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.

RC1: 'Comment on egusphere-2024-1941', Anonymous Referee #1, 01 Oct 2024

In general this is an interesting topic and a very valuable effort which I think warrant publication. However, the organization and writing of the paper is quite confusing and could really benefit from a good overhaul, making the paper and the code more useful and reproducible. There is also a lot of discussion on the results, but since this is a model description paper, I would have liked a lot more discussion on the modelling choices. Though some emphasis is put on the limitation of process parametrization and a global approach, the model is still quite complex and little to no time is spent discussing the additional value of this approach against something of lower complexity and with faster computational time. I would very much like to see some discussion, both of how much compute time this module adds to a typical MAGICC run, if some of the complexity could have been shaved off (maybe this is negligible and totally worth it, but from the paper I have no idea). Here is a list of additional questions that I feel should be addressed in a discussion section:

GT: Thank you for reviewing our manuscript and providing positive feedback on the topic and its value. We appreciate your constructive comments and suggestions, which will help improve the paper.

This manuscript, although originally titled as a "model description and calibration," aims to address two key aspects: (1) the model itself and its implementation and (2) the insights into nitrogen effects in ESMs as revealed by the emulation. Initially, we considered splitting these into two separate papers—one focusing on the model description and another on the nitrogen-related findings. However, we decided to combine them into a single, comprehensive paper to provide readers with a cohesive understanding of the subject. We have now revised the title to better align with the manuscript.

Regarding your specific concerns:

Modeling Choices and Complexity: The current model represents a balance between complexity, process representation, and emulation performance, as discussed in the model limitations and conclusion sections. We acknowledge that further discussion on comparisons with lower-complexity parameterizations could be beneficial. However, to balance the manuscript's length and focus, we prioritized results and discussions most relevant to the nitrogen cycle emulation.

Computational Cost: We have not yet tested the additional runtime added by the coupled carbon-nitrogen cycle model (CNit) to MAGICC runs. This is primarily because the CNit has not been implemented into the MAGICC Fortran code yet. However, the additional calculations introduced by CNit are relatively simple, and we expect the impact on runtime to be minimal. Future work will include quantitative assessments of this aspect.

1. The models being emulated show a big spread in their predictions and there is uncertain observational data. Is this the right setting and time to make an emulator of this kind? (The answer might be that this is the perfect time to have such an emulator to explore and push for better data, and be ready.)

GT: Thanks for the comment. Yes, the spread of the ESM carbon-nitrogen cycle and the lack of global observations (especially for nitrogen) are exactly the reasons for the emulator development. The emulator allows us to explore these variations systematically and provides insights that may help inform and potentially constrain the more complex models.

2. The model though simple in a way, is really quite complex with a really large number of parameters. The fact that it can emulate model behaviour relatively well using so many parameters is not very surprising. I think a more important question is whether there is enough input data to constrain such a large parameter space.

GT: Thank you for the comment. It is important to note that this paper also serves as a detailed documentation of the MAGICC carbon cycle. Apart from the nitrogen components and the updated land-use emission treatment, the remainder of the model description is derived from the previous MAGICC carbon cycle model (details available at https://doi.org/10.5194/acp-11-1417-2011). The added nitrogen cycle (consequently the parameters) is trying to find a balance between the reality (i.e., the process representation) and the computational efficiency. The successful emulation of CABLE and OCN, across diverse experiments (e.g., nitrogen on/off, varying temperature/CO_2 forcings), demonstrates that the nitrogen-related parameters are well-constrained despite the model's complexity.

3. Context. I imagine this will be used as part of MAGICCs default setup, or at least a fairly accessible version of it. How does this number of parameters add to the current number of free MAGICC parameters? Can they be fit in an online fit? Do they affect other MAGICC parameters? How easy is it to fit MAGICC parameter ensembles with this? What does it add to computational time for the full MAGICC? Etc. I see that you plan to discuss some of these issues in a different paper, but I expect at least some discussion of this here. Also can this module easily be coupled to other simple models? Perhaps the offline use demonstrated in this paper is just as interesting as online use?

