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 CH4 and N2O 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\(^{-1}\), although some laboratories improve upon this by having a large aqueous: gaseous phase ratio during the equilibration process (e.g. Upstill-Goddard et al., 1996). Depending upon the volume of sample analyzed, purge-and-trap analysis can have a detection limit much lower than 1 nmol kg\(^{-1}\) (e.g. Wilson et al., 2017).’

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

We apologize that the report which documented the production of the gas standards was not easily accessible. It is now accessible through the University of Delaware library and the citable URI is now included in the appropriate reference (http://udspace.udel.edu/handle/19716/23288). The report is also attached to this response for your convenience. On Pages 4-5 of this report, the calibrations for the nitrous oxide and methane WRS are described.

In response to the question, there is higher certainty for the ARS because the standards were cross-calibrated with National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory (NOAA/CMDL) and Advanced Global Atmospheric Gases Experiment (AGAGE) standards which have a similar mole fraction. In contrast, the mole fraction of the nitrous oxide WRS far exceeds that of the CMDL and AGAGE standards and the calibration curves are highly non-linear. Therefore, the reported 2-3% accuracy takes into consideration the likelihood of increased systematic errors.

Page 7, lines 158-182: The effects of sample volumes, type of septa used and storage should be assessed better since these differed between the laboratories involved in the intercomparison.
Reviewer #1 points out that there were sampling and storage variables which were not controlled for during the intercomparison exercise. These are responded to separately below.

Sample bottle size
We have taken the Reviewer’s comments into consideration and expanded Section 3.4 ‘Sample storage’ so that it now includes ‘Sample storage and sample bottle size’. Lines 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.’


Storage time
We have improved the wording of this section and Lines 448-459 now read ‘Because prolonged storage of samples can influence dissolved gas concentrations, including methane and nitrous oxide, the intercomparison dataset was analyzed for sample storage effects (Table S5 in the Supplement). It should, however, be noted that assessing the effect of storage time on sample integrity was not a formal goal of the intercomparison exercise and replicate samples were not analyzed at repeated intervals by independent laboratories, as would normally be required for a thorough analysis. Nonetheless our results did provide some insights into potential storage-related problems. Most notably, there were indications that an increase in storage time caused increased concentrations and
increased variability for methane samples with low concentrations, i.e. PAC1 and PAC2 samples which had median methane concentrations of 0.9 and 2.3 nmol kg$^{-1}$, respectively (Fig. 7). In comparison, for samples of nitrous oxide with low concentrations there was no trend of increasing values as observed for samples with low methane concentrations.

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

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

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

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

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

The reviewer raises the point that there are alternative preservatives to mercury(II) chloride. The issue with any preservative is to balance effectiveness at ceasing all relevant microbial activity, while minimizing toxicity from a human health and environmental perspective. In recent years, there have been a series of papers (Magen et al., 2014, Bussmann et al., 2015, Gloël et al., 2015) which have tested some of the alternatives to mercury(II) chloride. These include sodium azide, sodium, hydroxide, sulfuric acid, potassium hydroxide, benzalkonium chloride, and zinc chloride. These studies demonstrate the potential for alternative preservatives and show their effectiveness for a particular environment over a particular timeframe. However, they do not prove the applicability over a broad range of conditions, microbial communities, and storage times. The studies also do not provide a recommendation for the most superior preservative, nor do they always test both methane and nitrous oxide, and other substances such as dissolved inorganic carbon. Therefore, while we
agree that alternatives exist, they have not been extensively proven to be superior to the well-established use of mercuric chloride. After talking to a number of scientists about this issue, we understand that the community of scientists focused on dissolved inorganic carbon measurements are looking very carefully at alternatives to mercury(II) chloride. We have requested that measurements of methane and nitrous oxide be included in planned future tests of alternative preservatives. This will allow the whole community to switch to alternative preservatives at the same time.

We have revised the manuscript to reflect our perspectives and Lines 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.’


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 Hayese 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⁻³, yielding a final methane concentration of 2.1 and 2.8 nmol kg⁻¹ using the equations from Figure 3a and 3c, respectively.’

Page 14, lines 388-401: A sample with higher nitrous oxide concentrations could also be used in future intercomparison efforts. For instance, nitrous oxide concentrations of up to 1000 nmol/L were measured in coastal waters off Peru (Arévalo-Martínez et al., 2013).
The intercomparison of methane and nitrous oxide used typical shipboard sampling procedures as it replicated typical sampling and storage procedures. Future intercomparison exercises will have the ability to manipulate concentrations of methane and nitrous concentrations. The University of Hawaii is awaiting delivery of a large (200 liter) equilibrator unit. The 200 L capacity is smaller than the 760 L equilibrator used to produce reference material for dissolved inorganic carbon by Andrew Dickson, but it will allow us to produce reference material of varying concentrations on demand.

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

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

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

Page 19, lines 532-534: They discuss detection limits for methane but not for nitrous oxide analysis methods. What are the detection limits associated with the two different analysis methods (headspace equilibration versus purge and trap)?
We report on lines 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: 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 r2)? Ideally, to assess storage effects, samples collected the same way and using the same analysis method should be analyzed at different time points by the same laboratory. The r² value is included for each of the regression lines shown in Figure 7a and 7b. We completely agree with the Reviewer’s comment that the same laboratory needs to conduct a time-course set of measurements for a thorough analysis of storage effects. This was not part of the intercomparison work, but is clearly needed for a Best Practice Guide which is being planned.

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

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

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

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

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

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. A 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\(^{-1}\), 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.’
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\(^{-1}\) 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 N2O concentrations over time. Can you show that graph? When graphed, I see that BAL2 shows an increase with time but it also seems within the variability of the measurements.

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

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

In response to this comment and comments from other Reviewers, this section has been rewritten and Lines 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.’
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 7.

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 cation 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.

