

Responses to the reviewer

We would like to thank the reviewer for their thorough feedback. We answer the reviewer's comments below, indicated by italic text.

I wish to acknowledge the work done by the authors to revise the manuscript. I am glad that my review enabled you to identify major errors in the dataset that could be corrected before publication. I believe this is an example of the peer review process working. Given the manuscript was completely rewritten it took time for me to carefully review the new work. I appreciate your patience.

Thank you for re-revising the manuscript.

My biggest remaining with the manuscript is that they have not adequately explained how the atmospheric CH₄ analyzer was calibrated and quality was ensured. They state it was "benchmarked daily" but do not explain whether the benchmark data was used for any drift correction of the data. The composition of the gas is also confusing.

The author's response (page 14) states "The cylinder used in the field was filled with CO₂ mixed with ambient air, so that methane concentrations were not determined, but amounted to ~5 ppm with unknown uncertainty. Using this cylinder in the field was not planned, but due to some logistical issues, it was the only option for benchmarking."

The revised manuscript, line 86 says the analyzer was "benchmarked daily (except for the first two days due to logistical issues) with a certified standard gas mixture (from Praxair) of 450 ppmv CO₂ balanced with air containing 5000 ppbv CH₄, which was well within the analyzer's measurement range".

The first statement implies the final concentration was 5 ppmv and the second implies that the original concentration before dilution was 5 ppmv. Considering that the concentration of CH₄ in ambient air is 2 ppmv, I don't understand how a standard gas diluted with ambient air could end up with a final concentration of 5 ppmv CH₄ (2.5x the concentration of ambient air). What was the concentration of CO₂ in the final mixture (based on dilution this would tell us what was the original combination in the ambient air)?

I would suggest rewording it as "benchmarked daily (except for the first two days [list dates here] due to logistical issues) using a gas that was a mixture of a 450 ppm CO₂ standard from Praxair and air containing roughly ZZZ ppbv CH₄. The concentrations of CO₂ and CH₄ measured in the benchmark gas over the whole cruise were X \pm x and Y \pm y (mean \pm standard deviation), respectively. Add text here quantifying the drift in CH₄ and CO₂ observed during the cruise on the benchmark gas and explaining whether the benchmark gas data was used to perform any post-corrections to the atmospheric data or simply as a check to see whether drift was detectable. Add a statement that a post calibration with a certified reference gas was not done and not used to post-correct for drift (I think this is correct but I am not sure). State the date that the sensor was calibrated (currently it is stated as July 2021 and the cruise stated July 12).

Thank you for your suggestion. The gas cylinder's certificate indicated a concentration of 450 ppmv for CO₂. It was not certified for CH₄. Only based on our measurements, we determined a concentration of

about 5 ppmv. We clarified this in the text and followed the reviewer's suggestions by adding more details about the benchmarking procedure. We would like to note that the benchmarking was most suited to determine whether the analyzer and the setup was functioning properly, to some extent also to determine how precise measurements were during each benchmarking, and it may also be used to determine drift. However, benchmarking during field campaigns often shows a larger drift in the measurements than what we yield from in-lab tests. This may be attributed to different ambient conditions, vibration, power supply, etc. that could slightly affect measurements in the field, but these conditions are more controlled and difficult to reproduce in-lab. A post-expedition in-lab benchmarking was conducted, on which we based our decision not to post-correct our data. We also included this information in the text.

Other comments:

In my opinion, the current title is misleading: "Sea-air methane flux estimates derived from continuous atmospheric measurements and marine depth profiles in cold seep regions"

The previous title was somewhat better: "Methane flux estimates from continuous atmospheric measurements and surface-water observations in the northern Labrador Sea and Baffin Bay"

I believe the current title is misleading because

- A. The depth profiles are not used to calculate fluxes – only the surface measurements
- B. The continuous observations are not used to calculate fluxes: you take a 10 min average coinciding with the surface water measurements, as indicated on line 155. If I only read the title I would assume that you had used the eddy covariance technique to calculate fluxes.

I would recommend a title such as "Continuous atmospheric methane measurements and marine depth profiles in the northern Labrador Sea and Baffin Bay" OR

"Sea-air flux estimates derived from marine surface measurements and instantaneous atmospheric measurements in the northern Labrador Sea and Baffin Bay"

Thank you for the suggestion. We changed the title to "Sea-air methane flux estimates derived from marine surface observations and instantaneous atmospheric measurements in the northern Labrador Sea and Baffin Bay".

line 76: "Meteorological Tower" and "Model" should not be capitalized

We revised that.

Line 132: "The analytical precision was estimated from repeat measurements of standard gases and amounted to 0.5–0.8 % or better for dissolved CH₄ similar to previous studies (Punshon et al., 2014, 2019)." It is not appropriate to report analytical precision based on standard gases. You need to report precision based on samples, which have a completely different workflow and likely a lower precision.

We only took one sample per depth and location and only analyzed each sample once. This means we did not report the precision based on samples. What we reported here was analytical uncertainty of the instruments. We reworded the phrase to clarify this and added a statement that replicate samples were

not taken.

Line 134: change “and density” to “and potential density”

We revised that.

Line 144: “we also used seawater density and oxygen data” – but no oxygen is presented in the paper, nor in the data.

We do not show any figures including the oxygen data, but determined oxygen levels in the mixed layer and report a statistical relationship in the Discussion. Therefore, this data is mentioned in the Methods, but we don't think that it requires inclusion in the dataset, also because we don't present any other mixed layer data there. We added a few words to the text clarifying the use of oxygen data in this study and that the data is not shown further.

Line 270: “The mixing ratios measured in this study are higher than those determined from flask samples likely due to the influence of CH₄ seeps in our study area.”

I think it is misleading to claim this is the “likely” cause without any data demonstrating differences in methane fluxes between the regions where the flasks were collected and where your study occurred. The observed near-surface saturations of roughly 110 to 150% are similar to other Arctic seawater measurements. You state earlier in the article that nearly all of the seep-derived methane was oxidized before making it to the air-sea interface. What if the difference is just a calibration offset?

We revised the statement in the text. Flask samples usually undergo strict protocols to assure the correctness of gas mixing ratios. Our gas analyzer was calibrated prior to deployment and the “goodness” of this calibration was confirmed by benchmarking afterwards in-lab. Therefore, we are confident to state that flask samples showed lower mixing ratios compared to our measurements.

Data: I mentioned in my previous that the exact depth needs to be archived and this has not been fixed – it is still listed as the approximate depth e.g. 100 m, 800 m, etc.

We revised the depth and updated associated figures and the dataset.

Rows 76-78: three samples from three different locations are reported with exactly the same temperature and salinity (SE-1K, NE-1K, NE-5K). Is this a typo? I assume that you can obtain T and S data from the underway thermosalinograph if a rosette was not deployed.

Thanks for catching this. We used available data from the Rosette casts to fill missing data, revised the dataset and updated it accordingly. For your information, we carefully double-checked all salinity, pressure, temperature and depth data and found a few more typos which were revised, not noticeably affecting the results. We also revised the manuscript accordingly.