Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2018-274-AC1, 2018

Biogeosciences Discussions

Interactive comment on "An intercomparison of oceanic methane and nitrous oxide measurements" *by* Samuel T. Wilson et al.

Samuel T. Wilson et al.

stwilson@hawaii.edu

Received and published: 31 August 2018

We would like to thank the three Reviewers for their detailed comments, which have improved the manuscript. We have included a point-by-point response to their comments below, with our responses highlighted in bold text.

Reviewer #1 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? Thank you for these overall positive comments. We address the issue of storage artifacts below.

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. We have updated the text in the Introduction and Lines 89-95 now read 'The purge and trap technique is typically more sensitive by 2-3 orders of magnitude over headspace equilibrium. However, the purge and trap technique requires more time for sample analysis and it is more difficult to automate the injection of samples into the gas analyzer. Headspace equilibrium sampling is most suited for volatile compounds that can be efficiently partitioned into the headspace gas volume from the seawater sample. Its limited sensitivity can be compensated by large volume analysis (e.g. Upstill-Goddard et al., 1996).' The different merits of the two methods are also featured in the revised Discussion, where we highlight the detection limits for methane which are more of an issue than for nitrous oxide. Lines 519-522 read 'An approximate working detection limit for methane analysis via headspace equilibration is 1 nmol kg-1, although some laboratories improve upon this by having a large aqueous: gaseous phase ratio during the equilibration process (e.g. Upstill-Goddard et al., 1996). Depending upon the volume of sample analyzed, purge-and-trap analysis can have a detection limit much lower than 1 nmol kg-1 (e.g. Wilson et al., 2017).'

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) is not working. Why would the uncertainty be higher for the nitrous oxide WRS standard compared to the methane one? We apologize that the report which documented the production of the gas standards was not easily accessible. It is now accessible through the University of Delaware library and the citable URI is now included in the appropriate reference (http://udspace.udel.edu/handle/19716/23288). The report is also attached to this response for your convenience. On Pages 4-5 of this report, the calibrations for the nitrous oxide and methane WRS are described. In response to the question, there is higher certainty for the ARS because the standards were cross-calibrated with National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory (NOAA/CMDL) and Advanced Global Atmospheric Gases Experiment (AGAGE) standards which have a similar mole fraction. In contrast, the mole fraction of the nitrous oxide WRS far exceeds that of the CMDL and AGAGE standards and the calibration curves are highly non-linear. Therefore, the reported 2-3% accuracy takes into consideration the likelihood of increased systematic errors.

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. Reviewer #1 points out that there were sampling and storage variables which were not controlled for during the intercomparison exercise. These are responded to separately below

Sample bottle size We have taken the Reviewer's comments into consideration and expanded Section 3.4 'Sample storage' so that it now includes 'Sample storage and sample bottle size'. Lines 460-465 now read 'Another variable which differed between laboratories for the intercomparison exercise was the size of samples bottle, which ranged from 25 ml to 1 liter for the different laboratories. There was no observed difference between the methane and nitrous oxide values obtained from the various

sampling bottles and it was concluded that sampling bottles were not a controlling factor for the observed differences between laboratories. We note, however, the potential for greater air bubble contamination in smaller bottles'.

Septum We did not test for contamination (either production or adsorbtion) of methane and nitrous oxide by different septa. There are at least two recent articles presenting evidence that storing trace gas samples in bottles with rubber septa can cause contamination for methane (Magen et al., 2015, Niemann et al., 2015). The article by Magen et al (2014) also highlights the possibility of cleaning the septa, although they did not see any difference when this was conducted (albeit over an eight day period). We have amended the manuscript to address the issue of potential septa-derived contamination. This is included in the Discussion in Section 4.3 under General Recommendations. Lines 589-598 now read 'This study also revealed that sample storage time can be an important factor. The results from this study corroborate the findings of Magen et al. (2014) who showed that samples with low concentrations of methane are more susceptible to increased values as a result of contamination. The contamination was most likely due to the release of methane and other hydrocarbons from the septa (Niemann et al., 2015). Since the release of hydrocarbons occurs over a period of time, it is recommended to keep storage time to a minimum and to store samples in the dark. It should be noted that sample integrity can also be compromised due to other factors including inadequate preservation, outgassing, and adsorption of gases onto septa. For all of these reasons, it is recommended to conduct an evaluation of sample storage time for the environment that is being sampled.'

