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).

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.

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 assesment 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.

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 ¾ 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.

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.

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.

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.

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).

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.

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 D_{eff} profile

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.

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.

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.

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.

, and this analysis is only tangential towards the main result

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.

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 …"

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.

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

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

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."

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

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.

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.

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

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

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"

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)*

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*(ρ/ρcod -1)] for ρ <= ρcod (close-off density)
scl = s for ρ > ρcod

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)

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.”

Pg 10, Eq. 4. “p” is obviously pressure but not formally defined within the paper.

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”

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)”

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*”

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 […]”.

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.