Reply to Anonymous Referee #1

We are grateful for the comments and edits of the anonymous reviewer who invested time for the review of our manuscript. The reviewer's text is reported in Italic and our responses in roman.

The research topic is very timely, the manuscript is well written, and the model simulates seem to be well done. Furthermore, the authors have really made a lot of simulations and obviously went to great lengths. However, many things need to be addressed before the paper can be accepted. The most important points are: (a) Discussing Lacroix et al. (2021), (b) adding many more references to support the text, (c) evaluating the model results with observations and show where the model is good or bad, (d) evaluating the global news present-day river fluxes to understand how realistic they are, (e) discussing the far too low NPP in the Arctic Ocean, and (f) testing if results are statistically significant. In this state, the main conclusion that rivers might be of importance in coastal region has no underlying prove and large parts of the manuscript are not possible to review. I would focus more on regional coastal areas and discuss and analyze these further, but this is up to you.

Major comments

- *References are often missing. One of the most important references missing is probably Lacroix et al. (2021) who performed very similar simulations. Comparison to their results would be essential. Please revise the entire manuscript for the many missing references. Here some examples:*
    - *Lines 30/31: no references for this statement*

This sentence together with the sentence after is one complete statement, and the references are therefore after the second sentence in Line 32.

- *Lines 33/34: no references for this statement*

Reference (Chester, 2012) will be added.

- *Lines 34/35: no references for this statement*

This is our own statement.

- *Lines 37-48: Almost no references here!*

Line 37-38: Our own statement.

Line 39-41: Reference (Seitzinger et al., 2010) will be added.

Line 41-44: Reference was given.

Line 44-46: Reference (Meybeck and Vörösmarty, 1999) will be added.

- *Lines 49/50: No references*

That statement is according to our own knowledge, and it is supported by examples written in Line 50-55.

○ *…*

- *Lines 49-54: A more detailed explanation is missing on how ESMs simulate rivers. Seferian et al. (2020) gives a good overview. There were already 5 ESMs in CMIP5 that simulated riverine C, N, and P (3 also simulated Fe) fluxes and there are now 8 in CMIP6. Some of them even simulate dynamically changing riverine fluxes.*

We will add more information on the riverine implementation, especially with respect to new developments in some of the CMIP6 models (e.g. CESM2, CNRM-ESM2-1, MIROC-ES2L, IPSL-CM6A-LR):

"The latest generation of ESMs have implemented some forms of riverine inputs in their ocean biogeochemistry modules (Seferian et al., 2020). In general, those that implement riverine inputs, do it differently from constant contemporary fluxes based on data from GlobalNEWS (IPSL-SM6A-LR and CESM2; Aumont et al., 2015; Danabasoglu et al., 2020) to fully interactive with terrestrial nutrient leaching transported by dynamical river routing (CNRM-ESM2-1 and MIROC-ES2L; Seferian et al., 2019; Hajima et al., 2020). Models with interactive riverine transports do not consider biogeochemical processes in the freshwater, hence tracers are treated as passive tracers. Redfield ratio is typically used to convert from one chemical compound to the others."

- *In the main manuscript, a large space is given to the Arctic Ocean. Please introduce the Arctic Ocean accordingly and explain in the Introduction already why it might matter if you want to keep it as one of the regional seas that you want to discuss (see Terhaar et al., 2019 & 2021 and citations within).*

Thank you for the suggestion and references. We will add the paragraph below in the introduction after the first paragraph.

The Arctic Ocean, which accounts for only 4% of the global ocean area (Jakobsson, 2002), takes 11% of the global river discharge (McClelland et al., 2012), and it is estimated that about one third of its net primary production is sustained by nutrients originated from rivers and coastal erosion (Terhaar et al., 2021). Therefore, there is no surprise that Arctic primary productivity will be affected by altered riverine transport of nutrients and carbon under future climate changes. Previous studies have shown that enhanced riverine nutrient input increases primary production in the Arctic Ocean (Letscher et al., 2013; Le Fouest et al., 2013, 2015, 2018; Terhaar et al., 2019), while large riverine DOC delivery reduces CO_2 uptake in Siberian shelf seas (Anderson et al., 2009; Manizza et al., 2011).

- *The model is not evaluated. Only a reference to previous publications is mentioned (lines 111-113). However, to understand and discuss the changes in the future and the sensitivity to riverine fluxes a much more detailed model evaluation is needed. For example, NPP is far too low in the Arctic Ocean: for the 2nd half of the 20th century the simulated NPP is around 100 TgC/yr. However, the observation-based NPP in the last years of the 20th century is slightly above 450 Tg C/yr (Arrigo and van Dijken, 2015). If the model is not capable of simulating NPP in a part of the* *ocean, why should we trust any of the projections done by that model? Having demonstrated that this is the case in the Arctic Ocean, I cannot trust the other numbers. Especially, given that the model-obs differences (Fig 3b) are so much larger than the differences between rivers and no rivers (Fig. 3c). Please make a thorough comparison and evaluate your results on the background of the model performance and tell the reader about the models' strong points and weaknesses when it comes to ocean biogeochemistry.*

