

Author Comments (ACs)

In this Author Comments:

- The original referee comments are in black (directly copied from the comments).
- Our responses are in blue.
- *The text we quoted from the manuscript is in gray italics.*

We sincerely thank all referees for their constructive comments and feedback on our manuscript.

Best regards,

Gang Tang (GT, as referenced below)

on behalf of all co-authors

Top-Level Updates Before Addressing Individual Comments:

- Title Revision:

The manuscript title has been updated to “Synthesizing Global Carbon-Nitrogen Coupling Effects – the MAGICC Coupled Carbon-Nitrogen Cycle Model v1.0” This new title more accurately reflects the content and scope of the manuscript and is in line with the title used for other modules of MAGICC (e.g., Synthesizing long-term sea level rise projections – the MAGICC sea level model v2.0, <https://doi.org/10.5194/gmd-10-2495-2017>).

- Terminology Clarification:

To avoid confusion, we now exclusively use “MAGICC” to refer to the full model (the online model including all components) and “CNit” solely for the coupled carbon-nitrogen cycle model. This eliminates the previous ambiguity caused by the frequent use of “MAGICC” in varying contexts.

RC2: 'Comment on egosphere-2024-1941', Yann Quilcaille, 21 Oct 2024

This manuscript proposes the description of a new module for a coupled carbon-nitrogen cycle in MAGICC, with its calibration based on land surface models and a set of CMIP6 ESMs. Overall, the model is correctly described and well presented. Its design is common to climate emulators, with pools and fluxes representing the essential elements and processes. This well-established approach does indeed improve simplifications, but with appropriate reasons, and leaving room for improvements or sophistication in future works. The comparisons to the training data show relatively good performances, albeit lower for CMIP6. Besides, adding this CN module to MAGICC would be an important improvement on this key climate emulator, enhancing the robustness of future constraints on the land carbon cycle. To summarize, this work is then important, timely and relatively well presented, and I consider that it fits perfectly the scope of GMD.

Nevertheless, after careful consideration, I would recommend that this manuscript should go through major corrections before publication. The major reasons are the oversimplification on the plant uptake of N (comment 1) and the lack of clarity in the data handling (comment 2). The first one would require either a better justification or to account more explicitly for the inorganic N pool either in the plant uptake of N or in the limitation on the NPP. The second one is necessary for better understanding by the readers.

There are some additional minor points (comments 3-9), that matter for the overall quality of the manuscript, but are not sufficient to justify a major correction. Finally, some details are simply brought to the attention of the authors (comments 10-13), without requiring any action.

GT: Thanks a lot for reviewing our paper and providing us with constructive and positive comments, especially those regarding the model formulations. For the major concerns, actually we (the author team) had a lot of discussions on the first one (nitrogen plant uptake simulation and nitrogen limitation on NPP) at the model development stage. The data handling details are added now for clarification. Please see our justifications of both in the respective responses.

Major comments:

1. Modelling of the plant uptake of N

The plant uptake is currently a relationship based on the NPP, with a temperature dependency (equations 22, 24 & 27 in Section 2.4). First, a required plant uptake given the NPP is estimated. Then, the limited NPP is estimated. Thus the plant uptake given this limited NPP is obtained. The limitation on the NPP depends only on atmospheric deposition and the required plant uptake.

I was expecting to see either the plant uptake or the limitation to NPP to depend on N_{min} , the pool that provides the N. If the limitation on NPP would have depended on N_{min} here, it would have affected the actual plant uptake. Yet, N_{min} affects neither the limitation of the NPP nor the plant uptake. Having this disconnection between N_{min} and the plant uptake or the limitation of NPP could cause inconsistencies, for instance in the following examples:

- For an excessive fertilization, we should expect no limitation from the N cycle on the NPP. Yet, in the current modelling, N_{min} would be saturated but it would have no effect either on the limitation $\epsilon_{CN}(NPP)$ or on the P uptake. Thus excessive N_{min} isn't accounted for. It may be the reason for the discrepancy described Lines 457-458.

GT: Thank you for your comments. The key question here is whether the mineral nitrogen supplied by fertilizer application (the FT flux in our model) can enhance ecosystem NPP. We discussed whether to combine atmospheric deposition and fertilizer application as the external nitrogen forcing for the ecosystem.

