Dear Editor,

We feel that after addressing both reviewers' concerns the manuscript has improved and really shows the value of the developed CO_2 and CH_4 gliders. Please find the suggested edits in bold below each reviewer's comment and as track-change in the resubmitted version of the manuscript.

Thank you for your service as Editor and we are looking forward to your decision.

Best regards,

Claudine and co-authors

<u>Comments from Dariia Atamanchuk</u>

The manuscript by Hauri et al. details the integration of CO2 and CH4 sensors into the Seaglider and its performance during trials and rigorous data quality assessment. This work is an important contribution to the community effort to increase the number and type of observations in the global ocean, with a particular focus on the water column biogeochemistry and GHG fluxes.

The biggest obstacles to high-resolution carbonate system data from mobile platforms to date are the size, power consumption, and slow response time of the gas sensors. While pCO2 sensors have a relatively slow response, it is still possible to correct the glider-borne data to get decent quality, as nicely shown in this work. Please see my comment on Figure 8, though. As for the methane sensor, I've been wondering whether glider deployments are the best avenue for getting water-column data. The very slow response time of the CH4 sensor doesn't allow for full advantage of the glider-specific capabilities and allows for getting qualitative results, at best. In this respect, ROV-based surveillance or moored platforms would perhaps be a better and more economical option for methane monitoring. It all, of course, depends on the research objectives: qualitative vs quantitative assessment of CH4 distributions. The combined CH4/CO2 sensor is a great idea, too, but given the difference in response times, it will be hard to take full advantage of such a package on gliders. Perhaps an additional discussion on the complexity of the problem of ocean observing - there is no 'one size (sensor) fits all' approach when talking about autonomous sampling, sensors and platforms - would help to orient the readers in the field and help them appreciate the presented work even more.

These are some very interesting thoughts and align with Damian's (reviewer 2) comments, too. We reframed the discussion accordingly.:

"The SG HydroC CH_4 was successfully integrated into the Seaglider as part of this project. While tank experiments showed promising results, short field tests of the CH_4 Seaglider in shallow water revealed low and patchy methane concentrations near the detection limit. The CH_4 Seaglider requires further testing in environments with strong pCH_4 gradients during longer and deeper dives (to allow for equilibration) to assess the accuracy of its response time-corrected data in the field. The sensor's slow response time likely limits the glider to providing qualitative rather than quantitative results. However, due to the scarcity of oceanic CH_4 observations, deploying a CH_4 glider can help identify the location of methane sources and guide the placement of in situ observations to conduct a more quantitative assessment of CH_4 fluxes and dynamics."

Competing interests. Clearly declare the potential of a direct financial benefit to the co-authors affiliated with the private companies. That's what this section is for.

Thank you for pointing this out. We added the following statement:

"Authors Hayes and Abdi are employed by AOOI and CSCS (respectively) and their objective is to support the ocean research community by providing innovative, cutting-edge observing technological solutions. These include autonomous platforms and related services in unique configurations. Through the support of the National Science Foundation and the National Oceanographic Partnership Program, AOOI was able to jointly develop the CO2 and CH4 gliders and prove and improve the scientific utility of this approach. Authors Kinski and Kemme are employed by -4H-JENA engineering GmbH, the manufacturer of the HydroC CO2 sensor. The objective of -4H-JENA engineering GmbH is to provide best possible accuracy of dissolved gas measurements on any platform and at any environmental condition. Intensive collaboration with scientists is essential for the development of these products."

Minor comments:

L33: 'greater CH4 activity' – either specify the location or use something general e.g. 'field mission'.

We rephrased to: "The CH_4 Seaglider passed its flight trials in Resurrection Bay, yet needs to be tested during a field mission in an area with CH_4 concentrations beyond background noise."

L40: pCO2 and/or fCO2
Done.

L61: Some of the listed obstacles are the necessary steps to ensure high data quality. Please make a distinction here: those to assure QC and those unique to the glider/mobile/profiling integration or data processing.

We rephrased: "Although biogeochemical sensors deployed on autonomous platforms like moorings and Argo floats have become more prevalent in recent years, challenges such as high power requirements, sensor size, and data quality hinder their widespread use on underwater gliders."

L76: Here, at the start, some context as to why you mention pH sensors would be great.
We rephrased: "Ocean gliders autonomously collect water column data along planned waypoints, which allows for controlled exploration and adaptive sampling. To date, pH is the only carbon system parameter that has been successfully integrated into ocean gliders (Hemming et al., 2017; Saba et al., 2019; Possenti et al., 2021; Takeshita et al., 2021). The most promising results came from ISFET based pH sensors (Saba et al., 2019; Wright-Fairbanks et al., 2020; Takeshita et al., 2021).

