

Author's response

6 September 2018

We would like to thank the editor and the three Reviewers for their detailed comments, which have improved the manuscript. We have included a point-by-point response to the comments below, with our responses highlighted in bold text. The revised manuscript is attached after our response and the changes are highlighted in red text.

Sam Wilson, on behalf of all the coauthors

-----  
Editor's comments

Thank you for your detailed responses to three reviewer reports. Please revise your manuscript according to your responses and also considering my own suggestions provided below. To expedite the final review process, I would like to ask you to make all the changes "clearly visible" in a marked-up manuscript and a point-by-point reply to the reviewer and my own comments.

1. Contributing factors for large inter-laboratory discrepancies

In general agreement with the very positive evaluations provided by three reviewers, I think that your work would contribute greatly to raising awareness of inter-laboratory analytical discrepancies among research communities employing CH<sub>4</sub> and N<sub>2</sub>O measurements not only in the marine environment but also in various freshwater systems. Many researchers, including myself, would be interested in understanding major contributing factors for the reported (surprisingly large) inter-laboratory differences. The current version (and also your responses) do not provide an easy-to-follow summary of inter-lab differences in major procedures (starting from sampling storage) and an in-depth discussion of the relative importance of sample storage, gas extraction methods, and analytical accuracy. Details are now scattered through the text and in supplementary tables, so it is very difficult to attribute major findings to inter-lab differences in sample storage/preparation/analysis. As reviewers indicated, this issue of relative contribution is not trivial, because sample preparation and storage could override analytical inaccuracies. I wondered if you could provide an overview diagram or table (in the main manuscript) showing major inter-lab differences in sample storage (including gas sample storage after headspace equilibration), the used gas extraction techniques, and GC analysis. In discussing large analytical discrepancies, you could refer to this overview to allow readers to do some self-assessment of contributing factors.

**The editor points out that it would be helpful to have the information pertaining to all of the steps between sample collection and analysis and how these vary for each laboratory for both methane and nitrous oxide in a single Table. This is something that we originally attempted, however it became too unwieldy and could not have been accommodated within the main document. This is why we separated the methods (Supplementary Tables 6 and 7), storage times (Supplementary Table 5), and descriptions of gas standards (Section 4.1) and internal controls (Section 4.3). Therefore the information is dispersed throughout the manuscript as the Editor mentions, but it is difficult to see how it could be presented in an alternative way. Please also keep in mind that we wished to keep the identity of individual laboratories confidential. If we begin to match methods with some of the data, then this would compromise the anonymity of participating laboratories.**

**Finally, we hope that it is evident to the readers that improving the comparability of methane and nitrous oxide measurements takes time and this study is the first step towards making the improvements. As you point out, we have highlighted the variability that exists between laboratories and now we need to further develop the infrastructure to achieve a higher level of precision and accuracy for methane and nitrous oxide measurements. We are optimistic this can be achieved, particularly as the current level of variability is not dissimilar to the scale of variability observed for other oceanographic parameters (e.g. DIC, trace metals) when they first underwent similar exercises.**

**In talking with the scientists who have led these intercomparisons, there is often not a single ‘magic bullet’ that solves all of the problems. Sometimes the biggest advances were made when the independent scientists assembled in a laboratory to work through the methods collectively.**

## 2. Technical recommendations

In your response to a second reviewer comment, you just described your future plan for a "Best Practice Guide", indicating that this is not the primary objective. However, providing recommendations was implicated as an important goal in one of your four research questions (Q3). Please consider providing more detailed and specific recommendations, particularly in the concluding paragraph. For example, you could be more quantitative in your statements like “it is recommended to keep storage time to a minimum”.

**This is a good comment and thank you for pointing this out. We have re-structured Section 4.3 in the Discussion to better reflect our recommendations resulting from the intercomparison exercise.**

## 3. Uncertainties in addressing reviewer comments

Some of your responses might need more articulation in the revised text. For example, in your response to a first reviewer comment on stopper/septa contamination and leakage, you just mentioned potential contamination, not considering the leakage issue. Regarding a comment on preservative issue, you cited some potential problems but jumped to a TINA conclusion: “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.”. I wondered if this statement would be logically acceptable given the problems described in the preceding sentence. Please double check whether you have fully addressed all reviewer comments in preparing the revised version.

**We have revised the section of preservatives and Lines 187-193 now read ‘The choice of mercuric chloride as the preservative for dissolved methane and nitrous oxide was due to 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, it is not evident that they are a superior alternative to mercuric chloride’. It is possible that the ‘community-wide evaluation’ occurs in the near-future led by the scientists who use mercuric chloride to preserve Dissolve Inorganic Carbon samples. We also refer the editor to Section 4.3 of the Discussion which mentions the issues of both septa contamination and leakage during sample storage.**

**The underlying issue is that there is no perfect method to store dissolved gas samples. There is always the potential for a loss in sample integrity (i.e. change in concentration during storage). It is likely that loss of sample integrity is more prevalent in samples with higher particle loading, higher biomass, and either very high or very low concentrations.**

---

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 88-94 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 518-521 read '*An approximate working detection limit for methane analysis via headspace equilibration is 1 nmol kg<sup>-1</sup>, 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<sup>-1</sup> (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](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?

**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 459-464 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 adsorption) 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 586-595 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<sup>-1</sup>, 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 187-193 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, it is not evident that they are a superior alternative to mercuric chloride.'*

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 <sup>3</sup>H-CH<sub>4</sub> 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 nitrogen. 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 (-195°C) and either liquid nitrogen or cooled ethanol (-70°C) 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<sup>-3</sup>, yielding a final methane concentration of 2.1 and 2.8 nmol kg<sup>-1</sup> 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 549-551 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 equilibrium and purge-and-trap in the Introduction and Lines 88-94 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 602-604 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 175-176.**

Figure 7: Are these relationships significant (add r<sup>2</sup>)? 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 r<sup>2</sup> 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**

---

Referee #2

In their manuscript, Wilson et al. present data from a recent international intercomparison study which evaluated the analytical procedures used to measure the concentrations of methane and nitrous oxide dissolved in seawater. Specifically, seawater samples and gaseous standards were sent to several

different laboratories for analysis. Since the measurement of methane and nitrous oxide concentrations are mainly done in the gas, not liquid, phase, the different laboratories had different protocols to first separate the dissolved gas prior to analysis as well as the final analysis; while the different labs had different protocols, they mainly involved either headspace equilibration or a purge and trap technique. The results of this intercomparison are striking, with different laboratories reporting concentrations that could be different by several hundred percent. The highest percent differences were reported for the lowest concentration samples, and since low concentrations are typically reported in the near-surface waters, this inter-laboratory difference is particularly troubling for global extrapolation of sea-to-air fluxes for these two gases. The impact of this manuscript is that it identifies significant inconsistencies between laboratories, and while the data from any one laboratory is likely valid for testing hypotheses, combining data from multiple laboratories for global extrapolation or time series analysis will lead to significant unknowns. At the end of the manuscript, the reader is left hungry for more, wondering how these inconsistencies might be rectified with a hypothetical Standard Operating Procedure. But while the authors provide a few recommendations for how to lower uncertainties, they do not prove the major cause of these inconsistencies, and thus which procedure might be preferred. The authors appropriately did not attempt this recommendation as it was beyond what their data can illuminate. For example, a full analysis of the headspace equilibration procedure would require each laboratory to establish the accuracy and precision of each variable in Equation 1 (pressure, temperature, salinity, headspace volume, and water volume) using their procedures. The authors assess the calibration of the analytical instrument and the variability of the overall results, but not these specific variables. In addition, the authors recognize that storage time is a variable significantly influencing the results. Since these additional variables were not systematically investigated, the authors are correct in not recommending a preferred procedure, and instead choose to report overall inconsistencies.

**We thank Reviewer #2 for their comments. We are building on the results from this intercomparison exercise and in the future will have a Best Practice Guide for the measurements of dissolved methane and nitrous oxide.**

Sample storage: I recommend that the authors expand section 3.4. I found this section too brief on experimental details and I was left assuming how storage time was assessed. Was the sample storage time variable controlled in any systemic way or is this simply the time it took different labs to actually conduct their analyses? Is there any way to normalize the data in Figures 1 and 4 to sample storage time or would that be extending this data too far? Can the authors assess how much variation in the dissolved concentrations is due to storage vs. procedure?

**The specific questions are answered separately below. In response to the general comment, we have re-structured Section 3.4 to improve its clarity. Lines 447-458 now read “Because prolonged samples storage 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<sup>-1</sup>, 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.’**

Was the sample storage time variable controlled in any systemic way or is this simply the time it took different labs to actually conduct their analyses?

**The sample storage time represents the time taken for different laboratories to conduct the analysis. There was no control of the storage time.**

Is there any way to normalize the data in Figures 1 and 4 to sample storage time or would that be extending this data too far?

**We would be uncomfortable doing this conversion because it would insinuate a higher influence of sample storage on concentrations than what we can currently prove. We refer the readers to Figure 7 which shows concentration and coefficient variation against storage time for the samples with the lowest concentration of methane.**

Can the authors assess how much variation in the dissolved concentrations is due to storage vs. procedure?

**This would require a time-course set of measurements which was not conducted as part of this exercise. This would be a very interesting experiment and could feature in future intercomparisons. What we have noted in our response to the overall comment, is that contamination is considered most likely for the samples of methane collected from the Pacific Ocean. These samples had methane concentrations of 0.9 and 2.3 nmol kg<sup>-1</sup> and therefore were most sensitive to release of small quantities of hydrocarbons by the septa.**

The authors suggest that leakage may be a source of uncertainty for longer storage times, but they don't raise the possibility of inadequate preservation. Most groups analyzing these dissolved gases assume that adding enough mercuric chloride to a sample will halt all biological activity, but that may not be the case. In addition, what is the chance that gases are outgassing or adsorbing to the stopper? Since these are both possible influences on the final results, I suggest that the authors also briefly raise these possibilities.

**In response to the comments made by Reviewer #2, we have restructured the relevant part of the Discussion to specifically address the issue of sample storage. Lines 586-595 now read *'This study also revealed that sample storage time can be an important factor. Specially, the results from this study corroborate the findings of Magen et al. (2014) who showed that samples with low concentrations of methane and 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 which interfere with the dissolved methane in the sample (Niemann et al., 2015). Since the release of hydrocarbons occurs over a period 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. Due to all of these reasons, it is recommended to conduct an evaluation of sample storage time for the environment that is being sampled.'***

Please note that in response to comments by Reviewer #1 we addressed the issue about alternatives to mercuric chloride and Lines 187-193 now read *'The choice of mercuric chloride as the preservative for dissolved methane and nitrous oxide was due to 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, it is not evident that they are a superior alternative to mercuric chloride.'*

Overall, this investigation appears robust and the manuscript is well written. The authors have uncovered a significant result which will benefit the community.

**Thank you for your comments**

---

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.

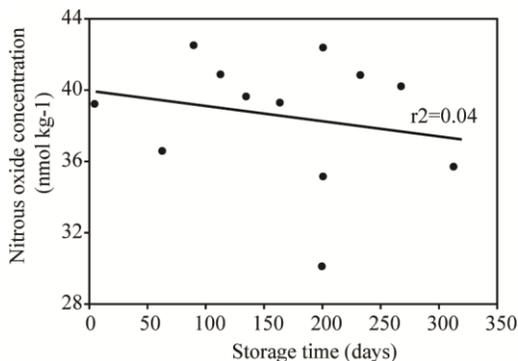
**The values of coefficient of variation (%) shown in Figure 7b are associated with methane concentrations measured by each lab for PAC1 and PAC2 samples (collected in February 2017). These specific values are not included in any of the Supplementary Material tables, where we instead report the mean 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 385-387 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 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.

**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 447-458 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 <sup>-1</sup> )	Coeffic. Variation (%)	Methane conc (nmol kg <sup>-1</sup> )	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 166-170 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 913 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 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.

**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 N<sub>2</sub>O. 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 N<sub>2</sub>O, 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.