Ultimately, we decided against this, as fertilizer use primarily boosts productivity in agricultural systems, which is harvested within the annual timescale. The current version of MAGICC focuses on the terrestrial ecosystem and does not simulate agricultural dynamics. The nitrogen that is not utilized by crops either releases to the atmosphere or leaches into the ocean, making it effectively unavailable to the ecosystem. Including fertilizer application in our model could introduce bias, as it might unrealistically fertilize the ecosystem. To avoid this, we excluded this flux from the nitrogen limitation.

- In the current modelling, without fertilization and atmospheric deposition, the only flux going in N_{min} is $LDN2M$, ie the fraction of the decomposition flux from litter to the inorganic N pool. According to equations (34) and (37), this flux could become very small. Thus, the N_{min} pool could be depleted by a continuous plant uptake and an insufficient decomposition flux, and then would become negative. Thus insufficient N_{min} isn't accounted for. There may be a link with the issue mentioned in Lines 666-668.

GT: Please note the nitrogen that predominantly goes into the mineral nitrogen pool is the SR_N (Fig. 1), which is the soil organic nitrogen decomposition (mineralization of soil organic matter). The sum of SR_N and $LDN2M$ is effectively the net mineralization nitrogen. In most of the cases, either from the complex models or our emulations, the net mineralization is comparable with (usually a bit higher than) the nitrogen plant uptake at the annual time step. This will prevent the mineral nitrogen pool from depletion. We have now made this clear in the equation explanation, as quoted below:

The sum of the fraction of litter decomposition nitrogen entering the mineral pool ($LD2Mn$, i.e., litter mineralization) and the nitrogen released during soil respiration (SRn , i.e., soil organic matter mineralization) constitutes the ecosystem's net mineralization is effectively the ecosystem's net mineralization.

So far, the only explanations are Lines 183-191, mentioning that plant uptake minus net mineralization is linked to the required plant uptake. However, this is insufficient to justify removing the dependency on the size of the inorganic N pool, either in the (required) plant uptake or in the N limitation on the NPP. I checked the papers proposed as sources, but I could not find a justification for this modelling of the plant uptake:

- Zaehle and Dalmonch, 2011, 10.1016/j.cosust.2011.08.008: The approaches outlined in sections "Nitrogen limitation on plant C uptake" and "Plant nitrogen uptake and competition with soil microbes" insist on the need to represent N availability/limitation, without introducing the approach shown in the reviewed manuscript.
- Zaehle et al., 2014, 10.1111/nph.12697: none of the equations introduce the equations of the reviewed manuscript. Eqn 4a-c would actually confirm a dependency of the plant uptake flux with N_{min} .

This point is crucial for the modelling of the N uptake and NPP limitation, thus I strongly recommend the authors to properly justify this modelling, or adapt it. Adequate sources making use of this modelling or observed relationships would be needed. Additionally, if these equations were still used, biophysical justifications would be needed in my opinion. Finally, discussions on the limits would be needed as well, e.g. in the two cases outlined earlier.

GT: Thanks for careful checking. We really appreciate the posted question. First, the cited works are simply to support "*linking plant nitrogen status with net primary production (or photosynthesis) is common in complex models*". Back to the question: Why not use mineral nitrogen pool size to explicitly represent the nitrogen availability and use it for the nitrogen limitation on NPP. The reason is simple: The mineral nitrogen pool size is magnitude smaller than the nitrogen plant uptake requirement at the annual-mean scale. This is explained in our manuscript:

This formulation is transformed from complex models with the key idea of comparing mineral nitrogen availability and plant nitrogen requirement. In complex carbon-nitrogen models, the nitrogen availability is typically based on the current mineral nitrogen pool size (with mass unit) and the nitrogen requirement is computed from the integrated fluxes in a given time step (with mass unit) (Thornton et al., 2007; Wiltshire et al., 2021; Zaehle et al., 2014). The competition from microbial immobilization is also considered in some complex models. However, in a model with a much longer time step (e.g., annually) like ours, such a system would be inherently unstable since the mineral nitrogen pool size would be orders of magnitude smaller than the annual nitrogen demand (i.e., the system would be unstable because the turnover of the mineral nitrogen pool would be substantially smaller than the time-step).

