Draft Response to ACPD NEEM firn CO paper reviews

Reviewer comments in blue. Responses in regular black font.

Reviewer 1 – Dr. Maarten Krol

This article presents and analyses firn CO mixing ratios from the recent NEEM ice core site in Greenland. The most important conclusion from the paper is that the peak CO was in the 1970s or early 1980s. This finding is in conflict with bottom-up emission estimates, which predict highest emission (and hence concentrations, although this relation depends on other factors) later in the 1980s. Given the firn mixing ratios, the authors claim that and emission decline should have started already in the late 1970s. They attribute this to reduced emissions from road transportation, in line with an analysis of CO isotopic signals in firn, as outlined in an accompanying paper by Wang et al..
Besides, it is concluded that the 1950 CO concentration was already higher than the current concentration. This latter result is surprising, since the production of CO from CH4 oxidation is considerably larger nowadays. If is assumed that emissions of CO scale with population (use of biofuels and fossil fuels) the emissions probably also in- creased, due to the large increases in population. However, technology of combustion plays an important role in the CO/CO2 emission ratio. Moreover, the shift in CO emis- sions from NH mid-lattitudes to more sub-tropical latitudes (India, SE Asia) plays also an important role. This factor is hardly quantified and discussed and this aspect should be improved in the final manuscript. The rest of the analysis in the paper is mostly sound and is back-upped by analysis of other gases H2, CH4, and VOCs. I have some comments and questions that should be addressed and are aimed at improving an already interesting paper.

We would like to thank Dr. Krol for his thorough and constructive review.

1. Major Comments

As discussed above, major CO sources probably decreased in Europe and North America since the 1970s, but increases are expected in SE Asia and India, due to large industrial developments since the 2000s. Since these sources are located fur- ther south, they are prone to faster oxidation by OH and might escape NH wintertime accumulation in the OH-void high northern latitudes and thus accumulation in firn air. This important caveat in the analysis of firn is hardly quantified and discussed in the paper and therefore deserves more attention. Given a lifetime of two month for CO and transport times from low NH latitudes to Greenland, it is obvious that the latitude of emission cannot be ignored.

We agree with Dr. Krol that the regional and latitudinal distribution of CO sources has changed over the last few decades. This is in fact already evident in our manuscript in Figure 8, which shows the RETRO and EDGAR-HYDE emissions inventories presented in a few different combinations of regions and latitude bands. We further agree with Dr. Krol that a shift of CO emissions to lower latitudes is likely to result in an overall decline of CO mole fractions over Greenland, owing to a larger fraction of CO being removed by

OH during transport. However, this latitudinal shift of CO emissions cannot explain the earlier (1970s) CO peak that we infer from firn measurements as compared to a later (~1990) CO peak expected from emission inventories. This is illustrated very clearly by the results from the CAM historical run (Figure 9), which uses the ACCMIP emission inventory (ACCMIP is based on a combination of EDGAR-HYDE and RETRO), accounts for latitudinal CO source changes and associated changes in what fraction of emitted CO actually reaches Greenland. The CAM historical run still predicts peak CO over the NEEM location at ~1990.

We are concerned that it might be possible that Dr. Krol was for some reason not able to see the results of the CAM historical run on figure 9, based on his comment near the end of his review about not being able to find the brown line.

All that said, we agree that effects of latitudinal shifts in CO sources should get more attention in the manuscript and will add further discussion of this to sections 5.1 and 5.7.

The most surprising conclusion of the paper is the higher 1950 high latitude CO concentration, compared to the present day concentration. Since (i) the CO production from CH4 must have increased, (ii) CO production from natural VOCs probably did not decrease, (iii) the use of bio and fossil fuel increased (albeit with better technology, at least at high NH latitudes), this calls for additional discussion and quantification. The interpretation of firn air remains a delicate exercise and, as recognized by the authors, in situ CO production in the firn close the the lock-in zone cannot be totally excluded.

We agree that the finding regarding CO mole fractions over Greenland in 1950 as compared to today is surprising, but this finding is robust. Every single one of the final set of 61 successful CO history scenarios shows this, regardless of the firn gas transport model (INSTAAR or LGGE-GIPSA) or site or combination of sites used for the reconstruction. Further, as we demonstrated in detail in Section 3.3 and associated figures and supplementary material, CO is well preserved in firn air at inland Greenland sites. Gas movement in the lock-in zone of the firn (where the interesting, older part of the CO signal is contained) is very limited (see e.g., Buizert et al., 2012, ACP), and it is extremely unlikely that CO records from 3 different sites with somewhat different snow accumulation rates, temperatures and trace organic loading would agree as well as they do if there was significant in situ CO production. This is already discussed in detail is Section 3.3 (main text and supplement). Our estimate of possible 5 ppb in situ CO contribution in deep firn is a conservative upper limit, as Reviewer 2 recognized. If there is indeed as much as 5 ppb of in situ CO present in the deepest samples, it still does not change the conclusion that CO in 1950 was at least slightly higher than today.

Some quantitative discussion of estimates of CO sources in 1950 as compared to today is already provided in the second paragraph of section 5.7. We will expand this discussion to include the natural VOC oxidation source of CO. We will further consolidate the discussion of 1950 vs today emissions and results from the CAM historical run into the same paragraph, to make this more clear for the reader. We will also add a statement to this paragraph (already in the Conclusions section) that our reconstruction is very difficult to reconcile with existing emissions inventories (unless unrealistically large changes in OH are considered). Unfortunately, we are not able to include a more quantitative discussion regarding which types of sources are driving CO trends over Greenland in the CAM historical run, because CO "tagging" is not available in this run.