1 **An intercomparison of oceanic methane and nitrous oxide measurements**

2  
3 Samuel T. Wilson<sup>1\*</sup>, Hermann W. Bange<sup>2</sup>, Damian L. Arévalo-Martínez<sup>2</sup>, Jonathan Barnes<sup>3</sup>,  
4 Alberto V. Borges<sup>4</sup>, Ian Brown<sup>5</sup>, John L. Bullister<sup>6</sup>, Macarena Burgos<sup>1,7</sup>, David W. Capelle<sup>8</sup>,  
5 Michael Casso<sup>9</sup>, Mercedes de la Paz<sup>10†</sup>, Laura Farías<sup>11</sup>, Lindsay Fenwick<sup>8</sup>, Sara Ferrón<sup>1</sup>, Gerardo  
6 Garcia<sup>11</sup>, Michael Glockzin<sup>12</sup>, David M. Karl<sup>1</sup>, Annette Kock<sup>2</sup>, Sarah Laperriere<sup>13</sup>, Cliff S.  
7 Law<sup>14,15</sup>, Cara C. Manning<sup>8</sup>, Andrew Marriner<sup>14</sup>, Jukka-Pekka Myllykangas<sup>16</sup>, John W.  
8 Pohlman<sup>9</sup>, Andrew P. Rees<sup>5</sup>, Alyson E. Santoro<sup>13</sup>, Philippe D. Tortell<sup>8</sup>, Robert C. Upstill-  
9 Goddard<sup>3</sup>, David P. Wisegarver<sup>6</sup>, Guiling L. Zhang<sup>17</sup>, Gregor Rehder<sup>12</sup>

10  
11 <sup>1</sup>University of Hawai'i at Manoa, Daniel K. Inouye Center for Microbial Oceanography:  
12 Research and Education (C-MORE), Honolulu, Hawai'i, USA

13 <sup>2</sup>GEOMAR Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20 24105 Kiel,  
14 Germany

15 <sup>3</sup>Newcastle University, School of Natural and Environmental Sciences, Newcastle upon Tyne,  
16 UK

17 <sup>4</sup>Université de Liège, Unité d'Océanographie Chimique, Liège, Belgium

18 <sup>5</sup>Plymouth Marine Laboratory, Plymouth, UK

19 <sup>6</sup>National Oceanic and Atmospheric Administration, Pacific Marine Environmental Laboratory,  
20 Seattle, Washington, USA

21 <sup>7</sup>Universidad de Cádiz, Instituto de Investigaciones Marinas, Departamento Química-Física  
22 Cádiz, Spain

23 <sup>8</sup>University of British Columbia, Vancouver, Department of Earth, Ocean and Atmospheric  
24 Sciences, British Columbia, Canada

25 <sup>9</sup>U.S. Geological Survey, Woods Hole Coastal and Marine Science Center, Woods Hole, USA

26 <sup>10</sup>Instituto de Investigaciones Marinas, Vigo, Spain

27 <sup>11</sup>University of Concepción, Department of Oceanography and Center for climate research and  
28 resilience (CR2), Concepción, Chile

29 <sup>12</sup>Leibniz Institute for Baltic Sea Research Warnemünde, Rostock, Germany

30 <sup>13</sup>University of California Santa Barbara, Department of Ecology, Evolution, and Marine  
31 Biology, Santa Barbara, USA

32 <sup>14</sup>National Institute of Water and Atmospheric Research (NIWA), Wellington, New Zealand

33 <sup>15</sup>Department of Chemistry, University of Otago, Dunedin, New Zealand

34 <sup>16</sup>University of Helsinki, Department of Environmental Sciences, Helsinki, Finland

35 <sup>17</sup>Ocean University of China, Department of Marine Chemistry, Qingdao, China

36

37 †Current address: Instituto Español de Oceanografía, Centro Oceanográfico de A Coruña, A

38 Coruña, Spain

39

40 \*corresponding author: stwilson@Hawaii'i.edu

41 **Abstract.** Large scale climatic forcing is impacting oceanic biogeochemical cycles and is  
42 expected to influence the water-column distribution of trace gases including methane and nitrous  
43 oxide. Our ability as a scientific community to evaluate changes in the water-column inventories  
44 of methane and nitrous oxide depends largely on our capacity to obtain robust and accurate  
45 concentration measurements which can be validated across different laboratory groups. This  
46 study represents the first formal, international, intercomparison of oceanic methane and nitrous  
47 oxide measurements whereby participating laboratories received batches of seawater samples  
48 from the subtropical Pacific Ocean and the Baltic Sea. Additionally, compressed gas standards  
49 from the same calibration scale were distributed to the majority of participating laboratories to  
50 improve the analytical accuracy of the gas measurements. The computations used by each  
51 laboratory to derive the dissolved gas concentrations were also evaluated for inconsistencies (*e.g.*  
52 pressure and temperature corrections, solubility constants). The results from the intercomparison  
53 and intercalibration provided invaluable insights into methane and nitrous oxide measurements.  
54 It was observed that analyses of seawater samples with the lowest concentrations of methane and  
55 nitrous oxide had the lowest precisions. In comparison, while the analytical precision for  
56 samples with the highest concentrations of trace gases was better, the variability between the  
57 different laboratories was higher; 36% for methane and 27% for nitrous oxide. In addition, the  
58 comparison of different batches of seawater samples with methane and nitrous oxide  
59 concentrations that ranged over an order of magnitude revealed the ramifications of different  
60 calibration procedures for each trace gas. Overall, this study builds upon the intercomparison  
61 results to develop a framework for improving oceanic methane and nitrous oxide measurements,  
62 with the aim of precluding future analytical discrepancies between laboratories.

## 63 1. Introduction

64 The increasing mole fractions of greenhouse gases in the Earth's atmosphere are causing long-  
65 term climate change with unknown future consequences. Two greenhouse gases, methane and  
66 nitrous oxide, together contribute approximately 23% of total radiative forcing attributed to well-  
67 mixed greenhouse gases (Myhre et al., 2013). It is imperative that the monitoring of methane  
68 and nitrous oxide in the Earth's atmosphere is accompanied by measurements at the Earth's  
69 surface to better inform the sources and sinks of these climatically important trace gases. This  
70 includes measurements of dissolved methane and nitrous oxide in the marine environment,  
71 which is an overall source of both gases to the overlying atmosphere (Nevison et al., 1995;  
72 Anderson et al., 2010; Naqvi et al., 2010; Freing et al., 2012; Ciais et al., 2014).

73 Oceanic measurements of methane and nitrous oxide are conducted as part of established  
74 time-series locations, along hydrographic survey lines, and during disparate oceanographic  
75 expeditions. Within low to mid-latitude regions of the open ocean, the surface waters are  
76 frequently slightly super-saturated with respect to atmospheric equilibrium for both methane and  
77 nitrous oxide. There is typically an order of magnitude range in concentration along a vertical  
78 water-column profile at any particular open ocean location (e.g. Wilson et al., 2017). In contrast  
79 to the open ocean, near-shore environments, which are subject to river inputs, coastal upwelling,  
80 benthic exchange and other processes, have higher concentrations and greater spatial and  
81 temporal heterogeneity (e.g. Schmale et al., 2010; Upstill-Goddard and Barnes, 2016).

82 Methods for quantifying dissolved methane and nitrous oxide have evolved and somewhat  
83 diverged since the first measurements were made in the 1960s (Craig and Gordon 1963;  
84 Atkinson and Richards 1967). Some laboratories employ purge-and-trap methods for extracting  
85 and concentrating the gases prior to their analysis (e.g. Zhang et al., 2004; Bullister and  
86 Wisegarver, 2008; Capelle et al., 2015; Wilson et al., 2017). Others equilibrate a seawater  
87 sample with an overlying headspace gas and inject a fixed volume of the gaseous phase into a  
88 gas analyzer (e.g. Upstill-Goddard et al., 1996; Walter et al., 2005; Farias et al., 2009). **The**  
89 **purge and trap technique is typically more sensitive by 2-3 orders of magnitude over headspace**  
90 **equilibrium. However, the purge and trap technique requires more time for sample analysis and**  
91 **it is more difficult to automate the injection of samples into the gas analyzer. Headspace**  
92 **equilibrium sampling is most suited for volatile compounds that can be efficiently partitioned**  
93 **into the headspace gas volume from the seawater sample. Its limited sensitivity can be**

94 compensated by large volume analysis (e.g. Upstill-Goddard et al., 1996). Additional  
95 developments for continuous underway surface seawater measurements use equilibrator systems  
96 of various designs coupled to a variety of detectors (e.g. Weiss et al., 1992; Butler et al., 1989;  
97 Gülzow et al., 2011; Arévalo-Martínez et al., 2013). Determining the level of analytical  
98 comparability between different laboratories for discrete samples of methane and nitrous oxide is  
99 an important step towards improved comprehensive global assessments. Such intercomparison  
100 exercises are critical to determining the spatial and temporal variability of methane and nitrous  
101 oxide across the world oceans with confidence, since no single laboratory can single-handedly  
102 provide all the required measurements at sufficient resolution. Previous comparative exercises  
103 have been conducted for other trace gases e.g. carbon dioxide, dimethylsulphide, and sulfur  
104 hexafluoride (Dickson et al., 2007; Bullister and Tanhua, 2010; Swan et al., 2014) and for trace  
105 elements (Cutter et al., 2013). These exercises confirm the value of the intercomparison concept.

106 To instigate this process for methane and nitrous oxide, a series of international  
107 intercomparison exercises were conducted between 2013 and 2017, under the auspices of  
108 Working Group #143 of the Scientific Committee on Oceanic Research (SCOR) ([www.scor-](http://www.scor-int.org)  
109 [int.org](http://www.scor-int.org)). Discrete seawater samples collected from the subtropical Pacific Ocean and the Baltic  
110 Sea were distributed to the participating laboratories (Table 1). The samples were selected to  
111 cover a representative range of concentrations across marine locations, from the oligotrophic  
112 open ocean to highly productive waters, and in some instances sub-oxic, coastal waters. An  
113 integral component of the intercomparison exercise was the production and distribution of  
114 methane and nitrous oxide gas standards to members of the SCOR Working Group. The  
115 intercomparison exercise was conceived and evaluated with the following four questions in  
116 mind:

117 Q1. What is the agreement between the SCOR gas standards and the ‘in-house’ gas standards  
118 used by each laboratory?

119 Q2. How do measured values of dissolved methane and nitrous oxide compare across  
120 laboratories?

121 Q3. Despite the use of different analytical systems, are there general recommendations to reduce  
122 uncertainty in the accuracy and precision of methane and nitrous oxide measurements?

123 Q4. What are the implications of inter-laboratory differences for determining the spatial and  
124 temporal variability of methane and nitrous oxide in the oceans?

125

## 126 **2. Methods**

### 127 **2.1 Calibration of nitrous oxide and methane using compressed gas standards**

128 Laboratory-based measurements of oceanic methane and nitrous oxide require separation of the  
129 dissolved gas from the aqueous phase, with the analysis conducted on the gaseous phase.

130 Calibration of the analytical instrumentation used to quantify the concentration of methane and  
131 nitrous oxide is nearly always conducted using compressed gas standards, the specifics of which  
132 vary between each laboratory. Therefore, the reporting of methane and nitrous oxide datasets  
133 ought to be accompanied by a description of the standards used, including their methane and  
134 nitrous oxide mole fractions, the declared accuracies, and the composition of their balance or  
135 ‘make-up’ gas. For both gases, the highest accuracy commercially available standards have  
136 mole fractions close to current day atmospheric values. These standards can be obtained from  
137 national agencies including National Oceanic and Atmospheric Administration Global  
138 Monitoring Division (NOAA GMD), the National Institute of Metrology China, and the Central  
139 Analytical Laboratories of the European Integrated Carbon Observation System Research  
140 Infrastructure (ICOS-RI). By comparison, it is more difficult to obtain highly accurate methane  
141 and nitrous oxide gas standards with mole fractions exceeding modern-day atmospheric values.  
142 This is particularly problematic for nitrous oxide due to the nonlinearity of the widely used  
143 Electron Capture Detector (ECD) (Butler and Elkins, 1991).

144 The absence of a widely available high mole fraction, high accuracy nitrous oxide gas  
145 standard was noted as a primary concern at the outset of the intercomparison exercise.  
146 Therefore, a set of high-pressure primary gas standards was prepared for the SCOR Working  
147 Group by John Bullister and David Wisegarver at NOAA Pacific Marine and Environmental  
148 Laboratory (PMEL). One batch, referred to as Air Ratio Standard (ARS), had methane and  
149 nitrous oxide mole fractions similar to modern air and the other batch, referred to as Water Ratio  
150 Standard (WRS) had higher methane and nitrous oxide mole fractions for calibration of high  
151 concentration water samples. These SCOR primary standards were checked for stability over a  
152 12 month period and assigned mole fractions on the same calibration scale, known as ‘SCOR-  
153 2016.’ A comparison was conducted with NOAA standards prepared on the SIO98 calibration  
154 scale for nitrous oxide and the NOAA04 calibration scale for methane. Based on the comparison  
155 with NOAA standards, the uncertainty of the methane and nitrous oxide mole fractions in the

156 ARS and the uncertainty of the methane mole fraction in the WRS were all estimated at better  
157 than 1%. By contrast, the uncertainty of the nitrous oxide mole fraction in the WRS was  
158 estimated at 2-3%. The gas standards were distributed to twelve of the laboratories involved in  
159 this study (Table 1). The technical details on the production of the gas standards and their  
160 assigned absolute mole fractions is included in Bullister et al. (2016).

161

## 162 **2.2 Collection of discrete samples of nitrous oxide and methane**

163 Dissolved methane and nitrous oxide samples for the intercomparison exercise were collected  
164 from the subtropical Pacific Ocean and the Baltic Sea. Pacific samples were obtained on 28  
165 November 2013 and 24 February 2017 from the Hawai'i Ocean Time-series (HOT) long-term  
166 monitoring site, Station ALOHA, located at 22.75 N, 158.00 W. The November 2013 samples  
167 are included in Figure S1 and S2 in the Supplement, but are not discussed in the main Results or  
168 Discussion because fewer laboratories were involved in the initial intercomparison, and the  
169 results from these samples support the same conclusions obtained with the more recent sample  
170 collections. Seawater was collected using Niskin-like bottles designed by John Bullister (NOAA  
171 PMEL), which help minimize contamination of trace gases, in particular chlorofluorocarbons  
172 and sulfur hexafluoride (Bullister and Wisegarver, 2008). The bottles were attached to a rosette  
173 with a conductivity-temperature-depth (CTD) package. Seawater was collected from two depths:  
174 700 m and 25 m, where the near-maximum and minimum water-column concentrations for  
175 methane and nitrous oxide at this location can be found. **The 25 m samples were always well  
176 within the surface mixed layer, which ranged from 100 to 130 m depth during sampling.**

177 Replicate samples were collected from each bottle, with one replicate reserved for analysis at the  
178 University of Hawai'i to evaluate variability between sampling bottles. Seawater was dispensed  
179 from the Niskin-like bottles using Tygon® tubing into the bottom of borosilicate glass bottles,  
180 allowing overflow of at least two sample volumes and ensuring the absence of bubbles. Most  
181 sample bottles were 240 mL in size and were sealed with no headspace using butyl-rubber  
182 stoppers and aluminum crimp-seals. A few laboratory groups requested smaller crimp-sealed  
183 glass bottles ranging from 20-120 mL in volume and two laboratories used 1 L glass bottles  
184 which were closed with a glass stopper and sealed with Apiezon® grease. Seawater samples  
185 were collected in quadruplicate for each laboratory. All samples were preserved using saturated  
186 mercuric chloride solution (100 µL of saturated mercuric chloride solution per 100 mL of

187 seawater sample) and stored in the dark at room temperature until shipment. The choice of  
188 mercuric chloride as the preservative for dissolved methane and nitrous oxide was due to its long  
189 history of usage. It is recognized that other preservatives have been proposed (e.g. Magen et al.,  
190 2014, Bussmann et al., 2015), however pending a community-wide evaluation of their  
191 effectiveness over a range of microbial assemblages and environmental conditions for both  
192 methane and nitrous oxide, it is not evident that they are a superior alternative to mercuric  
193 chloride.

