Response to reviewer 2

In this study, Linsy and co-authors perform a very thorough experimental study to assess the biogeochemical effects of sediment disturbance by bottom trawling in the Baltic Sea, where the authors provide valuable new evidence highlighting the complexity of such disturbance on biogeochemical pathways.

Notably, this is the first study that assesses the impacts of bottom trawling on total alkalinity fluxes from an experimental perspective rather than relying solely on modelling. It also stands out for addressing the effects of demersal fisheries from different perspectives, providing a deeper understanding of the biogeochemical consequences of demersal fisheries. By performing obtaining different sampling types (CTDs, sediment cores, landers) and analyzing a wide range of parameters, the authors provide a deeper understanding of the different biogeochemical processes affected by this sediment disturbance, while also recognizing the limitations of their approach.

I thoroughly enjoyed reading this in-depth study, and commend the authors for the work behind this. While the manuscript is both timely and highly relevant, I do have a number of comments and questions – particularly regarding the experimental design, data analysis, and its interpretation – which I hope will help strengthen the overall clarity and impact of this study.

We thank you, Sarah, for taking the time to thoroughly review our manuscript. We greatly appreciate your constructive feedback, which has helped us improve the clarity and quality of our work. Please find our detailed responses to each of your comments below.

Main comments:

1. While the description of the methodology is very detailed and could serve as a guideline for future studies that aim to better understand the biogeochemical impacts of demersal fisheries given its broad scope, it is not clear to me what kind of experimental design this study is following. I had to re-read the methods to properly identify if it was a Control-Impact experimental design, or a Before-After Control-Impact experimental design (sample all sites before the disturbance to account for temporal and site variability). I initially thought it was a Control-Impact experimental design, but when looking more closely at Table S1, I noticed that the authors also sampled the impact site before (July 19) the experimental trawl (July 20), sort of making it a BACI experimental design (only sampled the impact site before disturbance). The authors should be clearer about this experimental study design.

   Being a BACI experimental design, the authors should perform statistical analyses not only comparing Control-Impact, but also prior to the disturbance. In addition, the continuous sampling 16 days after the disturbance to assess the recovery is done in comparison to the control site, but it should also be done in comparison to pre-disturbance.

This also raises concerns with the statistical analysis used. From my understanding, the authors combine the sediment profiles of the cores (Figs. 5-6), or the fluxes (Figs. 7) in the impact and control sites. However, since the data was collected in different periods with respect to the disturbance, which the authors plot in Fig. 8 to assess the recovery after disturbance, then combining the data assumes that the temporal variation is not relevant. A quick look at the raw data in Table S1 doesn't show differences in the fluxes of the MUC in the impact site prior to the disturbance in comparison to the fluxes after the disturbance.

The following comments assume that the data processing was correctly done, but this should definitely be looked into.

Thank you for bringing this to our attention. Our study follows a control-impact experimental design. The two cores collected before the trawling event in the HI area are considered control samples, as now indicated in both the text and the table. However, performing statistical analysis on porewater fluxes based on only two data points is not advisable (BACI approach). Additionally, we did not have any BIGO deployments prior to the trawling experiment at these specific sites. We have now updated Figures 5 and 6 to include these samples within the control site category. As a result, the mean fluxes and porewater concentrations have changed slightly, but not substantially, as to alter the conclusions of the study.

Regards Fig. 8, the temporal changes are indeed included in the mean values. These are, of course, important. Yet, since we are comparing the CL and HI areas, both of which are undergoing short-term temporal changes, our results nonetheless show that the difference due to trawling is larger than those caused by natural temporal variability in the CL areas.

2. In relation to the statistical analyses (section 2.9), the authors consider that a p-value < 0.05 is indicative of statistically significant variability. After resolving the issue of my previous comment, I suggest the authors be more precise about the statistical significance of their results, and provide more detail about their statistical significance. For instance, the authors could add an asterisk in the figures to denote statistically significant differences with different confidence values, such as * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.05$ (or similar notation). I am surprised that there are statistically significant differences in the POC content of the surface sediment layers (Fig. 5a) considering that the limits of the error bars are touching (1 standard deviation, equivalent to 66 % of the variation of both samples). I have the similar doubts with the boxplots of TOU, TA, ammonium and phosphate of Fig. 7, since the upper and lower quartiles of the control and impact sites cross each other.

