Response to Referee #1

We are very grateful to both referees for in-depth understanding of the present study and constructive suggestions. We believe that we have made our best efforts to consider questions/suggestions by both referees. We have made the following major corrections:

1. For our firn modeling, we have additionally used the CMIP6 scenarios for various trace gases (Meinshausen et al., 2017). The differences from those for the NEEM modeling (Buizert et al., 2012) are presented and discussed in a dedicated section of the revised manuscript. Both scenarios are examined for consistency with the NGRIP and NEEM firn data sets. The comparison of both scenarios highlights that they show a clear disagreement and produce a significant difference in reproducing the firn depth profiles for CH_4, but not for other trace gases.
2. We have made additional simulations for the NEEM firn with various diffusivity profiles in the same manner as the NGRIP firn. The iterative dating reconstructions of the historical CH_4 variations were also made from the NEEM firn. The reconstructions from both firn data are now presented and compared with the above scenarios.
3. Constraints from different trace gases have been evaluated by using the NEEM firn data. It turned out that $^{14}CO_2$ data play an important role in constraining firn diffusivity in the LIZ and thus reducing uncertainty in reconstruction.
4. We have concluded that, for CH_4, the Buizert et al. (2012) scenario is in better agreement with the two sets of firn data (NGRIP and NEEM) than the CMIP6 scenario. In addition, we point out that the former scenario is more consistent with the current understanding of the change in the interpolar difference (IPD) of atmospheric CH_4.
5. We have corrected all the figures accordingly and added necessary figures. Associated texts in many places are also modified.

Our responses to the Referee #1 are detailed below, where *referee's comments* and our responses are in different styles. The line, section and figure numbers in our responses are for the revised manuscript.

General comments:

This paper uses firn air measurements from two Greenland sites (NEEM and NGRIP) to investigate the Arctic atmospheric methane history over the 20th century. A firn model is applied first to NEEM to demonstrate model performance, then to NGRIP to try to infer the Arctic methane atmospheric history.

A key assumption of this study is that the Arctic CH4 atmospheric history is uncertain but that the atmospheric histories of the other six gases (CO2, SF6, CFC-11, CFC-12, CFC-113 and CH3CCl3) are known with sufficient accuracy to constrain the firn model. More should be added to justify this assumption - I am not convinced. I do agree that it is a problem that the Arctic methane history is not known as well as has been assumed in previous firn model studies. However, the CO2 atmospheric history from Buizert et al is created in much the same way as CH4 (with an assumed offset from the SH ice core record), but has the added complication that CO2 has the possibility of elevated levels in NH firn due to in situ artefacts (e.g. mentioned in Buizert et al and elsewhere in the literature) - this could affect the ability of CO2 to constrain the firn model. The 14CO2 atmospheric history (relevant for NEEM, but unfortunatly not measured at NGRIP) is probably quite reliable when it is based on atmospheric or tree ring measurements. The halocarbon histories are based on estimates of emissions, but these also have inherent uncertainty (the emissions themselves are due to

reported production/sales, assumed emission functions, and atmospheric lifetimes and therefore have uncertainties). I am not convinced that the atmospheric history of CH4 is significantly more uncertain than these other gases, I think all are known to some extent, but not perfectly.

We agree with referee that the atmospheric histories of the many trace gases have significant uncertainties. To answer the referee's question about the degree of uncertainty for different gases, comparisons of the existing atmospheric synthetic histories by Buizert et al. (2012) and by Meinshausen et al. (2017) (hereafter referred to as the BZ and CMIP6 scenarios) are very helpful, which are now presented in the revised manuscript (section 3.2). We show in the revised manuscript that, for most trace gases, the atmospheric scenarios by the two independent studies are in good agreements, but the CH_4 scenarios are clearly inconsistent with each other (and small differences are obvious in some gases as well). We have made additional model simulations with the CMIP6 scenarios, which show that the modeled depth profiles differ significantly in CH_4 but with smaller magnitude in other trace gases. It is difficult to explicitly evaluate uncertainties of the atmospheric scenarios of individual trace gases, but the comparisons of the BZ and CMIP6 scenarios highlight considerable deficiency of our knowledge on the historical CH_4 variations in comparison to other trace gases.