GT: Thank you for the insightful questions. At this stage, we have not conducted any online calibration. Theoretically, this could be feasible given the simplicity of MAGICC. However, we generally prefer to calibrate individual modules separately rather than all at once. The primary reason is the lack of sufficient constraints to disentangle the counteracting effects of different modules within the larger model. Moreover, if such constraints were available, the difference between performing online calibration of the entire model and offline calibration of individual modules would be negligible.

The primary interaction between CNit and other components of MAGICC is through the net land-to-atmosphere carbon flux, which influences atmospheric CO_2 levels and consequently affects temperature. This temperature change then feeds back into the land carbon-nitrogen cycle, altering the land-to-atmosphere carbon flux. Beyond this interaction, CNit does not directly modify parameters in MAGICC's climate, atmospheric, or ocean modules. We have now included this information in the Section 2.1 Overview of MAGICC and CNit.

Regarding computational time, as noted in my general response, we have not yet explicitly tested the additional time required. However, we have performed coupled experiments and the increase in computational time is not noticeable (MAGICC's current computational time is around a tenth of a second and we don't notice big changes from this). Moreover, incorporating a nitrogen cycle addresses significant limitations in the model, making the added complexity worthwhile.

CNit can also function as a standalone model, requiring only inputs such as CO_2 concentration, temperature, land-use emissions, and nitrogen forcings. From this perspective, it should be straightforward to couple with other simple climate models.

To clarify these points further, we have updated the conclusion to explicitly address them. The relevant section, quoted below, discusses future work and the integration of CNit:

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 calibration of the full MAGICC structure

to CMIP6 ESMs (and/or observational data), evaluation of model performance with respect to computational efficiency and mechanistic insight, 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.

One more overall point before going over the paper from top to bottom: The structure and "story" of the paper is confusing to me. I often don't know where I am and feel like I'm scrambling for an overview when reading it. I think this can be solved without too much work using the following few principles: 1. Spoilers are great. Tell me what is going to happen and give me more of an overview on top, and on top of every section and subsection. Forcing yourself to write such miniature summaries might also help you understand how you've structured your text and see whether it makes sense. Sometimes it might not... 2. Tables are great, long lists in sentences are less so. Tables of models used, input data used to calibrate, parameters to be calibrated, experiments used etc. Also, there are so many variable abbreviations that a dictionary in the supplement would actually be useful. 3. What does MAGICC mean? It is highly unclear when you mean the full model, just this coupled carbon-nitrogen module, the new version of MAGICC the old version of MAGICC? This confusion is present in nearly every reference to the model. Please find a way to distinguish between these three things (MAGICC old version, MAGICC new version and the Coupled Carbon-Nitrogen Cycle model (MAGICC-CCNC?))

GT: Thank you for the comments and suggestions regarding the structure and clarity of the manuscript. In response:

We have added section overviews to improve the logical flow and help guide the reader through the paper. For instance, for the model description section, we now add:

The following sections outline the key components of CNit: Section 2.2 introduces the mass conservation framework and key fluxes; Section 2.3 details the NPP simulation; Section 2.4 explains carbon-nitrogen coupling, where we link 'nitrogen plant uptake' (PU) and NPP; Section 2.5 describes the litter production respiration flux; Section 2.6 focuses on carbon and nitrogen turnover calculations; and Section 2.7 addresses the implementation of land-use emissions.

For the model description section, we now add:

This section presents the offline calibration results for CNit, using prescribed land surface temperature and atmospheric CO_2 concentration from the original model outputs. We first describe the data acquisition and post-processing of land surface model outputs and CMIP6 ESM outputs (Section 3.2). Next, we define the calibration targets (major fluxes and pool sizes) and weight them to create a cost function. Finally, we apply optimization algorithms to identify the "best-estimate" parameter set (Section 3.3). For a single model, all experiments are calibrated simultaneously, resulting in one "best- estimate" parameter set that captures the model's behavior across experiments. Using these "best-estimate" parameter values, we evaluate CNit emulation against model outputs and calculate the 'root mean squared error' (RMSE) and normalized RMSE to assess model performance. The discussion of the calibration results for CABLE, OCN, CMIP6 ESMs are provided in Section 3.4 and Section 3.5.

