

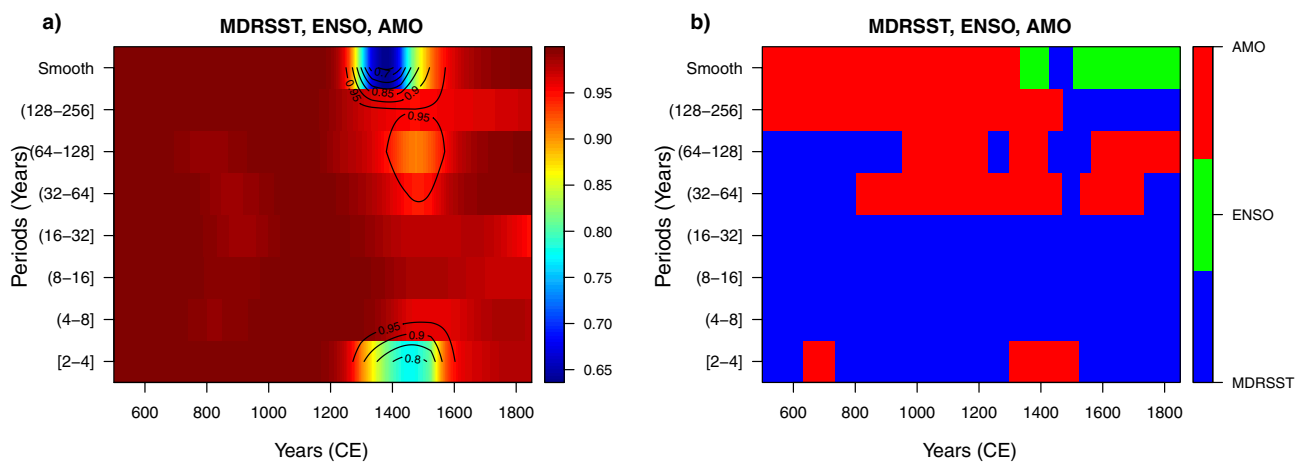
**Table 1.** Correlation matrix estimated through Spearman’s rank correlation. Statistically significant (95%) correlation coefficients are in bold.

**Wavelet local correlation for the bi-variate case for MDRSST, ENSO, and AMO.** The wavelet local multiple correlation for the bivariate case (Fig. 2) is used as proof of concept for the multivariate case ( $n > 2$ ), in order to determine the extent to which the climate variables by pairs are correlated with each other in the time and frequency (period) domains and thus avoid using the WLMC as a “black box”. The first evident result for this bivariate case is that MDRSST–AMO shows the highest WLMC among the three pairwise comparisons, in agreement with the result above using Spearman. However, one conspicuous feature of the dynamic correlation methods, such as the WLMC, is that they provide the evolution of correlation in the time and period domains. For example, the WLMC shows that the correlation of MDRSST–AMO is high (with coefficient values  $> 0.80$ ) for practically all the periods considered and during almost the full time interval of the records (CE 500–1850), except for a small interval between CE 1400 and 1600, for the shortest scales (2–4 years). This result must be regarded with caution, however, because climate time series are decadal smoothed<sup>36,37</sup>. The second important result is that although the correlations of MDRSST–AMO is generally high, the strength of correlations tends to “oscillate” in time and nearly in all periods, indicating that the degree of correlation is not constant through time. This result has important implications for diagnosing the causal relationships among these variables and for understanding their underlying large-scale dynamics<sup>36,43</sup>. Lastly, the WLMC of ENSO–MDRSST and ENSO–AMO shows moderate degree of correlation that is well localized in the frequency/period domain. As expected, the highest correlation among the variables occurs at 4–8 and 8–16 year periods, where ENSO has its main spectral signature<sup>36,40</sup>. Furthermore, there is a “slowdown” in correlation at the end of the heat maps for both pairs (more pronounced for ENSO–AMO), and their corresponding time intervals are not exactly the same as for MDRSST–AMO.

