Mutual Information and Entropy Transfer between Ice-Core Temperature time series and Stalagmite Precipitation time series

Juan Carlos Romero Gelvez

Department of Earth and Planetary Sciences

jromerogelvez@ucdavis.edu

Abstract

Distinguishing between climate-change triggering processes and secondary climatic processes is essential to advance in the understanding of last glacial period climate dynamics. However, differentiating between scenarios where two climatic processes respond to a high-order process, and scenarios where these two processes are intrinsically related because one process is driving changes in the other, is challenging. Comparisons between time series can produce high values of correlation without necessary implying that one of the processes is driving the other. Moreover, commonly used estimates of correlation between climatic time series, such as Pearson correlation coefficients, cannot capture the non-linear nature of the climate. Therefore, Measures of similarity and information transfer that can capture non-linear relationships and give a sense of causation between paleo-proxy climatic time series are needed to solve the aforementioned limitations. In here, I used measures of Mutual information and Entropy transfer to understand the relationship between temperature changes in high latitudes and precipitation in the tropics during the last glacial maximum interval (i.e. spanning from ~`18 to 22 ka). Time series were compared using NEST and RTransferentropy packages in R studio and DIT package in python. NEST package can give several estimates of information theory using as input continuous data, while for working in DIT discretization of the data was necessary. Data was discretized using Pandas package in python using different binning values until higher values of mutual information were obtained. My results show that during the last glacial maximum there was a higher mutual information correlation between tropical rainfall and high latitude temperatures of the southern hemisphere than with temperatures of the northern hemisphere, and that information flowed from tropical to extratropical locations.

Introduction

One major issue in paleo-climate analysis of time series is finding ways to compare time-uncertain and unevenly spaced observations with methods that adequately capture the linear and non-linear dependences inherent to climate dynamics (Rehfeld et al 2014). Given that in paleo-proxies such as stalagmites, ice cores, sediment cores, corals, etc. the sampling intervals cannot be guarantee, and thus observations usually are discontinuous and unequally spaced, the resulting time series most likely contain a set or unevenly spaced observations constrained by the aforementioned sampling limitations. Moreover, as dating techniques such as carbon dating, radiometrical dating, layer counting, have associated uncertainties, the resulting age-depth relationships used to place the observations in time will also have an associated age-uncertainty. These limitations have obstructed to adequately address similarity estimates between paleo-climatic time series, and thus have made extremely difficult interpreting the dynamics of the last glacial period climate system. Overcoming these limitation is a need is paleo-climatic analysis and in geosciences in general. Therefore, information theory can provide the tools that adequately can capture the non-linear nature of the climate and at the same time provide a sense of causation between different climatic processes interacting in periods of major climate change. For example,
the mutual information(MI) of two random variables is a measure of the mutual dependence between observations (). Since MI is a probabilistic measurement, it can capture non-linear dependences unlike commonly used estimates of similarity in geosciences such as Pearson coefficient correlations (Runge et al., 2019). Moreover, time-delayed mutual information (TDMI) can provide sense of causation in the climatic signals by lagging or leading the time series (Schreiber 2000), thus indicating the probability that one process drove the another. Similarly, transfer entropy states that if a signal X has a causal influence on a signal Y, then the probability of Y conditioned on its past is different from the probability of Y conditioned on both its past and the past of X. The same idea can be expressed observing that entropy on the present measurement of Y is reduced if knowledge of the past of X is added to knowledge of the past of Y. A great advantage of transfer entropy compared with the other methods is that it does not require any prior assumption on data generation (i.e., it is model-free) (Ursino et al., 2022). In here I investigated the controls from past temperature oscillations from both hemispheres over the position of the ITCZ (i.e. tropical rainfall that moves north or south towards the warmer hemisphere). Mutual information indicates the amount of information shared between X(t) and Y(t).