Magen, C., Lapham, L. L., Pohlman, J. W., Marshall, K., Bosman, S., Casso, M., and Chanton, J. P.: A simple headspace equilibration method for measuring dissolved methane, Limnol. Oceanogr.: Methods, 12, 637–650, 2014.

Niemann et al. (2015) Toxic effects of lab-grade butyl rubber stoppers on aerobic methane oxidation Limnol. Oceanogr.: Methods 13, 2015, 40–52

Storage time We have improved the wording of this section and Lines 448-459 now read 'Because prolonged storage of samples can influence dissolved gas concentrations, 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 into potential storage-related problems. 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.'

Page 7, line 171-173: Was there a difference between sampling bottles? No difference between sampling bottles was observed. This is now noted in the document on Lines 323-327 'Analysis conducted by the University of Hawaii of methane and nitrous oxide from each Niskin-like bottle used in the Pacific Ocean sampling did not reveal any bottle-to-bottle differences. Furthermore, analysis by Newcastle University showed there was no difference between the first and the last set of samples collected from the 1000 L tank used in the Baltic Sea sampling.'

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? These questions are answered separately below

Which kind of stopper? The 1 l glass bottles used a ground-glass stopper and Apiezon grease as widely used for dissolved inorganic carbon samples.

Also, what is the effect of different stoppers/seals used during storage? Are some

stoppers/seals leaking more than others? The recent publication by Niemann et al (2015) reported on the release of organic contaminants of five different commercially available, lab-grade butyl stoppers. Different stoppers release varying quantities of different compounds. It should be noted that the objective of the Niemann et al. (2015) study was to look at the effect on biological rate measurements (methane oxidation) and not concentrations. Magen et al (2014) also looked at the potential contamination by two stoppers, although their incubation period was for 3 days only.

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.

The reviewer raises the point that there are alternative preservatives to mercury(II) chloride. The issue with any preservative is to balance effectiveness at ceasing all relevant microbial activity, while minimizing toxicity from a human health and environmental perspective. In recent years, there have been a series of papers (Magen et al., 2014, Bussmann et al., 2015, Gloël et al., 2015) which have tested some of the alternatives to mercury(II) chloride. These include sodium azide, sodium, hydroxide, sulfuric acid, potassium hydroxide, benzalkonium chloride, and zinc chloride. These studies demonstrate the potential for alternative preservatives and show their effectiveness for a particular environment over a particular timeframe. However, they do not prove the applicability over a broad range of conditions, microbial communities, and storage times. The studies also do not provide a recommendation for the most superior preservative, nor do they always test both methane and nitrous oxide, and other substances such as dissolved inorganic carbon. Therefore, while we agree that

alternatives exist, they have not been extensively proven to be superior to the well-established use of mercuric chloride. After talking to a number of scientists about this issue, we understand that the community of scientists focused on dissolved inorganic carbon measurements are looking very carefully at alternatives to mercury(II) chloride. We have requested that measurements of methane and nitrous oxide be included in planned future tests of alternative preservatives. This will allow the whole community to switch to alternative preservatives at the same time. We have revised the manuscript to reflect our perspectives and Lines 188-193 now read 'The choice of mercuric chloride as the preservative for dissolved methane and nitrous oxide was based on its long history of usage. It is recognized that other preservatives have been proposed (e.g. Magen et al., 2014, Bussmann et al., 2015), however pending a community-wide evaluation of their effectiveness over a range of microbial assemblages and environmental conditions for both methane and nitrous oxide, we recommend continuing with a long-established method.' Magen, C., Lapham, L. L., Pohlman, J. W., Marshall, K., Bosman, S., Casso, M., and Chanton, J. P.: A simple headspace equilibration method for measuring dissolved methane, Limnol. Oceanogr.: Methods, 12, 637–650, 2014. Bussmann, I., Matousu, A., Osudar, R. and Mau, S., 2015. Assessment of the radio 3HâĂŘCH4 tracer technique to measure aerobic methane oxidation in the water column. Limnology and Oceanography: Methods, 13(6), pp.312-327. Gloël, J., Robinson, C., Tilstone, G.H., Tarran, G. and Kaiser, J., 2015. Could benzalkonium chloride be a suitable alternative to mercuric chloride for preservation of seawater samples?. Ocean Science Discussions, 12(4), pp.1953-1969. Page 8, line 188: I assume this tank was gas tight? The tank was sufficiently gas-tight for our purposes. The tank was made of high density polyethylene (same material as used for very large carboys). Prior to sampling, the seawater was gently stirred to ensure homogeneity. Subsampling was conducted from a port located at the lowest part of the tank and approximately one-tenth of the tank's contents were sampled. A headspace was created during the sampling and by the time the last sample was collected, there was approximately a 1 meter distance between the sampling port and the headspace interface.