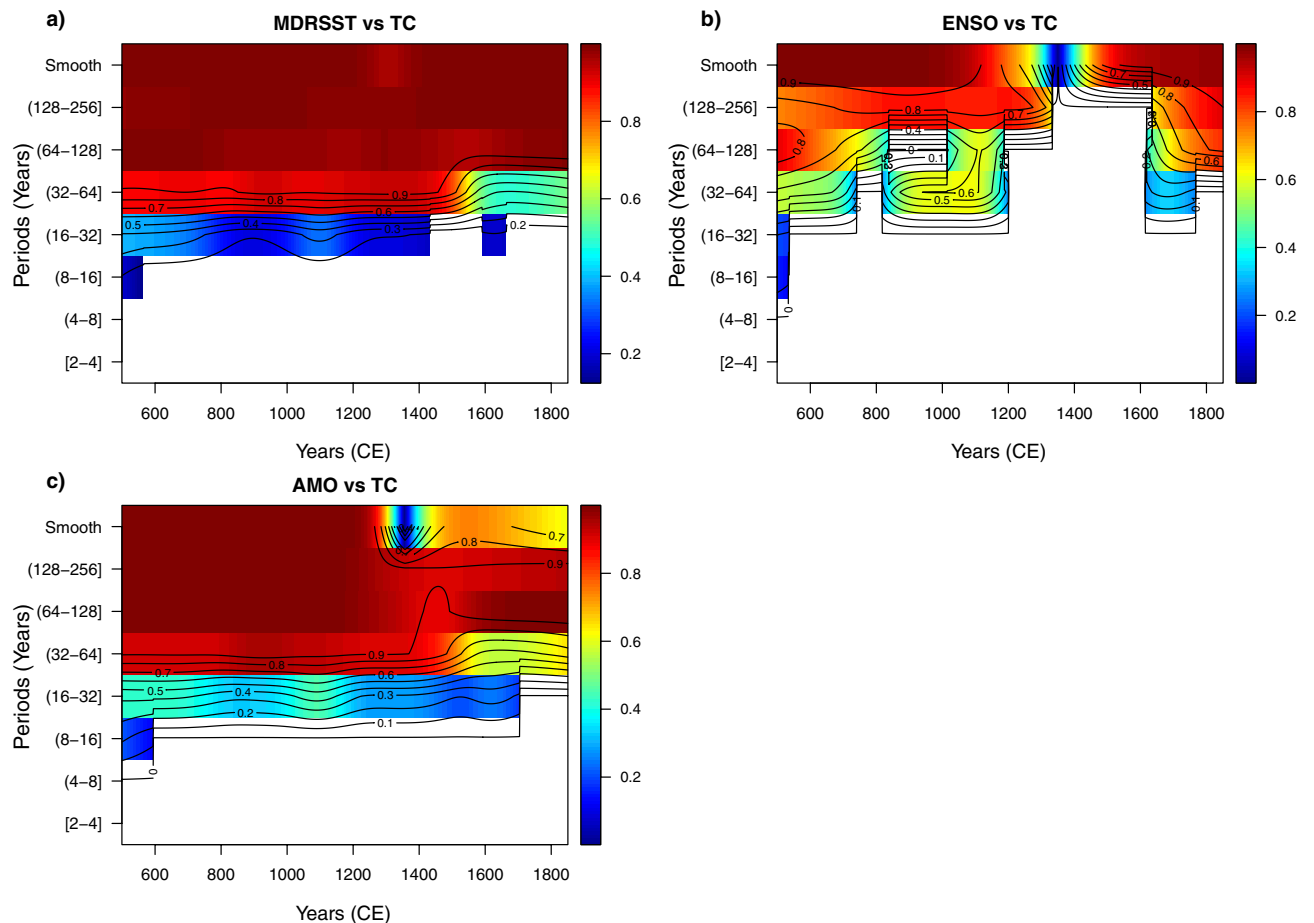
**WLMC for the three-variate case for MDRSST, ENSO, and AMO.** Figure 3 shows the WLMC for the three-variate case. We do not define *a priori* specific climate variable that would maximize the multiple correlation for each wavelet scale (parameter  $\gamma_{\max r} = \text{NULL}$ ) but instead let the WLMC select one. The reason for this choice is that although these climate variables are expected to be correlated with each other, their causal relationship in time and frequency domains remains a matter of research. The three-variate WLMC shows a significant level of correlation with variable coefficients (from  $\sim 0.8$  to  $\sim 0.99$ ), from short to long periods of variability, and for practically the full length of the records. The three-variate WLMC result is quite similar to the WLMC for the pair MDRSST–AMO from the bivariate case (Fig. 2). This means that the WLMC method developed by<sup>34</sup> is an “inclusive” multivariate correlation tool unlike the multiple wavelet coherence, that is “exclusive”<sup>32,44,45</sup>. For this reason, it is important to take into account, that the WLMC retains practically all the statistically significant correlations between dominant variables (e.g. MDRSST and AMO from this case study), but also considers the



**Figure 2.** Wavelet local multiple correlation (bivariate case) for climate variables MDRSST, ENSO, and AMO. The time-period points in blank indicate that these points are not statistically significant (outside of the 95% confidence interval). WLMC parameters:  $M = 1350/8$  (168) years, window = Gaussian, and wavelet filter (wf) = “la8”.



**Figure 3.** Wavelet local multiple correlation (tri-variate case) for climate variables MDRSST, ENSO, and AMO (left). Heat map that indicates the (“dominant”) variable(s) that maximizes the multiple correlation through time and scale (right) (blue, green, and red indicate MDRSST, ENSO, and AMO, respectively). The time-period points in blank indicate that these points are not statistically significant (outside of the 95% confidence interval). WLMC parameters:  $M = 1350/8$  (168) years, window = Gaussian, wavelet filter (wf) = “la8”, and  $y_{maxr} = NULL$ .