Therefore, high values of MI between northern/southern hemisphere temperatures and tropical precipitation will indicate coeval changes in Northern/Southern hemisphere temperatures and tropical precipitation. Since there is still ample of questions about climate dynamics in the last glacial period that haven't been addressed correctly, understanding these type of relationship (e.g. tropical rainfall and high-latitude temperatures) is important to fully understand glacial period climate dynamics. My results show that MI was higher, during the last glacial maximum, between tropical precipitation and Antarctica, and that information flowed from tropical to extratropical locations.

Background

The mutual information of two processes I and J with joint probability $_{pIJ}$(i,j) can be seen as the excess amount of code produced by erroneously assuming that the two systems are independent (i.e., using $_{qIJ}$(i,j)= $_{pI}$(i) pJ(j) instead of $_{pIJ}$(i,j)) (Schreiber 2000). The corresponding Kull-back entropy is:

$$M_{IJ} = \sum p(i,j) \log \frac{p(i,j)}{p(i)p(j)}, \quad \textbf{Eq.1}$$

Which is the well know formula for mutual information (Schreiber 2000). In addition, Mutual information can be given a directional sense in a somewhat ad hoc way by introducing a time lag in either one of the variables and compute (Schreiber 2000), e.g.,

$$M_{IJ}(\tau) = \sum p(i_n, j_{n-\tau}) \log \frac{p(i_n, j_{n-\tau})}{p(i)p(j)}. \quad \textbf{Eq.2}$$

Moreover, Transfer entropy is a non-parametric statistic measuring the amount of directed (time-asymmetric) transfer of information between two random processes. Transfer entropy from a process X to another process Y is the amount of uncertainty reduced in the future values of Y by knowing the past values of X given past values of Y. measured using Shannon's entropy (Schreiber 2000). The transfer entropy can be written as:

$$T_{X\to Y} = H\left(Y_t \mid Y_{t-1:t-L}\right) - H\left(Y_t \mid Y_{t-1:t-L}, X_{t-1:t-L}\right) \quad \textbf{Eq.3}$$

Where H(X) and H(Y) are the Shannon entropies of X and Y respectively.

In addition to the information theory measurements described above, it is pertinent to introduce the measurements and techniques used for reconstructing climatic variables in the different time-series used in this class project. Paleo-proxy reconstructions commonly use stable isotopes as the general indicator for climatic variables such as temperature and amount of precipitation. For example, in the stalagmite samples (i.e. cave deposits) which are made up of calcite (i.e. $CaCO_3$) the most important factor controlling the isotopic composition of $\delta^{18}O$ in calcite is the $\delta^{18}O$ of rainwater, which is controlled by the amount of rainfall. This relationship is known as the ''amount effect'' of isotopes in precipitation (i.e. inverse relationship between $\delta^{18}O$ content in rain samples and the amount of precipitation) (Fairchild and Baker, 2012), On the other hand, in ice cores the $\delta^{18}O$ values and temperatures are directly correlated. In stalagmites age-depth relationships are constructed used Uranium-thorium radiometrical dating yielding to uncertainties in the time-series that in this case are of no more than +/-200 yrs. Similarly, ice core age-depth relationships are constructed using layer counting techniques tied to layers where absolute dating can be applied (e.g. Layers with dust or volcanic ash). It is commonly thought that oscillation in amounts of precipitation during the last glacial period present in stalagmite records from the tropics were triggered by temperature oscillations in the northern hemisphere. However, Pearson correlation coefficients estimates from comparison between stalagmite records in the Colombian Andes and temperature records from ice cores in high latitudes of the southern hemisphere also show good correspondence. Therefore, this class project aims to compare the high-resolved and absolute dated climatic time-series which reconstruct amounts of precipitation from stalagmite in the tropics and temperatures from high-latitudes inferred from ice cores using mutual information, time delay mutual information and entropy transfer to get a sense of what signals are more alike and in what direction information flowed during some period of the last glacial period. This pilot data will serve as basis to construct a more robust analysis of causation and directionality of climatic change signals during the last glacial period necessary to understand future climate change.

