

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.

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 shows that temporality is not relevant, then it is justifiable to average the data in the figures. If the results of this analysis shows that temporality is actually relevant, then the authors can not 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.

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.

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.

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.

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.

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.