<table>
<thead>
<tr>
<th>Storage time (days)</th>
<th>PAC 1 Feb 2017 25 m samples</th>
<th>PAC 1 Feb 2017 25 m samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Methane conc (nmol kg⁻¹)</td>
<td>Coeff. Variation (%)</td>
</tr>
<tr>
<td>122</td>
<td>2.52</td>
<td>-</td>
</tr>
<tr>
<td>68</td>
<td>2.91</td>
<td>10.1</td>
</tr>
<tr>
<td>67</td>
<td>2.25</td>
<td>2.5</td>
</tr>
<tr>
<td>132</td>
<td>3.70</td>
<td>7.0</td>
</tr>
<tr>
<td>132</td>
<td>3.79</td>
<td>37.0</td>
</tr>
<tr>
<td>88</td>
<td>2.61</td>
<td>2.2</td>
</tr>
<tr>
<td>123</td>
<td>5.21</td>
<td>25.7</td>
</tr>
<tr>
<td>223</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>39</td>
<td>2.01</td>
<td>4.2</td>
</tr>
<tr>
<td>140</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>75</td>
<td>1.95</td>
<td>2.6</td>
</tr>
<tr>
<td>105</td>
<td>2.12</td>
<td>1.5</td>
</tr>
<tr>
<td>72</td>
<td>2.23</td>
<td>1.8</td>
</tr>
</tbody>
</table>

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 N2O for how each person dealt with water, like what was done for methane? Water is a huge issue for N2O precision, and there is no mention of how water was dealt with.
This is now included in Supplementary Table 7. As a quick response, water vapor is removed by most laboratories using a drying agent frequently in combination with Nafion tubing.

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

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

11. In table S5, “red” is listed as having measured something on the PAC samples 140 days after collection. But when I try to cross reference this in table 2, it looks like “red” didn’t measure for methane. It might help to know if the storage times in table S5 are for methane and/or N2O. Overall, I think this table needed revisiting.
Reviewer#3 is correct, ‘red’ Laboratory M only made nitrous oxide measurements. There was also one laboratory (Laboratory D, beige) that only measured methane. We have improved the Table heading to make this clearer and it now reads ‘The reported storage times are for both methane and nitrous oxide (Laboratory M ‘red’ measured methane only and Laboratory D ‘beige’ measured nitrous oxide only).’

12. Figure S1, what is the gray dashed line? What do colors represent?
Individual data points are plotted sequentially in increasing value with the same color symbol for each laboratory in all plots for the main text and Supplementary Material. The dashed grey line represents the value of methane at atmospheric equilibrium as stated in the Figure legend.

13. Figure S2, are a and b shallow water and c and d deep water? Make that clear in the first description of the figure. It says “same location” but what you mean is at the same lat/long but two different depths. Also, caption says “In contrast, the concentration of nitrous oxide in the deep-water samples (Figure S2c and d) was more consistent and the data values for the laboratories that measured samples from 2013 and 2017 are shown together in Figure S2d.” is that also supposed to be shown by a gray dashed line? Can you make the scales the same for both sides?
We have now plotted Figure S2c on the same scale as Figure S2d. Each subplot also includes a description of depth as well as the actual Figure legend.
The Figure S2 legend has been improved and now reads ‘Supplementary Figure S2: Nitrous oxide concentrations in seawater samples collected at the same location but varying depths in the North Pacific Ocean on February 2017 (Fig. S2a and c) and November 2013 (Fig. S2b and d). The dashed grey line represents the value of nitrous oxide at atmospheric equilibrium for the 25 m seawater samples (Figure S2a and b). The February 2017 plots are discussed in the main manuscript and are replicated here to facilitate comparison with the November 2013 data, particularly for comparison with the 700 m samples (Figure S2d).’

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

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

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

Samuel T. Wilson, Hermann W. Bange, Damian L. Arévalo-Martínez, Jonathan Barnes,
Alberto V. Borges, Ian Brown, John L. Bullister, Macarena Burgos, David W. Capelle,
Michael Casso, Mercedes de la Paz, Laura Farías, Lindsay Fenwick, Sara Ferrón,
Gerardo García, Michael Glockzin, David M. Karl, Annette Kock, Sarah Laperriere,
Cliff S. Law, Cara C. Manning, Andrew Marriner, Jukka-Pekka Myllykangas,
John W. Pohlman, Andrew P. Rees, Alyson E. Santoro, Philippe D. Tortell,
Robert C. Upstill-Goddard, David P. Wisegarver, Guiling L. Zhang, Gregor Rehder

1 University of Hawai‘i at Manoa, Daniel K. Inouye Center for Microbial Oceanography:
2 Research and Education (C-MORE), Honolulu, Hawai‘i, USA
3 GEOMAR Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20 24105 Kiel,
   Germany
4 Newcastle University, School of Natural and Environmental Sciences, Newcastle upon Tyne,
   UK
5 Université de Liège, Unité d’Océanographie Chimique, Liège, Belgium
6 Plymouth Marine Laboratory, Plymouth, UK
7 National Oceanic and Atmospheric Administration, Pacific Marine Environmental Laboratory,
   Seattle, Washington, USA
8 Universidad de Cádiz, Instituto de Investigaciones Marinas, Departamento Química-Física
   Cádiz, Spain
9 University of British Columbia, Vancouver, Department of Earth, Ocean and Atmospheric
   Sciences, British Columbia, Canada
10 U.S. Geological Survey, Woods Hole Coastal and Marine Science Center, Woods Hole, USA
11 Instituto de Investigaciones Marinas, Vigo, Spain
12 University of Concepción, Department of Oceanography and Center for climate research and
   resilience (CR2), Concepción, Chile
13 Leibniz Institute for Baltic Sea Research Warnemünde, Rostock, Germany
14 University of California Santa Barbara, Department of Ecology, Evolution, and Marine
   Biology, Santa Barbara, USA
15 National Institute of Water and Atmospheric Research (NIWA), Wellington, New Zealand
15 Department of Chemistry, University of Otago, Dunedin, New Zealand
16 University of Helsinki, Department of Environmental Sciences, Helsinki, Finland
17 Ocean University of China, Department of Marine Chemistry, Qingdao, China

