Response to reviewer 1

The authors have addressed most of my concerns, and I am happy to see it published after consideration of a last few comments, listed below.

Thank you for your positive feedback, Sebastiaan. Please find our detailed response below.

1. The use of Bottom trawling versus mobile-bottom contact fishing. If you acknowledge that it is the more precise term, why would you not just use it? Bottom trawling might be used in other papers, but that is not a reason not to use the correct term.
   Thank you, we have corrected the terminology throughout.

2. The use of 'severe' in the title. Severe is a subjective term, and your paper concludes the impact of the fluxes is not that important. I would urge to remove the term 'severe'.
   We have modified the title.

3. L36: I would add 'potential' to this sentence, as you assumed a rate from a different experiment. So, you have no evidence to say that pyrite oxidation is definitely important at your site.
   Added.

4. The explanation of the model setup:
   "In this study, the model was modified in two ways; (i) to include constant values of carbonate dissolution at the derived rates of 9 mmol TA m-2 d-1 and 4.5 mmol DIC m-2 d-1, and (ii) during each trawling event, the benthic DIC and TA fluxes were reduced by 9.2 and 6.7 mol m-2 d-1 in accordance with the data, and for a period of 21 days."
   This does not make sense to me – the TA flux from the sediment is a consequence of carbonate dissolution. So, do you mean you have a baseline flux of 9 mmol TA m-2 d-1, and reduce this by 6.7 mmol m-2 d-1 (not mol m-2 d-1 I hope?)?
   If this is the case, please rephrase the description.
   If not, then the model setup is probably incorrect and needs to be reconsidered

   Thanks for noting the typo in our manuscript (line 695, missing m in front of mol). We corrected this typo and now give the decrease in the TA flux during trawling events in correct units as 6.7 mmol m^{-2} d^{-1}. We also expanded our model description (line 692 – 695) as follows: ". Benthic DIC fluxes in the original model were exclusively due to POC degradation, whereas TA fluxes were driven by pyrite burial and oxidation. In this study, the model was modified in two ways; (i) to include baseline values of carbonate dissolution at the derived rates of 9.4 mmol TA m^{-2} d^{-1} and 4.7 mmol DIC m^{-2} d^{-1}, and (ii) during each trawling event, the benthic DIC and TA fluxes were reduced from the baseline values by 9.2 and 6.8 mmol m^{-2} d^{-1} in accordance with the flux data in Fig. 7, and for a period of 21 days. With this approach, we consider that carbon degradation and carbonate dissolution were

diminished since most of the degradation and dissolution takes place in surface sediments that were partly removed by trawling."

Response to reviewer 2

I thank the authors for taking their time to address all of my comments, which I admit were many. The impressive number of samples and analyses performed for this study is a step forward to understand the biogeochemical effects of mobile demersal fisheries, which is why I have thoroughly revised the response to my first review. I believe that there are still several aspects that are not quite clear and would like the authors to further clarify and address in their manuscript before publication.

Thank you, Sarah, for reviewing the manuscript again and providing your comments. Please find our detailed responses to each of them below.

I have separated the different comments based on the same numbering I employed in the first review, in order to avoid repetition and to indicate what comment and reply I am referring to. In this review, I only address the aspects that I believe the authors should further clarify. In my opinion, all the other several comments have been addressed.

