### **Response to Referee #1**

We are grateful to the referee for thorough re-review of our manuscript. Again, comments are very helpful for us to improve the manuscript. Our responses are detailed below, in correspondence to *referee's comment*.

My concerns about the original version of the paper have generally been addressed. The most important part of this paper is questioning the assumption that NH CH4 is well known over the past century and can be used without considering uncertainties to tune firn diffusivity. For the paper's exploration of this topic, I believe it is worthy of publication. The limitations of the approach are now well spelled out.

In the first review on the original manuscript, the referee provided important criticisms, which significantly contributed to our preparation of the revised version. We are very pleased to hear general agreement of the referee for the value of this study.

#### Minor comments

*line 16 - would it be more appropriate to say 'we try to reconstruct CH4 for that period'?* According to suggestion by Referee #2, we have deleted the sentence.

### line 19 - the mention of NEEM at this point in the abstract understates the use of NEEM data in the paper. NEEM should be discussed more equally with NGRIP (details can be less, but at least mention the two sites, as it can be confusing as to how they are both used).

According to this suggestion as well as comment from Referee #2, we have revised the sentences as follows:

"We newly report a data set of trace gases from the air trapped in firn (an intermediate stage between snow and glacial ice formation) collected at the NGRIP (North Greenland Ice Core Project) site in 2001. We also use a set of published firn-air data at the NEEM (North Greenland Eemian ice Drilling) site. The two Arctic firn air data sets are analyzed with a firnair transport model, which translates historical variations to depth profiles of trace gases in firn."

Line 49 - The data in the figure cover the last 300, not 200, years

The caption has been corrected to 300 years.

*Fig 1 - the gray symbols for firn and ice core are hard to see on a printed copy* According to this comment and suggestion by Referee #2, we have modified the color style of Figure 1.

# Line 72 - the Law Dome CH4 record has been updated by Rubino et al 2019 (https://doi.org/10.5194/essd-11-473-2019)

We have updated the data in Figure 1 and cited Rubino et al. (2019).

## *Line 77 and 81 - I don't understand use of the word 'synthetic' here, wouldn't reconstructions, scenarios or histories be better?*

We hope to keep the word here. We use the word to mean a data set that is to large extent supported by underlying assumptions as observational data are limited.

## Section 2 - NGRIP is described in this section, yet NEEM is barely mentioned. At least refer to Buizert et al in this section for NEEM details.

At the end of this section, we have inserted the following paragraph.

"We also use a suit of trace gas measurement data from the NEEM site (77.45° N, 51.06° W). The firn air samples were collected in July 2008. The details of firn air sampling and gas measurements have been described by Buizert et al. (2012). The depth profile data of all the above trace gases (CH<sub>4</sub>, CO<sub>2</sub>, SF<sub>6</sub>, CFC-11, CFC-12, CFC-113 and CH<sub>3</sub>CCl<sub>3</sub>) from the NEEM firn are used in this study as well as <sup>14</sup>CO<sub>2</sub> data that are available from NEEM but not for NGRIP."

# Section 2 - Make it clear here that some gases (stating which) were measured and modelled previously by Ishijima, and which new gases are measured in this study.

The following sentence has been added at the end of third paragraph of this section. "Although not presented in this study, the firm-air samples were analyzed for nitrous oxide (N<sub>2</sub>O) and its isotope ratios (Ishijima et al., 2007)."

## Section 3 - could mention depths of top of LIZ and deepest sampling at NEEM as well as NGRIP.

These data have been added with the paper Buizert et al. (2012) cited.

### Line 169 - Schwander mis-spelt

Corrected.

### line 182 - could quantify the maximum difference between the 2 scenarios

The sentence has been corrected as follows: "In contrast, the Arctic CH<sub>4</sub> histories by the two studies differ considerably with maximum difference of ~85 ppb around 1910 (Figure 2b)."

## *Line 184 - could mention that the CMIP6 Antarctic histories are based on the Law Dome ice cores.*

The following sentence has been inserted:

Note that the CMIP6 scenario for the Antarctic latitude was constructed based on the Law Dome ice core data (see agreement in Figure 1).