L111: Zero correction is good, but span correction is also needed in post-deployment treatment.
Thank you for the comment, span correction was included in the post-processing when a post-calibration was conducted but was not clearly stated in the text and has been corrected:

"SG HydroC CO_2 (SG HydroC CO2T-0422-001) data from the tank experiment (Table 1, Figure 4) and rosette mounted CTD casts (Table 2 and 3, Figure 7 and S1) were post-processed to correct for baseline drift (change in the zero signal reference) and span drift (changes in the sensor's concentration dependent characteristics) using pre- and post-calibration coefficients interpolated over the deployment (Fietzek et al., 2014). For the May 2022 Seaglider integrated SG HydroC CO_2 sensor (SG HydroC CO2T-0718-001, Table 3, Figures 8 and 9), data were post-processed with pre-calibration coefficients only (no span drift correction) because the sensor was damaged during the return shipment for post-calibration."

L116: Indicate rpms of the pump used.

RPM is not stated on Seabird's documentation. Seabird states a 5M has a flow rate of 25 ml/s (https://www.seabird.com/modular/sbe-5m-mini-submersible-

pump/family-cms.block?productCategoryId=54627473799) but this was not measured directly when powered by the Seaglider.

L133: Please provide more details here about the testing, it is not clear from the description.

We rephrased to: "In situ comparison of the orientation of the sensor and close examination of pCO_2 and internal pressure data suggested the highest data quality was achieved with this mounting design.

L143-onward: are the drivers available for users in open-access mode should anyone decides to outfit their glider with the sensor? The same for SIRMA. Please clarify.
We now state in the data availability section: "The HydroC specific SIRMA code is available on Github (Cyprus-Subsea, 2024) and the CNF file is available as supplemental material. More detailed information on the HydroC – glider integration and operation can be found in the OceanGliders CO_2 Standard Operating Procedures (Irving et al., 2024)."

L168: Please specify the wavelengths of the ecopuck channels.
We now specify in the manuscript: "... Wetlabs Ecopuck measuring chlorophyll fluorescence at 695 nm."

L178: Maybe paste a link to the glider specs here. Please provide a detailed description of the mission here (distance, duration, energy use, number of dives, depth, number of water samples collected and their depth, etc).
We don't think a link to the glider specs is necessary nor appropriate at this point. We added the requested information in section 2.3:
"2.3 Spring and winter CO_2 Seaglider missions

Both versions of the CO_2 Seaglider (rated to 300 m versus 1000 m) were tested in separate missions (Figure 3, Table S1) in spring (53 dives, May 4 – 7, 2022, Figure 5) and winter (310 dives, February 8 – 21, 2023, Figure 6). The 300 m version with integrated POM housing was tested during a five-day-long mission in May 2022. The glider followed along a transect within Resurrection Bay. CTD casts near the glider path allowed for in-depth evaluation of the data quality. The 1000 m depth-rated CO_2 Seaglider with integrated titanium housing was tested in February 2023. Estimated energy consumption during the CO_2 Seaglider missions was 19 out of 135 Ah and 75 out of 120 Ah for the 24 V which powered the SG HydroC CO_2 sensor battery for the spring and winter missions, respectively."

L190: on rocks??? Maybe a sketch diagram will help picture the flow-through setup.
We don't think that a sketch up is necessary
We simplified "were secured next to the Seaglider".

L197: It's not clear what you mean here.
We reworded to: "The SBE-55 and SG HydroC CO2 were powered by a SBE-33 carousel deck unit. The SG HydroC CO2 interfered with the communication stream and thereby prevented real-time data acquisition and control of the SBE-55, however HydroC data were internally logged as required."

L201: Which depths were targeted? And why? Please describe the CTD cast, including the number of stops, their duration, and the reasoning behind such an experiment design.
We rephrased: "The SG HydroC CO_2 interfered with the communication stream and thereby prevented real-time data acquisition and control of the SBE-55, however HydroC data were internally logged as required. Depth of the rosette package was monitored directly on the winch and the timing of firing of the sample bottles, after an approximate 15 min hovering period to (allow for equilibration), was programmed in advance based on time intervals. Target depths for discrete water sample collection were 5 m, 20 m, 40 m, 60 m, and 80 m. However, only samples from the upper 20 m of the water column were usable due to issues with manually measuring the depths and the sample collection."

L205-208: This paragraph probably fits better with the description of the tank experiment above.

We think that this paragraph fits best where it is at the moment, since it describes discrete water samples in general.

L283-286: Was the span correction applied too? Was the detector calibrated in post-processing beyond zero calibration?

Yes, span drift was corrected when post-calibration was carried out (see comment above and corresponding addition to paper). Span drift correction was not performed for the Seaglider mounted HydroCs because post-calibration was not possible (in one instance the sensor was damaged on transport back to 4H Jena, and the other instance the sensor was lost) so the temporal stability of the sensor response slope (span drift) could not be determined.