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? No difference was observed between the first and last samples. Please see our description about sampling from the tank in our previous response.

Page 9, lines 223-225: "headspace collected into a gas tight syringe and injected": How is this different than the physical injection? This sentence highlighted the fact that the headspace had been subsampled into a separate syringe. However, this is a very subtle point and as the Reviewer points out, by including physical injection in the previous sentence, this extra description is not needed. We have removed this sentence from the manuscript.

Page 9, lines 228-229: How many standards were typically used? The number of standards used by each laboratory ranged from 2-4. This information is provided in the Supplementary Information in Tables 6 and 7.

Page 9, line 248: Why does the tubing need to be maintained at low temperatures? The majority of scientists install gas purifiers on the gas supply lines which feed any gas analyzer. This is a preventative measure in case the commercially sourced compressed gas cylinders vary in quality, which can occur for even the high-purity gases. The majority of the gas purifiers are commercially available, however a homemade purifier consisting of a length of tubing packed with Porapak or Hayesep material and immersed in liquid nitrogen is recommended for methane analysis when measurements are made using purge-and-trap. The larger volume of purge gas used during purge-and-trap causes trace contaminants to become concentrated which affects the methane chromatogram. This does not appear to be an issue when analyzing methane using the headspace equilibrium technique. We have improved the text to clarify these additional steps for methane analysis. Lines 257-260 now read 'In addition to commercially available scrubbers, purification of the sparge gas was achieved by passing it through stainless steel tubing packed with Poropak Q and immersed in liquid nitro-

gen. This is a recommended precaution to consistently achieve a low blank signal of methane.'

Page 9, line 249: Low blank for what? Methane, nitrous oxide, or both? We have clarified this in response to the previous comment.

Page 10, line 251-252: Be more specific: "liquid nitrogen (-165oC) for methane or cooled ethanol (-70oC) for nitrous oxide." This sentence has been improved and lines 262-263 now read 'Cryotrapping was achieved for methane using liquid nitrogen (-195oC) and either liquid nitrogen or cooled ethanol (-70oC) for nitrous oxide.'

Page 11, line 303: By "comparable values" do you mean peak area? Not quite. The text has been improved to make this clearer. Lines 314-315 now read 'For the two laboratories with an in-house standard of comparable mole fraction to the WRS, an offset of 3% and a >20% offset was reported.'

Page 13, lines 362-371: This point comes across more clearly in the Fig. 3's legend. Perhaps rewrite? We agree this section was awkwardly written and Lines 376-382 now read 'The relevance to final methane concentrations is demonstrated by considering the values reported by the University of Hawaii for PAC2 samples (Fig. 1b). An almost 30% increase in final methane concentration occurs from the use of the calibration equation in Figure 3c, compared to Figure 3a. This derives from a measured peak area for methane of 62 for a sample with a volume of 0.076 L and a seawater density of 1024 kg m-3, yielding a final methane concentration of 2.1 and 2.8 nmol kg-1 using the equations from Figure 3a and 3c, respectively.'

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). The intercomparison of methane and nitrous oxide used typical shipboard sampling procedures as it replicated typical sampling and storage procedures. Future intercomparison exercises will have the ability to manipulate concentrations of methane

and nitrous concentrations. The University of Hawaii is awaiting delivery of a large (200 liter) equilibrator unit. The 200 L capacity is smaller than the 760 L equilibrator used to produce reference material for dissolved inorganic carbon by Andrew Dickson, but it will allow us to produce reference material of varying concentrations on demand.