### line 185 - "indicates almost constant values at ~ 130 ppb" - add "of IPD" i.e. "indicates almost constant values of IPD at ~ 130 ppb" Corrected.

### line 199 - the authors state that the two scenarios indicate that the other gases are at least better known than CH4 - could it be that they are based on the same underlying data? I'm not disagreeing, just be careful with the comparison.

The referee is correct. We have corrected the sentence and added another sentence as follows: "The scenarios of the individual trace gases have inherent uncertainties, but the comparisons of the two scenarios indicate that the data sources for other gases do not show inconsistent variations as seen in  $CH_4$ . It should be however noted that except  $CO_2$ , many trace gases lack observational data for the early 20th century, thereby both scenarios to a large extent being based on same data sources."

line 209 - weren't simulated and observed profiles of more gases than CO2 used?

No. Only CO<sub>2</sub> was used in the previous studies.

*line 210 - for clarity, state specifically the effective diffusivity as a function of depth. e.g. 'initial guess of the depth profile of effective diffusivity for CO2, Dinit(z)'* Corrected.

page 10 - say earlier in this paragraph that CH4 was not used here, e.g. around line 222 We think that the explanation fits better at the original place. The sentence has been however modified to make it clearer that  $CH_4$  was not used in the evaluation.

line 260 - specify the trends are of CH4, and the two sets of scenarios are of CH4.

This sentence explains trends and scenarios of other trace gases as well. For clarity, the first clause of the sentence has been corrected as:

"To represent atmospheric trends of different trace gases in the Arctic region, ..."

# *line 293 - for clarity, specify that the initial diffusivity was from Ishijima (fitted to which gases)*

The following sentence has been added:

"It is again noted that the initial diffusivity profile was tuned only for the depth profile of CO<sub>2</sub> (Ishijima et al., 2007)."

# *line 298 - CFC-11 and CFC-113 have not increased monotonically, they have been decreasing since maxima in the 1980s and 1990s, respectively.*

We thank the referee for this correction. We have corrected the sentences as follows: "It is known that, since the mid 20th century, the atmospheric mole fractions of the five trace gases (CO<sub>2</sub>, SF<sub>6</sub>, CFC-11, CFC-12 and CFC-113) have increased either monotonically or shown peak/slowed increase in the early 1990s (Sturrock et al., 2002; Martinerie et al., 2009). In contrast, CH<sub>3</sub>CCl<sub>3</sub> has increased until the early 1990s and has rapidly decreased since then (Sturrock et al., 2002; Rigby et al., 2017), which is also observed in Figure 2."

## line 227 - could add 'shown by the dotted lines in Fig 5' at the end of the sentence that finished in line 227

Corrected. We consider that this comment referred to line 327 of the former version of the manuscript.

line 344 - could add 'for gases other than CH4' at the end of the sentence that finished in line 344

Corrected.

*line 365 - "initial diffusivity profile of the BZ scenario" - doesn't make sense as written* Corrected as "initial diffusivity profile and the BZ scenario".

*line 365 - it is important to know here whether Ishijima used CH4 to tune the model* Ishijima et al. (2007) used only CO<sub>2</sub> and it has been clearly explained earlier.

*line 368 - 'also' rather than 'commonly'* Corrected.

line 372 - 'is used to the firn' - word missing? force or drive maybe?

The word "force" has been added. The sentence has then been reformulated and moved section 4.5.

## *Fig 9a - how significant is it that CO2 in the deep firn at NEEM is not particularly well modelled with either scenario?*

As seen in the figure, we do not find the difference between simulations with the different scenarios significant, and thus cannot identify which scenario better reproduces depth profile of CO<sub>2</sub>.

line 401 - when looking at the horizontal rows of circles, it may be hard to understand what the 'upper bounds' is, and may need better explanation

The following words have been inserted: "...the upper bounds of the reconstructions (line connecting far-left red circles)..."

line 422 - 'decreased CH4 mole fraction from the 1950s to 1970s' could be misunderstood as a decrease in time, it would be clearer to say 'lower CH4 mole fraction than BZ or CMIP6'. Similarly at lines 440 and 441, 'lower than' would be clearer than decrease Corrected.

line 484 - Put \sigma\_{age}(yr) into the figure caption, 'spread of effective age (sigma\_{age}{yr}, maximum minus minimum)' This has been added.