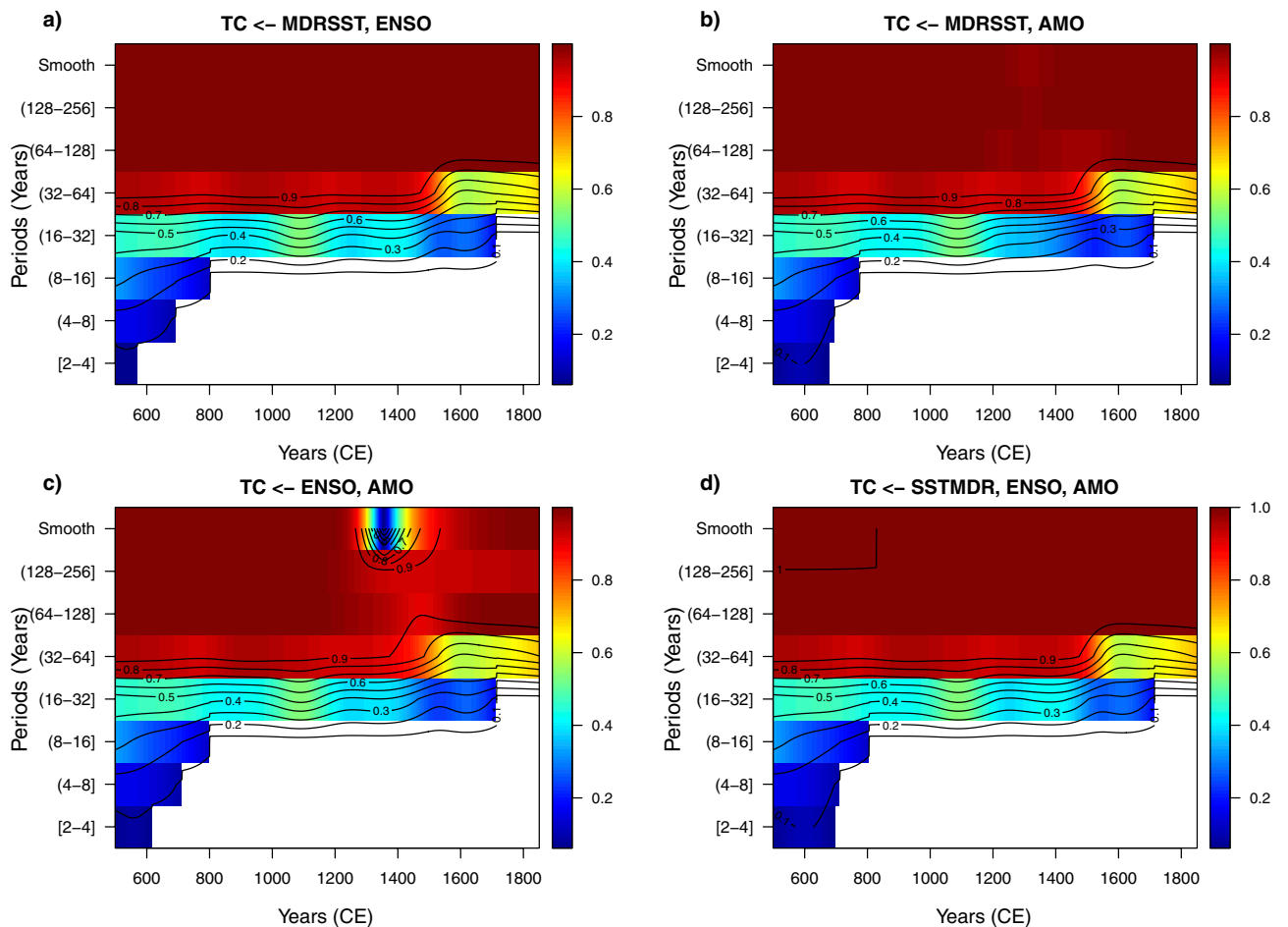


**Figure 4.** Wavelet local multiple correlation (bivariate case) for climate variables MDRSST, ENSO, AMO and TC. The time-period points in blank indicate that these points are not statistically significant (outside of the 95% confidence interval). WLMC parameters:  $M = 1350/8$  (168) years, window = Gaussian, and wavelet filter (wf) = “la8”.

correlation among other variables (e.g. MDRSST–ENSO and ENSO–AMO). Importantly, the WLMC provides the “dominant” variable (the one that maximizes the multiple correlation and can be used to explain the other variables across each period and time interval) (Fig. 3 right). For example, for the triad MDRSST–ENSO–AMO, the MDRSST represents the dominant climate variable (followed to a lesser extent by AMO). This result is explained because SST fields are used to build MDRSST and also to estimate the climate indices of AMO and ENSO (see Supplementary Material in<sup>37</sup>).

**WLMC for the bi-variate case for MDRSST, ENSO, and AMO vs. TC.** The WLMC for the bivariate case (Fig. 4), when the TC record is included shows three interesting results. The WLMC only shows statistically significant correlation coefficients with relatively high coefficients (approximately greater than 0.6) for medium (32–64 years) and long (128–256 years) periods. This result indicates that the WLMC tool works properly since the TC record is smoothed multidecadally (> 40 years) and therefore is not expected to reveal significant variability at periods lower than 40 years. The second important result is that apart from the high and statistically significant correlation for the pairs SSTMDR–TC and AMO–TC (please note that this result is in agreement with the correlation matrix presented previously in Table 1), the coefficient values tend to vary as a function of time, which is more evident at the periods 32–64 and 64–128 years for the pair SSMDR–TC and at the periods 32–64 and 128–256 years for the pair AMO–TC. The tool also shows an almost complete lack of statistically significant correlation between ENSO and TC, except for the long-term periods between 128 and 256 years. This lack of correlation reflects the smoothing of the TC record (> 40 years) relative to the main periods of ENSO variability (3–7 years).

**WLMC for the three- and four-variate cases for MDRSST, ENSO, and AMO vs. TC.** Figure 5 shows the three- and tetra-variate WLMC cases between the examined climate variables (MDRSST, ENSO and AMO) relative to the TC record. In these cases, the TC record is chosen to be the dependent variable ( $y_{\max r} = TC$ ). The three- and tetra-variate cases confirm some of the previous results obtained for the bivariate case; e.g. statistically significant correlation from medium (16–32) to long (128–256) periods. An outstanding



**Figure 5.** Wavelet local multiple correlation (tri- and tetra-variate cases) for climate variables MDRSST, ENSO, AMO and TC. The time-period points in blank indicate that these points are not statistically significant (outside of the 95% confidence interval). WLMC parameters:  $M = 1350/8$  (168) years, window = Gaussian, wavelet filter (wf) = “la8”, and  $y_{max} = TC$ .

feature observed in these WLMC heat maps is that, despite the low or lack of correlation between ENSO and TC over several periods, the correlation between MDRSST and AMO relative to TC is preserved. This corroborates that the WLMC is an “inclusive” multivariate correlation tool, as previously pointed out. This means that, at least with these climate variables, the tetra-variate case does not provide “extra” information to be considered.

## Discussion

To summarize, this study introduces for the first time to the climate community and related fields the use of wavelet local multiple correlation (WLMC)<sup>34</sup> to perform dynamic correlation analyses for multivariate climate time series. This method is suitable to compare time series that are linear, non-linear, non-stationary, containing (quasi)periodic events or transient signals, and that are a composition of numerous packages of information in time-scales ranging from days to millennia (i.e. multiscale phenomena).