The inverse method, described in section 4.2, states that the method used is similar, but also slightly different than Rommelaere et al. (1997). One of the differences is in the aim to match the experimental error estimate (page 19008, line 26). The authors should show, e.g. in the Appendix, how the methods differ exactly. Now they use a minimum of the sum of two error terms as is done in Rommelaere et al. (1997). So the difference between the methods is not totally clear at the moment.

This point will be clarified in the revised manuscript and supplement. In the case of gases (isotope pairs were not included in Rommelaere et al., 1997), the only important difference with the original method is the choice of the optimal solution. Rommelaere et al. (1997) adjusted the regularization term using a chi-square test based on the a-priori experimental error (Rommelaere et al., 1997, last sentence of section 5.2). An important limitation of this method is that the a-priori uncertainty is difficult to assess precisely especially for multi-site simulations which are sensitive to differences between single site data series (e.g. inter-calibration biases or geographic variations). Thus an alternative method was developed, based on a compromise between two model diagnostics already available in the Rommelaere et al. (1997) model and described at page 19008 lines 19-24 of our Atmos. Chem. Phys. Discuss. paper. As shown on Figure 11 of Rommelaere et al. (1997), the choice of what is interpreted as noise or signal in the firn data leads to drastically different reconstructed scenarios in terms of smoothness. Tests performed on multiple species and site combinations (based on results in Witrant et al., 2012) showed that the new method systematically leads to reasonable scenario smoothness whereas the original method by Rommelaere et al. (1997) does not. Since the submission of our Atmos. Chem. Phys. Discuss. paper, we made new progress on the choice of the optimal solution, taking advantage of a recent mathematical study (Lukas, 2008). This new robustness-oriented definition has been implemented in our model (Witrant and Martinerie, 2013). The new results are consistent with our previous range of scenarios (see Figure 1 at end of this document) and will be shown in our revised supplement. They are further discussed below (answers to next two comments).

Furthermore, some arbitrary decisions are taken later in the manuscript. First, at page 19009, line 22, an exemption is made to the "all data points must be matched"-rule. Even if the excluded point is caused by the 1998-1999 biomass burning event, the inverted atmospheric concentration scenario should resolve this event, since the signal is recorded in the firn.

There were several reasons besides the possible effect of the 1998-99 biomass burning anomaly that led us to make an exception for the 42.6 m NGRIP data point. First, we note that all data points above 40 meters depth were excluded due to their sensitivity to seasonality (as explained p 19008 lines 5-10 and p19009 line 10). We further note that this 40 m threshold is very close to the 42.6 m depth of the NGRIP data point in question.

The scenarios which would be excluded by taking into account the 42.6 m data point are shown in grey on Figure 1 (at end of this document). Excluding these scenarios would not significantly affect our overall scenario envelope, as other very similar scenarios are not excluded (in black). On the other hand, taking into account the NGRIP data point at 42.6 meters depth would exclude most of the two-site scenarios based on NEEM-EU + NEEM-US data. As overall the firn at NEEM is probably the best characterized of all sites (more species, replicate measurements from more labs and overall more recent, higher-quality analyses), we decided to make the exception for this data point.

The “arbitrary” aspect of the two decisions pointed out by Dr. Krol (see also answer to next comment) partly stems from the fact that assessing the multi-site consistency of firn signals (at least for more than 2 sites) is at a very early stage of development, and no established protocols exist. Our approach to combined model-data uncertainty estimation (page 19009, line 13) may underestimate forward model uncertainties (imperfect model physics) arising from applying a model to different sites, as well as from applying two different models. The inverse model also neglects geographic CO variations within Greenland, which may be significant.

We will add some further detail to the manuscript with regard to the above points to better explain all the reasons for making an exception for the 42.6m data point.

Second, at page 19010 some scenarios are claimed to be visibly too irregular and are excluded from the analysis while they pass the data-model comparison test. So either the error criteria chosen wrongly and retuning of the method is required, or all the scenarios that pass the test should be included. Now, some unwanted arbitrariness is introduced, and repeatability is not guaranteed.

The smoothness of a scenario reconstructed from firn data results from both the physical and mathematical nature of the problem:

(1) The deep firn imposes a large amount of smoothing on the atmospheric signal, with the degree of smoothing increasing with depth. For example, at the start of the lock-in zone at NEEM, the width at half-height for the CO gas age distribution is already ~8 years, and it is ~35 years at the deepest sampled level (Figure S4; see also Figures 7 and 8 and Table 4 in Buizert et al, 2012, ACP for similar analysis for CO_2).

(2) The mathematical inverse problem best estimate smoothness is related to what the model considers as noise and as signal in the firn records. This smoothness thus depends on the quality of the data, but also on the model quality (capacity to fit the data well) and assumptions (e.g. the atmospheric trend scenario is the same at all modeled sites).

Atmospheric history scenarios with low smoothing (a large amount of temporal variability) can be generated that match the firn data better when run through our two forward models than smoother (more robust) scenarios. We cannot rule such scenarios out for CO, as a large amount of interannual variability is likely given the relatively short lifetime of CO. However, the firn does not contain information of sufficient temporal resolution to validate the higher frequency components of such scenarios.