Further explanation and background:

In complex models, nitrogen limitation is commonly represented by comparing the current mineral nitrogen availability to nitrogen requirements. In such models, the availability may be based on the mineral nitrogen pool size, as the required nitrogen is typically comparable to the pool size over short time steps. However, in a global-mean, annual-averaged box model, this direct comparison is not feasible, as converting nitrogen requirements (measured in N per year) to finer resolutions would either introduce unnecessary assumptions (e.g., monthly or daily estimates) or require additional parameters specific to each land surface model, which is undesirable.

The current formulation, using the required plant uptake (PUreq), is more reasonable when considering the sources of mineral nitrogen: net mineralization (predominant), biological nitrogen fixation (which is already fixed by plants within one year and thus not included in our nitrogen mass balance), atmospheric deposition, and fertilizer application (primarily relevant to agriculture). As explained in the manuscript, net mineralization directly supports plant uptake, which is why the mineral nitrogen pool size does not fluctuate significantly—plant uptake and net mineralization tend to balance each other.

Using net mineralization, which is on the same order of magnitude as plant uptake at the annual-mean scale, is therefore a more reasonable approach than using the annual mean size of the mineral nitrogen pool. The key assumption in our model is that net mineralization is linearly correlated with plant uptake, meaning the unmet plant uptake requirement from net mineralization can be expressed as a linear function of the required plant uptake (i.e., required plant uptake minus net mineralization = $f_2 \times \text{PUreq}$, where f_2 is a constant). This assumption is supported by the relationship between plant uptake and net mineralization data (below).

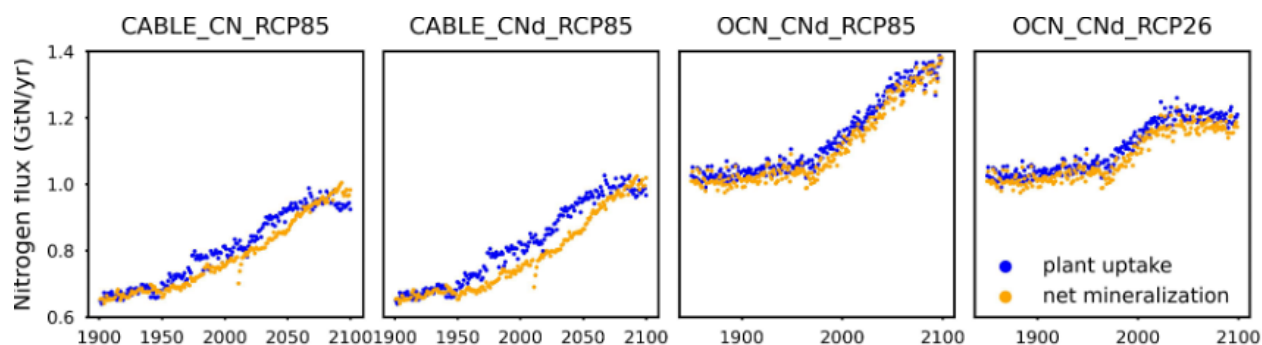


Figure A2. Relationship between nitrogen plant uptake and net mineralization as simulated by CABLE and OCN.

This point is the main reason for switching the manuscript from Minor to Major Correction. If I have missed something, I am sorry. I triple-checked, but could not find anything to prove me wrong. Therefore even if I were indeed wrong, other readers may also miss the point, and it would be necessary to clarify.

GT: We really appreciate your careful reading of our manuscript and posing this question. We agree that there might be something missing in our writing. Thus, I have rearranged the writing and added new subtitles for better clarity.

2. Lack of clarity on data handling

The Section 3.1 “Data acquisition and processing” isn’t clear enough at the moment. CABLE & OCN provided CMIP5 runs on RCPs. Besides, some ESMs provided CMIP6 runs on SSPs. Both sources are used for calibration, hence the analysis in Section 3.4 and further. However, the authors write that:

“Unfortunately, a robust and feedback-specific emulation is not feasible for CMIP6 ESMs, as the results from experiments without the nitrogen effect are unavailable.” (lines 363-364)

Apparently, the authors still managed to train and emulate the CN module of MAGICC? Is it about the robust and feedback-specific part on the training & emulation, and the ensuing workaround outlined Section 3.2?

