Response to reviewer 1:

Firstly, we thank Dr Sims for his very thorough and helpful review. Our responses are in line, in italics, with the specific comments below. In some cases we have responded once following a group of comments on a particular section of the manuscript. In most cases we have made the suggested changes, and where we have chosen not to, we give our reasons. We attach a file to this response containing the figures modified following the suggestions of both reviewers.

- Line 14 – This sentence should be improved as it sets the context of the paper. It would be better to describe the issue in terms of the importance of the CAO in the carbon cycle. It may also be necessary to describe the uncertainty as a separate point or in a separate sentence. It is not clear that climate forcing gases refers to CO2 and CH4 which are mentioned just below.
- Line 15 – 16 – There needs to be a point mentioning the need for baseline data before this. What about baseline direct CO2 flux data for CO2?
- Line 18 – The season in which the measurements were collected is very relevant for the abstract
- Line 20 – This point is not clear at all on first reading without consulting the text. On the surface this median value looks like it is based on conditions you experienced during the cruise. What you mean is that this value is appropriate for all low wind speed conditions in sea ice leads.
- Line 23 – How do you measure snow flux when you only access sea ice leads with floating chambers? The basic methodology is missing from the abstract for this result.
- Line 24 -25 – Presumably you mean carbon/CO2 here and not CH4? The C vs CO2 distinction is important as your fluxes are in Tg.
- Line 24 – How does this flux of presumably carbon compare to previous measurements?
- Line 25 – Over the whole cruise? Is this the CAO? Presumably this is August and September? Or the whole ice free season?

We have modified the abstract following the suggestions above, and addressed the questions. Other modifications to the abstract follow suggestions from the other reviewer.

- Line 29 – I would welcome a broader introduction than this. Discussing anthropogenic climate change, the need to know the global budget for these two gases and why the Arctic Ocean is considered important.
- Line 33-35 – this is important and should be above line 29 in the introduction.
- Lines 35-39 – Again the uncertainty around parametrisations in sea ice should be introduced before you introduce the CAO as your study area
- Lines 38-39 – this also needs to come earlier
- Lines 40-42 – this belongs in the methodology
- Lines 42-45 – move this up and after introducing the role of the arctic say that the CAO is deep and accounts for 47 % of the surface area of the entire Arctic Ocean.

- Line 50 – I would finish the paragraph with this point again and state the importance of finding this out
- Line 60 – do you need to say "the authors" here?
- Lines 81-95 – this is very nicely written overall but you don't touch upon stratification which is now thought to have a big impact on fluxes. Please add references to Miller 2019 or Ahmed 2020. You also don't make reference to the uncertainties with the two prevailing methods EC and chambers, this is key background info that I now realise is in the discussion but it must come earlier to set the scene for the reader.
- Lines 96-101 – please explain why this is relevant to a reader, presumably because gas escape from the sediments saturates the above water column.
- Lines 104 – ok but can you possibly tease what analysis you do and what you discuss?

  We have modified the introduction to address the above points.

- Line 110 /Figure 1– What is on this figure is clear but all the locations you include in the text here need to be marked on the figure. Additionally you need to indicate on the map where you were at certain points in the cruise.

  We have added locations to the map, and also numbered the measurement stations with corresponding numbers added to Table S1.

- Line 122 – Were any actions put in place to reduce turbulence? E.g. captain told to use props at the other end of the ship?

  We've added a brief description of what actions were taken.

- Line 140 – What is the uncertainty of this setup?

  We have expanded the discussion of chamber uncertainties, rearranged where in the paper these issues are brought up, moving material from the discussion to the introduction. We have also given a more detailed description (including supplemental photos) of the chamber design. This chamber design has been shown to have agreement with k determined from other methods, minimising the biases observed in earlier chamber studies. There may of course still be biases with the chamber. But we are not able to quantify these. All methods to do so that we are aware of require a baseline measurement of k to compare with (e.g. Mannich et al., 2019), likely requiring laboratory measurements which may have a questionable applicability to the field conditions we measured in, and are in any case beyond the scope of this work.

  In order to better represent the uncertainty in our measurements, and following the suggestion of the second reviewer, we have determined a sampling uncertainty for the surface partial pressure measurements, and propagated this uncertainty through to the gas transfer velocities, combining it with the sampling uncertainty for the chamber flux measurements.

