

Reviewer #3

The authors present a very important result of an intercomparison between many labs for measuring methane and nitrous oxide levels in ocean water samples. Overall, I think this paper is well written and will be a great contribution to the field. A lot of planning and work went into this study, and is worthy of publishing. The main focus is to look at standards, calibration issues, but don't really address how with the large variability of how people process water samples affects the results. I think this paper highlights some very important issues regarding trace gas analysis in open ocean settings, and could be transferred to other environments. Section 4.3 will be regarded as a huge step forward, once this group is able to produce a Good Practice Guide to the community. While I was left wanting to know about how best to make these measurements, I acknowledge that this group is on the way to doing that and will do that. This paper is the first step. The conclusion that calibration issues are a huge problem in this field, and the recommendation to produce reference material for both trace gases is a wonderful contribution.

1. They mention on line 587 for all labs to do internal checks by measuring an air-equilibrated seawater. They mention needing a water bath and stirrer. Since this is a main finding that could be implemented in the community ASAP, could they provide true details of the setup? This might be appropriate in the supplementary materials.

We reference four studies which report using air-equilibrated seawater as an internal control. Each of these studies had slightly different procedures and at this stage we refer the readers to these publications for further information. We would like to conduct a more thorough analysis of how robust these measurements are (e.g. sensitivity to temperature fluctuations) before publishing more detailed recommendations as part of a planned Best Practice Guide.

2. Line 220: Why is there such variation in equilibration time for the gases; between 20 min to 24 hours? Has anyone done a time series of equilibration times to show what the time needs to be? This could be part of the recommendations.

The longer equilibration times are due to overnight equilibrations in water baths. All laboratories should test equilibration time for the headspace analysis or the sparge time for the purge-and-trap technique, when establishing their own personal protocols for different sample volumes, temperatures, and sampling habitat.

3. Line 272: Where do the CV values come from that are plotted in figure 7b? In table S2, there is one column for "mean CV" which seems to be related to each lab, and not specifically for PAC1 and PAC2. Maybe those CVs are just not reported in the table, in which case, please report them.

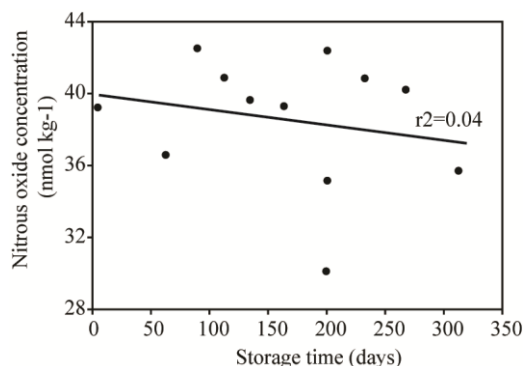
The values of coefficient of variation (%) shown in Figure 7b are associated with methane concentrations measured by each lab for PAC1 and PAC2 samples (collected in February 2017). These specific values are not included in any of the Supplementary Material tables, where we instead report the mean coefficient of variation associated with each laboratory. We also report the coefficient of variation for the whole batch of samples in Table 2 in the main document.

4. Line 371: it is not clear to me what they mean by "sample contamination, discussed below (datasets J and K)." Where do they discuss below? Could they call out the specific sections they want the reader to refer to?

This sentence has been improved and Lines 385-388 now read '*In contrast, the datasets with a higher offset at low methane concentrations (Datasets J and K) could be due to the use of incorrect intercepts as well as other factors including sample contamination, discussed in Section 3.4.*'

5. Line 430 and on: The storage section really added a nice dimension to the paper, even though it was not a main focus. On line 445, you state that BAL2 shows a decrease in N₂O concentrations over time. Can you show that graph? When graphed, I see that BAL2 shows an increase with time but it also seems within the variability of the measurements.

Reviewer #3 has highlighted an error in the manuscript as we meant to say BAL5, not BAL2. We apologize for the error. Because there is not a significant decrease of nitrous oxide with time, we did not initially include this Figure in the manuscript. We now feel that it is inappropriate to include this comment and we have removed the sentence ‘There was some indication of a decrease in concentration for seawater samples with higher concentration of nitrous oxide (i.e. BAL5), which could have been caused by gas leakage’ from the manuscript.