Page 15: Why was the variability higher for the BAL5 dataset? Could this be related to sampling and/or storage? The BAL5 samples had the highest concentrations of nitrous oxide sampled from the Baltic Sea and were associated with high inter-laboratory variability. We believe that the high variability is caused to a large extent by the non-linear response of the ECD. Differences in calibration procedures by the different laboratories, as shown in Figure 6, become exacerbated for high concentrations of nitrous oxide. If sampling and/or storage were the primary causes of the variability, we would have expected to see equally high variability in the samples with lower concentrations.

Page 16, lines 438-439: Was this only true for samples with methane concentrations less than atmospheric concentrations? Yes, it appears that low concentration samples are more susceptible to an increase due to contamination.

Page 18, line 512: What would be their maximum recommended storage time? For samples with very low or high concentrations, analysis within 2 months is recommended. For samples with concentrations equivalent to or exceeding atmospheric equilibrium, analysis could be conducted within a slightly longer timeframe e.g. 6 months.

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)? We report on lines 550-552 that 'The low concentrations of nitrous oxide still exceed detection limits by at least an order of magnitude for even the less-sensitive headspace method due to the high sensitivity of the ECD.' In response to an earlier comment by Reviewer 1, we have now included a brief comparison of the detection limits for headspace equilib-

rium and purge-and-trap in the Introduction and Lines 89-95 now read 'The purge and trap technique is typically more sensitive by 2-3 orders of magnitude over headspace equilibrium. However, the purge and trap technique requires more time for sample analysis and it is more difficult to automate the injection of samples into the gas analyzer. Headspace equilibrium sampling is most suited for volatile compounds that can be efficiently partitioned into the headspace gas volume from the seawater sample. Its limited sensitivity can be compensated by large volume analysis e.g. (Upstill-Goddard et al., 1996).'

Page 20, lines 560-565: Other important points, e.g., sample volume, septa/seals used, preservative used, should also be included in future efforts. We agree with this comment, and have modified Section 4.3 in the Discussion to address this point.

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). We agree with this comment which is why we also suggested using air from compressed gas cylinder after cross-checking its concentration. This is more likely to be relevant for methane than nitrous oxide. Lines 605-607 read 'The air used in the equilibration process could be sourced from the ambient environment if sufficiently stable or from a compressed gas cylinder after cross-checking the concentration with the appropriate gas standard.'

Page 20, line 586: Bourbonnais et al. (2017) also used air-equilibrated seawater standards to calculate water-column nitrous oxide concentrations off Peru. Thank you for this reference, it is now included in the manuscript

Figures 1: Are values of methane at atmospheric equilibrium expected at 25 m depth? Is this in the mixed layer? At Station ALOHA, the mixed layer depth nearly always exceeds 25 m during the winter months (November-March). During the expedition in February 2017 when the samples were collected, the mixed layer depth ranged from 110-130 m. We have now reported this in the text on Lines 176-177.

Figure 7: Are these relationships significant (add r2)? 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. The r2 value is included for each of the regression lines shown in Figure 7a and 7b. We completely agree with the Reviewer's comment that the same laboratory needs to conduct a time-course set of measurements for a thorough analysis of storage effects. This was not part of the intercomparison work, but is clearly needed for a Best Practice Guide which is being planned.

Tables 6 and 7: Add detection limits for each laboratory. We considered including detection limits, but did not include them in this Supplementary Table. This is because detection limits can be lowered (improved) by increasing the sample volume (for purge-and-trap method) or altering the ratio of water to headspace (for the headspace equilibrium method). In Column 3 of Tables 6 and 7, published references have been included for the majority of the laboratories. These include more in-depth description of the individual methods than can be provided here.

Add last name "Macarena Burgos" as done for all other researchers. Done

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

Page 18, line 501: change "equilibration" for "equilibrated". Changed

Page 19, line 545: change to "switching between different calibration curves." Changed

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-274, 2018.