   Thank you for this useful comment. We have now clearly stated the significance thresholds in the Methods section (*$p < 0.05$, **$p < 0.01$, *** $p < 0.001$) and consistently indicated the significance of p-values in the Results section. The

POC values show a statistically significant difference in the surface (0–1.5 cm) section, with p-values of 0.046 and 0.036, respectively. The total fluxes of TA, ammonium, and phosphate also show significant differences between the control and HI treatments. TOU does not show a statistically significant difference. Regarding diffusive fluxes, TA is not significantly different, whereas ammonium (P value - 0.02) shows a significant difference even though the boxes overlap slightly.

3. I am also missing some more background information of the study area, more specifically in relation to the fishing history. As pointed out in a data compilation of studies assessing the biogeochemical impacts of demersal fisheries, Paradis et al. (2024) identified that the control site in the majority of studies have been historically fished and were not being trawled during the study due to a seasonal closure or the recent establishment of a trawl ban. I am aware that the Baltic Sea has been extensively impacted one way or another (HELCOM, 2018; Bradshaw et al., 2024; Díaz-Mendoza et al., 2025), so what is the fishing history and current condition of the study area?

   In addition, how does the experimental fishing conducted in this study compare to the bottom trawling activities that usually take place in the Baltic Sea in terms of gear type, fishing intensity, fishing season? This is especially important to clarify and apply for the last modelling exercise (see comment 9). What is the distance between trawl tracks (red lines in Fig. 1), what is the width of the trawl nets and sweeplines? The authors have a schematic diagram of the gear type used in Figure 4, but this one is too small to annotate these elements (e..g, width between otter doors). This would be especially beneficial considering that the authors target the wide area between the otter boards (it would give additional perspective of why they target this area and not the furrows caused by the heavy otter doors).

   The study area lies within the 3-nautical-mile zone, where fishing activities require special permission. Prior to the experiment, a detailed bathymetric survey was conducted, revealing no visible fishing tracks or seabed disturbances. However, according to HELCOM 2021 data, the broader Mecklenburger Bight experiences a maximum trawling intensity of 2 yr^{-1} with 38% of the seafloor affected. We agree that the control area is defined as "control" only with regard to our experiment and is not intended to infer that the control area has never been trawled.
   It should be noted that this work represents a case study, and direct comparison with the entire Baltic Sea fishery is challenging, as different types of fishing gears are employed across the region. For this study, we used a standardized bottom trawl commonly applied in Baltic demersal surveys (TV-3#520 × 80 mesh size). The gear was fitted with ThyborØn Type 2 Standard trawl doors, each with a surface area of 1.78 m². The distance between the otter boards on either side of the trawl net was approximately 60 m. The length of the sweep line was 75 m. Additional technical specifications of the trawl configuration are provided in the ICES Baltic International Trawl Surveys (BITS) Manual. We have now updated

the manuscript to include this information and added a schematic diagram of the otter trawl. The sampling position was determined based on the trawl track recorded by the multibeam bathymetry.

4. The contact of demersal fishing gear with the seafloor has several effects: it can resuspend sediment and hence erode the seafloor, create furrows associated to this sediment resuspension and erosion and adjacent sediment piles, and/or mix the sediment. The magnitude of each of these impacts is difficult to quantify, and the effects on sediment biogeochemistry will differ depending on these processes.

   The authors observe a combination of these processes in this study: defined furrows, sediment piles next to the trawl tracks, visible scrapes, and a sediment plume implying sediment resuspension (lines 392-399; Fig. 4). What were the sizes (width and depth) of each of these features?

   The depth of the furrows ranged between 10–12 cm, and the width of a single trawl mark was approximately 85 cm. The precise physical impact of this disturbance is currently being prepared in a forthcoming manuscript.

   Later on, the authors conclude that bottom trawling has removed the upper 2 cm of sediment since there are statistically significant differences of POC in the impact and control site in these surficial sections (lines 414-416, Fig. 5). However, this reduction could also be due to remineralization, or mixing of the high OC in surficial 0-1 cm with the lower OC in deeper sediment sections (affecting the shape of the POC profile). Morys et al. (2021) observe that there is an upward 2.5 cm shift in the profiles of Chl-a, OM, and water content in the impact site (IN) with respect to the control site (OUT), which they attribute to erosion of these 2.5 cm. In this study, you also determined the water content and porosity of the sediment cores. This metric could be used to determine mixing (constant porosity in mixed layers) as well as erosion (removal of the less-consolidated surface sediment as seen by Morys et al. (2021)).The different physical effect of bottom trawling (mixing and erosion) and its biogeochemical effects should be discussed in more detail. For instance, the authors relate the lower flux of nutrients in the impact site due to erosion, but it could also be caused by mixing, which would accelerate the diffusion of porewaters to the overlying water.