Our two-step approach thus aims at better exploring the interplay of non-modelled uncertainties and mathematical under-determination of the problem by testing reasonably smooth scenarios through all available datasets and models. We recognize that finding an objective criterion to eliminate irregular scenarios that fit the data well is difficult. Another article in the ACPD special issue (Trudinger et al., 2012) discusses similar issues of discriminating between the noise and signal in the data used for firn diffusivity calculation. At the time of manuscript submission, we suspected that our optimal solution was leading to solutions that were too irregular; we therefore only used smoother scenarios than this optimum. The new development of a robustness-oriented solution (Witrant and Martinerie, 2013) has confirmed this. Importantly, the model-calculated uncertainty on the new robust scenarios constrained with three sites (dashed green lines on Figure 1) are similar to the range of scenarios from our two-step approach. This confirms the overall magnitude of the uncertainties on the reconstructed CO trends.

As mentioned above, Figure 1 and some discussion of the new Witrant and Martinerie (2013) approach will be added to the revised Supplement.

Another note about the inverse modeling concerns the prior information that is used. In the source inversion world, a different kind of of prior information is used. Starting from a "best-guess" scenario of the emissions, improvements are sought that better fit the available atmospheric measurements. In In the method described by Rommelaere et al. (1997), no information about the atmospheric concentrations are used. Instead, a smoothness constraint is used, that requests that the atmospheric concentration sce- nario is not too noisy. I would argue that, especially after 1980, good prior information is available about the atmospheric concentrations over Greenland (seasonal cycle, and special events, such as the 1998 biomass burning). In fact, this is illustrated perfectly in figure S2. The authors should argue more clearly why this information is not used in their reconstruction, or better, present a results in which this valuable prior information is used.

Direct atmospheric measurements were not used as a “prior” in the inverse model because the LGGE-GIPSA model is not yet configured to use such a constraint. We agree with the reviewer that this would be a valuable improvement, involving the addition of an extra term in the cost function and subsequent calculations. However, this would entail a major modification of the model that we consider to be beyond the scope of the current work.

An important point in the method used in our manuscript is that a unique solution is directly obtained for each chosen value of the regularization factor without iterations to progressively minimize the cost function (Eq. 35 in Rommelaere et al., 1997). It has the advantage to be computationally cheap but including atmospheric constraints would require to re-formulate the model solution to include the effect of the modified cost function. Very recent work on the optimal solution and isotopic ratios (Witrant and Martinerie, 2013) provides the basis for integrating constraints from atmospheric data in the model. The cost function (Eq. 12 in Witrant and Martinerie, 2013) could be updated with an extra term involving the weighted squared error between the known atmospheric trend and the calculated one (u), integrated over the time interval when measurements are available, and the analytical solution (Eq. 16 in Witrant and Martinerie, 2013) updated accordingly. The major additional difficulty would be to properly weight this extra term and how it affects the expected regularization factor, as more knowledge on this would implicitly be added in the inversion strategy. Thus a multi-site/multi-gas analysis should be carried out to validate such change in the method.

We further note that both our inverse scenarios and atmospheric measurements from Arctic stations match the firn data similarly well when run through the forward models (Figures S2 and S5; see revised version of Figure S2 that includes Barrow OGI data at bottom of this document). This suggests that our inverse scenarios overall would not change significantly if the direct atmospheric measurement constraint was added to the firn data constraint.

We will add a brief version of this discussion to the revised manuscript to clarify why direct atmospheric measurements were not used as a prior.

A final point about the inverse method concerns the error estimate. I think it is extremely important that for the problem at hand an error estimate in the exact chronology is provided. From figure 7 it is clear that some scenarios show CO mixing ratio maxima around 1970, while in other scenarios the peak is shifted towards 1980. The timing error will depend on time and will probably increase for older air. Discussion about the time-error estimate is therefore vital for the interpretation of the data.

We estimate the age uncertainty of the atmospheric CO peak in the smoothed CO history we reconstructed by considering the full range of peak dates provided by all successful scenarios. Sampling and analytical uncertainties as well as uncertainties arising from firn gas transport models are already reflected in this range of peak dates, because we use data from multiple labs, 3 different drilling sites and employ two different firn gas models. This is already discussed to some extent in the second paragraph of Section 5.1, and we will clarify this further in the revised manuscript, as well as identify the limitations (e.g., our scenarios reconstruct a single smoothed peak and would not capture a possible more complex double-peak CO history accurately).

At the start of section 5.2, the CO record is compared to the d[CH4]/dt record. The authors argue that CH4 would react slowly to OH changes, but that the methane growth rate is more sensitive to OH. In box-formula form:

$dCH4/dt = E - k.OH.CH4$

Most people believe that the leveling-off towards 2000 of the methane growth-rate is due to an approach to steady state, e.g. due to stabilizing emissions.The role of OH changes is hard to quantify and the shape of the methane growth-rate curve can be explained perfectly by assuming constant OH. The correspondence between the re- constructed CO mixing ratios and the methane growth rate curve appears to me co- incidental and I see no physical arguments to show both curves together. Of course, increasing OH cannot be excluded as a possible cause for both for both CO decline and decline methane growth rate, but the analysis should then include (i) a model in which the impact of the OH change is calculated on both CO and CH4, and (ii) realistic scenarios of the methane (and CO) emission histories (see equation above).

One point that we have tried to make through the discussion (and will try to do so in a more organized way in the revised manuscript, as per suggestions from reviewer 2) is that there are currently no data sets that allow to decisively rule out significant changes in OH in the decades prior to about 1985. We are not arguing that changing OH is the explanation for observed changes in trace gas trends (we think emissions are the likely main driver of both CO and CH4 trends), but rather we are highlighting (with the use of Figure 7) that it cannot be ignored as one of the possibilities.