This study exemplifies the use of WLMC to analyze a multivariate set of climate time series. Specifically, it examines relationships among various climate proxy reconstructions spanning the last millennium that are known to influence tropical Atlantic hurricane frequency and intensity. These variables come from<sup>36,37</sup> and represent: (1) sea surface temperatures anomalies in the main developmental region for tropical cyclones (MDRSST); (2) the El Niño-Southern Oscillation sea surface temperature anomalies (ENSO); (3) the Atlantic multidecadal Oscillation (AMO) sea surface anomalies; and (4) the tropical cyclone counts (TC).

The WLMC results obtained for the multivariate paleoclimate time series confirm that MDRSST and AMO are the highest correlated variables among the climate data examined and with respect to TC. The MDRSST is found to be the dominant climate variable that can be used to explain the other variables examined. We would like to highlight that, the WLMC method developed by<sup>34</sup>, is an “inclusive” multivariate correlation tool. For this reason, it is important to take into account that the WLMC retains the statistically significant correlation between dominant variables (e.g. MDRSST and AMO), but also considers the correlation (statistically significant) among other variables (e.g. MDRSST and ENSO or ENSO and AMO). The WLMC is shown to be a suitable tool to investigate correlation among climate time series in a multivariate context.

## Methods

This section presents and discusses the methodological approach proposed and used in this study; the wavelet local multiple correlation (WLMC)<sup>34</sup>. We will follow<sup>34,35,46,47</sup> to summarize the method. Readers interested in obtaining more detailed information can consult these references. The WLMC method is implemented by a freely available R computer package called *wavemulcor* with its corresponding documentation to perform the WLMC analysis<sup>35,47</sup>.

**Local multiple regression.** The WLMC is based on the notion of multiple regression and in this paper we use the concept of wavelet local multiple regression introduced by<sup>34,46</sup>.

Let  $X$  be a multivariate time series of dimension  $n$  observed at times  $t = 1, \dots, T$ . According to<sup>34</sup>, for some  $x_i \in X$  a local regression at a fixed  $s \in [1, \dots, T]$  can be used to minimize the weighted sum of squared errors

$$S_s = \sum_t \theta(t-s)[f_s(X_{-i,t}) - x_{it}]^2 \quad (1)$$

where  $\theta(x)$  is a given moving average weight function that depends on the time lag between observations  $X_t$  and  $X_s$  and  $f_s(X_{-i})$  is a local function of  $\{X_{x_i}\}$  around  $s$ . Letting  $s$  move along time, the corresponding local coefficients of determination are given by

$$R_s^2 = 1 - \frac{RwSS_s}{TwSS_s}, \quad s = 1, \dots, T, \quad (2)$$

where  $RwSS_s$  and  $TwSS_s$  are the residual and total weighted sum of squares respectively.

**Definition and estimation of the wavelet local multiple correlation.** Let  $W_{jt} = (w_{1jt}, \dots, w_{njt})$  be the wavelet coefficients for scale  $\lambda_j$  (where  $j = 1, \dots, J$ , and  $J$  indicates the maximum level of the wavelet transform decomposition) obtained by applying the MODWT to each time series  $x_i \in X$ , where  $i = 1, \dots, n$ . Following to<sup>34</sup>, at each wavelet scale  $\lambda_j$  the wavelet local multiple correlation coefficients  $\varphi_{X,s}(\lambda_j)$  can be estimated as the square roots of the regression coefficients of determination for that linear combination of variables  $w_{ij}$ ,  $i = 1, \dots, n$ , where such coefficients of determination are maxima. That is, from Eq. (2)

$$\tilde{\varphi}_{X,s}(\lambda_j) = \sqrt{R_{js}^2}, \quad j = 1, \dots, J, \quad s = 1, \dots, T. \quad (3)$$

On the other hand, since the  $R^2$  coefficient in the regression of a  $z_i$  on the rest of variables in the system is equivalent to the square correlation between the observed and the fitted values  $\hat{z}_i$  obtained from such regression, according to<sup>34</sup>, it is possible to express the consistent sample estimator of the WLMC as

$$\tilde{\varphi}_{X,s}(\lambda_j) = \text{Corr}(\theta(t-s)^{1/2}w_{ij}, \theta(t-s)^{1/2}\hat{w}_{ij}) \quad (4)$$

$$s = 1, \dots, T,$$

where  $w_{ij}$  is chosen so that its local regression on the set of regressors  $\{w_{kj}, k \neq i\}$  maximizes the corresponding coefficient of determination and  $\hat{w}_{ij}$  denotes the corresponding vector of fitted values.

**Weight (window) functions.** The *wavemulcor* software<sup>35,47</sup> used to estimate the WLMC includes six of the most commonly used weight functions (or windows) for averaging and smoothing: the uniform window, Bartlett's triangular window, Cleveland's tricube window, Wendland's truncated power window, Epanechnikov's parabolic window, and the Gaussian window. All six windows  $\theta(x)$  satisfy  $\int_{-\infty}^{\infty} \theta(x)dx = 1$  and have compact support in  $|x| \leq M$ , where  $M$  is the length of the weight function  $\theta(x)$ , except the Gaussian window that takes values for  $x \in (-\infty, \infty)$ . However<sup>34</sup>, suggested that the uniform, Cleveland's tricube and Epanechnikov's parabolic windows are not recommended due to the presence of negative values in their corresponding spectral windows and, he also suggested to use the Bartlett's triangular, Wendland's truncated power, and Gaussian windows because these windows are most adequate for signal extraction and smoothing and their spectral windows are non-negative. By default, *wavemulcor*<sup>35,47</sup> uses the Gaussian window for the following reasons: it is closest to the uniform weights (windows) in the time domain within a certain bandwidth, its Fourier transform is also Gaussian, it has near compact support in the frequency domain, and its spectral window is always positive<sup>34</sup>.

