Response to Referee #2

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 #2 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.

Umezawa et al. used a suite of gas measurements from NGRIP firn air (CO2, CH4, SF6, CH3CCl3, CFC-11, CFC-113, and CFC12) in combination with a firn model to reconstruct the atmospheric history of CH4 in the northern hemisphere (NH). Although the firn air samples were collected close to 20 years ago (in 2001), a great care has been taken to use state-of-the-art (or close to state-of-the-art) measurement techniques to achieve analytical precisions that are comparable or better than present-day modern atmospheric measurements. This is not a trivial merit and I think the authors should be commended. Following precedents set by previous studies of firn air (e.g., Rommelaere et al., 1997; Trudinger et al., 2002; Witrant et al., 2012; Buizert et al., 2012), Umezawa et al. used a forward gas transport firn model that takes in a "known" atmospheric history of a certain gas as an input and produce the expected mole fraction of that gas vs. depth profile in the open porosity of the firn. The difference between the expected mole fraction depth profile vs. measurements is then used to tune the "effective diffusivity" for this particular firn air sampling borehole (which is the Japanese firn sampling borehole at NGRIP).

A previous study by Buizert et al. (2012) set a precedent by including CH4 as part of the suite of gases used to tune the effective diffusivity at the NEEM ice core site. Buizert et al. (2012) achieved this by first making an educated guess about the "known" atmospheric history of CH4 in the NH. However, in this study Umezawa et al. challenge this assumption, treat the

NH atmospheric history of CH4 as an unknown, and only used the other six gas measurements (CO2, SF6, CH3CCl3, CFC-11, CFC-113, and CFC12) to tune the effective diffusivity profile for NGRIP. As a result, the atmospheric CH4 history reconstructed by Umezawa et al. has larger uncertainties; from this, Umezawa et al. argue that we cannot take the NH CH4 history for granted as a known variable to tune effective diffusivity profile for ice cores collected in the northern hemisphere and to certain extent, we also do not know the true atmospheric history of NH CH4 before the 1970s.

The main conclusion from of Umezawa et al. study (to which precision do we know the NH atmospheric history of CH4) is potentially an important one, so I would recommmend the manuscript for publication if the following comments are sufficiently addressed.

We thank the referee for in-depth understanding of the present study.

Major comments:

1. As reviewer #1 pointed out, it is not immediately clear whether the atmospheric histories for the other six gases outside of CH4 (CO2, SF6, CH3CCl3, CFC-11, CFC-113, and CFC12) used to tune the effective diffusivity profile are sufficiently known as well. Why focus on CH4 and not say, the uncertainty on NH CO2 history? I think a discussion or even a specific section addressing this question is warranted. Fortunately, given the current state-of-science knowledge, I think Umezawa et al. should be able justify their assumption in using CO2, SF6, CH3CCl3, and CFCs to tune effective diffusivity profile. Meinshausen et al. (2017) took a great care in synthesizing all available data from historical atmospheric measurements, firn and ice cores from several sites to best reconstruct the GHGs (including CO2, CH4, SF6, CH3CCl3 and the CFCs measured by Umezawa et al.) mole fraction, interhemispheric gradient, and seasonal variabilities for the purpose of CMIP6 model runs. This would be a great starting point. The justification for treating the NH histories of CO2, SF6, and the CFCs as "known" parameters, or at least better known parameters than NH CH4 history in my opinion should revolve around a discussion about the interpolar gradients of these gases (which are relatively small owing to their long atmospheric lifetimes), but I will leave the exact formulation of this argument to Umezawa et al.