In response to Dr. Krol's comment, we will add a brief discussion to the effect that most literature has indeed attributed the leveling-off of the CH4 growth rate to stabilizing emissions. The comparison of dCH4/dt with CO is still valuable, however, in the sense that it is consistent with a significant role for OH and highlights uncertainties in the conclusions about both the CH4 and CO budgets. We also infer from the comments of Reviewer 2 that this reviewer considers the dCH4/dt – CO comparison valuable.

The analysis that Dr. Krol suggests for the CO – CH4 comparison would indeed be valuable, but to do it properly is beyond the scope of this paper. A proper analysis of this kind would explore a number of possible CH4 and CO emission scenarios. While a 2-box or 3-box model may be sufficiently informative for CH4, for CO a chemistry-transport model is needed because of CO's relatively short atmospheric lifetime. As we stated near the end of Section 5.1, a model capable of such repeated historical runs is currently not available to us.

The main strength and novelty of this study is really the multi-site Greenland firn air CO record itself (first of its kind), and the demonstration that it is robust (i.e., not significantly affected by alteration in the firn). Full exploration of the interesting implications of this record for emissions inventories and possibly for atmospheric chemistry processes would require repeated CTM historical runs and could form the basis for a separate study, which we sincerely hope the atmospheric modeling community will undertake.

2 Minor Comments

On page 18997 the OGI record is discussed (figure 1). This record should have peaked in 1984. However, this notion seems to be based on the wintertime values only. I suggest to plot the data (both OGI and NOAA) also as 12-month running means (NOAA provides great tools for this) to avoid the visual dominance by the wintertime maxima.

We will add 12-month means to the Figure 1 plot.

In section 2, the sampling and analytical methods are described in (too much) detail. Yet, more information is given in the appendix. I suggest to move more text to the appendix.

We respectfully disagree. For papers reporting analytical data, the expectation is that if a method has been published, a short summary plus a reference is given. For unpublished methods (NOAA NEEM measurements, for example) and modifications to previously published methods, the new details need to be provided. This is exactly what has been done, and there is quite a lot to describe given the number of field and laboratory methodologies involved. The methods are a key and necessary foundation for this data set and therefore have a place in the main body of the paper. We further note that Reviewer 2 commented that the Methods section was well written.

Furthermore, the claimed agreement in figure 3 still contains the know cali- bration offsets between the different labs. Since these offsets are known, I do not see any reason not to apply them in figure 3 to show better the remaining discrepancies after correction.

While it is generally known that calibration offsets for both CO and H2 exist between the different labs, these offsets are not well characterized. This is already stated in Section 3.1 of the Supplement. For combining the data sets (to generate a single data set for each borehole that is used for atmospheric history reconstructions) we utilized calibration offsets as determined from the data itself, by comparing values for flasks filled simultaneously at the same depth levels but analyzed in different labs. This is already explained in Section 3.1 of the Supplement.

Part of the value of Figure 3 in its present form is that it allows the reader to evaluate the data as measured, without any processing, as well as to see the agreement in the general features of the different labs' curves prior to calibration scale adjustments. We note that reviewer 2 commented that this entire section is well written as well as that "The overall agreement between the CO data from five different groups is pretty remarkable, considering the different sampling methods and equipment used at three different firn sites, and the potential calibration offsets between different laboratories". Taking this into consideration, our preference is to keep Figure 3 in its current form.

Also, the discussion about data-selection on page 19003 is rather vague and the selected procedure seems rather arbitrary.

A detailed discussion of why the UEA and Stony Brook data were excluded from the final combined data sets is already provided in the Supplement (Section 3.1). We will add a more explicit statement pointing the reader to the Supplement. We note that excluding a single data set because it has larger analytical uncertainties than other data sets is scientifically justifiable when multiple other data sets are already available (in our case, there are 4 sets of CO measurements available for the NEEM EU borehole).

Maybe it can be shown in an Appendix what the consequences are of including the UEA and Stony Brook CO data.

The UEA data are not suitable for a long historical reconstruction because the S4 borehole did not sample the lock-in zone (where the old air is contained). This is already stated in the Supplement (Section 3.1)

The effect of including the Stony Brook data is very minor, is not expected to affect the inverse scenarios significantly, and will be shown in a revised figure S5.

Now the inclusion of these data in the paper seems driven by politeness rather than by scientific considerations.

The inclusion of these data adds to the overall confidence of the NEEM CO record, as it allows the confirmation of the firn CO trends by more laboratories. This is already discussed in section 3.1. Also, we infer from Reviewer 2's comment about the agreement between data sets from different labs being " pretty remarkable" that this reviewer appreciated the inclusion of data from all the labs.

On page 19011 the OGI record is discussed. I wonder how the OGI record (corrected for offsets) would translate in the firn history (like figure S2).

We will include the OGI record in the forward model runs shown in Figure S2, as the reviewer recommends. A preliminary version of this figure appears at the bottom of this document. The OGI record appears very consistent with the firn air data, as the revised Figure S2 illustrates. Because of the importance of Figure S2 in illustrating that the firn air really does accurately capture the [CO] atmospheric history, this figure will be moved to the main text in the revised manuscript.
The Barrow OGI record (adjusted to NOAA scale using the period of overlap) is also shown in Figure 1 at the end of this document; it appears overall consistent with our reconstructions.