Methods

Mutual Information & Time-Delay Mutual Information Using NEST package in R

The mutual information calculation applied in here follows the procedure described in Rehfeld et al., 2011 which is described as follows: A local reconstruction of the signal is performed by estimating for each point I in the time series $X=(t^x, x)$ a corresponding observation from $Y=(t^y, y)$, by estimating a local, observation-time weighted mean y^{lr}_j around a time point t^x_i in Y,

$$y_j^{lr} = \sum_{i=1}^{N_y} b(d,k,h)\, y_i, \qquad \textbf{Eq.4}$$

$$b(d,k,h) = \frac{1}{\sqrt{2\pi h}} e^{-|d|^2/2h^2}, \qquad \textbf{Eq.5}$$

with the Gaussian-kernel based local weight b (d, k, h) defined as in Eq. (5). For MI the standard deviation of the Gaussian weight function is set to h=0.5. If there are no observations y_i available in a time window $\pm\tau\Delta t$ around t_x^i this reconstruction is not performed. Repeating this for each time point j =1, . . . , N_x in X one obtains a new, bivariate set of observations:

$$\mathbf{Y}^x = \left(t_i^x, x_i, y_i^{lr}\right). \qquad \textbf{Eq.6}$$

Afterwards the procedure is repeated by stepping through t_y^j, which yields

$$\mathbf{X}^y = \left(t_j^y, x_j^{lr}, y_j\right). \qquad \textbf{Eq.7}$$

Later the local reconstruction Y^x and the original observations Y are then concatenated into one vector $Y^r=\{Y \cup Y^x\}$ combining locally reconstructed and original observations. Similarly, a vector $X^r=(X \cup X^y)$ is obtained. Finally, based on this set of bivariate observations (X^r, Y^r) the joint density of X and Y

can be estimated using standard binning estimators for MI.

After mutual information values are obtained, one set of observations are lagged in time to obtained the TDMI.

Entropy Transfer from RTransferentropy in R Studio

The transfer entropy was calculated following the procedure described in the RTransfer entropy documentation in R studio. Let *log* denote the logarithm to the base 2, then informational gain is measured in bits. Shannon entropy (Shannon 1948) states that for a discrete random variable *J* with probability distribution

$p(j)$, where j stands for the different outcomes the random variable J can take, the average number of bits required to optimally encode independent draws from the distribution of J can be calculated as

$$H_J=-\sum_j p(j)\cdot \log(p(j)) \quad \textbf{Eq.8}$$

Strictly speaking, Shannon's formula is a measure for uncertainty, which increases with the number of bits needed to optimally encode a sequence of realizations of J. In order to measure the information flow between two processes, Shannon entropy is combined with the concept of the Kullback-Leibler distance (Kullback and Leibler 1951) and by assuming that the underlying processes evolve over time according to a Markov process (Schreiber 2000). Let I and J denote two discrete random variables with marginal probability distributions $p(i)$ and $p(j)$ and joint probability distribution $p(i,j)$, whose dynamical structures correspond to stationary Markov processes of order k (process I) and l (process J). The Markov property implies that the probability to observe I at time $t+1$ in state i conditional on the k previous observations is $p(i_{t+1}|i_t,\ldots,i_{t-k+1})=p(i_{t+1}|i_t,\ldots,i_{t-k})$. The average number of bits needed to encode the observation in $t+1$ if the previous k values are known is given by

$$h_I(k)=-\sum_i p(i_{t+1},i^{(k)}_t)\cdot \log(p(i_{t+1}|i^{(k)}_t)) \quad \textbf{Eq.9}$$

where $i^{(k)}_t=(i_t,\ldots,i_{t-k+1})$. $h_J(l)$ can be derived analogously for process J. In the bivariate case, information flow from process J to process I is measured by quantifying the deviation from the generalized Markov property $p(i_{t+1}|i^{(k)}_t)=p(i_{t+1}|i^{(k)}_t,j^{(l)}_t)$ relying on the Kullback-Leibler distance (Schreiber 2000). Thus, (Shannon) transfer entropy is given by