GT: Thanks for the comment. Here we mean, since CABLE and OCN provided the nitrogen on/off experiments, the calibration would recognize the difference caused solely by nitrogen. But for CMIP6 ESMs with a nitrogen cycle, none of them provide the nitrogen off experiment. Thus the disentangled nitrogen effect in the CMIP6 ESMs is solely based on our emulation, though we applied constraints that were informed by the CABLE and OCN calibration. We cannot say the nitrogen effect in CMIP6 ESMs from our emulation is 100% robust. But when the emulation turned out matching the ESM outputs, it gave us confidence that at least our model is not too wrong. The only way to get a robust emulated nitrogen effect is to calibrate the CMIP6 ESMs with the nitrogen off experimental data - which is, unfortunately, not available for the current ESMs. We hope the modeling groups will run such experiments in the future but we also understand that this might be too computational demanding. That is also why we have a lot of discussion and further analysis about the limitation of the disentangled nitrogen effect in Section 5.3.

Besides, it remains unclear how data from CABLE & OCN is used for training in comparison to CMIP6. Is it a two-step training, first on CABLE & OCN, then on CMIP6, i.e. using the parameters obtained from CABLE & OCN as a first guess for the optimization on CMIP6 data? Or is it a one-step training, pulling all samples together? Although the results of the calibration are shown for CABLE & OCN in Section 3.3 and for CMIP6 models in Section 3.4, the Section 3.2 explaining the calibration setup does not mention CABLE or OCN.

I strongly suggest clarifying how both datasets are used precisely, and the questions that I present here. Ideally, a reader should not have to re-read this section to understand the data flow. It may require the authors to adapt the structure of Section 3.1 and 3.2, but it would be worthwhile for the readers.

GT: Thanks for the comment and suggestion. We realized this calibration section is not clear. We have now rewritten this section with details about the processes we calibrate our model.

Minor comments:

3. Land-use & deforestation

I do appreciate the effort in modelling land use in the C & N cycles, but I have to flag two important limits.

The first one is in the parametrization of the regrowth flux (Lines 313-318). At the moment, the regrowth depends only on the gross deforestation, with two constant parameters φ and $rrgr$ (equation 40). Yet, I would rather expect the regrowth of a deforested parcel to depend on its NPP rather than how much was

deforested. In other words, a primary forest would have a high C stock; deforesting would be a strong C flux because of its past unperturbed growth under a favourable climate; but its regrowth under a less favourable climate would be towards a lower C stock. A potential correction would be to approximate the deforested area using the ratio of *LUgrsd* with *CP* (thus neglecting *CL* and *CS*). The regrowth of this area would depend on the NPP, with some parameters to account that the regrowth on a deforested area is not exactly the same than the one aggregated on all biomes as modelled by MAGICC-CN.

Additionally, these fluxes are not only due to deforestation (lines 308-312). Biomass extraction from croplands will also matter a lot, especially for the N cycle. This is an important limit of this current modelling, that must be mentioned.

To be clear, I'm not asking the authors to modify the C-N modelling to account for both effects. I am aware that it would require an extensive work (example with OSCAR as illustrated in Gasser et al, 2017: 10.5194/gmd-10-271-2017). This manuscript already provides a significant modification. However, I suggest to explicitly mention both limits in the manuscript, and keep them for future developments of the model.

GT: Thank you for the thoughtful suggestions. Regarding the first limitation, I agree with the mechanisms you mentioned - ecosystems with different NPP will exhibit different regrowth patterns. However, implementing this in a global box model is quite challenging. In my view, it is nearly impossible to model global-averaged NPP as a whole while accounting for such dynamics.

The correction you suggested, if I understand correctly, involves manually separating part of the deforestation flux and assuming it occurs in primary forests, then calculating regrowth based on NPP. This makes sense biophysically and could work, but I have some concerns about its implementation:

a) If regrowth is linked to NPP, it would likely need to reference the NPP from the previous time step (i.e., the unperturbed NPP before deforestation). This requires a reference NPP for each deforestation event, which would change with each time step due to the independent nature of deforestation perturbations. Determining this reference NPP is complex, and it raises the unresolved question in complex models: How do we isolate the effects of different perturbations over time? In an ideal scenario, complex models would need to run an undisturbed experiment first to establish a reference state for each time step, but this is not feasible for our emulator.

b) The fraction of deforestation dependent on NPP is difficult to determine.

c) Ensuring mass conservation with this parameterization would be challenging.

d) We lack sufficient data to constrain the additional parameters needed for regrowth.

In conclusion, while the current formulation is not perfect, it represents a balance between simplicity, functionality, and the availability of calibration data.

