

General comments:

Wilson et al. present the first intercomparison of oceanic methane and nitrous oxide measurements across numerous ($n = 11$) international laboratories. This is a timely and important contribution for the community. The paper is scientifically sound, well-written and clear. I have few (generally minor) comments/suggestions below.

While this intercomparison is a first step toward being able to compare the concentrations of these gases measured by different laboratories in marine environments, I have some recommendations to improve the paper. First, while they could discern some trends, I don't think the effect of storage can easily be isolated if the samples are not collected the same way (e.g., using same vial sizes, stoppers) and analyzed using the same analytical method. Although admittedly not being the focus of the present paper, a storage experiment should be repeated where samples in each dataset would be sequentially analyzed at different time points by the same laboratory (all other things being equal). Different type of stoppers/seals should also be compared to determine which one is best.

Also, because water budgets are often limited, they should better assess the effect of different sample volumes on precision and exactitude if possible. For instance, are samples with larger volumes yielded better results?

Minor comments

Page 4, lines 85-89: Which method is the most sensitive (purge and trap versus headspace equilibration)? Discuss the advantages/inconveniences of using one over the other a bit more.

Page 6, lines 140-156: The part describing how they determined the absolute mole fractions for these standards is not clear and the link ([www.scor-int.org /SCOR_Publications](http://www.scor-int.org/SCOR_Publications)) is not working. Why would the uncertainty be higher for the nitrous oxide WRS standard compared to the methane one?

Page 7, lines 158- 182: The effects of sample volumes, type of septa used and storage should be assessed better since these differed between the laboratories involved in the intercomparison.

Page 7, line 171-173: Was there a difference between sampling bottles?

Page 7, line 178: Which kind of stopper? Also, what is the effect of different stoppers/seals used during storage? Are some stoppers/seals leaking more than others?

Page 7, lines 180-182: They used mercuric chloride for preservation, which is probably acceptable for water-column samples. However, mercuric chloride is toxic and difficult to ship and use at sea due to safety concerns. Future efforts should test alternative types of preservatives (sodium hydroxide, formaldehyde) to evaluate their suitability to preserve these samples in different marine environments. Also, mercuric chloride might not be suitable for some marine samples as Ostrom et al (2016) suggest that it could enhance nitrous oxide production by chemodenitrification in Fe-rich environments.

Page 8, line 188: I assume this tank was gas tight?

Page 8, lines 196-198: Was there a difference between this first and last samples? Any change in temperature during sampling would affect gas concentrations. Also, I suppose a headspace was created in the 1000 L water tank as samples were drawn?

Page 9, lines 223-225: “headspace collected into a gas tight syringe and injected”: How is this different than the physical injection?

Page 9, lines 228-229: How many standards were typically used?

Page 9, line 248: Why does the tubing need to be maintained at low temperatures?

Page 9, line 249: Low blank for what? Methane, nitrous oxide, or both?

Page 10, line 251-252: Be more specific: “liquid nitrogen (-165°C) for nitrous oxide or cooled ethanol (-70°C) for methane.”

Page 11, line 303: By “comparable values” do you mean peak area?

Page 13, lines 362-371: This point comes across more clearly in the Fig. 3’s legend. Perhaps rewrite?

Page 14, lines 388-401: A sample with higher nitrous oxide concentrations could also be used in future intercomparison efforts. For instance, nitrous oxide concentrations of up to 1000 nmol/L were measured in coastal waters off Peru (Arévalo-Martínez et al.,2013)..

Page 15: Why was the variability higher for the BAL5 dataset? Could this be related to sampling and/or storage?

Page 16, lines 438-439: Was this only true for samples with methane concentrations less than atmospheric concentrations?

Page 18, line 512: What would be their maximum recommended storage time?

Page 19, lines 532-534: They discuss detection limits for methane but not for nitrous oxide analysis methods. What are the detection limits associated with the two different analysis methods (headspace equilibration versus purge and trap)?

Page 20, lines 560-565: Other important points, e.g., sample volume, septa/seals used, preservative used, should also be included in future efforts.

Page 20, line 576-577: This assumes that the air in the laboratory where the measurements are done is not contaminated by other sources of nitrous oxide (non-atmospheric).

Page 20, line 586: Bourbonnais et al. (2017) also used air-equilibrated seawater standards to calculate water-column nitrous oxide concentrations off Peru.

Figures:

Figures 1: Are values of methane at atmospheric equilibrium expected at 25 m depth? Is this in the mixed layer?

Figure 7: Are these relationships significant (add r^2)? Ideally, to assess storage effects, samples collected the same way and using the same analysis method should be analyzed at different time points by the same laboratory.

Supplementary Materials:

Tables 6 and 7: Add detection limits for each laboratory.
Add last name “Macarena Burgos” as done for all other researchers.

Few technical points

Page 4, lines 76 to 78: Typically is used twice in these two sentences – remove one instance.

Page 18, line 501: change “equilibration” for “equilibrated”.

Page 19, line 545: change to “switching between different calibration curves.”

Additional references:

Bourbonnais, A., Letscher, R.T., Bange, H.W., Echevin, V., Larkum, J., Mohn, J., Yoshida, N. and Altabet, M.A., 2017. N₂O production and consumption from stable isotopic and concentration data in the Peruvian coastal upwelling system. *Global Biogeochemical Cycles*, 31(4), pp.678-698.

Ostrom, N.E., Gandhi, H., Trubl, G. and Murray, A.E., 2016. Chemodenitrification in the cryoecosystem of Lake Vida, Victoria Valley, Antarctica. *Geobiology*, 14(6), pp.575-587.