

Review of “High resolution measurements of carbon monoxide along a late Holocene Greenland ice core: evidence for in-situ production”

The manuscript by Fäin *et al.* presents first ice core CO measurements made with a laser instrument coupled to a continuous ice core melter. The main conclusion of the paper is that there is *in situ* (processes that occur in the ice matrix) production of CO in Greenland ice cores. The authors also suggest that ice core CO measurements, despite being altered *in situ*, may be used as a proxy for biomass burning. The argument being, the deposition of the precursor chemicals is linked to biomass burning much like atmospheric CO (especially in the preindustrial atmosphere) itself. I should say, I am not really clear whether they take this position for changes in the baseline levels or just the abrupt spikes.

Given that there are not many reported measurements of CO in ice cores, and that laser spectroscopy for CO analysis has only recently become available, it is to be expected that this manuscript would be heavy on experimental and analytical methods. One of the areas I struggled with was figuring out how the calibration was accomplished and how issues like instrument stability were addressed. The relevant information is mostly there but scattered between multiple sections in the main text and the supplement. I consider this to be an important problem for a methods heavy paper (specific comments below). Some reorganization is necessary, although it was too complicated a task for me to provide specific suggestions about what to move where.

I had similar issues with the results and discussion. There is a section in the beginning about the interpretation of some of the long-term trends in the data as a biomass record and then focus shifts entirely to exploration of *in situ* CO production: abrupt CO spikes, drill fluid effects, pyrogenic aerosols etc. I struggled with this logical flow. There is a lot of interest in recovering a CO atmospheric history from ice cores, although the results presented in this paper suggest this may not be possible in the case of the Greenland ice cores. The authors suggest the measurements can potentially be used as a biomass burning proxy. In my opinion, the manuscript would be better served if the discussions ended with what the long term trends in this record, along with the prior Eurocore and firn air measurements, might be telling us about biomass burning. I should add, I am not convinced that *in situ* production is the only plausible explanation for all the unrealistic features observed in the record (see detailed comments).

In short, although I find the data and the subject matter of this paper interesting, I have many questions and recommendations for improvements. Major changes seem necessary before it should be accepted for publication.

Detailed comments

Calibration of the system:

It is stated on pages 2827 and 2828 that “An ideal calibration curve plotting full loop measurements against those on dry gas would have a slope reflecting dissolution losses and an offset reflecting the degree of contamination from the system.” I agree

with this statement and do not understand why the system behavior is not better explored within this context, given the rather odd calibration results (Table 1).

They calculate an average difference between measured and known values of the STDs (52+-26 ppb) and apply this as a fixed blank correction, implicitly assuming that the standard reproducibility issue is a bias (blank) problem, which is constant within a range of variability. First, it appears to me that the 52 ppb number is weighed towards the STD1 results, by virtue of the fact that STD1 was analyzed 19 times vs. only 4 times for STD2. What is done here amounts to a one-point calibration near the more dilute 50 ppb std when even the baseline CO trace is in 100-300 ppb range (spikes measure up to 1 ppm).

What I see in the full-loop results shown in Table 1 can also be explained by a combination an initial dissolution loss on the order of 40-50%, followed by a contamination on the order of 75-80 ppb. I'm not suggesting this is exactly what is happening in their system (hard to conclude with just two points) but it is a possibility. There is a big difference between subtracting 52 ppb from raw data instead of assuming 50% of the signal was lost before 80 ppb was added to it. If this were the right method to use, most measurements may be underestimates. Now, is 50% a reasonable number for loss to dissolution? I don't know. Dissolution loss is 7% for CH₄. What would we expect it to be for CO, assuming solubility equilibrium?

I do not fully understand (and not agree with the parts that I understand) the statements on lines 1-18 pg. 2828:

- Any two points will yield a perfect linear line, no?
- If there were multiple calibration points, one can estimate the dissolution loss with some assumptions about where it is happening. As I mentioned above, they could easily be underestimating the CO levels, especially the high-end numbers (also says so on ln.4 pg.2840, contradicting these earlier statements).
- How do we define a reasonable slope, and if the slope is unreasonable, should we not explore why?

I would say the right thing to do is to add more points to the calibration curve and push its upper end to 500 ppb or higher (depending on what the outcome is for richer standards and ideally covering the entire measurement range). If they don't want to make assumptions about when the loss and production takes places with respect to each other (i.e. which correction to make first), they can interpolate from the nearest standard runs to determine how much correction each measurement requires (better than extrapolating from 50 ppb). There is verbiage in section 2.3.1 about regular flow of synthetic air-water mixture through the system in full-loop mode (the loop explanations come later but I presumed). Was this also done with a zero (no CO) tank to independently establish the blank? I cannot tell from the text. What were the blank levels observed in those experiments?