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

*corresponding author: stwilson@Hawai’i.edu
Abstract. Large scale climatic forcing is impacting oceanic biogeochemical cycles and is expected to influence the water-column distribution of trace gases including methane and nitrous oxide. Our ability as a scientific community to evaluate changes in the water-column inventories of methane and nitrous oxide depends largely on our capacity to obtain robust and accurate concentration measurements which can be validated across different laboratory groups. This study represents the first formal, international, intercomparison of oceanic methane and nitrous oxide measurements whereby participating laboratories received batches of seawater samples from the subtropical Pacific Ocean and the Baltic Sea. Additionally, compressed gas standards from the same calibration scale were distributed to the majority of participating laboratories to improve the analytical accuracy of the gas measurements. The computations used by each laboratory to derive the dissolved gas concentrations were also evaluated for inconsistencies (e.g. pressure and temperature corrections, solubility constants). The results from the intercomparison and intercalibration provided invaluable insights into methane and nitrous oxide measurements. It was observed that analyses of seawater samples with the lowest concentrations of methane and nitrous oxide had the lowest precisions. In comparison, while the analytical precision for samples with the highest concentrations of trace gases was better, the variability between the different laboratories was higher; 36% for methane and 27% for nitrous oxide. In addition, the comparison of different batches of seawater samples with methane and nitrous oxide concentrations that ranged over an order of magnitude revealed the ramifications of different calibration procedures for each trace gas. Overall, this study builds upon the intercomparison results to develop a framework for improving oceanic methane and nitrous oxide measurements, with the aim of precluding future analytical discrepancies between laboratories.
1. Introduction
The increasing mole fractions of greenhouse gases in the Earth’s atmosphere are causing long-term climate change with unknown future consequences. Two greenhouse gases, methane and nitrous oxide, together contribute approximately 23% of total radiative forcing attributed to well-mixed greenhouse gases (Myhre et al., 2013). It is imperative that the monitoring of methane and nitrous oxide in the Earth’s atmosphere is accompanied by measurements at the Earth’s surface to better inform the sources and sinks of these climatically important trace gases. This includes measurements of dissolved methane and nitrous oxide in the marine environment, which is an overall source of both gases to the overlying atmosphere (Nevison et al., 1995; Anderson et al., 2010; Naqvi et al., 2010; Freing et al., 2012; Ciais et al., 2014).

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

Methods for quantifying dissolved methane and nitrous oxide have evolved and somewhat diverged since the first measurements were made in the 1960s (Craig and Gordon 1963; Atkinson and Richards 1967). Some laboratories employ purge-and-trap methods for extracting and concentrating the gases prior to their analysis (e.g. Zhang et al., 2004; Bullister and Wisegarver, 2008; Capelle et al., 2015; Wilson et al., 2017). Others equilibrate a seawater sample with an overlying headspace gas and inject a fixed volume of the gaseous phase into a gas analyzer (e.g. Upstill-Goddard et al., 1996; Walter et al., 2005; Farias et al., 2009). 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). Additional developments for continuous underway surface seawater measurements use equilibrator systems of various designs coupled to a variety of detectors (e.g. Weiss et al., 1992; Butler et al., 1989; Gülzow et al., 2011; Arévalo-Martínez et al., 2013). Determining the level of analytical comparability between different laboratories for discrete samples of methane and nitrous oxide is an important step towards improved comprehensive global assessments. Such intercomparison exercises are critical to determining the spatial and temporal variability of methane and nitrous oxide across the world oceans with confidence, since no single laboratory can single-handedly provide all the required measurements at sufficient resolution. Previous comparative exercises have been conducted for other trace gases e.g. carbon dioxide, dimethylsulphide, and sulfur hexafluoride (Dickson et al., 2007; Bullister and Tanhua, 2010; Swan et al., 2014) and for trace elements (Cutter et al., 2013). These exercises confirm the value of the intercomparison concept.

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

Q1. What is the agreement between the SCOR gas standards and the ‘in-house’ gas standards used by each laboratory?
Q2. How do measured values of dissolved methane and nitrous oxide compare across laboratories?
Q3. Despite the use of different analytical systems, are there general recommendations to reduce uncertainty in the accuracy and precision of methane and nitrous oxide measurements?
Q4. What are the implications of inter-laboratory differences for determining the spatial and temporal variability of methane and nitrous oxide in the oceans?
2. Methods

2.1 Calibration of nitrous oxide and methane using compressed gas standards

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

The absence of a widely available high mole fraction, high accuracy nitrous oxide gas standard was noted as a primary concern at the outset of the intercomparison exercise. Therefore, a set of high-pressure primary gas standards was prepared for the SCOR Working Group by John Bullister and David Wisegarver at NOAA Pacific Marine and Environmental Laboratory (PMEL). One batch, referred to as Air Ratio Standard (ARS), had methane and nitrous oxide mole fractions similar to modern air and the other batch, referred to as Water Ratio Standard (WRS) had higher methane and nitrous oxide mole fractions for calibration of high concentration water samples. These SCOR primary standards were checked for stability over a 12 month period and assigned mole fractions on the same calibration scale, known as ‘SCOR-2016.’ A comparison was conducted with NOAA standards prepared on the SIO98 calibration scale for nitrous oxide and the NOAA04 calibration scale for methane. Based on the comparison with NOAA standards, the uncertainty of the methane and nitrous oxide mole fractions in the
ARS and the uncertainty of the methane mole fraction in the WRS were all estimated at better than 1%. By contrast, the uncertainty of the nitrous oxide mole fraction in the WRS was estimated at 2-3%. The gas standards were distributed to twelve of the laboratories involved in this study (Table 1). The technical details on the production of the gas standards and their assigned absolute mole fractions is included in Bullister et al. (2016).

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

Dissolved methane and nitrous oxide samples for the intercomparison exercise were collected from the subtropical Pacific Ocean and the Baltic Sea. Pacific samples were obtained on 28 November 2013 and 24 February 2017 from the Hawai’i Ocean Time-series (HOT) long-term monitoring site, Station ALOHA, located at 22.75 N, 158.00 W. 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. Seawater was collected using Niskin-like bottles designed by John Bullister (NOAA PMEL), which help minimize contamination of trace gases, in particular chlorofluorocarbons and sulfur hexafluoride (Bullister and Wisegarver, 2008). The bottles were attached to a rosette with a conductivity-temperature-depth (CTD) package. Seawater was collected from two depths: 700 m and 25 m, where the near-maximum and minimum water-column concentrations for methane and nitrous oxide at this location can be found. The 25 m samples were always well within the surface mixed layer, which ranged from 100 to 130 m depth during sampling.

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

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

### 2.3. Sample analysis
Each laboratory measured dissolved methane and nitrous oxide slightly differently. A full description of each laboratory’s method can be found in Table S6 and Table S7 in the Supplement for methane and nitrous oxide, respectively.