I think a sensitivity analysis comparing what mole fraction should we expect in the open porosity of NGRIP firn if we put in NH vs. SH history from Meinshausen et al. (2017) for CO2, SF6, and the halocarbons is warranted to further drive the point home. I might be wrong, but I would expect the mole fraction vs. depth profiles for these suite of gases in the firn open porosity would not be as sensitive to NH vs. SH difference, at least relative to their respective measurement precisions compared to CH4 given their long atmospheric lifetimes and relatively low interhermispheric gradient. Given Umezawa et al. already had their forward firn model setup, hopefuly this does not require a lot of additional work. Furthermore, as a more general comment, I would also recommend Umezawa et al. to use gas histories from Meinshausen et al. (2017) for their overall firn gas transport and effective diffusivity tuning because the GHGs histories proposed by Meinshausen et al. (2017) represent more updated, better-educated "guesses" than the gas histories previously used by Buizert et al. (2012).

We thank the referee for the constructive suggestions. In addition to the atmospheric scenarios by Buizert et al. (2012) (hereafter referred to as BZ scenario), we now also use the

synthetic atmospheric histories by Meinshausen et al. (2017) (hereafter referred to as CMIP6 scenario) for our firn transport model simulations. In the revised manuscript, we have added a section in which both scenarios are compared and their differences are described (section 3.2 and Figure 2). We highlight that, while the two scenarios show general agreements to each other for most trace gases, difference between the two scenarios is outstanding for CH_4. The CH_4 difference comes from the underlying datasets and assumptions for producing the respective synthetic data. The BZ scenario was produced by adding IPD to the Law-Dome-based Antarctic history, whereas the CMIP6 scenario employed the data from the NEEM-S1 ice core (Rhodes et al. 2013). While the BZ scenario assumed that IPD increased with the CH_4 growth rate (thus, with time) over the 20th century, the CMIP6 scenario suggests IPD to be almost constant and >100 ppb over the period. Except CH_4, IPDs of the other trace gases are sufficiently consistent with each other.

According to the referee's suggestion, we have made series of forward modelings for both NGRIP and NEEM firn sites using both historical scenarios (Figures 7, 8 and 9). We show in the revised manuscript that, the simulations with the CMIP6 scenario tend to overestimate the depth profiles of CH_4 to larger degree at both firn sites than those with the BZ scenario. For other six trace gases, the simulations using both histories do not produce significant differences.

We have also made the NGRIP firn simulations with the atmospheric scenarios for SH from Meinshausen et al. (2017) (Figures 5 and 8). We have found that input of the SH history for the NGRIP simulation resulted in significant differences even for the trace gases excluding CH_4. However, as the referee presumed, it has been found that relative differences between the simulations with the NH and SH scenarios are most pronounced for CH_4; while the differences for the other six gases are no more than 20 times the respective measurement precisions, that for CH_4 reaches about 40 times. This also emphasizes strong impact of IPD of CH_4 in comparison to other species.

In summary, we have concluded that (1) uncertainty of atmospheric history of CH_4 is manifest, while those of the other trace gases are relatively small so that they consistently reconcile the NGRIP and NEEM firn profiles, and (2) the atmospheric CH_4 history of CMIP6 is likely too high for the first half of the 20th century. Regarding the latter, we note that the increasing trend of IPD over the 20th century in the BZ scenario, in comparison to the constant IPD in the CMIP6 scenario, is more consistent with increasing anthropogenic emissions in the northern hemisphere suggested by earlier studies (Dlugokencky et al., 2003; Ghosh et al., 2015; Chandra et al., 2021). We have therefore corrected our argument. The BZ CH_4 scenario (Buizert et al., 2012) is the current best synthetic scenario, albeit large uncertainty and its use for tuning firn diffusivity unproven, and the firn air data are not consistent with the alternative CMIP6 scenario (Meinshausen et al., 2017).

