We thank Referee #2 to support publication of our paper. In the following, we address the referee's comments in the order of appearance.

Page 2, line 15-16 (of the manuscript with no track changes revisions)

Comment from referee: **I suggest to rephrase "it is extremely difficult to separate the impacts of anthropogenic increases in CO_2 on carbon sinks from the impacts of global warming or increased CO_2 concentration on these sinks.". This statement was unclear for me**

Author's response: We have change the statement to make it clear.

Author's changes in manuscript: We have modified the sentence at page 2 lines 15-17 to: "*Additionally, temperature and CO_2 have both increased almost continuously through the 20th century, making it difficult to separate the impacts of CO_2 on carbon sinks from the impacts of temperature increase on these sinks.*"

Page 3, line 7 (of the manuscript with no track changes revisions)

Comment from referee: **Rhodes et al., 2016 (Climate of the Past) include more CH4 data from Greenland.**

Author's response: We thank the referee for this important comment, which has helped us address an important weakness of our study when comparing the Law Dome CH_4 records with CH_4 records from other sites. As the referee points out, Rhodes et al. (2013) shows the GISP2 CH_4 record as well, published in Mitchell et al. (2013). We have added this record to figure 3c (now 4c) and have cited Mitchell et al. (2013).

Author's changes in manuscript: We have modified the sentence at page 3 lines 8 adding "… and from GISP2 (Mitchell et al., 2013)…" and have added the GISP2 CH_4 record to figure 3c (now 4c).

Page 5, section 2.1

Comment from referee: **use the same unit for all accumulation data, for homogeneity**.

Author's response: We agree with the reviewer that it is better to have all accumulation rates expressed with the same units. We have decided to convert all accumulation rates to kg m^{-2} yr^{-1}, which is also consistent with the units used in Figure 1. We have also left the accumulation rate at DE08/DE08-2 as meters of ice equivalent per year in brackets, as it is nice to give a better feel to the non-specialist reader for what the accumulation looks like.

Author's changes in manuscript: We have expressed all units of accumulation rate at page 5 in kg m^{-2} yr^{-1}.

Page 6, line 23 (of the manuscript with no track changes revisions)

Comment from referee: typo "I think".

Author's response: fixed

Page 11, line 20 (of the manuscript with no track changes revisions)

Comment from referee: **I understand that the atmospheric mixing between Northern hemisphere and Southern hemisphere is fast enough so [CH4] would exhibit almost simultaneous trend in both hemispheres. Here the shift in in LIA CH4 decrease seems to be about 40 yrs (Fig. 3). Similar shift**

seems to exist at the onset of the industrial period CH4 increase. Can we explain such shift with Age Scale uncertainty? Maybe discuss this shift by providing more quantitative estimation.

Author's response: We thank the reviewer for this important comment, which, together with a similar comment of reviewer 1, and has helped us improve this part of the manuscript. We have updated the original NEEM CH_4 gas age scale published in Rhodes et al. (2013), with the new NEEM ice age scale published in Sigl et al. (2015) and the revised Δage (gas age-ice age) after Buizert et al. (2014). The different timing is now partially resolved (15 years now, instead of 30 years). Having added the CH_4 GISP2 record, we have commented on how the Northern (GISP2) and Southern (WAIS) records have been synchronised by Mitchell et al. (2013) based on the following reasoning: "The multidecadal events observed in both ice core records must have occurred simultaneously since the durations of the events were much larger than the atmospheric mixing time (~1 year).". As already written above, there are multiple possible reasons, associated with the differences of the sites and/or the sampling resolution of the records, to explain the discrepancy found between Northern (NEEM) vs Southern (Law Dome/WAIS) Hemisphere CH4 records, as discussed in the following:

