This is an interesting article exploring the long-term impact of riverine nitrogen inputs on the global ocean nitrogen inventory and associated primary production. The authors show that in simulations that have reached equilibrium, impacts on the global N inventory and primary production are highly limited due to feedbacks on N fixation and denitrification. I appreciated the candidness of the discussion on the modelling approach’s strengths/weaknesses and how the results differ from those previously published in this area. The main discussion that I think might be expanded upon is how indicative simulations that have reached equilibrium are for policy relevant timescales (interannual to decadal). Put another way, how do the authors’ main conclusions change over the course of these 10000-year simulations? Are the implications different for watersheds that are experiencing rapid increases or decreases in nitrogen export at present? Most of my comments and suggestions relate to how the manuscript text could be better structured and the figures could be made much easier to interpret. Subject to these changes, I would be happy to recommend for publication.

**Minor comments**

L23. “atmospheric” –should probably be dissolved/aqueous.

L26. Aren’t these “model concepts” observationally/experimentally derived? Some reference to the empirical evidence would be useful here. Maybe also its limitations if it’s heavily based on given species (eg Trichodesmium).

L31. concentrations “of” fixed N

L32. “the consumption of O2”- Do you mean the consumption of O2 during remineralisation? If so, be explicit.

L47. The Séférian et al. 2020 reference is perhaps worth citing here as it summarises the
inclusion of riverine inputs in recent models.

L51. “real” should probably be realistic or observed.

L59. Maybe clarify what is meant by N export here. Riverine delivery? For many in the ocean biogeochemistry community export is instinctively a vertical flux.

Figure 1. More detail is needed on the tracers in the figure legend.

L100. “atmospheric” – should probably be dissolved/aqueous.

L101. Are they limited by a max NO3 concentration? I know much of this will be in the cited references but more detail is required on N-fixation in the model. Highlight perhaps that most models don’t have explicit diazatrophs and this is an advantage of using uvic. What is the diazatroph PFT based on? How does N-fixation compare to observations where they exist?

L105-110. As with the above comment, some comparison on how denitification in the model compares to observations would be very useful. I think a global map of N-fixation and denitification in the model CTR is required.

L140. It should be made clearer on first use that NEWS etc are simulation names.

Table 1. For clarity I would remove UVic from the simulation names as this is not repeated in the main text. I would also add a CTR row.

L156. “vary a little” – please quantify this

L151. This should really be called a “Results and Discussion” section.

L160. The wording here needs to be clearer. “At smaller scales...globally higher.” This reads like it contradicts L167-168.

Figure 3. Axes are missing labels and units here.

Figure 5. I find it very difficult to see differences between positive and negative anomalies using this colorbar. I suggest changing to something far more distinctive (e.g. red for negative anomalies). The same applies to other figures using this scale.

L175-177. This is difficult to see in Figure 5 maybe cite figure 6 here.

Figure 6. The depths given in the figure don’t match the legend.

L181-183. This sentence is confusing and needs rephrasing.

Figure 7. Label missing from panel c.

L187. typo. “amounts to an increase of only 1.1...”

L209. I’m not sure the language here is accurate. Presumably the model is not explicitly trying to compensate anything. Wouldn’t this be better described as enhanced denitification sinks promoting conditions that favour N-fixers over the other PFT type and consequently global N-fixation rates are higher?

L210-213. The discussion of other literature here before properly explaining your model results is confusing. Where these papers have used the same model this should be clear.
L216. “where NO3 concentrations are substantially reduced relative to the CTL”

Figure 8- See earlier comments on how it would be nice to see global Nfix in the control.

Figure 9. Suggest using different color palettes for mean states and anomalies.

L219-227. The balance between results of the model simulations and the discussion of other literature needs to be more organised. The presentation of discussion before results is quite confusing.

L221-222. This is a bit rushed and therefore confusing. I think more detail is needed here on this mechanism, the difference between the stoichiometry of N-fixation and denitrification and how spatial and temporal coupling is important for the positive feedback to occur.

Table 3. Benthic denitrification appears to be twice the magnitude of that in the water column. This doesn’t seem to be reported and discussed in the manuscript. Does this have implications for models lacking benthic denitrification?

L246. “...and vary little between...”

L251. “smaller scales” is ambiguous. Here and elsewhere I recommend being more specific e.g basin/watershed/coastal scales etc.

L257-259. Differences in spatial patterns between these simulations are difficult to discern it looks more like the magnitude of change is the only difference.

L259. Are these subtropical and tropical regions where N-fixers are predominately confined to?

L297. Maybe “exported again” should be “recycled” here.

L301. And presumably not all the N can be consumed via local primary production due to other constraints.

Table 6. Maybe the increase in NPP per quantity of additional N would be a useful metric to add to this table given each watershed provides different total N delivery.

L325. For clarity I think “inhibiting additional NPP” should be “limiting increases in NPP”.

Section 4. Given the extensive discussion of the Lacroix et al. 2020 paper, I think a few more details are required to compare the studies properly. What was their model resolution, did they have explicit N-fixers and benthic denitrification?

L358. The emphasis here and in the conclusions (L376) doesn’t really match the findings. I would say the feedbacks do much more than “partly compensate” the riverine fluxes. Maybe this would be clearer if you gave the % of added N that is retained in the inventory at equilibrium or some other metric of feedback strength.

L388-389. I’m not sure this “upper limit” conclusion would hold if N fertilisation were targeted spatially and temporally in regions of N limitation. Perhaps this should be toned down a little.