194 Samples from the western Baltic Sea were collected during 15-21 October 2016, onboard the  
195 R/V *Elisabeth Mann Borgese* (Table 2). Since the Baltic Sea consists of different basins with  
196 varying concentrations of oxygen beneath permanent haloclines (Schmale et al., 2010), a larger  
197 range of water-column methane and nitrous oxide concentrations were accessible for inter-  
198 laboratory comparison compared to Station ALOHA. For all seven Baltic Sea stations, the  
199 water-column was sampled into an on-deck 1,000 L water tank that was subsequently  
200 subsampled into discrete sample bottles. At three stations (BAL1, BAL3, and BAL6), the water  
201 tank was filled from the shipboard high-throughput underway seawater system. For deeper  
202 water-column sampling at the stations BAL2, BAL4, and BAL5, the water tank was filled using  
203 a pumping CTD system (Strady et al., 2008) with a flow rate of 6 L min<sup>-1</sup> and a total pumping  
204 time of approximately 3 h. For the final deep water-column station, BAL7, the pump that  
205 supplied the shipboard underway system was lowered to a depth of 21 m to facilitate a shorter  
206 pumping time of approximately 20 mins. Subsampling the water tank for all samples took  
207 approximately 1 h in total and the total sampling volume was less than 100 L. To verify the  
208 homogeneity of the seawater during the sampling process, the first and last samples collected  
209 from the water tank were analyzed by Newcastle University onboard the research vessel. In  
210 contrast to the Pacific Ocean sampling, which predominantly used 240 mL glass vials, each  
211 laboratory provided their own preferred vials and stoppers for the Baltic Sea samples. Seawater  
212 samples were collected in triplicate for each laboratory. All samples were preserved with 100  
213 µL of saturated mercuric chloride solution per 100 ml of seawater sample, with the exception of  
214 samples collected by U.S. Geological Survey, who analyzed unpreserved samples onboard the  
215 research vessel.

216

### 217 **2.3. Sample analysis**

218 Each laboratory measured dissolved methane and nitrous oxide slightly differently. A full  
219 description of each laboratory's method can be found in Table S6 and Table S7 in the  
220 Supplement for methane and nitrous oxide, respectively.

221 The majority of laboratories measured methane and nitrous oxide by equilibrating the  
222 seawater sample with an overlying headspace and subsequently injecting a portion of the gaseous  
223 phase into the gas analyzer. This method has been conducted since the 1960s when gas  
224 chromatography was first used to quantify dissolved hydrocarbons (McAuliffe, 1963). The  
225 headspace was created using helium, nitrogen, or high-purity air to displace a portion of the  
226 seawater sample within the sample bottle. Alternatively, a subsample of the seawater was  
227 transferred to a gas-tight syringe and the headspace gas subsequently added. The volume of the  
228 vessel used to conduct the headspace equilibration ranged from 20 ml borosilicate glass vials to 1  
229 L glass vials and syringes used by Newcastle University and U.S. Geological Survey,  
230 respectively. The dissolved gases equilibrated with the overlying headspace at a controlled  
231 temperature for a set period of time that ranged from 20 min to 24 h. The equilibration process  
232 was typically enhanced by some initial period of physical agitation. After equilibration, an  
233 aliquot of the headspace was transferred into the gas analyzer (GA) by either physical injection,  
234 displacement using a brine solution, or injection using a switching valve. Some laboratories  
235 incorporated a drying agent and a carbon dioxide scrubber prior to analysis. The gas sample  
236 passed through a multi-port injection valve containing a sample loop of known volume, which  
237 transferred the gas sample directly onto the analytical column within the oven of the GA.  
238 Calibration of the instrument was achieved by passing the gas standards through the injection  
239 valve.

240 The final gas concentrations using the headspace equilibration method was calculated by:

241

$$242 \quad [1] \quad C_{gas} [\text{nmol L}^{-1}] = \left( \beta x P V_{wp} + \frac{xP}{RT} V_{hs} \right) / V_{wp}$$

243

244 where  $\beta$  is the Bunsen solubility of nitrous oxide (Weiss and Price, 1980) or methane  
245 (Wiesenburg and Guinasso, 1979) in  $\text{nmol L}^{-1} \text{atm}^{-1}$ ,  $x$  is the dry gas mole fraction (ppb)  
246 measured in the headspace,  $P$  is the atmospheric pressure (atm),  $V_{wp}$  is the volume of water  
247 sample (mL),  $V_{hs}$  is the volume (mL) of the created headspace,  $R$  is the gas constant (0.08205746

248 L atm K<sup>-1</sup>mol<sup>-1</sup>), and  $T$  is equilibration temperature in Kelvin (K). An example calculation is  
249 provided in Table S8 in the Supplement.

250 In contrast to the headspace equilibrium method, five laboratories used a purge-and-trap  
251 system for methane and/or nitrous oxide analysis (Table S6 and Table S7 in the Supplement).  
252 These systems were directly coupled to a Flame Ionization Detector (FID) or ECD, with the  
253 exception of University of British Columbia, where a quadrupole mass spectrometer with an  
254 electron impact ion source and Faraday cup detector were used (Capelle et al., 2015). The  
255 purge-and-trap systems were broadly similar, each transferring the seawater sample to a sparging  
256 chamber. Sparging times typically ranged from 5-10 min and the sparge gas was either high  
257 purity helium or high purity nitrogen. **In addition to commercially available gas scrubbers,**  
258 **purification of the sparge gas was achieved by passing it through stainless steel tubing packed**  
259 **with Poropak Q and immersed in liquid nitrogen. This is a recommended precaution to**  
260 **consistently achieve a low blank signal of methane.** The elutant gas was dried using Nafion or  
261 Drierite, and subsequently cryotrapped on a sample loop packed with Poropak Q to aid retention  
262 of methane and nitrous oxide. **Cryotrapping was achieved for methane using liquid nitrogen (-**  
263 **195°C) and either liquid nitrogen or cooled ethanol (-70°C) for nitrous oxide.** Subsequently, the  
264 valve was switched to inject mode and the sample loop was rapidly heated to transfer its contents  
265 onto the analytical column. Calibration was achieved by injecting standards via sample loops  
266 using multi-port injection valves. Injection of standards upstream of the sparge chamber allowed  
267 for calibration of the purge-and-trap gas handling system, in addition to the GA. Calculation of  
268 the gas concentrations using the purge-and-trap method was achieved by application of the ideal  
269 gas law to the standard gas measurements:

270 [2]  $PV = nRT$

271 where  $P$ ,  $R$ , and  $T$  are the same as Equation 1,  $V$  represents the volume of gas injected (L),  
272 and  $n$  represents moles of gas injected. Rearranging Equation 2 yields the number of moles of  
273 methane or nitrous oxide gas for each sample loop injection of compressed gas standards. These  
274 values were used to determine a calibration curve based on the measured peak areas of the  
275 injected standards, and thereafter derive the number of moles measured for each unknown  
276 sample. To calculate concentrations of methane or nitrous oxide in a water sample, the number  
277 of moles measured were divided by the volume (L) of seawater sample analyzed. An example  
278 calculation is provided in Table S8 in the Supplement.

279

## 280 **2.4 Data analysis**

281 The final concentrations of methane and nitrous oxide are reported in  $\text{nmol kg}^{-1}$ . The analytical  
282 precision for each batch of samples obtained by each of the individual laboratories was estimated  
283 from the analysis of replicate seawater samples and reported as the coefficient of variation (%).  
284 The values reported by each laboratory for all the batches of seawater samples are shown in  
285 Tables S1 to S4 in the Supplement. Due to the observed inter-laboratory variability, it is likely  
286 that the median value of methane and nitrous oxide for each batch of samples does not represent  
287 the absolute *in situ* concentration. As this complicates the analytical accuracy for each  
288 laboratory, we instead calculated the percentage difference between the median concentration  
289 determined for each set of samples and the mean value reported by an individual laboratory. The  
290 presence of outliers was established using the Interquartile Range (IQR) and by comparing with  
291 one standard deviation applied to the overall median value.

292

## 293 **3. Results**

### 294 **3.1 Comparison of methane and nitrous oxide gas standards**

295 Six laboratories compared their existing ‘in-house’ standards of methane with the SCOR  
296 standards. This was done by calibrating in-house standards and deriving a mixing ratio for the  
297 SCOR standards which were treated as unknowns. Four laboratories reported methane values for  
298 either the ARS or WRS within 3% of their absolute concentration, whereas two laboratories  
299 reported an offset of 6% and 10% between their in-house standards and the SCOR standards  
300 (Table S6 in the Supplement). For those laboratories who measured the SCOR standards to  
301 within 3% or better accuracy, observed offsets in methane concentrations from the overall  
302 median cannot be due to the calibration gas.

303 Seven laboratories compared their own in-house standards of nitrous oxide with the prepared  
304 SCOR standards. Six laboratories reported values of nitrous oxide for the ARS which were  
305 within 3% of the absolute concentration, with the remaining laboratory reporting an offset of  
306 10% (Table S7 in the Supplement). The majority of these laboratories (five out of six groups)  
307 compared the SCOR ARS with NOAA GMD standards, which have a balance gas of air instead  
308 of nitrogen. Some laboratories with analytical systems that incorporated fixed sample loops (*e.g.*  
309 1 or 2 ml loops housed in a 6-port or 10-port injection valve) had difficulty analyzing the WRS,

310 as the peak areas created by the high mole fraction of the standard exceeded the signal typically  
311 measured from in-house standards or acquired by sample analysis, by an order of magnitude.  
312 The high mole fraction of the WRS was not an issue when multiple sample loops of varying  
313 sizes were incorporated into the analytical system, which was the case for purge-and-trap based  
314 designs. For the two laboratories with an in-house standard of comparable mole fraction to the  
315 WRS, an offset of 3% and a >20% offset was reported.

316

### 317 **3.2 Methane concentrations in the intercomparison samples**

318 Overall, median methane concentrations in seawater samples collected from the Pacific Ocean  
319 and the Baltic Sea ranged from 0.9 to 60.3 nmol kg<sup>-1</sup> (Table 2). Out of 101 reported values, 3  
320 outliers were identified using the IQR criterion and were not included in further analysis. The  
321 methane data values for each batch of samples analyzed by each laboratory, including the mean  
322 and standard deviation, the number of samples analyzed, and the % offset from the overall  
323 median value are reported in Table S1 and Table S2 in the Supplement. Analysis conducted by  
324 the University of Hawai'i of methane and nitrous oxide from each Niskin-like bottle used in the  
325 Pacific Ocean sampling did not reveal any bottle-to-bottle differences. Furthermore, analysis by  
326 Newcastle University showed there was no difference between the first and the last set of  
327 samples collected from the 1000 L collection used in the Baltic Sea sampling.

328 The two Pacific Ocean sampling sites had the lowest water-column concentrations of  
329 methane (Fig. 1a and 1b). The PAC1 samples collected from within the mesopelagic zone,  
330 where methane concentrations have been reported to be less than 1 nmol kg<sup>-1</sup> (Reeburgh, 2007;  
331 Wilson et al., 2017), showed a distribution of reported concentrations skewed towards the higher  
332 values. For the PAC1 samples, seven out of twelve laboratories reported values ≤1 nmol kg<sup>-1</sup>  
333 and the mean coefficient of variation for all laboratories was 11% (Table 2). In contrast to the  
334 mesopelagic samples, the methane concentrations for the near-surface seawater samples (PAC2)  
335 were close to atmospheric equilibrium (Fig. 1b). Measured concentrations of methane for PAC2  
336 samples ranged from 1.9 to 3.8 nmol kg<sup>-1</sup> and the mean coefficient of variation for all  
337 laboratories was 7%. Similar to the PAC1 samples, PAC2 also had a distribution of data skewed  
338 towards the higher concentrations.

339 Three Baltic Sea sampling sites (BAL1, BAL3, and BAL6) had median methane  
340 concentrations that ranged from 4.1 to 5.7 nmol kg<sup>-1</sup> (Fig. 1c). The BAL1 samples also showed a

341 skewed distribution of reported values towards higher concentrations, as seen in PAC1 and  
342 PAC2 samples. However, this was not evident in BAL3 or BAL6, which had the closest  
343 agreement between the reported methane concentrations. For these three sets of Baltic Sea  
344 samples, the mean coefficient of variation for all laboratories ranged from 4% (BAL3) to 9%  
345 (BAL1). The next three Baltic Sea samples (BAL4, BAL5, and BAL7) had methane  
346 concentrations that ranged from 18.8 to 35.4 nmol kg<sup>-1</sup> (Fig. 1d). These three sets of samples  
347 had a normal distribution of data and the closest agreement between the reported concentrations  
348 for all of the Pacific Ocean and Baltic Sea samples. Furthermore, for these three sets of samples,  
349 the mean coefficient of variation for all laboratories was 4% (Table 2). The final Baltic Sea  
350 sample (BAL2) had the highest concentrations of methane, with a median reported value of 60.3  
351 nmol kg<sup>-1</sup>, and a large range of values (45.2 to 67.2 nmol kg<sup>-1</sup>; Fig. 1e). The BAL2 samples had  
352 the lowest overall mean coefficient of variation for all laboratories; 2% (Table 2).