- Age scale issues, as suggested by the referee. Usually, age scale issues are not very significant for the last centuries. Over such a short time scale, the age of the ice is established through annual layer counting. However, there can be significant uncertainty associated with the delta_age (ice age-gas age).
- Smoothing of the atmospheric signals due to air diffusion in the firn open porosity. As shown for the Law Dome CO2 record smoothed through the gas age distribution of DML in Figure 2 of Rubino et al. (2016), the smoothing causes a shift in age of the max and min values of the LIA CO2 decrease. For the Law Dome vs DML comparison, the shift amounts to a few years. NEEM has higher accumulation than DML, so firn smoothing should cause an even smaller shift.
- Inadequate sampling resolution. The NEEM CH4 records plotted in figure 3 (now figure 4) is a 5 year average of the high resolution record published in Rhodes et al. (2013). The Law Dome CH4 record has a lower sampling resolution (5 datapoint over a 30-year period of decreasing CH4 concentration, so on average 1 datapoint every 6 years). By increasing the sampling resolution, it is possible that the time shift between the NEEM and the Law Dome CH4 event decreases.

A thorough investigation of the cause of the differences in the LIA CH4 event in the Northern (as recorded in NEEM) and the Southern (as recorded in Law Dome/WAIS) Hemisphere is out of the scope of our paper. However, in the future, this discrepancy should be resolved to obtain a precise synchronisation of all ice core records available over the LIA.

Author's changes in manuscript: We have modified the discussion at page 12 line 18-27: "*The age scale of the NEEM CH_4 record published in Rhodes et al. (2013) has been revised with the updated ice age scale published in Sigl et al. (2015) and the new estimate of Δage provided by Buizert et al. (Buizert et al., 2014). Mitchell et al. (2013) have synchronised the GISP2 CH_4 record with the WAIS CH_4 record to investigate changes of the Inter Polar Difference in the Pre-Industrial based on the reasoning that "the multidecadal events observed in both ice core records must have occurred simultaneously since the durations of the events were much larger than the atmospheric mixing time (~1 year)" (Mitchell et al., 2013). The NEEM CH_4 record has not been synchronised with the others, and there are multiple possible reasons, including age scale issues, different smoothing of the atmospheric signals at the different sites and inadequate sampling resolution, to explain the*

discrepancy found between the NEEM and the GISP2/Law Dome/WAIS CH_4 records during the LIA. A thorough investigation is out of the scope of this paper, but, in the future, this discrepancy should be resolved to obtain a precise synchronisation of all ice core records available over the LIA.".

Figures:

Comment from referee: **I find the figures difficult to read: I would advise to increase size for labels and titles.**

Author's response: We have chosen the size for labels and titles based on the font size in the template provided for ESDD articles for consistency. We would be happy to increase the size if the editorial office allows us to do so.

Author's changes in manuscript: We have not changed the size because we need approval from the ESDD editorial office.

Figure 4:

Comment from referee: **this figure highlight two important past findings (Rubino et al., 2016 ; Ferreti et al., 2005). It also includes data that can potentially be relevant for further studies and interpretations. I am not sure these data (b, d, and e) need to be plotted; likely description in the manuscript is enough. If the authors want to keep these data as part of plot 4, I would recommend to clarify the figure, e.g. the panel b and c can be shifted, so the figure reports first ice core data, and second complementary climatic data**.

Author's response: We agree with the reviewer that swapping figures 4b and 4 c (now 5b and 5c) makes sense for clarity of presentation, but also under the point of view of citation order.

Author's changes in manuscript: We have modified figure 4 (now figure 5) by swapping figure 4b and 4c (now 5b and 5c). We have modified the figure caption and the citation to figure 5b and 5c in the text accordingly.

Technical comments related to Supplement:

Comment from referee: **My main concern is about the way uncertainties are calculated by multiplying the blank uncertainty with a factor u.f. I find this process complex, not fully understandable for data users, and maybe operator-dependent. To me, the blank uncertainty is independent from other sources of uncertainties (e.g., dispersion of results observed for replicated measurements on of the same sample). Independent uncertainties when propagated do not multiply each others**.