**Estimation of statistical significance.** One of the main advantages of wavelet correlation obtained through the discrete wavelet transform (DWT), or its improved versions such as the MODWT, as compared to the continuous wavelet transform (CWT) is that for the (MO)DWT, it is possible to construct an analytical confidence interval (CI), while for the CWT it is practically impossible to do so.

In particular, for the WLMC<sup>34,46</sup>, obtained the CI by means of the Fisher's transform as follows.

Let  $\tilde{\varphi}_{X,s}(\lambda_j)$  be the sample wavelet local multiple correlation (WLMC) calculated from Eq. (4). Then, from<sup>34</sup>, Theorem 1],

$$\tilde{Z}_{j,s} \stackrel{a}{\sim} \mathcal{F}\mathcal{N}(Z_{j,s}, (T/2^j - 3)^{-1}) \quad (5)$$

where  $\tilde{Z}_{j,s} = \text{arctanh}(\tilde{\varphi}_{X,s}(\lambda_j))$ ,  $Z_{j,s} = \text{arctanh}(\varphi_{X,s}(\lambda_j))$  and  $\mathcal{F}\mathcal{N}$  stands for the folded normal distribution. Thus, since  $\varphi_{X,s}(\lambda_j)$  is the correlation between weighted observations from two Gaussian variates of which  $T/2^j$

are serially uncorrelated asymptotically, applying the Fisher's transformation to  $\epsilon_j$  such that  $\text{abs}(\epsilon_j) = \tilde{Z}_{j,s}$  in Eq. (5)<sup>34</sup>, obtained the CI for the WLMC as

$$CI_{1-\alpha}(\varphi_{X,s}(\lambda_j)) = \tanh \left[ \tilde{Z}_{j,s} \pm \phi_{1-\alpha/2}^{-1} / \sqrt{T/2^j - 3} \right] \quad (6)$$

where  $\phi_p^{-1}$  is the 100p% point of the standard normal distribution.

**Computational and practical aspects.** In practice, the WLMC<sup>34</sup> method consist of the following steps (the R code used to generate all the figures in this paper are in the Supplementary Information (Material)):

1. Estimation of the MODWT via the R package *waveslim*<sup>48</sup> for each of the time series  $x_i$  in the multivariate set ( $i = 1, \dots, n$ , where  $n$  is the number of time series) and all scales  $\lambda_j$  ( $j = 1, \dots, J$ , where  $J$  is the maximum level of the MODWT decomposition with  $J \leq \log_2(T)$  and  $T$  is the length of the time series. Note that although the theoretical maximum level of the MODWT decomposition is given by  $\log_2(T)$ , in practice  $J$  should be much smaller in order to avoid boundary effects since the number of feasible wavelet coefficients becomes critically small for higher levels). As a wavelet filter or function, we chose the Daubechies *LA(8)* (or "la8") that is atleast asymmetric wavelet filter of length  $L = 8$ <sup>49</sup>. We use *LA(8)* since<sup>20</sup> proposed this filter to analyze climate time series and<sup>34,49,50</sup> showed that the use of a relatively long wavelet filter (e.g. *LA(8)* or *LA(4)*) is adequate to analyse non-stationary time series but also correlation structures that are not stationary. We used both wavelet filters (*LA(8)* and *LA(4)*) in all the examples presented in this paper and the WLMC heat maps are quite similar. In addition to these wavelet filters we used other eight filters with different lengths: "Haar" ( $L=2$ ), "d4" ( $L=4$ ), "d6" ( $L=6$ ), "fk8" ( $L=8$ ), "bl14" ( $L=14$ ), "mb16" ( $L=16$ ), "la20" ( $L=20$ ), and "fk22" ( $L=22$ ) and the WLMC heat maps (results not shown, but can be obtained through the R code included in the Supplementary Information (Material) or upon request to the corresponding author) are considerably similar to the corresponding one when the "la8" is used, except for "bl14", "mb16", and "fk22" and for the longest wavelet scales (periods), which can be explained mainly due to the "excessive" length of these wavelet filters. As a good practice, it is highly recommendable to try several wavelet filters (the MODWT from the R package *waveslim*<sup>48</sup> includes 21 wavelet filters that can be used in the estimation of the WLMC) with different lengths ( $L$ ), from short, medium to long, to study the sensitivity of the wavelet functions in order to corroborate the stability of the WLMC.
2. Application of the rolling time window or weight function  $\theta(x)$  to the MODWT components. We used the default Gaussian window due to the aforementioned features. However, the other two recommendable options, Bartlett's triangular or Wendland's truncated power, provide similar results. Following to<sup>35,47</sup>, the recommended length of the weight function or rolling window is given by  $T/8$ , where  $T$  is the number of elements of time series. For the example presented in this paper we also choose a window length of  $T/8$  (or 168 years since  $T = 1350$ ) years as a compromise. A shorter window length would not have had enough data points to study longer climate time-scales phenomena and also could introduce a high degree of variability in the time domain, and a longer window length would have isolated climate phenomena that take place at short time scales. However, it is highly recommendable to try different window lengths, from short, medium, to long sizes (see e.g.<sup>51-53</sup>). For instance, in the example presented in this paper, in addition to the window length of  $T/8$  (168 years), we tried other four window lengths ( $M = 42, 84, 337$ , and 675 years) and the WLMC heat maps for 84 and 337 years are quite similar (whereas the "extreme"  $M$  values of 42 and 675 years are not very different) to that of the corresponding one with a value of  $M = 168$  years (results not shown, but can be obtained through the R code included in the Supplementary Information (Material) or upon request to the corresponding author).
3. Estimation of local least-squares regression applied to the wavelet coefficients  $W_{ij}$  as described in<sup>34</sup>. At each wavelet level the implementation of WLMC in *wavemulcor*<sup>35,47</sup> automatically chooses the variable maximizing the multiple correlation. Alternatively, the user may provide information about the number of the variable (parameter "ymaxr") whose correlation against a linear combination of the others is to be calculated<sup>35,47</sup>. This may be useful if a relationship between the variables under study is known a priori.
4. Estimation of the wavelet local multiple correlation coefficients  $\tilde{\varphi}_{X,s}(\lambda_j)$  applying the consistent estimator in Eq. (3) and estimation of the confidence interval (Eq. 6) to establish the statistical significance of  $\tilde{\varphi}_{X,s}(\lambda_j)$ .

