Computers & Geosciences 91 (2016) 11–18

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

REDFIT-X: Cross-spectral analysis of unevenly spaced paleoclimate time series

Kristín Björg Ólafsdóttir [a,b,*], Michael Schulz [b], Manfred Mudelsee [a]

[a] Climate Risk Analysis, Heckenbeck, 37581 Bad Gandersheim, Germany
[b] MARUM - Center for Marine Environmental Sciences, University of Bremen, 28334 Bremen, Germany

ARTICLE INFO

Article history:
Received 13 July 2015
Received in revised form
23 December 2015
Accepted 1 March 2016
Available online 4 March 2016

Keywords:
Cross-spectral analysis
Coherency
Phase spectrum
Irregular sampling
Monte Carlo simulations
Paleoclimate time series

ABSTRACT

Cross-spectral analysis is commonly used in climate research to identify joint variability between two variables and to assess the phase (lead/lag) between them. Here we present a Fortran 90 program (REDFIT-X) that is specially developed to perform cross-spectral analysis of unevenly spaced paleoclimate time series. The data properties of climate time series that are necessary to take into account are for example data spacing (unequal time scales and/or uneven spacing between time points) and the persistence in the data. Lomb–Scargle Fourier transform is used for the cross-spectral analyses between two time series with unequal and/or uneven time scale and the persistence in the data is taken into account when estimating the uncertainty associated with cross-spectral estimates. We use a Monte Carlo approach to estimate the uncertainty associated with coherency and phase. False-alarm level is estimated from empirical distribution of coherency estimates and confidence intervals for the phase angle are formed from the empirical distribution of the phase estimates. The method is validated by comparing the Monte Carlo uncertainty estimates with the traditionally used measures. Examples are given where the method is applied to paleoceanographic time series.

1. Introduction

Cross-spectral analysis is often used to estimate the relationship between two time series as a function of frequency. Of particular importance are the coherency and phase spectrum. Coherency is a dimensionless measure on how well two time series co-vary at different frequencies while the phase spectrum shows if the variations happen synchronously at each frequency or if there is a phase difference between them. Cross-spectral analysis is used in climate research to identify joint variability between two variables and to assess the phase (lead/lag) between them.

Paleoclimate proxy time series come from various archives such as marine sediments, ice caps, lake sediments, speleothems, tree rings or corals. Usually the sampling is carried out at constant length intervals and then transferred into the time domain, either by direct dating, or by aligning with other dated time series. Most cross-spectral analysis methods require that the two time series are sampled at identical times and have constant spacing between time points (evenly spaced). This is rarely the case with paleoclimate time series as the archives do normally not accumulate at constant rate, which makes in many cases some kind of interpolation necessary prior to the analysis. Unfortunately, interpolation can bias the spectral results substantially as the spectral power may be shifted from higher to lower frequencies, that is, the spectrum becomes redder (Schulz and Stattegger, 1997). The interpolation can be avoided by estimating the spectrum directly from unevenly spaced time series with the Lomb–Scargle Fourier transform (Lomb, 1976; Scargle, 1982) as done for example in the computer programs by Schulz and Stattegger (1997), Schulz and Mudelsee (2002) and Pardo-Igúzquiza and Rodríguez-Tovar (2012).

Generally, climate time series include persistence (serial correlation) or memory as there is natural inertia in the climate system. Due to the persistence, the spectra of climate time series are characterized by greater amplitude values at lower frequencies (red noise). To distinguish the signals (spectral peaks) in the spectrum of climate time series from background variability they need to be tested against red noise. First-order autoregressive or AR(1) process can be used to model the climate noise (Hasselmann, 1976). The model is normally fitted to the observed time series and the estimated AR(1) parameter is used to form the red noise spectrum (Allen and Smith, 1996). In REDFIT (Schulz and Mudelsee, 2002), the AR(1) parameter is estimated directly from the unevenly spaced time series, so there is no need to interpolate the time series, which can bias the estimated value. The estimated

* Corresponding author at: Climate Risk Analysis, Heckenbeck, 37581 Bad Gandersheim, Germany.
E-mail address: olafsdottir@climate-risk-analysis.com (K. Björg Ólafsdóttir).

http://dx.doi.org/10.1016/j.cageo.2016.03.001
0098-3004/

AR(1) parameter is used to form a theoretical AR(1) spectrum and false-alarm level for testing the significance of spectral peaks via the χ^2 distribution. In addition Monte Carlo simulations can be used, where a large number of red noise processes are generated with the same estimated AR(1) parameter, to form false-alarm level as percentiles of the Monte Carlo ensemble (Schulz and Mudelsee, 2002).