Author's response: We thank the referee for the detailed comments provided on the way we calculate the uncertainty associated with the results we produce. We agree with the referee that the process we have developed is quite complex. Unfortunately, our rule-based selection procedure needs to take into account all the little pieces of information that are normally evaluated by an operator based on the experience matured over 20+ years at ICELAB-GASLAB. Our aim, though, was to develop an automatic system, which would minimise operator-dependent judgements, to make it consistent in time and between different operators. We understand that it is very difficult for an external data user to understand the details of the procedure we have developed. We have tried to make the description of it as clear as possible.

As the referee states, the blank uncertainty is independent from other sources of uncertainties, such as those mentioned by the referee. We have propagated the uncertainty using the guidelines from

the Joint Committee for Guides in Metrology (2008). The uncertainty associated with the blank correction dominates over the other sources of uncertainty.

Author's changes in the Supplement: We have added a citation to the Joint Committee for Guides in Metrology (2008), Evaluation of measurement data—Guide to the expression of uncertainty in measurement, Bur.Int. des Poids et Mesures Pavillon de Breteuil, France, at page 5. We have also added a couple of sentences at page 4: "*The uncertainty associated with gravity and the diffusion corrections are negligible compared to the uncertainty associated with the blank correction*" and at page 5 "*However, the variability associated with replicates of the same sample is negligible compared with the uncertainty associated with the blank correction*." to explain that the blank uncertainty dominates over the other sources of uncertainty.

Comment from referee: **We can observe here that for a sample where u.f. = 1 (i.e., qf and mq = fair or good), the data uncertainty is reduced to only the blank uncertainty, ignoring for example that different replicated measurements of the same sample will likely not be exactly all the same**.

Author's response: The referee is right: different replicated measurements of the same sample will not show the same result. However, the variability associated with replicates of the same sample is negligible compared to the uncertainty associated with the blank correction, which quantifies the variability associated with multiple tests run together with the samples. Additionally, because of the way the ICELAB-GASLAB system has been conceived (measurements of different species on the same ice sample), replicates of the same sample measured for the same species are so rare that there is no uncertainty associated with replicates for the vast majority of the samples analysed.

Author's changes in the Supplement: We have modified a sentence at page 3: "*Only a few measurements of concentration have been replicated. For isotopes, there has never been enough air to measure replicates*". We have also modified a sentence at page 4 of the Supplement: "*Because of the small size of ice core samples generally available and the need for large air volume to measure multiple species, the vast majority of samples have no replicated measurements.*" and another sentence at page 5: "*However, the variability associated with replicates of the same sample is negligible compared with the uncertainty associated with the blank correction*."

Comment from referee: **The u.f. factor includes many parameters (flags, and criteria associated to weights), and some of them are not related to uncertainty. As an example, the first parameter is "melt layer". A melt layer can, e.g., results in high methane concentration (due to in situ production), but the measurement uncertainty of such high concentration should not be different from a regular sample. Just higher concentration will be measured. A melt layer sample does not have the quality required for reconstructing past CH_4, but its measurement could be of great quality! Overall, I would advise that the authors identify more clearly what causes uncertainties, instead of considering everything.**

Author's response: Our database is in constant development and it is possible that we will find better ways of including factors affecting the total uncertainty in the future. However, for now, when a melt layer is found, a sample is rejected because, as the referee suggests, melt layers are known to alter the concentration of the species we measure.

Author's changes in the Supplement: We have added a sentence at page 3: "*For example, a melt layer is classified as evidence of a fatal problem of sample quality and provides a q.s. = 3 and a q.f. = "reject"*."

Comment from referee: **Some of the flags and criteria associated to weights seem subjective, and maybe operator-dependent in their evaluation (at least this is what I feel when reading the Supplement).**

Author's response: The thresholds for flags and weights have been decided based on a "calibration". The idea was to replicate with an automatic procedure what had been done in the past manually. The aim was to make the procedure consistent over time and as independent from the operator as possible.