$$T_{J\rightarrow I}(k,l)=\sum_{i,j} p(i_{t+1},i^{(k)}_t,j^{(l)}_t)\cdot \log\left(\frac{p(i_{t+1}|i^{(k)}_t,j^{(l)}_t)}{p(i_{t+1}|i^{(k)}_t)}\right) \quad \textbf{Eq.10}$$

where $T_{J\rightarrow I}$ consequently measures the information flow from J to I ($T_{I\rightarrow J}$ as a measure for the information flow from I to J can be derived analogously). In order to assess the statistical significance of transfer entropy estimates, we rely on a Markov block bootstrap as proposed by Dimpfl and Peter (2013). In contrast to shuffling, the Markov block bootstrap preserves the dependencies within each time series. Thereby, it generates the distribution of transfer entropy estimates under the null hypothesis of no information transfer, i.e. randomly drawn blocks of process J are realigned to form a simulated series, which retains the univariate dependencies of J but eliminates the statistical dependencies between J and I. Shannon or Rényi transfer entropy is then estimated based on the simulated time series. Repeating this procedure yields the distribution of the transfer entropy estimate under the null of no information flow. The p-value associated with the null hypothesis of no information transfer is given by $1-\hat{q}\,TE$, where $\hat{q}\,TE$ denotes the quantile of the simulated distribution that corresponds to the original transfer entropy estimate.

Dynamical System

Climate is considered a non-linear dynamical system. The climate response to perturbations can include feedback mechanisms that affect the linearity of the climatic response. For example, climatic regimen shifts (e.g. glacial vs interglacial periods) that are triggered by changes in solar insolation will cause a decrease in global temperatures. However, the temperature response not only depends on the reduced amount of solar insolation, but on other climatic variables such as the amount of ice in the planet. As temperature drops glacier buildup takes place which in turns increase the planetary albedo decreasing even more global temperatures. The system is also time dependent example, the amount of solar radiation varies periodically and depending on the position of the earth against the sun. The cause of stadials and interstadials is still matter of debate. It has been hypothesized that during stadials (cold periods) the strength of the Atlantic meridional overturning circulation(AMOC) (i.e. an oceanic current that transfer heat in the Atlantic from south to north) is reduced. Warm waters from tropical locations that usually get into the northern hemisphere are reduced, therefore cooling the north Atlantic. As a result, cold conditions are developed in the northern hemisphere. At the same time, the reduced AMOC caused buildup of warm waters in the southern hemisphere warming it up. The mechanism is known as the Bipolar seesaw. The oscillations in temperature present in the Greenland and Antarctica ice core time-series are direct result of the Bipolar seesaw mechanic. As AMOC slow down Greenland cools down and Antarctica warms up. Consequentely, tropical precipitation in Colombia decreases due to precipitation migrating towards warmer parts of the planet.

Results

For the interval selected (i.e. 15 to 22 ka) TDMI values show that there is a closer relationship between Antarctica and Colombia. The highest MI correlation between both time series is reached at a lag of +/- 200 yrs. Greenland and Colombia also exhibit good MI correlation although smaller than between Colombia and

Antarctica. It is interested however that the MI graph of Greenland vs Colombia show symmetry at -100 lag. However, the time interval covered in this work doesn't allow to see how this symmetry looks like going back further in time. Future work will try to lag the calculations of MI even more to see how symmetrical are these values. In the Case