In section 5.1, page 19011 at the bottom, it is said that no model is available to explore the implications of the reconstructed history. However, some kind of analysis as presented in the accompanying Wang et al. (2012) paper (Table 1) is needed in this paper also (see major comments).

We will expand the discussion of the results from the Wang et al (2012) isotopic mass balance analyses with respect to their findings about changes in individual CO sources. Unfortunately, the modeler who performed the MOZART runs for the Wang et al (2012) paper (Key Hong Park) and provided model estimates of different CO source contributions at Iceland is no longer part of the Stony Brook group and is not available to contribute to this manuscript. We note that while the CAM historical run results we present do not allow for tagging of individual CO sources, in some ways they are more applicable than the MOZART runs from the Wang et al paper because the CAM results are calculated for the exact location of NEEM rather than for Iceland.

In section 5.5 it is argued that an increase in OH would lead to an increase in H2 due to an increase in production from hydrocarbons. This seems to quickly argued, since part of the H2 is also oxidized by OH and NMVOCs quickly reach a new and lower concentration, feeding back negatively on the H2 production.

That is true; all other things being equal, if OH were to increase the NMHCs would be quickly pulled down to a lower steady-state, with the net effect on H2 source from NMHC oxidation being nil ([NMHC]x[OH] stays constant).

We will modify these sentences to say instead that for H2 (as compared to CO) an increase in OH would be expected to produce a smaller and delayed decline. The smaller response for H2 is due to the much lower fraction of H2 that is removed by OH (as compared to CO). The delayed response is due to the longer lifetime of H2, as well as to the fact that initially the CH4 oxidation source of H2 will increase in response to an OH increase, and may entirely cancel or even temporarily overwhelm the effect of the increased OH sink.

On page 19015 and 19016 the work of Dentener et al. (2003) is wrongly interpreted. In this work a full chemistry simulation was performed in which the methane concentrations we forced to follow the observations (mass-balance approach). The analysis of OH in this long term simulation showed a small long term positive trend.

We will correct and clarify the Dentener et al approach in the revised manuscript.

Methane variability is explored in Monzka et al. (2011), a reference that is missing (Montzka et al. (2011). Small interannual variability of global atmospheric hydroxyl. Science, 331(6013), 67–69. doi:10.1126/science.1197640).

Our main focus is pre-1990, while Montzka et al address OH changes mainly after this period. We will include their results in the discussion, however.

3 Textual Comments

page 18996, line 10: replace "atmospheric chemistry" by "atmospheric composition"
will do as suggested

page 19006, line 3: "discussed below": please state where exactly (next section?).
This should really point to the Supplement; we will include this

caption figure 9: Brown line not found.
This displays as brown for us (curve on lower half of plot) – perhaps a pdf reader issue?

Table S5: A more recent H2 budget estimate is provided by Pieterse et al., Atmos. Chem. Phys., 11, 7001–7026, 2011.
We will add information from this reference

Figure S2/S4/S6, caption: The blue line? I would call this purple.
Again, seems like a pdf reader issue – this shows as blue for us.

Reviewer 2

The manuscript presents new CO measurements in firn air from 3 different sites in Greenland and uses the fir air CO record to reconstruct atmospheric histories for the last 60 years. The manuscript also includes the first measurements of H2 in Greenland firn air that can help interpret the CO record. The CO atmospheric histories reveal an interesting trend: CO was considerably higher in 1950 than it is today and there has been a decline during the last 20-30 years of the 20th century. One important implication of these trends is that historical CO emissions appear to be seriously un- derestimated in the emission inventories.

The paper is very well written, especially sections 1-4. The overall agreement between the CO data from five different groups is pretty remarkable, considering the different sampling methods and equipment used at three different firn sites, and the potential calibration offsets between different laboratories. One fundamental scientific question is whether the main features of the firn air CO records (specifically the long term trends) reflect real atmospheric changes. The authors take a conservative approach and put heavy emphasis on this issue. They highlight a few instances of apparent measure- ments discrepancies and explore the possible explanations to conclude that there may be room for in situ production of about 5 ppb in the firn. I would argue ~5ppb is well within the range of uncertainties due, for example, to the potential differences and vari- ability in transport of polluted air to the different firn sites and the surface sites they are using in their comparisons. They also thoroughly explore the modeling uncertainties that arise from the signal smoothing in the firn. Overall, there is robust and convincing evidence that CO is well preserved in Greenland firn air and the atmospheric histories display real long-term atmospheric trends. This is a unique data set that will enhance our understanding of changes in fossil fuel use that results in CO emissions (and pos- sibly NMHCs) and may also provide some constraints on OH variability in the NH. In contrast, the interpretation of the H2 firn air measurements is very challenging and it is not yet possible to develop H2 atmospheric histories with much confidence.
The manuscript fits well within the scope of ACP and warrants publication, although I do have a few reservations that I would like to see addressed before it is accepted I its final form. I found some of the discussions to be overly qualitative and somewhat out of focus on what the authors think was driving the observed CO trends. For example, I had trouble getting a clear message from sections 5.2, 5.5, and 5.6. These three sections seem to be written in a way to leave an element of doubt about the role of OH. With that, I mean doubt about what the authors think the role of OH is, otherwise, I understand that there is a lot of unknowns about past OH variability. For example, it is very difficult for the reader to follow the train of thought that gets us to the conclusion on the role of OH based on H2 measurements: “Overall, the examined evidence does allow for a modest increase in OH during the 1980s to be part of the explanation for the [CO] trend.” (lns. 26-27, pg. 19018)

I also found the supplemental material to be well organized, appropriate, and helpful, except the section about H2 in NEEM firn air (details below).