- Line 145 – Cite Yang 2016 who show nicely the limit of detection for the two main suppliers of fast CH4 analysers
- Line 147 – Were the CO2 also below the limit of detection? The ΔpCO2 for most of the cruise was ~100 µatm, you would expect that would help provide detectable fluxes. More information would help an interested reader!
- Fig S1b – Note that on your Matlab plot at 08/22 there is a straight line for pCO2 which is not real data, this looks like a data outage. Please change the marker style on this plot to avoid this.
- Line 153 – please state the maker of the CO2 sensor
- Line 164 – this methodology is fine but can you just be clear that you stored the samples in the cold and dark temporarily and then heated them to 25 when you needed to do the measurement.
- Line 183-189 – Please state the material of the plastic used for your headspace equilibration as this will have an impact on leakage through the wall.
- Line 183 – Please can you reference this method?

  We have modified the methods section following the above suggestions with responses / answers in the text as required.

- Line 195 – 200 – Please can you provide a citation for this method? Or state differences from other similar methods e.g. Else 2022. Gas exchange out of the plastic bags over 3-4 days is not insignificant here, why could this not be done in <12 hours? Did you do any replicates to check this method? Or any nearby adjacent cores?

  We have chosen to remove the ice core data and description from this manuscript, as described below.

- Line 221 – This is of course not wrong but the subscript x use here is not very clear. Perhaps just have two equations for flux of CH4 and CO2. This would make this much easier.

  We have decided to keep the 'X' notation signifying gas species as we think the meaning is clear here and later in the results.

- Line 223 – Join these two sentences. Ideally you would have made this correction, if you chose not to you at least need to state the % change when you state the correction is very small.
- Line 234 – I think the solubility equation for CO2 is from Weiss, please cite that as the original literature source.

  Changed as suggested.

- Figure 3 – Consider putting this in the supplement it is not that informative.

  We decided to leave this figure in the manuscript as we believe it gives useful context to the measurements.

- Figure 4 /Fig S1 – These plots are very similar. Why have the supplementary plot at all? You don't actually cite it in the text?

  We have made reference to this plot in the text and choose to keep it as it more clearly shows the differences between the different sampling methods for the seawater concentrations, which the other reviewer requested be emphasised.

- Figure 5 number these a/b/c etc. Can you also improve the visualisation of this plot? You need bigger axis labels for salinity, depth and methane. You can also make the markers much bigger. You ideally need to add error bars to these data points

  Panels are already individually labelled with sampling dates, and so a/b/c labels are redundant. We have increased the size of the labels as suggested.

- Fig 6 – Please get rid of the Fx and just make it FCH4 in panel a.

  The flux, F in this plot is of both species as indicated in the legend, and so Fx is more appropriate.

- Line 354-356 – any indication of any of these processes at play here from data you collected e.g. underway chl-a?

  Surface water samples were unfortunately not analysed for chl-a or other biological indicators. Extensive biological sampling was performed from the CTD and we do have calibrated chl-a data from the CTD published at PANGAEA (links below). Initial examination did not reveal a clear link to the flux measurements published. Further analysis of biological influence will be pursued in future work but is beyond the scope of the current manuscript.

  Profiles: https://doi.pangaea.de/10.1594/PANGAEA.951266

  Bottles: https://doi.pangaea.de/10.1594/PANGAEA.951264

- Line 359 – it would help to elaborate on this hypothesis for a sea ice based methane source as I think the context would be helpful to the discussion. You are in a position to report on the flux of this to the seawater as you have already collected all the data you need.

  Ice core data was brought into this manuscript at an early stage but as the analysis progressed, the focus here became the surface fluxes, while a more detailed analysis of the ice core measurements and deeper CTD-derived trace gas concentrations and related oceanography from SAS will be the focus of future publications. To keep this focus, and as the ice cores were only described in the methds and then very briefly mentioned in the results, we have removed them entirely from this paper, and also

removed comments such as this one in the discussion that can be more properly addressed in an analysis of the ice core measurements.

- Line 359 – Can you provide the dates for the stratification in addition to pointing to the figure?

Added as suggested.

- Line 359 – Based on the depth of the mixed layer can you calculate what the flux of ch4 has been from the sea ice to the surface mixed layer?

  As per the ice core comment above, we choose to keep the current manuscript focussed on the air-sea exchange. Mixed layer depths are not presented in this paper, though are likely to form part of future publications based on data from this expedition. While the back-of-the envelope estimated suggested would add some context to the current manuscript, we believe it would not change the results or conclusions presented and so we omit it here.

- Line 369 – p on pCO_2 should be in italics throughout.

  Here we are quoting a previously published paper verbatim.

- Line 363-372 – This feels like it belongs in the introduction

  Moved as suggested.

- Line 374 – Can you really say this without accounting for sea ice melt? You have values for the sea ice so you can account for mixing of sea ice and seawater to get the mixing endmember as they do in Meire et al., 2015 and also Sims et.al 2023. If you are saying the 90 µatm gradient is due to biological activity you need to calculate what productivity rate would be needed to create these surface CO2 gradients.