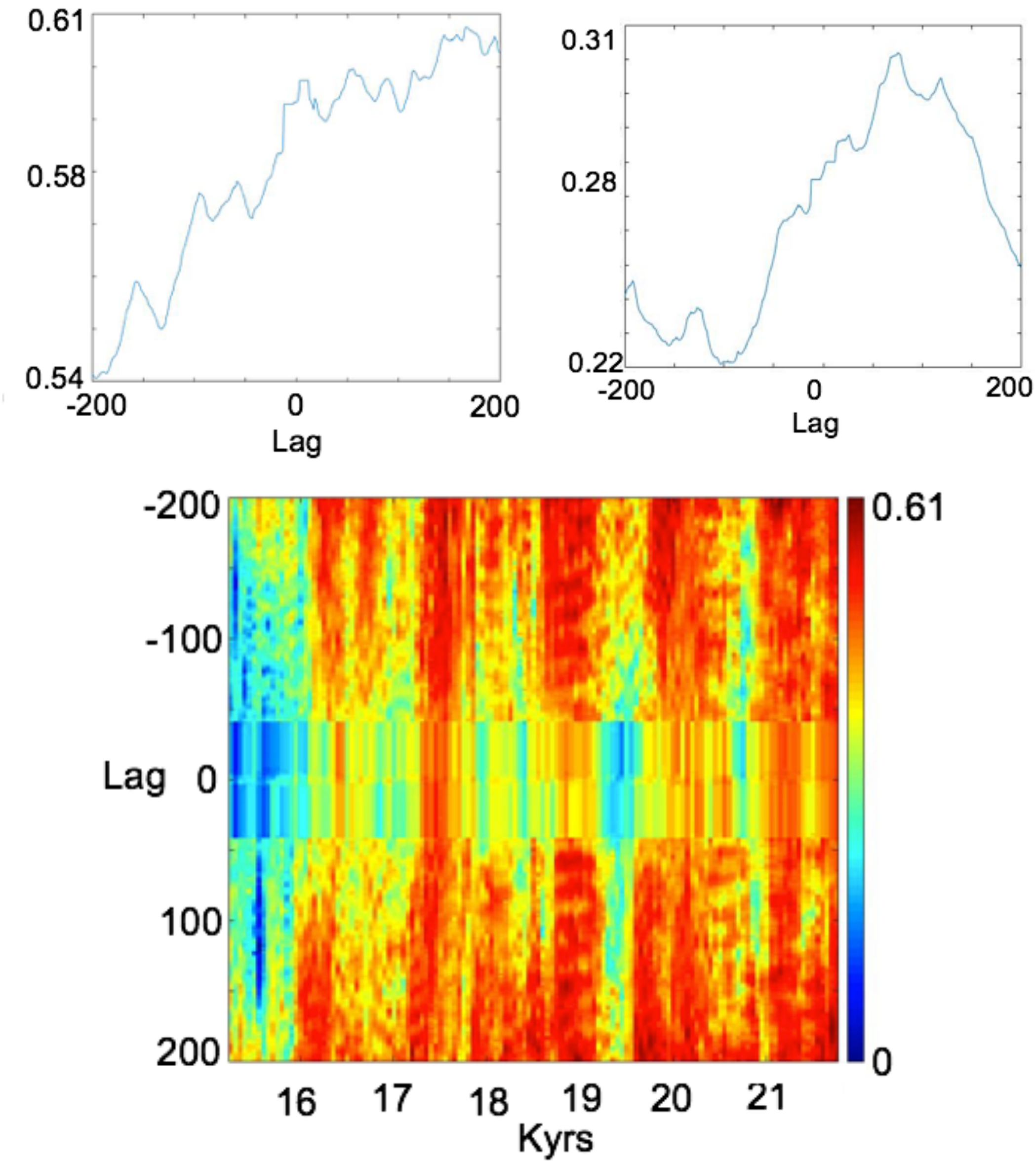

Fig 1. Top left) Mutual information and time delay mutual information for Antarctica and Colombia time series. Top right) Mutual information and time delay mutual information for Greenland and Colombia time series. Bottom) Heat map showing the time delay mutual information distributed on time between Colombia and Antarctica.

Values of entropy transfer only show significant estimates from tropical to subtropical locations. In both experiments, after the bootstrapping, Shannon entropy transfer results point to an information flow from Colombia to the extra-tropics.

