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. '... While this study is worthy of publication, I do have some concerns, mostly regarding purging of the samples during the 15 N-labelled incubations, which removes H_2S - an electron donor during chemolithoautotrophic denitrification. Furthermore, since mostly all O_2 is removed during purging with He prior to incubating the samples, their measured rates are not representative of in-situ conditions (O_2 concentrations in bottom waters were generally higher than 20 $\mu molL^{-1}$, which is too high for N_2O conversion to N_2....'

RESPONSE: We will address the issues raised by the reviewer in the specific responses below.

Specific comments of Reviewer #1

Line 9: This was in contrast to realized rates in the surrounding Pacific". This sentence is unclear as the author later claim that rates measured as part of the mesocosm experiment and in surrounding waters were comparable.

RESPONSE: We will clarify this statement with 'In the surrounding Pacific measured denitrification rates were similar, although no indications of substrate limitation were detected.'

Lines 39-43: This sentence about ocean acidification seems a bit out of place since the connection with OMZs and nitrogen cycling is not clear. I suggest removing.

RESPONSE: We will clarified the connection by adding the following 'Changes in upwelling frequency/intensity, oxygen availability, temperature and pH can influence planktonic food web functioning in EBUS, with repercussions for nitrogen loss processes.'

Line 73: How long were the samples stored before analysis?

RESPONSE: It was about 2-4 hours. This information will be added.

Line 74: O_2 concentrations should ideally be monitored during 15N-labelled incubations using non-invasive O_2 measurement technology .... This is also to ensure that no O_2 is infiltrating during the incubations from the stoppers.

RESPONSE: We did not monitor the oxygen inside the exetainers in this experiment. However, data from other very similar exetainer experiments, in which O_2 was monitored with the Lumos sensors (Sun et al. 2020), show that the O_2 level varied between essentially zero and up to 100 nM over a period of two days. Even that highest concentration is below the lowest thresholds reported to inhibit anammox and denitrification (>200 nM Dalsgaard et al. 2014; see discussion in the manuscript, section 4.2.2).

Line 75: I think such a high O_2 offset is problematic and not representative of anoxic coastal

waters off Peru – where O_2 concentrations are generally well below 10 μM. I suspect that O_2 is also introduced during sampling from a Niskin bottle (de Brabandere et al.. 2012). Furthermore, O_2 in the nanomolar range has been shown to influence N_2 production rates (Dalsgaard et al., 2014). Since their samples are purged with He, the estimated rates are only putative and likely not representative of real in-situ conditions. I understand the limitation, but in-situ incubations would be preferable.

RESPONSE: The reviewer is correct, it would be wonderful to be able to do in situ incubations in order to assess true rates of denitrification and anammox. Unfortunately, at present, we are not aware of any realistic way to do that. The in situ incubation methods are cumbersome and allow at most a few measurements per day (e.g., Collins et al. 2018, Ward et al. 2019), and at present would be difficult to deploy in a mesocosm and hard to scale up to the numbers of experiments required for the experimental design used here. The non incubation approach of measuring in situ gas concentrations (gas tension device, Reed et al. 2018) has great attractions. This method was published after the study under review had been completed. The GTD has not been widely used yet, but based on the presentation by Altabet et al. at the Ocean Sciences meeting in 2020, it promises to be very useful on the oceanographic scale, although with substantial assumptions and constraints. So much as we would like to, making in situ rate measurements in the mesocosms is not yet feasible.

The purging approach reduces the gas concentrations for all gases prior to the incubations (except helium), including $N_2, N_2O, O_2, CO_2, H_2S,$ etc. By lowering the oxygen concentration, purging may enhance the rates of anaerobic processes such as denitrification and anammox. But as observed by de Brabandere et al. (2013), the fact that the processes occurred in anoxic incubations in water collected from oxic layers indicates the presence of viable populations of microbes capable of those processes. That implies the definite potential for in situ activities.

We have also re-calculated oxygen concentrations as measured by the optical CTD sensor by applying a 1 second response time hysteresis correction as described in Fiedler et al. (2013), rather than just mentioning the issue that 'raw' CTD data will over-estimate oxygen concentrations in a particular depth during a down cast, when moving from high to low oxygen concentrations. Now, corrected oxygen concentrations at the sampling depth are hovering around 20 $\mu molL^{-1}$ for most of the time. Together with the +13 $\mu molL^{-1}$ offset in comparison to Winkler titrations in these samples, this suggests that in-situ oxygen concentrations reached indeed below 10 $\mu molL^{-1}$, being representative for anoxic coastal waters off Peru.