6. Line 432: The explanation of the results from Magen 2014 are a bit misleading. That paper shows that at methane concentrations less than ~1ppm in the headspace, there could be a storage issue after 1 year. And the issue is that concentrations increase. There should be more context to your statement “because prolonged sample storage adversely affects dissolved methane and nitrous oxide samples (Magen et al., 2014)...”

In response to this comment and comments from other Reviewers, this section has been rewritten and Lines 448-459 now read ‘Because prolonged samples storage can have an adverse affect on dissolved gases, including methane and nitrous oxide, the intercomparison dataset was analyzed for sample storage effects (Table S5 in the Supplement). It should however be noted that assessing the effect of storage time on sample integrity was not a formal goal of the intercomparison exercise and replicate samples were not analyzed at repeated intervals by independent laboratories, as would normally be required for a thorough analysis. Nonetheless our results did provide some insights. Most notably, there were indications that an increase in storage time caused increased concentrations and increased variability for methane samples with low concentrations, i.e. PAC1 and PAC2 samples which had median methane concentrations of 0.9 and 2.3 nmol kg⁻¹, respectively (Fig. 7). In comparison, for samples of nitrous oxide with low concentrations there was no trend of increasing values as observed for samples with low methane concentrations.’

7. Line 439: Storage for methane. Where did the data come from for figure 7? From the supplemental tables, the only storage time data shown is from Feb for PAC 2, and Nov for PAC1. Just from a first look, there are only 7 reported values for methane for PAC1 Nov in Table S2, but 11 points plotted in figure

The questions in 7, 7a, and 7b are dealt with below

7a. Where is the extra data coming from? Data looks consistent for PAC2. If I replot the storage days from table S5 vs the concentrations from table S2, I get the following graphs. (For the graphs below,

methane concentrations are plotted over storage time with the outliers and without.) Those outliers were identified in figure 7a with () around the symbols, which is stated in the figure caption to be taking out of the regression. For PAC2, I reproduce what was reported in figure 7a, but for PAC1, the story is completely different. Please address this inconsistency.

a. After agonizing over the mismatch of this data, it looks like they plotted PAC1 Feb 2017 in figure 7a, not PAC1 Nov 2013. If that's the case, the storage time data presented in table S5 is not right.

I think the confusion exists because the Supplementary Table 5 included the storage times for samples collected in November 2013 (Pacific_1) and February 2017 (Pacific_2). However, we also referred to the sampling depths as PAC1 (25 m depth) and PAC2 (700 m depth). Therefore, there is too much similarity between date (Pacific_1 and Pacific_2) and depth (PAC1 and PAC2). After consideration, we have removed the column in Table S5 which lists the storage time for the November 2013 samples. Since we do not refer to the November 2013 samples in the main document, there is no loss of information by not including their storage times and there will be less confusion.

The Table of data used to create Figure 7 is shown below.

Storage time (days)	PAC 1 Feb 2017 25 m samples		PAC 1 Feb 2017 25 m samples	
	Methane conc (nmol kg ⁻¹)	Coeffic. Variation (%)	Methane conc (nmol kg ⁻¹)	Coeffic. Variation (%)
122	2.52	-	2.96	14.1
68	2.91	10.1	1.65	9.3
67	2.25	2.5	0.64	2.8
132	3.70	7.0	2.35	6.6
132	3.79	37.0	2.33	28.0
88	2.61	2.2	1.03	9.3
123	5.21	25.7	31.4	11.2
223	-	-	-	-
39	2.01	4.2	0.65	8.6
140	-	-	-	-
75	1.95	2.6	0.62	19.5
105	2.12	1.5	0.63	6.6
72	2.23	1.8	0.82	12.4

Where did the data come from for figure 7?