We clarified: "SG HydroC CO_2 (SG HydroC CO2T-0422-001) data from the tank experiment (Table 1, Figure 4) and rosette mounted CTD casts (Table 2 and 3, Figure 7 and S1) were post-processed to correct for baseline drift (change in the zero signal reference) and span drift (changes in the sensor's concentration dependent 2023 the May 2022 Seaglider integrated SG HydroC CO_2 sensor (SG HydroC CO2T-0718-001, Table 3, Figures 8 and 9), data were post-processed with pre-calibration coefficients only (no span drift correction) because the sensor was damaged during the return shipment for post-calibration."

L291: It's not clear here why the real-time data is needed for post-processing. Wasn't the sensor recording data internally? How do you expect this lack of calibration to affect data quality?

The data were recorded internally, but the glider was lost at sea during the winter mission. One of the key parameters for post processing (*p_NDIR*) was not included in the real-time data. Since the HydroC was only collecting data for ~4 days during the spring mission and ~2 days during the winter mission, we do not believe this lack of calibration affected data quality.

We clarified: "The *p*CO_2 data from February 2023 were not post-processed because a required parameter was not relayed in real-time and the glider was lost. Lack of post-calibration most likely had no negative effect on the quality of data since the HydroC was only collecting data for ~4 days during the spring mission and ~2 days during the winter mission."

We addressed the loss of the glider in more detail at the end of section 2.3 - Spring and winter CO2 Seaglider missions:
"Before the February mission the on board modem was replaced with a different model, with different input voltage requirements, which were probably not met as the mission evolved. As a result, the glider could not communicate anymore and was lost. While this was an unfortunate mistake, the loss of the glider had nothing to do with the HydroC CO2 integration."

L305-307: you refer here to the detector response time, not the sensor response time, correct? Please be clear. Because in the next paragraph, you give 106,108 and 109s as the response times. This could be confusing to the reader.
We clarified what we mean by response time:
"The ability to determine the in situ response time (t_63 of the HydroC, which took into account membrane characteristics and the rate of water exchange over the membrane, i.e. pump characteristics) of the sensor made correction for

hysteresis through data post processing possible. This is critical for a sensor operating on profiling platforms, especially in the Gulf of Alaska, where strong environmental gradients were encountered.

L315: Was it 1min running average or? Why was 1min was chosen? With a response time of ca 2 min, 1 min sampling resolution is not sufficient to apply RTC reliably, especially when resolving gradients. I think Miloshevish (2004) touches on that.

Thank you for this thoughtful comment. After carefully reviewing the Miloshevich et al. (2004) and Dølven et al. (2022) papers again, HydroC CO2 data were reprocessed accordingly. Instead of 1-minute averaging, smoothing was applied to keep the original 2 second HydroC resolution before RTC was applied. The RTC resolution was changed to 8 seconds following the L-curve analysis included in the RTC code by Dølven et al. (2022). Figures and related text were updated accordingly.

Correction: Lines 311-321

"Response times determined during calibration at -4H-JENA were used for response time correction (RTC) and found to be 106 seconds for the HydroC mounted on the rosette in May 2022 and 108 seconds when it was integrated into the Seaglider in February 2023 (HydroC CO2T-0422-001). The response time of the HydroC integrated into the Seaglider in May 2022 (HydroC CO2T-0718-001) was 109 seconds. Since field verification of the response time was recommended to ensure the highest quality post-processed data product, we verified the sensor response time at deployment. After the glider was stationary for approximately 15 minutes, a zeroing interval was performed with the HydroC CO_2. The response time was determined by reviewing the time it took for the signal to recover to the ambient concentration. Our in situ response time tests suggested to be within 5 seconds of the response time found during calibration (not shown). Before RTC was applied, HydroC CO_2 data were smoothed using a quadratic regression (MATLAB's smoothdata.m function with the loess method) over a 2 minute window. This was done to eliminate erroneous spikes in the RTC data and retain the original 2 second resolution of the pCO_2 data. The RTC resolution of 8 seconds was determined with the L-curve analysis included in the publicly available code from Dølven et al. (2022). The Dølven et al. (2022) RTC method produced more realistic profiles than an RTC method (Miloshevich et al., 2004, not shown) previously used for HydroC CO2 correction from a profiling float (Fielder et al. 2013), so we opted to use the Dølven et al. (2022) algorithm. In addition, Dølven et al. (2022) developed their algorithm with equilibrium-based sensors in mind and was proven with a sensor with a long response time (HydroC CH4 $\tau 63 \cong 23$ minutes). HydroC CO2 data were linearly interpolated onto the Seaglider timestamp and 1-meter binned data were calculated by first averaging 1 meter (+/- 0.5 m) upcast and downcast data independently, linearly interpolating

over gaps, then averaging the interpolated 1-meter binned upcast and downcast together."