For the second limit, I think the current formulation, with the regrowth fraction (ϕ), has somewhat captured the biomass extraction from harvest.

To avoid the discussion going too far away, here I also share some justifications for the current formulation. First of all, we have to admit the land use perturbation interacts with nearly every part of the land ecosystem and the interactions are largely heterogeneous. As an emulator focusing on the coupled carbon-nitrogen cycle, we are not aiming to tackle the complexity of land use perturbation itself. The current formulation adapts the idea of regrowth as a linear function of regrowth time (constant regrowth flux \Rightarrow integration is linear biomass growth), which is not too wrong considering that in the natural world the plant grows as a sigmoidal function. I also had discussions with Thomas Gasser from the OSCAR team, but the preliminary conclusion is, instead of further updating the formulation in MAGICC, using the direct output from OSCAR (or other sources) might be more straightforward. So in summary, we keep the current

formulation for its functionality. The further improvements will be possible only when the data is available (from ESMs, UNFCCC, LUMIP etc).

As suggested, we have now added some discussions on the limitations, as quoted below:

The formulation presented here has certain limitations. For example, it does not account for variations in regrowth rates among different ecosystems or across ecosystem successional stages - an inherent constraint of the global box model approach. Additionally, it aggregates deforestation and harvest fluxes into a single LU input, although the regrowth fraction parameter may provide some indication of harvest activities that do not result in regrowth.

4. Calibration

a. Ensemble members

It is not detailed which ensemble members are used for the calibration of the CN module in Sections 3.1-3.2. Is that only the first one, e.g. r1i1p1f1, or all available ones? If all, is there an averaging? I recommend the authors to give some information on these questions.

Additionally, is there some form of weighting on the sample from the samples from the SSPs and the historical, to account for a varying density of points in the space of predictors. To be clear, I'm not asking the authors to apply such a weighting, but I'm asking whether they apply it, and simply suggest to mention it. There are imperfect solutions, like accounting for the length of the runs, and more sophisticated ones, like the inverse of the density in the predictor space, but such solutions may not be feasible for simple climate models.

GT: Thanks for the suggestion. We have now added the ensemble information ("variant_label" in CMIP6 global attribute). We did not use all the ensemble members as they are of different realizations, initializations, physics, and forcings. The ones we used are now specified in Table 1.

b. Base period for calibration

At the moment, the base period is the first year, at least for temperatures (Lines 399-400). Due to internal variability in ESMs runs, I would recommend taking an average over a longer base period, e.g. 1851-1900.

GT: Thanks for the suggestion. For this calibration we used the values in the year 1850 for the base period (i.e., the initial state). We will consider using the average over a longer base period for our future calibrations.

5. Model design

a. CO2 fertilization

I appreciate the approach on the CO2 fertilization (Section 2.3.1), especially how to deal with an overreliance on the rectangular hyperbolic function. Yet, I would appreciate having a Figure in appendix showing the response of CO2 fertilization with CO2, for different values of the method factor, e.g. 0, 0.25, etc to 2. It would help the readers get a better idea on the impact of this parametrization.

GT: Thanks for the suggestion. We have now added the illustrative figures for the method factor for both CO2 fertilization and temperature effect on NPP.