Thank you for the comment. The NorESM model has been evaluated in different publications. For the Arctic domain, its skill in simulating the observed primary production was done in Lee et al. (2016) and the reviewer is correct that the NorESM is biassed low against observations. However, it is on par with other global ocean models. For instance, in their paper, Lee et al. (2016) assessed the relative skills of 21 regional and global biogeochemical models in reproducing the observed contemporary Arctic primary production. The NorESM has a negative bias of -0.49, and is well within the multi-model mean bias of -31+/-0.39. In another study, the NorESM model is compared with a regional model that comprises part of the Arctic region, and it shows that the NorESM simulates too late and too short bloom period than the regional model (Skogen et al., 2018), hence the annual integrated primary production is too low. Global high resolution models also show considerably lower NPP in the Arctic (e.g., 165Tg C/yr; Terhaar et al., 2019). Such common shortcomings in global scale marine ecosystem models can partly be attributed by the simplified parameterization, which can be improved through data assimilation (Tjiputra et al., 2007; Gharamti et al., 2017).

We want to show the direction of improvement toward observation by showing the differences in PP between with rivers and no rivers (Fig. 3c). Despite the model-obs differences in the Arctic Ocean PP, we would like to see whether there is a difference in the sensitivity of PP and carbon uptake to the riverine input at global scale and in the Arctic (Figs. 5 and 7).

We will add discussion on model confidence in the Arctic Ocean in the 'Limitations and Uncertainties' section.

- *Four different scenarios for the future riverine fluxes are introduced in line 125. However, it is impossible to know which scenario is which. Please introduce the scenarios carefully so that the reader knows these scenarios are.*

We will add a table similar to Figure 2 in the paper by Seitzinger et al. (2010) to introduce these four scenarios.

- *Riverine data in general: How good is the data that you use? I would like to see a comparison to observations of larger important rivers such as the Amazon or the rivers in the Arctic that are observed by ArcticGRO (https://arcticgreatrivers.org/). Nutrient fluxes from Global News 2 in the Arctic can be off by 300% (Kaiser et al., 2017; Thibodeau et al., 2017; Terhaar et al., 2019). Without knowing the quality of the present-day riverine fluxes, it is not possible to evaluate the results.*

We didn't find any evaluation of the Global NEW 2 data in the publications that the reviewer mentioned here (Kaiser et al., 2017; Thibodeau et al., 2017; Terhaar et al., 2019). However, the Global NEWS 2 riverine dataset has been calibrated and assessed against measured yields (Mayorga et al., 2010) and has been widely used and evaluated for different river estuaries (A number of publications can be found in a special section of Global Biogeochemical Cycles: https://agupubs.onlinelibrary.wiley.com/doi/toc/10.1002/(ISSN)1525-2027.NUTRIENT1). For example, van der Struijk et al. (2010) compared the Global NEWS nutrient yields to observed values for South American rivers. They stated that "For some rivers (such as the Amazon), the model performs better than for others. In general, the model seems to do better for DIN, DON and DOC than for DIP, although for the Amazon also modeled DIP yields also compare well to measured values. The variations in yields among rivers are described well by the NEWS models… Nevertheless, we may argue that the NEWS models in general perform reasonably well for South American rivers."

Although the evaluation of the riverine data is out of the focus of our paper, we will mention the evaluation of the dataset in section 2.2 and refer the readers to relevant publications.

- *Some of the organic nutrients are remineralized directly due to the fixed stoichiometric ratio in the marine organic matter (line 140). Please tell the reader how much is remineralized directly and discuss later if that influences the results. The lability of terrestrial organic matter is an important factor for the impact of riverine nutrient fluxes on NPP and carbon fluxes on air-sea CO2 fluxes (Terhaar et al., 2021).*

We have calculated the proportion of directly remineralized matter, including dissolved organic matter (DOM) and particulate (inorganic and organic) matter (PM), i.e., $X/(DOM_{riv}+PM_{riv})$ (X is the directly remineralized dissolved organic and particulate matter). The directly remineralized part on average accounts for 64.8%, 27.8% and 62.8% of the total riverine organic and particulate matter of phosphorus, nitrogen and carbon, respectively. This approach may lead to bias in the enhanced primary production. We further calculated the contribution of the directly remineralized part on the enhanced primary production, by comparing X with the corresponding total riverine-induced dissolved nutrient additions [X/(X+DIXriv)*100%], which accounts to 80.5%, 33.3%, and 41.1% for phosphate, nitrate, and carbon, respectively. Assuming that all coastal regions are nutrient limited, this direct remineralization could be responsible for 33.3%-80.5% of the enhanced primary production, depending on which nutrient species is limiting the primary production.

- *Is there a particular reason why you use fluxes from 1970 for the FIX run? Later you compare the NPP results to observation-based NPP from after 2000. Wouldn't it be better to use the 2000 fluxes for the FIX run?*

In Section 3.3, we assess the effect of future changes in riverine inputs comparing the future period 2050–2099 to the recent past reference period 1950–1999. This assessment was our main objective when we designed the experiments. We chose the 1970 fluxes because they are more representative for the 1950–1999 period than the 2000 fluxes (Beusen et al., 2016).