In addition to the question of how well atmospheric histories are known, it is also relevant to consider how well different gases can constrain firn diffusivity. Halocarbons measured in the deepest few firn samples at both NEEM and NGRIP are very close to zero, so do not provide a strong constraint on diffusivity in that region of the firn. As discussed at line 307, it is the region below about 74m at NGRIP that is used to infer the CH4 atmospheric reconstruction before 1980. The blue, orange and red lines in Figs 5 and 6 have a large spread below 74m for CH4 and CO2, but there is not a very large spread for the other gases, with the spread for some of these gases dropping rapidly to zero as depth increases. This shows that the modeled mole fraction profile for the CFCs, SF6 and CH3CCl3 in the deep firn is not very sensitive to the diffusivity profile, and consequently that the diffusivity profile is not as well constrained by these gases. It has been pointed out in previous studies that methane provides a strong constraint on diffusivity in the deep firn, but, as the authors note, only if the atmospheric history is well known, and unfortunately the authors are correct that it is not well known in the Arctic. CO2 would provide a similarly strong constraint on diffusivity, but I would suggest that the Arctic CO2 atmospheric history is also not well known and has the possibility of in situ artefacts in Arctic firn, as mentioned above. Thus, calibrating the firn model without CH4 for NGRIP, then expecting to reconstruct atmospheric CH4 is risky, and I believe the results show that it has not been successful (the model appears not to have been well constrained by the observations used).

We thank the referee for this comment. To answer the referee's question, and as suggested by Referee #2, we have made additional model simulations for the NEEM firn (where CH_4 is excluded for diffusivity tuning as we made for NGRIP) to understand degree of constraints provided by different gases (section 5). The series of model simulations showed that large part of constraints to diffusivity in the deepest layers comes from the $^{14}CO_2$ data, which indicates that diffusivity in the NGRIP firn is relatively underconstrained in comparison to the NEEM firn, thereby subject to large uncertainty in reconstructing trace gas histories. Constraints from halocarbons (in addition to that from CO_2 as previously made by Ishijima et al., 2007) are found to be relatively weak in contradiction to our expectation at measurements. Confirming this, as well as suggested by Referee #2, we have decided to

conduct iterative dating reconstruction of CH_4 also for the NEEM firn in the same manner as the NGRIP firn (section 4.2 and Figure 10). We now therefore infer the historical CH_4 variations consistent with both firn data to compensate incompleteness of the NGRIP data set only.

The most important contribution of this paper is questioning the assumption of a known Arctic atmospheric methane history for constraining firn models for Greenland firn sites. This has consequences both for calibrating firn models and for interpreting the CH4 north- south gradient in terms of emissions, as the authors discuss. However, as I have said, I believe the Arctic atmospheric histories for the other gases should be similarly questioned. I am not convinced that substantial conclusions have been reached in this study. The result that it is difficult to identify the atmospheric CH4 history that consistently reproduces the depth profiles of CH4 in NEEM and NGRIP firn is due to the fact that the firn model has not been adequately constrained by the other gases. The last 2 sentences of the abstract say that a consequence of this result is that the Artic CH4 history should be considered preliminary - it may be true that the methane history is not well known, however is not a consequence of that result. Rather, it is a prior assumption that has not changed as a result of the study.

While this study does highlight the deficiency that we don't know Arctic atmospheric CH4 well, in my opinion it doesn't go any way towards solving it. This makes me question the value of the study as it is currently presented.

We thank the referee for the in-depth understanding and critical evaluation of this study. According to suggestions by both referees, we have examined additional model simulations and we believe that the present study made one step forward from the original manuscript to clarify the current best capability based on the two available firn data sets in Greenland (NGRIP and NEEM). The new series of simulations have confirmed difficulty in reconstructing the Arctic CH_4 history with small uncertainties. Albeit not a perfect success (as the referee points out), we consider that all of our efforts (measurements and modelings) and open issues are of importance and worth documenting so as to support future studies on Arctic firn and ice cores. We have concluded that for CH_4, the two Greenland firn data sets as well as the current understanding of the IPD prefer the BZ scenario over the CMIP6 scenario and that subtle CH_4 modeling for Arctic firn sites is still challenging, thereby suggesting that a new measurement of shallow ice cores may be the only way to significantly reduce uncertainty of the CH_4 history (the last paragraph in section 5). With the revision, these arguments are now better supported by the new model simulations. We are confident that our conclusions in this study are original and that this study has provided a path to better solve the question of reconstruction of the Arctic CH_4 history. We are very grateful to the referees for comments to improve our manuscript.