The REDFIT program can only be used for univariate spectral analysis or the autospectrum. Given the need for cross-spectral analysis for unevenly spaced data where the significance is evaluated with Monte Carlo simulations, we present a computer program REDFIT-X, in which cross-spectral analysis has been implemented. Until now the computer program SPECTRUM (Schulz and Stattegger, 1997) has been available for cross-spectral analysis for unevenly spaced climate time series. However the significance measurements in SPECTRUM do not allow for the persistence included in paleoclimate time series. Therefore it is necessary to combine the two approaches, to perform both auto- and cross-spectral analysis with reliable uncertainty estimates.

2. Method

2.1. Cross-spectral analysis – background

2.1.1. Coherency and phase spectrum

The most important features of the cross-spectrum are coherency spectrum and phase between the two signals. Coherency is defined as:

$$c_{xy}^2(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)\, G_{yy}(f)}, \tag{1}$$

where $G_{xx}(f)$ and $G_{yy}(f)$ are the autospectra of the signals $x(t)$ and $y(t)$ (with t being time), respectively and $G_{xy}(f)$ is the cross-spectrum between them (e.g., Bendat and Piersol, 2010). It is a dimensionless measure that informs about the degree of linear relationship between two time series, as a function of frequency (f). Coherency is in the range from 0 (no relationship) to 1 (perfect relationship) and can be thought of as a squared correlation coefficient depending on frequency (von Storch and Zwiers, 2003).

Coherency is estimated as:

$$\hat{c}_{xy}^2(f_k) = \frac{\left|\hat{G}_{xy}(f_k)\right|^2}{\hat{G}_{xx}(f_k)\, \hat{G}_{yy}(f_k)}, \tag{2}$$

where $\hat{G}_{xx}(f_k)$ and $\hat{G}_{yy}(f_k)$ are the estimated autospectra and $\hat{G}_{xy}(f_k)$ is the estimated cross-spectrum of two weakly stationary time series $\{t_x(i), x(i)\}_{i=1}^{n_x}$ and $\{t_y(i), y(i)\}_{i=1}^{n_y}$ (n_x and n_y are numbers of data points in each series) (e.g., Bendat and Piersol, 2010). The frequency f_k is in the range from the fundamental frequency $\bar{f} = 1/(n\bar{d})$ to the average Nyquist frequency $f_{\mathrm{Nyq}} = 1/(2\bar{d})$, where n is the number of data points and $\bar{d} = [t(n) - t(1)]/(n-1)$ is the average spacing of the time series (Mudelsee, 2010). Apparently when two time series do not have the same sampling points, the average spacing ($\bar{d}_x$, $\bar{d}_y$) and the fundamental frequency ($\bar{f}_x$, $\bar{f}_y$) for each time series can differ. To ensure that the time series with the lower resolution determines these variables, we use $\bar{d}_{xy} = \max(\bar{d}_x, \bar{d}_y)$, $\bar{f}_{xy} = \max(\bar{f}_x, \bar{f}_y)$ and the average Nyquist frequency is determined as $f_{\mathrm{Nyq}} = 1/(2\bar{d}_{xy})$ (Schulz and Stattegger, 1997).

The auto- and cross-spectra are estimated with the Lomb–Scargle Fourier transform (Lomb, 1976; Scargle, 1982) in combination with the "Welch's Overlapped Segment Averaging" (WOSA) procedure (Welch, 1967) as done in Schulz and Stattegger (1997).