We agree with the reviewer that the use of 2000 fluxes for the FIX run would have been preferred when comparing to observation-based NPP from after 2000. In hindsight, choosing 1950–1999 as a present day reference was not an optimal choice considering better availability of observations estimates during the early 21st century. We will mention this in the discussion of caveats in the revised manuscript version.

- *As mentioned above, nutrient fluxes often do not scale at all with runoff as concentrations can decrease strongly when discharge increases. Furthermore, apart from DOM and DIP the global news scenarios all give very different future scenarios compared to RUN. I think the simulation RUN hence really makes very little sense and I do not know what its value is. I would certainly not make such strong statements in the Discussion (line 321). I am happy to be convinced otherwise.*

We agree with the reviewer that the RUN experiment has its uncertainties. It is an idealistic scenario, which assumes that future changes in riverine carbon and nutrient transports are directly linked to changes in riverine freshwater transports, ignoring other anthropogenic effects related to land surface processes. The comparison of RUN to the GNS experiments demonstrates that the anthropogenic and natural changes in nutrients and carbon on land are equally or more important than the direct effect of the changes in the hydrological cycle. Including RUN along with the more realistic GNS scenarios thus still provide valuable information, which e.g., may caution other modelling groups against adopting an over-simplified coupling of riverine nutrient and carbon transports to the hydrologic cycle. Another motivation for RUN was to introduce seasonal and interannual variability in nutrient and carbon transports that is linked to variability in riverine freshwater transport. Future work should explore if GNS and RUN can be integrated to produce more realistic long-term trends in riverine nutrient and carbon transports as well as short-term variability.

We will clarify that to explore the best practical way of implementing the riverine inputs for modelling groups is one of the aims of this work. We will delete the sentence in line 321.

- *It seems that large changes are always simulated in the Black Sea. However, results from ESMs in enclosed or semi-enclosed seas usually make no sense. Did you mask these seas, including the Mediterranean Sea? If not, how does masking these seas change your results?*

The reviewer is right that in global models the results in those enclosed or semi-enclosed seas are often largely biassed. We did not mask those seas during simulation, but we masked them for plotting.

- *I am really struggling with the significance of the results. For example, in line 192, I would like to see the inter-annual variability as a measure of the standard deviation to see how much they are really different. Similar, is there an uncertainty estimate for the observation-based estimates in line 197? Overall, the differences in annual NPP and RMSE seem to be so small that I am not sure if it makes sense to use terms such as 'better' or 'improve'. Can you find any statistical way to evaluate if the changes are significant? This comment should be addressed to other numbers throughout the manuscript. Please be also careful with the word 'significant' as used in line 238 if it has no statistical meaning.*

We thank the review for the comment. We will follow the suggestion and add more information on the robustness of changes in the revised manuscript. We will test the statistical significance with respect to sampling error resulting from interannual variability and add this information in the text and on the figures. In addition to formal significance, we will introduce a signal-to-noise measure that provides further information on the importance of the signal in a real-world context (more on that below). We will detail the significance assessment in a new section "2.4 Statistical robustness of riverine impacts" and move the last paragraph of section 2.3 to the new section.

As stated in section 2.3, we expect even small values to be statistically significant because the interannual climate variability is the same in all simulations and thus most of the interannual signal is removed in the computation of the experiment differences. We will illustrate this behaviour with a Supplementary Figure showing two time series from two different experiments in one panel and the time-evolving difference of these series in another panel, together with confidence intervals of the time-means. Differences between two temporal periods (i.e., future minus past) are still affected by interannual variability, but the sampling effect will be similar for all experiments and therefore should not impact much the comparison of the temporal differences between the experiments.

In the revised manuscript, we will formally test the statistical significance of changes by doing the following: 1) construct time series of differences (either from two periods or two experiments), 2) compute the standard-deviation of the mean from the time-evolving differences, 3) compute the p-value from the time-mean difference and the standard-deviation of the mean, 4) reject null-hypothesis if $p<alpha$. This will test only for local significance and the probability of falsely rejecting the null-hypothesis somewhere on the map is generally higher than alpha. While performing more rigorous multi-hypothesis and field significance testing would be desired, it is beyond the scope of this study and we have not observed it in similar studies. Despite limitations, the local significance test provides basic information on statistical robustness.

Small but statistically significant differences between the experiments (where the effect of interannual variability has been removed) are not necessarily large enough to have real-world implications and be detectable in observations. Therefore, we will introduce a second measure. We will define a signal-to-noise ratio S2N as the time-mean difference over the standard-deviation of the mean of the original field (and not the difference field with the interannual variability removed). On our difference maps, we will mark areas with S2N>1 as regions where the signal emerges from interannual noise, indicating that the signal is large enough to have real-world implications.

In addition to including robustness information in the difference maps, we will add a table that summarises all globally integrated changes mentioned in text and figures. We will add statistical significance (expressed as p-values) and S2N values to this table.