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

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

\[ C_{gas} \text{ [nmol L}^{-1}] = \left( \beta \times PV_{wp} + \frac{xP}{R_T} V_{hs} \right) / V_{wp} \]

where \( \beta \) is the Bunsen solubility of nitrous oxide (Weiss and Price, 1980) or methane (Wiesenburg and Guinasso, 1979) in nmol L\(^{-1}\) atm\(^{-1}\), \( x \) is the dry gas mole fraction (ppb) measured in the headspace, \( P \) is the atmospheric pressure (atm), \( V_{wp} \) is the volume of water sample (mL), \( V_{hs} \) is the volume (mL) of the created headspace, \( R \) is the gas constant (0.08205746
L atm K^{-1}mol^{-1}), and \(T\) is equilibration temperature in Kelvin (K). An example calculation is provided in Table S8 in the Supplement.

In contrast to the headspace equilibrium method, five laboratories used a purge-and-trap system for methane and/or nitrous oxide analysis (Table S6 and Table S7 in the Supplement). These systems were directly coupled to a Flame Ionization Detector (FID) or ECD, with the exception of University of British Columbia, where a quadrupole mass spectrometer with an electron impact ion source and Faraday cup detector were used (Capelle et al., 2015). The purge-and-trap systems were broadly similar, each transferring the seawater sample to a sparging chamber. Sparging times typically ranged from 5-10 min and the sparge gas was either high purity helium or high purity nitrogen. In addition to commercially available gas 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. The elutant gas was dried using Nafion or Drierite, and subsequently cryotrapped on a sample loop packed with Porapak Q to aid retention of methane and nitrous oxide. Cryotrapping was achieved for methane using liquid nitrogen (-195°C) and either liquid nitrogen or cooled ethanol (-70°C) for nitrous oxide. Subsequently, the valve was switched to inject mode and the sample loop was rapidly heated to transfer its contents onto the analytical column. Calibration was achieved by injecting standards via sample loops using multi-port injection valves. Injection of standards upstream of the sparge chamber allowed for calibration of the purge-and-trap gas handling system, in addition to the GA. Calculation of the gas concentrations using the purge-and-trap method was achieved by application of the ideal gas law to the standard gas measurements:

\[
P V = nRT
\]

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

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

3. Results

3.1 Comparison of methane and nitrous oxide gas standards

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

Seven laboratories compared their own in-house standards of nitrous oxide with the prepared SCOR standards. Six laboratories reported values of nitrous oxide for the ARS which were within 3% of the absolute concentration, with the remaining laboratory reporting an offset of 10% (Table S7 in the Supplement). The majority of these laboratories (five out of six groups) compared the SCOR ARS with NOAA GMD standards, which have a balance gas of air instead of nitrogen. Some laboratories with analytical systems that incorporated fixed sample loops (e.g. 1 or 2 ml loops housed in a 6-port or 10-port injection valve) had difficulty analyzing the WRS,
as the peak areas created by the high mole fraction of the standard exceeded the signal typically measured from in-house standards or acquired by sample analysis, by an order of magnitude. The high mole fraction of the WRS was not an issue when multiple sample loops of varying sizes were incorporated into the analytical system, which was the case for purge-and-trap based designs. 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.

3.2 Methane concentrations in the intercomparison samples

Overall, median methane concentrations in seawater samples collected from the Pacific Ocean and the Baltic Sea ranged from 0.9 to 60.3 nmol kg\(^{-1}\) (Table 2). Out of 101 reported values, 3 outliers were identified using the IQR criterion and were not included in further analysis. The methane data values for each batch of samples analyzed by each laboratory, including the mean and standard deviation, the number of samples analyzed, and the % offset from the overall median value are reported in Table S1 and Table S2 in the Supplement. Analysis conducted by the University of Hawai’i 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 collection used in the Baltic Sea sampling.

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

Three Baltic Sea sampling sites (BAL1, BAL3, and BAL6) had median methane concentrations that ranged from 4.1 to 5.7 nmol kg\(^{-1}\) (Fig. 1c). The BAL1 samples also showed a
skewed distribution of reported values towards higher concentrations, as seen in PAC1 and PAC2 samples. However, this was not evident in BAL3 or BAL6, which had the closest agreement between the reported methane concentrations. For these three sets of Baltic Sea samples, the mean coefficient of variation for all laboratories ranged from 4% (BAL3) to 9% (BAL1). The next three Baltic Sea samples (BAL4, BAL5, and BAL7) had methane concentrations that ranged from 18.8 to 35.4 nmol kg\(^{-1}\) (Fig. 1d). These three sets of samples had a normal distribution of data and the closest agreement between the reported concentrations for all of the Pacific Ocean and Baltic Sea samples. Furthermore, for these three sets of samples, the mean coefficient of variation for all laboratories was 4% (Table 2). The final Baltic Sea sample (BAL2) had the highest concentrations of methane, with a median reported value of 60.3 nmol kg\(^{-1}\), and a large range of values (45.2 to 67.2 nmol kg\(^{-1}\); Fig. 1e). The BAL2 samples had the lowest overall mean coefficient of variation for all laboratories; 2% (Table 2).

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

The reasons behind the trends for each dataset became more apparent when considering the effect of the inclusion or exclusion of low standards in the calibration curve on the resulting derived concentrations (Fig. 3). The FID has a linear response to methane at nanomolar values and therefore a high level of accuracy across a relatively wide range of *in situ* methane concentrations can be obtained with the correct slope and intercept. To demonstrate this,
calibration curves for methane were provided by the University of Hawai’i. These revealed minimal variation in the slope value when calibration points were increased from low mole fractions (Fig. 3a) to higher mole fractions (Fig. 3b). However, the intercept value was sensitive to the range of calibration values used, and this effect was further exacerbated when only the higher calibration points were included (i.e., Fig. 3c). The relevance to final methane concentrations is demonstrated by considering the values reported by the University of Hawai‘i for PAC2 samples (Fig. 1b). An almost 30% increase in final methane concentration occurs from the use of the calibration equation in Figure 3c, compared to Figure 3a. This derives from a measured peak area for methane of 62 for a sample with a volume of 0.076 L and a seawater density of 1024 kg m\(^{-3}\), yielding a final methane concentration of 2.1 and 2.8 nmol kg\(^{-1}\) using the equations from Figure 3a and 3c, respectively. With this understanding on the effect of FID calibration, we consider it likely that the increased deviation from median values at high methane concentrations (Datasets B, C, and H) results from differences in calibration slope between each laboratory. In contrast, the datasets with a higher offset at low methane concentrations (Datasets J and K) could be due to erroneous low standard values causing a skewed intercept. In addition, there may be other factors including sample contamination, discussed in Section 3.4.

