Author comment on “Evaluation and optimization of ICOS atmospheric station data as part of the labeling process” by Camille Yver-Kwok et al.

Anonymous Referee #2
Received and published: 29 August 2020

We thank the reviewer for their positive review and their constructive and helpful comments. We answer them below, highlighted in bold and italic.

Accurate measurements of greenhouse gas (GHG) concentrations in the atmosphere is the first step to understanding and mitigating the impacts of climate change, and a full assessment of the temporal and spatial variabilities and trends in these concentrations requires a wide network of observations. The GHG measurement community has worked hard to ensure that the GHG measurements operated around the world are “compatible” (data meet quality standards necessary for addressing the scientific questions and can be compared under common calibration scales) through the efforts such as the recommendations of the WMO/IAEA GGMT meetings (most recent report at; https://library.wmo.int/index.php?lvl=notice_display&id=21758#.Xzxuv1QzaUk). But more focused efforts such as the Integrated Carbon Observation System (ICOS) present an opportunity to improve on these practices and derive a more coherent dataset required for the regional focus of such projects. As such, this current work, documenting the key aspects of the ICOS station labeling process and data quality control, is a significant contribution to the field, and suitable for publication in this journal.
Overall, the manuscript is well-written and major revisions aren’t necessary. However, I would like to make some suggestions that, when properly addressed, I feel the community will benefit in knowing, especially given the extensive and high-quality dataset that this work can derive it’s conclusions from.

[P5L13-29] Cylinder calibration requirements
I feel that the authors can further elaborate on the following questions.
- Line 14: Can the authors specify what range of calibration frequencies are typically tested in the initial test period?
During the initial test period, a minimal frequency of 15 days is applied. Depending on the stability of one calibration to the other, we can determine if the frequency can be reduced but at the present time we require at least one calibration sequence per month. In addition we recommend to perform a calibration also after any power cut. In such a case the protocol is to let the instrument measuring ambient air for several hour after the restart of the instrument, before starting the calibration sequence. According to the stability of the analyzers for CO_2 and CH_4 we could consider relaxing the frequency. However, the lifetime of the calibration scales at stations is already long (at least 10 years) so the cost of the calibrations is not very high. Also the CO measurements performed by the same analyzers (CRDS picarro model G2401) or by other analyzers (OA-ICOS or CRDS for N_2O and CO) are generally less stable in time. In such cases we add to the calibration protocol one tank (called REF or short-term working standard) measured every few hours in order to correct for short term drifts often observed for CO and N_2O measurements.

- Are there any requirements for the order in which the 3-4 cylinders are measured in a sequence?
No order is required but in most sites, the calibration goes from lower to higher mixing ratio.

- Line 18: I would prefer that the authors clarify that the 30 minutes is the run time per cylinder, even if that becomes implied when reading through other parts of the text.
Similarly, I think it’s worth further clarifying that the calibration cylinders will be run in series, leading to an estimated length of some continuous hours of calibration per sequence.

- When stating the total length of the calibration sequence, I feel it would be more coherent if a short reference is made to the discussions in 2.3.1 regarding the how the instrument will dry out during the long calibration sequence.
We will clarify these points.

- Figure 2 indicates a REF tank which isn't explained in the text. Is this part of the ICOS requirements, and if so, can an explanation be provided for the role of the REF tank?
We will add an explanation. A REF or short-term working standard is strongly recommended for instruments showing a short-term drift in the order of hour/day such as the instruments measuring N_2O/CO. It is measured as often or more than the target and is used to correct the short-term drift (see previous point on calibration frequency).