- *Please refrain from making strong claims about the Arctic. Indicating a ~76% increase in NPP is misleading giving how bad the model simulates the present-day NPP. Based on an observation-based NPP of 450 TgC/yr, a change of 70-80 TgC/yr is only an increase in 17%. Moreover, the very low present-day NPP suggests either* *strong light or strong nutrient limitation. If it is strong nutrient limitation, riverine fluxes would have an overly strong effect because all nutrients would be used immediately. So maybe even the 17% are still too high. This goes back to the point that the reader must know how good the model performs locally.*

We agree that stating a relative change (in percentage) could be misleading given the biassed low primary productivity simulated in our model. As with other Earth system models, the increased Arctic primary production in NorESM, which is not strongly nutrient limited, is associated with sea-ice loss. Our projected PP increase in the reference run is roughly 70 Tg C yr-1 by the end of the 21st century. This is in the same order of magnitude estimated from 11 ESMs with mean change of 59Tg C yr-1 (individual ESM ranges from -110 to +253 Tg C yr-1; Vancopenolle et al., 2013). We will include the following statements in the revised version.

We note that the relative change in PP in the Arctic is likely to be overestimated since the NorESM simulates a biassed low PP under the contemporary climate. Nevertheless, the projected absolute change of 70Tg C yr-1 is well within the range estimated from other Earth system models (Vancopenolle et al., 2013).

General comments

- *Often 'biogeochemistry is used as a synonym for PP and air-sea CO2 fluxes. But the word biogeochemistry also includes acidification, carbon and nutrient cycles, and other things. Please just say PP and air-sea CO2 fluxes. (for example line 250).*

We will make respective specifications in order to be clear in these occasions.

- *Please adhere to the best practice guide (https://www.ncei.noaa.gov/access/ocean-carbon-data-system/oceans/Handbook_2007/Guide_all_in_one.pdf) and use CT and AT instead of DIC and ALK.*

The best practices guide by Dickson et al. (2007) refers to best practices for measurements. The symbols AT and CT for alkalinity and total dissolved inorganic carbon are not sacrosanct. However, we will follow this recommendation and replace DIC by CT and ALK by AT.

- *I find the name 'reference run' misleading. It is rather a control run. Reference should be the best case or something.*

The term "reference experiment" is widely used in the Earth system modelling community in the way used in our study. It is standard for ESM modelling studies to have a "reference" experiment and one or several "sensitivity" experiments. The response is typically evaluated as the difference sensitivity minus reference. The term "control experiment" is typically reserved for a simulation where external forcings are fixed at pre-industrial (or some other) levels. However, the external forcings used in this study are transient ones, therefore using "control" may confuse readers.

Furthermore, the original NorESM1-ME (Tjiputra et el., 2012) and many other ESMs do not include riverine nutrient and carbon transports to the oceans. Therefore, we find the use of "reference" for the experiment without transports appropriate in our study and prefer to keep it in the revised manuscript version.

- *Significant digits should always be the same. For example, in lines 204 and 205 you cite air-sea CO2 fluxes and use different number of digits.*

We will check and change all numbers accordingly.

Minor comments

- *Lines 19/20: Can one speak of improve based on such small changes in the RMSE? Is it significant?*

Referring to the response to major comment on the significance, we will do more analysis and edit it accordingly.

- *Line 13: Suggest changing “not only regionally but also globally” to “regionally and globally”*

We will change it accordingly.

- *Line 18: Suggest changing “modelled” to “simulated”*

We will change it accordingly.

- *Line 22: Unclear what you mean by depending on the riverine configuration. There is no range in the numbers given in this sentence.*

We meant that by adding the riverine nutrients input, the projected future decline in PP can be alleviated maximum by 0.6 PgC yr-1 (27.3%) globally in our experiments. The range is given in Conclusion in line 367-369 (Riverine nutrient inputs into surface coastal waters alleviate the nutrient limitation and considerably lessen the projected future decline in PP from -5.4% without riverine inputs to -4.4%, -4.1% and -3.6% in FIX, RUN and GNS (averaged over four scenarios), respectively.)

- *Lines 23/24: the last part of the sentence should be rewritten*

We will change it to “The riverine impact on projected C uptake depends on the net effect of riverine nutrient induced PP increase and riverine C input induced outgassing.”

- *Lines 22-25: A lot of words that do not tell much. Nutrients increase CO2 uptake, CT fluxes decreases it. But where does it increase it and where does it decrease it? Maybe shorten this or explain.*

Please see the response to the last comment.

- *Line 26: Can you be more quantitative?*

We will change this sentence to “Simulations with high-resolution global or regional models with an adequate representation of shelf processes are required to accurately assess the impact of future riverine scenarios.”

- *Line 31: Not sure if you can count runoff.*

We will change it to “river runoff plays an essential role in …”.

- *Lines 31/32: transporting nutrients where? Suggest adding "into the ocean" after "transporting nutrients".*

We will add it accordingly.

- *Line 34: What do you mean by "absolutely dominant" source? More than 50%? Please be clear.*

We will add the information as follows. "For some substances such as total phosphorus (~90.0%) and total silicon (>70.0%), riverine input even acts as the absolutely dominant source."

