Thanks a lot to the authors for their considerable revisions of the original manuscript. In my view, the new version is a great improvement. The structure gained clarity and the refined focus on the time scale makes the entire paper appear more concise. Again, this manuscript might require some language editing or input by the native speaker of the author list. On top of the good development, I recommend major revisions on several aspects of the new manuscript. These include i) the description of the analytical system and its uncertainties, ii) the use of IPD-1 and IPD-2, iii) Section 3.3, Comparison with late Holocene variability, iv) conclusions, v) Supplements.

i) the description of the analytical system and its uncertainties

I find the part on the analytical uncertainty of the methods section extremely hard to follow and even after careful re-reading, I am not sure what data were used to calculate the stated averages, pooled standard deviations etc. The authors have chosen to publish the data paper before their new method is published in a separate publication. In my point of view, this requires a sufficient description of method and uncertainty if it is claimed that the new method is superior to existing methods, even if a dedicated publication follows. It seems to me that the authors should increase their resistance to the temptation to choose statistical methods that understate the uncertainty of their analytical system, as this leaves the reader rather suspicious.

Please make sure it is well described how you achieve the values you present and what data they were calculated from. For example on page 42, line 6-7 of the document including the tracked changes, you state a standard error of the mean (SEM) of 2.0±1.0 ppb. To my understanding, the SEM is the uncertainty measure, it has no uncertainty itself. What has been calculated here?

In general, SEM's may be used as measure of uncertainty when a sufficient number of samples have been measured. In this manuscript, the authors seem to apply SEM to sample sizes of n=2. If the authors have a good reason to use SEM for n=2, they should clearly state why and account for the obvious lack of sample size (student's t). I feel very strongly that using SEM on n=2 is not OK and needs changing in every uncertainty calculation (e.g. p. 42, L10). Presenting a formula may help clarify what has been calculated. An understandable and ethically correct uncertainty demonstration will provide more credibility to the technical aspects of this manuscript than unrealistically small values. If the system is a new game changer, it should be possible to show outstanding performance using appropriate statistical methods.

This manuscript version has a lot of emphasis on the use of bubble-free blank ice to determine the system blank. The authors state that based on their method to produce the blank ice, they can exclude that CH_4 is introduced with the blank ice. From personal experiences, I wouldn't trust that statement. Dedicated blank ice tests including several labs showed that while it is easily possible to produce bubble free ice, there is no such thing as gas-free blank ice. The gas content of the blank ice varied both between the labs and between the batches produced in each lab. All labs used deionised water and pumped on the water for 90 minutes and froze large crystals bottom up… While blank ice allows for a valuable tests, great care has to be taken in the use and interpretation of the results. Based on 4 daily blank ice analyses, the authors state an intra-day SEM of 2.0±1.0 ppb and a variation of the inter-day mean between 5 and 15 ppb. I think the authors intention to present these data is to show that even if the inter-day variation is as large as 15 ppb, the daily blank can be quantified with low uncertainty (SEM 2.0±1.0 ppb) and a correction can therefore be accurately determined. I find this hard to understand. If I chose 4 random values between 5 and 15 (as representative for the large variability of the inter-day blank) and calculate the SEM for these 4 values, the SEM is <2.5. The authors state that the blank estimate of OSU has a high uncertainty of 10% due to the small peak size, but don't seem to think this is a problem when quantifying the blank ice contribution with higher relative uncertainties. Please clarify.

I would think that introducing a known amount of air with a known CH4 mixing ratio into the melt chambers and to analyse that air in the same way as an ice core sample could provide a useful measure of analytical error. This could be done in a dry chamber and over melt water to test for differences. Such experiments could be useful to determine the **accuracy of the measurements**.

