Final Response to RC2 on "Results of a Long-Term International Comparison of Greenhouse Gas and Isotope Measurements at the Global Atmosphere Watch (GAW) Observatory in Alert, Nunavut, Canada"

By Referee #2: Martin Steinbach (received and published: 14 Jul. 2023)

Thanks very much for the editor's effort to coordinate the reviewing and for the reviewer's constructive feedback and comments. We will answer the questions and address the concerns point by point raised by *Referee #2* below in the format of "reviewer's comments/ author's responses".

General comments:

The manuscript presents a comprehensive analysis of results from a co-located long-term flask sampling comparison in the high Arctic. The flask samples cover a period of 17 years. Seven international institutions, all well renowned in the atmospheric monitoring of greenhouse gases, participated in the comparison. The paper is well structured and written. It is very comprehensive and provides an excellent overview of what can be achieved in terms of precision and accuracy by the leading experts in greenhouse gas observations. The manuscript includes an excessive number of tables. The authors may consider moving Tables 6 to 22 into the supplementary material for the sake of conciseness of the main part.

Thank you for the suggestion. Tables 6-22 have been moved to the supplement in the revised version.

The paper is definitely within the scope of 'Atmospheric Measurement Techniques'. I have only a few minor comments and suggestions for improvement, all listed below.

Chapter 2 briefly summarizes the different sampling approaches and the analytical techniques used. I do not suggest to elaborate on many additional details here since relevant references to the existing literature are provided in the bibliography. However, I was wondering if some specific sampling and analytical methods do systematically contribute to the differences described in the results section. Factors could be the pressure in the flasks during filling, drying/no drying/the way of drying (cryo vs. Mg (ClO4)2) of the air prior to sampling, the time between sampling and analysis (which I assume can be rather long considering the remoteness of the site; rough numbers could be given), effects of single-stopcock vs. double-stopcock flasks (i.e. flushing of the flasks), or (changes) in the analytical technique. Relevant specifications should be mentioned accordingly. What are the most striking lessons-learnt after all these years?

Oftentimes, we discovered that the causes behind the observed differences were not readily apparent and couldn't be easily explained. As a result, we chose to primarily concentrate on the outcomes rather than delving extensively into the underlying reasons. Nevertheless, we did incorporate explanations when they were relevant. For instance, the change from $Mg(ClO4)_2$ to a cryo-cooler for MPI-BGC's system in 2015 was included.

For example, the change of the analytical technique from NDIR to CRDS at SIO in 2012 seems to be observable in Fig. 1 b (panel with the purple symbols, less noise in the difference in the more recent data).

Thank you for your careful observation. A comment has been added accordingly.

Chapter 3 is rather descriptive. There, reference could be made to the different sampling and analysis methods (see my comment just above) if they can explain some of the observed differences. See comments above.
Thanks for your suggestions. We've moved Tables 6-22 to the supplement and modified the contents accordingly in the revised version.

Due to the characteristics of ALT, almost exclusively clean-air samples were analyzed and compared. Can you make a (qualitative) statement about the to-be-expected agreement under more polluted conditions?
In principle, the measurement agreement should be better, as polluted air masses would result in a higher signal to noise ratio.

At least for some of the 17 years, continuous in-situ measurements (for some of the species) were performed. How do the continuous data compare with the flask data? Please add at least a short (quantitative) statement, if applicable.
We can certainly provide you with some information, even though it might not be within the scope of this study. The median CO_2 *in- situ* vs NOAA flask data comparison for 1999-2016 is 0.025 ppm (-0.298, 0.38) at the 10^{th} and 90^{th} percentile. For CH4, the median difference is it's -0.05ppb (-3.35, 4.09) at the 10^{th} and 90^{th} percentile.

Are the sampling procedures at MLO and CGO identical to the described methods for ALT?
Further information about the sampling at MLO and CGO has been integrated into the respective laboratory sampling descriptions. The sampling system used by SIO remains consistent across all sites, while CSIRO utilizes different types of samplers and dryers, including $Mg(ClO_4)_2$. Additionally, some of NOAA's samplers , like the one at CGO, incorporate dryers, which is not the case at ALT or MLO.

Specific comments:

Line 182ff.: low operating costs are mentioned as a benefit of the flask sampling program. I tend to agree in terms of investment, but the logistics, shipping costs, and manpower requirements for the off-line analysis are significant. Can you somehow quantify the effort over all these years? 8'000 flasks are mentioned in the abstract but I assume that there are also other quite impressive numbers. Numbers of boxes shipped to ALT? Total distance travelled by the flasks? … Any idea about the manpower required for flask preparation, sampling, and analysis? Total costs of the endeavour? No reference is made to any funding agency in the acknowledgements. This makes it even more impressive (that you were able to keep up this comparison exercise over all these years).
The annual shipping cost covered by ECCC for all flask programs was ~$30k annually. This encompasses round trip travel to and from Alert, as well as the return shipping to participating laboratories. It is important to note that this does not include shipping expenses to Toronto, which are covered by the participating laboratories themselves.

A very rough estimate for Alert site operator labour is around 4 person-hours at $25/hr, totaling $200 per week (covered by ECCC). Additionally, there are analysis expenses incurred at each of the laboratories.

The flasks travel a distance of over 9,000 km on their round trip journey from Toronto to Alert and back. When factoring in the trip to the participating laboratories, the distance increases significantly (with an additional ~33,000km for CSIRO)! Managing the annual supply of flasks to be transported to Alert requires substantial logistical coordination, especially as they are carried via the sealift organized by the military.

Line 281 – 283: "Flask air samples were collected at Alert during persistent southwesterly wind conditions, when wind speeds were greater than 1.5 m s-1 for several hours prior to sample air collection." In other words, there was no sampling when the wind conditions were inappropriate? If so, sampling was retried as soon as the conditioned matched the requirements, or did you skip this sampling and stick to the regular (weekly, bi-weekly) schedule (the following week)?
That's correct. They waited for the appropriate conditions. If the planned sampling day (Wednesday) proved unfeasible, they would postpone it to the following day. If the conditions remained unfavorable by Friday, sampling was conducted regardless, and a note was written indicating that conditions weren't ideal.

Lines 344 – 345: "When meteorological conditions are favorable for sampling, the NOAA sampler is taken outside and several meters away from the GAW laboratory to collect the air samples." What happened when conditions were unfavorable? Flasks for the other institutions were filled but no NOAA flasks? How is "favorable" defined?
The conditions had to be within the parameters stated above, for all flasks to be sampled. No flasks were sampled on their own. We have modified that section accordingly for clarity.

Lines 367 ff.: Is it correct that SIO flasks are not pressurized at all (only the valve of the evacuated is opened for filling)?
Yes, the flask is under vacuum, and the valve is opened for filling.

Line 387: "… pressurized, vented and re-pressurized …" To which pressure?
Most flasks are filled to an overpressure of 5psi (NOAA) or 15psi (everyone else, except SIO).

Line 398 "double-stopcock flasks" vs. line 405 "double-valve stopcock flasks". Please harmonize the wording.
The wording has been fixed.

Table 4 reads "GC/ECD" etc. while "GC-ECD" etc. is used in the text.
These expressions have been fixed.

Calibration scales given in Table 4: most scales are WMO/GAW scales hosted by the GAW Central Calibration Laboratory at NOAA (X2007, X2004A, …); SIO uses X08A, which does not seem to be an official WMO/GAW scale. Is there a reference that documents the WMO CO2 X2007 to X08A consistency?
We've included a new reference that documents the consistency with the WMO scale.