To enhance clarity, we have incorporated tables where appropriate, such as for the models used and parameters to be calibrated.

We have standardized the terminology to avoid confusion. Specifically:

"MAGICC" now refers exclusively to the online model, representing the full structure with all components.

"CNit" is used solely to refer to the coupled carbon-nitrogen cycle module described in this paper.

Specific comments from text:

Title: Is the version number correct? Is this version number for an upcoming python version of MAGICC? Isn't MAGICC on a much higher version number. This is already confusing, and it shouldn't be.

GT: Thank you for the comment. The version number v1.0.0 specifically refers to the coupled carbon-nitrogen cycle model (CNit) described in this paper, rather than the overall MAGICC model. We have revised the title to: "Synthesizing global carbon-nitrogen coupling effects – the MAGICC coupled carbon-nitrogen cycle model v1.0."

Line 68: "Section 2 presents a detailed descritpion ... in MAGICC". The new version of MAGICC? Also, as far as I can understand it is only the land carbon-nitrogen cycle and carbon-coupling which is described and not how they actually feed in to the wider MAGICC code, though I'd prefer the text in the section to change to reflect that rather than this sentence.

GT: It is only the land carbon-nitrogen cycle model (CNit) itself rather than how it fits into the full MAGICC structure. We have revised it to "*Section 2 presents a detailed description of the CNit model.*" for clarity. We have also added descriptions on how CNit is linked with MAGICC, as quoted below. Technically CNit can be coupled with any model.

CNit is a globally integrated, annually averaged box model (Fig. 1) designed to simulate terrestrial carbon and nitrogen dynamics. It includes carbon and nitrogen pools for 'plant' (P), 'litter' (L), and 'soil' (S), along with an inorganic 'mineral' (M) nitrogen pool. The 'atmosphere' (A) exchanges carbon with the land carbon pools via 'net primary production' (NPP), 'heterotrophic respiration' (RH), and 'land-use or other anthropogenic fluxes' (LUC). Similarly, the atmosphere exchanges nitrogen with the land nitrogen pools via 'nitrogen atmospheric deposition' (AD), 'biological nitrogen fixation' (BNF), 'gaseous nitrogen loss' (LS2A), and land-use or anthropogenic fluxes LUN. CNit takes the land use emissions of carbon and nitrogen, 'nitrogen fertilizer application' (FT), AD, and BNF, as the inputs. Then, it models key fluxes and solves a system of mass conservation equations to determine the fluxes and states for carbon and nitrogen. The resulting net land-to-atmosphere carbon and nitrogen fluxes are then used to estimate atmospheric concentrations, which subsequently inform radiative forcing and climate responses. These climate responses, in turn, interact with the carbon-nitrogen cycle, creating a feedback loop (see details in Meinshausen et al., 2011a).

Line 72: "In future work...". Is this specific planned work or just an aspirational statement? Both are fine, but clarity is preferable.

GT: It is specially planned work. Thanks.

Line 75: Add a sentence or two on what the model description will involve. This is a great opportunity to prepare the reader for the overall model structure. Mass balance equations including pools for plants, soil litter and mineral nitrogen with feedbacks from temperature and CO2 etc…

GT: Thank you for the suggestion. We have now added additional description, as quoted below:

The following sections outline the key components of CNit: Section 2.2 introduces the mass conservation framework and key fluxes; Section 2.3 details the NPP simulation; Section 2.4 explains carbon-nitrogen coupling, where we link nitrogen plant uptake and NPP; Section 2.5 describes the litter production

respiration flux; Section 2.6 focuses on carbon and nitrogen turnover calculations; and Section 2.7 addresses the implementation of land-use emissions.

Line 75/ section 2 overall: The Figure 1 Flowchart shoud appear much earlier, it should be explained, and it should be annotated with equations and sections. It should also be much clearer how the code flows through the various equations. Does each section refer to a method or function or are they not built that way? This is a model description paper, I expect to get some idea about this from reading it.