*line 485 - specify in the caption that \sigma\_{age}(yr) uses the right axis.* This has been added.

lines 492-494 - the way this sentence is written, it sounds like the result that consistent and accurate reconstruction is achievable only back to the mid 1970s is a universal result. Beginning the sentence with 'based on NGRIP and NEEM', doesn't make it clear enough that the situation could be quite different for other sites. I suggest reviewing this sentence.

Following comments from both referees and revision at this time, we have modified the sentence as follows:

"Based on the currently best firn CH4 data from NGRIP and NEEM, we demonstrated that consistent reconstruction of the Arctic CH4 mole fraction is achievable back to the 1950s, but the uncertainty of reconstruction is still large (>30 ppb) for the 1950s to 1970s."

line 503 rather than 'prefer', which sounds too informal, use 'are more consistent with', like in the abstract.

Corrected.

### **Response to Referee #2**

We are very grateful to the referee for thorough reading and comments/suggestions to improve the manuscript. Our responses are detailed below, in correspondence to *referee's comment*.

Umezawa et al. did significant work on improving the clarity and justifying their study. In particular, a lot of additional sensitivity analyses were conducted to clearly show why one needs to be careful when using CH4 as a "known" history to tune effective diffusivity in Greenland firn. Although the conclusion of this study is not very satisfactory (unfortunately we still don't know NH CH4 history that well), this conclusion is obtained through meticulous and comprehensive analyses of all available data and reconstruction techniques. Furthermore, this study provides a very conclusive result that shows that the NH CH4 history reconstruction from the "educated guess" used to tune NEEM diffusivity profiles (Buizert et al. 2012) is better than the one officially used for CMIP6 (Meinshausen et al. 2017). I recommend the manuscript for publication after some minor revisions (mostly stylistic). We are very pleased to see such positive evaluation of our work after the last revision. The earlier suggestions by the referee had great contribution and significantly improved our manuscript.

### **Major comments**

I think one aspect that is not sufficiently addressed is why the authors do not combine the result of NGRIP and NEEM (e.g., figure 10.c) to say something more definitive about NH CH4 history reconstruction inferred from both NGRIP and NEEM. It seems to me that there is a narrow band of CH4 history that satisfies both NGRIP and NEEM, and this at the moment should be considered our best NH CH4 history reconstruction. Umezawa et al. probably have good reasons on why they are reluctant to overinterpret this combined history, but I think the reasoning should be justified and discussed in the paper, otherwise readers like me are left wondering.

We thank the referee for this suggestion. We agree to the referee that the narrow band that satisfies both firn data corresponds to the likely range of the CH<sub>4</sub> history. Due to limitation of constraint to diffusivity in this study, we cannot identify the best reconstructed CH<sub>4</sub> history, but we consider that it would fall within the range. We have added a paragraph which explains our interpretation in the discussion section. The inserted paragraphs are as follows: "Whereas uncertainties in the reconstructions from the individual firn sites are large, Figure 11c could suggest a relatively narrow range of the CH<sub>4</sub> history that satisfies both NGRIP and NEEM reconstructions. Although the overlapping range of the two reconstruction is ~90 ppb in the 1970s, it is as small as ~30 ppb in the 1950s. This suggests that the combined NGRIP and NEEM firn data could provide a stronger constraint to the range of the CH<sub>4</sub> mole fraction e.g. 1185–1215 and 1225–1260 ppb in 1950 and 1955, respectively (Figure 10c inset). It is again noted that only the BZ scenario fall within these ranges for the period, suggesting that it is likely closer to the true atmospheric CH<sub>4</sub> history than the CMIP6 scenario for the mid 20th century."

#### **Minor comments**

Manuscript length, formatting, and redundancy I think the manuscript can be streamlined for clarity. It is clear that the writing on the new additions (seen in the track changes) is unfortunately not as concise as the original manuscript and sometimes somewhat redundant with other sections (often with the other new additions). Below, I provided some specific suggestions on parts that I personally think can be written more clearly, but this is more of a personal stylistic recommendation from me that in no way an assessment on the scientific merit of the paper. I also encourage the authors to look further and consider other parts of the manuscript can also be streamlined beyond my suggestions at their own discretion.