1. Type of experimental design
   In my first review, I asked the authors to clarify what type of experimental design they followed, since it was hard to identify based on the methods and results of the manuscript:
   1. Control-Impact (CI), when an experimentally trawled site is compared with a control site
   2. Before-After (BA), when the same site is sampled before the disturbance and after the disturbance
   3. Before-After-Control-Impact (BACI), when (at least) one site is sampled before and after the disturbance (impact site), and (at least) another site is sampled at the same time (before and after) but is not disturbed. This approach also includes collecting additional samples in time after the disturbance to get a better temporal variability. Clarifying and identifying what type of experimental design this study followed is crucial to then know what type of statistical analysis to employ.
   The authors clarify that their study follow a CI approach, but then mention that they collected samples in the impact site before the disturbance. Hence, it should be either a BA or a BACI approach, not a CI approach.
   In their statistical analysis, the authors aggregate all their "control" (before impact?) and "impact" (after impact?) sites to assess if there are any statistically significant differences in the different parameters they studied. This aggregation is done with samples collected in different periods. The authors argue that they can aggregate this temporal data because "their results […] show that the difference due to trawling is larger than those caused by natural temporal variability in the CL areas". What statistical analyses did the authors do to arrive to these conclusions? This should be clarified.
   I would first perform a linear mixed-effects model that accounts for the disturbance (impact/control) time, and the interaction of both. This would allow the authors to properly

assess if natural temporal variability is relevant. If the results of this analysis show that temporality is not relevant, then it is justifiable to average the data in the figures. If the results of this analysis show that temporality is actually relevant, then the authors cannot average the data in the figures. Moreover, this would imply that temporality should be addressed when assessing the biogeochemical impacts of demersal fisheries, which is seldomly done.
The interpretations and conclusions of this manuscript depend on this, which is why I insist on a proper statistical analysis given the large number of samples collected.

We apologize for the confusion. Our study is designed as a control–impact experiment, in which one area was trawled (HI) and the other untrawled (control). Samples were collected from both areas during the study period to assess the impact relative to the control. As mentioned, two MUCs were taken from the HI area before trawling. Two data points does not adequately represent the pre-disturbance condition required for a Before–After (BA) design. For this reason, we stated that the two cores (MUC1 and MUC2) collected from the HI area prior to the trawling experiment were considered control samples.

Regarding the statistical analysis of temporal variability, we had not previously conducted such an analysis. As suggested, we performed a linear mixed-effects model analysis, and the results indicate that TOU, TA, and DIC exhibited temporal variability. In contrast, the remaining parameters (NH_4, PO_4, $SiOH_4$, NO_3, and NO_2) did not show significant temporal variability. Please note that the data points shown in the figure (Fig. 8) are not averages but individual measurements.

2. Statistical analyses. Assuming that the statistical analysis employed is correct, the authors mention that TOU and TA are not significantly different among the impact and control sites. However, in several occasions in both the manuscript and the reply to my reviews, the authors then argue that there is a decrease in the TA flux in the HI site – this decrease is not statistically significant, right? Please be consistent. This is especially important in this study given the large number of proxies measured.

   We estimated fluxes in two ways: one using landers (total fluxes) and the other based on porewater/diffusive fluxes. The lander fluxes showed significant variability between the control and HI areas, whereas the diffusive flux of TA was not significantly different. For the box model, we only considered the total fluxes measured by the lander, as these results are statistically significant and total fluxes best represent the natural net flux. This was stated in the manuscript: *'Total fluxes of TA (p value), DIC (*p value), NH_4^+ (*p value), PO_4^{3-} (*p value), and H_4SiO_4 (**p value) exhibited statistically significant differences between the HI and CL areas.'**

   8. Alkalinity fluxes and their reasons

First of all, I want to re-iterate that according to the statistical analysis performed (which should also be revised), there is no statistical difference in the TA fluxes between the HI and CI sites (see my comments above). However, the authors go in great detail to explain that sediment disturbance has an effect on TA fluxes. Assuming that this is a relevant process, the authors point that this reduction in TA fluxes is due to changes in the sulfate reduction. However, in my previous review I noted that the authors identified that there are no statistically significant changes in sulfate reduction or pyrite content. Isn't this contradictory? Please clarify. In addition, in my previous review, I asked about the influence of carbonate dissolution. To this, the authors replied: "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." I had to re-read these sentences several times since I was confused in the contradiction, which I point below to clarify: - You don't calculate carbonate dissolution for the HI site because you associate the reduction in TA and DIC due to carbonate dissolution? So, carbonate dissolution should be calculated, no? - You include denitrification in the mass balance to calculate carbonate dissolution. So, is carbonate dissolution calculated? As you can see, it is not clear to me whether carbonate dissolution is calculated. If so, what are the values, and is it relevant? In their reply to my comment 7, the authors mention "Yes, the reduction in TA and DIC may be attributed to a decrease in the rate of POC remineralization and calcite dissolution." As you can see, this is not at all clear to me. Again, all of this is assuming that the changes in TA is statistically significant, which the authors previously say it is not.