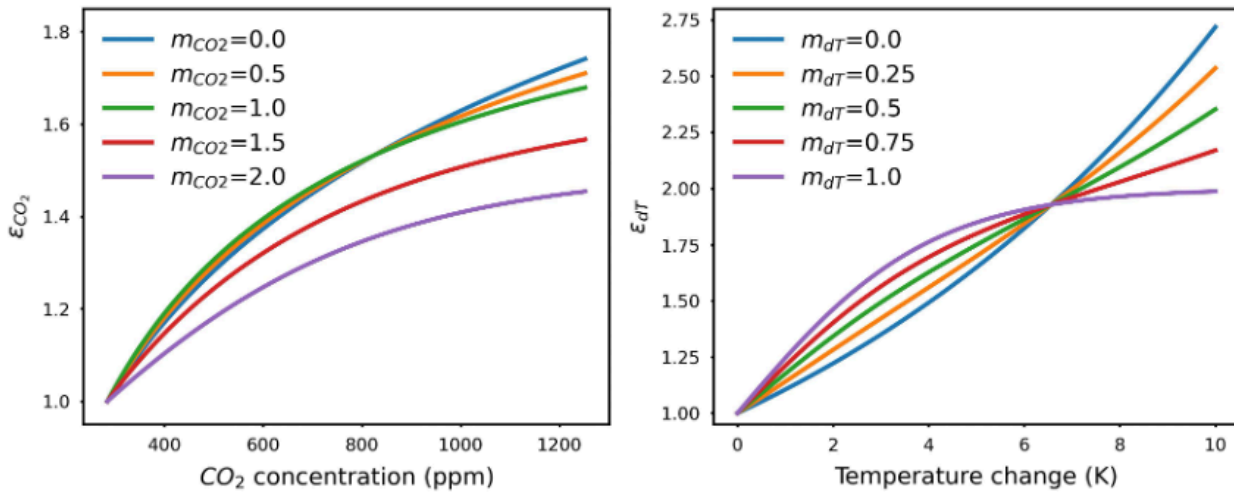


Figure A1. Illustration of the functionality of the method factor for CO₂ fertilization (m_{CO_2}) and temperature feedback (m_{dT}). An m_{CO_2} of 0, 1, and 2 represents the logarithmic, rectangular, and sigmoidal CO₂ fertilization formulations, respectively (Eqs. 17 and 18). Similarly, an m_{dT} of 0 and 1 corresponds to the exponential and sigmoidal temperature response formulations, respectively (Eqs. 19 and 20). Intermediate values represent a linear combination of the two formulations.

b. Overfitting?

The model proposed for the CN module is very well designed, I appreciate the representation of the crucial fluxes and pools in a synthetic approach. Yet, the high flexibility in the parametrizations of the fluxes make me wonder about overfitting. For instance, to what extent should the BNF flux be split between the plant, litter and soil pools? Given all fluxes being split, isn't there a risk to have a spurious & non-physical parametrization of the cycles?

To answer these questions, I would have two recommendations. First, the Table A.1 should include the significance of the coefficients, with a discussion in the manuscript. Then, the differences in the N cycles in ESMs could be further discussed, be it for the partitioning or the behaviours. Of course, an exhaustive analysis would make a full paper, but I would suggest to keep it to one paragraph.

GT: Thanks for the suggestions. This manuscript aimed to present the calibration only from the best-estimate parameters (i.e., we only extract the parameter sets for best-fitting). We were planning to do a more comprehensive analysis on the parameter uncertainties using MCMC, with which the posterior distribution would be more straightforward to see if there is overfitting, correlation, or redundancy of parameters. For nitrogen behaviors, we have done some preliminary analysis for the parameters (Section 5.3), which supported that the nitrogen effect parameters are necessarily constrained in the calibration. For the current calibration, the fractionation factors are set completely free in this calibration (the only constraint is the sum of partition==1), which is why we did not discuss them in the manuscript.

6. Performances for CMIP6?

In Section 3.2, the authors write that for MIROC-ES2L and UKESM1-0-LL, the NPP over 1pctCO₂ is higher than in SSP126, while the opposite is seen for the plant uptake. They conclude in an inconsistency in their modelling. I would argue that it is not necessarily inconsistent for two potential reasons. First, the 1pctCO₂ does not assume any change in land management, thus no increase in fertilization, while SSP126 does.

Figure 2 shows good performances for the CN module on CABLE & OCN. For CMIP6 models, Figure 3 shows a more contrasted image. The authors explain issues for instance related to the N_{min} pool of the

ESMs, but there are still important fluxes that seem not adequately modelled. For instance, MIROC- ES2L exhibit differences on the NPP.

I would be interested in seeing the comparison up to 2300, which is provided for MIROC-ES2L.

GT: Thanks for the comments. “the 1pctCO₂ does not assume any change in land management, thus no increase in fertilization, while SSP126 does.”, if I understand it correctly, it actually supports our assertion that “for MIROC-ES2L and UKESM1-0-LL, the NPP over 1pctCO₂ is higher than in SSP126, while the opposite is seen for the plant uptake. They conclude in an inconsistency in their modelling.” There is no fertilization in 1pctCO₂, the PU should be lower (agrees with the model output), the NPP should also be lower (not enough N for NPP). I have rechecked our discussion (as quoted below). Our key point here is: NPP and PU should be compatible with each other (which is the assumption of our formulation and it makes biophysical sense).