- Figure 2 suggests that all the samples go through the selection valve and straight into the instrument, which means that the inlet pressure will likely vary between samples, for example the air inlets will likely be slightly sub-ambient, while the calibration cylinders will be dependent on the regulator settings for each tank, and this design is dependent on the instrument's internal cell pressure control to compensate accordingly. Is this correct? Is there any guidance on the regulator pressure settings?
It is correct. We advise to set the pressure on the regulators so that the pressure and flow are slightly higher than the air measurements. For example, on CRDS instrument, this is done by comparing the outletvalve value. During the initial test of the instrument, an acceptable range for the pressure, and outletvalve is determined to help the PI set the regulators. During the initial tests at the MLab, biases are mostly seen when changing the pressure then over time at the same pressure they decrease, this means if the difference between air and cylinder is too large then it will take too long to disappear over the time we are measuring the samples. This useful comment will be added in the manuscript.

- Regarding regulators, I feel the community would greatly benefit from a list of recommended regulator models, similarly to how the the authors have stated a specific recommendation for the sample switching rotary valve later in the text.
The recommendation is listed in the ICOS specification document (SCOTT MODEL 14 M-14C (or -14B) Nickel-plated brass or TESCOM Serie 64-3400 Stainless steel electropolish with PCTFE valve seat and the use of Stainless steel High purity gas pressure gauge (e.g. Bourdon Haenni UPG2)). We may add this information in the paper, at least mentioning that it is provided in the ICOS specifications which is a public document (see the point about this reference below).

[P6L9-20] Moisture effects on the short-term target measurements
- The effect of moisture on the short-term target is very interesting. How long does it take for the moisture level to return to steady state once the air measurements resume, and is there an attempt to flag out the air measurements while the moisture levels are still low?
The first minutes of each sample are flagged out, the length is determined individually. Usually, without a Nafion drier (which was the case in this study for all CRDS at the time of the labeling), the humidity levels are back to steady state almost instantaneously. With a Nafion drier, it can take up to 2 hours but the humidity influence is corrected.

- Also, what are the implications of calibrating the instrument in a dry state while air measurements are done wet, especially in cases where a dryer is not used? Does what's shown for the short-term target measurements indicate there could there be a (small) bias between the calibration and air measurements?
Indeed, for this type of instrument, measuring in super dry conditions leads to a bias. This is why we ask the PIs to measure the target tank just after the calibration to be able to estimate this bias. Depending on the instrument, this bias is more or less pronounced but do not exceed 0.05 ppm

for CO_2 and 0.4ppb for CH_4. It is also part of the uncertainty of the water vapor correction estimated during the initial test at the Mlab. Moreover, one of the final test is to compare the instrument with the Mlab reference instrument whose samples (air and cylinder gas) are all dried.

[P7L15-27] Inlet line tests
- Just making sure I understand this process correctly: For the instrument test, do I understand correctly that there's an electronically actuated 12-port switching valve, and that one of the ports (#12 in this case) is being left open to facilitate direct measurement of the test tank? Is this port normally plugged?
Usually, we use 16 port valve so even with 7 cylinders and up to 5 levels, there are some ports that are free. We can use any free port, it does not need to be the same all the time. They are usually closed when not used.

- For the shelter test, I presume the downstream overflow pump for the inlet line is kept on, then the pressure of the test tank connected to the inlet line is adjusted to the point that the slight sub-ambient pressures in the inlet lines are reproduced? If these presumptions are correct, I would advise that these details be included in the manuscript.
We turn the flushing pump off and adjust the regulator to reproduce the ambient pressures. We will add theses details in the text.

[P8L12-24] PI data validation
- Can the authors clarify the temporal resolution in which the PI's perform data validation? I presume either raw or 1-min level?
The PIs validate the data on the raw level first then every one to two months on the hourly level. The second validation on the hourly dataset aims to verify the longer term consistency of the time series. Every time a data flagging is performed either on the raw or hourly data, a reprocessing is automatically applied to the other aggregated levels (raw, minute, hour) for consistency.

- What are the criteria that go into validating the data at hourly timescales as shown in Figure 3? For example, is there a % or minimum data amount requirement to flagging the hourly data as valid/invalid?
The ATC performs simple filtering on the raw data based on valid ranges (min/max values) for the five mandatory species which are pressure, temperature, relative humidity, wind speed and wind direction. Except for relative humidity, the data are also marked as invalid if the measurement is constant for more than X minutes in a row. X is set to ten for the wind variables and to 60 for the other species. The second criterion is used to cope with blocked sensors.