353 Further analysis of the data was conducted to better comprehend the factors that caused the  
354 observed inter-laboratory variability in methane measurements. The deviation from median  
355 values was calculated for each sample collected from the Baltic Sea (Fig. 2). The Pacific Ocean  
356 samples (PAC1 and PAC2) were not included in this analysis due to the skewed distribution of  
357 data. There were also some instances in the Baltic Sea samples, where the median concentration  
358 might not have realistically represented the absolute *in situ* methane concentration. This was  
359 most likely to have occurred at low concentrations due to the skewed distribution of reported  
360 concentrations (*e.g.* BAL1) or at high concentrations where there was a large range in reported  
361 values (*e.g.* BAL2). The results revealed that a few laboratories (Datasets D, F, and G) were  
362 consistently within or close to 5% of the median value for all batches of seawater samples (Fig.  
363 2). Some laboratories (*e.g.* Datasets B, C, and H) had a higher deviation from the median value  
364 at higher methane concentrations. Two laboratories (Datasets J and K) had a higher deviation  
365 from the median value at lower methane concentrations. Finally, in some cases it was not  
366 possible to determine a trend (Datasets A and E), due to the variability.

367 The reasons behind the trends for each dataset became more apparent when considering the  
368 effect of the inclusion or exclusion of low standards in the calibration curve on the resulting  
369 derived concentrations (Fig. 3). The FID has a linear response to methane at nanomolar values  
370 and therefore a high level of accuracy across a relatively wide range of *in situ* methane  
371 concentrations can be obtained with the correct slope and intercept. To demonstrate this,

372 calibration curves for methane were provided by the University of Hawai'i. These revealed  
373 minimal variation in the slope value when calibration points were increased from low mole  
374 fractions (Fig. 3a) to higher mole fractions (Fig. 3b). However, the intercept value was sensitive  
375 to the range of calibration values used, and this effect was further exacerbated when only the  
376 higher calibration points were included (*i.e.* Fig. 3c). **The relevance to final methane**  
377 **concentrations is demonstrated by considering the values reported by the University of Hawai'i**  
378 **for PAC2 samples (Fig. 1b). An almost 30% increase in final methane concentration occurs**  
379 **from the use of the calibration equation in Figure 3c, compared to Figure 3a. This derives from a**  
380 **measured peak area for methane of 62 for a sample with a volume of 0.076 L and a seawater**  
381 **density of 1024 kg m<sup>-3</sup>, yielding a final methane concentration of 2.1 and 2.8 nmol kg<sup>-1</sup> using the**  
382 **equations from Figure 3a and 3c, respectively. With this understanding on the effect of FID**  
383 **calibration, we consider it likely that the increased deviation from median values at high methane**  
384 **concentrations (Datasets B, C, and H) results from differences in calibration slope between each**  
385 **laboratory. In contrast, the datasets with a higher offset at low methane concentrations (Datasets**  
386 **J and K) could be due to erroneous low standard values causing a skewed intercept. In addition,**  
387 **there may be other factors including sample contamination, discussed in Section 3.4.**

388

### 389 **3.3 Nitrous oxide concentrations in the intercomparison samples**

390 Overall, median nitrous oxide concentrations in seawater samples collected from the Pacific  
391 Ocean and the Baltic Sea ranged from 3.4 to 42.4 nmol kg<sup>-1</sup> (Table 2). Of the 113 reported  
392 values, ten outliers were identified using the IQR criterion and were not included in further  
393 analysis. The nitrous oxide data values for each batch of samples analyzed by each laboratory,  
394 including the mean and standard deviation, the number of samples analyzed, and the % offset  
395 from the overall median value are reported in Table S3 and Table S4 in the Supplement.

396 For six sets of seawater samples, BAL1, BAL2, BAL3, BAL6, BAL7, and PAC2, the  
397 concentrations of nitrous oxide were close to atmospheric equilibrium. The reported values  
398 ranged from 7.7 to 12.7 nmol kg<sup>-1</sup> in the Baltic Sea (Fig. 4a) and from 5.9 to 7.6 nmol kg<sup>-1</sup> in the  
399 Pacific Ocean (Fig. 4b). For the Pacific Ocean near-surface (**mixed layer**) sampling site (PAC2),  
400 the theoretical value of nitrous oxide concentration in equilibrium with the overlying atmosphere  
401 is also shown (Fig. 4b). For these six samples with concentrations close to atmospheric

402 equilibrium, the mean coefficient of variation for all laboratories ranged from 3% (BAL3 and  
403 PAC2) to 5% (BAL1) (Table 2).

404 For the three other sets of samples (BAL4, BAL5, and PAC1), the nitrous oxide  
405 concentrations deviated significantly from atmospheric equilibrium (Fig. 4c, 4d, and 4e). At one  
406 sampling site, BAL4 (Fig. 4c), nitrous oxide was under-saturated with respect to atmospheric  
407 equilibrium and reported concentrations ranged from 2.1–5.5 nmol kg<sup>-1</sup>. As observed in the low  
408 concentration Pacific Ocean methane samples, there was a skewed distribution of the data  
409 towards the higher nitrous oxide concentrations. The BAL4 samples also had the highest  
410 variability (*i.e.* lowest precision), with a mean coefficient of variation of 8% (Table 2). The two  
411 remaining samples (PAC1 and BAL5) had much higher concentrations of nitrous oxide, as  
412 expected for low-oxygen regions of the water-column. In contrast to the samples with near  
413 atmospheric equilibrium concentrations of nitrous oxide, there was a low overall agreement  
414 between the independent laboratories for PAC1 and BAL5 nitrous oxide concentrations (Fig. 4d,  
415 4e). At PAC1 and BAL5, nitrous oxide concentrations ranged from 34.3–45.8 nmol kg<sup>-1</sup> (Fig.  
416 4d) and 30.1–45.9 nmol kg<sup>-1</sup>, respectively (Fig. 4e). The mean coefficient of variation for all  
417 laboratories was 4% for BAL5 samples compared to 3% for PAC1 samples.

418 The deviation of individual nitrous oxide concentrations from the median value provides  
419 insight into the variability associated with their measurements (Fig. 5). The BAL1 dataset was  
420 not included in this analysis due to its skewed data distribution and the high inter-laboratory  
421 variability for BAL5 indicated that the median value may differ from the absolute nitrous oxide  
422 concentration for this sample. For the low nitrous oxide Baltic Sea and Pacific Ocean samples  
423 (Fig. 5a), the majority of data points were within 5% of the median values. Furthermore, for the  
424 majority of laboratories, the data points for separate seawater samples clustered together  
425 indicating some consistency to the extent they varied from the overall median value. Exceptions  
426 to this observation include Datasets E, C, L, and K (Fig. 5a) which demonstrated varying  
427 precision and accuracy. At high nitrous oxide concentrations (Fig. 5b), there are fewer data  
428 points within 5% of the median value compared to low nitrous oxide concentrations (Fig. 5a).  
429 Therefore, for PAC1 and BAL5 samples, 6 and 7 data points fall within 5% of the median value,  
430 respectively. Furthermore, only three laboratories (Datasets F, G, and K) had data for both  
431 Pacific Ocean and Baltic Sea samples within 5% of the median value. This could have been

432 caused by inconsistent analysis between different batches of samples or by variable sample  
433 collection and transportation.

434 The likely factors that caused these offsets in nitrous oxide concentrations among  
435 laboratories include sample analysis and calibration of the gas analyzers. Calibration of the ECD  
436 is nontrivial and at least two prior publications have discussed nitrous oxide calibration issues  
437 (Butler and Elkins, 1991; Bange et al., 2001). The laboratories participating in the nitrous oxide  
438 intercomparison employed different calibration procedures (Fig. 6). Some used a linear fit and  
439 kept their analytical peak areas within a narrow range (Fig. 6a), while others used a step-wise  
440 linear fit and therefore used different slopes for low and high nitrous oxide mole fractions (Fig.  
441 6b). Finally, some applied a polynomial curve (Fig. 6c) and sometimes two different polynomial  
442 fits, for low and high concentrations. The difficulty in calibrating the ECD was evidenced by the  
443 deviation from median values as multiple datasets show good precision but consistent offsets at  
444 the lowest (Fig. 5a) and highest (Fig. 5b) final concentrations of nitrous oxide.

445

### 446 **3.4 Sample storage and sample bottle size**

447 Because prolonged storage of samples can influence dissolved gas concentrations, including  
448 methane and nitrous oxide, the intercomparison dataset was analyzed for sample storage effects  
449 (Table S5 in the Supplement). It should, however, be noted that assessing the effect of storage  
450 time on sample integrity was not a formal goal of the intercomparison exercise and replicate  
451 samples were not analyzed at repeated intervals by independent laboratories, as would normally  
452 be required for a thorough analysis. Nonetheless our results did provide some insights into  
453 potential storage-related problems. Most notably, there were indications that an increase in  
454 storage time caused increased concentrations and increased variability for methane samples with  
455 low concentrations, *i.e.* PAC1 and PAC2 samples which had median methane concentrations of  
456 0.9 and 2.3 nmol kg<sup>-1</sup>, respectively (Fig. 7). In comparison, for samples of nitrous oxide with  
457 low concentrations there was no trend of increasing values as observed for samples with low  
458 methane concentrations.

459 Another variable which differed between laboratories for the intercomparison exercise was  
460 the size of samples bottle, which ranged from 25 ml to 1 liter for the different laboratories.  
461 There was no observed difference between the methane and nitrous oxide values obtained from  
462 the various sampling bottles and it was concluded that sampling bottles were not a controlling

463 factor for the observed differences between laboratories. We note, however, the potential for  
464 greater air bubble contamination in smaller bottles.

465

#### 466 **4. Discussion**

467 The marine methane and nitrous oxide analytical community is growing. This is reflected in the  
468 increasing number of corresponding scientific publications and the resulting development of a  
469 global database for methane and nitrous oxide (Bange et al., 2009). Like all Earth observation  
470 measurements, there is a need for intercomparison exercises of the type reported here, for data  
471 quality assurance, and for appropriate reporting practices (National Research Council, 1993). To  
472 the best of our knowledge, the work presented here is the first formal intercomparison of  
473 dissolved methane and nitrous oxide measurements. Based on our results, we discuss the lessons  
474 learned and our recommendations moving forward, by addressing the four questions that were  
475 posed in the Introduction.

476

##### 477 **4.1 What is the agreement between the SCOR gas standards and the ‘in-house’ gas 478 standards used by each laboratory?**

479 It is typical for laboratories to source some, or all, of their compressed gas standards from  
480 commercial suppliers. National agencies, such as NOAA GMD or National Institute of  
481 Metrology China, also provide standards to the scientific community. The national agencies  
482 typically offer a lower range in concentrations than commercial suppliers, but their standards  
483 tend to have a higher level of accuracy. Of the twelve laboratories participating in the  
484 intercomparison, eight reported using national agency standards, with seven of them using gases  
485 sourced from NOAA GMD. Since the methane and nitrous oxide mole fractions of these  
486 national agency standards are equivalent to modern-day atmospheric mixing ratios, they are  
487 similar to the SCOR ARS distributed to the majority of laboratories in this study. Laboratories  
488 in receipt of the SCOR standards were asked to predict their mole fractions based on those of  
489 their own in-house standards. For the majority that conducted this exercise, there was good  
490 agreement (<3% difference) between the NOAA GMD and the SCOR ARS for both methane  
491 and nitrous oxide. For three laboratories, a larger offset was observed between the NOAA GMD  
492 and the SCOR ARS. There was also a good prediction for the higher methane content SCOR  
493 WRS, facilitated by the linear response of the FID (Fig. 3). In contrast, the nitrous oxide mole

494 fraction in the SCOR WRS exceeded the typical working range for several laboratories and it  
495 was difficult for them to cross-compare with their in-house standards. This reflects an analytical  
496 set-up that involves on-column injection via a 6-port or 10-port valve with one or two sample  
497 loops, respectively. The sample loops have a fixed volume and their inaccessibility makes it  
498 difficult to replace them by a smaller loop size. Therefore either dilution of the standard is  
499 required, or smaller loops need to be incorporated into the calibration protocol. The two  
500 laboratories that compared their in-house standards with the SCOR WRS reported an offset of  
501 3% and >20%. This indicates that variability between standards can be an issue for obtaining  
502 accurate dissolved concentrations and provides support for the production of a widely available  
503 high concentration nitrous oxide standard. We strongly recommend that all commercially  
504 obtained standards are cross-checked against primary standards, such as the SCOR ARS and  
505 WRS. This should be conducted at least at the beginning and end of their use to detect any drift  
506 that may have occurred during their lifetime. With due diligence and care, the SCOR standards  
507 provide the capability for cross-checking personal standards for years to decades (Bullister et al.,  
508 2016).

509

#### 510 **4.2 How do measured values of methane and nitrous oxide compare across laboratories?**

511 **Methane:** The methane intercomparison highlighted the variability that exists between  
512 measurements conducted by independent laboratories. At low methane concentrations, a skewed  
513 distribution of methane data was observed, which was particularly evident in PAC1 (Fig. 1a).  
514 Potential causes include calibration procedures (Section 3.2) and/or sample contamination which  
515 is more prevalent at low concentrations (Section 3.4). For some laboratories, the low methane  
516 concentrations are close to their detection limit, which is determined by the relatively low  
517 sensitivity of the FID and the small number of moles of methane in an introduced headspace  
518 **equilibrated** with seawater. An approximate working detection limit for methane analysis via  
519 headspace equilibration is  $1 \text{ nmol kg}^{-1}$ , although some laboratories improve upon this by having  
520 a large aqueous: gaseous phase ratio during the equilibration process (*e.g.* Upstill-Goddard et al.,  
521 1996). Depending upon the volume of sample analyzed, purge-and-trap analysis can have a  
522 detection limit much lower than  $1 \text{ nmol kg}^{-1}$  (*e.g.* Wilson et al., 2017). Methane measurements  
523 in aquatic habitats with methane concentrations near the limit of analytical detection include  
524 mesopelagic and high latitude environments distal from coastal or benthic inputs (*e.g.* Rehder et

525 al., 1999; Kitidis et al., 2010; Fenwick et al., 2017). Of additional concern is that the skewed  
526 distribution of methane concentrations also occurs in samples collected both from the surface  
527 ocean (PAC2; Fig. 1b) and coastal environments (BAL1; Fig. 1c). Methane concentrations  
528 between 2–6 nmol kg<sup>-1</sup> are within the detection limit of all participating laboratories. To address  
529 this we recommend that laboratories restrict sample storage to the minimum time required to  
530 analyze the samples and incorporate internal controls into their sample analysis (Section 4.4).