2. The suite of CFCs measurements (CFC-11, CFC-113, and CFC12), CH3CCl3 and SF6 do not provide good constraints for reconstructing effective diffusivity for the deep firn just because the concentration of these gases are all very low and close to zero. Usually, the gases that are most useful to reconstruct the effective diffusivity in this firn region are CH4, CO2, and 14CO2 due to their respective unique atmospheric histories. 14CO2 is especially useful as its atmospheric history can be validated from tree rings and historical atmospheric measurements. Furthermore, 14CO2 has a unique profile from the "bomb pulse" in the 1950s that provides a strong and unique constraint on the effective diffusivity. Unfortunately, 14CO2 measurements for NGRIP are not available. Because the CH4 history in

this study is treated as unknown, the effective diffusivity in the lower part of the NGRIP Japanese borehole presented by Umezawa et al. is almost solely constrained by CO2 data. This made me question whether the conclusion obtained by Umezawa et al. regarding how we cannot accurately reconstruct NH CH4 history from firn air samples is a unique problem pertaining to NGRIP (and its suite of gas measurements) or is it more general problem to other Greenland ice core sites as well. I don't think the current version of the manuscript sufficiently answer this question and additional work might be warranted to justify the conclusion put forward by Umezawa et al.

In particular, I think it would be especially useful to revisit the NEEM data from Buizert et al. (2012) with the same firn model and iterative dating algorithm presented in this study, but also excluding CH4 as part of the suite of gases to tune the effective diffusivity of the NEEM site. This would provide a more fair comparison rather than putting in the atmospheric history reconstruction from likely underconstrained NGRIP site into NEEM with a forward firn model. It would be interesting to see whether additional constraints from 14CO2 data at NEEM will allow for reconstruction of NH CH4 history with a better uncertainty and to what extent the uncertainty is better. For this experiment, I would recommend using the updated "known" 14CO2 history from Graven et al. (2017). Given Umezawa et al. already had their firn model tuned for the NEEM EU borehole as part of their model validation, I don't think this extra calculation would require significant amount of additional work.

We thank the referee for the suggestions. As the referee points out, the constraints by halocarbons (CFCs and CH_3CCl_3) are relatively weak in contradiction to our expectations at measurements. From series of the simulations which we have made after the referee's suggestion, we found that large uncertainty in reconstructing the CH_4 history is a particularly pronounced problem for the NGRIP firn, and that the NEEM firn data set provides reconstruction with smaller uncertainty (section 5). This critical difference is ascribed to availability of the $^{14}CO_2$ data, by which we feel very regrettable for lack of $^{14}CO_2$ measurements for the NGRIP firn.

More specifically, according to the referee's suggestion, we have made forward model calculations also for the NEEM site. We have evaluated range of diffusivity profiles by trace gases excluding CH_4 and made reconstructions of historical CH_4 variations in the same manner as made for NGRIP (Figure 10). The result shows that the NGRIP-based reconstructions have larger uncertainties than the NEEM-based reconstructions. While the estimate of uncertainty of effective age at the two deepest depths (which roughly corresponds to the time period 1950–1970) exceeds 10 years, those at the corresponding NEEM depths (four deepest depths) are estimated to be less than 10 years (Figure 11). If $^{14}CO_2$ data were excluded for evaluation of the diffusivity profiles, we found that the uncertainties of effective age at the two deepest depths at NEEM would be increased to 15–20 years. Therefore, as the referee points out, the constraint from $^{14}CO_2$ data is strong, and the NGRIP reconstruction would have been different if $^{14}CO_2$ data were available. Relevant figures (Figures 10 and 11) and discussion have been added in the revised manuscript.

3. I think the uncertainty analysis/discussion regarding the conclusion is a bit lacking. It is not immediately clear to me whether conclusion reached by Umezawa et al., that NH CH_4 history in general should be considered preliminary and should not be used to tune effective diffusivity is sufficiently justified. From the study, it is clear that reconstructing CH_4CH4 history from NGRIP firn air samples, when CH4 is excluded from the suite of gases used to

tune the effective diffusivity result in large uncertainties. But I think we know the NH CH4 history slightly better than just the reconstructed history from NGRIP firn air presented in this study.