We would like to thank this reviewer for their detailed and constructive review.

More specific comments

Introduction: Somewhere in the in the introduction there should be a justification of why Greenland is representative of high NH and a description of what is meant by high NH.

We will add a brief discussion of this, as well as point the reader to section 3.3 where this is discussed in more detail

I also think there should be a little more on the CO budget: A more quantitative analysis of the best estimate CO sources in the NH, for example.

We will point the reader to Table S5 (which presents a global CO budget) and add more detail on global versus NH CO sources.

Section 5.2: The correlation in fig. 7 is very strong and intriguing. The authors report a +-10 years age uncertainty for the dCH4/dt based on NOAA (flasks) and Law Dome (ice core). Can we not do better if we try to get an estimate of the relative uncertainty (we don't really care about the absolute uncertainty in this context) between the firn inversions alone?

Unfortunately, we can't reliably estimate the relative uncertainties on the peak timing of CO and dCH4/dt from inverse scenarios because the magnitudes of the non-modelled uncertainties are very different (e.g. geographic variations, seasonality effect). Further, because CH4 is a strong constraint on the firn diffusivities in the forward models, the model uncertainties for CO and CH4 are also of a different nature.

I agree with the authors that Fig. 7 is suggestive of a large role for OH in driving CO and dCH4/dt trends, although I cannot readily dismiss a connection (may be environmental policy related) on the source side very easily. An immediate question is: what would be the impact of large OH changes on methane emission estimates? I cannot think of an easy way to answer this (my guess is very large and unrealistic), not without getting into modeling, which the authors don't deem within the scope of this paper. However, I still did expect a qualitative discussion. What seems logical to me is to carry out that discussion for 1950-1970 and 1980-2000 periods when things are changing a lot. Instead, the focus here is on 1970s and 1980s (lns. 26-29 on pg. 19012 and continuing onto pg. 19013). I was quite puzzled by this at first, but after reading the whole manuscript it occurred to me that may be they are trying to set up the stage for later OH arguments based on the CO-H2 comparison? I have to say this is very confusing for the reader. Why not add a separate section about the role of OH to the very end of the discussion section (I'm assuming they feel strongly about getting a point across about a potential role for OH during 70s and 80s), after they are done with all the comparison sections.

Part of the issue, as the reviewer noted above, is that there are still large uncertainties regarding OH, particularly before 1990. While this study provides a valuable constraint via the atmospheric CO history, by itself this study does not resolve the OH questions. We will follow the reviewers suggestion and consolidate the discussion of possible changes in OH at the end of the discussion part of the paper. This can be combined with the CAM historical run comparison into a new section 5.8, which would focus on the explanation of why the Greenland CO history looks the way it does. We will also expand the discussion to decades beyond 1970s and 1980s as the reviewer suggests.

The last sentence of this section is problematic because the information in parenthesis applies only to the decline period (1980s -1990s). As a side note, assuming methane steady state by ~2000 with constant sinks, total methane emissions must have stabi- lized (not just slowed their growth) some years before.

We will revise this sentence to take into account that CO sources are only declining during the period of [CO] decline, as well as that CH4 sources may have already stabilized.

Section 5.3: It would be nice to provide at least a scaling on how much CO is directly emitted vs. how much is coming from oxidation of methane and light hydrocarbons in the present day. The arguments are very qualitative and a bit taxing on the mind.

In the revised text, we will point the reader to Table S5 (which has a quantitative breakdown of CO sources) and try to find and include more quantitative estimates of the fraction of CO source attributable to light non-methane hydrocarbon oxidation.

Section 5.4: This section relies on the Wang et al. (2012) study, essentially providing a short summary of the relevant information from that paper. Given that there is no original isotope analysis in this manuscript, I think this section would be better placed as the last of their comparison sections. The purpose would be to bring in an isotope perspective from Wang et al. (2012) and discuss whether those results support or weaken their analyses of other comparisons.

We will follow this straightforward suggestion.

Again, I cannot see any justification for the specific focus on the 1970-1990 period. I realize the CO peak is within that period but the largest changes in the CO histories are during 1950-1970 and 1980-2000. Is it not more logical to look for a signal in isotopes when things are changing the fastest?

We will follow this suggestion and expand the discussion to decades beyond 1970s and 1980s.

Another issue here is the fact that Stony Brook CO data were excluded from the CO inversions. I understand and accept the reasoning in section 3.1, but it seems neces- sary to state (or show in a figure) how a Stony Brook only inversion compares with the ensemble results.

The Wang et al (2012) companion paper analysis of CO isotopic results actually used the CO concentration reconstruction presented in this paper (average of the 61 successful scenarios) to aid in the interpretation of CO isotopic data. So in this sense the two studies are completely consistent with each other. A Stony Brook only inversion was not performed and would not be consistent with the methodology used in our manuscript, which averages several datasets. The effect of including Stony Brook data in the combined NEEM EU dataset is small and will be shown in the revised Fig. S5 (see also answer to Stony Brook data related comment from Maarten Krol).