- *Line 34: Suggest adding "into the ocean" after "transport of carbon".*

We will add it accordingly.

- *Line 34: Suggest writing air-sea CO2 exchange instead of air-sea C exchange. It CO2 and not C that is exchanged across the air-sea interface.*

We will change it accordingly.

- *Line 36: What is "it"?*

We will change this sentence to "Despite our limited understanding on the riverine carbon fluxes, they could play an important role in closing the global carbon budget (Friedlingstein et al., 2021) and could be very sensitive to regional and global changes such as weathering, land cover and climate (Meybeck and Vörösmarty, 1999)." They refer to riverine carbon fluxes.

- *Line 36: Do you mean global ocean carbon cycle or really global carbon cycle?*

We have removed this term. Please see above.

- *Line 36: Global and regional changes of what?*

Please see the response above.

- *Line 37: Suggest starting a new paragraph here*

We will follow the suggestion in the revised version.

- *Line 55: In the Arctic, Terhaar et al. (2019) started to assess future changes.*

Thank you for the reference. We will include it in the introduction.

- *Line 56: Please say which datasets you are referring to.*

We will add "Global NEWS2 dataset" in this sentence.

- *Line 58: Why does more data make the impact study more 'desirable'?*

"Taking the advantage of the latest improvement of global river nutrient/carbon export datasets and responding to the demand of development of ESMs with increasing model resolution, the assessment of the impact of riverine nutrients and carbon on future projections of marine biogeochemistry becomes feasible and desired, especially for impact studies along continental margins."

Here more data makes the assessment feasible, and increasing model resolution makes the assessment desirable.

- *Line 58: Please say why it is now feasible? One could argue that the CMIP6 horizontal resolution is still not good enough to resolve the global ocean.*

Please see the response above.

- *Line 68: Please already say here why you use RCP4.5.*

We will change it to "the RCP4.5 (middle-of-the-road) scenario" here. We think section 2.3 Experimental design is a suitable place to state the detailed reason for the choice of future scenario. We will also add references to support our choice. Please see the response to comment on Line 153 below.

- *Line 84: Configured sounds strange here.*

We will change "configured" to "implemented".

- *Line 88: 'd' is missing in 'based'.*

We will add it.

- *Lines 88-113: Does the ocean biogeochemical component also have a name? It is very confusing that you say it is based on HAMOCC and then you only describe HAMOCC. That makes it impossible to understand what has changed.*

In the NorESM1-ME, the biogeochemical model is still called HAMOCC5, despite some development from the original HAMOCC5 version, which was developed in Hamburg. However, in the newer version of NorESM2, the biogeochemical model is renamed as iHAMOCC.

- *Line 129: What is the motivation to use CT and AT data from Hartmann (2009)? What kind of data is that? Modeled, observed, extrapolated? What is the underlying data? Please explain what you use.*

The reason that we used CT and AT data from Hartmann (2009) is that the Global NEWS2 dataset does not include CT and AT data. The CT and AT data from Hartmann (2009) are produced from a high-resolution model for global CO2-consumption by chemical weathering. The dataset contains different forms of riverine carbon (dissolved and particulate, inorganic and organic), implemented to the Global NEWS2 river basin map.

- *Lines 129-133: Why do you use iron riverine fluxes from 1990? Is there nothing newer including observations since 1990? Does it make sense to weight iron river fluxes by runoff? Often nutrients do not scale at all with runoff (Holmes et al., 2012), so I would like to see some support for this assumption.*

To the best of our knowledge, the available global riverine iron dataset is rare. Previous studies have used various approximation approaches, e.g., constant Fe to dissolved inorganic carbon (DIC) ratio (Aumont et al., 2015), Fe to phosphorus ratio (Lacroix et al., 2020). In the study by Aumont et al. (2015), the Fe: DIC ratio is determined so that the total Fe supply equals 1.45 Tg Fe yr^{-1} as estimated by Chester (1990). We are aware that our approximation likely has bias in regional scales, and we will discuss this in the Limitation section.

- *Line 135: Is there a reason why you use 1000 km and 300 km here?*

The short answer is that NorESM1 is based on NCAR's Community Earth System Model (CESM; Hurrell et al., 2013). The configuration for distributing riverine runoff into ocean grid cells along with the 1-degree resolution ocean grid have been both inherited from CESM without modification.

The exact reasoning behind the NCAR's choice of using a 1000 km e-folding length scale and 300 km cutoff value is not known to us. The effect of using an e-folding length scale that is considerably larger than the cutoff value is that approximately equal weights are used when distributing the runoff to ocean grid cells that lie within the cutoff range of the river mouth. The cutoff value must be large enough that at least one ocean grid cell midpoint lies within the range. For a 1-degree resolution ocean grid, a value of 100 km should be sufficient to satisfy this. Possibly, the large value of 300 km was chosen to also satisfy coarser grids, such as the 3-degree resolution grid used in NCAR's computationally efficient model configuration, and to avoid numerical instabilities by more smoothly distributing the runoff.