  As per the comments above relating to ice cores, this calculation will be more appropriate in future publications focussed on the ice core and oceanographic data from SAS. Here, our analysis of data beneath the sea ice was limited, and so we have instead omitted this comment.

- Line 390 -402 this is quite key information that I felt was missing from the introduction.

  We have moved this section to the introduction.

- Line 420-425 – this is a big result. What I feel is missing is a bigger discussion around the physics. How does this compare to the model of Bigdeli for example? Can they model at this scale? Where are the differences and what processes might the model be missing or parametrising incorrectly?

We have added the Bigdeli et al., 2018 WAGT model to the results and expanded the discussion of the gas transfer results as suggested.

- Line 426-488 – I would include the region fluxes in the discussion and then have a separate conclusion where you summarise the entire paper. At the moment the lack of a separate conclusion does not highlight all the results and findings of this great work.

  We have moved the regional flux section to the discussion and created a separate conclusion as suggested.

Response to reviewer 2:

We thank Dr Else for his thorough, insightful and helpful review. Our responses to his criticisms and comments are in line and italicised. A file is attached to the response to reviewer 1 containing the figures modified following the suggestions of both reviewers.

1. Major criticisms

1.1 I'm a bit concerned with over-interpreting the results shown in Fig. 7. There are relatively few data points here, and the r2 regressions are not good (and presumably not statistically significant). I think the figure should stay in the paper, but the authors need to make sure to exercise caution in their interpretations, and be sure to advise future readers against using the results to widely. Some specific comments that I think should be revised include:

L406: "The wind speed dependence of all measurements (Fig. 7) is close to quadratic and in close agreement with that from a large Arctic lead of ~100-400m in width (Prytherch and Yelland, 2021)"

Can we really say that it's close to the Pryherch & Yelland curve? There are statistical tests that are available to test if a regression-derived curve is significantly different than another curve, or if differences are likely due to sampling error. In this case, the number of samples is too small to do proper statistics, so real caution needs to be applied here when comparing curves.

We agree with Dr Else's comments. We have modified the text describing the results in section 3.4 and the text highlighted by Dr Else in the discussion (4.2) to emphasise that the small number of data points and the weak correlation of the regression to those data.

L424: "the results reported here show that gas exchange can be appropriately represented with a constant k600, 2.5 cm hr-1"

The way this is worded it seems like this could be applied anywhere. That seems like a stretch to me considering the number of data points. I would recommend rewording this to be clear that this is a result only applicable to this study.

We agree and have clarified this comment so that it more clearly applies only to the measurements reported in the manuscript.

1.2 I understand the desire to take the observations made in this study and apply them over a large area. In fact, I think it should be done; it helps characterize the relative importance of leads and sea ice to overall budgets, at least to an order of magnitude. But there are a few problems with the approach used in section 5 that may make one question whether or not the estimate is accurate even to order of magnitude resolution:

L440 - this estimate essentially assumes that the air-sea pCO2 gradient will remain constant all year. While some discussion about this is presented, it doesn't fully capture the complexity of the problem. For example, it is quite possible that the air-sea pCO2 gradient will invert at some point during the winter due to net respiration, and offset some of the CO2 uptake. The seasonal pattern of pCO2 in ice covered waters has been discussed quite a bit, but of course I'm most familiar with my own work (Else et al., 2019: https://doi.org/10.1029/2018JC013899, Else et al. 2012: https://doi.org/10.1029/2011JC007346) and one of the earliest papers to discuss this: Yager et al. (1995: https://doi.org/10.1029/94JC01962).

We have modified the text to expand on the poorly constrained annual cycle of pCO2 (using one of the suggested references) and clarify that the measurements we report can properly only provide regional estimates for the season in which they are measured.

L450 - again, extrapolating the sea ice fluxes through the entire year assumes that the same mechanisms and gradients will be driving exchange. This is in contrast with observations derived primarily from the Nomura papers (that the authors cite) that show thin (new) ice releasing CO2 to the atmosphere, cold (thick ice) exchanging very little with the atmosphere, and melting ice absorbing CO2 from the atmosphere. This seasonal pattern needs to be considered, as the different seasons (to some degree) will cancel each other out. There is not enough data in this paper to tackle this; I know Dr. Nomura is working on compiling seasonal observations to come up with a more rigorous estimate.

We have removed the extrapolation of the sea ice fluxes to an annual estimate.

L461 - again, extrapolating the CH4 fluxes year round assumes a constant supply of CH4, even though the origin of dissolved CH4 is not identified in the paper, and is therefore highly uncertain.