The underestimation is mainly from the inconsistent behavior of these two ESMs in the idealized 1pctCO₂ experiment and hist_SSP experiments. Both ESMs have simulated higher NPP at the end of their 1pctCO₂ runs (~100 GtC/yr in 1999 for both ESMs) than that at the end of their SSP runs (e.g., SSP126, <80 GtC/yr in 2100 for both ESMs), which is contradictory to their plant uptake results (lower in 1pctCO₂ and higher in SSPs, Fig. A4). Such behavior is in direct contradiction with our assumption that higher NPP requires a higher PU (section 2.4, Eq. 26).

We have also explained why “MIROC- ES2L exhibit differences on the NPP”, which is quoted below:

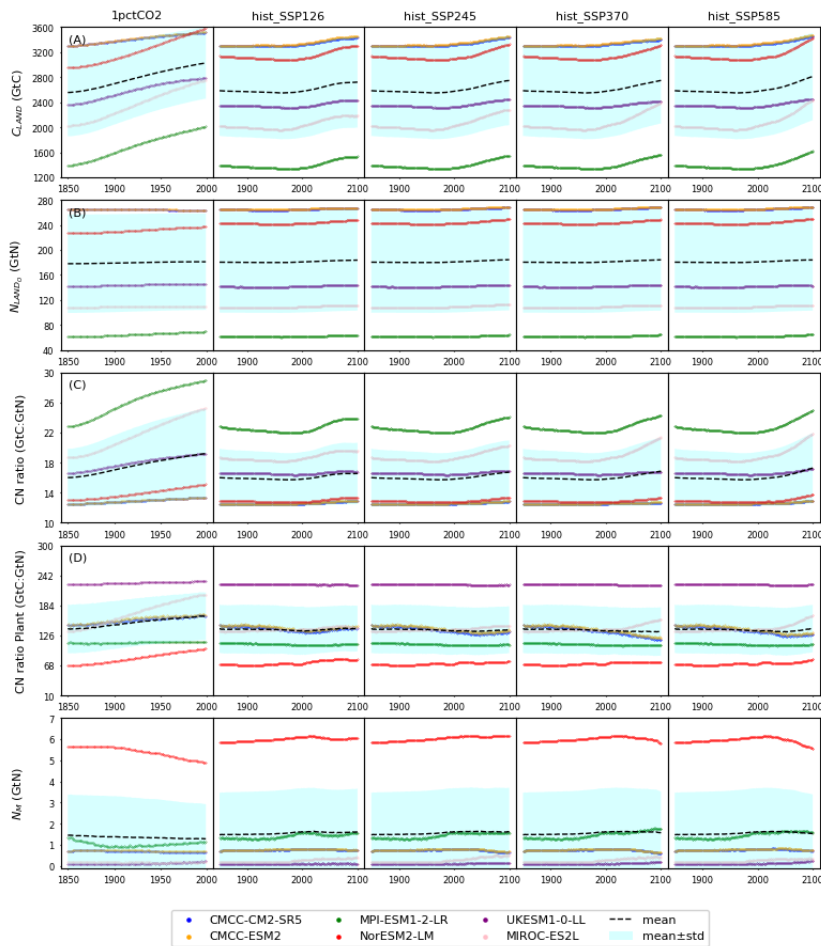
Since NPP and PU are both set as calibration targets, MAGICC has tried to minimize the gap between the emulated fluxes and targets, resulting in the simultaneous underestimation of NPP and overestimation of nitrogen plant uptake for UKESM1-0-LL and MIROC-ES2L (Fig. 3 and Fig. A4). However, such different behavior is only observed in the 1pctCO₂ experiments for these two ESMs, indicating there could be either some model response nonlinearities between their 1pctCO₂ and SSP runs that our model is not capturing or some regional distinct effects that we are not seeing in the global, annual averages.

Please also note we have provided the results to 2300 for MIROC-ES2L (where available) in Fig. A4.

7. Showing the C:N ratio of plants

I would be curious to see the C:N ratio of the plant pool in SSPs. There is one mention Line 552, but this is for the land, while I would consider the one for the plant pool to have a stronger interpretation. Current Figure A4 seems to suggest a varying C:N, in particular in 1pctCO₂. I would appreciate such a figure in the appendix, if the authors agree that it would provide worthwhile inputs for the manuscript.