GT: Thank you for your comments. We have moved Figure 1 to the beginning of the model description section for better clarity. We have also updated Figure 1 to annotate the specific sections. Additionally, we have included descriptions that provide an overview of the model, explain how the code flows through the equations, and describe its integration with the full MAGICC structure. The updated text is quoted below.

The implementation of the code flow may differ slightly from the model description presented here. For example, while the first section of the model description addresses overall mass conservation, the individual calculations prioritize solving for fluxes, which do not always align precisely with the described sections.

CNit is a globally integrated, annually averaged box model (Fig. 1) designed to simulate terrestrial carbon and nitrogen dynamics. It includes carbon and nitrogen pools for 'plant' (P), 'litter' (L), and 'soil' (S), along with an inorganic 'mineral' (M) nitrogen pool. The 'atmosphere' (A) exchanges carbon with the land carbon pools via 'net primary production' (NPP), 'heterotrophic respiration' (RH), and 'land-use or other anthropogenic fluxes' (LUC). Similarly, the atmosphere exchanges nitrogen with the land nitrogen pools via 'nitrogen atmospheric deposition' (AD), 'biological nitrogen fixation' (BNF), 'gaseous nitrogen loss' (LS2A), and land-use or anthropogenic fluxes LUN. CNit takes the land use emissions of carbon and nitrogen, 'nitrogen fertilizer application' (FT), AD, and BNF, as the inputs. Then, it models key fluxes and solves a system of mass conservation equations to determine the fluxes and states for carbon and nitrogen. The resulting net land-to-atmosphere carbon and nitrogen fluxes are then used to estimate atmospheric concentrations, which subsequently inform radiative forcing and climate responses. These climate responses, in turn, interact with the carbon-nitrogen cycle, creating a feedback loop (see details in Meinshausen et al., 2011a).

Line 76: I would like to see a flowchart of the workings of MAGICC which shows me where the new coupled carbon-nitrogen model fits in. Also this section does very little in the way of giving me an overview of the workings of MAGICC, a sentence or two to explain a flowchart would really improve this.

GT: Thank you for the comments. While CNit is designed to integrate into MAGICC, this paper primarily focuses on the carbon-nitrogen cycle itself. As explained above, the net land-to-atmosphere carbon flux from the CNit and the temperature response of the carbon-nitrogen cycle are the key interactions between CNit and MAGICC's other modules. The description added (please see the above quotes) should make this clear.

Line 85: "intial design..." give some reference to equations and sections coming up that the describe the design that you landed on, if I want to flip over and have a look from here. Also since this is an update to MAGICC, what did MAGICC have to treat this before? A sentence on that would be helpful before the initial design of this model.

GT: Thank you for the feedback. This line serves as a disclaimer about the model design principles of MAGICC (and CNit). We have added examples to clarify these principles, as quoted below. Please note there is no nitrogen cycle in MAGICC before, which is the

However, during model parameterization and refinement, some processes were simplified or integrated with others to improve efficiency. For instance, biological nitrogen fixation is directly allocated to organic nitrogen pools, bypassing the intermediate step of mineral nitrogen enrichment and subsequent plant uptake (Fig. 1). Additionally, certain representations, such as land-use emissions, were updated to achieve a balance between model simplicity and mechanistic insight (Section 2.7). These refinements align with MAGICC's design philosophy of being as simple as possible but as mechanistic as necessary.

Lines 95-110: Please be much more explicit about which equation or equations you are explaining when. Also referring to the different parts of the flowchart of figure 1 might be helpful. Again I'd very much like the supplement to include a vocabulary for reference as the amount of shorthand used is daunting. Also giving more overview would be helpful. For instance is the sentence from line 99 to 100 describing equation 4? Is equation 4 a sum of equations 1 to 3? Is equation 9 the sum of equations 5-8? Is the sentence from line 104 to 105 explaining equation 9?