**Datasets.** To illustrate the WLMC method, we analyse the dynamic relationship among three large-scale climate variables closely related to each other and to tropical Atlantic cyclone activity from a multivariate perspective. These climate variables (Fig. 1) are based on paleoclimate reconstructions and cover the interval CE 500–1850<sup>36,37</sup>. They represent: (1) a record of sea surface temperature (SST) anomalies ( $^{\circ}\text{C}$ ) from the main developed region (MDR) for tropical cyclones<sup>36</sup>. The MDRSST record shows interannual to multidecadal scale variability and reflects the "favourability" of the local thermodynamic environment to tropical cyclones formation<sup>36,54-56</sup>. However, we are aware that there has been a considerable debate as to how SST should be viewed in relation to North Atlantic TC activity<sup>57,58</sup>. For instance<sup>58-60</sup>, showed that hurricane frequency is strongly correlated with the so-called "relative SST", i.e. the difference between SSTs averaged over the MDR (MDRSST) and global tropical mean SSTs, and that the "relative SST" is a good predictor of North Atlantic TC counts. (2) The El Niño-Southern Oscillation (ENSO) SST anomalies ( $^{\circ}\text{C}$ ) (the El Niño 3 region)<sup>37</sup>, which is governed by large-scale ocean dynamics and coupled ocean-atmosphere interactions from seasonal to interannual variability. ENSO reduces Atlantic TC activity since ENSO induces changes in tropospheric vertical wind shear and the temperature of the upper troposphere<sup>36,38,40,61</sup>. (3) The North Atlantic Multidecadal Oscillation (AMO) sea surface temperature anomalies ( $^{\circ}\text{C}$ ) averaged over the North Atlantic ocean<sup>37</sup>. AMO induces changes in tropical atmospheric



circulation which alter the tropospheric vertical shear in the MDR, and positive (warm) and negative (cold) of the AMO can decrease or increase the effect of ENSO on Atlantic TC activity<sup>40,62,63</sup>. This case study also aims at examining how these large-scale climate variables (MDRSST, ENSO, and AMO) are correlated with Atlantic tropical cyclone activity for the interval CE 500–1850. To address this aspect, we use annual paleo-hurricane reconstructions of TC counts (smoothed  $\sim 40$  years) from<sup>36</sup> (Fig. 1). We are aware that there are other climate variables (e.g. the North Atlantic Oscillation—NAO, the stratospheric Quasi-Biennial Oscillation—QBO, the Atlantic Meridional Overturning Circulation—AMOC, among others) that have influence on Atlantic tropical cyclone activity, but we have limited our analysis for the following four reasons: (1) the ultimate goal of this case study is to exemplify the use of a novel statistical and computational tool, the wavelet local multiple correlation (WLMC), to analyse multivariate climate time series; (2) availability of data spanning the time interval in question; (3) paleoclimate reconstructions are complex and it can be problematic (from a technical or methodological point of view) to combine data sets from different studies and archives, and (4) these climate variables have been invoked to explain climate dynamics inferred from paleoclimate records<sup>26</sup>.

## Data availability

In this study we have used previously released, freely available datasets. The datasets were obtained from<sup>36,37</sup>. These are freely available at [http://www.meteo.psu.edu/holocene/public\\_html/Nature09/index.htm](http://www.meteo.psu.edu/holocene/public_html/Nature09/index.htm) and [http://www.meteo.psu.edu/holocene/public\\_html/supplements/MultiproxySpatial09/results/](http://www.meteo.psu.edu/holocene/public_html/supplements/MultiproxySpatial09/results/), and in the open repository [https://github.com/jomopo/WLMC\\_climate\\_time\\_series](https://github.com/jomopo/WLMC_climate_time_series). In case of any difficulty in obtaining the datasets mentioned above, the corresponding author can provide the data used upon request.

## Code availability

R codes to produce the figures are available at the open repository [https://github.com/jomopo/WLMC\\_climate\\_time\\_series](https://github.com/jomopo/WLMC_climate_time_series) and as Supplementary Information (please look at the Additional information). In case of any difficulty in obtaining the R code mentioned above, the corresponding author can provide the code used upon request.