### 3.3 Nitrous oxide concentrations in the intercomparison samples

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

For six sets of seawater samples, BAL1, BAL2, BAL3, BAL6, BAL7, and PAC2, the concentrations of nitrous oxide were close to atmospheric equilibrium. The reported values ranged from 7.7 to 12.7 nmol kg\(^{-1}\) in the Baltic Sea (Fig. 4a) and from 5.9 to 7.6 nmol kg\(^{-1}\) in the Pacific Ocean (Fig. 4b). For the Pacific Ocean near-surface (mixed layer) sampling site (PAC2), the theoretical value of nitrous oxide concentration in equilibrium with the overlying atmosphere is also shown (Fig. 4b). For these six samples with concentrations close to atmospheric
equilibrium, the mean coefficient of variation for all laboratories ranged from 3% (BAL3 and PAC2) to 5% (BAL1) (Table 2).

For the three other sets of samples (BAL4, BAL5, and PAC1), the nitrous oxide concentrations deviated significantly from atmospheric equilibrium (Fig. 4c, 4d, and 4e). At one sampling site, BAL4 (Fig. 4c), nitrous oxide was under-saturated with respect to atmospheric equilibrium and reported concentrations ranged from 2.1–5.5 nmol kg\(^{-1}\). As observed in the low concentration Pacific Ocean methane samples, there was a skewed distribution of the data towards the higher nitrous oxide concentrations. The BAL4 samples also had the highest variability (i.e. lowest precision), with a mean coefficient of variation of 8% (Table 2). The two remaining samples (PAC1 and BAL5) had much higher concentrations of nitrous oxide, as expected for low-oxygen regions of the water-column. In contrast to the samples with near atmospheric equilibrium concentrations of nitrous oxide, there was a low overall agreement between the independent laboratories for PAC1 and BAL5 nitrous oxide concentrations (Fig. 4d, 4e). At PAC1 and BAL5, nitrous oxide concentrations ranged from 34.3–45.8 nmol kg\(^{-1}\) (Fig. 4d) and 30.1–55.9 nmol kg\(^{-1}\), respectively (Fig. 4e). The mean coefficient of variation for all laboratories was 4% for BAL5 samples compared to 3% for PAC1 samples.

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

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

3.4 Sample storage and sample bottle size

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

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.

4. Discussion

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

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

It is typical for laboratories to source some, or all, of their compressed gas standards from
commercial suppliers. National agencies, such as NOAA GMD or National Institute of
Metrology China, also provide standards to the scientific community. The national agencies
typically offer a lower range in concentrations than commercial suppliers, but their standards
tend to have a higher level of accuracy. Of the twelve laboratories participating in the
intercomparison, eight reported using national agency standards, with seven of them using gases
sourced from NOAA GMD. Since the methane and nitrous oxide mole fractions of these
national agency standards are equivalent to modern-day atmospheric mixing ratios, they are
similar to the SCOR ARS distributed to the majority of laboratories in this study. Laboratories
in receipt of the SCOR standards were asked to predict their mole fractions based on those of
their own in-house standards. For the majority that conducted this exercise, there was good
agreement (<3% difference) between the NOAA GMD and the SCOR ARS for both methane
and nitrous oxide. For three laboratories, a larger offset was observed between the NOAA GMD
and the SCOR ARS. There was also a good prediction for the higher methane content SCOR
WRS, facilitated by the linear response of the FID (Fig. 3). In contrast, the nitrous oxide mole
fraction in the SCOR WRS exceeded the typical working range for several laboratories and it was difficult for them to cross-compare with their in-house standards. This reflects an analytical set-up that involves on-column injection via a 6-port or 10-port valve with one or two sample loops, respectively. The sample loops have a fixed volume and their inaccessibility makes it difficult to replace them by a smaller loop size. Therefore either dilution of the standard is required, or smaller loops need to be incorporated into the calibration protocol. The two laboratories that compared their in-house standards with the SCOR WRS reported an offset of 3% and >20%. This indicates that variability between standards can be an issue for obtaining accurate dissolved concentrations and provides support for the production of a widely available high concentration nitrous oxide standard. We strongly recommend that all commercially obtained standards are cross-checked against primary standards, such as the SCOR ARS and WRS. This should be conducted at least at the beginning and end of their use to detect any drift that may have occurred during their lifetime. With due diligence and care, the SCOR standards provide the capability for cross-checking personal standards for years to decades (Bullister et al., 2016).

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

**Methane:** The methane intercomparison highlighted the variability that exists between measurements conducted by independent laboratories. At low methane concentrations, a skewed distribution of methane data was observed, which was particularly evident in PAC1 (Fig. 1a). Potential causes include calibration procedures (Section 3.2) and/or sample contamination which is more prevalent at low concentrations (Section 3.4). For some laboratories, the low methane concentrations are close to their detection limit, which is determined by the relatively low sensitivity of the FID and the small number of moles of methane in an introduced headspace equilibrated with seawater. An approximate working detection limit for methane analysis via headspace equilibration is 1 nmol kg\(^{-1}\), although some laboratories improve upon this by having a large aqueous: gaseous phase ratio during the equilibration process (e.g. Upstill-Goddard et al., 1996). Depending upon the volume of sample analyzed, purge-and-trap analysis can have a detection limit much lower than 1 nmol kg\(^{-1}\) (e.g. Wilson et al., 2017). Methane measurements in aquatic habitats with methane concentrations near the limit of analytical detection include mesopelagic and high latitude environments distal from coastal or benthic inputs (e.g. Rehder et
Of additional concern is that the skewed distribution of methane concentrations also occurs in samples collected both from the surface ocean (PAC2; Fig. 1b) and coastal environments (BAL1; Fig. 1c). Methane concentrations between 2–6 nmol kg\(^{-1}\) are within the detection limit of all participating laboratories. To address this we recommend that laboratories restrict sample storage to the minimum time required to analyze the samples and incorporate internal controls into their sample analysis (Section 4.4).