We thank the referee for the specific suggestions. We have made attempt to reformulate the manuscript so as to make it as concise as possible.

In its current state, the manuscript is unfortunately not very well formatted – it is hard to tell paragraphs apart because there is no indent or increased spacing between paragraphs. This made it quite challenging to review. I'm sure this will be fixed in the final proofs version. For example, the later 3/4 of page 23 is a solid wall of text with no (obvious) break in paragraphs. If there is really no paragraph breaks, then I would recommend splitting the wall of text on page 23 and 24 into several paragraphs.

We again thank the referee for thorough reading despite incomplete format. According to the comment, the corresponding paragraph has been split. We have also reformulated sections 4 and 5. We hope these reformulations have improved readability.

In the initial review, the two referees have asked for justification in why the authors are picking CH4 reconstruction above other trace gases. I personally think the authors have successfully shown throughout the paper that the motivation of this study is well justified. However, these explanations are scattered around throughout the paper. For example, one idea that is brought up in the introduction is the discrepancy between NEEM-S1 and other older ice cores, as well as BZ and CMIP6 reconstructions. Then not until page 16 that the authors run sensitivity analysis with regards to IPD to show that CH4 is uniquely uncertain. Finally, later on page 19 it is also discussed how the BZ and CMIP6 reconstructions result in different CH4 mole fraction over depth in NGRIP firn.

While I understand that all of the above are separate topics, I think the discussion to justify and show how CH4 is uniquely uncertain can be consolidated into one specific section, probably after Section 4.1 once the firn model and iterative methods are thoroughly explained. Then the authors can refer to this section from early in the introduction when they describe the motivation of the study, why they think (and how they show) that CH4 is particularly uniquely underconstrained compared to other trace gases.

We thank the referee for this suggestion. We agree that the corresponding discussion was not well incorporated in the revised manuscript after the first review. According to the referee's suggestion, we have inserted a sentence that mention to the CH<sub>4</sub>'s uniqueness in introduction and added a paragraph that describes how series of modeling results demonstrated it before the section of CH<sub>4</sub> reconstruction (section 4.7).

#### *More specific comments*

Page 3, figure 3: This figure has the NEEM-S1 scenario in red line overlapping directly with Arctic and Antarctic composite by Meinshausen et al. (2017) in green line. Please use a colorblind-friendly color palette to improve accessibility.

We have modified the colors in the figure to make it more colorblind-friendly.

*Page 4, line 77-88: Here the authors already refer to BZ and CMIP6 and discuss their discrepancy. This is (in some part) redundant with Section 3.2. where the authors formally* 

## designate the BZ and CMIP6 scenarios and also discuss their discrepancy. I think both parts can be combined.

We have shortened descriptions about scenarios in this introduction part to refer section 3.2 for details.

Page 6, Eq. 1. I find that the authors are inconsistent in sometimes providing the units of all terms in the equations and sometimes they don't. For clarity, I would recommend giving the units to all terms in the equations (even the unitless term like porosity can be explicitly specified to be unitless).

We have added units for every term of equations at its first appearance.

Pg 10, line 222: "The various diffusivity profiles were constructed by modifying the original profiles at a certain range of depths in a stepwise manner." I think this needs to be elaborated further. What certain range of depth? What is the range of the perturbation? "In a stepwise manner" how many steps? In figure 3 left panel, it is clear that the set of prepared diffusivity profiles is only perturbed between 50-65m. We have modified the sentence as follows:

"The various diffusivity profiles were constructed by modifying the original profiles at a certain range of depths in a stepwise empirical manner; depth range of the diffusivity profile key to improve reproducibility of trace gas depth profiles was first diagnosed and then the diffusivity was perturbed up and down in the depth range to the degree in which the corresponding simulated profiles do not deviate substantially."

### Pg 11, line 247-248: "[...] profile that was prepared by modifying the profile originally optimized for the CIC (Centre for Ice and Climate) model at a certain range of depths." Same questions as above. What certain depths and what is the range of perturbation away from the original Deff profile

We hope the above explanation answers this question as well.

