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?
- Are there any requirements for the order in which the 3-4 cylinders are measured in a sequence?
- 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.
- 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?
- 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?
- 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.

[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?
- 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?

[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?
- 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.

[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?
- 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?

[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.
- 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?
- 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.

[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.

[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.

[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?

[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?
- 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?

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.
- 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 Atmospheric Gases 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.

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., ICOS Atmosphere Monitoring Station Assembly and ICOS Atmosphere Thematic Centre (ATC)”. My suggestion would be for “Laurent et al.” or “ATC” with Laurent noted as an editor 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.

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.

Table 4. KRE Ambient air (%) is 5069?

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?