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

Nitrous oxide: Some of the trends discussed for methane were also evident in the nitrous oxide data. For the samples with the lowest nitrous oxide concentrations a skewed data distribution was observed, as found for methane (Fig. 4c). Such low nitrous oxide concentrations are typical of low-oxygen water-column environments (<10 µmol kg\(^{-1}\)). Therefore, the analytical bias towards measuring values higher than the absolute in situ concentrations is particularly pertinent to oceanographers measuring nitrous oxide in oxygen minimum zones and other low-oxygen environments (Naqvi et al., 2010; Farías et al., 2015; Ji et al., 2015). 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. Therefore, the bias towards reporting elevated values for low concentrations of nitrous oxide is related less to analytical sensitivity and is more a consequence of calibration issues. During the intercomparison exercise ECD calibration was identified as a nontrivial issue for all participating laboratories and it deserves continuing attention. In particular, the nonlinearity of the ECD means that low and
high nitrous oxide concentrations are more vulnerable to error since the values fall outside of the most frequented part of the calibration curve. This is particularly true if a linear fit is used to calibrate the ECD (Fig. 6a). To circumvent this problem, one laboratory used a step-wise linear function while other laboratories used a quadratic function. The usefulness of multiple calibration curves for low and high nitrous oxide concentrations was highlighted during the intercomparison exercise, although this necessitates some consideration of the threshold for switching between different calibration curves.

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

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

There are several analytical recommendations resulting from this study. The use of highly accurate standards and the appropriate calibration fit is an essential requirement for both headspace equilibration and the purge-and-trap technique. It was shown that both analytical approaches can yield comparable values for methane and nitrous oxide, with the main differences observed at low methane concentrations. At sub-nanomolar methane concentrations, four out of the six laboratories that reported methane concentrations \(<1 \text{ nmol kg}^{-1}\) used a purge-and-trap analysis.
This study also revealed that sample storage time can be an important factor. Specifically, 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 (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 these reasons, it is recommended to conduct an evaluation of sample storage time for the environment that is being sampled.

One useful item that was not included as part of the intercomparison exercise but can help decrease uncertainty in the accuracy and precision of methane and nitrous oxide measurements are internal control measurements. Internal controls represent a self-assessment quality control check to validate the analytical method and quantify the magnitude of uncertainty. Appropriate internal controls for methane and nitrous oxide consist of air-equilibrated seawater samples. Their purpose is to provide checks for methane concentrations ranging from 2–3 nmol kg$^{-1}$ and for nitrous oxide concentrations from 5–9 nmol kg$^{-1}$. 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. Air-equilibrated samples provide reassurance that the analytical system is providing values within the correct range. Air-equilibrated samples also indicate the certainty associated with calculating the saturation state of the ocean with respect to atmospheric equilibrium. This is particularly relevant when the seawater being sampled is within a few percent of saturation. Finally, these air-equilibrated samples provide an estimate of analytical accuracy, which is infrequently reported for methane or nitrous oxide. At present, only a few studies report the analysis of air-equilibrated seawater alongside water-column samples (Bullister and Wisegarver, 2008; Capelle et al., 2015; Bourbonnais et al., 2017; Wilson et al., 2017). It is likely that wider implementation would facilitate internal assessment of the analytical system. Since the main equipment required is a water-bath and an overhead stirrer, the production is not cost-prohibitive. A recommendation of this intercomparison exercise is that laboratories routinely use air-equilibrated seawater samples to provide an estimate of analytical accuracy.
In addition to the self-assessments provided by the analysis of air-equilibrated seawater, this study revealed the need for reference seawater to help assess the accuracy of high concentration methane and nitrous oxide measurements. Reference seawater in this instance refers to batches of dissolved methane and nitrous oxide samples prepared in the laboratory using an equilibrator set-up, as used for dissolved inorganic carbon (Dickson et al., 2007). In the absence of plans for additional intercomparison exercises, the provision of reference seawater will allow laboratories to continue evaluating their own measurements. **Finally, the lessons learned during the intercomparison exercises will be the basis for a forthcoming Good Practice Guide for dissolved methane and nitrous oxide.**

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

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

The inter-laboratory differences highlighted by this study should be viewed in the context of numerous individual efforts to assess temporal and/or spatial trends in methane and nitrous oxide by way of time-series observations (Bange et al., 2010; Farías et al., 2015; Wilson et al., 2017; Fenwick and Tortell, 2018), repeat hydrographic survey lines (de la Paz et al., 2017), and single expeditions. While the value of these in integrating the behaviour of methane and nitrous oxide into the hydrography and biogeochemistry of local-regional ecosystems is beyond question, their value would be enhanced by the rigorous cross-validation of analytical protocols. Without this, perceived small temporal and/or spatial changes in water-column concentrations in any given region are difficult to verify unless the data all originate from a single laboratory. In addition,
the value of a global methane and nitrous oxide database (e.g. Bange et al., 2009) would to some extent be compromised by the uncertainty. Taking due account of the analytical variability between laboratories will clearly be vital to any future assessment of the changing methane and nitrous oxide budgets of the oceans.

5. Conclusions

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

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

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

The methane and nitrous oxide intercomparison exercise was conducted as a Scientific Committee on Ocean Research (SCOR) Working Group which receives funding from the U.S. National Science Foundation (OCE-1546580). Pacific Ocean seawater samples were collected on HOT cruises which are supported by NSF (including the most recent OCE-1260164 to DMK). Baltic Sea seawater samples were collected during Cruise #142 of the RV Elisabeth Mann Borgese, with the ship-time provided by the Leibniz Institute for Baltic Sea Research Warnemünde. We thank Liguo Guo for help with sampling during the Baltic Sea cruise. The methane and nitrous oxide gas standards were produced via a Memorandum of Understanding between the University of Hawai‘i and NOAA-PMEL. Funding for the gas standards was provided by for the Center for Microbial Oceanography: Research and Education (C-MORE; EF0424599 to DMK), SCOR, the EU FP7 funded Integrated non-CO₂ Greenhouse gas Observation System (InGOS) (Grant Agreement #284274), and NOAA’s Climate Program Office, Climate Observations Division. Additional support was provided by the Gordon and Betty Moore Foundation #3794 (DMK), the Simons Collaboration on Ocean Processes and Ecology (SCOPE; #329108 to DMK), and the Global Research Laboratory Program (#2013K1A1A2A02078278 to DMK) through the National Research Foundation of Korea (NRF).