Meinshausen et al. (2017) decided against providing uncertainties to the reconstruction of GHG histories that they did, arguing that the CMIP6 models would not have the computational resources to run multiple scenarios and sensitivity analysis from multiple GHG histories on top of the envisaged SSPs. I think an assessment about the uncertainty of historical CH4 reconstruction is very valuable and Umezawa et al. is in a unique position to take a first attempt at this. How about reconstructing NH CH4 history from NEEM (with its additional 14CO2 constraint) like discussed above, how about combining NGRIP, NEEM history inversion results to make a best-estimate of NH CH4 history and its uncertainties, and how about including CH4 in the suite of gases used for effective diffusivity tuning, but through iterative method starting first with larger uncertainty for the RMSD calculation to account for uncertainty in the CH4 history? There are still many avenues to explore beyond the reconstructed NH CH4 history from NGRIP firn samples before one can conclusively claim that we don't know the NH CH4 history to such a degree that it should not be included in the suite of gases used to tune effective diffusivity in firn profiles. I don't demand Umezawa et al. to do all of the above, as it might constitute a whole different study entirely, but a preliminary exploration on this and an honest assessment about how well can we reconstruct the NH CH4 history would significantly strenghten the manuscript and provide very valuable insights to the community.

We thank the referee for the constructive suggestions. According to the referee's comment, we have additionally made reconstructions of historical CH_4 variations from the NEEM data (Figure 10). The CH_4 reconstructions from both NGRIP and NEEM are now combined and then compared with the BZ and CMIP6 scenarios. We realize that both scenarios were prepared with great care and used maximum number of data available at each time of the production, but they show significant differences in CH_4 for the early 20th century and earlier, as described above. While the BZ scenario follows the overlapping range of the reconstructions from NGRIP and NEEM back to around 1950, the CMIP6 scenario shows excursion to higher CH_4 mole fraction. Albeit large uncertainties of the reconstructions, we have concluded that the BZ CH_4 scenario better reconcile the currently available firn data from the NGRIP and NEEM sites.

We agree with the referee that it is of great value to assess uncertainty of the historical CH_4 scenario for climate modeling studies. However, such exact evaluation is still difficult because of the large uncertainties in reconstructing the CH_4 history from the firn data sets. A current possible conclusion is that the available NGIRP and NEEM firn data sets are in agreement with the BZ scenario better than the CMIP6 scenario. Considering that the CMIP6 scenario relies on the NEEM-S1 ice core data, this study highlight inconsistency between the ice core and two sets of firn data in Greenland. Rigorous evaluation/discussion of these available data sets is an important open question, but it is beyond the scope of this study.

Minor comments:

I find that in general, the description about the firn gas transport models and the iterative method is very brief and might be bit hard to follow. The brevity is fine for the main manuscript, but the authors might want to consider a supplementary material where they

will have more room to describe the gas transport model, iterative methods, and especially additional data treatments. For example
Line 212 "Effective age at each sampling depth was calculated..." Several steps are clearly skipped here. It is not immediately clear to me, from the description of the model and equations above how one can determine the effective age at each sampling depth, as all the description before this line only pertains to the forward firn model. Did Umezawa et al. calculated a depth-age transfer function similar to Rommelaere et al. (1997) or through other means? Either way this needs to be elaborated.

We have made descriptions of our modeling approach enriched in the revised manuscript (section 3), rather than adding a supplemental material that could make readers go back and forth.