Section 5.5: This is probably the most problematic section in the manuscript from my perspective because I'm not sure what to make out of the H2 data. It seems to me that the first order scientific questions about H2 appear to be on the sampling, measure- ments, and modeling sides. I think the authors' interpretation of the lag between CO and H2 peaks as being supportive of "the sink hypothesis" is overly simplistic. First, they have to carefully frame it in time when OH could be rising to decrease CO and increase H2. This is difficult because the firn inversions are highly uncertain, and they are not even shown in the main body of the manuscript, which makes is harder for the reader to follow these arguments. Given the lack of understanding of the H2 behaviour in the firn, I recommend this section to be modified to focus more heavily on the chal- lenges in developing H2 atmospheric histories from firn measurements. To this end, it would make sense to move all the supplemental info on H2 into the main text (removing redundant info of course). The figure (fig. S6) especially, since timing is central to the OH arguments.

We will follow this suggestion and move more material from the Supplement to the main text (including what is now Figure S6), as well as focus this section more heavily on the challenges involved in reconstructing H2 atmospheric histories from firn air.

One other mechanism that comes to mind is the exchange between neighboring open and close pores within the lock-in zone. This could become significant if the age differ- ence between air in open and closed pores is large enough.

We will add a brief discussion of this mechanism (and relevant recent studies) into section 5.5.

Section 5.6: The discussion on the OH literature needs some revisions. Prinn et al. (2005) published methyl chloroform based OH estimates for 1978-2004 and Montzka et al. (2011) push it out to 2008, including a synthesis of their results with results of Prinn et al.

(2005) and Bousquet et al. (2005). The uncertainties are large as discussed in the text but Prinn et al. (2005) inversions show an increase during 1978-1985 (likely not significant) and a decline in OH from mid to late 1980's to late 1990's, both of which are on the order of 10% or so. I'm not sure if anyone thinks methyl chloroform inversions imply a 30% change in OH between 1979-1990.

Our main focus is on the period before 1990, which is why some of the papers that mainly addressed post-1990 OH variability have been left out. We will expand our discussion to include these references.

Conclusions: Lns. 10-12: There is no way to fully reconcile the discrepancy presented in fig. 9 with OH changes that are within range of sanity. I would scrap the second half of the sentence: everything starting with "under the assumption. . ." Lns. 23-25: I already stated my reservations about section 5.5. What goes into conclusions about H2 may need revisiting based on what happens with 5.5.

We will remove the second part of that sentence as well as further revise the conclusions regarding H2 (also based on comments from Maarten Krol).

Pg. 19000, lns. 19-21: The sentence that starts with "NOAA [H2] overall...' needs revision
We will revise this sentence for better readability

Pg. 19001, lns. 6-9: I don't understand the implication of this sentence.
This was meant to convey the point that exact calibration offsets between CSIRO and NOAA are not well determined; we will clarify this further.

Pg. 19003, ln. 17: I think "a best estimate combined data set" is more appropriate than "the best estimate combined data sets". You may have to write another sentence to tell there are more than one data set because of the US and Europe holes etc.
We will follow this suggestion.

References not cited in the Atmos. Chem. Phys. Discuss. paper

Lukas, M. A., 2008, Strong robust generalized cross-validation for choosing the regularization parameter, Inverse Problems, 24, 034006, doi:10.1088/0266-5611/24/3/034006.

Trudinger, C. M. et al., 2012, How well do different tracers constrain the firn diffusivity profile? Atmos. Chem. Phys. Discuss., 12, 17773–17834.

Witrant et al., 2012, A new multi-gas constrained model of trace gas non-homogeneous transport in firn: evaluation and behaviour at eleven polar sites, Atmos. Chem. Phys., 12, 11465-11483.

Witrant and Martinerie, 2013, Input Estimation from Sparse Measurements in LPV Systems and Isotopic Ratios in Polar Firns, Firns",Proc. of the IFAC Joint Symposium on SSSC, TDS and FDA, Grenoble, France, Feb. 4-6, 2013, http://www.gipsa-lab.grenoble-inp.fr/~e.witrant/papers/13_delta_sssc.pdf).

