Response to Reviewer #1

We thank Reviewer #1 for their insightful comments and suggestions. Below please find our detailed point-by-point response.

General Comments

1. 'The authors revised their paper in response to many of the reviewers' comments, but they didn't really try to address with my main complaint. Why this study? They still have only a rather vague statement about wishing "to quantify the importance of nitrogen loss processes in overall nitrogen cycling." Maybe a more precise question is, why do the mesocosms? ...'

RESPONSE: As suggested, we have expanded on the aims of the mesocosm experiment (please see our response to the Line 48 comment below). We furthermore added '– a mesocosm approach' to the title, to highlight the complimentary novelty of our experiment.

Specific comments of Reviewer #1

Line 9: The argument why substrate limitation explains the difference is not clear. See below my comment on L265.

RESPONSE: There is probably a misunderstanding here i.e. measured and potentially realised rates in the mesocosms and Pacific were similar. Substrate limitation is most likely to explain that measured rates were higher than theoretically sustainable in the mesocosms. The confusion might have been caused by using the term in-situ, to distinguish processes in the mesocosms from those during bottle incubations. We have now changed the respective sentence to 'However, actual in-situ rates in the mesocosms, estimated via Michaelis-Menten kinetic scaling, did most likely not exceed 0.2–4.2 $\mathrm{nmol\ N_2\ L^{-1}h^{-1}}$ (interquartile range), due to substrate limitation.'

Line 33: I'm not sure what this means: "the net balance in terms of bioavailable nitrogen is negative." All of the processes discussed before this sentence are about how N is lost, so of course, they are "negative." Why is this sentence needed?

RESPONSE: We agree, and have removed the sentence.

Line 48: The authors here could add a couple of sentences about what they hope to learn from their mesocosm experiment.

RESPONSE: As suggested, we have changed the last paragraph of the introduction to: ' To better understand the events following the coastal upwelling of oxygen and nitrogen depleted deep waters, we make use of an off-shore mesocosm setup. Such approach allows simulating upwelling and tracing biogeochemical element cycling and associated trophic interactions. We were specifically aiming to address the question of nitrogen cycling, i.e. the build-up and turnover of organic nitrogen pools, their export from the surface to depth, and most importantly, potential loss processes. Because such approach enables budgeting of the various pools, it will be an alternative and independent assessment of the nitrogen balance in coastal ODZs, next to classical shipboard transects.'

Line 84: The N/P ratio, to be picky, isn't given in Table 1, just N/P/Si. More obvious than the difference in N/P ratios between the two waters is the presence or absence of NO3 and NO2. That's seen in the first two columns of Table 1. I think it's a better way to label the two waters.

RESPONSE: The N/P/Si ratio given in Table 1 allows to directly read off the N/P ratio. The reason why we would prefer to keep the 'low N/P' and 'very low N/P' terminology (besides that it appears valid, with 1.7/1 and 0.1/1), is that it will be consistent with the accompanying paper by Bach et al. 2020.

Line 265: If denitrification were substrate-limited in the mesocosms, wouldn't you expect rates to be lower, not higher than in situ? The authors' argument is not clear.

RESPONSE: This is probably the same misunderstanding as above, i.e. in-situ is used not to describe conditions outside the mesocosms but within, contrasting the lab incubations. We hope to clarify this by changing to :'Most interestingly, measured denitrification rates in incubations of mesocosm waters during the second half of the experiment exceeded those theoretically sustainable at in-situ substrate availability.'

Line 270: The authors changed "measured/maximum" to "measured/maximum sustainable rates", in response to my comment that it's ambiguous. In their response to the reviewers, they say they made the change to make "clear that in cases where substrate limitation was encountered, maximum-sustainable rather than measured rates were used for in-situ N-loss estimates." But that's not clear in the Discussion. "measured/maximum sustainable rates" is still ambiguous by itself. Use of "/" often leads to ambiguity.

RESPONSE: We have adopted the suggestion by reviewer #2 to use Michaelis-Menten kinetic scaling, rather than maximum sustainable rates derived from substrate concentrations, to estimate in-situ from measured rates in incubations. This should avoid any confusion.

Line 279: The authors say inorganic N didn't differ much between the two mesocosms, citing Table 1. But that table has fairly high NO3 and NO2 (>1 uM) for one mesocosm and zero NO3 and NO2 for the other. That seems like a big difference to me. Maybe over time, the difference disappeared?

RESPONSE: We state that the differences in dissolved inorganic nitrogen between these water masses are relatively small in relation to the overall nitrogen pool in the mesocosms. In other words, adding 3 $\mu\mathrm{molL}^{-1}$ of DIN to a 30 $\mu\mathrm{molL}^{-1}$ background of bioavailable nitrogen will change the overall nitrogen, potentially available for denitrification and anammox, by only 10%, as opposed to no addition.