In this study, we ensured that riverine nutrients and carbon are distributed in the same way as riverine freshwater. We considered this important especially for the RUN configuration that couples the variability of the riverine nutrient and carbon transports to the variability of the riverine freshwater transport. How sensitive the ocean biogeochemistry impacts are to the details of how riverine runoff is distributed into coastal grid cells and generally how well ocean shelf processes are represented warrants further investigation.

We will add some text summarising the above to the discussion section or the Supplementary Information of the revised manuscript.

- *Line 145: Any reason why not GLODAPv2 is used?*

Our NorESM1-ME model configuration was finalised in the early 2010s, and GLODAPv2 was not available at that time. While there are improvements in the GLODAPv2 due to higher volume of data and improved interpolation scheme (Lauvset et al., 2016), since the model was then spun up for nearly 1000 years, we expect that these differences in initialization will not be significant for our purpose.

- *Line 147: Is the additional spin-up of 200 years is sufficient to get into a new equilibrium.*

The nutrient drift after 200 years spin-up is small (in the order of 1%/100 years).

- *Line 153: In what sense is RCP4.5 the most representative scenario? Most likely? Based on what?*

We will add the following sentence here. "Here, we consider RCP4.5 as the representative future scenario following the CO2 emission rate based on the submitted Intended Nationally Determined Contributions (INDC), which projects a median warming of 2.6-3.1 degC by 2100 (Rogelj et al., 2016)."

- *Line 160: What means not considered? Deactivated?*

We will change “not considered” to “deactivated”.

- *Lines 168-170: It is not entirely clear that you make 4 simulations. Can you be a little bit more explicit?*

We will change Line 166 to “GNS: Four different transient inputs following future projections of NEWS 2.”

- *Line 194: In the figure it does not look as if only 15% of the increase is in the coastal shelf seas. What do you mean by predominantly, can you be quantitative here?*

The number 15% is calculated from continental shelf, where seafloor is shallower than 300m. The increased primary production occurs on the continental shelves in the North Atlantic will be quantified in the revised version.

- *Line 214: Is this a result or a speculation. If you cannot prove it, I suggest to either add it to the Discussion and add literature that supports this point or delete it.*

This is not a direct result. We will remove it.

- *Line 244-245: Does it really make sense to say slightly higher if the difference is that small?*

We will add more information on the robustness of changes in the revised manuscript. Please also see the response to the comment on the significance of the results.

- *Line 272: Please be quantitative*

We will change it to "The projection of global total PP shows up to 27.3% less decrease, if riverine inputs are present in the model."

- *Line 279: Is not nitrogen the limiting nutrient in the Arctic Ocean (Tremblay et al., 2015)?*

Our model shows that a large area of the Arctic Ocean is nitrogen limiting, while some part is iron limiting.

- *Line 286: Why do you not add CMIP6 data?*

We will add comparison with CMIP6 results (Kwiatkowski et al., 2020; Tagliabue et al., 2021) in the revised version.

- *Why do you speculate in the Discussion in lines 290 to 293: You can show that with your model results.*

Thank you for the comment. The statement was inferred from model results (i.e., FIX simulation produces higher NPP than REF simulation, which suggests that riverine-induced nutrient addition alleviate the stronger nutrient limitation in the future. We have rephrased the sentence to clarify this.

- *Figure 4c is almost impossible to read.*

We will improve this figure by smoothing the contour lines and using different colours.

- *Figure 5a: Could you highlight the +38 more? I was confused first.*

We will highlight it in the revised version.

Reference

Anderson, L. G., Jutterström, S., Hjalmarsson, S., Wåhlström, I., & Semiletov, I. P. (2009). Outgassing of CO2 from Siberian Shelf seas by terrestrial organic matter decomposition. *Geophysical Research Letters*, 36, L20601. https://doi.org/10.1029/2009GL04004

Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., and Gehlen, M.: PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies, Geosci. Model Dev., 8, 2465-2513, 10.5194/gmd-8-2465-2015, 2015.

Chester, R.: The transport of material to the oceans: the river pathway, in: Marine Geochemistry, Springer, Dordrecht, https://doi.org/10.1007/978-94-010-9488-7_3, 1990.

Chester, R.: The Input of Material to the Ocean Reservoir, in: Marine Geochemistry, Wiley Online Books, 7-10, https://doi.org/10.1002/9781118349083.ch2 https://doi.org/10.1002/9781118349083.ch3, 2012.

Danabasoglu, G., Lamarque, J. F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2 (CESM2), Journal of Advances in Modeling Earth Systems, 12, e2019MS001916, https://doi.org/10.1029/2019MS001916, 2020.

