1 Motivation:

Several approaches are commonly used to quantify the potential association between two evenly spaced time series, e.g., Pearson’s and Spearman’s correlation or the cross-correlation function (CCF). However, it is not straightforward when the time series are unequally spaced

- Particularly for the case when the two time series under analysis are not sampled on identical time points, which is usual in paleoclimate research [1,2].
- The most common way to tackle this problem is to interpolate in time the original unequally spaced climate time series in order to obtain equidistant and same times. Then, these series can be analysed using existing conventional correlation analysis techniques.
- However, experience shows that interpolation has its drawbacks: it is dependent of the method of Gaussian-Kernel-based cross-correlation (gXCF) and its associated software EEDON [3].

Thus, interpolation should be avoided as much as possible.

- Fortunately, at least for the case of unequally spaced climate time series sampled on identical time points, there exist some algorithms and software to carry out this task [4,5].
- However, when the two time series are not sampled on identical time points, there are few statistical techniques to estimate the correlation between these. One exception is the method of Gaussian-Kernel-based cross-correlation (gXCF) and its associated software EEDON [3] and the extended version [4], or the binned correlation proposed by BINCOR [6]. The software of this method, however, is not freely available on the Internet.

2 Method:

The procedure is described as follows:

1. Input: two ununequally spaced climate time series \( X(t_i), T_{X(i)} \) and \( Y(t_i), T_{Y(i)} \), where \( T_{X(i)} \) and \( T_{Y(i)} \) are the time domains and the sample sizes of each series, respectively.

2. Compute the average spacing among samples:
   
   \[
   \bar{d}_X = \frac{1}{n_{X}} \sum_{i=1}^{n_{X}} |T_{X(i)} - T_{X(i-1)}|, \quad \bar{d}_Y = \frac{1}{n_{Y}} \sum_{i=1}^{n_{Y}} |T_{Y(i)} - T_{Y(i-1)}|, \quad \bar{d}_{max} = \max(\bar{d}_X, \bar{d}_Y)
   \]

3. Estimate the bin-width \( b \) taking into account the persisitence (memory) estimated of each unequally spaced climate time series, \( X \) and \( Y \), which are denoted as \( T_{X(i)} \) and \( T_{Y(i)} \), respectively.

   - Estimate the persistence, an AR1 model [7] is fitted to each unequally spaced time series.

   - The bias-estimated equivalent autocorrelation coefficients
     
     \[
     \hat{a}_X = \exp(-\frac{\bar{d}_X}{2b}), \quad \hat{a}_Y = \exp(-\frac{\bar{d}_Y}{2b}) \quad \text{with} \quad \hat{a}_X < \hat{a}_Y
     \]

   - The estimate bin-width as
     
     \[
     b = -\frac{\ln(\hat{a}_X)}{\bar{d}_X} \quad \text{and} \quad b = -\frac{\ln(\hat{a}_Y)}{\bar{d}_Y}
     \]

4. Determine the number of bins: \( N_{X} = \max(T_{X_{max}}/b) \)

5. Set: \( n_{X}(n) = 1 \), \( n_{Y}(n) = 1 \). Then, for \( n = 1, 2, \ldots , N_{X} \) define:

   - \( \text{lim}_{n} = \text{lim}_{n}(n) = T_{X(n)} - T_{X(n-1)} \)

   - \( \Delta T_{X} = \frac{1}{\text{lim}_{n}} \text{lim}_{n} \geq \text{lim}_{n}(n) \) and \( T_{X} \leq \text{lim}_{n}(n) \)

   - \( \Delta T_{Y} = \frac{1}{\text{lim}_{n}} \text{lim}_{n} \geq \text{lim}_{n}(n) \) and \( T_{Y} \leq \text{lim}_{n}(n) \)

6. Compute the correlation between the two binned time series through a correlation function (CCF) or through a smoothing function (SMF) or through BINCOR functions cor and ccf.

Monte Carlo Experiments:

We conducted Monte Carlo experiments to study how the specific rules chosen for calculating the bin-width based on persistence reduces the error to arbitrarily choosing a bin-width and we have used bivariate Gaussian AR1 process for uneven time spacings to generate 5000 replicates.

3 Case study: abrupt climate changes - Last Glacial

4 The BINCOR R package

5 Case study: abrupt climate changes - Last Glacial

- Before applying the correlog function, a linear trend was removed. From the two time-series, the time period containing the optimum correlation was determined and the correlation was calculated using this out-of-phase relationship.

6 Final remarks:

We presented a computational package named BINCOR (BiNitudinal CORrelation) that can be used to estimate the correlation between two unequally spaced climate time series and is based on a novel estimation approach proposed by BINCOR. The package contains four functions (bin_cor, cor_b, ccf, and plot_bsd). It is available in R and is available from CRAN.

- BINCOR requires the concept of nonzero persistence times, thus allowing to recover the mixing information, even when the two examined timescales differ from each other.

- The results from applying BINCOR to real climate data sets suggest that the BINCOR properly in detecting relationships between paleoclimate records.

Key references

- [2] A. D. Wintle, “Primary” (unevenly spaced) and binned pollen time series under analysis. The number of elements or time steps for the time series from site MD04-2845 is \( N = 685 \) and \( N = 3500 \) for the time series from site MD95-2039 and \( N = 685 \) and \( N = 1300 \) years.
- [3] BINCOR (BiNitudinal CORrelation) is available on CRAN (http://CRAN.R-project.org/package=BINCOR) and contains four functions:
  - bin_cor: the main function to build the binned time series.
  - plot_bsd: to plot and compare the “primary” and binned time series.
  - cor_b: to estimate the correlation between the binned time series.
  - ccf: to estimate the cross-correlation between binned time series.

The graphical outputs can be displayed on the screen or saved as PDF/PNG. BINCOR depends on the dipr2 and pracma packages. The dipr2 package is used by the function bin_cor to calculate the persistence for the climate time series under analysis. The pracma package is used by the functions cor_b and ccf to remove the linear trend before estimating the correlation.

Acknowledgments: JPFM was funded by a Basque Govt post-doctoral fellowship. MM’s work was supported by the European Commission via Marie Curie Initial Training Network LICIN (project number 289547) under the Seventh Framework Programme.

ABSTRACT

We present a computational program named BINCOR (BiNitudinal CORrelation) to estimate the correlation between two unevenly spaced climate time series, which is based on a novel estimation approach proposed by [1]. The idea is that autocorrelation (e.g., AR1) has memory, which allows to correlate values obtained on different time points. The binned correlation is performed by resampling the time series under study into time bins on a regular grid that are assigned the mean values of the variable under scrutiny within those bins. To exemplify its use, we apply BINCOR to two temporally unevenly-sampled pollen records from two marine sediment cores (MD04-2845 & MD95-2039) collected on the southwestern European margin. These series from a global pollen & charcoal database [2] developed under the framework of the INQUA International Focus Group ACER (Abrupt Climate Changes and Environmental Responses). BINCOR works properly in detecting the well-established relationships between the compared climate records.