Specific comments:
A conclusion in the abstract and at line 374 that "We find that, given the currently available firn air data sets from Greenland, reliable reconstruction of the Arctic CH4 mole fraction is possible only back to the mid 1970s" - atmospheric observations began around 1980, so this isn't much of a result. The title of the paper is 'Towards reconstructing the Arctic atmospheric methane history ...', but the study doesn't move very far towards that goal.

According to the new additional model experiments, we believe that we have made one step forward towards the title of this study, which we hope the referee could agree. Please refer also to our response described earlier.

Line 61 - "The NEEM-S1 data are notably higher than the ice core data after ~1850." The NEEM-S1 data after 1850 are fairly consistent with the Blunier et al 1993 data in Fig 1. The NEEM-S1 data are definitely higher than the Nakazawa data, but some of the Nakazawa data are lower than the SH Law Dome data which is unrealistic. Rhodes et al note that the uncertainty in absolute mole fraction of the NEEM-S1 data is about 6-9 ppb, and that that is a limitation to deducing the interpolar gradient, but perhaps the NEEM-S1 are our best chance at the moment to reconstruct NH methane between 1850 and 1945, seemingly better than the firn reconstruction presented here. The NEEM-S1 data were mentioned once in this study but otherwise dismissed (unfairly, in my opinion).

We agree that the NEEM-S1 ice core data need to be more discussed, because it is the only available data that cover the period of interest at high resolution. In the revised manuscript, we did so by investigating the CMIP6 scenario which to a great extent relies on the NEEM-S1 data, and we found the following issues. First, our firn modeling results from NGRIP and NEEM sites indicated that the Arctic CH_4 history in line with the NEEM-S1 data (i.e. the CMIP6 scenario) cannot reconcile both NGRIP and NEEM profiles (Figures 8 and 9). Second, the NEEM-S1 data suggest CH_4 IPD of ~130 ppb around 1900, which is almost equal to that observed in the 1980s (Figure 2). Constant IPD between 1900 and 1980 is highly unlikely, because increase of CH_4 emissions for the intermediate period is considered to have occurred in the NH and it requires IPD to increase with time, as discussed in the previous studies (Dlugokencky et al., 2003; Ghosh et al., 2015; Chandra et al., 2021). Accordingly, we conclude that the CMIP6 CH_4 scenario (that is, the NEEM-S1 data) is consistent neither with the firn data nor the current our understanding on the atmospheric IPD. The data set may have an issue in data quality or dating, but such in-depth discussion is beyond the scope of this study. We have included above discussions in the revised manuscript.

The strategy with prior and calibrated diffusivity profiles is not clear to me. For example: Line 177 - "The diffusivity profile optimised for the CIC model was tuned for our model" - what does this mean? Was the CIC profile used as a prior then improved by comparing to observations?

We have reformulated the relevant sentences, after we made various diffusivity simulations also for the NEEM firn according to suggestion by Referee #2 (section 3.4). In the simulations presented in the original manuscript, we achieved good reproducibility with the diffusivity profile that was modified from the profile originally optimised for the CIC model. In the revised manuscript, we have used various diffusivity profiles including those prepared by likewise modifying the profiles optimised for other models.

Line 182 - What diffusivity profile gave the RMSD value for NEEM of 0.83? Is this the same as the case shown in Fig 2?

We have added the following sentence in the revised manuscript.

Line 244: “This RMSD value was achieved with an effective diffusivity profile that was prepared by modifying the profile originally optimised for the CIC (Centre for Ice and Climate) model at a certain range of depth.”

Line 197 - was that the initial diffusivity from equation 4 or Ishijima et al (2007)?

In the manuscript, the initial diffusivity corresponds to that was used in Ishijima et al. (2007). This has been made clear as follows and throughout the revised manuscript.
Line 215: “In this study, the effective diffusivity profile prepared for the NGRIP firn by Ishijima et al. (2007) is referred to as the initial diffusivity and it was modified to improve the reproducibility of our newly measured trace gas profiles.”