X=Antarctica Temperatures Y=ITCZ rainfall

```
Shannon Transfer Entropy Results:
Direction  TE     Eff. TE  Std.Err.  p-value  sig
X->Y    0.0001  0.0000  0.0004  1.0000
Y->X    0.0017  0.0008  0.0003  0.0100   *
Bootstrapped TE Quantiles (100 replications):
Direction   0%     25%     50%     75%     100%
 X->Y  0.0000  0.0001  0.0002  0.0004  0.0021
 Y->X  0.0002  0.0005  0.0007  0.0011
Number of Observations:
p-values: < 0.001 '***', < 0.01 '**', < 0.05 '*', < 0.1 '.'
```

X=Greenland Temperatures Y=ITCZ rainfall

```
Shannon Transfer Entropy Results:
Direction      TE     Eff. TE  Std.Err.  p-value
X->Y    0.0010  0.0005  0.0003  0.1100
Y->X    0.0013  0.0006  0.0003  0.0100    *
Bootstrapped TE Quantiles (100 replications):
Direction   0%     25%     50%     75%     100%
X->Y  0.0000  0.0001  0.0003  0.0005  0.0017
Y->X  0.0001  0.0003  0.0005  0.0007
Number of Observations: 7000
p-values: < 0.001 '***', < 0.01 '**', < 0.05 '*', < 0.1 '.'
```

Fig 2. Top) Entropy transfer between Antarctica and Colombia. Bottom) Entropy transfer between Greenland and Colombia.

DIT results were not included because discretization the time series always yielded to values of mutual information of 0. Future work will focus in optimizing the discretization procedure so highest values of mutual information can be found.

Conclusions

Mutual information and entropy transfer seem to be a promising tool to understand climate dynamics during the last glacial period. The sense of causation that some of these tools can provide is valuable to understand how the components of the climate system react to major periods of climate change, and most importantly to understand in what order they do it. In this project I found that contrary to the proxy-based paradigm that the Northern hemisphere controlled the changes in tropical precipitation during the last glacial period, the Southern hemisphere temperatures seem to have an important an central role for some intervals. Also, my estimates of entropy transfer show a different view of how was transmitted the climatic signal during the last glacial maximum. It is believed that this periods of climate change were originated in the Northern hemisphere but according to my entropy transfer estimates it was from the tropics to the extratropics. Efforts to calculate other variables form information theory that can help to support or challenge these findings will be part of future work.

Bibliography

Rehfeld, K., & Kurths, J. (2014). Similarity estimators for irregular and age-uncertain time series. *Climate of the Past*, *10*(1), 107-122.

James, R. G., Barnett, N., & Crutchfield, J. P. (2016). Information flows? A critique of transfer entropies. *Physical review letters*, *116*(23), 238701.

James, R. G., Ayala, B. D. M., Zakirov, B., & Crutchfield, J. P. (2018). Modes of information flow. *arXiv preprint arXiv:1808.06723*.

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., ... & Zscheischler, J. (2019). Inferring causation from time series in Earth system sciences. *Nature communications*, *10*(1), 1-13.

Behrendt, S., Dimpfl, T., Peter, F. J., & Zimmermann, D. J. Introduction to RTransferEntropy.

James, R. G., Ellison, C. J., & Crutchfield, J. P. (2018). ``dit``: a Python package for discrete information theory. *Journal of Open Source Software*, *3*(25), 738.

Schreiber, T. (2000). Measuring information transfer. *Physical review letters*, *85*(2), 461.

Ursino, M., Ricci, G., & Magosso, E. (2020). Transfer entropy as a measure of brain connectivity: A critical analysis with the help of neural mass models. *Frontiers in computational neuroscience*, *14*, 45.

Fairchild, I. J., & Baker, A. (2012). *Speleothem science: from process to past environments* (Vol. 3). John Wiley & Sons.