Dickson, A.G.; Sabine, C.L. and Christian, J.R. (eds) (2007) Guide to best practices for ocean CO2 measurement. Sidney, British Columbia, North Pacific Marine Science Organization, 191pp. (PICES Special Publication 3; IOCCP Report 8). DOI: https://doi.org/10.25607/OBP-1342

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Bakker, D. C. E., Hauck, J., Le Quéré, C., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Bates, N. R., Becker, M., Bellouin, N., Bopp, L., Chau, T. T. T., Chevallier, F., Chini, L. P., Cronin, M., Currie, K. I., Decharme, B., Djeutchouang, L., Dou, X., Evans, W., Feely, R. A., Feng, L., Gasser, T., Gilfillan, D., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, Ö., Harris, I., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Luijkx, I. T., Jain, A. K., Jones, S. D., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lienert, S., Liu, J., Marland, G., McGuire, P. C., Melton, J. R., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S. I., Niwa, Y., Ono, T., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan, T. M., Schwinger, J., Schwingshackl, C., Séférian, R., Sutton, A. J., Sweeney, C., Tanhua, T., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F., van der Werf, G., Vuichard, N., Wada, C., Wanninkhof, R., Watson, A., Willis, D., Wiltshire, A. J., Yuan, W., Yue, C., Yue, X., Zaehle, S., and Zeng, J.: Global Carbon Budget 2021, Earth Syst. Sci. Data Discuss., 2021, 1-191, 10.5194/essd-2021-386, 2021.

Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., Ohgaito, R., Ito, A., Yamazaki, D., Okajima, H., Ito, A., Takata, K., Ogochi, K., Watanabe, S., and Kawamiya, M.: Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks, Geosci. Model Dev., 13, 2197-2244, 10.5194/gmd-13-2197-2020, 2020.

Hartmann, J.: Bicarbonate-fluxes and CO2-consumption by chemical weathering on the Japanese Archipelago — Application of a multi-lithological model framework, Chemical Geology, 265, 237-271, https://doi.org/10.1016/j.chemgeo.2009.03.024, 2009.
Hurrell et al. (2013). The Community Earth System Model: A Framework for Collaborative Research, Bulletin of the American Meteorological Society, 94(9), 1339-1360, https://doi.org/10.1175/BAMS-D-12-00121.1.

Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J. R., Dunne, J. P., Gehlen, M., Ilyina, T., John, J. G., Lenton, A., Li, H., Lovenduski, N. S., Orr, J. C., Palmieri, J., Santana-Falcón, Y., Schwinger, J., Séférian, R., Stock, C. A., Tagliabue, A., Takano, Y., Tjiputra, J., Toyama, K., Tsujino, H., Watanabe, M., Yamamoto, A., Yool, A., and Ziehn, T.: Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections, Biogeosciences, 17, 3439-3470, 10.5194/bg-17-3439-2020, 2020.

Lacroix, F., Ilyina, T., & Hartmann, J. (2020). Oceanic CO_2 outgassing and biological production hotspots induced by pre-industrial river loads of nutrients and carbon in a global modeling approach. *Biogeosciences*, **17**, 55– 88. https://doi.org/10.5194/bg-17-55-2020

Lauvset, S. K., Key, R. M., Olsen, A., van Heuven, S., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishii, M., Perez, F. F., Suzuki, T., and Watelet, S.: A new global interior ocean mapped climatology: the 1° × 1° GLODAP version 2, Earth Syst. Sci. Data, 8, 325-340, 10.5194/essd-8-325-2016, 2016.

Le Fouest, V., Babin, M., & Tremblay, J.-E. (2013). The fate of riverine nutrients on Arctic shelves. *Biogeosciences*, 10(6), 3661–3677. https://doi.org/10.5194/bg-10-3661-2013

Le Fouest, V., Manizza, M., Tremblay, B., & Babin, M. (2015). Modelling the impact of riverine DON removal by marine bacterioplankton on primary production in the Arctic Ocean. *Biogeosciences*, 12(11), 3385–3402. https://doi.org/10.5194/bg-12-3385-2015

Le Fouest, V., Matsuoka, A., Manizza, M., Shernetsky, M., Tremblay, B., & Babin, M. (2018). Towards an assessment of riverine dissolved organic carbon in surface waters of the western Arctic Ocean based on remote sensing and biogeochemical modeling. *Biogeosciences*, 15(5), 1335–1346. https://doi.org/10.5194/bg-15-1335-2018

Lee, Y.J., P.A. Matral, A.M. Friedrichs, V.S. Saba, O. Aumont, M. Babin, E. Buitenhius, M. Chevallier, L. de Mora, M. Dessert, D. Feldman, R. Frouin, M. Gehlen, T. Gorgues, T. Ilyina, M. Jin, J. John, J. Lawrence, C. Perruche, V. Le Fouest, E. Popova, A. Romanou, A. Samuelsen, J. Schwinger, R. Séférian, C. Stock, J. Tjiputra, B. Tremblay, K. Ueyoshi, M. Vichi, A. Yool, and J. Zhang, 2016: Net primary productivity estimates and environmental variables in the Arctic Ocean: An assessment of coupled physical-biogeochemical models. *J. Geophys. Res. Oceans*, **121**, no. 12, 8635-8669, doi:10.1002/2016JC011993.

