Clim. Past Discuss.,
https://doi.org/10.5194/cp-2020-136-RC2, 2021

Climate of the Past Discussions

Open Access

CPD

Interactive comment

Interactive comment on "A statistical approach to the phasing of atmospheric reorganization and sea ice retreat at the onset of Dansgaard-Oeschger events under rigorous treatment of uncertainties" *by* Keno Riechers and Niklas Boers

Anonymous Referee #2

Received and published: 4 January 2021

Summary

This is an interesting study that re-investigates the temporal order of abrupt changes in different Greenland ice core proxies leading up to Dansgaard-Oeschger events. Compared to a previous paper by Erhardt et al, the authors use different methods and an alternative interpretation of the timing differences as a random variable that can be different from event to event. With this, they arrive at a different conclusion, namely

Printer-friendly version

Discussion paper

that due to the uncertainties in estimating the onsets of the abrupt transitions, we cannot exclude that there was no systematic lag of the abrupt transitions in either of the Greenland proxies. The paper is quite heavy on the mathematical and methodological aspects. To be more appealing to CP readers, I hope with the following comments to encourage the authors to better motivate why their approach can indeed be beneficial in resolving the problem at hand, or similar problems that readers might encounter in their own research.

General Comments

1. The authors do not specifically discuss how their approach is different to Erhardt et al 2019. While the manuscript gives the impression that previous studies completely disregarded uncertainties and used expectation values, this is not true for Erhardt et al, who included the uncertainties in a rigorous manner in their own right. The difference lies in the interpretation of the estimated onsets as random variables. Erhardt et al consider the uncertain samples as measurements of the same fixed quantity (a fixed time lag of Ca and Na at DO onsets), which allows them to simply multiply the individual MCMC posteriors to obtain a single posterior distribution that represents the measurement uncertainty of the fixed time lag. In contrast, the present study interprets the time lag to be a random variable, varying in between DO events. Thus, they cannot multiply the individual posteriors. In the discussion, the authors contrast their approach with the results one would obtain when completely discarding the uncertainties in the onset determination. It is not very surprising that including uncertainty yields hypothesis tests which are no longer significant. Instead, it would be more relevant to highlight the contrasting results of their approach to Erhardt et al.

2. The authors speak of a rigorous propagation of uncertainties to p-values, among other things. To achieve this, they introduce probability distributions of the mean and other test statistics, which are formally represented using delta distributions arising from the empirical sample. In practice, these distributions are all computed by summing a random sample from the individual MCMC posteriors, which is essentially boot-

strapping if I understand correctly. Can the authors comment on how the results of this work would be different if they simply summed the individual MCMC posteriors (which would be equivalent to bootstrapping as well), and then looked at the arising distribution of the lag, determining the probability of a lag >=0? This would be much simpler and the obvious alternative approach to the multiplication of the posteriors by Erhardt et al. It would probably also give a non-significant result regarding a Ca2+ lead.

3. I am wondering in what way the hypothesis tests introduced in Sec. 3.6-3.7 are necessary and add to the results? When reading the manuscript, I was surprised that hypothesis tests were introduced after distributions for the sample or population mean had already been given, which would directly allow to test the hypothesis of a mean >=0?

4. Why did the authors restrict the analysis to Na and Ca, and omitted an analysis of the offsets of d18O and lambda? This would be an important consistency test, and might even be more relevant since lambda has a very direct meaning as a proxy, and d18O is the most important of all proxies.

Specific Comments

L15ff: I think the conclusions should be stated differently. In my interpretation, the analysis does not contradict a lead or lag. Rather, as a result of the large uncertainties in estimating the individual onset timings, it cannot be excluded that there are in fact no leads or lags. Similarly, on the grounds of the uncertainties in the individual event timings, there is not enough evidence to conclude that atmospheric reorganization systematically preceded sea ice retreat for all events.

L34-35: The interstadials lasted up to 10 millennia, with 1.5 millennia being the average.