531 There was an improvement in the overall agreement between the laboratories for samples  
532 with higher methane concentrations. However, some of the highest variability between the  
533 laboratories was observed at the highest concentrations of methane analyzed (BAL2; Fig. 1e).  
534 This high degree of variability resulted in significant uncertainty in the absolute *in situ*  
535 concentration. Methane concentrations of this magnitude and higher are found in coastal  
536 environments (Zhang et al., 2004; Jakobs et al., 2014; Borges et al., 2017) and in the water-  
537 column associated with seafloor emissions (*e.g.* Pohlman et al., 2011). These environments are  
538 considered vulnerable to climate induced changes and eutrophication, and therefore it is  
539 necessary that independent measurements are conducted to the highest possible accuracy to  
540 allow for inter-laboratory and inter-habitat comparisons. To address this we recommend that  
541 reference material be produced and distributed between laboratories.

542  
543 **Nitrous oxide:** Some of the trends discussed for methane were also evident in the nitrous oxide  
544 data. For the samples with the lowest nitrous oxide concentrations a skewed data distribution  
545 was observed, as found for methane (Fig. 4c). Such low nitrous oxide concentrations are typical  
546 of low-oxygen water-column environments (<10 μmol kg<sup>-1</sup>). Therefore, the analytical bias  
547 towards measuring values higher than the absolute *in situ* concentrations is particularly pertinent  
548 to oceanographers measuring nitrous oxide in oxygen minimum zones and other low-oxygen  
549 environments (Naqvi et al., 2010; Farías et al., 2015; Ji et al., 2015). The low concentrations of  
550 nitrous oxide still exceed detection limits by at least an order of magnitude for even the less-  
551 sensitive headspace method due to the high sensitivity of the ECD. Therefore, the bias towards  
552 reporting elevated values for low concentrations of nitrous oxide is related less to analytical  
553 sensitivity and is more a consequence of calibration issues. During the intercomparison exercise  
554 ECD calibration was identified as a nontrivial issue for all participating laboratories and it  
555 deserves continuing attention. In particular, the nonlinearity of the ECD means that low and

556 high nitrous oxide concentrations are more vulnerable to error since the values fall outside of the  
557 most frequented part of the calibration curve. This is particularly true if a linear fit is used to  
558 calibrate the ECD (Fig. 6a). To circumvent this problem, one laboratory used a step-wise linear  
559 function while other laboratories used a quadratic function. The usefulness of multiple  
560 calibration curves for low and high nitrous oxide concentrations was highlighted during the  
561 intercomparison exercise, although this necessitates some consideration of the threshold for  
562 switching **between** different calibration curves.

563 The majority of seawater samples analyzed had nitrous oxide concentrations ranging from 7–  
564 11 nmol kg<sup>-1</sup> (Fig. 4a, 4b), which are close to atmospheric equilibrium values, as shown for the  
565 Pacific Ocean (Fig. 4b). Collective analysis of these samples gives insight into the precision and  
566 accuracy associated with surface water nitrous oxide analysis (Fig 5a). This is discussed further  
567 in the context of implementing internal controls for methane and nitrous oxide (Section 4.4). For  
568 samples with the highest nitrous oxide concentrations, *i.e.* exceeding 30 nmol kg<sup>-1</sup>, there was  
569 high variability between the concentrations reported by the independent laboratories. This was  
570 most evident for the BAL5 samples (Fig. 4e) and similar to the variability observed at the highest  
571 methane concentrations analyzed (Fig. 1e). It is difficult to assess how much of this variability  
572 was specifically due to the differences in calibration practices between the laboratories and the  
573 differences in gas standards with high nitrous oxide mole fractions, but at least some of it can be  
574 attributed to this. These results form the basis for a proposed production of reference material  
575 for both trace gases.

576

### 577 **4.3 Are there general recommendations to reduce uncertainty in the accuracy and** 578 **precision of methane and nitrous oxide measurements?**

579 **There are several analytical recommendations resulting from this study. The use of highly**  
580 **accurate standards and the appropriate calibration fit is an essential requirement for both**  
581 **headspace equilibration and the purge-and-trap technique. It was shown that both analytical**  
582 **approaches can yield comparable values for methane and nitrous oxide, with the main**  
583 **differences observed at low methane concentrations. At sub-nanomolar methane concentrations,**  
584 **four out of the six laboratories that reported methane concentrations <1 nmol kg<sup>-1</sup> used a purge-**  
585 **and-trap analysis.**

586 This study also revealed that sample storage time can be an important factor. Specifically,  
587 the results from this study corroborate the findings of Magen et al. (2014) who showed that  
588 samples with low concentrations of methane and more susceptible to increased values as a result  
589 of contamination. The contamination was most likely due to the release of methane and other  
590 hydrocarbons from the septa (Niemann et al., 2015). Since the release of hydrocarbons occurs  
591 over a period of time, it is recommended to keep storage time to a minimum and to store samples  
592 in the dark. It should be noted that sample integrity can also be compromised due to other  
593 factors including inadequate preservation, outgassing, and adsorption of gases onto septa. For all  
594 these reasons, it is recommended to conduct an evaluation of sample storage time for the  
595 environment that is being sampled.

596 One useful item that was not included as part of the intercomparison exercise but can help  
597 decrease uncertainty in the accuracy and precision of methane and nitrous oxide measurements  
598 are internal control measurements. Internal controls represent a self-assessment quality control  
599 check to validate the analytical method and quantify the magnitude of uncertainty. Appropriate  
600 internal controls for methane and nitrous oxide consist of air-equilibrated seawater samples.  
601 Their purpose is to provide checks for methane concentrations ranging from 2–3 nmol kg<sup>-1</sup> and  
602 for nitrous oxide concentrations from 5–9 nmol kg<sup>-1</sup>. The air used in the equilibration process  
603 could be sourced from the ambient environment if sufficiently stable or from a compressed gas  
604 cylinder after cross-checking the concentration with the appropriate gas standard. Air-  
605 equilibrated samples provide reassurance that the analytical system is providing values within the  
606 correct range. Air-equilibrated samples also indicate the certainty associated with calculating the  
607 saturation state of the ocean with respect to atmospheric equilibrium. This is particularly  
608 relevant when the seawater being sampled is within a few percent of saturation. Finally, these  
609 air-equilibrated samples provide an estimate of analytical accuracy, which is infrequently  
610 reported for methane or nitrous oxide. At present, only a few studies report the analysis of air-  
611 equilibrated seawater alongside water-column samples (Bullister and Wisegarver, 2008; Capelle  
612 et al., 2015; Bourbonnais et al., 2017; Wilson et al., 2017). It is likely that wider implementation  
613 would facilitate internal assessment of the analytical system. Since the main equipment required  
614 is a water-bath and an overhead stirrer, the production is not cost-prohibitive. A  
615 recommendation of this intercomparison exercise is that laboratories routinely use air-  
616 equilibrated seawater samples to provide an estimate of analytical accuracy.

617 In addition to the self-assessments provided by the analysis of air-equilibrated seawater, this  
618 study revealed the need for reference seawater to help assess the accuracy of high concentration  
619 methane and nitrous oxide measurements. Reference seawater in this instance refers to batches  
620 of dissolved methane and nitrous oxide samples prepared in the laboratory using an equilibrator  
621 set-up, as used for dissolved inorganic carbon (Dickson et al., 2007). In the absence of plans for  
622 additional intercomparison exercises, the provision of reference seawater will allow laboratories  
623 to continue evaluating their own measurements. **Finally, the lessons learned during the**  
624 **intercomparison exercises will be the basis for a forthcoming Good Practice Guide for dissolved**  
625 **methane and nitrous oxide.**

626

#### 627 **4.4 What are the implications of interlaboratory differences for determining the spatial and** 628 **temporal variability of methane and nitrous oxide in the oceans?**

629 The key outcome of this study was the identification of differences in methane and nitrous oxide  
630 concentrations for the same batch of seawater samples measured by several independent  
631 laboratories. Emergent from this is the distinct possibility that any given laboratory will  
632 incorrectly report data, thereby increasing uncertainty over the saturation states of both gases.  
633 The tendency to over-estimate methane concentrations close to atmospheric equilibrium means  
634 that marine emissions of methane to the overlying atmosphere will be also overestimated (Bange  
635 et al., 1994; Upstill-Goddard and Barnes, 2016). In contrast, for nitrous oxide there does not  
636 appear to be either an under-estimation or over-estimation of concentrations. Consequently, there  
637 is generally a lower inherent uncertainty in its surface ocean saturation state, as previously  
638 proposed (Law and Ling, 2001; Forster et al., 2009).

639 The inter-laboratory differences highlighted by this study should be viewed in the context of  
640 numerous individual efforts to assess temporal and/or spatial trends in methane and nitrous oxide  
641 by way of time-series observations (Bange et al., 2010; Farías et al., 2015; Wilson et al., 2017;  
642 Fenwick and Tortell, 2018), repeat hydrographic survey lines (de la Paz et al., 2017), and single  
643 expeditions. While the value of these in integrating the behaviour of methane and nitrous oxide  
644 into the hydrography and biogeochemistry of local-regional ecosystems is beyond question, their  
645 value would be enhanced by the rigorous cross-validation of analytical protocols. Without this,  
646 perceived small temporal and/or spatial changes in water-column concentrations in any given  
647 region are difficult to verify unless the data all originate from a single laboratory. In addition,

648 the value of a global methane and nitrous oxide database (*e.g* Bange et al., 2009) would to some  
649 extent be compromised by the uncertainty. Taking due account of the analytical variability  
650 between laboratories will clearly be vital to any future assessment of the changing methane and  
651 nitrous oxide budgets of the oceans.

652

## 653 **5. Conclusions**

654 Overall, the intercomparison exercise was invaluable to the growing community of ocean  
655 scientists interested in understanding the dynamics of dissolved methane and nitrous oxide in the  
656 water-column. The level of agreement between independent measurements of dissolved  
657 concentrations was evaluated in the context of several contributing factors, including sample  
658 analysis, standards, calibration procedures, and sample storage time. Importantly, the  
659 intercomparison represents a concerted effort from the scientists involved to critically assess the  
660 quality of their data, and to initiate the steps required for further improvements.

661 Recommendations arising from the intercomparison include routine cross-calibration of working  
662 gas standards against primary standards, minimizing sample storage time, incorporating internal  
663 controls (air-equilibrated seawater) alongside routine sample analysis, and the future production  
664 of reference seawater for methane and nitrous oxide measurements. These efforts will help  
665 resolve temporal and spatial variability, which is necessary for constraining methane and nitrous  
666 oxide emissions from aquatic ecosystems and for evaluating the processes that govern their  
667 production and consumption in the water-column.

668 *Acknowledgements:*

669 **During the final stages of this work, our coauthor John Bullister passed away. The**  
670 **intercomparison exercise greatly benefited from John's scientific expertise on dissolved gases.**  
671 **He will be greatly missed by the oceanographic community.**

672 The methane and nitrous oxide intercomparison exercise was conducted as a Scientific  
673 Committee on Ocean Research (SCOR) Working Group which receives funding from the U.S.  
674 National Science Foundation (OCE-1546580). Pacific Ocean seawater samples were collected  
675 on HOT cruises which are supported by NSF (including the most recent OCE-1260164 to  
676 DMK). Baltic Sea seawater samples were collected during Cruise #142 of the RV *Elisabeth*  
677 *Mann Borgese*, with the ship-time provided by the Leibniz Institute for Baltic Sea Research  
678 Warnemünde. We thank Ligu Guo for help with sampling during the Baltic Sea cruise. The  
679 methane and nitrous oxide gas standards were produced via a Memorandum of Understanding  
680 between the University of Hawai'i and NOAA-PMEL. Funding for the gas standards was  
681 provided by for the Center for Microbial Oceanography: Research and Education (C-MORE;  
682 EF0424599 to DMK), SCOR, the EU FP7 funded Integrated non-CO<sub>2</sub> Greenhouse gas  
683 Observation System (InGOS) (Grant Agreement #284274), and NOAA's Climate Program  
684 Office, Climate Observations Division. Additional support was provided by the Gordon and  
685 Betty Moore Foundation #3794 (DMK), the Simons Collaboration on Ocean Processes and  
686 Ecology (SCOPE; #329108 to DMK), and the Global Research Laboratory Program (#  
687 2013K1A1A2A02078278 to DMK) through the National Research Foundation of Korea (NRF).  
688 AVB is a senior research associate at the FRS-FNRS. AES would like to acknowledge NSF  
689 OCE-1437310. MP would like to acknowledge the support of the Spanish Ministry of Economy  
690 and Competitiveness (CTM2015-74510-JIN). LF received financial support by FONDAP  
691 1511009 and FONDECYT N°1161138. Any use of trade names is for descriptive purposes and  
692 does not imply endorsement by the U.S. government