GT: Thanks for the comments. For Line 552 we are talking about the delta land C / delta land N rather than the absolute land C/N to emphasize the stoichiometric relationship between carbon and nitrogen in the carbon or nitrogen accumulation (with exceptions like CMCC model's 1pctCO₂, the loss of organic N). The absolute land C/N ratio is mentioned and plotted in Fig. A6 and Text A1, where we described the varying CN ratio in the 1pctCO₂ experiment. I have tried to plot the plant C/N ratio (the figure below) but it has the same trend as the land C/N ratio (the multi-model mean, with considerable model difference).



8. Mentioning before the limit on resolutions

The aggregation to global & annual resolutions is an usual limit of the simple climate models. This is typical from these models, because their model design is not meant to analyse spatial heterogeneity, but rather the Earth system modelling through the interaction of many processes. It should be the first limit reminded in the Section 5, yet it is for now the last point in Section 5.3 (Lines 797-800). These lines do apply to the content of Section 5.3, but it applies as well to Sections 5.1 and 5.2. Thus, it would make sense to mention the issue of resolutions from the beginning of Section 5.

GT: Thanks for the suggestion. We have now moved this to the beginning of the limitation section.

9. Limits on modelling to mention as potential future works

In my opinion, the Section 6 “Conclusion and future works” should remind the limits mentioned in Section 5 as potential future works. For instance, the comments in Lines 712-720 clearly suggest that this modelling is just a first step. It is common for simple climate models to be designed that way, to start with a first simple version, and then to sophisticate where necessary. The authors mention oversimplifications, I mention others in comments 1 and 3, such limits can be future works.

GT: Thanks for the comments. We have now specified these points you mentioned in the future work part. The revision is quoted below:

Therefore, the current formulation and treatment of these aspects in MAGICC may have to be updated too, while aiming to continue to strike a balance between model simplicity, process representation, and emulation performance, reflecting a fundamental design principle for RCMs and MAGICC in particular.

Future work on MAGICC's carbon-nitrogen cycle will focus on the online calibration of the full MAGICC structure to CMIP6 ESMs (and/or observational data), evaluation of model performance with respect to computational efficiency, incorporation of additional constraints, uncertainty quantification, sensitivity analysis, application of probabilistic projections, and continued model development (e.g., land use emission implementation and nitrogen process representation) to align with advances in complex models and emerging theoretical frameworks.

Details:

10. Position of Figure 1

The Figure 1 is crucial to visualize the design of the CN module. It should appear early for the readers to structure its understanding of the model. At the moment, it is only at the very end, in Section 2.7, which is too late. I strongly suggest shifting Figure 1 to the Section 2.2 for improved clarity.

GT: Thanks. We have moved this figure and revised the writing to provide an overview of the model.

11. Difficulties in calibration due to data reporting by ESMs

I congratulate the authors for acknowledging that, and explaining how. This is a recurring issue in CMIP exercises. Although technical, it does matter a lot for calibration, and it may be useful to raise awareness on this issue.

GT: We have written another paper to discuss this issue in detail and hopefully it could get attention from the CMIP and ESM community (Investigating Carbon and Nitrogen Conservation in Reported CMIP6 Earth System Model Data, <https://doi.org/10.5194/egusphere-2024-3522>). I am also planning to contact as many CMIP6 ESMs as I can to specifically elucidate their land carbon-nitrogen data.

12. Code of MAGICC-CN

The code is well structured, relatively well commented. However, the code of MAGICC v7 itself remains openly but not anonymously available. pymagicc is available for the v6, but not the v7. The requirement for this manuscript is met, with the -CN module provided. However, I would simply suggest that future versions of MAGICC itself should be openly AND anonymously available. Additionally, development on GitHub would provide an open perspective on the developments on MAGICC and foster collaborations.

GT: Thanks for the suggestions. I will later make the carbon-nitrogen cycle publically available as a python package (though now the code is already publicly available in Zenedo). As for the MAGICC full structure, the team is working on it to make it open-sourced. The full MAGICC code is openly and anonymously available at <https://gitlab.com/magicc/magicc>.

13. RCM vs SCM

As a simple reminder, the acronym RCM may not necessarily be great for models like MAGICC, FaIR, OSCAR, HECTOR, etc. I acknowledge that we used this acronym for the RCMIP phase 1 & 2 papers, but this choice was criticized by researchers using Regional Climate Models, thus RCMs as well... At some point, the community of climate emulators should decide what to do, RCMs, or SCMs (Simple Climate Models), or else.

GT: Yea let's see. I personally do not have a preference for RCM or SCM.