Review of “Inter-comparison of O$_2$/N$_2$ Ratio Scales Among AIST, NIES, TU, and SIO Based on Round-Robin Using Gravimetric Standard Mixtures” by Aoki et al.

In this manuscript, Aoki and colleagues present the results from their intercomparison program for O$_2$/N$_2$. They show the details of the gravimetric scale established at NMJI/AIST, in comparison also to the internationally used Scripps scale. They show the impact of the scale conversion for the measurements at the Hateruma Island measurement site and the impact on global carbon budget calculations. The work done is worthwhile for publication in AMT, taking into account the following comments.

General comments:

Language: in some places the English writing could be improved, I have listed some suggestions in the detailed comments below.

Abstract: the abstract would benefit for having a number to represent the differences in precision between the laboratories, next to the offsets. The span offsets given in % are not directly clear, and it would be helpful if they could be expressed differently. It is also not clear from the abstract how this span offsets leads to a value of 0.29 PgC/yr in the carbon budget, and in which direction the shift is. It would be good to elaborate on this. It would be helpful to specify what the term zero offsets represents, so that the abstract is easier to read without reading the manuscript first. It would be good to include a quantification of the comparison of the 2 records at HAT in the final sentence. The “temporal drift” in line 21 could also be further explained, e.g. with details on the time period.
Page 3, lines 25-28: it is not fully clear to me how the correction for the expansion rates are applied and how these are measured.

Page 4, line 10: it would be good to add information on the choice of the range, it seems quite a large range in comparison to observed values.

Page 5, line 3: how are these values determined? (5000, 1000, 500 times).

Page 5, line 28: what is sufficiently small?

Page 6, line 13: why are these values constant?

Page 7, line 2: can the authors also provide the long-term stability for the other labs?

Page 7, line 27: it would be good to explain that the SIO scale is defined to be 0 per meg, because it is used internationally.

Page 9, line 11: (and other places in the text): what does the “expanded uncertainty” represent?

Page 9, line 124: this seems a large drift in a couple of years’ time, so “slightly” might not be the appropriate word here. The explanations for the drift because of the oxidation inside the cylinders seems to be different for each cylinders, are the corrections made for each cylinder separately? Are these the regular cylinders, also used for maintaining the NMJJ/AIST scale? Or is this only used for the round-robins?

Page 9, line 32: Where do we see the long-term drift of each laboratory’s scale?

Page 10, line 10: what do these percentages represent?

Page 10, line 12: could you elaborate on the filling years?
Page 10, line 16: how are the results consistent with the GOLLUM program? Can this be quantified?

Page 10, line 17: could you quantify “slightly bigger”?

Page 10, line 22: can you quantify how this study shows that the labs can be compared?

Page 10, line 25: how was it confirmed that the isotope ratios did not differ significantly?

Page 10, line 29: why not both against the SIO scale?

Page 11, line 2: how is the value of -6.6 per meg derived, and should it be compared to the goal of intercompatibility of 5 per meg?

Page 11, lines 16-20: the GCB paper has been updated in the meantime twice, and it would be best to use the numbers from Friedlingstein et al. 2020. Line 18 does not seem to be a full sentence and it is not clear to me what the 0.29 PgC/yr correction is (e.g. from land to ocean, or the other way around?). Table 5: it would be good to add the numbers from Friedlingstein et al. 2020 in the table for reference.

Page 11, line 30: “first time in the world”: what is the first time in the world? The GOLLUM program is also an intercomparison program between laboratories.

Page 11, line 31: rewrite the 0.29 PgC/yr, to be more specific what the number means. See comment above.

Page 12, line 3: what does “other four” mean in comparison to the GOLLUM program?

Page 12, lines 4-6: it would be good to be more specific here on the implications, rather than repeating the causes for the decrease.

Page 12, line 11: can you quantify the bias?
**Page 12, lines 11-12:** how do the results improve the carbon budget and OHC increase?

**Conclusions:** it would be useful to include an outlook. Will this intercomparison continue in the future? Will other laboratories be invited to participate?

**Table 1:** can the authors add more information about the expansion rates? What is meant with the standard uncertainty?

**Figure 2:** what is on the x axis?

**Table 3:** how is the standard uncertainty determined?

**Figure 3:** should the y-axis in panel a read NMJI/AIST instead of grav? Maybe also include the average residuals to compare lab precisions?

**Figure 4:** maybe add a panel with the differences/bias? Why not on the Scripps scale? It would be good to include in the caption that these are duplicate samples, not measurements of the same flasks.

**Table 5:** change to Friedlingstein et al. 2020 (see comment above), and add numbers for comparison. Which numbers are Tohjima et al. 2019? Rewrite “changing rate”.

**Detailed comments:**

**Page 2**

Line 2: “molar fraction” could be changed to “mole fraction”, which is more commonly used in the field, throughout the text.

Line 3: explain “some” stoichiometric ratios.

Lines 2-5: references to earlier studies would be appropriate here.
Line 6: “changes” should be “measurements”.

Line 11: rewrite “the ocean uptakes”

Line 14: “precision” -> “precise” and rewrite micro-mole-per-mole

Line 27: “per meg” instead of “per meg units”

Page 3

Line 6: “remain”? 

Line 22: explain “round-robin”

Page 4

Line 6: “hereafter the” -> “hereafter called the”

Line 24: why are the Ar values much more variable compared to tropospheric air?

Page 7

Line 10: what is meant with “unequal to”?

Page 8

Lines 5-8: the sensitivity factor and interference factors could be further explained to be clearer.
Line 1: what are "changing rates"?

Lines 29-30: what is meant here by selecting mixtures from the round-robin?

Page 10:

Line 5: rewrite "Figure 3a plots"

Page 11

Line 10: "corrected" -> "improved"?

Line 13: rewrite "secular changing rate"