693 **References**

- 694 Anderson, B., Bartlett, K., Frolking, S., Hayhoe, K., Jenkins, J., and Salas, W.: Methane and  
695 nitrous oxide emissions from natural sources, Office of Atmospheric Programs, US EPA,  
696 EPA 430-R-10-001, Washington DC, 2010.
- 697 Arévalo-Martínez, D. L., Beyer, M., Krumbholz, M., Piller, I., Kock, A., Steinhoff, T.,  
698 Körtzinger, A., and Bange, H. W.: A new method for continuous measurements of oceanic  
699 and atmospheric N<sub>2</sub>O, CO and CO<sub>2</sub>: performance of off-axis integrated cavity output  
700 spectroscopy (OA-ICOS) coupled to non-dispersive infrared detection (NDIR), *Ocean Sci.*,  
701 9, 1071–1087, 2013.
- 702 Atkinson, L. P., and Richards, F. A.: The occurrence and distribution of methane in the marine  
703 environment, *Deep-Sea Res.*, 14, 673–684, 1967.
- 704 Bange, H. W., Bartell, U. H., Rapsomanikis, S., and Andreae, M. O.: Methane in the Baltic and  
705 North Seas and a reassessment of the marine emissions of methane, *Global Biogeochem.*  
706 *Cycles*, 8, 465–480, doi:10.1029/94GB02181, 1994.
- 707 Bange, H. W., Rapsomanikis, S., and Andreae, M. O.: Nitrous oxide cycling in the Arabian Sea,  
708 *J. Geophys. Res.: Oceans*, 106, 1053–1065, 2001.
- 709 Bange H.W, Bell, T.G., Cornejo, M., Freing, A., Uher, G., Upstill-Goddard, R.C., and Zhang G.  
710 MEMENTO: a proposal to develop a database of marine nitrous oxide and methane  
711 measurements. *Env. Chem*, 6, 195–197, 2009.
- 712 Bange, H.W., Bergmann, K., Hansen, H.P., Kock, A., Koppe, R., Malien, F., and Ostrau, C.:  
713 Dissolved methane during hypoxic events at the Boknis Eck Time Series Station  
714 (Eckernförde Bay, SW Baltic Sea), *Biogeosciences.*, 7, 1279–1284, 2010.
- 715 Borges, A.V., Speeckaert, G., Champenois, W., Scranton, M.I., and Gypens, N.: Productivity and  
716 temperature as drivers of seasonal and spatial variations of dissolved methane in the Southern  
717 Bight of the North Sea, *Ecosystems*, 1–17, 2017.
- 718 Bourbonnais, A., Letscher, R. T., Bange, H. W., Échevin, V., Larkum, J., Mohn, J., N. Yoshida,  
719 N., and Altabet, M. A.: N<sub>2</sub>O production and consumption from stable isotopic and  
720 concentration data in the Peruvian coastal upwelling system, *Global Biogeochem. Cycles*, 31,  
721 678–698, 2017. doi:10.1002/2016GB005567.
- 722 Bullister, J. L., and Wisegarver, D. P.: The shipboard analysis of trace levels of sulfur  
723 hexafluoride, chlorofluorocarbon-11 and chlorofluorocarbon-12 in seawater, *Deep-Sea Res.*,  
724 55, 1063–1074, 2008.

725 Bullister, J. L., and Tanhua, T.: Sampling and measurement of chlorofluorocarbons and sulfur  
726 hexafluoride in seawater, IOCCP Report No. 14 ICPO Publication Series No. 134, Version 1,  
727 2010.

728 Bullister, J. L., Wisegarver, D. P., and Wilson, S. T.: The production of methane and nitrous  
729 oxide gas standards for Scientific Committee on Ocean Research (SCOR) Working Group  
730 #143, <http://udspace.udel.edu/handle/19716/23288>, 2016.

731 Bussmann, I., Matousu, A., Osudar, R. and Mau, S.: Assessment of the radio  $^3\text{H-CH}_4$  tracer  
732 technique to measure aerobic methane oxidation in the water column, *Limnol. Oceanogr.:*  
733 *Methods*, 13, 312–327, 2015.

734 Butler, J. H., Elkins, J. W., Thompson, T. M., and Egan, K. B.: Tropospheric and dissolved  $\text{N}_2\text{O}$   
735 of the west Pacific and east Indian Oceans during the El Niño Southern Oscillation Event of  
736 1987, *J. Geophys. Res.*, 94, 14,865–14,877, 1989.

737 Butler, J. H., and Elkins, J. W.: An automated technique for the measurement of dissolved  $\text{N}_2\text{O}$   
738 in natural waters, *Mar. Chem.*, 34, 47–61, 1991.

739 Capelle, D. W., Dacey, J. W., and Tortell, P. D.: An automated, high through-put method for  
740 accurate and precise measurements of dissolved nitrous oxide and methane concentrations in  
741 natural waters, *Limnol. Oceanogr.:* *Methods*, 13, 345–355, 2015.

742 Ciais, P., Dolman, A.J., Bombelli, A., Duren, R., Pregon, A., Rayner, P.J., Miller, C., Gobron,  
743 N., Kinderman, G., Marland, G., and Gruber, N.: Current systematic carbon-cycle  
744 observations and the need for implementing a policy-relevant carbon observing system,  
745 *Biogeosciences*, 11, 3547–3602, 2014.

746 Craig, H. and Gordon, L. I.: Nitrous oxide in the ocean and the marine atmosphere, *Geochim.*  
747 *Cosmochim. Acta* 27, 949–955, 1963.

748 Cutter, G. A.: Intercalibration in chemical oceanography - getting the right number. *Limnol.*  
749 *Oceanogr.:* *Methods*, 11, 418–424, 2013.

750 de la Paz, M., García-Ibáñez, M.I., Steinfeldt, R., Ríos, A.F., and Pérez, F.F.: Ventilation versus  
751 biology: What is the controlling mechanism of nitrous oxide distribution in the North  
752 Atlantic?, *Global Biogeochem. . Cycles*, 31, 745–760, doi: 10.1002/2016GB005507, 2017.

753 Dickson, A. G., Sabine, C. L. and Christian, J. R.: Guide to best practices for ocean  $\text{CO}_2$   
754 measurements, PICES Special Publication 3, 2007.

755 Farías, L., Castro-González, M., Cornejo, M., Charpentier, J., Faúndez, J., Boontanon, N. and  
756 Yoshida, N.: Denitrification and nitrous oxide cycling within the upper oxycline of the

757 eastern tropical South Pacific oxygen minimum zone. *Limnol. Oceanogr.*, 54, 132–144,  
758 2009.

759 Farías, L., Besoain, V., and García-Loyola, S.: Presence of nitrous oxide hotspots in the coastal  
760 upwelling area off central Chile: an analysis of temporal variability based on ten years of a  
761 biogeochemical time series, *Environ. Res. Lett.*, 10, 044017, 2015.

762 Fenwick, L., and Tortell, P. D.: Methane and nitrous oxide distributions in coastal and open  
763 ocean waters of the Northeast Subarctic Pacific during 2015–2016, *Mar. Chem.*, 200, 45–56,  
764 2018.

765 Fenwick, L., Capelle, D., Damm, E., Zimmermann, S., Williams, W. J., Vagle, S., and Tortell, P.  
766 D.: Methane and nitrous oxide distributions across the North American Arctic Ocean during  
767 summer, 2015, *J. Geophys. Res.: Oceans*, 122, 390–412, doi:10.1002/2016JC012493, 2017.

768 Forster, G., Upstill-Goddard, R. C., Gist, N., Robinson, C., Uher, G. and Woodward, E. M. S.:  
769 Nitrous oxide and methane in the Atlantic Ocean between 50 N and 52 S: latitudinal  
770 distribution and sea-to-air flux, *Deep-Sea Res.*, 56, 964–976, 2009.

771 Freing, A., Wallace, D. W. R., and Bange, H. W.: Global oceanic production of nitrous oxide,  
772 *Phil. Trans. R. Soc. B*, 367, 1245–1255, 2012.

773 Gülzow, W., Rehder, G., Schneider, B., Schneider, J., Deimling, V., and Sadkowiak, B.: A new  
774 method for continuous measurement of methane and carbon dioxide in surface waters using  
775 off-axis integrated cavity output spectroscopy (ICOS): An example from the Baltic Sea,  
776 *Limnol. Oceanogr.: Methods*, 9, 176–184, 2011.

777 Jakobs, G., Holterman, P., Berndmeyer, C., Rehder, G., Blumenberg, M., Jost, G., Nausch, G.,  
778 and Schmale, O.: Seasonal and spatial methane dynamics in the water column of the central  
779 Baltic Sea (Gotland Sea), *Cont. Shelf Res.*, 91, 12–25, 2014.

780 Ji, Q., Babbín, A. R., Jayakumar, A., Oleynik, S., and Ward, B. B.: Nitrous oxide production by  
781 nitrification and denitrification in the Eastern Tropical South Pacific oxygen minimum zone,  
782 *Geophys. Res. Lett.*, 42, 10,755–10,764, doi:10.1002/2015GL066853, 2015.

783 Kitidis, A., Upstill-Goddard, R. C., and Anderson, L. G.: Methane and nitrous oxide in surface  
784 water along the North-West Passage, Arctic Ocean, *Mar. Chem.*, 121, 80–86, 2010.

785 Law, C. S., and Ling, R. D.: Nitrous oxide flux and response to increased iron availability in the  
786 Antarctic Circumpolar Current, *Deep-Sea Res.*, 48, 2509–2527, 2001.

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

790 McAuliffe, C.: Solubility on water of C<sub>1</sub>-C<sub>9</sub> hydrocarbons, *Nature*, 200, 1092–1093, 1963.

791 Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestedt, J., Huang, J., Koch, D.,  
792 Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura,  
793 T., and Zhang, H.: Anthropogenic and Natural Radiative Forcing, In: *Climate Change 2013:*  
794 *The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report*  
795 *of the Intergovernmental Panel on Climate Change* [Stocker, T. F., Qin, D., Plattner, G.-K.,  
796 Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M.  
797 (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA,  
798 2013.

799 Naqvi, S. W. A., Bange, H. W., Farías, L., Monteiro, P. M. S., Scranton, M. I., and Zhang, J.:  
800 Marine hypoxia/anoxia as a source of CH<sub>4</sub> and N<sub>2</sub>O, *Biogeosciences*, 7, 215–2190, doi:  
801 10.5194/bg-7-2159-2010, 2010.

802 National Research Council: *Applications of analytical chemistry to oceanic carbon cycle studies.*  
803 Washington DC. National Academy Press, 1993.

804 Nevison, C. D., Weiss, R. F., and Erickson, D. J.: Global oceanic emissions of nitrous oxide, *J.*  
805 *Geophys. Res.*, 100, 15809–15820, doi: 10.1029/95JC00684, 1995.

806 Niemann, H., Steinle, L., Brees, J., Bussmann, I., Treude, T., Krause, S., Elvert, M., and  
807 Lehmann, M. F.: Toxic effects of lab-grade butyl rubber stoppers on aerobic methane  
808 oxidation, *Limnol. Oceanogr.: Methods*, 13, 40–52, 2015.

809 Pohlman, J. W., Bauer, J. E., Waite, W. F., Osburn, C. L., and Chapman, N. R.: Methane  
810 hydrate-bearing seeps as a source of aged dissolved organic carbon to the oceans, *Nature*  
811 *Geosci.*, 4, 37–41, 2011.

812 Reeburgh, W. S.: Oceanic methane biogeochemistry, *Chem. Rev.*, 107, 486–513, doi:  
813 10.1021/cr050362v, 2007.

814 Rehder, G., Keir, R. S., Suess, E., and Rhein, M.: Methane in the northern Atlantic controlled by  
815 microbial oxidation and atmospheric history, *Geophys. Res. Lett.*, 26, 587–590, doi:  
816 10.1029/1999GL900049, 1999.

817 Schmale, O., Schneider von Deimling, J., Gülzow, W., Nausch, G., Waniek, J. J. and Rehder, G.:  
818 Distribution of methane in the water column of the Baltic Sea. *Geophys. Res. Lett.*, 37,  
819 L12604, doi: 10.1029/2010GL043115, 2010.

820 Strady, E., Pohl, C., Yakushev, E. V., Krügera, S., and Hennings, U.: PUMP–CTD-System for  
821 trace metal sampling with a high vertical resolution. A test in the Gotland Basin, Baltic Sea,  
822 *Chemosphere*, 70, 1309–1319, 2008.

- 823 Swan, H. B., Armishaw, P., Iavetz, R., Alamgir, M., Davies, S. R., Bell, T. G., and Jones, G. B.:  
824 An interlaboratory comparison for the quantification of aqueous dimethylsulfide. *Limnol.*  
825 *Oceanogr.: Methods*, 12, 784–794, 2014.
- 826 Upstill-Goddard, R. C., Rees, A. P., and Owens, N. J. P.: Simultaneous high-precision  
827 measurements of methane and nitrous oxide in water and seawater by single phase  
828 equilibration gas chromatography, *Deep-Sea Res.*, 43, 1669–1682, 1996.
- 829 Upstill-Goddard, R. C., and Barnes, J.: Methane emissions from UK estuaries: Re-evaluating the  
830 estuarine source of tropospheric methane from Europe, *Mar. Chem.*, 180, 14–23, 2016.
- 831 Walter, S., Peeken, I., Lochte, K., Webb, A., and Bange, H. W.: Nitrous oxide measurements  
832 during EIFEX, the European Iron Fertilization Experiment in the subpolar South Atlantic  
833 Ocean, *Geophys. Res. Lett.*, 32, doi:10.1029/2005GL024619, 2005.
- 834 Weiss, R. F., and Price, B. A.: Nitrous oxide solubility in water and seawater, *Mar. Chem.*, 8,  
835 347–359, doi: 10.1016/0304-4203(80)90024-9, 1980.
- 836 Weiss, R. F., Van Woy, F. A., and Salameh, P. K.: Surface water and atmospheric carbon  
837 dioxide and nitrous oxide observation by shipboard automated gas chromatography: Results  
838 from expeditions between 1977 and 1990, Scripps Institution of oceanography Reference 92-  
839 11. ORNL/CDIAC-59, NDP-044. Carbon Dioxide Information Analysis Center, Oak Ridge  
840 National Laboratory, Tennessee, 1992.
- 841 Wiesenburg, D. A., and Guinasso, N. L.: Equilibrium solubilities of methane, carbon monoxide  
842 and hydrogen in water and seawater, *J. Chem. Eng. Data*, 24, 354–360, doi:  
843 10.1021/je60083a006, 1979.
- 844 Wilson, S. T., Ferrón, S., and Karl, D. M.: Interannual variability of methane and nitrous oxide in  
845 the North Pacific Subtropical Gyre, *Geophys. Res. Lett.*, 44, doi: 10.1002/2017GL074458,  
846 2017.
- 847 Zhang, G. L., Zhang, J., Kang, Y. B., and Liu, S. M.: Distributions and fluxes of methane in the  
848 East China Sea and the Yellow Sea in spring, *J. Geophys. Res.*, 109, C07011,  
849 doi:10.1029/2004JC002268, 2004

850 **Table 1.** List of laboratories that participated in the intercomparison. All laboratories measured  
 851 both methane and nitrous oxide except U.S. Geological Survey (methane only), U.C. Santa  
 852 Barbara (nitrous oxide only), and NOAA PMEL (nitrous oxide from the Pacific Ocean). Also  
 853 indicated are the twelve laboratories that received the SCOR gas standards of methane and  
 854 nitrous oxide.