L45-49: While I understand that this discussion is primarily about the abrupt onset of DO events, it might be worthwhile to mention that there is evidence that sea ice and atmospheric proxies already start to gradually change much earlier than d18O in stadials (see Sadatzki et al, Sci Adv 5, 2019 and Lohmann, GRL 46, 2019).

L130: Why does the transition detection fail for some events?

L150: How would this depend on the autocorrelation of the noise? The Na and Ca records might have different noise structure and thus there is the potential for systematic biases indeed.

Eq. 7: Maybe the authors can elaborate specifically in the text what this approximation does. Since the order of MCMC samples for each event is arbitrary, by associating the i-th MCMC sample for every event, it seems like it is just a random sampling of m=6000 points in the joint space. If this is the case, why not simply sample randomly in the first place, and why not choose many more than m=6000 points? Or is this rather done in order to simplify the notation?

L278ff: I am not sure why the authors say that they are only given "relative" data, since they estimate the onsets in the two proxies. Furthermore, since until this point the data was already given exclusively as onset timing differences, I wonder why it is necessary to introduce "paired samples" now? This is a bit confusing to the reader.

L348: I might have missed this somewhere, but how is the distribution of the test statistic under the null hypothesis constructed?

L396: Could the authors explain more explicitly why the sample and population mean distributions are so different, and what this means for the interpretation of the results? Which distribution should be preferred?

L439: "...might simply be a stochastic feature." This is bit unclear. Maybe it would be better to write "...might simply occur by chance due to the small sample size."

L441ff: Here the authors introduce another simple method to address the likelihood of a systematic Ca2+ lead. I think it would be more coherent if this were moved to the Results Section. Furthermore, I find the conclusions from this simple calculation too confident, and in spirit contradictory to the interpretation of the other results. The probability of obtaining exactly n=16 events with a Ca lead will always be very small when there is a relatively large measurement uncertainty of the individual lags (spanning both positive and negative values), even if all individual posteriors would be clearly centered at negative values. Just like the probability of flipping 16 out of 16 heads is still very low for a strongly biased coin. This does not allow one to contradict the hypothesis that all events would follow the same pattern with a Ca^{2+} lead. For the data at hand, all but two events show posterior distributions centered at a negative value. Maybe the authors could write instead that from the MCMC posteriors it seems unlikely that all DO events occurred with a preceding abrupt change in Ca^{2+}. However, this might merely reflect the fact that due to the large uncertainty in estimating the onsets, the individual MCMC posteriors have significant support for positive lags as well. Arguing like this would also be much more in line with the authors' earlier statements that they cannot infer an absence of causality from their non-significant tests.

Technical Comments

Figure 1 and Table 1: Just to be sure, can the authors confirm that the time scale they use really is years BP (before the year 1950 AD), and not years b2k (before the year 2000), which is what is commonly used in GICC05?

L18: ...holds true, the we conclude...

L98: Instead of "compression" rather say: ...due to the thinning of the annual layers in the core.

Figure 3: Maybe it would be good to choose a different color for the null hypothesis in panel b. Otherwise it gives the impression that it corresponds in some way to the blue distribution in panel a.

Eq. 14 and Eq. 16: I am wondering whether "u" in the second delta function should be replaced by the empirical mean sum(y_i)/n?

L235: Maybe "marginal distribution" would correspond better to the nature of this distribution?

Eq. 17: What is the notation u_j and s_j?

Eq. 20: I am unfamiliar with this definition of a p-value. Is the integration not simply over all phi < phi_0 (for a one-sided test)? The definition here could lead to rather strange results for very asymmetric and long-tailed distributions.

L387: missing delta in the sum.

Figure 4: Larger fonts and panels would be nice for better visibility.

Eq. 35: It would be good if the authors could point to where the individual probabilities in the product come from.

Interactive comment on Clim. Past Discuss., https://doi.org/10.5194/cp-2020-136, 2020.