We agree and have clarified that this annual estimate is highly uncertain.

Overall, I have to question the veracity of trying to take these seasonal-specific measurements, and applying them year round. I do not have so much trouble with the authors approach of scaling them spatially. But I think it would be wise to limit temporal extrapolation to the season of observation.

For both CO2 and CH4 we have left the actual calculated values in the paper as a reader may want them at east for comparative purposes, but we have clarified that these estimates are very uncertain and limited their emphasis. We have omitted mention of the annual values from the abstract and conclusions.

2. Minor criticisms

L130 - I would have appreciated a bit more description of the floating chambers. Are they commercially available? Were the custom built? A photograph (even in the supplemental) would go a long way here. The use of floating chambers in these environments is relatively new, and other members of the community might find it useful to have more details on their application here.

We have added further description of the chambers to the methods section and photographs in the supplemental materials.

L143 - The authors mention that ice chamber measurements were made with and without a snow cover. A justification for why a measurement without a snow cover should be presented. I have never understood why this is commonly done. It seems to me that removing the snow cover immediately changes the temperature regime of the surface ice, leading perhaps to freezing (under low temperature conditions) or melt (warmer conditions), both of which may induce an artificial flux. This should also be considered in analysis of these results.

While other studies have attempted to compare flux measurements with and without snow cover to determine the impact of snow on the flux (e.g. Nomura et al., 2018), on reflection, the ice-air flux measurements in our study do not enable any particular insights, and as they are not used in the later analysis and so just create unnecessary complication, we have decided to remove them from the manuscript.

L156 - There are a lot of different ways that water samples were collected here, and I wonder if they all did indeed adhere to the Dickson et al. (2007) protocols. Seems almost by definition that the couldn't have. The one that I would most like to see explained is the note about "bottles submerged at depths of 0-10cm and syringes at depths of 0-5 cm". Under these conditions, were the samples added to the bottles allowing a 1.5-2x overflow of sample, as per the Dickson protocol? I understand that in these situations not all of the protocols are sometimes maintained, but it's important to identify which ones are not so that the data can be interpreted appropriately.

We agree this is important. For the surface water bottle sampling, the priority was sampling as close to the surface as possible and bottles were submerged by hand. As such it was not possible to exactly follow the Dickson protocols, although as much as possible air-water mixing was minimised during the submerging process. The syringe sampling is not described in Dickson, but air-water equilibration is an intrinsic part of the process and was performed

immediately after sampling. Further description of the sampling and where it does and does not follow Dickson protocols has been added.

L164 - Define the storage time (at least a range)

Samples of TA and DIC where typically analysed within 12 hours. We have clarified this in the text.

L150 - L190 - It seems like there were A LOT of different ways to measure pCO2/CH4 utilized here. How comparable are these different methods? Was a subset of samples used for intercomparison? For example, I would expect potentially different results from pCO2 derived from CO2SYS vs. pCO2 from the headpace shaking method. This is a bit of an issue that might need to be explained when looking at the results.

We agree that we hadn't described the differences resulting from the different measurement methods. We have expanded our description of the differences in the methods, results and supplemental materials. We have furthermore determined the standard error of the surface partial pressure measurements. As for both pCO2 and pCH4, the surface partial pressure presented in the paper is (usually) an average of two measurements (a syringe measurement and a bottle measurement), the standard error is equivalent to half the difference of the two measurements. This standard error makes the method difference apparent, and is included in the results tables and figures. We also combine this error with that of the flux measurements in the uncertainties given for the gas transfer velocities. We note that low-salinity samples (~0.2 – 2) of low DIC and TA content (~50-120 µmol/kg) are inherently associated with large uncertainties in the computed pCO_2w.

L200 - the specific analyses conducted on the ice cores are not noted. Was it DIC/TA? pCH4/pCO2? both?

While ice cores were analysed for DIC/TA and pCH4 during the expedition, only the pCH4 results are referred to in the paper. This has been clarified in the text.

L313 - It is stated that "grease ice impedes gas exchange", but it's not clear whether this is a result of this paper, or a statement of known fact from a previous study. Please clarify.

We have modified the text to clarify that this is a finding from this paper, and one that may not be of significance as no effect is found on CO_2 flux, and the effect on CH_4 is within the standard deviation of the measurements.

L347 - Here is where some discussion on the potential influence of clearing snow from the sea ice may be relevant. I would expect for example, an artificial positive flux associated with snow clearing on sub-freezing days, which induces surface freezing, brine rejection, and outgassing.

L406 - specify how lead width was measured

Lead widths were measured by laser ranging (section 2.4) – we have added an additional note to this effect to this paragraph.