Author's changes in Supplement: We have added a line at page 3: "*The flagging and weighting thresholds are tuned by calibrating the rule-based selection on the manual selection used before the database was conceived. The idea was to replicate with an automatic procedure what had been done in the past manually. The aim was to make the procedure consistent over time and as independent from the operator as possible. In summary, the rule-based selection converts qualitative judgments on the robustness of sample preparation, extraction and analysis into quantitative scores in order to consistently select/reject the results and quantify uncertainty.*"

Comment from referee: **When all criteria reports "reject" (i.e., u.f. = 4), the data is not rejected, but the uncertainty is increased more. It seems to me that these data should be excluded (as suggested by the wording "fatal problem").**

Author's response: When q.f. = "reject", the sample is marked with a rejection flag. However, the result of that sample is retained in the database and a u.f. = 4 is associated with it. The reason why we have decided to do that is to give enough flexibility for the operator to overrule the automatic selection, providing a convincing justification that the operator's choice is appropriate. If a data user wants to use that sample, the uncertainty associated with the results is very high (4 × the blank uncertainty and then propagated as explained).

Author's changes in the Supplement: We have added a line at page 5: "*When q.f. = "reject", the sample is marked with a rejection flag. However, the result of that sample is retained in the database and a u.f. = 4 is associated with it. The reason why we have decided to do that is to provide enough flexibility for the operator to overrule the automatic selection when there is a strong reason (e.g. new insight developed over time). If a data user wants to use that sample, the uncertainty associated with the results is very high (4 × the blank uncertainty and then propagated as explained below).*"

Comment from referee: **What are the typical blanks observed, and typical blank uncertainties observed?**

Author's response: This comment, together with a similar comment from referee 1, has helped us provide important information, which was missing from the previous version of the manuscript. We thank the referees for that.

Author's changes in the Supplement: We have added a line at page 5: "*The blank correction can vary significantly depending on the conditions of the extraction line and on how experienced the line operator is. Typical values are within the following ranges: 0.5-1.5 ppm (uncertainty 0.5-2 ppm, 1σ) for CO_2, 3-10 ppb (uncertainty 3-15 ppb, 1σ) for CH_4, 0.5-3 ppb (uncertainty 0.5-4 ppb, 1σ) for N_2O and 0.03-0.1 ‰ (uncertainty 0.04-0.13 ‰, 1σ) for $\delta^{13}C$-CO_2.*" and a sentence at the end of page 4 of the Supplement: "*The uncertainty associated with gravity and the diffusion corrections are negligible compared to the uncertainty associated with the blank correction*"

Comment from referee**: I am not convinced that CO concentration is a good tool to evaluate the quality of a measurement (similarly to "melt layer", see before), or the quality of a sample. CO can**

be produced by chemical processes (the authors mention biological production of CO, citation is missing for that), but to my knowledge no collocated productions of CO and, e.g. CO_2, CH_4 or N_2O have been reported so far in ice cores. The processes involved could be different, and a sample compromised for CO can be of good quality for others analyses. Ambient CO is often higher than what is in ice core bubbles, but this is also clear with CH_4.

Author's response: High CO is quite a reliable indicator of contamination by the ingress of lab/storage facility air into the sample, or by leaks during processing the sample. High CO has also been reported together with high CO_2 and low $\delta^{13}C\text{-}CO_2$ values in a study of Antarctica vs Greenland CO_2 by Francey et al. (1997).

Author's changes in the Supplement: At page 2, we have cited Francey et al. (1997) showing evidence of in-situ production of CO in Greenland ice. We have modified the associated sentence: "*High CO values have been measured in Greenland ice (Francey et al., 1997) together with high CO_2 and low $\delta^{13}C\text{-}CO_2$, suggesting in situ oxidation of organic material deposited on the Greenland icecap (Francey et al., 1997)*".