<b>Institution</b>	<b>Lead Scientist</b>	<b>SCOR Standards</b>
University of Hawai'i, USA	Samuel Wilson	Yes
GEOMAR, Germany	Hermann Bange	Yes
Newcastle University, UK	Robert Upstill-Goddard	Yes
Université de Liège, Belgium	Alberto Vieira Borges	No
Plymouth Marine Laboratory, UK	Andrew Rees	Yes
NOAA PMEL, USA	John Bullister	Yes
IIM-CSIC, Spain	Mercedes de la Paz	Yes
CACYTMAR, Spain	Macarena Burgos	No
University of Concepción, Chile	Laura Farías	Yes
IOW, Germany	Gregor Rehder	Yes
University of California Santa Barbara, USA	Alyson Santoro	Yes
National Institute of Water and Atmospheric Research, NZ	Cliff Law	Yes
University British Columbia, Canada	Philippe Tortell	Yes
U.S. Geological Survey, USA	John Pohlman	No
Ocean University of China, China	Guiling Zhang	Yes

855

856

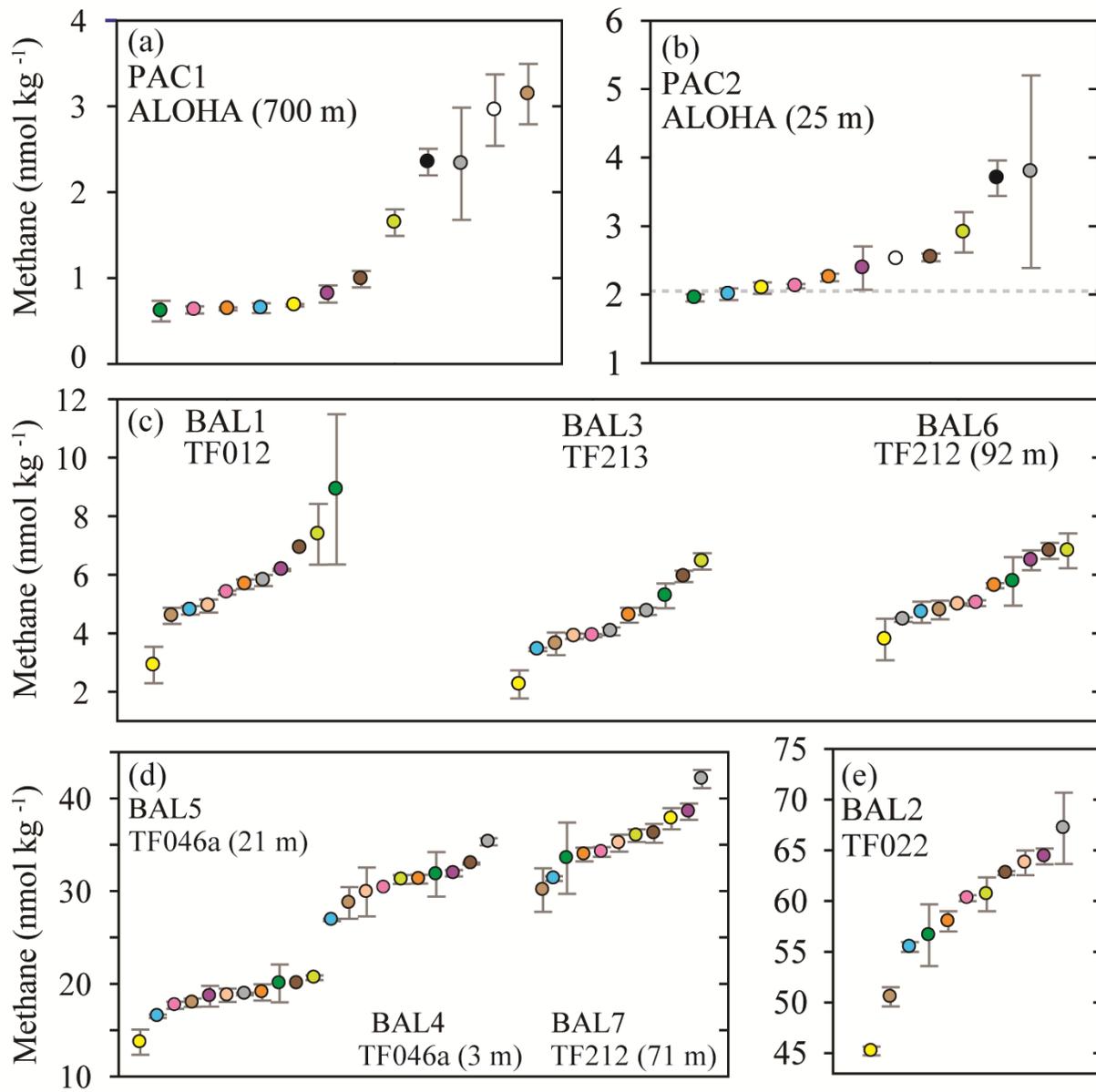
857 **Table 2.** Pertinent information for each batch of methane and nitrous oxide samples. This  
 858 includes contextual hydrographic information, median and mean concentrations of methane and  
 859 nitrous oxide, range, number of outliers, and the overall average coefficient of variation (%).  
 860

<b>Sampling parameters</b>									
Sample ID	PAC1	PAC 2	BAL1	BAL2	BAL3	BAL4	BAL5	BAL6	BAL7
Location	22.75N 158.00W	22.75N 158.00W	54.32N 11.55E	54.11N 11.18E	55.25N 15.98E	55.30N 15.80E	55.30N 15.80E	54.47N 12.21E	54.47N 12.21E
Location name	Station ALOHA	Station ALOHA	TF012	TF022	TF213	TF212	TF212	TF046a	TF046a
Sampling date	24.2.17	24.2.17	16.10.16	17.10.16	18.10.16	19.10.16	20.10.16	21.10.16	21.10.16
Sampling depth (m)	25	700	3	22	3	92	71	3	21
Seawater temperature (°C)	23.6	5.1	12.0	13.6	12.2	6.6	6.7	11.8	13.4
Salinity	34.97	34.23	13.85	17.37	7.87	18.40	18.08	8.81	17.65
Density (kg m <sup>-3</sup> )	1024	1027	1010	1013	1006	1014	1014	1006	1013
<b>Nitrous oxide</b>									
Number of datasets	13	13	12	13	12	13	12	13	12
Outliers	0	1	2	1	1	0	1	2	2
Median N <sub>2</sub> O conc. (nmol kg <sup>-1</sup> )	42.4	7.0	11.0	9.4	11.1	3.4	40.2	11.0	9.6
Mean N <sub>2</sub> O conc. (nmol kg <sup>-1</sup> )	41.3	7.0	11.1	9.2	11.0	3.4	39.0	10.8	9.5
Range	34.3-45.8	5.9-7.6	10.1-12.7	7.7-11.0	9.6-11.6	2.1-5.5	30.1-45.9	9.5-11.5	8.0-10.4
Average coeff. variation (%)	2.8	4.4	4.5	4.2	2.7	7.5	4.0	2.6	4.4
<b>Methane</b>									
Number of datasets	12	12	11	11	11	11	11	11	11
Outliers	0	1	0	0	0	1	1	0	0
Median CH <sub>4</sub> conc. (nmol kg <sup>-1</sup> )	0.9	2.3	5.7	60.3	4.1	31.3	18.8	5.0	35.2
Mean CH <sub>4</sub> conc. (nmol kg <sup>-1</sup> )	1.8	2.6	5.8	58.6	4.4	31.1	18.8	5.4	35.4
Range	0.6-3.1	1.9-3.8	2.9-8.9	45.2-67.2	2.5-6.5	26.9-35.3	16.5-20.7	3.8-6.8	30.1-42.1
Average coeff. variation (%)	10.9	7.2	8.6	2.1	4.3	3.5	4.2	6.5	3.5

861

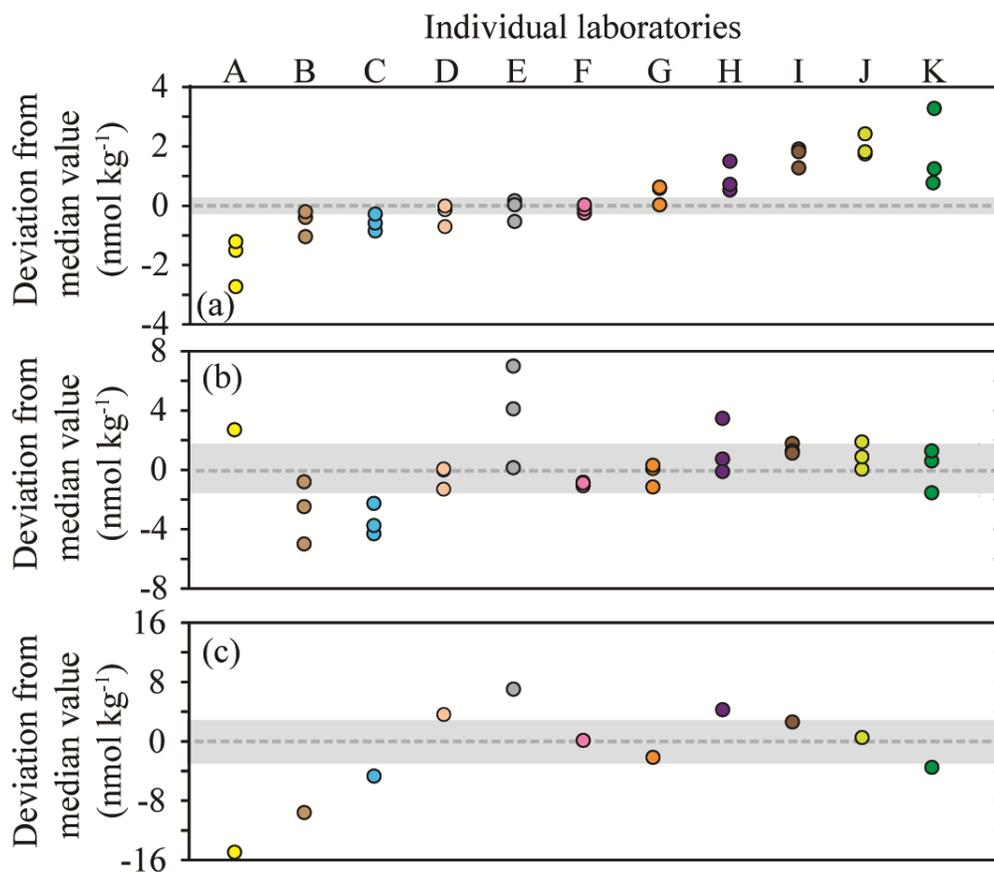
862

863 **Figures**



864  
865  
866 Figure 1. Concentrations of methane measured in nine separate seawater samples collected from  
867 the Pacific Ocean (Fig. 1a, 1b) and the Baltic Sea (Fig. 1c, 1d, 1e). The dashed grey line  
868 represents the value of methane at atmospheric equilibrium (Fig. 1b.) Individual data points are  
869 plotted sequentially by increasing value. The same color symbol is used for each laboratory in  
870 all plots.

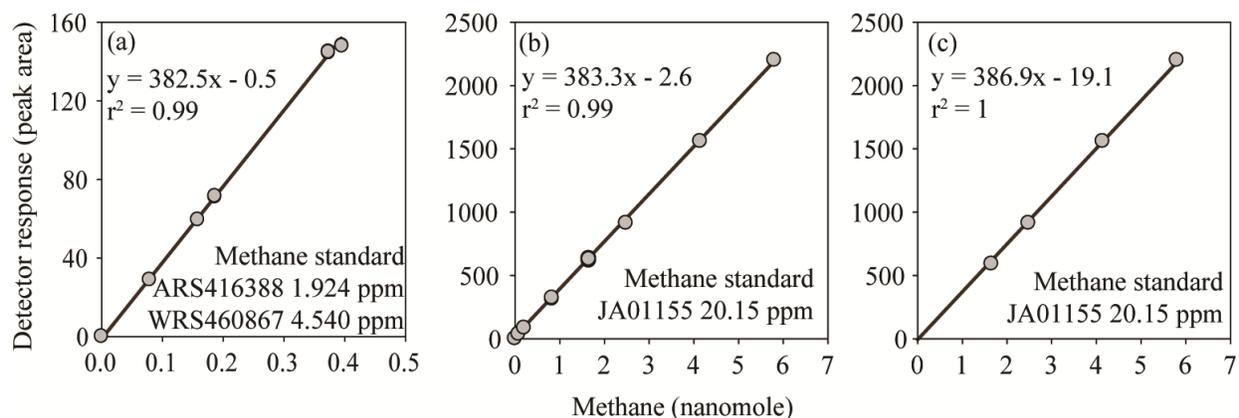
871  
872



873  
 874 Figure 2. Deviation from the median methane concentration (reported as absolute values in nmol  
 875  $\text{kg}^{-1}$ ) for the seven Baltic Sea samples. The batches of seawater samples include BAL1, BAL3,  
 876 and BAL6 (Fig. 2a), BAL4, BAL5, and BAL7 (Fig. 2b), and BAL2 (Fig. 2c). The shaded grey  
 877 area indicates values  $\leq 5\%$  of the median concentration. The color scheme for each laboratory  
 878 dataset is identical to that used in Figure 1 and the letters allocated to each dataset are to facilitate  
 879 cross-referencing in the text. Note that the y-axis scale varies between the Figures.