   Thank you for your insightful comment. In line with the observations reported by Morys et al. (2021), we observed an upward shift of approximately 2 cm in the porosity profile at the HI site, strongly indicating erosion of the upper sediment layer. We have now incorporated this into Figure 5 of the revised manuscript to support our interpretation.

   Although bottom trawling can also lead to mixing of surface sediments, our data more strongly support erosion as the dominant process. While some degree of mixing may occur, it does not appear to be the primary factor influencing nutrient fluxes in this study. If mixing were the main mechanism, we would expect an increase in total solute fluxes measured by the lander, as mixing enhances the release of porewater solutes into the overlying water. However,

the observed reduction in total fluxes, particularly at the HI site, is more consistent with the removal of surface sediments rather than enhanced mixing. The BIGO measurements, which capture both diffusive and non-diffusive fluxes, further support our interpretation by providing a more comprehensive assessment of sediment–water exchange. Additionally, Figure 4 clearly shows the net area (area affected by back strop, sweep lines, chains and bridles as well as by the foot rope and the fishing line), including the lines caused by the back strop as well as a thin layer of brown phytodetritus and underlying dark grey anoxic sediments. This suggests that the upper sediment layer has been removed. Further details on this observation will be provided in the aforementioned forthcoming manuscript.

5. In this study, the authors identify that sediment disturbance has minimal effects on TOU, in comparison to lowered O_2 consumption rates in other studies (Tiano et al., 2019; Bradshaw et al., 2024), implying that it is the first time this has been observed. However, as portrayed in a recent compilation of biogeochemical studies of the impacts of demersal fisheries (Paradis et al., 2024), there are several other studies that have shown that there is a minimal effect of demersal fisheries in oxygen consumption (Warnken et al., 2003; Polymenakou et al., 2005; Trimmer et al., 2005; Goldberg et al., 2014; Meseck et al., 2014). These studies were done in continental margins with contrasting dissolved oxygen concentrations, so the relationship between TOU and BW oxygen observed in this study are not necessarily applicable to those other studies.

   Thank you for making us aware of your manuscript. The effect of trawling on TOU appears minimal in our study, consistent with findings from previous research. While we are not claiming this as a novel observation, we have now added a sentence in the revised text acknowledging that similar results have been reported in earlier studies, and we have cited the relevant literature accordingly.

6. When discussing the mechanisms affecting the fluxes of nutrients, the authors discuss that phosphate could be released to the water column after disturbance, leading to a lower flux, but this is not the case for nitrate fluxes, since the fluxes of this latter nutrient did not vary between the control and impact sites (Fig. 7). If phosphate is released to the water column, nitrate should have been released as well (Breimann et al., 2022). Maybe the lack of significant difference of nitrate flux between the control and impact sites could be due to the counterbalance of nutrient release (as suggested for a decrease in phosphate flux) and decreased denitrification rates as reported in other studies (Ferguson et al., 2020).

   Good suggestion that we have now included in the text.

7. The authors find that in the impact sites, fluxes of DIC and TA decrease. They then mention that the biogeochemical explanation for this decrease is unclear (lines 610-612). Isn't it simply because there is a decrease in the RPOC? (Fig. 7h).

Also, the fluxes of TA in this study are within the range of TA fluxes in other regions (see lines 615-616). Hence, is there really an impact of bottom trawling in terms of TA flux?

Yes, the reduction in TA and DIC may be attributed to a decrease in the rate of POC remineralization and calcite dissolution. Accordingly, we have removed the sentence stating that "the biogeochemical explanation for this decrease is unclear." The revised text now reads: "The observed reduction in TA and DIC in the present study may be attributed to decreased remineralization of POC and calcite dissolution (Fig. 7)." Although TA values in the present study are comparable to those reported from other regions, this does not imply that bottom trawling has no impact on TA fluxes. When compared to the control site, a clear reduction is evident.

8. This is the first study that looks at the effects of bottom trawling in alkalinity fluxes. Van de Velde et al. (2025) performed a global modelling study of the effects of sediment disturbance on benthic alkalinity fluxes, and the causes behind it (carbonate dissolution, sulfate reduction and pyrite burial, denitrification). How do your observed results compare to those seen in that study? In that modelling study, they observe that the majority of the alkalinity reduction is due to changes in sulfate reduction and pyrite burial, but in your study, there are no statistically significant changes in sulfate reduction nor pyrite content. However, you mention that "sediments transition from a TA source to a large TA sink during the first trawl due to pyrite oxidation" (lines 658-659). I also don't agree with this sentence, since the sediments do not transition to a TA sink. Their TA flux is simply reduced in comparison to the control sites, so they are simply a "less strong TA source", to put it one way. What about carbonate dissolution? You calculate the carbonate dissolution rates in the control site, but what about the impact site? What about denitrification? See earlier comment 6.