Lines 80-82: Why is this treatment referred to as a moderate treatment if the N:P values between this and the extreme treatment are similar? I would rename this treatment as it is clearly not representative of moderate N loss conditions.

RESPONSE: We will re-name the treatments to 'low N/P' and 'very low N/P', respectively, which will also keep it consistent with the terminology introduced in the accompanying paper by Bach et al. (2020).

Lines 82-83: How did they prevent gas exchange and minimize O_2 contamination in these

waters during collection/injection? As this is an important detail for their experiment, a brief explanation should be added here (without having to refer to the Bach et al. (2020) paper).

RESPONSE: We will add the following piece of information: 'The deep waters were collected into 100 m^3 bags at the respective depths and sealed once brought back to the surface. Deep water was added by first removing about 20 m^3 from each mesocosm and replacing it with the respective deep water that was injected into the bottom layer between 14-17 metres on day 11, and the surface layer between 1-9 metres on day 12. To minimise changes to deep water gas concentrations during injections, water was pumped from several meters depth out of the deep water bags.'

Lines 87-88: Why not use 15 N-labeled NO_3^- to measure denitrification rates? Nitrate concentrations are generally higher than NO_2^- and NH_4^+ (thus, the substrate is less limiting).

RESPONSE: Previous experience has shown repeatedly that lower rates of N_2 production result from parallel incubations with $^{15}NO_3^-$ vs $^{15}NO_2^-$. We have interpreted that difference to the exchange of NO_2^- as an intermediate in the complete denitrification pathway, i.e. $^{15}NO_2^-$ produced from $^{15}NO_3^-$ is diluted with residual $^{14}NO_2^-$ in the medium and reduces the amount of label that makes it all the way to N_2. Thus, the rates measured from $^{15}NO_2^-$ are a better estimate of the actual rate of N_2 production. The fact that we routinely detect $^{15}NO_3^-$ reduction (as $^{15}NO_2^-$ production) in these same incubations shows that NO_2^- dilution does indeed occur.

Line 89: Why 3 $\mu molL^{-1}$? It seems to be a bit arbitrary.

RESPONSE: 3 $\mu molL^{-1}$ tracer addition is pretty standard. For NO_2^-, in particular in ODZ conditions, it is a level commonly occurring in natural waters, so does not represent a big perturbation, but ensures enough substrate to obtain a signal in the product. At the beginning of the experiment similar concentrations were present in the bottom layer. However, being closed systems, NO_2^- concentrations decreased over time, so the tracer addition could have stimulated measured rates, in particular after the N-depleted deep water addition. This is consistent with the statements in L134-138 (original MS) about the measured rates exceeding the rates that would be possible at in situ substrate concentration.

Lines 86-93: On a cautionary note, other studies (de Brabandere et al., 2013; Chang et al., 2014), observed that more $^{29}N_2$ is sometimes produced than could be accounted for assuming a binomial distribution after taking the production due to anammox. They propose that "nitrite shunting" where NO_3^- is converted to N_2 completely intracellularly without exchange with the external ambient NO_2^- pool could lead to that $^{29}N_2$ excess. I am curious to know if such $^{29}N_2$ excess was also observed in this study. Using NO_3^- as a tracer could help to account for this process.

RESPONSE: As we have only made incubations with labelled NO_2^- we have no comparison, or binomial expectation. Please also refer to our response to the Lines 87-88 comment.

Lines 93-101: I understand that purging with He is necessary since these are anoxic incubations, but since H_2S was present in bottom waters, removing all gases (including H_2S) would underestimate chemoautotrophic denitrification. I strongly recommend complementation of

stripped gases involved in N cycling metabolisms.

RESPONSE: The reasoning behind He purging is to reduce background $^{14}N_2$ concentrations (to more easily pick up label incorporation into the N_2 produced during denitrification or anammox), not to strip the solution off dissolved gases. And indeed, we saw in earlier studies that we can observe massive rates of chemoautotrophic denitrification using this approach (Kalvelage et al., 2013), suggesting that a significant portion of dissolved gases such as H_2S is retained. Hence, this method doesn't appear to necessarily underestimate rates resulting from He bubbling.

