Referee's report on:

Coupled Carbon-Nitrogen Cycle in MAGICC v1.0.0: Model Description and Calibration

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.

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 Nmin, the pool that provides the N. If the limitation on NPP would have depended on Nmin here, it would have affected the actual plant uptake. Yet, Nmin affects neither the limitation of the NPP nor the plant uptake. Having this disconnection between Nmin 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, Nmin would be saturated but it would have no effect either on the limitation $\epsilon_{CN(NPP)}$ or on the P uptake. Thus excessive Nmin isn't accounted for. It may be the reason for the discrepancy described Lines 457-458.

- In the current modelling, without fertilization and atmospheric deposition, the only flux going in Nmin is LD_N2M, 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 Nmin pool could be depleted by a continuous plant uptake and an insufficient decomposition flux, and then would become negative. Thus insufficient Nmin isn't accounted for. There may be a link with the issue mentioned in Lines 666-668.

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 Dalmonech, 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 Nmin.

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.

This point is the main reason for switching the manuscript from Minor to Major Correction. If I have missed something, I am sorry. I tripled-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.

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?

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.

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 τ_{rgr} (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 LU_{grsd} with C_P (thus neglecting C_L and C_S). 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.

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.

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.

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.

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.

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 Nmin 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.

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 interpreraton. 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.

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.

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.

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.

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.

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.

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.