Table 2: The author revised this table in response to most of my criticisms. I still think it would be better if they label the rows in the table, not give explanations in the caption, to indicate the expected maximum rates and the anammox values. I had to think too much to interpret the * and the italics used to indicate the anammox values. What's the "overall mean"? It needs to be explained.

RESPONSE: We actually had added '*anammox**' to the row label, as suggested. We have

changed 'Mean' to 'Treatment Mean', which should clarify that the 'Overall Mean' is across all mesocosms. Finally, the table was further simplified as now only reporting in-situ rates, and measured rates being presented in a separate table in the appendix.

Also, I suggest putting a "-" in front of the N-loss estimates (make them negative—they are losses) to make clear they can be compared directly with the N-budget numbers.

RESPONSE: Done.

Table 3 and Figure 5: Although the authors have simplified Table 3, I still think it and Figure 5 aren't needed. The authors don't really use these to make any points. I still don't understand Figure 5.

RESPONSE: We actually do use information provided in this table and figure to make an important argument, i.e. 'The primary drivers of in-situ denitrification rates appeared to be the availability of $\mathrm{NO_2^-}$ and $\mathrm{NO_3^-}$, followed by particulate organic matter (nitrogen), as indicated by multiple linear regression and effect size analysis (Fig. 5). This suggests that heterotrophic rather than chemolithotrophic denitrification was dominant (compare section 4.2.2), as both are substrates for the former process and eventually limit rates.'

Concerning Figure 5, what might be confusing is our approach to not simply use all seven measured potential predictors for an MLR. In order to find the sweet spot between model complexity (number of predictors in the model out of the seven possible) and avoid overfitting, we employed multiple stepwise MLRs. To explain the rational and benefit we have added the following to the figure caption: 'The red arrows denote the relatively small increase in maximum $\mathrm{R^2}$ beyond 5 main variables and the relatively large decrease below, indicating this to be the sweet-spot in terms of balancing model complexity with predictive power, avoiding overfitting (see Methods for details).' We have also slightly modified a sentence in the methods section, hopefully providing more clarity on the approach.

Response to Reviewer #2

We thank Reviewer #2 for their insightful comments and suggestions. Below please find our detailed point-by-point response.

General Comments

1. ... However, I am still a bit dissatisfied with the quality of their oxygen measurements, which should have been more carefully done in-situ, and also during their incubations. Denitrification and anammox are highly dependent on oxygen concentrations, even at nanomolar levels. I do not quite understand why they would observe such a high offset (+13 $\mu\mathrm{molL^{-1}}$) between Winkler titrations and their optical CTD sensor. Were Winkler titrations performed at all depths? Oxygen concentrations are such important measurements for their study, it is a pity that more attention wasn't dedicated to calibrating their CTD sensor better.

RESPONSE: After factory calibration prior to the experiment, the optical oxygen sensor was two point field calibrated as per the manufacturer's instructions (for 0 and 100%) every other deployment. The most likely explanation for the offset to the two Winkler samples that

were measured is the extremely steep oxycline, basically going from 100% at the surface to almost 0% within 10 to 15 metres. Hence, any sensor delay will cause the probe to read higher oxygen concentrations than actually present. That is the reason why we employed a hysteresis correction, assuming a response time of 1 second (based on information provided by the manufacturer). However, if the response time was actually 2 to 2.5 seconds, an offset of $13 \mu \mathrm{mol} \mathrm{L}^{-1}$ would be explained. So, the most likely explanation is that the response time of the optical oxygen sensor was actually slower than we thought. We have added this piece of information to the methods section. In hindsight, we should have measured the actual hysteresis of our sensor, as well as taken more Winkler samples. The realisation that getting precise oxygen concentration measurements from CTD casts when gradients are extremely steep, only materialised when carefully analysing the data after the experiment.

2. ... Second, I would like to see a more thorough assessment of how important gases (O_2, $\mathrm{H}_2\mathrm{S}$) influence nitrogen loss. They added a short discussion in their manuscript, but perhaps back of the envelope calculations would be more helpful to understand the effect of purging with He on the concentration of these gases. In the case of $\mathrm{H}_2\mathrm{S}$, a reduction of 80% after purging is quite a lot and is most likely to affect chemoautotrophic nitrogen loss ($\mathrm{H}_2\mathrm{S}$ oxidation coupled to NO_3^- reduction).