L327-329: did the pumps both have the same rpms? This result is surprising. There should be no difference between 5T and 5M, only the pressure rating.
RPM for 5M is not stated on Seabird's documentation. Seabird states a 5M has a flow rate of 25 ml/s (https://www.seabird.com/modular/sbe-5m-mini-submersible-pump/family-cms.block?productCategoryId=54627473799) but this was not measured directly when powered by the Seaglider.

L359: Check Equation 1, must be % diff = (delta/pCO2 disc) *100%
Thank you for catching this! The denominator should be divided by two so it should read:
% diff = (delta/(pCO2HydroC - pCO2 disc)/2) *100%
The values are within ~0.1% using the equation you recommend.

L366: Sensor-pCO2 or discrete pCO2?
We will add "... *p*CO_2 measured with the HydroC …"

L375: The dashed black line shows the downcast data. I think you should be comparing both upcast and downcast to the discrete samples, especially given the fact that it seems to be an unresolved time lag if using lab-derived time constraints (see my comment to Fig. 8 below for some ideas). How far apart in time were glider upcast and downcast?
It is standard practice to give differences based on a ~1 meter depth averaged profile (e.g. Saba et al., 2019) and since we were outside the recommended distance for comparison (~100 m Thompson et al., 2021) we believe this is sufficient.

L386: It would be easier to read this paragraph if the May glider mission is compared to the May Sunny Cove data and the Feb glider data to the Feb buoy data first. The agreement is quite impressive, and you should emphasize this fact more clearly.
We rewrote to: "NOAA's moored sensor located in Sunny Cove (59.911 °N, -149.35 °W), near the CO_2 Seaglider trial site, measured an average surface *p*CO_2 of 240.7 +/- 10.4 µatm during the time of the May 2022 mission, which compares remarkably well with the glider based measurements. A minimum of 140 µatm was measured in Sunny Cove in mid-April (3-day average) (Figure 12, Monacci et al., 2023), suggesting that the peak of the spring bloom happened three weeks prior to our glider mission."

L415: What could be done to increase data accuracy further? Some recommendations and the vision forward would be useful for the reader.
We added: -4H-Jena is reassessing their sensor calibration methodology and data post-processing algorithm to further improve the HydroC's data accuracy."

L417: Not only the pCO2 gap but also carbonate system dynamics.
Was added to text.

L452: Could this pCO2/pCH4 sensor solution easily be integrated into other gliders (e.g. Slocums)? Would data processing be similar? Could other glider users easily replicate this integration?
This was already addressed in lines 429-437.

Figure 8: How was 1-meter binned calculated? It seems like RTC applied to downcast/upcast profiles using lab-derived time constants is insufficient. If you were to find an in situ time constant that would collapse the profiles on each other, that would be a really useful exercise.
We looked into this at first, similar to the methods used for thermal lag correction (Garau et al., 2011) but chose to use the response time correction based on the known sensor response time. This was done because, though the assumption of measuring the same water column on the downcast and upcast for shallow dives may be justifiable (we did not quantify this), we wanted our methods to translate to deep dives and missions. Additionally, using a published correction method with the provided t_63 of the sensor and pump, eases the burden of future glider users and data processors and was found to produce good results (see Tables 1-3).

We added the following to the Response Time Correction section: "HydroC data at the original resolution (2 s) and RTC resolution (8 s) were linearly interpolated onto the Seaglider timestamp and 1-meter binned data were calculated by first averaging 1 meter (+/- 0.5 m) upcast and downcast data independently, linearly interpolating over gaps, then averaging the interpolated 1-meter binned upcast and downcast together. "

L913: add legend to the plot
The different lines are described in the caption and do not need to be shown in a legend.

L930: add Legend to the plot

The different lines are described in the caption and do not need to be shown in a legend.

Figure 10: The glider dataset is indeed impressive; however, it is of no relevance to this paper. Therefore, please show only the period when Contros was on.
Since this paper also discusses scientific results and provides the first ever winter high resolution O2 and carbon measurements in Resurrection Bay, we believe that this figure and discussion of the full mission is valuable to the community studying this region.

Comments from Damian Arévalo-Martínez

Summary: The manuscript by Hauri et al. presents the development and comprehensive testing of a glider-based sensor package for measuring water column partial pressures of CO2 and CH4. This is a timely task because it directly addresses the need of increasing observations of the spatial and temporal variability of these major greenhouse gases.