The WOSA segmenting is used to smooth the estimated raw spectrum and make it consistent (the raw spectrum is an inconsistent estimator as the variance does not decrease with increasing data size). The time series of length n_x and n_y are split into a number n_{50} of overlapping segments of length n_{seg}^x and n_{seg}^y (with 50% overlap)

$$\begin{aligned} x_i(j) &= x[(i-1)(n_{\mathrm{seg}}^x/2) + j], \quad j = 1, \ldots, n_{\mathrm{seg}}^x, \\ y_i(j) &= y[(i-1)(n_{\mathrm{seg}}^y/2) + j], \quad j = 1, \ldots, n_{\mathrm{seg}}^y, \end{aligned} \tag{3}$$

where $i = 1, \ldots, n_{50}$. A linear trend is subtracted from each segment to avoid possible artifacts at low frequencies (i.e., resulting from periods, which exceed the segment length). Of course, interpretation of the low-frequency part of a spectrum requires sufficiently long segments. The segments are multiplied by a taper $w(j)$, $j = 1, \ldots, n_{\mathrm{seg}}$ (see different types of spectral windows in Harris, 1978) to reduce spectral leakage and then Fourier transformed

$$\begin{aligned} X_i(f_k) &= \mathcal{F}_{\mathrm{LS}}\{x_i(j)w(j)\}, \\ Y_i(f_k) &= \mathcal{F}_{\mathrm{LS}}\{y_i(j)w(j)\}, \end{aligned} \tag{4}$$

where $\mathcal{F}_{\mathrm{LS}}$ denotes the Lomb–Scargle Fourier transform (Lomb, 1976; Scargle, 1982). Finally the n_{50} segments are averaged to form consistent auto- and cross-spectrum

$$\hat{G}_{xx}(f_k) = \frac{2}{n_{50}\bar{f}_{xy}\, n_{\mathrm{seg}}^x} \sum_{i=1}^{n_{50}} |X_i(f_k)|^2, \tag{5}$$

$$\hat{G}_{yy}(f_k) = \frac{2}{n_{50}\bar{f}_{xy}\, n_{\mathrm{seg}}^y} \sum_{i=1}^{n_{50}} |Y_i(f_k)|^2, \tag{6}$$

$$\hat{G}_{xy}(f_k) = \frac{2}{n_{50}\bar{f}_{xy}\sqrt{n_{\mathrm{seg}}^x n_{\mathrm{seg}}^y}} \sum_{i=1}^{n_{50}} \left[X_i(f_k) Y_i^*(f_k)\right], \tag{7}$$

where $*$ denotes the complex conjugate (Schulz and Stattegger, 1997).

The coherency estimate is biased, where the coherency between two uncoupled time series are expected to be greater than zero (Benignus, 1969). We use the bias approximation from Bendat and Piersol (2010) to form a bias-corrected coherency estimate

$$\begin{aligned} \mathrm{bias}\left[\hat{c}_{xy}^2\left(f_k\right)\right] &\approx \left[1 - \hat{c}_{xy}^2\left(f_k\right)\right]^2 / n_{\mathrm{eff}}, \\ \hat{c}_{xy}'^2\left(f_k\right) &= \hat{c}_{xy}^2\left(f_k\right) - \mathrm{bias}\left[\hat{c}_{xy}^2\left(f_k\right)\right], \end{aligned} \tag{8}$$

where n_{eff} is the effective number of segments (defined in Section 2.1.2).

The phase spectrum is estimated as

$$\hat{\phi}_{xy}(f_k) = \tan^{-1}\left(\frac{\hat{Q}_{xy}\left(f_k\right)}{\hat{C}_{xy}\left(f_k\right)}\right), \tag{9}$$

where $\hat{Q}_{xy}(f_k)$ and $\hat{C}_{xy}(f_k)$ are the real and imaginary parts of the estimated cross-spectrum $\hat{G}_{xy}(f_k)$, respectively (e.g., Bendat and Piersol, 2010). We use the four-quadrant inverse tangent function in Fortran (atan2), which returns the result in appropriate quadrant. Therefore the estimated phase angle falls in the range from $[-180°, 180°]$, where zero value means that the two time series are in phase while non zero value means out of phase.