RESPONSE: There is a misunderstanding here, and we have changed our wording in the Methods section to be more clear about it. The He flow comparison with Holtappels et al. suggests that the reduction of O_2 was not to less than 20% of in-situ concentrations but by less than 20%. Hence, that is also the case for all the other gases, and even more so for $\mathrm{H}_2\mathrm{S}$, as it is buffered. Hence, given the relatively high $\mathrm{H}_2\mathrm{S}$ concentrations, potential chemolithotrophic denitrification should not have been limited by $\mathrm{H}_2\mathrm{S}$. In fact, measured denitrification was best explained by nitrite, nitrate and particulate organic nitrogen concentrations, and $\mathrm{H}_2\mathrm{S}$ had hardly any explanatory power (see MLR results), suggesting that chemolithotrophic denitrification was a minor contributor, if any.

3. As was pointed out by reviewer #2, I think the authors need to be more careful in interpreting these rates at the larger scale, especially since it was recognized that rates were stimulated due to substrate limitation during the 15 N-labeled incubations. These are potential rates that provide information about processes; hence these rates should be interpreted with caution and the limitations of this study should be better acknowledged. I agree with reviewer 2 that the authors should focus on the importance of different processes (denitrification versus anammox) rather than trying to compare their rates to the full nitrogen budget in the mesocosms (any strong agreement between the two is likely fortuitous).

RESPONSE: As already pointed out in our previous response to reviewer #1 comments, the overall nitrogen loss over the first 38 days of the experiment (prior to ornieutrophication) had been calculated using maximum sustainable rates, when substrate limitation was encountered. Hence, the finding that the addition of labelled NO_2^- stimulated measured rates, especially during the second half of the experiment, had been removed from for the overall nitrogen loss estimate. We have now also adopted the suggested approach to estimate in-situ rates from Michaelis-Menten kinetic scaling (see below). Together with the fact that there is an overall

good agreement of measured N loss with the full nitrogen budget, lends confidence to interpret findings in a larger context. And as we have pointed out, there is no statistically significant correlation between individual measured N loss and the full nitrogen budget, which would indeed be fortuitous given the number of entities in the mass balance. However, overall mean N loss and N budget are indistinguishably close, again, lending confidence to our findings and resulting conclusions.

4. I also agree with reviewer #2 that NO_2^- cannot be deemed to be "more important" than NO_3^- for denitrification.... The authors need to cite relevant study that observed clear discrepancies between ^{15}N-labeled rate measurements using both NO_2^- and NO_3^- and perhaps briefly discuss these differences.

RESPONSE: We have added the follow explanation and reference to the manuscript: 'The reason for NO_2^- rather than NO_3^- concentrations explaining rates of denitrification in one of the MLRs could be found in the following: As denitrification from NO_3^- to N_2 involves multiple and independent steps and organisms, the correlation between N_2 production rate and a substrate concentration should become better the closer one gets to the end of this chain (Fig. 1). For example, there should be a perfect correlation between N_2O concentrations and N_2 production rate, and NO_3^- concentrations and their turn-over to NO_2^- are meaningless if the intermediate steps to nitric and nitrous oxide are blocked or constitute a bottle-neck. This also would contribute to the finding that denitrification rate measurements based on $^{15}NO_3^-$ can be lower than those based on $^{15}NO_2^-$ (Hamersley et al. 2007).

5. On this note, I do not think the authors understood my argument regarding the production of excess $^{29}N_2$ during ^{15}N-NO_2^- labeled incubations (see Chang et al. 2019 paper). This process needs to be discussed in the manuscript (in connection with their data).

RESPONSE: We have added the following statement to the methods section: 'It is noted that there have been studies which found discrepancies between denitrification calculated using $^{29}N_2$ as above, and $^{30}N_2$, due to non-binomial distributions (De Brabandere et al., 2014; Chang et al., 2014). This has been attributed to so-called intra-cellular 'NO_3^-/NO_2^--shunting', which leads to an error in the calculations of labelled to unlabelled substrate, as based on known additions and measured seawater concentrations. As of noisy $^{30}N_2$ data, we can not check if that was an issue here, yet it would lead to an underestimation of denitrification rates in both cases. Given the good agreement between our rate measurements and a full nitrogen budget (see section 3.2 for details), however, it appears that potential 'NO_3^-/NO_2^--shunting' did not affect our rate measurements significantly.'

Specific Comments

Line 40: Adding "of deep waters" after frequency does not make sense as they are referring to upwelling frequency.

RESPONSE: It is both, the upwelling intensity and upwelling frequency of deep waters.

Line 46-48: At least one reference needs to be added at the end of this sentence.

RESPONSE: We have changed 'can' into 'could' as there are no studies to date yet.

Line 63-64: What were measured H_2S concentrations using this sensor for incubated waters? That could help address the He purging issue if H_2S concentrations were negligible.

RESPONSE: The CTD sensor is too large to fit into the 12 ml Exetainers, hence H_2S could not be measured directly for the incubations. Please see also our response the the reviewer's general comment #2.