Overall assessment: This contribution is significant for the field in that paves the way for large-scale surveys of greenhouse gases during process studies ranging from mid-water to the deep (1000 m) ocean, which in turn could be useful for the combined investigation of physical and chemical oceanographic variability. The manuscript is well written, figures and tables are mostly of adequate quality and the literature choice is appropriate. However, a major caveat I see in the study is the applicability of this approach given the different “readiness” level of the sensors for in-situ measurements of CO2 and CH4. While for the former the authors clearly show that the data obtained during both laboratory and at-sea conditions matches the desired accuracy and resolution (albeit uncertainties), for the latter both aspects rise questions on whether gliders are actually the right platform to be used (which is also an issue raised by Dr. Atamanchuk on her comment to this manuscript). The authors indicate that their pCH4 measurements have an uncertainty of +/- 2 µatm, which is problematic because it would not allow fully distinguishing between under- and supersaturated conditions in the water column. Although I would expect this to be a more serious issue in open rather than in near-coastal settings, I would still expect that in the latter part of the seasonal variability (and possibly a large fraction of the water column) could not be adequately resolved. Recent intercomparison efforts from different groups (see Wilson et al., 2020; https://doi.org/10.5194/bg17-5809-2020) came to the conclusion that although there is

no consensus on the threshold for “high-quality” seawater CH4 measurements, the achieved accuracy should be such that it allows tracking the ocean’s response to increasing tropospheric CH4 inn time scales of 5 years (which translates into an analytical agreement of <1 % between independent observations). I would argue that, at this point, this should be the target for marine water column measurements. Furthermore, the slow response time of the CH4 sensor poses a practical constraint for vertical profiling, at least if combined with the CO2 sensor. The CH4 sensor used in this study is therefore rather suitable for stand-alone (and potentially shorter) deployments in gliders, or long-term applications in, for instance, moored observatories where seasonal and longer time scales of variability are the target. I noticed that the description of the validation conducted by the authors is mostly centered around CO2 and my impression is that, at least in parts of the manuscript, the work done for CH4 was addressed somewhat qualitatively. Since the authors do have data to substantiate their thorough tests, I would kindly invite them to discuss the above mentioned caveats in their manuscript. In the following, I list general and specific comments not only to support my general assessment, but also in the hope that this is useful for the authors in view of a potential revision.

We thank both reviewers for their thoughtful comments on the methane glider. We revised the discussion paragraph on methane:

“The SG HydroC CH_4 was successfully integrated into the Seaglider as part of this project. While tank experiments showed promising results, short field tests of the CH_4 Seaglider in shallow water revealed low and patchy methane concentrations near the detection limit. The CH_4 Seaglider requires further testing in environments with strong pCH_4 gradients during longer and deeper dives (to allow for equilibration) to assess the accuracy of its response time-corrected data in the field. The sensor’s slow response time likely limits the glider to providing qualitative rather than quantitative results. However, due to the scarcity of oceanic CH_4 observations, deploying a CH_4 glider can help identify the location of methane sources and guide the placement of in situ observations to conduct a more quantitative assessment of CH_4 fluxes and dynamics.”

General comments: Introduction:

• The manuscript is strongly biased towards CO2 and this is reflected also in the introduction. If the platform is to be presented as relevant for both gases, the key processes for CH4 cycling in the ocean should at least be briefly mentioned. This is important because for a reader not familiar with trace gases, it will be hard to grasp where CH4 comes from in the ocean and why processes such as seepage and permafrost thawing (as mentioned by the authors) are so relevant.

We now start the introduction with:

“Understanding the distribution and dynamics of carbon dioxide (CO_2) and methane (CH_4) in the ocean is crucial for predicting and mitigating climate change and ocean acidification impacts.”

After a paragraph on CO2 we added:
“Over 100 years, CH_4 possesses a global warming potential approximately 28 times greater than that of CO_2 (IPCC AR5; Myhre et al., 2013). Sediments along the seafloor at continental margins contain large amounts of CH_4, with about ten times as much carbon than the atmosphere (Kessler, 2014). CH_4 is biologically produced in anoxic sediments and the surface mixed layer or released from geological sources like hydrocarbon seeps and degrading methane hydrate deposits (Barnes and Goldberg; Du et al, Skarke 2014). This powerful greenhouse gas is emitted to the atmosphere through bubbling (ebullition) or diffusive gas transfer (Reeburgh, 2007; McGinnis et al., 2006), which is limited by rapid oxidation to CO_2 during transport through the water column (Leonte et al., 2017). CH_4 occurs generally at low levels (background concentrations) throughout oceans, unless close to a source. Positive feedback mechanisms, like warming induced CH_4 seepage from destabilizing hydrates and thawing subsea permafrost, may further accelerate ocean acidification and climate change (Garcia-Tigreros et al., 2021; Sparrow et al., 2018; Shakhova et al., 2010; Rees et al., 2022).