[P9L19-28] Calibration drift due to regulator effects
- I find Figure 5 fascinating, in terms of understanding regulator contamination effects. As expected, the short-term target, which is measured more frequently, shows faster stabilization time. On the other hand, for the long-term target and calibration tanks, all measured at the same 15-day cycles (as far as I understand), regulator flushing problems are much more severe in the the long-term target. My guess is that the calibration runs are an average of the 3~4 cycles, hence the variability is substantially reduced by the 2nd~4th runs. Is this correct? If so, comparing the first calibration runs against the long-term target run may provide a more meaningful comparison of regulator contamination effects. Also, it would be interesting to know if stability is reached much more quickly in the 2nd~4th calibration runs.
It is correct, however we have focused in this plot in providing a stabilization time for each sample to potentially save gas without losing quality. This is something we could look at in the future.

- Do all the cylinders use the same model of regulators? If not, could that be a consideration in interpreting the results shown in Figure 5?- I might expect that each of the calibration tanks would show a somewhat different contamination effect based on its concentration, but have the authors looked as this?
All cylinders should use the regulators deemed compliant in the specification document. Only two types were selected: SCOTT MODEL 14 M-14C (or -14B) Nickel-plated brass or TESCOM Serie 64-3400 Stainless steel electropolish with PCTFE valve seat and the use of Stainless steel High purity gas pressure gauge (e.g. Bourdon Haenni UPG2) after study presented during MSA meetings.

- I find it interesting that each compound shows a different pattern in reaching steady-state. Looking at the long-term target where the signals are most amplified, CO_2 is initially very low, CH_4 starts low and slightly overshoots prior to reaching steady state, while CO is initially very high. Do the authors have an explanation for these results? Some of these discussions may be more suitable in section 4.2.
We do not have an explanation for this but this at least is partly instrument specific.

[P9L19-28] Calibration drift
- Have the authors estimated whether the instrument drift is "linear"? I.e. When looking at the drift rates of the 4 calibration cylinders at different concentrations, is there indication that the drift rate is similar/different at each of the concentrations? Looking at Figure 6, it certainly seems like the 4 tanks drift at a similar rate, but it's hard to be quantitive. I would love to see an assessment of instrument linearity changes quantified and compared across instruments, and the vast dataset accumulated in this ICOS experiment would be an excellent for this.
From the linearity tests performed at the Mlab and the figures like Figure 6 made for each site, it indeed seems that the drift is linear. However the drift can change every time the instrument is restarted which makes the analysis challenging. We think it will be interesting to perform such analysis when all the sites will have a longer period to investigate (some sites only had a few months of data).

[P19L7-12] Meteorological measurements
- Can the authors specify what checks the ATC performs on the meteorological measurements? My understanding is that verifying the instrumentation and checking the accuracy of the meteorological data is quite challenging, and it would be interesting to know the ATC's specific procedures.
The hourly means are computed automatically using minutes means which are themselves computed using raw data. If there is at least one valid raw data within a given minute, the corresponding minute mean is considered as valid. Similarly, if there is at least one valid minute mean to compute an hourly mean, the hourly mean is consider valid.
There is no automatic quality control criterion applied to the hourly means, only the criteria are applied on the raw data.
We are working on a comparison with the ECMWF data to highlight potential drifts or outliers. In terms of instrumentation, we are working on instating a two-year recalibration for the humidity sensors that are the one drifting the fastest over time.

[P10L22] Alignment of timestamps
- One thing that hasn't been mentioned here is whether there are procedures to ensure that the timestamps of the various data streams on site are matched. For example, when combining the data stream from the CRDS and the meteorology measurements, is there a process to ensure that the timestamps for these two data streams are matched? Or similarly in case there are multiple CRDS instruments on site?
All instruments have access to internet and a time server that is updated at least once a day. This information will be added to the manuscript.