Received: 6 October 2020; Accepted: 17 November 2020

Published online: 04 December 2020

## References

- Peixoto, J. P. & Oort, A. H. *Physics of Climate* (American Institute of Physics, New York, 1992).
- Shao, Y., Peng, G. & Leslie, L. M. *The Environmental Dynamic System* (Springer, Berlin, 2002).
- Von Storch, H. & Zwiers, F. W. *Statistical Analysis in Climate Research* (Cambridge University Press, Cambridge, 2001).
- Donner, R., Barbosa, S., Kurths, J. & Marwan, N. Understanding the Earth as a Complex System—recent advances in data analysis and modelling in Earth sciences. *Eur. Phys. J. Special Top.* **174**, 1–9 (2009).
- Stocker, T. F. *et al.* *Climate Change 2013: The Physical Science Basis (IPCC)* (Cambridge University Press, Cambridge, 2013).
- Rial, J. A. *et al.* Nonlinearities, feedbacks and critical thresholds within the Earth's climate system. *Clim. Change* **65**, 11–38 (2004).
- Bojinski, S. *et al.* The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Am. Meteorol. Soc.* **95**, 1431–1443 (2014).
- Mudelsee, M. *Climate Time Series Analysis: Classical Statistical and Bootstrap Methods* Second edn. (Springer, Berlin, 2014).
- Weedon, G. P. *Time-Series Analysis and Cyclostratigraphy: Examining Stratigraphic Records of Environmental Cycles* (Cambridge Univ Press, Cambridge, 2003).
- Wilks, D. S. *Statistical Methods in the Atmospheric Sciences* Vol. 100 (Academic Press, London, 2011).
- 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.* **107**, 177–183 (2011).
- Polanco-Martínez, J. M. & Faria, S. Estimation of the significance of the Foster's wavelet spectrum by means of a permutation test and its application for paleoclimate records. *Boletín Geológico y Minero* **129**, 549–564 (2018).
- Reljin, B., Reljin, I. & Jovanovic, G. Signal processing methods in analyzing of spatial-temporal climate data. In *Signals, Circuits and Systems, 2003. SCS 2003. International Symposium on*, Vol. 1, 49–52 (IEEE, New York, 2003).
- Donner, R. V. & Barbosa, S. M. Nonlinear time series analysis in the geosciences. *Lecture Notes in Earth Sciences* **112**, (2008).
- Polanco-Martínez, J. M., Medina-Elizalde, M. A., Sánchez-Goñi, M. F. & Mudelsee, M. BINCOR: An R package for estimating the correlation between two unevenly spaced time series. *R J.* **11**, 170–184 (2019).
- Babu, P. & Stoica, P. Spectral analysis of nonuniformly sampled data—a review. *Digit. Signal Process.* **20**, 359–378 (2010).
- Rehfeld, K., Marwan, N., Heitzig, J. & Kurths, J. Comparison of correlation analysis techniques for irregularly sampled time series. *Nonlinear Process. Geophys.* **18**, 389–404 (2011).
- Marwan, N. & Kurths, J. Nonlinear analysis of bivariate data with cross recurrence plots. *Phys. Lett. A* **302**, 299–307 (2002).
- Marwan, N. & Kurths, J. Cross recurrence plots and their applications. In *Mathematical physics research at the cutting edge*, 101–139 (Nova Science Publishers, Hauppauge, 2004).
- Whitcher, B., Guttorp, P. & Percival, D. B. Wavelet analysis of covariance with application to atmospheric time series. *J. Geophys. Res. Atmos.* **105**, 14941–14962 (2000).
- Whitcher, B., Byers, S. D., Guttorp, P. & Percival, D. B. Testing for homogeneity of variance in time series: Long memory, wavelets, and the Nile River. *Water Resour. Res.* **38**, 12-1–12-16 (2002).
- Torrence, C. & Compo, G. P. A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* **79**, 61–78 (1998).
- Torrence, C. & Webster, P. J. Interdecadal changes in the ENSO-monsoon system. *J. Clim.* **12**, 2679–2690 (1999).
- Grinsted, A., Moore, J. C. & Jevrejeva, S. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Process. Geophys.* **11**, 561–566 (2004).
- Maraun, D. & Kurths, J. Cross wavelet analysis: Significance testing and pitfalls. *Nonlinear Process. Geophys.* **11**, 505–514 (2004).
- Medina-Elizalde, M., Polanco-Martínez, J. M., Lases-Hernández, F., Bradley, R. & Burns, S. Testing the “tropical storm” hypothesis of Yucatan Peninsula climate variability during the Maya Terminal Classic Period. *Quat. Res.* **86**, 111–119 (2016).
- Sang, Y.-F. A review on the applications of wavelet transform in hydrology time series analysis. *Atmos. Res.* **122**, 8–15 (2013).
- Casagrande, E., Mueller, B., Miralles, D. G., Entekhabi, D. & Molini, A. Wavelet correlations to reveal multiscale coupling in geophysical systems. *J. Geophys. Res. Atmos.* **120**, 7555–7572 (2015).