The depth-age transfer function presented by Rommelaere et al. (1997) is indeed an interesting approach, and we have once examined similar age distributions for various trace gases at different depths at the NGRIP site (not shown). However, the effective age in this study was calculated in a simpler manner according to Trudinger et al. (2002), which was also used in Ishijima et al. (2007). The modeled CH_4 mole fraction at each sampling depth was compared to the input atmospheric scenario of the forward modeling, and the time at which the modeled CH_4 mole fraction agrees to the scenario was determined to be the effective age at the depth. As in Trudinger et al. (2002), the measurement data against the effective ages produce a renewed atmospheric scenario, which is then used for a forward modeling again. These steps are repeated to observe convergence of modeling results (iterative dating). We have reformulated the sentence as follows:
Line 276: "The iterative dating for CH_4 was performed as follows:
(*I*) Depth profile of CH_4 was calculated with the initial atmospheric CH_4 scenario.
(*II*) The modeled CH_4 mole fraction, calculated in step *I*, was compared to the input atmospheric CH_4 scenario, and effective age at each sampling depth was determined as the time when the modeled CH_4 agreed with a value in the atmospheric CH_4 scenario. 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.
(*III*) A new atmospheric CH_4 scenario was constructed by assigning the observed CH_4 mole fraction, at each depth, to the effective age determined in step *II*. The observed CH_4 versus the effective age data set was interpolated by a smoothing spline function and it is considered as a revised atmospheric CH_4 scenario.
(*IX*) Depth profile of CH_4 was again calculated with the revised atmospheric CH_4 scenario constructed in step *III*.
(*X*) The above steps *II*–*IX* were repeated until the model-data difference converged within an acceptable range (typically after a few iterations) (Trudinger et al., 2002; Ishijima et al., 2007). In this study, we made five iterations for each modified diffusivity case as we confirmed sufficient convergence of the result."

Fig.3. From the text it says "Figure 3 presents the initial simulations ..." Does this mean this is the initial effective diffusivity profile? It might also be beneficial to have the other effective diffusivity profiles like Fig.5 shown in Fig. 3.

In the original manuscript, we intended to begin from the starting point of our modeling. The initial simulations in the original manuscript were made with the effective diffusivity profile

used for the previous study (Ishijima et al., 2007). After the reformulation of the manuscript according to the referees, we have now added the simulation results for the NGRIP firn with the atmospheric scenario for Antarctica (Figures 5 and 8), in order to highlight importance of IPD for different trace gases.

There are several data treatment steps that is missing/the authors did not explain in sufficient details, or if the authors didn't do it, it is not well justified why they choose not to. For example, in their supplementary material Buizert et al. (2012) discussed how they added additional uncertainties for CO2 to account for possible in situ production and bubble close-off fractionation. In Buizert et al. (2012), uncertainty in atmospheric histories is accounted during the tuning of effective diffusivity by running the uncertainties through the forward model when the tuning of effective diffusivity is near complete to transfer the uncertainties from time domain to depth domain. I might miss it somewhere, but I think it is not immediately clear to me how the uncertainties of "known" atmospheric gases used to tune the effective diffusivity is treated in this study.

We agree that our methodology of the data treatment, in particular for estimation of uncertainties, was insufficiently explained in the original manuscript. For the NGRIP firn, we only included the analytical precisions as the uncertainties in the RMSD evaluations, and did not include the possibility of in-situ production and close-off fractionation as done by Buizert et al. (2012) for the NEEM firn, because they are minor contributors with insufficient quantitative understandings (the atmospheric scenarios and analytical precisions are the two largest contributors to the total uncertainties). Therefore, the difference between the uncertainties of the NGRIP and NEEM firn reconstructions is largely due to the atmospheric scenarios. In theory, it would be possible to estimate the additional uncertainties for the NGRIP firn in the same manner as Buizert et al. (2012) and it would be a straightforward approach for comparison. However, in revising the manuscript according to the referee comments, we largely expand the simulations and now use the two independent atmospheric scenarios for all the trace gases, so that the uncertainties in the atmospheric scenarios are appreciably examined through the comparisons of the simulation results using the two scenarios. In addition, the complete comparison our results with those of Buizert et al. (2012) requires time-varying uncertainty estimates for the CMIP6 scenarios for all the gases, which is by itself a quite complex problem and beyond the scope of this study.