[P12L29] Section on tank stabilization time
- Again, can the authors confirm that the stabilization time for the calibration combines the data for all 3~4 cycles? I would imagine that the 1st cycle shows the longest stabilization time, while the others would be relatively short, and averaging these together would lead to a relatively short stabilization time. In fact, I wonder if the 1-min stabilization time mentioned in page 13 line1 refers to this special case of 2nd~4th cycles in a calibration run?
Indeed, we combine the data from all cycles, including the first one. As the first one is not used to calculate the calibration coefficients, its weight on the stabilization time is less important.

- Please refer to questions above on "Calibration drift due to regulator effects" on some questions I hope the authors can address, perhaps in this section.
- It is interesting that looking at CH_4 in Figure 15, the short-term target stabilization times are higher than the long-term target stabilization times at many sites. I'm a little surprised by this since the short-term target regulator would be flushed more frequently and hence I would expect it to show less regulator artifacts. Do the authors have an explanation for this?
We do not have a definitive explanation for this. However, in most cases, the difference is small. As both target flushing time is the same in the database, we choose the largest value to be conservative.

Minor comments (Denoted by Page# and Line#)
P2L10-19 Introductory paragraph: This introductory paragraph is a summary of the long history of greenhouse gas measurements, which understandably is not an easy task.
Here are some thoughts:
- "Continuous"?: First, it is somewhat vague what "continuous" refers to in this case. I presume the authors mean measurements of very high frequency, but technically speaking, no measurement is truly "continuous". While I understand the authors' emphasis on measurements made at high-frequency, I do feel that the significant scientific contributions from flask based measurements of GHGs deserves to be acknowledged. Also, I would note that the AGAGE measurements (referenced in the Prinn et al.) are GC based and I'm not sure that one would describe them as "continuous" (although I believe some prior publications have described the measurements as "quasi-continuous").
- Also, I find it odd that the more recent megacities efforts are not referenced here, nor are any of the flux measurement networks, as both are high-frequency measurements, and the authors note the importance of networks to address regional and local fluxes (line 17).
- Given these comments, I would ask that the authors reconsider the framing of this introductory sentence, and rewrite as necessary.
We will rephrase to acknowledge the different frequencies and type of sampling as well as different networks.

- Line 13, Prinn et al. 2000: Please update to Prinn et al 2018. Prinn, R., Weiss, R., Arduini, J., Arnold, T., DeWitt, H., Fraser, P., Ganesan, A., Gasore, J., Harth, C., Hermansen, O., Kim, J., Krummel, P., Li, S., Loh, Z., Lunder, C., Maione, M., Manning, A., Miller, B., Mitrevski, B., Mühle, J., O'doherty, S., Park, S., Reimann, S., Rigby, M., Saito, T., Salameh, P., Schmidt, R., Simmonds, P., Steele, L., Vollmer, M., Wang, R., Yao, B., Yokouchi, Y., Young, D., Zhou, L. (2018). History of chemically and radiatively important atmospheric gases from the Advanced Global AtmosphericGases Experiment (AGAGE) Earth System Science Data 10(2), 985 - 1018. https://dx.doi.org/10.5194/essd-10-985-2018
P3L8: Does the "data center" have any special acronym like MLab or MobileLab? At first reading, the data center didn't seem distinguished enough as the second component of ATC.
The data center is definitively as distinguished as the MLab but it does not have a special acronym. The need for such acronym never seemed to come.