The authors use table R1 to show the reproducibility of the analytical system. I find this is a very useful way to state the potential uncertainty/reproducibility of the system and would suggest to improve the clarity and possibly give this more weight in the formulation of the uncertainty estimate. On p. 42 L. 12, the authors refer to the data in table R1 and state, that the differences between the averages from the 1st and the 2nd measurement pair is on average 1.9 ppb. Even though this is nice, it ignores the uncertainty of each single measurement as well as the differences within each sample pair. Again, the authors do not clearly state how the uncertainty in R1 is calculated for 1st and 2nd measurement, is it SEM again for n=2 or 1σ? The authors provide one depth interval for all for samples in table R1, where the depth is sometimes stated to the mm. Are really all four samples from the same mm depth? In my view, this table should include all 4 measurements per depth level. You can then add columns for 1σ or SEM for all four, and calculate the pooled standard deviation (PSD) of the whole lot afterwards as total uncertainty estimate for **measurement precision**. I am not sure if this is what the authors describe on p. 42, L. 13 (PSD of 1.4 ppb)? Did you calculate the PSD based on 4 individual measurements per depth or based on the averages from 1st and 2nd measurements? If this is the "final" uncertainty that the authors intend to state for their analytical system, they should clearly say so in a dedicated sentence, not hidden in a sentence and furthermore in brackets.

P. 42, L 30: You mention the difference OSU-SNU of 3ppb and present a hypothesis on what might have caused this offset. However, this dilemma shows that an analytical uncertainty of 1.4 ppb cannot include the entire uncertainty to the NOAA scale. It would be great to see this resolved, i.e. precision and accuracy, or at least clearly discussed if unresolvable.

On p. 42 L. 34, the authors state that their new Siple Dome data have the third highest temporal resolution. However, they don't say how high it is, as they do for the others. Please quantify, otherwise your statement of 3rd highest is meaningless.

<u>ii) Use of IPD-1 and IPD-2</u>

The authors calculate IPD's based on NEEM discrete and Siple discrete (IPD-1) and NEEM discrete and WAIS continuous (IPD-1). Unfortunately, the authors do not describe how the uncertainty envelope is calculated. Furthermore, you do not state the temporal resolution of the Siple data, as compared to the WAIS continuous (p. 42, L. 34). There are a few periods between 11.4 and 10.5 ka where WAIS and Siple are different by ~20 ppb, which produces the different IPD shapes. While WAIS continuous includes several 100 data points (mean resolution of 2 years), I manually count ~15 in the Siple record in Figure 3. Moreover, both the WAIS and NEEM records share several features, that are not resolved at all in the Siple record. It seems to me, that focussing your IPD interpretation on IPD reconstructions that are best constraint makes most sense. I understand the temptation to focus your analysis on the new data. However, I would strongly recommend to shift IPD-2 from Supplements into the main figure and to use IPD-2 for interpretation between 11.5 and 10.0 ka. This will still leave you with the IPD increase until 9.5 ka, but is better constraint and has less artefacts. In the box model, this would create rather stable emissions from tropics until 10.0 and a decrease thereafter. To me, this would seem the best IPD reconstruction possible.

iii) Section 3.3, Comparison with late Holocene variability

This section feels completely unattached to the rest of the manuscript. The figures are described but any quantification is hard to follow. The section title is meaningless and confusing.

iv) Conclusions

Please quantify your statements and avoid generalisations, such as high resolution (how high), agree well with previous measurements (how well), four CH4 minima (how big are anomalies, how long in duration), first reporting of IPD increase from 11.5 to 9.9 (how much increase), elevated emissions from NH extra-tropics (how much did emissions from NH extra tropics increase), RMS amplitude smaller (how much) and what does this even mean in conclusions?

v) Supplements

I would like to encourage the authors to shift all these important information on analytical reproducibility, gas age uncertainty, IPD-2, age model effect on previous gas age scale and box model data in the main text. Figures can be reduced size, e.g. one column etc, but including those information will increase value of paper! I don't recall being frequently referred to supplements in the main text. Having a lot to read every day, I would probably never get to read supplements that aren't constantly advertised in the main paper, unless this is the most critical paper for my current work.