GT: Thank you for pointing this out, and we apologize for any confusion caused by the writing. We have now reorganized the equations and revised the descriptions to improve clarity - now equations are always accompanied by their descriptions. Regarding the vocabulary table, we have rechecked the manuscript to ensure that all abbreviations are introduced in full the first time they appear in a new section or figure (including figure captions). This work introduces four new abbreviations—LP (litter production), LD (litter decomposition), SR (soil respiration), and LS (mineral nitrogen loss)—which are essential for formulating the equations. All other abbreviations refer to commonly used variables. The updated Table A1 provides the full names of all parameters, linking abbreviations to their complete forms, effectively serving as a vocabulary table.

Lines 110-123: Can the equations have an explanatory text to the side which would make them easier to come back to and review, such as Plant, Litter, Soil, Atmosphere to land?

GT: Thanks for the suggestion. The whole section is now revised to make it clear. Please see the Section 2.2 Carbon and nitrogen mass conservation in CNit.

Lines 162-165: In my opinion this should be a single equation with two domains or something like that \epsilon_CO2 = { and then two lines with different conditions. I find that easier to read.

GT: Thanks for the suggestion. I have revised the equations.

Line 172: That is a mouthful... A table of free parameters with a longer descriptive name/ explanation in the supplement would be a really good thing. Perhaps with possible ranges and ranges in the calibration set? Mean + std for ESMs and actual values in the OCN and CABLE calibrations?

GT: Thanks for the comment. The Table A1 (now Table A2) lists all the parameters and their values for the calibration, which is necessary because we referenced the values in the discussions. The descriptive long name and range are now added as a new table (new Table A1, as it comes before the model calibration results).

Line 226-228: I would like a reference either external or internal (i.e. to a specific upcoming section/figure) for this statement.
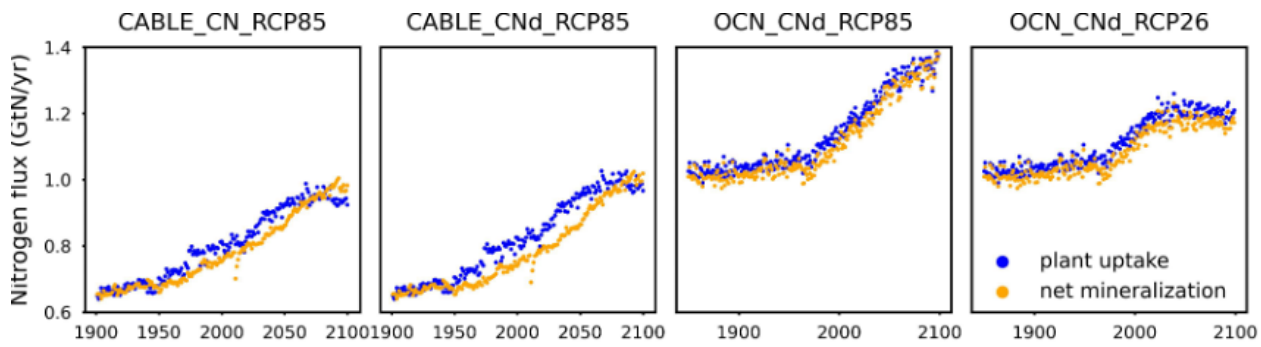
Figure A2. Relationship between nitrogen plant uptake and net mineralization as simulated by CABLE and OCN.

Lines 264-265: I assume recurring parameters are the same as presented before, but if there was a supplemental table to peruse, that would be much easier to understand.

GT: Thank you for your comment. The reason we used the full descriptive names here is exactly as you mentioned—to enhance readability. While we have now added a reference table for parameter names and ranges in the supplement (the new Table A1), we do not expect all readers to rely on the table while reading. Therefore, we retain the practice of repeating the full name and description at the start of a new section.

Lines 277-282: To me these look like they are more or less the same equation. That could be used to explain it in a schematic way as in equation 37, and these individual equations could be moved to the supplement. Even equation 36 is the same equation, only it doesn't include a carbon-nitrogen coupling because there is no corresponding carbon mineral pool. This also becomes such a long list of independent parameters, where a table would be useful, and I would like a discussion on whether so many separate parameters can be meaningfully calibrated from the available data (maybe they can and it's no big deal, then tell me).

GT: Thank you for the feedback. We have chosen not to include equations in the appendix because the model is not overly complex. While these equations are similar, they describe turnover processes for different pools, each with distinct turnover times. We will revise it soon for simplification.