Page 16, line 326-333. Here the authors are running a sensitivity analysis to show that CH4 is uniquely sensitive to IPD. I might be wrong, but I think here Figure 8 is cited before Figure 6 and 7. Furthermore, the authors need to state the purpose of this IPD sensitivity test before describing the result so that readers can have an idea of where the paper is going towards. In the current manuscript, the purpose of this sensitivity test is only just sort-of inferred after the result of the sensivity test is discussed in detail.

We have reformulated the section to move the discussion about IPD later so that Figure 8 now appear in a correct order. Now the discussion is made in section 4.5. We have added motivation of the sensitivity test at the beginning of the section.

Page 17, figure 6. Here the black lines are clearly stated as initial diffusivities. The initial diffusivities are inconsequential but they are distracting because at first glance of the figure, presentation-wise the black lines seem like "best-fit" diffusivities. I would suggest the author to change the initial diffusivities into something that is more visually inconsequential (for example, dashed lines) and then use solid black lines as "best-fit" diffusivities if the authors are comfortable in showing the best-fit diffusivities on the plot.

According to the suggestion, we have modified the black lines from solid to dashed in Figure 6. However, we do not show "best-fit" result because our methodology in this study cannot specify one single profile that gives a best match e.g. the lowest RMSD.

Page 18, figure 7: Please add a note on the figure caption that this is NGRIP firn. I believe that all figures should be as standalone as possible, especially in a paper where there are a lot of similar figures like this one. The light blue of BZ and dark blue of CMIP6 are also hard to tell apart and I recommend the authors to use different and more distinct color palette to differentiate the two. Same with figure 9.

The caption has been modified to clearly state that this is NGRIP firn. We have also turned color of the CMIP6 scenarios blue to red, which has been applied for Figure 9, too.

Page 22, line 440. I'm a bit confused on why the authors spend 5-6 lines to emphasize how in NGRIP reconstruction CH4 can be ~100 ppb off from the BZ scenario just to later say how highly unlikely the end-member CH4 reconstruction that is ~100 ppb off from the BZ scenario is. I understand that this is to illustrate that NGRIP is really underconstrained when compared to NEEM, but this idea can be conveyed in a more concise manner and this analysis is only somewhat tangential to the main result.

We agree to the referee that the following sentences are indeed tangential. We have deleted those sentences because similar discussion appears later.

Page 25 line 509-onwards: I think the discussion about atmospheric d13C-CH4 reconstruction and how it is uncertain because uncertainty in CH4 history belong to the introduction as one of the motivation why we want to constrain NH CH4 history specifically, not the end of the discussion as it is only somewhat tangential to the result of this study.

We have deleted the paragraph from the discussion section and instead inserted shortened sentences in introduction to explain the motivation in terms of  $\delta^{13}$ C-CH<sub>4</sub>.

#### Other comments

*Pg 1, line 15: "[...] the early 1980s, and whilst CH4 measurements from Greenland ice cores ..." I would split the sentence into two. "[...] the early 1980s. Although CH4 measurements from Greenland ice cores ..."* Corrected.

*Pg* 1, line 16-17: "In this study, we reconstruct the atmospheric CH4 for that period …" In my opinion, for brevity this sentence is not necessary and can be combined with the following sentence.

We have deleted this sentence.

*Pg 1, line 17: "We use a data set of trace gases, measured from the air trapped in firn …" to "We use a suite of trace gas measurements from firn air …" again for brevity* 

We have corrected the sentence as follows. The correction is slightly different from suggestion as a comment from Referee #1 has been also considered.

"We newly report a data set of trace gases from the air trapped in firn (an intermediate stage between snow and glacial ice formation) collected at the NGRIP (North Greenland Ice Core Project) site in 2001."

*Pg 1, line 24: "reproduce" to "reproduces"* Corrected.

Pg 1, line 26: "It is considered ...". This sentence is one of the main result of the paper, but unfortunately is written in passive voice and unclear manner. I would suggest writing something like "The atmospheric CH4 scenario used for NEEM firn air modeling is often considered the current best choice for Arctic CH4 history, but our study shows that until verified by further measurements it should not be used to tune firn models." This sentence has been revised according to the suggestion.

*Pg 1, line 23: "mid 1970s" to "mid-1970s" with a hyphen (-)* Corrected.