Lines 198-203 - This paragraph is a little hard to follow, it became clearer as you read further, but could be improved. For example, line 198 "We examined the different sets of profiles" .. which different sets? (It becomes clearer, but is confusing at this point). Line 201 - "We prepared 100 different sets" - at this point the reader wonders how they are prepared, this also becomes clearer (page 12), but if this information was given when the steps are first discussed, it would improve readability.

We have reorganized the explanations of the diffusivity profiles. The description of preparation of the 100 different sets of profiles have been now merged into section 3.3.

Line 222 - which diffusivity profile was used in Fig 3? Eqn 4, Ishijima or a tuned profile? Why aren't the NGRIP results corresponding to the diffisivity profile giving RMSD=0.51 shown as a case (e.g. dashed black line) in Figs 4, 5, 6, 7 and 8? This would be good to see.

The initial effective diffusivity profile from Ishiijma et al. (2007) was used here. The diffusivity profile giving the smallest RMSD was shown as part of red lines in the series of the figures. We do not think that the smallest RMSD case should be highlighted, because identification of a single case does not make a strong sense in this study where both diffusivity and atmospheric scenario have significant uncertainties.

Fig 2 - It is difficult to see some of the observations, particularly CH4 in the deep firn. Could the observations be shown more clearly?

We have made the symbols of the observations closed and larger in the figure (Figure 4).

line 212 - the atmospheric history is not quite monotonic, so there could be more than one time with atmospheric mole fraction matching the mole fraction at the firn depth - how is that handled? Was the atmospheric history smoothed?

We have added the following sentence in the revised manuscript.
Line 280: “It is noted that the smoothing spline curve applied to the BZ CH_4 scenario was used for calculation of the effective age, as the input scenario with seasonal variation (Figure 2) would not allow the effective age to be uniquely determined.”

Line 230 - at this point I'm already wondering what the modeled CH4 depth profile at NGRIP looks like with the Buizert et al atmospheric scenario in the model, but I need to wait....

The colored lines (red, orange, blue) cover many different diffusivity profiles, some of which don't fit the firn data well at all, particularly the blue cases. Is it worth showing the blue cases at all? At line 293, they are described as "less likely", but many of them simply do not fit the observations. Could the red group be split into two to highlight the really good cases? Do the better diffusivity profiles tend to fit all gases well, or do some profiles fit some of the gases well and others not so well, and vice versa (for groups of gases)?

We hope to keep the light blue cases in the figures (Figures 3, 6, 8, 10 and 11) because it helps to see that the acceptable range was well narrowed after series of our simulations. As the referee also concerns, we have struggled between uncertainties of the diffusivity profile and the atmospheric scenario, and at this stage where neither can be tightly fixed, we do not wish to clearly highlight "really good cases" as it could lead misinterpretation of the present study. We hope the referee also understands that we try to present honest assessment of our current best use of the data. For the last question above, one "good" diffusivity profile tends to reproduce all gases well (except CH_4) as seen in the similarity of model-data differences for those gases in the initial simulations (Figure 5).

line 297 "suggesting that the CH4 mole fraction may have been lower than the initial modeling scenario" - I am not convinced that this is a robust result. I am not convinced that the atmospheric scenarios are known more accurately for the other gases, or that they provide sufficient constraint on the model so that it can be used to infer the CH4 history, as discussed above.

Please refer to our response described earlier.

Line 309 - "The differences between the initial and corrected atmospheric CH4 scenario from these three deepest data are up to ~ 100ppb" - because the model is not well constrained by the other gases.

Please refer to our response described earlier.

Line 313 - "NGRIP firn data suggests decreased CH4 mole fraction from the 1950s to 1970s in any case, albeit with large uncertainty" - I do not believe this is a robust result, for reasons given above.

Please refer to our response described earlier.

Line 320-322 - if I understand this correctly, the CH4 history reconstructed from NGRIP gives a larger model-data difference at NEEM than the original history, is that not indicating an inconsistency?

The referee is correct. The reconstructed CH_4 history from the NGRIP shows to some degree inconsistency with the NEEM firn data. To make fairer treatment, and as suggested by Referee #2, we have also made historical CH_4 reconstruction from the NEEM firn data (but CH_4 excluded from the diffusivity evaluation), and the reconstructions from both firn sites are compared in the revised manuscript (Figure 10).