Line 99: What is the recommended flow rate? I assume purging time is adjusted to leave enough background N_2 for GC-IRMS measurements?

RESPONSE: As mentioned previously, the idea of purging is to reduce background N_2 in order to enhance the ability to detect the small isotopic signal of the labelled N_2 produced during the incubation. Since N_2 concentration doesn't vary much in seawater at the level we can detect with these methods, we used a standard flow rate (monitored by assuring 1 – 2 psi at the exetainer level, rather than flow rate at the cylinder level) and purging time, previously calculated to assure at least 10-fold replacement volumes for the exetainers. These methods have been described in great detail elsewhere by ourselves and others, which is why we did not repeat all the details and rationale here.

Lines 102-103: At what temperatures were samples equilibrated?

RESPONSE: Samples were equilibrated at room temperature. This will be mentioned in the text.

Lines 110-114: This approach could potentially affect their rate calculations if "nitrite shunting" produces $^{29}N_2$ excess. See above comment (lines 86-93).

RESPONSE: Please see our response to the previous comment(s).

Lines 113-115: In figure A1, it seems like there might be an exponential increase from time point 6 hrs and 20 hrs. In this case, I would only use the linear portion for rate calculations. Is this observed for all other incubations? It would have been useful to obtain another time points somewhere in between (at 12 hrs) to better disregard this possibility.

RESPONSE: Given the clear increase between 0 and 2 hours, we would rather argue for the lower 7 hour data point to be off. Furthermore, deviations from a linear increase over the entire incubation period was rather observed at times to be resulting from a lag phase during the first 2 hours, as acknowledged in the figure caption. We agree that having an extra time point between 8 and 20 hours would have been ideal in hindsight, but there are always logistical constraints. Furthermore, having the majority of samples in the first half of the incubation proved to be quite valuable as the lag phase issue could be clearly detected and accounted for, when necessary. Finally, restricting the incubation period to just under one day ensured to avoid what is usually considered to introduce potential bottle-effects. We will make reference to the latter issue in the methods section.

Lines 134-138: A more correct approach would be to construct Michaelis-Menten curves and calculate the half saturation constants and maximum denitrification rates from the measured in-situ rates (see Michiels et al., 2019).

RESPONSE: There is surprisingly little information on the dependence of denitrification on the concentration of nitrate or nitrite. This is probably because denitrification is often shown to be limited by organic matter concentration in oceanic systems. Again, in the coastal mesocosms, this may not have been the case – OM may not have been the limiting factor – and as noted above, the tracer addition could have stimulated measured rates. Lacking direct experiments on the MM kinetics of denitrification in this system, the approach used here is probably a reasonable compromise.

Lines 146-147: Change to: "24 hrs per day x 38 days x 2 (conversion between N 2 to N) divided by 3 (contribution of bottom layer water to overall mesocosm volume), ..."

RESPONSE: We agree and will make the suggested changes.

Lines 189-191: The oxygen concentrations shown in Fig. 3E are generally above 20 μmolL^{-1}, which would be too high for N_2O conversion to N_2 (see Dalsgaard et al., 2014 and Frey et al., 2020). Therefore, their measured rates are potential and denitrification was likely only observed because the samples were purged with helium - removing mostly all O_2.

RESPONSE: We agree, like in any assay incubation, measured rates should be taken with a grain of salt when extrapolating to in-situ conditions. Please also see our replies on oxygen response time hysteresis correction (suggesting oxygen levels below 10 μmolL^{-1} for most of the experiment), as well as helium purging.

Line 194: I don't think it makes sense to call this a "moderate" treatment (see above comment lines 80-82).

RESPONSE: As suggested, we will change the terminology.

Lines 199-200: Again, these relatively high H_2S concentrations indicate that chemoautotrophic denitrification might be an important process that was not measured (since samples were purged with He before the incubations).

RESPONSE: Please see our response to the Lines 93-101 comment. Furthermore, please also see our response to the L226 comment from reviewer #2 on nitrite and organic matter being the main drivers of denitrification.

Lines 211-214: The near perfect agreement between the two approaches is a bit surprising considering that measured rates are potential and likely not representative of in-situ conditions.

