Review for "Comparison of Holocene temperature reconstructions based on GISP2 multiple-gas-isotope measurements", Döring and Leuenberger, 2019, Climate of the Past Discussions

Summary:

Döring and Leuenberger present the first application of a new gas-based temperature inversion procedure to previously published data from GISP2 during the Holocene. This is an interesting and timely contribution to the problem of constraining site temperature from histories of inert gas isotopes that has been gaining traction as an alternative to the classic water isotope paleothermoter. The type of technique they describe is particularly useful as it promises a more subjective and quantitative treatment of uncertainty imparted by the inversion.

Overall, I was overwhelmed by the volume of information provided in the manuscript. I would challenge the authors to consider a significant reduction in length in a new submission. Considering that a separate, lengthy methods paper has already been published, I was slightly surprised to see the volume of additional information required to establish the robustness of their method.

The key questions I had after reading the paper are as follows. What information is crucial and to what accuracy must it be known (e.g. accumulation, bore-hole thermometry, water isotopes temperature reconstructions for initialization, etc.)? What are the knowns that go into the inversions to produce estimates of the unknowns? How over or under determined is the system?

Additionally, there's a thread that runs throughout the paper about the hazards of using 15Nexcess. Because of the length and wide range of topics, the arguments end up being very dispersed. I would have appreciated a more concise structure to this thread as it took me a number of reads to find all the salient points.

Because of the length I have decided not to make line-by-line suggestions. Rather, I've broken the paper down into sections and highlighted what my main "take-home messages" were as a reader and any key questions I had.

Abstract

I count five major points in the abstract.

1) Comparison of 15N, 40Ar 15Nex
2) Gas loss as an artefact in 40Ar
3) Comparison of two different densification models
4) Treatment of analytical uncertainty
5) Comparison of Buizert et al., Kobashi et al, with this study

I personally think the paper could be reduced to about five figures along these lines and a few tables; whereas it currently stands at 14 figures and 10 tables.

The use of ellipses (…) in the uncertainty on the temperature is unclear to me until you read deep in into the paper (section 4?)

1 Introduction

Please define 15N, 40Ar, and crucially 15Nex. Without stating 15Nex is a function of both 15N and 40Ar it is not clear to the reader that you are in some sense doing a joint inversion when you use 15Nex. Additionally, this would be great time to discuss want information you "lose" when using 15Nex (e.g. LID), which is currently buried deep in the section on signal-to-noise.

Please remove references to a general body of work as simply "Kobashi et al.," or "his work".

2.1 Firn-models and inversion algorithm

This is fine but it tends to re-hash the published methods paper. Also, given that the Goujan and Schwander models actually show small difference when fitting 15N and 40Ar, I'm not sure so much detail is needed about basic things like the different zones of densification and depth resolution. It would suffice to say you simply used the two classic firn models.

The point about no geothermal heat flux in the Schwander is an interesting aside – particularly if you were examining a shallow core (<1000 m) - but my take home was that it just imparted a constant offset in the absolute temperature. Given that we are more interested in changes in temperature relative to the modern (when we can actually measure it), I was a bit lost as to why this was so important.

2.2 Timescale, and necessary data

Why do the text and nearly all the figures refer to chronology as the "ice age" of GICC05A? Shouldn't it be the gas-age chronology for the gas-based data?

Accumulation-rate input data: the difference in these scenarios needs to be defined. Later on there are many references to the 50km, 100km, etc scenarios but no information about what they mean.

2.3 Model Spin-Up

I see divergences of maybe 0.01 K in the plot so either I've completely missed the point or I would suggest they are negligible and insignificant relative to the larger uncertainties you go on to treat.

3.1 Signal to noise analysis of isotope data

I wasn't sure what the point of this section was. If it is simply to "evaluate the suitability of the three different inversion targets (δ15N, δ40Ar, δ15Nexcess)" prior to inversion, it would be sufficient to state that these data have been used like this before.

On the other hand, the statements about 15Nexcess as losing information about LID is really interesting but needs more development.

3.2 Gas-isotope fitting results

The procedure for the 10 different fittings runs is not properly described. What is varied so that the 10 runs differ (analytical error, accumulation, etc.)? Also, 10 iterations is generally not a large enough suite to arrive at a statistically robust distribution. How sensitive are your mean and standard deviations to the number of runs? This is crucial information as it used to then to conclude the ultimate uncertainty in the inversion.

3.2.2 Final misfits:

My take home was that the mean solution was within the analytical uncertainty, which is not surprising, given that you'd developed this algorithm to fit the data. I think these could easily be incorporated into the above sections. A reordering of these lines of argument might help the readability here.

3.2.3 Boundary effect

I've struggled a bit to understand this section and why you have a boundary problem starting about 1000 years before the last data point. Perhaps a useful analogy would to consider a present day firn site. Why would we need information that's 1000 years "in the future" to understand the gas isotopes of today?

Otherwise, **y**our description of the competing effects makes it sound like your system is severely underdetermined and could easily lead to spurious fits elsewhere.

3.2.4 Influence of different accumulation-rate estimates .

My first impression from the figure is that the effect of accumulation in a well-known constrained core like GISP2 is small. Yet on the other hand I missed the specific effect of accumulation on the inversion because the different scenarios are insufficient described. My questions would be:

-What would erroneously high or erroneously low accumulation impart on the temperature inversion? Can you robustly define the frequency of centennially-scale variability if an inaccurately variable delta-age is constant stretching or squeezing your periodic signals? Is the temperature inversion an intractable problem if you don't know accumulation, for example in the deepest sections of the core where the thinning function dominates the annual layer thickness.

This section could be folded into the section about the overall reproducibility (e.g. combined with 3.2.1. and 3.2.2).

4.1 Solution comparison

This was the heart of the paper to me. However, I found little utility in the tables with correlation analysis. The figures are sufficient.

4.2 Bandpass analysis of isotope data

This section seems quite tangential or is either misplaced and should go into the raw data analysis.

4.3 Model comparison

Just a note on the figures that relates to the discussion of the absolute temperature: having chosen to plot all data as anomalies and thus nearly identical it was confusing to see the subplot with the 2 K temperature difference. I think it would be a safe to state early on that all your temperature estimates are accurate only relative to modern as thus will focus the discussion on anomalies.

4.5 Comparison of T(δ15N) with the reconstructions of Kobashi et al. 2017 and Buizert et al. 2018

When rewriting this section please consider the saying "a picture is worth a thousand words". This is the exciting part of the paper but at the moment it reads like description of the figures and tables. Please consider bringing in a broader yet brief discussion of Holocene climate and the implications of your results. Also, a single paragraph over two pages in excessive.