General comments

P36 L11: human influence was substantially smaller

P37 L1: is geological CH4 really the 2nd most important natural source? Not sure if most recent 14C-CH4 tells the same story. The authors might want to consider to tone this statement down a bit.

P37 L4: The CH4 flux from the ocean to the atmosphere

P37 L7: not to forget CH4 itself

P37 L15: …during the past climate changes… could you be more specific on time scales?

P37 L15-16: This sentence should be in the section that describes the box model method

P37 L18: polar firn and ice

P37 L20: cite Loulergue 2008

P37 L29: Greenland temperature change

P37 L30: …around Greenland is linked to the….

P37 L32: with abrupt Northern Hemispheric warming during DO…

P38 L8-11: Please develop this sentence, I cannot follow the line of argumentation

P38 L25: cover only a part…

P38 L34: To my knowledge, there is no plural for “ice”, “ices” doesn’t exist. Please correct here and everywhere else.

P38 L37: On the other hand…

P39 L20: record from the early Holocene and investigate…

P40 L30: starts below 400 m and continues to the bottom of the core at 1004 m

P41 L3: SEM of n=2???

P41 L25-27: How do you expand the gas into the GC and ensure 100% sample transfer into the GC? Do you flush the headspace with He? Could this He create a blank that varies with He cylinder?

P42 L6-7: SEM of 2.0±1.0ppb… please explain your calculations

P42 L10: SEM of n=2???

P42 L12: Everything will agree well if you average often enough. Please develop transparent and unbiased approach.

P42 L13: If this is your final uncertainty estimate, please state this in a clear and dedicated sentence.

P42 L17: You might want to consider leaving the “M” at the beginning of the sentence.

P42 L34: Please quantify the temporal resolution of your Siple Dome data

P44 L31: from Asian and Amazon wetlands

P47 L29: Another argument to use NEEN discrete is that these data were measured at OSU as well. No?

P48 L9: Please provide a transparent description of how you calculate 1.4 ppb.

P48 L13: “regarded as more accurate”, not exactly a scientific term. Haven’t the uncertainties of 1.4 ppb and 1.5 ppb just been stated for Siple and WAIS, respectively, in the previous sentence? I strongly recommend to re-think the presented uncertainty model. Accuracy is part of the uncertainty but doesn’t seem to be considered in the uncertainties presented in this manuscript. If ice core specific accuracy problems infiltrate IPD analysis, how reliable is the magnitude of IPD reconstructions?

P48 L17: Describe the calculation of the envelope

P49 L3: State lifetime and transport time you assume in model

P49 L7: 15 Tg

P49 L8: IPD-2, 134 to 115 Tg, I don’t see anything >125 Tg in Figure S3. I am confused with some of the quantifications in the following text, often, the stated numbers don’t seem to match the values in the figures.

P49 L9: What trend? I can't see a trend in Fig. S3.

P49 L12: Where are these numbers from? Fig S3? Fig 4?

P49 L12: The minima at 10.7 ka is only a feature in IPD-1, not in IPD-2. Again, this is misleading, as IPD-1 is based on much lower temporal resolution. Please use IPD-2 where possible.

P49 L16: "…from 29 to 35% during the 11.5 to 10.0 interval." I cannot even see a value <30% or >34% in Fig. S3, even including the envelope.

P49 L7-16: This entire section needs major revision. The numbers don't seem to match the presentation in the figures, the description jumps back and forth between Figures in supplements and main text, the text flow makes it hard to understand.

P49 L22: What conclusion?

P49 L30: Your 13 Tg estimate is based on IPD-1, which matches the 8.2 Tg within uncertainties. What value would you get from IPD-2?

P51 L4: Please find a meaningful section title

P51 L27ff: please generalise less and quantify more.

P62, Figure 2: Please add next to axes what these proxies actually show, e.g. warmer, wetter, colder dryer with arrows.

P64, Figure 4: Please add reconstructions based on IPD-2

P65, Figure 5: Please synchronize x-axes directions in top panels

P66 L2: Uncertainties or errors?

P67ff: Please include supplements in main text