AVB is a senior research associate at the FRS-FNRS. AES would like to acknowledge NSF OCE-1437310. MP would like to acknowledge the support of the Spanish Ministry of Economy and Competitiveness (CTM2015-74510-JIN). LF received financial support by FONDAP 1511009 and FONDECYT N°1161138. Any use of trade names is for descriptive purposes and does not imply endorsement by the U.S. government.
References


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


Swan, H. B., Armishaw, P., Iavetz, R., Alamgir, M., Davies, S. R., Bell, T. G., and Jones, G. B.:
An interlaboratory comparison for the quantification of aqueous dimethylsulfide. Limnol.
Oceanogr.: Methods, 12, 784–794, 2014.

Upstill-Goddard, R. C., Rees, A. P., and Owens, N. J. P.: Simultaneous high-precision
measurements of methane and nitrous oxide in water and seawater by single phase

Upstill-Goddard, R. C., and Barnes, J.: Methane emissions from UK estuaries: Re-evaluating the

during EIFEX, the European Iron Fertilization Experiment in the subpolar South Atlantic


dioxide and nitrous oxide observation by shipboard automated gas chromatography: Results
from expeditions between 1977 and 1990, Scripps Institution of oceanography Reference 92-
11. ORNL/CDIAC-59, NDP-044. Carbon Dioxide Information Analysis Center, Oak Ridge


2017.

Zhang, G. L., Zhang, J., Kang, Y. B., and Liu, S. M.: Distributions and fluxes of methane in the
East China Sea and the Yellow Sea in spring, J. Geophys. Res., 109, C07011,
Table 1. List of laboratories that participated in the intercomparison. All laboratories measured both methane and nitrous oxide except U.S. Geological Survey (methane only), U.C. Santa Barbara (nitrous oxide only), and NOAA PMEL (nitrous oxide from the Pacific Ocean). Also indicated are the twelve laboratories that received the SCOR gas standards of methane and nitrous oxide.

<table>
<thead>
<tr>
<th>Institution</th>
<th>Lead Scientist</th>
<th>SCOR Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Hawai’i, USA</td>
<td>Samuel Wilson</td>
<td>Yes</td>
</tr>
<tr>
<td>GEOMAR, Germany</td>
<td>Hermann Bange</td>
<td>Yes</td>
</tr>
<tr>
<td>Newcastle University, UK</td>
<td>Robert Upstill-Goddard</td>
<td>Yes</td>
</tr>
<tr>
<td>Université de Liège, Belgium</td>
<td>Alberto Vieira Borges</td>
<td>No</td>
</tr>
<tr>
<td>Plymouth Marine Laboratory, UK</td>
<td>Andrew Rees</td>
<td>Yes</td>
</tr>
<tr>
<td>NOAA PMEL, USA</td>
<td>John Bullister</td>
<td>Yes</td>
</tr>
<tr>
<td>IIM-CSIC, Spain</td>
<td>Mercedes de la Paz</td>
<td>Yes</td>
</tr>
<tr>
<td>CACYTMAR, Spain</td>
<td>Macarena Burgos</td>
<td>No</td>
</tr>
<tr>
<td>University of Concepción, Chile</td>
<td>Laura Farías</td>
<td>Yes</td>
</tr>
<tr>
<td>IOW, Germany</td>
<td>Gregor Rehder</td>
<td>Yes</td>
</tr>
<tr>
<td>University of California Santa Barbara, USA</td>
<td>Alyson Santoro</td>
<td>Yes</td>
</tr>
<tr>
<td>National Institute of Water and Atmospheric Research, NZ</td>
<td>Cliff Law</td>
<td>Yes</td>
</tr>
<tr>
<td>University British Columbia, Canada</td>
<td>Philippe Tortell</td>
<td>Yes</td>
</tr>
<tr>
<td>U.S. Geological Survey, USA</td>
<td>John Pohlman</td>
<td>No</td>
</tr>
<tr>
<td>Ocean University of China, China</td>
<td>Guiling Zhang</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Table 2. Pertinent information for each batch of methane and nitrous oxide samples. This includes contextual hydrographic information, median and mean concentrations of methane and nitrous oxide, range, number of outliers, and the overall average coefficient of variation (%).

<table>
<thead>
<tr>
<th>Sampling parameters</th>
<th>PAC1</th>
<th>PAC 2</th>
<th>BAL1</th>
<th>BAL2</th>
<th>BAL3</th>
<th>BAL4</th>
<th>BAL5</th>
<th>BAL6</th>
<th>BAL7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>22.75N 158.00W</td>
<td>22.75N 158.00W</td>
<td>54.32N 11.55E</td>
<td>54.11N 11.18E</td>
<td>55.25N 15.98E</td>
<td>55.30N 15.80E</td>
<td>55.30N 15.80E</td>
<td>54.47N 12.21E</td>
<td>54.47N 12.21E</td>
</tr>
<tr>
<td>Location name</td>
<td>Station ALOHA</td>
<td>Station ALOHA</td>
<td>TF012</td>
<td>TF022</td>
<td>TF213</td>
<td>TF212</td>
<td>TF212</td>
<td>TF046a</td>
<td>TF046a</td>
</tr>
<tr>
<td>Sampling date</td>
<td>24.2.17</td>
<td>24.2.17</td>
<td>16.10.16</td>
<td>17.10.16</td>
<td>18.10.16</td>
<td>19.10.16</td>
<td>20.10.16</td>
<td>21.10.16</td>
<td>21.10.16</td>
</tr>
<tr>
<td>Sampling depth (m)</td>
<td>25</td>
<td>700</td>
<td>3</td>
<td>22</td>
<td>3</td>
<td>92</td>
<td>71</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>Seawater temperature (°C)</td>
<td>23.6</td>
<td>5.1</td>
<td>12.0</td>
<td>13.6</td>
<td>12.2</td>
<td>6.6</td>
<td>6.7</td>
<td>11.8</td>
<td>13.4</td>
</tr>
<tr>
<td>Salinity</td>
<td>34.97</td>
<td>34.23</td>
<td>13.85</td>
<td>17.37</td>
<td>7.87</td>
<td>18.40</td>
<td>18.08</td>
<td>8.81</td>
<td>17.65</td>
</tr>
<tr>
<td>Density (kg m(^{-3}))</td>
<td>1024</td>
<td>1027</td>
<td>1010</td>
<td>1013</td>
<td>1006</td>
<td>1014</td>
<td>1014</td>
<td>1006</td>
<td>1013</td>
</tr>
</tbody>
</table>