29. Gabor, D. Theory of communication. Part 1: The analysis of information. *J. Inst. Electr. Eng. Part III Radio Commun. Eng.* **93**, 429–441 (1946).
30. Kaiser, G. *A Friendly Guide to Wavelets* (Springer, Berlin, 1994).
31. Cazelles, B. *et al.* Wavelet analysis of ecological time series. *Oecologia* **156**, 287–304 (2008).
32. Velasco-Herrera, V., Soon, W., Herrera, G. V., Traversi, R. & Horiuchi, K. Generalization of the cross-wavelet function. *New Astron.* **56**, 86–93 (2017).
33. Chavez, M. & Cazelles, B. Detecting dynamic spatial correlation patterns with generalized wavelet coherence and non-stationary surrogate data. *Sci. Rep.* **9**, 7389 (2019).
34. Fernández-Macho, J. Time-localized wavelet multiple regression and correlation. *Phys. A Stat. Mech. Appl.* **492**, 1226–1238 (2018).
35. Fernández-Macho, J. Package wavemulcor: Wavelet multiple regression and correlation in R. *Comput. Sci. Eng.* **21**, 63–73 (2019).
36. Mann, M. E., Woodruff, J. D., Donnelly, J. P. & Zhang, Z. Atlantic hurricanes and climate over the past 1,500 years. *Nature* **460**, 880–883 (2009).
37. Mann, M. E. *et al.* Global signatures and dynamical origins of the Little Ice Age and Medieval Climate Anomaly. *Science* **326**, 1256–1260 (2009).
38. Goldenberg, S. B. & Shapiro, L. J. Physical mechanisms for the association of el Niño and West African rainfall with Atlantic major hurricane activity. *J. Clim.* **9**, 1169–1187 (1996).
39. Tang, B. H. & Neelin, J. ENSO influence on Atlantic hurricanes via tropospheric warming. *Geophys. Res. Lett.* **31**, 1–4 (2004).
40. Patricola, C. M., Saravanan, R. & Chang, P. The impact of the El Niño-Southern Oscillation and Atlantic meridional mode on seasonal Atlantic tropical cyclone activity. *J. Clim.* **27**, 5311–5328 (2014).
41. Zanchettin, D. *et al.* A decadal delayed response of the tropical Pacific to Atlantic multidecadal variability. *Geophys. Res. Lett.* **43**, 784–792 (2016).
42. Geng, X., Zhang, W., Jin, F.-F., Stuecker, M. F. & Levine, A. F. Modulation of the relationship between ENSO and its combination mode by the Atlantic Multidecadal Oscillation. *J. Clim.* **33**, 4679–4695 (2020).
43. Sun, C., Li, J. & Jin, F.-F. A delayed oscillator model for the quasi-periodic multidecadal variability of the NAO. *Clim. Dyn.* **45**, 2083–2099 (2015).
44. Aguiar-Conraria, L. & Soares, M. J. The continuous wavelet transform: Moving beyond uni- and bivariate analysis. *J. Econ. Surv.* **28**, 344–375 (2014).
45. Soon, W. *et al.* A review of Holocene solar-linked climatic variation on centennial to millennial timescales: Physical processes, interpretative frameworks and a new multiple cross-wavelet transform algorithm. *Earth Sci. Rev.* **134**, 1–15 (2014).
46. Fernández-Macho, J. Wavelet multiple correlation and cross-correlation: A multiscale analysis of Eurozone stock markets. *Phys. A Stat. Mech. Appl.* **391**, 1097–1104 (2012).
47. Fernández-Macho, J. wavemulcor: Wavelet routines for global and local multiple correlation (2018). <https://CRAN.R-project.org/package=wavemulcor>. R package version 2.2.1.
48. Whitcher, B. & Whitcher, M. B. The waveslim package (2019). <https://CRAN.R-project.org/package=waveslim>. R package version 1.7.5.1.
49. Daubechies, I. *Ten lectures on wavelets*, vol. 61 (SIAM - Society for Industrial and Applied Mathematics, 1992).
50. Gençay, R., Selçuk, F. & Whitcher, B. *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics* (Academic Press, London, 2002).
51. Polanco-Martínez, J. M., Fernández-Macho, J., Neumann, M. & Faria, S. H. A pre-crisis vs. crisis analysis of peripheral EU stock markets by means of wavelet transform and a nonlinear causality test. *Phys. A Stat. Mech. Appl.* **490**, 1211–1227 (2018).
52. Polanco-Martínez, J. M. Dynamic relationship analysis between NAFTA stock markets using nonlinear, nonparametric, non-stationary methods. *Nonlinear Dyn.* **97**, 369–389 (2019).
53. Polanco-Martínez, J. RolWinMulCor: An R package for estimating rolling window multiple correlation in ecological time series. *Ecol. Inform.* **60**, 101163 (2020).
54. Emanuel, K. Increasing destructiveness of tropical cyclones over the past 30 years. *Nature* **436**, 686–688 (2005).
55. Hoyos, C. D., Agudelo, P. A., Webster, P. J. & Curry, J. A. Deconvolution of the factors contributing to the increase in global hurricane intensity. *Science* **312**, 94–97 (2006).
56. Trenberth, K. E. & Fasullo, J. Energy budgets of Atlantic hurricanes and changes from 1970. *Geochem. Geophys. Geosyst.* **9**, 1–12 (2008).
57. Hall, T. & Yonekura, E. North American tropical cyclone landfall and SST: A statistical model study. *J. Clim.* **26**, 8422–8439 (2013).
58. Mei, W., Kamae, Y., Xie, S.-P. & Yoshida, K. Variability and predictability of North Atlantic hurricane frequency in a large ensemble of high-resolution atmospheric simulations. *J. Clim.* **32**, 3153–3167 (2019).
59. Vecchi, G. A. & Soden, B. J. Effect of remote sea surface temperature change on tropical cyclone potential intensity. *Nature* **450**, 1066–1070 (2007).
60. Johnson, N. C. & Xie, S.-P. Changes in the sea surface temperature threshold for tropical convection. *Nat. Geosci.* **3**, 842–845 (2010).
61. McPhaden, M. J., Zhang, X., Hendon, H. H. & Wheeler, M. C. Large scale dynamics and MJO forcing of ENSO variability. *Geophys. Res. Lett.* **33**, 1–5 (2006).
62. Goldenberg, S. B., Landsea, C. W., Mestas-Núñez, A. M. & Gray, W. M. The recent increase in Atlantic hurricane activity: Causes and implications. *Science* **293**, 474–479 (2001).
63. Knight, J. R., Folland, C. K. & Scaife, A. A. Climate impacts of the Atlantic multidecadal oscillation. *Geophys. Res. Lett.* **33**, 1–4 (2006).

## Acknowledgements

J.M.P.M. was funded by the PIC 444/18 – EU Interreg project MOSES (EAPA 224/2016), FEDER funds and the SEPE (Spanish Public Service of Employment). J.F.M. acknowledges research funding received from UPV/EHU Econometrics Research Group (Basque Government Dpt. of Education grant IT-1359-19) and Spanish Ministry of Economy and Business (grant MTM2016-74931-P).

## Author contributions

J.M.P.M. designed this study, analysed data and results, and prepared the figures. J.F.M. developed the wavelet local multiple correlation and wrote the method section. M.M.E. provided expertise in the analysis of the paleoclimate data and helped to interpret the results. J.M.P.M. drafted and led the manuscript writing and revision with continuous inputs from all authors. All authors contributed and reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-77767-8>.

**Correspondence** and requests for materials should be addressed to J.M.P.-M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020