   Our results for calcite dissolution are near-identical to the global model analysis of coastal muds by van de Velde et al. (2025), which is now included in the paper. Pyrite burial in our model equaled 1.5 mmol m^{-2} d^{-1}; again, very similar to the global study (note that we corrected a typo in Eq. 8 whereby the new calcite dissolution rate is 9.4 mmol m^{-2} d^{-1}compared to the previously reported value of 9 mmol m^{-2} d^{-1}). We believe it is correct to say that with trawling the sediment transitions to a TA sink because of the impact of proton release due to pyrite oxidation. The way the model is configured, pyrite oxidation represents a negative alkalinity source to the water column. We do not calculate carbonate dissolution for the HI site since here we are interested in the reduction in TA and DIC following trawling, which we ascribed to a reduction in POC degradation and carbonate dissolution (now mentioned). Denitrification (i.e. NO3 flux) is included in the mass balance (Eq. 8) to calculate carbonate dissolution.

9. Regarding the seafloor-water-air box model of the impacts of DIC and TA fluxes, I don't understand the point of doing a trawling disturbance with and without changes

in DIC and TA fluxes. The experimental data show a reduction of DIC and TA fluxes, so why make a scenario called “No impact” but still with a trawling disturbance event?

This was misleading. We have now renamed the “no impact” run to “No directly imposed TA reduction”.

The experimental approach was performed in summer, and the parameters observed during that season were applied in the box model for several years. Wouldn’t the conditions change over time? Wouldn’t the “baseline” CO_2 and TA fluxes also change seasonally?

Most likely yes, but unfortunately, we only have data from one time point.

Finally, is this fishing intensity (once per year) representative for this study area?

Following Kalapurakkal et al. (2025), we have now obtained trawling frequency (2 yr-1) and disturbed areas (38%) from the ICES Vessel Monitoring by Satellite (VMS) database.

10. Regarding future work, the authors modestly acknowledge in different points of the manuscript that they are not capable of discerning the causes behind the trends they see, and that they would need more information. What kind of information would they need to properly understand the causes behind the trends observed? For instance, the authors discuss why they don’t see changes in nitrate flux in comparison to other studies, or decreases in phosphate, DIC and TA fluxes, and mention that they would need more information.

    The authors also acknowledge the need to study the fate of resuspended sediment to get a broader understanding of the biogeochemical consequences of sediment disturbance. Another aspect that should be studied is the effect of repetitive bottom trawling activities. Depending on the region, fishing grounds are disrupted almost on a daily basis, which would limit the capacity of deploying landers to properly calculate porewater fluxes, especially since the authors mention that porewater fluxes could be flawed (lines 489-500). This adds to the complexity of studying the biogeochemical impacts of demersal fisheries.

    These are good points. More data for this kind of study is always welcome. Carefully controlled laboratory incubations to investigate Fe and P cycling in the resuspended layer is one potential research avenue that we are considering. We agree that repetitive trawling makes it problematic to properly assess the trawling impact on benthic fluxes. Recently, MPAs have come into force in the Baltic Sea, and continuous monitoring is now needed to compare future undisturbed fluxes with those in the same area that have been previously measured in our project. However, the temporal and spatial variability of the

fluxes within the relatively small control area complicates the characterization of an undisturbed baseline. We have now added this caveat to the conclusions.

Minor comments:

- Line 22. The authors refer to studying the impacts on the "benthic ecosystem", which includes benthic communities (e.g., meiofauna). However, they did not study the benthic community. I would suggest to replace "benthic ecosystem" to "benthic biogeochemical pathways" or something similar.

  We have replaced the benthic ecosystem with benthic biogeochemical pathways.

- Line 28. Define which nutrients, since not all nutrients showed a decrease in flux.

  Thank you for spotting this. Now we have mentioned that nutrients (PO_4^{3-}, NH_4^+, and H_4SiO_4) showed a reduction.

- Line 29. Change "variations" to "decreases"

  We have modified the text.

- Line 32. Convoluted sentence. Suggest to modify to "[…] had not returned to baseline levels by the conclusion of the 16-day observation period, indicating prolonged effects of the disturbance, although natural temporal variations may have an influence." Or something similar.

  We have revised the text as suggested.

- Fig. 1. Add location of CTD deployment(s).

  We have modified the figure to include the CTDs.

- Lines 145-155. The authors extract porewater using two different approaches: centrifuge and Rhizon samplers. How do the results vary between both methods?