*Pg 4, line 68: "but their data are notably higher than the ice core data" to "higher than other ice core data" then cite the Eurocore and Site J paper.* Corrected.

Pg 4, line 86: "data set (red). Figure 1 shows that the two …" I would just say "data set (red) and inconsistent with the Buizert et al. (2012) scenario." and then delete the following "Figure 1 shows that the two …" sentence because it prematurely described the result of the study, which is repeated multiple times later. Corrected as suggested.

*Pg 4, line 92: "using the iterative" to "using an iterative"* Corrected.

*Pg 5, line 96-97: "[...] in May-June 2001. Accumulation, surface density, mean temperature ..." to "[...] in May-June 2001. Mean accumulation, surface density, temperature ..."* Corrected.

*Pg 6, line 139: "occurs with depth, which stops at the top of LIZ" to "occurs with depth and stops at the top of LIZ"* Corrected.

*Pg 6, line 150: "Namely, a trace gas flux in the firn" to "Namely, a trace gas flux (F) in the firn" add the (F)* Corrected.

Pg 7, line 168: "The closed porosity sc was calculated by empirical equation given by Schawander (1989)". First, Schwander's name is misspelled here. Also Schwander (1989) is unfortunately not an easily accessible chapter of a book (not available online through regular academic institution access). I presume the parameterizations Umezawa et al. refer to here are

scl (closed porosity) =  $s^{exp}$  [75\*( $\rho/\rho cod -1$ )] for  $\rho <= \rho cod$  (close-off density) scl = s for  $\rho > \rho cod$ 

*I would recommend Umezawa et al. to just show the equations or cite other papers that show these equations together with Schwander et al. (1989)* 

We thank the referee for this suggestion. We also found an error in our original manuscript about the prosity parametarization for NEEM, which has been corrected as follows. The previously published modeling on the NGRIP firn air (Sugawara et al. 2003; Ishijima et al.

2007) employed the porosity parameterization by Schwander (1989), and it is also the case in this study (e.g. the one correctly written by the referee). For NEEM, we employed the parameterization used by Buizert et al. (2012) i.e. the equation proposed by Goujon et al. (2003), for consistenvy with the previous NEEM firn air study and validation of our model. We confirmed that the simulations presented in this study were all performed accordingly, and added both equations in the text with references. We note that the choice of closed porosity parameterization should not have discernible impacts on the atmospheric reconstructions because the diffusivity profile is tuned for the respective cases.

Pg 8, line 195: "Therefore, CH4 is the only compound, with an available atmospheric history, which shows a clear disagreement, thus highlighting ...". To improve clarity I would just say something like "Thus CH4 is unique because currently it has two diverging synthetic histories that are only loosely constrained by observational data."

The sentence has been corrected according to the suggestion.

*Pg 10, Eq. 4. "p" is obviously pressure but not formally defined within the paper.* "p" has been now defined.

*Pg 13, line 284: "A new atmospheric CH4 scenario was constructed by assigning the observed CH4 mole fraction, at each depth" to "[...] at each sample depth"* Corrected.

*Pg 19, line 385-386: "[...] produced enlarged overestimate in the LIZ (>63m) in the modeled profiles ... ". "enlarged overestimate" is redundant. I would just say "[...] results in overestimation of CH4 mole fraction in the LIZ (>63m)"* Corrected as suggested.

Pg 22, line 440: "[...] suggested that the CH4 mole fractions over the period 1950-1980 could be decreased in comparison to the original BZ scenario." For clarity, I would recommend changing "decreased in comparison to" to "~100 ppb lower" Corrected.

*Pg 22, line 441: "The decrease of up to 100 ppb from the BZ [...]" also for clarity, to "The ~100 ppb lower CH4 mole fraction from the BZ [...]".* This sentence has been deleted to incorporate the earlier comment.

Pg 26, line 518: "but we regrettably report that the reconstruction of d13C-CH4 has not been possible despite our best modeling efforts (not shown)."I think this sentence does not add anything to the manuscript, as neither the d13C-CH4 and dD-CH4 data nor "our best modeling efforts" are presented in the paper. I would suggest removing it entirely. This sentence has been deleted according to the suggestion.