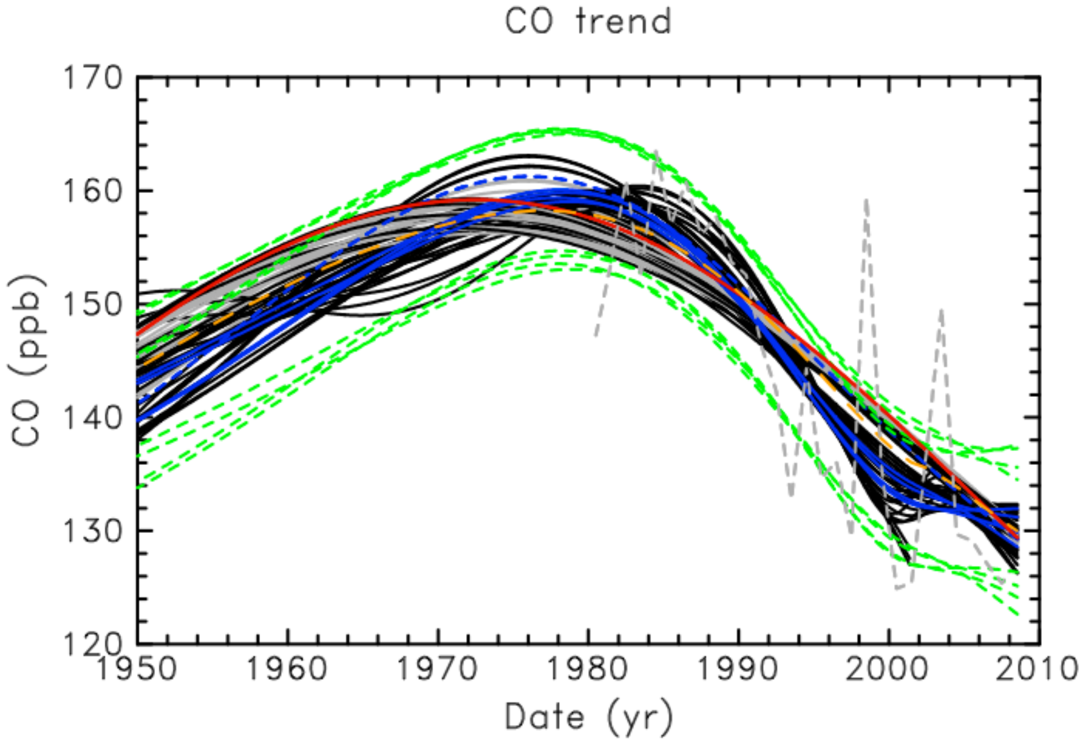
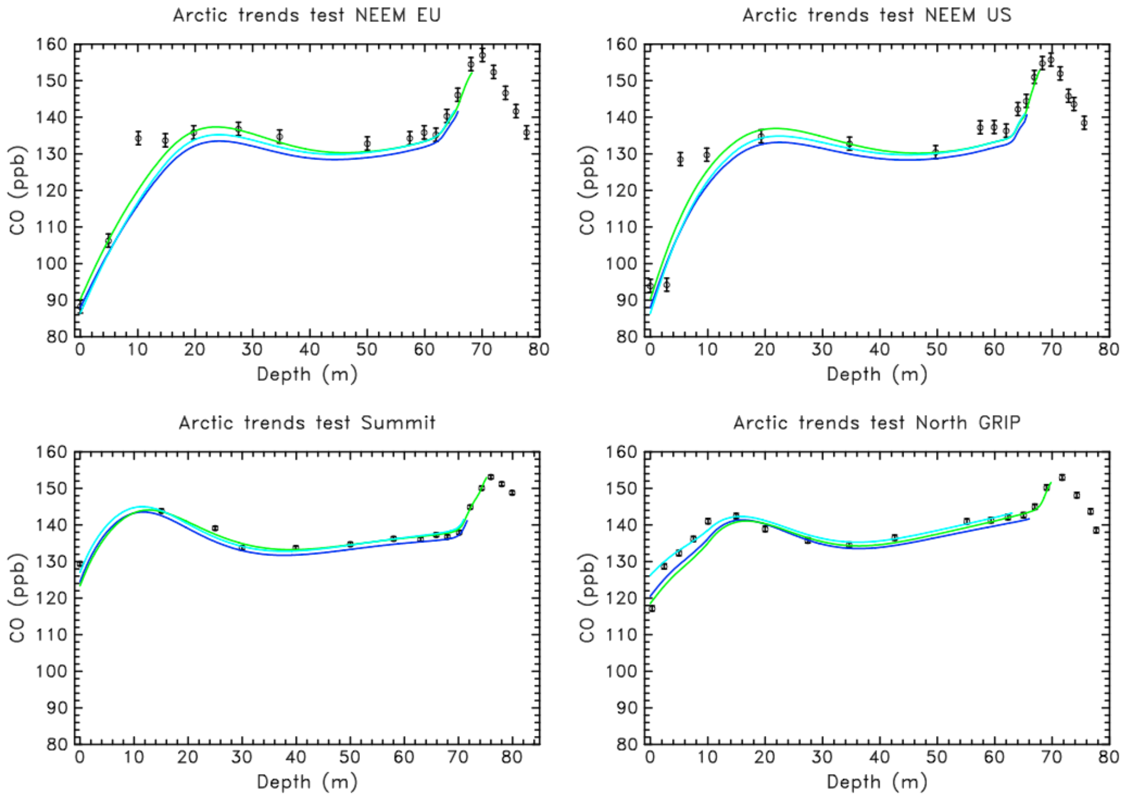


Fig. 1: Comparison of the previous range of scenarios (black and grey continuous lines) with multi-site scenarios based on the new robustness-oriented optimal solution (in blue and red). Grey lines show the scenarios which would be excluded if the NGRIP biomass burning influenced data point was taken into account. Red lines show the NEEM-EU + NEEM-US constrained scenario obtained with the INSTAAR model. Dashed blue lines show the NEEM-EU + NEEM-US constrained scenario obtained with the LGGE-GIPSA model. Continuous blue lines show the NEEM-EU + Summit + North GRIP and NEEM-US + Summit + North GRIP constrained scenarios obtained with the LGGE-GIPSA model. The dashed green lines show the model-determined uncertainty on the scenarios constrained with 3 sites (continuous blue lines). The long dashed orange line shows the average of the 61 selected scenarios in the manuscript (shown in black and grey here), it is the best estimate CO scenario used in Wang et al. (2012). The dashed grey line shows the annual mean CO atmospheric trend at Barrow (combined NOAA + OGI data).



A preliminary version of revised Figure S2. As mentioned above, this figure will be moved to the main text in the revised manuscript. Black markers represent the finalized [CO] data for each borehole. Lines represent runs of atmospheric [CO] histories from Arctic monitoring stations through the LGGE-GIPSA forward firn model for each borehole. Runs using Barrow data (combined NOAA + OGI) in green, Alert in dark blue, Ny Alesund in light blue. The lines end at the depth at which the mean age of CO in the firn air is the same as the start date of the atmospheric record (1980 for Barrow, 1992 for Alert and 1994 for Ny Alesund).