RESPONSE: Good point. We have wondered about that ourselves. Considering the many caveats for incubation-based rate measurements, how can they make so much sense in comparison to processes estimated from several other independent in-situ measurements? The only conclusion we can come up with is that the rates measured in the essay incubations must have been similar to what was happening in-situ.

Line 266: The calculated overall nitrogen loss could also be overestimated since in-situ denitrification rates were likely lower. Samples collected in the mesocosms and surrounding Pacific waters were purged before the incubations, removing mostly O_2 and thus creating conditions more conducive to N_2 loss. The O_2 concentrations observed in bottom waters were too high for N_2O conversion to N_2 (see Dalsgaard et al., 2014).

RESPONSE: Please see our responses to a number of comments above, in particular the one on the new CTD-oxygen-optode response time correction - Line 75 comment.

Lines 313: Why did C/N values not increase in that one mesocosm?

RESPONSE: Deep water additions were followed by a bloom of the dinoflagellate *Akashiwo sanguinea*, fixing carbon without significant net nitrogen assimilation, in all except this one mesocosm. We will mention this fact, which is described in more detail in the accompanying Bach et al. (2020) paper, in the discussion.

Lines 319-321: It is also possible that the measured DON pool was mostly recalcitrant, with fast cycling of labile DON.

RESPONSE: We agree and, as stated in the text, it would require preferential N over C remineralisation.

Lines 340-341: Denitrification/anammox linked to microenvironments around particles would not be captured by 15N-labeled incubations, especially if these are not performed in situ.

RESPONSE: As the incubation seawater was not filtered and hence contained particles, microenvironments around those are likely to have been re-established during the 20 hours of incubation.

Lines 343-345: It is unclear how H_2S would inhibit anammox in their incubations, since samples were purged (hence H_2S was removed).

RESPONSE: Please see our responses to various comments above.

Lines 350-353: It would be relevant to include these data (i.e., anammox functional marker gene hzo) in the manuscript.

RESPONSE: Unfortunately, no genomic data can be presented at this stage, as of ongoing Nagoya Protocol negotiations. However, as there is no discrepancy between our rate measurements and described gene abundance observation, there shouldn't be a need to explicitly show the latter.

Line 375: Why is the contribution from the Arabian Sea, where significant N-loss occurs, not taken into account here?

RESPONSE: We will include reference to the Arabian Sea and Bay of Bengal in a revised version of our manuscript. This will not change the main findings and conclusions of our calculations.

Lines 376-379: I don't think there is anything new in this statement. Due to the large uncertainties associated with these estimates, it is still unclear if the majority of the N-loss occurs in the water-column or sediments.

RESPONSE: The discussion in the last paragraph is only about water column denitrification, i.e. a comparison of globally assembled in-situ estimates with our mesocosm-derived measurements.

Lines 379-380: Why is export production projected to decrease if upwelling intensity and frequency (and thus nutrient supply) is expected to increase (Hauri et al., 2013 and Wang et al., 2015 papers cited in the introduction)?

RESPONSE: This is a valid point raised by the reviewer. We will be more specific here and explain that projected reductions in global export production are thought to result from changes to community structure. Furthermore, in regards to export production in ODZs and OMZs, as of a potentially counter-acting increase in upwelling intensity and frequency, we will be more cautious about the expected sign of change.

Table 2: I would rename the "moderate" treatment to "extreme" since the degree of N-loss is similar in both treatments. It is odd to express individual N-budgets for each mesocosms as negative values and present the mean as a positive value. I suggest renaming these columns N-loss from 15N-labelled incubations and N-loss from N-budget.

RESPONSE: We agree and will change the terminology to 'low' and 'very low' N/P. We will also change the N-budget mean to a negative number to match the individual mesocosm values.

Figure 2: What was bottom depth at the mooring site?

RESPONSE: The depth at the mooring site was between 25 to 30 metres. This information will be added to the figure caption.

Figure 1A: It is difficult to tell if the last time point around t = 20 hrs represents an exponential increase (as often observed for 15N-labelled incubations).

RESPONSE: Please see our response to the Lines 113-115 comment.

Technical corrections

Line 49: define nm (i.e., nautical miles).

RESPONSE: We agree.

Line 63: change acoording to "according"

RESPONSE: We will do.