Line 77-82: This section is still a bit vague. Oxygen concentration is very important in controlling nitrogen loss rates. Were oxygen concentrations also measured at all depths using Winkler titrations? If so, I think these values should be used instead if they experienced issues with their CTD sensor calibration. I don't think it is sufficient to say: "Hence oxygen concentrations .... are likely to have been significantly lower". How much lower? Are they sure that conditions were truly anoxic, and conducive to nitrogen loss?

RESPONSE: We have added a likely explanation for the observed off-set (please also see our response to the reviewer's general comment #1). This also should clarify by how much, i.e. exactly the off-set.

Line 86: Remove "However" at the beginning of sentence.

RESPONSE: The 'However' makes sense, as having very similar deep water masses in terms of N-deficit goes against our intent described a sentence earlier, i.e. to simulate upwelling of deep water with two distinct OMZ signatures.

Line 103-106: Also cite the new manuscript by Bourbonnais et al. (2020) in Frontiers that describes these types of incubations in detail as well as provide calculation templates: Bourbonnais, A, C. Frey, X. Sun, L. A. Bristow, A. Jayakumar, N. E. Ostrom, K. L. Casciotti, and B. B. Ward. (2021), Protocols for assessing transformation rates of nitrous oxide in the water column, Frontiers in Marine Science 8, 293.

RESPONSE: Done.

Lines 102-103: I think this sentence is ambiguous. Change to "Our calculations have shown that exchanging the bottle volume at least 24 times is required to reduce the O_2 concentration to less than 20% of in-situ conditions.

RESPONSE: It is a reduction BY less than 20%, not TO less than 20%. We have clarified the text accordingly.

Line 103: "Observered" is misspelled!

RESPONSE: Thanks for finding this typo.

Lines 113-115: This sentence is a bit vague. Could they provide a back of the envelope calculation to better estimate how He purging would affect H_2S concentrations? This is important to assess the role of chemoautotrophic versus heterotrophic denitrification.

RESPONSE: Please see our response to the reviewer's general comment #2.

Lines 118-122: This sentence is too long – I suggest breaking in two.

RESPONSE: We have split it in two.

Line 125-131: Why was $^{30}N_2$ noisy? I think that their rates are high enough to get a good $^{30}N_2$ signal. Calculating denitrification rates using mass 29 can be problematic as other studies reported production of excess 29 N 2 that could not be accounted by assuming binomial distribution (after considering the contribution from anammox). I strongly recommend the authors to read de Brabandere et al. (2013) and Chang et al. (2014) for more information regarding this process....

RESPONSE: Why it was noisy, we do not know. For the other concerns raised, please see our responses above.

Lines 168 and 177: Remove the word "please"

RESPONSE: Done.

Lines 226-230: I think this similarity is fortuitous since only potential rates are measured using ^{15}N-labeled incubations.

RESPONSE: Please see our response to the reviewer's general comment #3 above.

Lines 261-263: I do not think complete nitrogen loss occurs at such high O_2 concentrations (30-40 $\mu molL^{-1}$).

RESPONSE: We simply cite the findings by Farías et al. 2009 here. We also do not speculate on whether the nitrogen loss would be 'complete'.

Lines 269-278: It would have been best to construct the Michaelis-Menten curves as in Michiels et al. (2019) or use published Michaelis-Menten parameters (for the same or similar environments – as published in Michiels et al. (2019)) to estimate denitrification rates at in-situ NO_2^-/nitrate concentrations (rather than using a maximum nitrogen loss rates based on nutrient concentrations). The authors did not well address this point in their response to my previous review.

RESPONSE: We thank the reviewer for bringing this up again, and totally agree. We have now calculated in-situ denitrification as suggested, using the half-saturation rate constant by Michiels et al. (2009), which is the highest published for water column denitrification. Hence, this conservative approach should rather under- than overestimate in-situ rates. We have also updated Tab. 2 accordingly, which furthermore simplified it, and added a new table to the appendix with the measured rates. We also re-ran the MLRs, using in-situ denitrification rates. Overall, the new in-situ rates are slightly lower than previously estimated, but all our conclusions remain.

Line 286-287: Would that rather be an upper boundary estimate (relative to true environmental conditions), since orni-eutrophication and using maximum-sustainable denitrification

rates would artificially increase their nitrogen loss rate estimates?

RESPONSE: The N-loss from the budget approach would underestimate losses if ornieutrophication would have increased nitrogen pools in the mesocosms prior to T38. However, as there is no indication that this was the case, as well as it appeared to be confusing, we removed this statement.

Line 343: Change to "over-consumption"

RESPONSE: Done.

Lines 364-375: The authors need to acknowledge that purging with He before their 15 N-labeled incubations would reduce the H_2S concentrations. Hence, these rates should be interpreted with caution.

RESPONSE: Please see our detailed response to the reviewer's general comment #2.