880

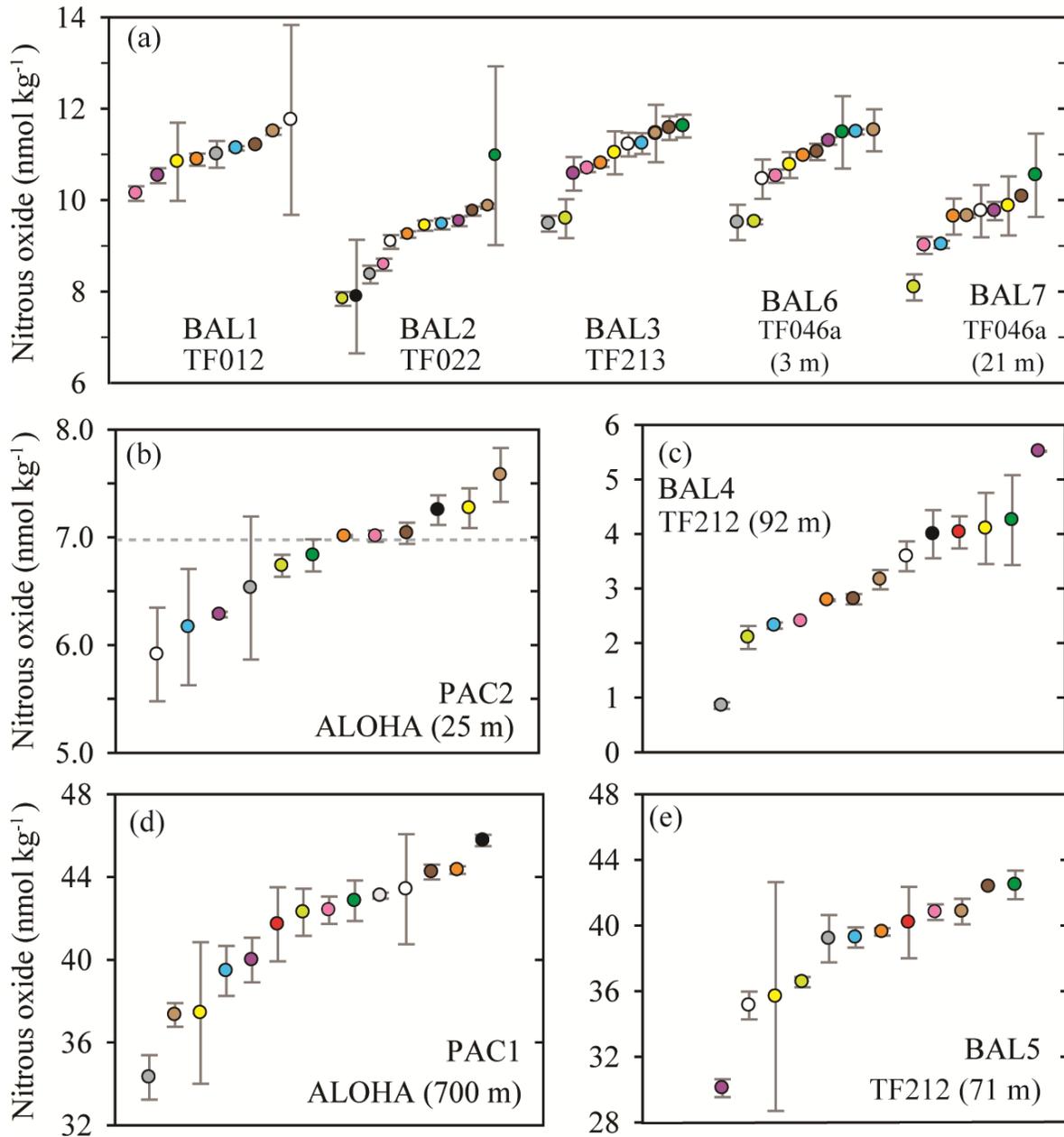
881



882

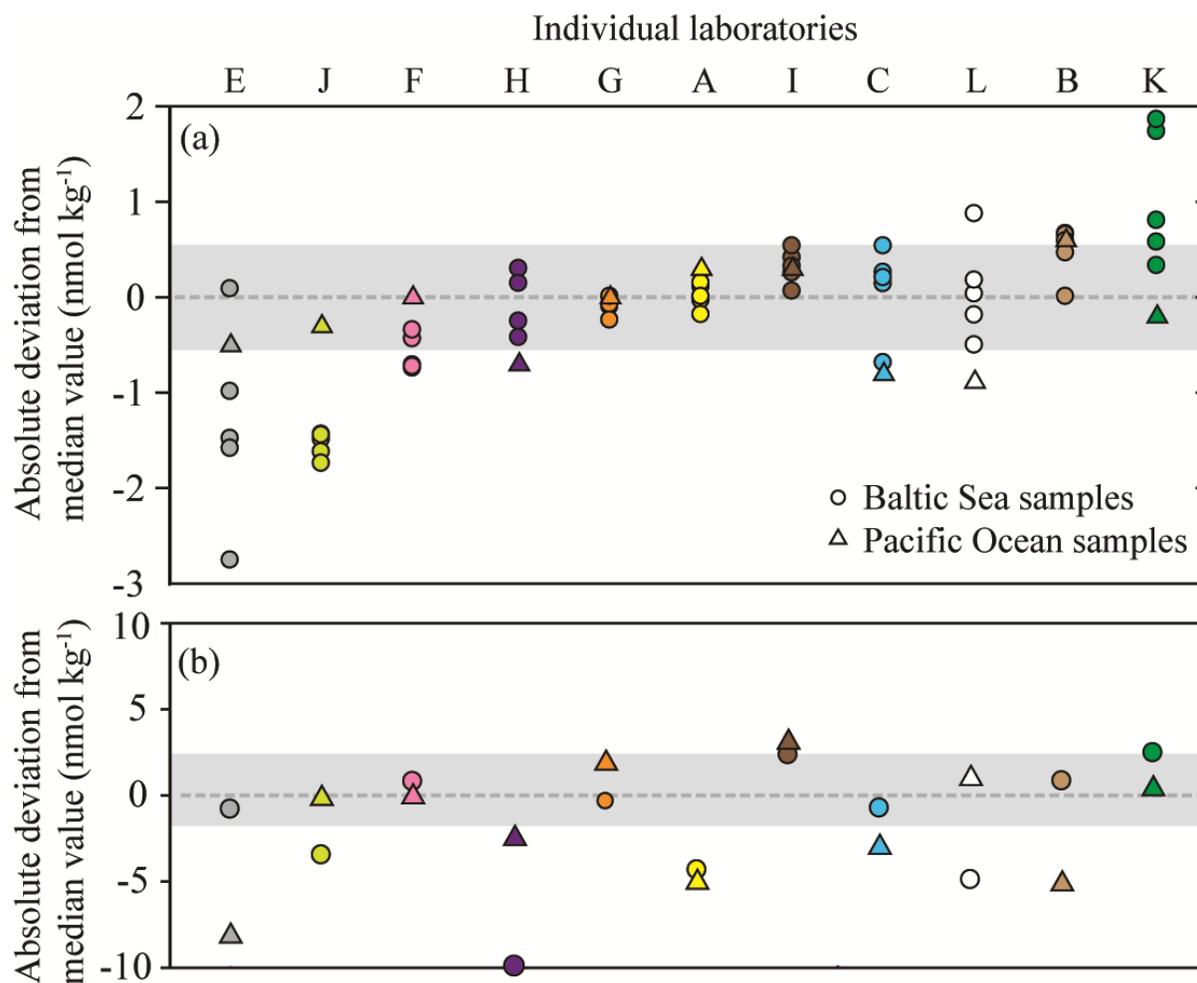
883 Figure 3. FID response to methane, fitted with a linear regression calibration. The inclusion  
884 (Fig. 3a and Fig. 3b) or exclusion (Fig. 3c) of low methane values cause the calibration slope and  
885 intercept to vary. However, the observed variation in the calibration slope does not have a  
886 significant effect on the final calculated concentrations of methane. In contrast, variation in the  
887 intercept does have an effect on the final concentrations of methane.

888



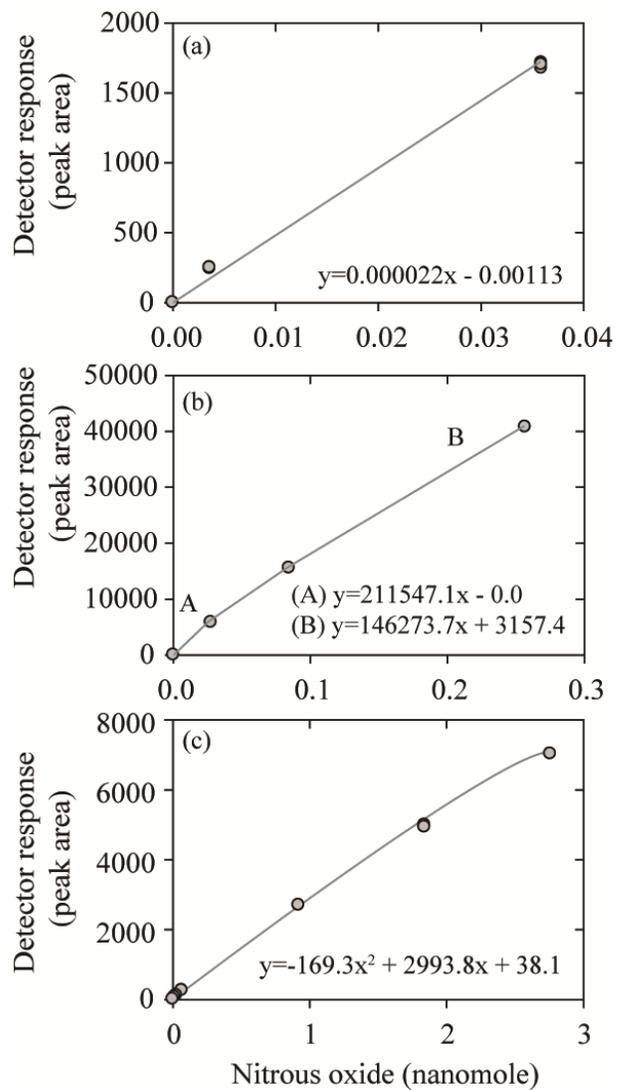
889  
 890  
 891  
 892  
 893  
 894  
 895

Figure 4. Concentrations of nitrous oxide measured in nine separate samples from the Baltic Sea and the Pacific Ocean. The dashed grey line represents the value of nitrous oxide at atmospheric equilibrium (Fig. 4b). Individual data points are plotted sequentially by increasing value. The same color symbol is used for each laboratory in all plots.



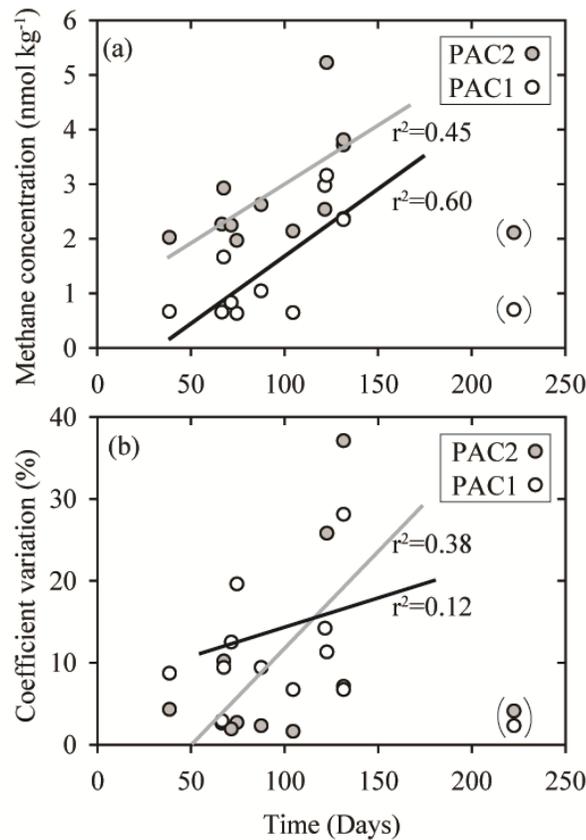
896  
897

898 Figure 5. Deviation from the median value (reported in absolute units) for nitrous oxide datasets.  
 899 The batches of samples include BAL1,2,3,6,7 (Fig. 5a) and PAC2 and BAL5 (Fig. 5b). The  
 900 Baltic Sea samples are represented by circles and the Pacific Ocean samples are represented by  
 901 triangles. The shaded area indicates a deviation  $\leq 5\%$  from the median value, based on a water-  
 902 column concentration of  $11 \text{ nmol kg}^{-1}$  and  $42 \text{ nmol kg}^{-1}$  for Fig. 5a and 5b, respectively. The  
 903 color scheme for each laboratory dataset is identical to that used in Figure 4 and the letters  
 904 allocated to each dataset are to facilitate cross-referencing in the text. Note the y-axis for Fig 5a  
 905 and 5b are plotted on a different scale.



906

907 Figure 6. Three calibration curves for nitrous oxide measurements using an ECD including linear  
 908 (Fig. 6a), multilinear (Fig. 6b), and quadratic (Fig. 6c) fits.



910

911 Figure 7. Comparison of sample storage times with measured concentrations of methane (Fig.  
 912 7a) and coefficient variation (Fig. 7b) for two sets of seawater samples (PAC1 and PAC2)  
 913 collected in February 2017. These two sets of seawater samples had the lowest methane  
 914 concentrations and appear to be influenced by the duration of storage time. The data points  
 915 enclosed in parentheses were not included in the regression analysis. The PAC1 regression line  
 916 is black and the PAC2 regression line is grey. All of the storage times are included in the  
 917 Supplementary Material.