• The motivation of the study (beyond the notorious technological advances) is not clearly specified. A line of thought which – in my opinion- would help introducing that motivation would be to first mention the extent of the ocean contribution to the natural sources/sinks of these two gases, see e.g.: Global Carbon Budget (https://doi.org/10.5194/essd-15-5301-2023) The global Methane Budget 2000 – 2017 (https://essd.copernicus.org/articles/12/1561/2020/) After which the issue of undersampling would underpin why extending the usage of autonomous platforms to climate-relevant trace gases is urgent. Overall, I would recommend revising the structure of this section to show more clearly why this is great progress that needs to be further developed and implemented. While I appreciate that the focus of the manuscript is more technical, doing this would increase the impact of this work.

We now start the intro with: “Understanding the distribution and dynamics of carbon dioxide (CO_2) and methane (CH_4) in the ocean is crucial for predicting and mitigating climate change and ocean acidification impacts.”

We believe that this was also addressed in the preprint with the following lines on CO2: “On top of this natural variability, the ocean has absorbed about one third of the CO_2 emitted by humans since the industrial revolution (Sabine et al., 2004; Gruber et al., 2019). In doing so, it has played an important role in mitigating climate change (Sabine and Tanhua, 2010).”

We also added a new paragraph on CH4:
"CH_4 possesses a global warming potential over a 100-year period that is approximately 28 times greater than that of CO_2 (IPCC AR5; Myhre et al., 2013). Sediments along the seafloor at continental margins contain large amounts of CH_4, with about ten times as much carbon than the atmosphere (Kessler, 2014). CH_4 is biologically produced in anoxic sediments and the surface mixed layer or released from geological sources like hydrocarbon seeps and degrading methane hydrate deposits (Barnes and Goldberg; Du et al, Skarke 2014). This powerful greenhouse gas is emitted to the atmosphere through bubbling (ebullition) or diffusive gas transfer (Reeburgh, 2007; McGinnis et al., 2006), which is limited by rapid oxidation to CO_2 during transport through the water column (Leonte et al., 2017). CH_4 occurs generally at low levels (background concentrations) throughout oceans, unless close to a source. Positive feedback mechanisms, like warming induced CH_4 seepage from destabilizing hydrates and thawing subsea permafrost, may further accelerate ocean acidification and climate change (Garcia-Tigreros et al., 2021; Sparrow et al., 2018; Shakhova et al., 2010; Rees et al., 2022)."

• In line 57 and following, the authors mention how key parameters are under-sampled. Most of the variables measured by the platform presented in this new application including gases are actually classified as essential ocean variables of the Global Ocean Observing System (see https://goosocean.org/what-we-do/framework/essentialocean-variables/). I would suggest the authors to mention this framework in their manuscript in order to emphasize how scientific and technical developments in this direction do address a timely task at an international level.

To address this and the next comment we rewrote this paragraph: "To effectively observe and understand the complex processes and feedback mechanisms regulating Earth's systems, certain key parameters, defined by the Global Ocean Observing System as essential ocean variables (EOVs,), must be measured accurately. However, these variables are often vastly undersampled across time and space due to traditional sampling methods, which rely mainly on discrete water sample collections from dedicated research cruises, underway measurements from transiting vessels, or time series measurements from in situ sensors on fixed moorings. Although biogeochemical sensors deployed on autonomous platforms like moorings and Argo floats have become more prevalent in recent years, challenges such as high power requirements, sensor size, and data quality hinder their widespread use on underwater gliders. Autonomous, spatially resolved surface measurements of pCO_2 and pH are collected using wave gliders and sail drones (Chavez et al., 2018; Nickford et al.,

2022; Manley and Willcox, 2010). The state-of-the-art biogeochemical (BGC) Argo floats measure variables like pH, O_2, NO_3, chlorophyll-a, suspended particles, and downwelling irradiance in subsurface waters (Claustre et al., 2020). These floats can last for several years at low sampling resolutions, such as a 2000-meter depth profile every ten days, or they can be programmed for high-resolution and shallow sampling. They can even sample beneath seasonal sea ice (Briggs et al., 2018). Despite their capabilities, their trajectory is hard to control, and they are usually not recovered after their mission, which prevents sensor calibration and post-mission corrections."

• With regards to undersampling, it would be important to mention that this does not only refer to spatial, but also temporal coverage. This would emphasize further the potential contribution of the approach presented in this manuscript, as the largest sampling deficits occur in winter.

We agree. We now write: "However, these variables are often vastly undersampled across time and space due to traditional sampling methods, which rely mainly on discrete water sample collections from dedicated research cruises, underway measurements from transiting vessels, or time series measurements from in situ sensors on fixed moorings."

• Since the authors aim to show the advantages of glider-based, large-scale measurements of CO2 and CH4, it would be good to provide the reader examples of the current approaches and their limitations (added as citations). Some suggestions (which include further examples therein) are as follows: Behncke et al (2024): A detectable change in the air-sea CO2 flux estimate from sailboat measurements, https://www.nature.com/articles/s41598-024-53159-0 Resplandy et al (2024): A Synthesis of Global Coastal Ocean Greenhouse Gas Fluxes, https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023GB007803
We believe that the section about wavegliders, saildrones, and BGC Argofloats addresses this comment.