Letscher, R. T., Hansell, D. A., Kadko, D., & Bates, N. R. (2013). Dissolved organic nitrogen dynamics in the Arctic Ocean. *Marine Chemistry*, 148, 1–9. https://doi.org/10.1016/j.marchem.2

Manizza, M., Follows, M. J., Dutkiewicz, S., Menemenlis, D., McClelland, J. W., Hill, C. N., Peterson, B. J., & Key, R. M. (2011). A model of the Arctic Ocean carbon cycle. *Journal of Geophysical Research*, 116, C12020. https://doi.org/10.1029/2011JC006998

Meybeck, M. and Vörösmarty, C.: Global transfer of carbon by rivers, Global Change Newsletter, 37, 18-19, 1999.

McClelland, J. W., Holmes, R. M., Dunton, K. H., & Macdonald, R. W. (2012). The Arctic Ocean estuary. Estuaries Coasts, 35(2), 353–368. https://doi.org/10.1007/s12237-010-9357-3

Rogelj, J., den Elzen, M., Höhne, N., Fransen, T., Fekete, H., Winkler, H., Schaeffer, R., Sha, F., Riahi, K., and Meinshausen, M.: Paris Agreement climate proposals need a boost to keep warming well below 2 °C, Nature, 534, 631-639, 10.1038/nature18307, 2016.

Seitzinger, S. P., Mayorga, E., Bouwman, A. F., Kroeze, C., Beusen, A. H. W., Billen, G., Van Drecht, G., Dumont, E., Fekete, B. M., Garnier, J., and Harrison, J. A.: Global river nutrient export: A scenario analysis of past and future trends, Global Biogeochemical Cycles, 24, https://doi.org/10.1029/2009GB003587, 2010.
Jakobsson, M. (2002). Hypsometry and volume of the Arctic Ocean and its constituent seas. Geochemistry, Geophysics, Geosystems, 3(5), 1028. https://doi.org/10.1029/2001GC000302

Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., Decharme, B., Delire, C., Berthet, S., Chevallier, M., Sénési, S., Franchisteguy, L., Vial, J., Mallet, M., Joetzjer, E., Geoffroy, O., Guérémy, J.-F., Moine, M.-P., Msadek, R., Ribes, A., Rocher, M., Roehrig, R., Salas-y-Mélia, D., Sanchez, E., Terray, L., Valcke, S., Waldman, R., Aumont, O., Bopp, L., Deshayes, J., Éthé, C., and Madec, G.: Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth System Processes in Present-Day and Future Climate, Journal of Advances in Modeling Earth Systems, 11, 4182-4227, https://doi.org/10.1029/2019MS001791, 2019.

Séférian, R., Berthet, S., Yool, A. et al. Tracking Improvement in Simulated Marine Biogeochemistry Between CMIP5 and CMIP6. Curr Clim Change Rep 6, 95–119 (2020). https://doi.org/10.1007/s40641-020-00160-0

Skogen, M. D., Hjøllo, S. S., Sandø, A. B., and Tjiputra, J.: Future ecosystem changes in the Northeast Atlantic: a comparison between a global and a regional model system, ICES Journal of Marine Science, 75, 2355-2369, 10.1093/icesjms/fsy088, 2018.

Tagliabue, A., Kwiatkowski, L., Bopp, L., Butenschön, M., Cheung, W., Lengaigne, M., and Vialard, J.: Persistent Uncertainties in Ocean Net Primary Production Climate Change Projections at Regional Scales Raise Challenges for Assessing Impacts on Ecosystem Services, Frontiers in Climate, 3, 2021.

Terhaar, J., Orr, J. C., Ethé, C., Regnier, P., & Bopp, L. (2019). Simulated Arctic Ocean response to doubling of riverine carbon and nutrient delivery. Global Biogeochemical Cycles, 33, 1048– 1070. https://doi.org/10.1029/2019GB006200

Terhaar, J., Lauerwald, R., Regnier, P. et al. Around one third of current Arctic Ocean primary production sustained by rivers and coastal erosion. Nat Commun 12, 169 (2021). https://doi.org/10.1038/s41467-020-20470-z

Tjiputra, J. F., Polzin, D., and Winguth, A. M. E. (2007), Assimilation of seasonal chlorophyll and nutrient data into an adjoint three-dimensional ocean carbon cycle model: Sensitivity analysis and ecosystem parameter optimization, *Global Biogeochem. Cycles*, 21, GB1001, doi:10.1029/2006GB002745.

Tjiputra, J. F., Roelandt, C., Bentsen, M., Lawrence, D. M., Lorentzen, T., Schwinger, J., Seland, Ø., and Heinze, C.: Evaluation of the carbon cycle components in the Norwegian Earth System Model (NorESM), Geosci. Model Dev., 6, 301-325, 10.5194/gmd-6-301-2013, 2013.

van der Struijk, L. F. and Kroeze, C.: Future trends in nutrient export to the coastal waters of South America: Implications for occurrence of eutrophication, Global Biogeochemical Cycles, 24, https://doi.org/10.1029/2009GB003572, 2010.

Vancoppenolle, M., Bopp, L., Madec, G., Dunne, J., Ilyina, T., Halloran, P. R., and Steiner, N. (2013), Future Arctic Ocean primary productivity from CMIP5 simulations: Uncertain outcome, but consistent mechanisms, *Global Biogeochem. Cycles*, 27, 605– 619, doi:10.1002/gbc.20055.