  We employed two different methods due to a centrifuge malfunction that occurred midway through the cruise, and therefore could not compare the two methods. To minimize variability between the methods, instead of inserting Rhizons directly into the intact sediment cores, which is known to have artifacts due to the extraction of porewater from adjacent sediment layers, we first sliced the cores as usual and transferred the sediment into centrifuge tubes before inserting the Rhizons. While we cannot definitively quantify how this procedural difference may have influenced the results, we assume that its impact is minor.

- Line 221. Were these geochemical analyses performed on the same sample (the same 1 cm interval of the same core)? I'm asking this because some samples were treated (e.g., with ascorbic acid) and some were left untreated. Hence, aliquots would have had to be subsampled for each analysis, and I'm surprised you would have gotten sufficient volume for all of these analyses. Please clarify.

Yes, all analyses were performed on the same set of samples. Approximately 10 mL of porewater was extracted from each sample and used for all measurements.

- Lines 325-336. The diffusion fluxes were obtained from fitting the porewater data using the FindFit function in Mathematica software. As I'm not familiar with this software, and other readers may not be either, is this fitting done weighing the uncertainties of the measurement? Does this fitting give you a measure of uncertainty within a specific confidence interval? Was it error-propagated?

  Thank you for this comment. The FindFit function in Mathematica was used solely to obtain a best-fit line through the porewater concentration data in the upper sediment layer. This approach allowed us to calculate a more representative concentration gradient for estimating diffusive fluxes, rather than relying on a simple linear gradient between the 0–1 cm depth interval. The fitting was not weighted by measurement uncertainties, and it does not provide confidence intervals or error propagation. Our primary aim was to extract a reliable slope from the data for flux calculation, rather than conducting a full statistical analysis of the fit.

- Figure 2. Instead of plotting wind direction on a y axis that goes from 0 to 359 (this value is cyclical), it should be plotted as an "arrow graph" (see example in Fig. 3 of Puig et al. (2003)). Another alternative would be to choose a cyclic colormap for the wind direction and use it on wind speed data (Fig. 2b).

  In both Fig. 2d and e, there seems to be artifacts in the data (spikes of SST, BW Temp, and BW O_2). The O2 concentration obtained from Winkler method in Fig. 2e is not sufficiently visible.

  We have updated the figure by adding an arrow graph to indicate both wind direction and wind speed. The previously visible spikes were not artifacts but represented the lowering and retrieval of the ROVER. These have now been removed from the figure, making the Winkler-analysis-derived O_2 concentration more clearly visible.

- Finally, to give Fig. 3 a bit more context, consider adding an arrow (or similar marker) above Fig. 2 for each CTD deployment. That way, the reader will not have to find the environmental conditions during each CTD deployment in Fig. 2.

  We have modified the figure and added an arrow for CTD deployment.

- Lines 410-416. Add a more detailed description of the differences, or lack of, of POC, PON and CaCO3, as done in lines 417-418 for pyrite contents.

  We have added the surface values and described the variation between the control (CL) and HI sites. The following sentence has been included in the revised manuscript: 'The surface concentrations of POC, PON, and $CaCO_3$ were consistently higher at the CL site (POC: 3.58 ± 0.74 wt.%; PON: 0.54 ± 0.11 wt.%; $CaCO_3$: 5.39 ± 1.12 wt.%) compared to the HI site (POC: 2.77 ± 0.78 wt.%; PON: 0.41 ± 0.12 wt.%; $CaCO_3$: 3.55 ± 0.48 wt.%).

- Line 431. Depth-integrated SRR at the control and impact sites should have a measure of uncertainty, no?

  Yes, the values include uncertainty. We have revised the text to reflect this, now reporting the depth-integrated SRR as 3.9 ± 3.0 and 4.9 ± 2.9 mmol S m^{-2} d^{-1} at the CL and HI sites, respectively.

- Line 468. Remove "fluxes of" in "showing the fluxes of solutes fluxes across"

  We have modified the figure caption.

- Line 490. The authors mention that the two approaches have "slight differences in magnitude for TA and NH4+". Are these differences statistically significant? And the lower H4SiO4 fluxes of BIGO, were they significantly lower?

  In the case of TA and $NH4^+$, the differences are not statistically significant, whereas for silicate, the difference is statistically significant. Now included in the manuscript.

- Line 563. Add "it" in "[…] since it will affect the rates of […]"

  We have corrected the sentence.

- Line 577. Remove the comma after "as well as, the local"

  Corrected

- Line 672. The surface pyrite content in the impact and control site are not statistically significantly different.

  Correct, but as model results show, only a few percent of ambient pyrite pool needs to be oxidized to account for the TA losses. Now clarified in the text.

- Line 676. Sentence is missing a verb

  Corrected.