Methods: • Some statements in this section (mostly regarding CH4) are vague or require additional explanation in order for the readers to be able to fully understand the methodological approach of the study (and ensure reproducibility).
We revised large sections of the methods- for example we added: " Discrete CH_4 were converted from the concentration of dissolved gas in water (mol L^{-1}) to partial pressure (µatm) using the solubility coefficient following Sarmiento and Gruber (2006)."

• As part of the validation discrete samples for both gases were collected and analysed. While it is clear that the glider sensors provide CH4 values as partial pressures, the

discrete samples were measured as dissolved gas concentration in an aqueous matrix. This means that partial pressures needed to be calculated. I would suggest the authors to clarify this aspect in the manuscript.

Rephrase 2.7.3 pCH4 post-processing

"SG HydroC *p*CH_4 data were response time corrected using a τ_{63} of 43 minutes (Dølven et al., 2022; Figure 4c). Discrete CH_4 were collected during the tank experiment (Figure 4c) and analyzed at John Kessler's laboratory at the University of Rochester. Discrete CH_4 were converted from the concentration of dissolved gas in water (mol L^{-1}) to partial pressure (µatm) using the solubility coefficient following Sarmiento and Gruber, 2006."

• In lines 301–302 the authors mention that the response times of their gas sensors do not have a temperature dependency. I think this is an important advance, which should be shown with data in the manuscript.

There are results from laboratory experiments that describe the (compared to PDMS) very low t_63(T_water) dependency for TOUGH membranes. But as T_water is just one parameter influencing t_63 it is more valuable to highlight that the HydroC CO2 sensor provides the opportunity to determine t_63 during the deployment.

We reworded: "For this study, the SG HydroC CO_2 sensors were deployed with the new robust TOUGH membrane, which had Teflon AF2400 as the active separation layer with a low temperature dependence on the permeability coefficient (Pinnau and Toy, 1996). The response time with the HydroC CO_2 TOUGH membrane is generally stable but can be affected by the speed of water exchange across the membrane (e.g. pump speed, tube length, etc.). Response times determined during calibration at -4H-JENA were used for response time correction (RTC) and found to be 106 seconds for the HydroC mounted on the rosette in May 2022 and 108 seconds when it was integrated into the Seaglider in February 2023 (HydroC CO2T-0422-001). The response time of the HydroC integrated into the Seaglider in May 2022 (HydroC CO2T-0718-001) was 109 seconds. Since field verification of the response time was recommended to ensure the highest quality post-processed data product, we verified the sensor response time at deployment. After the glider was stationary for approximately 15 minutes, a zeroing interval was performed with the HydroC CO_2. "

• I suggest the authors to make sure they write this section consistently in past tense.

We rewrote the methods section and used past tense wherever possible.

Results and discussion:

• Besides withstanding high pressure and probably the need of extended battery power, I could not not see why differences in the performance of the deep water glider are to be

expected (or it is at least not described in the manuscript). Even if there would be any changes, the tests shown on Figure 6 depicts a deployment that was even shallower than the first one, and therefore does not necessarily substantiates the author's argument.

Section 3.1 aims to convince the reader that the glider was able to fly well because of our careful trim and ballast efforts (to account for the large amount of mass and length added). The fact that we could even confirm stable flight in a shallow dive shows that the glider quickly reached a stable equilibrium. Critically, this means that the flow met by the sensors was undisturbed and data quality maximized.

We rephrased: "Example flight profiles with the POM and Titanium integrated sensors are shown in Figures 5 and 6, respectively."

• Section 3.2 (comparison with underway measurements): Neither during the methods, nor the results and discussion sections there is mention of how underway systems were used to cross-check the glider CO2 and CH4 sensors.

We rephrased: "The quality of the CO_2 Seaglider data was thoroughly tested with discrete measurements during a tank experiment and glider missions."

• Section 3.3.1 (tank experiments): the description provided here is rather qualitative. I recommend the authors to present the corresponding results on this validation for CH4, as they nicely did for CO2.

We wrote: "The SG HydroC CH4 was evaluated during the tank experiment described in section 2.4 (Figure 4c). Relative differences (Eq. 1) between discrete pCH_4 (average of triplicate samples) and pCH_4^{RTC} were 6.3 to 14.6 % (Table 1). During the experiment there was a decrease in salinity from 30.95 to 29.88 where $pCO2$ correspondingly decreased by 80 µatm. The corresponding pCH_4^{RTC} signal decreased by 25.4 µatm from 32.3 to 6.9 µatm. Although the triplicate discrete pCH4 water samples were slightly lower than the sensor measured pCH4 values, they also reflected this step change."