Nitrous oxide

<table>
<thead>
<tr>
<th>Number of datasets</th>
<th>13</th>
<th>13</th>
<th>12</th>
<th>13</th>
<th>12</th>
<th>13</th>
<th>12</th>
<th>13</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outliers</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Median N(_2)O conc. (nmol kg(^{-1}))</td>
<td>42.4</td>
<td>7.0</td>
<td>11.0</td>
<td>9.4</td>
<td>11.1</td>
<td>3.4</td>
<td>40.2</td>
<td>11.0</td>
<td>9.6</td>
</tr>
<tr>
<td>Mean N(_2)O conc. (nmol kg(^{-1}))</td>
<td>41.3</td>
<td>7.0</td>
<td>11.1</td>
<td>9.2</td>
<td>11.0</td>
<td>3.4</td>
<td>39.0</td>
<td>10.8</td>
<td>9.5</td>
</tr>
<tr>
<td>Range</td>
<td>34.3-45.8</td>
<td>5.9-7.6</td>
<td>10.1-12.7</td>
<td>7.7-11.0</td>
<td>9.6-11.6</td>
<td>2.1-5.5</td>
<td>30.1-45.9</td>
<td>9.5-11.5</td>
<td>8.0-10.4</td>
</tr>
<tr>
<td>Average coeff. variation (%)</td>
<td>2.8</td>
<td>4.4</td>
<td>4.5</td>
<td>4.2</td>
<td>2.7</td>
<td>7.5</td>
<td>4.0</td>
<td>2.6</td>
<td>4.4</td>
</tr>
</tbody>
</table>

Methane

<table>
<thead>
<tr>
<th>Number of datasets</th>
<th>12</th>
<th>12</th>
<th>11</th>
<th>11</th>
<th>11</th>
<th>11</th>
<th>11</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outliers</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Median CH(_4) conc. (nmol kg(^{-1}))</td>
<td>0.9</td>
<td>2.3</td>
<td>5.7</td>
<td>60.3</td>
<td>4.1</td>
<td>31.3</td>
<td>18.8</td>
<td>5.0</td>
</tr>
<tr>
<td>Mean CH(_4) conc. (nmol kg(^{-1}))</td>
<td>1.8</td>
<td>2.6</td>
<td>5.8</td>
<td>58.6</td>
<td>4.4</td>
<td>31.1</td>
<td>18.8</td>
<td>5.4</td>
</tr>
<tr>
<td>Range</td>
<td>0.6-3.1</td>
<td>1.9-3.8</td>
<td>2.9-8.9</td>
<td>45.2-67.2</td>
<td>2.5-6.5</td>
<td>26.9-35.3</td>
<td>16.5-20.7</td>
<td>3.8-6.8</td>
</tr>
<tr>
<td>Average coeff. variation (%)</td>
<td>10.9</td>
<td>7.2</td>
<td>8.6</td>
<td>2.1</td>
<td>4.3</td>
<td>3.5</td>
<td>4.2</td>
<td>6.5</td>
</tr>
</tbody>
</table>
Figure 1. Concentrations of methane measured in nine separate seawater samples collected from the Pacific Ocean (Fig. 1a, 1b) and the Baltic Sea (Fig. 1c, 1d, 1e). The dashed grey line represents the value of methane at atmospheric equilibrium (Fig. 1b.) Individual data points are plotted sequentially by increasing value. The same color symbol is used for each laboratory in all plots.
Figure 2. Deviation from the median methane concentration (reported as absolute values in nmol kg$^{-1}$) for the seven Baltic Sea samples. The batches of seawater samples include BAL1, BAL3, and BAL6 (Fig. 2a), BAL4, BAL5, and BAL7 (Fig. 2b), and BAL2 (Fig. 2c). The shaded grey area indicates values ≤5% of the median concentration. The color scheme for each laboratory dataset is identical to that used in Figure 1 and the letters allocated to each dataset are to facilitate cross-referencing in the text. Note that the y-axis scale varies between the Figures.
Figure 3. FID response to methane, fitted with a linear regression calibration. The inclusion (Fig. 3a and Fig. 3b) or exclusion (Fig. 3c) of low methane values cause the calibration slope and intercept to vary. However, the observed variation in the calibration slope does not have a significant effect on the final calculated concentrations of methane. In contrast, variation in the intercept does have an effect on the final concentrations of methane.
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.
Figure 5. Deviation from the median value (reported in absolute units) for nitrous oxide datasets. The batches of samples include BAL1,2,3,6,7 (Fig. 5a) and PAC2 and BAL5 (Fig. 5b). The Baltic Sea samples are represented by circles and the Pacific Ocean samples are represented by triangles. The shaded area indicates a deviation ≤5% from the median value, based on a water-column concentration of 11 nmol kg⁻¹ and 42 nmol kg⁻¹ for Fig. 5a and 5b, respectively. The color scheme for each laboratory dataset is identical to that used in Figure 4 and the letters allocated to each dataset are to facilitate cross-referencing in the text. Note the y-axis for Fig 5a and 5b are plotted on a different scale.
Figure 6. Three calibration curves for nitrous oxide measurements using an ECD including linear (Fig. 6a), multilinear (Fig. 6b), and quadratic (Fig. 6c) fits.
Figure 7. Comparison of sample storage times with measured concentrations of methane (Fig. 7a) and coefficient variation (Fig. 7b) for two sets of seawater samples (PAC1 and PAC2) collected in February 2017. These two sets of seawater samples had the lowest methane concentrations and appear to be influenced by the duration of storage time. The data points enclosed in parentheses were not included in the regression analysis. The PAC1 regression line is black and the PAC2 regression line is grey. All of the storage times are included in the Supplementary Material.