P3L11 “ATC and Laurent (2017)”: The notation of this reference seems a bit strange to me. Looking at the original document, I see that it is edited by Laurent with many contributors listed. This metadata page (https://meta.icos-cp.eu/objects/_fDB4nDzrYYG9uu6fPsvfiG9) asks that the citation be for “Laurent, O., ICOSAtmosphere Monitoring Station Assembly and ICOS Atmosphere Thematic Centre (ATC)”. My suggestion would be for “Laurent et al.” or “ATC” with Laurent noted as aneditor in the reference, but I ask the authors to confirm the appropriate way to reference this report. Also note that the full reference for this item seems to include some typographical errors.
A DOI has recently been attributed to this document and the reference will be updated in the manuscript:
ICOS RI (2020): ICOS Atmosphere Station Specifications V2.0 (editor: O. Laurent). ICOS ERIC. https://doi.org/10.18160/GK28-2188

For the technical corrections, we will apply them.
P3L13 “Downstream”: This usage seems a little awkward, as I would associate the term with a physical connected process (as in instrumentation), and not for a separate task carried out at a later point in time. Perhaps something like “Afterwards” or “Once the labeling process has been approved,”?
P3L30-33: Should the first word after the “:”s be capitalized?
P4L4 “exchanges first with” -> “first contacts”?
P4L6 “in a limited number” -> “a limited number of”?
P4L10 “sites” -> “site”?
P4L15 “thanks to” -> “through”?
P4L33 “class 1” -> “Class 1”?
P5L14 “a lot of” -> “more frequent”?
P6L15 “depending” -> “dependent”?
P6L15 “really important” -> “important”?
P6L18-20: “If the instrument variability should be assessed with...”: I’m not exactly sure what the authors refer to in this sentence, please clarify.
P6L22 “are going through” -> “go through”?
P6L22 “different criteria” -> “various criteria”?
P7L1-5: Capitalize each “to”?
P8L6 “are meeting” -> “meet”?
P8L27 “evaluating that” -> “evaluating whether”
P10L16 “flow rates”: Can the authors clarify what specific flow rate is being referenced here, and where this flow rate measured?
P11L20 “who” -> “which”?P11L28-30 “The other ten use either solenoid values either for...” : The meaning of this
sentence seems unclear, with the two “either”s. Please revise.
P12L2 “experience often” -> “often experience”?
P12L2 “do no use” -> “do not use”
P13L9 “was concerned” -> “was of concern”
P13L9 “as said in” -> “as discussed in”
P13L9-10 “GAT CRDS lines showed a long stabilization time” -> “Long stabilization times were found for the GAT CRDS lines”?
P14L21-22 “This can be due to ... analyzer inlet,...” -> “This can be due to factors such as ... and the filter at the analyzer inlet.”
P14L30 “either a cryogenic” -> “or a cryogenic”?
P15L11: Use of acronyms vs full names for the sites?
P15L16 “no CO bias was observed anymore” -> “CO bias was no longer observed”?
P15L25: Should the first letter of each bullet item be capitalized?

Table 1. Based on P4L27, is the “*” in [Class 1 ; Gases, periodical] applied to the wrong item? I.e., shouldn’t the 14 C item be starred?
Table 3. Coordinates: I would find it helpful if the lat/lon coordinates could be given with more significant digits to allow for pin-pointing site locations in interactive mapping tools such as Google Earth.
We used the values provided by the PIs during the Step1 on the CarbonPortal, we will check them for their accuracy.

Table 4. KRE Ambient air (%) is 5069?
This is of course a typo, it will be corrected.

All figures: Some figures missing specific description of how the error bars were calculated. Also, please note the differing concentration units in the plots/text (ex. Ppm vs µmol/mol), and any missing subscripts in CO 2 and CH 4 .
Figure 2. Perhaps arrows would be more apt in identifying the specific injection points? Also, the plot is difficult to interpret without legends for the symbols.
Figure 15. Can the font size of the x-axis labels be adjusted for better legibility?
Figure 16-18: It seems like the bottom Bias plot in each figure has 26 ticks on the x-axis, while the CMR/LTR plots above have 27 ticks. Is there a station/instrument missing in the bottom Bias plot? Also, the pink/blue boxes are very hard to distinguish for many of the box plots, is there a way to improve legibility?
We will improve the figures according to comments. In the case of the box plots, their size is due to the small range (which show the good quality of data). We can try to show some outliers differently to reduce the overall range.