We also added pCH4 to Table 1.

Triplicate Date Time (UTC)	pCO_2^{disc} ± uc (µatm)	$pCO_{2,sn422}^{RTC}$ - pCO_2^{disc}	$pCO_{2,sn0718}^{RTC}$ - pCO_2^{disc}	pCH_4^{disc} ± u (µatm)	pCH_4^{RTC} - CH_4^{disc}
5/2/2022 3:25	298.7 ± 10.2	-0.9 (0.3%)	-	-	-
5/2/2022 7:32	227.1 ± 7.8	4.3 (1.9%)	2.4 (1.1%)	-	-
5/2/2022 11:27	223.3 ± 7.7	0.7 (0.3%)	-2.6 (1.2%)	-	-
5/2/2022 15:30	227.8 ± 7.9	-1.1 (0.5%)	-3.3 (1.5%)	-	-
5/2/2022 00:11	-	-	-	25.4 ± 2.1	4.0 (14.6%)
5/2/2022 12:06	-	-	-	7.3 ± 1.3	0.5 (6.3%)

Specific comments:
l. 21–22: “The key parameters to observing and understanding (…)”. Here I recommend revising the syntax.
We will leave as is since it is just a question of writing style.

l. 31: (…) provides (…). I would write “provided” instead.
Done.

l. 38–41: Spatial variability appears twice. I would check here to avoid redundance.
We rephrased to: ” Understanding the distribution and dynamics of carbon dioxide (CO_2) and methane (CH_4) in the ocean is crucial for predicting and mitigating climate change and ocean acidification impacts. Within the ocean, CO_2 levels (measured as the partial pressure of CO_2, pCO_2 and/or fugacity of CO_2) are spatially and temporally variable as they are influenced by a myriad of highly dynamic physical, chemical, and biological processes.”

l. 67: The abbreviation “BGC” should be inserted here.
Done.

l. 68: This is unclear. Of which variables? In the sentence above only pCO2 and pH are mentioned.
We rephrased to: "The state-of-the-art biogeochemical (BGC) Argo floats measure variables like pH, O_2, NO_3, chlorophyll-a,"

l. 91: Here it should be indicated what is the measurement standard (and/or reference) that substantiates this statement (e.g. accuracy, detection limits, etc.). Is this referred to Newton et al (2015) as included in the caption of figure 7?
We deleted this sentence because we couldn't find supporting literature. Thanks for pointing this out.

l. 128: "POM". This abbreviation should be spelled in full upon first usage.
Done.

l. 154: "headspace" instead of "head space"
Done.

l. 163: "percentage" instead of "%"
Done.

l. 223–224: Here the specifics of the CRM should be indicated (e.g. exact denomination, literature if any, etc).

We rephrased to: "This method requires Certified Reference Material (CRM, Batch #198 from A. Dickson's Certified Reference Materials Laboratory) to create a three-point calibration line.

l. 238: I would suggest "processing" as more appropriate than "manipulation" here.
Agreed.

l. 258,259: "CRM" instead of "Certified Reference Material" as the abbreviation was already defined
Done.

l. 292: "required parameter". This is rather vague. Here it should be stated which parameter was missing to be able to carry out the post-processing.
We rephrased to:"The *p*CO_2 data from February 2023 were not post-processed because a required parameter (*p_NDIR*) was not relayed in real-time…"

l. 450–451: Considering the comparatively low accuracy and long response time, I would have to disagree, unless it is clearly stated for which types of studies this is the case (see also overall assessment above).
We rephrased to: "CO_2 gliders are perfectly suited to contribute data for understanding relevant inorganic carbon processes in coastal shelf and boundary regions where mesoscale or sub-mesoscales dominate."

The flaws of CH4 gliders are now discuss in the paragraph just above:
"**While tank experiments showed promising results, short field tests of the CH_4 Seaglider in shallow water revealed low and patchy methane concentrations near the detection limit. The CH_4 Seaglider requires further testing in environments with strong *p*CH_4 gradients during longer and deeper dives (to allow for equilibration) to assess the accuracy of its response time-corrected data in the field. The sensor's slow response time also limits the glider to providing qualitative rather than quantitative results. However, due to the scarcity of oceanic CH_4 observations, deploying a CH_4 glider can help identify the location of methane sources and guide the placement of in situ observations to conduct a more quantitative assessment of CH_4 fluxes and dynamics."**

l. 492: There should be a link in the revised version.
The new data accessibility section now reads:

"The CO_2 Seaglider data is publicly available (Hauri et al., 2022; 2023). The HydroC-specific SIRMA code and CNF file are available on Github (Cyprus-Subsea, 2024a and 2024b). More detailed information on the HydroC – glider integration and operation can be found in the CO_2 Seaglider Standard Operating Procedures (Irving et al., 2024).