None of the November 2013 Pacific_1 data are shown in Figure 7. We state on Lines 167-171 that ‘The November 2013 samples are included in Figure S1 and S2 in the Supplement, but are not discussed in the main Results or Discussion because fewer laboratories were involved in the initial intercomparison, and the results from these samples support the same conclusions obtained with the more recent sample collections.’ To make this clearer for the readers, we have repeated this text in the Figure 7 legend and Line 909 now reads ‘...collected in February 2017’

Where is the extra data coming from? **There are no extra data. For the February 2017 Pacific_2 Column in Table S5 there are 14 labs in total and 2 of these labs (Red and Beige) did not measure**

methane in the Pacific Ocean. The 12 datasets are represented by the 12 data points are shown in Figure 7.

8. Can you add a column in the supplemental table for N₂O for how each person dealt with water, like what was done for methane? Water is a huge issue for N₂O precision, and there is no mention of how water was dealt with.

This is now included in Supplementary Table 7. As a quick response, water vapor is removed by most laboratories using a drying agent frequently in combination with Nafion tubing.

9. Line 507, if your intent is to show some examples, you should add “for example” to your reference list here. There are many other papers that show this.

Changed

10. Line 557: extra space between “proposed” and “production”

Changed

11. In table S5, “red” is listed as having measured something on the PAC samples 140 days after collection. But when I try to cross reference this in table 2, it looks like “red” didn’t measure for methane. It might help to know if the storage times in table S5 are for methane and/or N₂O. Overall, I think this table needed revisiting.

Reviewer#3 is correct, ‘red’ Laboratory M only made nitrous oxide measurements. There was also one laboratory (Laboratory D, beige) that only measured methane. We have improved the Table heading to make this clearer and it now reads ‘The reported storage times are for both methane and nitrous oxide (Laboratory M ‘red’ measured methane only and Laboratory D ‘beige’ measured nitrous oxide only).’

12. Figure S1, what is the gray dashed line? What do colors represent?

Individual data points are plotted sequentially in increasing value with the same color symbol for each laboratory in all plots for the main text and Supplementary Material. The dashed grey line represents the value of methane at atmospheric equilibrium as stated in the Figure legend.

13. Figure S2, are a and b shallow water and c and d deep water? Make that clear in the first description of the figure. It says “same location” but what you mean is at the same lat/long but two different depths. Also, caption says “In contrast, the concentration of nitrous oxide in the deep-water samples (Figure S2c and d) was more consistent and the data values for the laboratories that measured samples from 2013 and 2017 are shown together in Figure S2d.” is that also supposed to be shown by a gray dashed line? Can you make the scales the same for both sides?

We have now plotted Figure S2c on the same scale as Figure S2d. Each subplot also includes a description of depth as well as the actual Figure legend.

The Figure S2 legend has been improved and now reads ‘Supplementary Figure S2: Nitrous oxide concentrations in seawater samples collected at the same location but varying depths in the North Pacific Ocean on February 2017 (Fig. S2a and c) and November 2013 (Fig. S2b and d). The dashed grey line represents the value of nitrous oxide at atmospheric equilibrium for the 25 m seawater samples (Figure S2a and b). The February 2017 plots are discussed in the main manuscript and are replicated here to facilitate comparison with the November 2013 data, particularly for comparison with the 700 m samples (Figure S2d).’

14. Supp table 1: what is the point of the far right columns in this table? What is the mean CV of? For example, for lab A, it says 9.2% CV. Did you take CV for each BAL1, BAL2, etc, and then average that? Since we don't see the BAL1 CV, this is not clear. That being said, I'd like to see the CV for the standards run in the lab. From my experience with N₂O, I can have ~10% CV if there is still water in the sample.

The purpose of the Supplementary Tables 1-4 is to provide further information about the data values provided in Figure 1 and Figure 4 in the main document. The far right-hand columns provide a measure of variability for each laboratory as shown by the mean coefficient of variation (%) and the mean offset (%). We now state in the Table heading that these values are for all sampling stations shown in each respective Table, 'based on all 7 sampling stations'.

Reviewer #3 also indicates that it would be helpful to see the coefficient of variation (%) for standards as well as the samples. In our experience, there is always higher precision associated with analysis of standards. This is because sample analysis includes multiple steps of sample handling, gas extraction/equilibration. Therefore we prefer to report the precision associated with sample analysis, as the precision associated with standards will be lower than this value.