As noted above, total TA fluxes were significantly different between control and high-impact areas. We did not state that the reduction in TA fluxes is due to changes in sulfate reduction. Regarding carbonate dissolution, we calculated it for the CL area using the TA mass balance in Eq. 8, to provide a baseline estimate of this process for the purpose of including it in the box model. We cannot do this for the HI sites since the sediments were disturbed. Denitrification is included in the mass balance, along with the ammonium flux and pyrite burial. The net result of these fluxes gives carbonate dissolution (RCaDiss, Eq. 8). In the text after Eq. 8, the dissolution rate is provided.

9. Seafloor-water-air box model I still find confusing adding two kind of disturbance (with and without changes in DIC and TA fluxes). A more thorough explanation is needed. This is beyond my area of expertise, and I believe other readers would like to understand this better.

In our experimental study, we observed a reduction in TA and DIC fluxes as a result of trawling (comparing CL and HI areas). To assess whether these reductions influence air–sea CO_2 exchange, we applied a previously established model developed by Kalappurakkal et al. (2025) that did not include carbonate dissolution (not clarified). We adapted this model by incorporating our baseline carbonate dissolution rates (i.e., using

data from the control area, Eq. 8). In one simulation run ("standard run"), we reduced the TA and DIC fluxes according to our flux measurements (Fig. 7). In a second run ("No directly imposed TA reduction"), the reduction in TA and DIC fluxes due to trawling was ignored. By comparing these two simulations, the impact of TA and DIC flux reduction on CO_2 exchange can be evaluated.

Lines 325-336. Calculation of diffusion fluxes using FindFit function. The authors should use a fitting that provides measures of uncertainties within a specific confidence interval to be able to compare the diffusive fluxes across sites and see if they are significant while accounting for their confidence interval.

We appreciate the suggestion regarding the use of fitting methods that provide uncertainties within a specific confidence interval for diffusive flux estimates. We recognize the importance of illustrating variability across sites, and rather than incorporating formal confidence intervals into the model fitting, we chose to present the variability across replicates for each site by averaging the flux values and visualizing their distribution using box-and-whisker plots. This approach provides a clear, interpretable representation of the spread and central tendency of fluxes across the six datasets, allowing for direct visual comparison of site-to-site variability. We have updated the manuscript to reflect this methodological decision in methods section, and we believe this presentation adequately captures the uncertainty inherent in the measurements for the purpose of inter-site comparison:

"To assess variability in porewater fluxes across datasets and to facilitate comparison with the BIGO fluxes, we computed the average flux per site and visualized the distribution using box-and-whisker plots. This descriptive approach provides a visual representation of the spread in the data, facilitating comparisons between the HI and CL areas."

Lines 410-416. Description of POC, PON, and CaCO3 variations. The authors provided the revised sentence with a description of POC, PON and CaCO3 variations across sites, which is exactly what I meant in my previous comment. However, this revised sentence should also include the results of the statistical analysis which are not statistically significant.

Thank you for pointing this out. We have now included the results for TS and C/N ratios. The revised sentence reads as follows: "In the surface sediment layer, the concentrations of TS and C/N ratios showed no significant differences between the CL (TS: 0.62 ± 0.13 wt.%; C/N: 7.75 ± 0.26) and HI (TS: 0.49 ± 0.17 wt.%; C/N: 7.93 ± 0.38) sites."