I acknowledge that the precision and the reproducibility of the system are good. I suspect this is why the authors chose not to address the calibration issue more thoroughly. I disagree with this logic. One can surely live with less than perfect

accuracy, however, the issue is as much about the internal consistency as accuracy. Additionally, I think that one of the core arguments in the paper (the CO spikes being produced *in situ*) can only be proven with comparisons between well calibrated data sets (see my comments below). Calibration issues also hurt the discussions on the trends and comparisons with firn air and Eurocore measurements.

In situ CO production:

I agree that there is strong evidence for *in situ* enhancement of CO in Greenland ice cores and all measurements may be elevated above atmospheric levels to a certain degree. However, can we single out *in situ* production as the only plausible explanation for the CO spikes with the given data, or even the trends in the baseline? Can we rule out the possibility that there is CO production during the melting and extraction from precursors that exist in the ice? I don't think replicate core reproducibility alone is sufficient for yes or no answers.

The comparison of the baseline with the Eurocore data is suggestive of a common long term trend, I agree. But the truth is, there is no reason to leave out the spikes from this comparison (i.e. no reason to compare with a baseline line instead of a running mean) unless we are assuming the *in situ* mechanisms that lead to CO spikes are specific to the NEEM core. Even without the spikes, CO measurements from NEEM are somewhat higher than the Eurocore ones (some caveat should be provided where ever they say the results from the two cores agree because I think they are only referring to general trends). This is true even during 1700-1800 when the effects of *in situ* production should be very low. Is there any independent evidence (ice chemistry) that can be used to show we should expect stronger *in situ* production of CO at the NEEM site than at Summit? It could be very useful to measure CO in discrete samples from NEEM (same analysis method as Eurocore). A limited number of discrete samples should be enough to see if discrete samples from NEEM would average over the continuous flow measurements like they should. I presume the capability exists given the measurements in Alpine ice cores. Obviously, site-to-site or continuous-discrete comparisons require some confidence that the measurements are on the same (or close enough) calibration scales.

Drill fluid contamination:

I'm not sure how close to the rind these ice core sticks were but it would be reasonable to expect some level of drill fluid contamination on outer layers of fluid-drilled cores. There may also be memory effects. I suspect the water solubility of the drill-fluid is fairly low and it could take a long flush to clean it all out. In a continuous analysis, this may lead to artifactual trends after a contamination event.

Chemical processes that drive *in situ* CO production:

This is a difficult section to follow, especially the last three paragraphs (pgs. 2836-2837). There are many issues being addressed all at once and it does not work for me. The firn or ice question (essentially where the enhancement occurs) and

related discussions are scattered and confusing. If this is a firm process near lock-in, why would the firm data make sense? Is this not a contradiction with conclusions of earlier firm works by Petrenko et al. (2012) and Wang et al. (2012)? Is it possible to quantify the arguments for continual production based on the depth dependence of spike amplitudes, may be with a separate figure of deviation from the median vs. depth? I can see that they are more intense between years 400-800 but the get smaller again on the deepest layers.

Figure 4 needs tick labels for the upper x-axis since tick marks do not line up and the axis breaks are at a different positions. I cannot really follow the discussions about which peaks line up and which do not. The figure also needs a better caption: which trace belongs to which axis etc.

Figure 5 axes should be reversed, CO is the dependent variable here. The answer (based on figure 5) is it is not clear (no statistically significant correlation?). After that, things gets nearly incomprehensible. Figure 6 and all the discussions centered around it do not provide conclusive evidence on what is going on, which makes me question whether figure 6 should be shown in the paper at all. Does it make sense to focus on one peak unless what we learn from it will lead us to the promised land? I would argue the real advantage of the higher resolution measurements is that it provides many instances of abrupt changes and increases the chances of finding a statistically significant correlation between the spikes (all of them) and one of the biomass burning markers. Another thing that puzzles me is why there is hardly any focus on changes in baseline levels in this section? Would long term trends in pyrogenic aerosols not influence the long term trends in CO?

There is a misplaced reference to figure 5 on ln.26 pg. 2836.

CO as an integrated biomass burning signal (atmospheric CO + pyrogenic aerosols):

There are two complications I can think of right of the bet. One is the obvious ice age – gas age issue for the atmospheric signal vs. the ice side measurements. The other one is the more obscure footprint issue. I understand that this is a somewhat speculative suggestion at this point but like I mentioned in the beginning of my comments, the biomass burning perspective is likely to draw the most interest from the readers of this journal. I would have liked to see a more together discussion. In the conclusions (lns. 6-23 pg. 2842), right after they suggest CO could be a biomass burning proxy, may be they could provide some brief comments.

On ln.14 they mean “baseline” when they say “not spiked”, correct?

Comparison with an alpine core:

There is not a real comparison here, at least not in the sense that CO measurements from Greenland are compared with measurements in Alpine cores. Similar to figure 6, I do not see a strong piece of evidence that provides some sort of closure on the *in situ* production question. I would suggest summarizing the winter and summer layer differences and how that might be relevant for the NEEM results in a few sentences (within an already existing section) and omit this section.