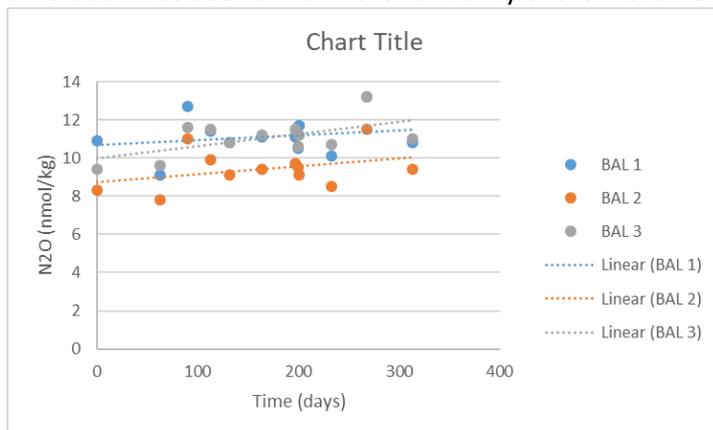


The authors present a very important result of an intercomparison between many labs for measuring methane and nitrous oxide levels in ocean water samples. Overall, I think this paper is well written and will be a great contribution to the field. A lot of planning and work went into this study, and is 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.

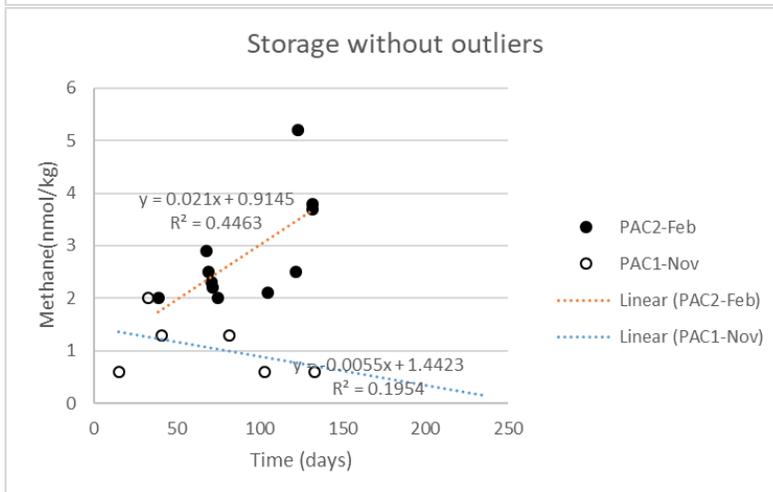
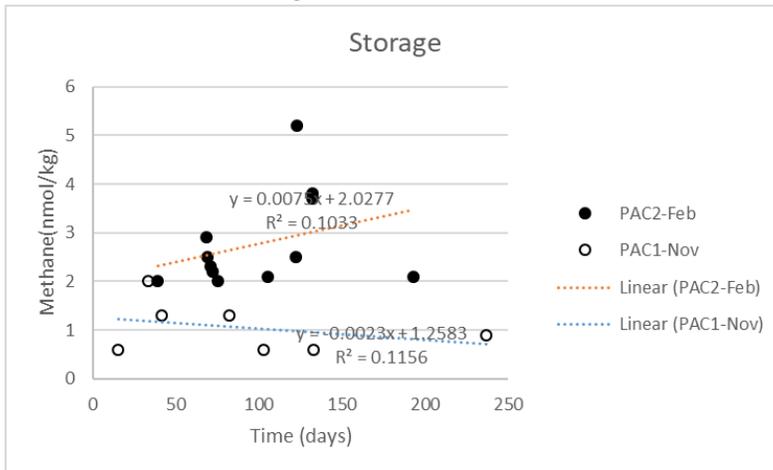
1. They mention on line 587 for all labs to do internal checks by measuring an air-equilibrated seawater. They mention needing a water bath and stirrer. Since this is a main finding that could be implemented in the community ASAP, could they provide true details of the setup? This might be appropriate in the supplementary materials.
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.
3. Ling 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.
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?
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 N<sub>2</sub>O 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.



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)....”

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 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.



8. Can you add a column in the supplemental table for N<sub>2</sub>O for how each person dealt with water, like what was done for methane? Water is a huge issue for N<sub>2</sub>O precision, and there is no mention of how water was dealt with.
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.
10. Line 557: extra space between “proposed” and “production”
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 N<sub>2</sub>O. Overall, I think this table needed revisiting.
12. Figure S1, what is the gray dashed line? What do colors represent?
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?
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 N<sub>2</sub>O, I can have ~10% CV if there is still water in the sample.