As mentioned earlier, the model design principle is to keep it as simple as possible while remaining as mechanistic as necessary. From a mechanistic modeling perspective, I do not find the formulations overly complex; in fact, they might even be too simple. The temperature response is clearly important. I assume we all agree on the necessity of the carbon-nitrogen coupling effect for the turnovers, otherwise complex models would not model them and and we would not develop the emulator for them. If your concern is about the nitrogen effect in the emulator, I believe it is also critical. Without it, we risk misattributing such effects to other factors, such as the temperature response. The emulator should retain the ability to disentangle different effects, just as complex models do. Specifically regarding the nitrogen effect, it is unrealistic to require all the complex models to perform nitrogen-off experiments and explore the nitrogen effect due to the high computational cost. Therefore, using an emulator to at least quantify the effect—despite the uncertainty arising from data limitations (as discussed in our limitations section)—becomes critical.

Regarding the necessity of these free parameters and whether the data can constrain them, I do not expect all turnover processes to have the same sensitivity to plant nitrogen states or atmospheric nitrogen deposition. Since this is a global box model, nitrogen effects may behave more diversely. For instance,

using plant uptake as a single proxy for nitrogen status may not be sufficient—other factors, such as plant carbon:nitrogen stoichiometry, may also play significant roles. However, as an emulator, CNit aims to provide strong emulation performance, as shown by the results, while maintaining relative simplicity. Thus, the emulator's performance justifies these formulations to some extent. The data we have, which includes experiments with nitrogen on/off from various models and scenarios, covers a range of carbon-nitrogen cycle behaviors. This allows us to disentangle the nitrogen effect and calibrate the nitrogen parameters effectively. The data provides constraints for these parameters.

Line 289: "PU and atmospheric deposition" this statement mixes a shorthand and a full writing out of something that has a similar shorthand. Please do one or the other, at least in a single term like this. With a supplementary vocabulary, I think using the shorthands systematically would make a lot of sense. When you mix terms like this I also get confused on how to interpret "atmospheric deposition", does it now mean something else than the "atmospheric deposition = AD" shorthand? What would that be?

GT: Apologies for the confusion. We have revised the text accordingly (quoted below). As mentioned earlier, we prefer to introduce abbreviations at the beginning of each section rather than using them throughout the paper, as this approach is generally clearer for readers.

The carbon-nitrogen coupling feedback takes current 'nitrogen plant uptake' (PU) and 'nitrogen atmospheric deposition' (AD) as proxies to represent the plant nitrogen status and the nitrogen forcing, respectively.

Figure 1: I'd like to reiterate that this figure should come sooner, and that the caption, and possibly the arrows themselves could do with references to sections and equations that they depict.

GT: Thanks a lot for the suggestion. We have now moved the figure to the overview of CNit.

Line 352: Before you start the subsections, give an overview of the calibration process and datasets involved so I know what I am getting into in this section.

GT: Thanks for the comments. We have now added an overview of the calibration section, as quoted below:

3.1 Overview of the calibration process and results

This section presents the offline calibration results for CNit, using prescribed land surface temperature and atmospheric CO_2 concentration from the original model outputs. We first describe the data acquisition and post-processing of land surface model outputs and CMIP6 ESM outputs (Section 3.2). Next, we define the calibration targets (major fluxes and pool sizes) and weight them to create a cost function. Finally, we apply optimization algorithms to identify the "best-fitting" parameter set (Section 3.3). For a single model, all experiments are calibrated simultaneously, resulting in one "best-fitting" parameter set that captures the model's behavior across experiments. Using these "best-estimate" parameter values, we evaluate CNit emulation against model outputs and calculate the 'root mean squared error' (RMSE) and normalized RMSE to assess model performance. The discussion of the calibration results for CABLE, OCN, CMIP6 ESMs are provided in Section 3.4 and Section 3.5.

Lines 355-360: Are these datasets publicly available? Where can I get them? If I wanted to reproduce your paper how would I do that? This should be very clear either here or in the data availability statement. I find it in neither.

GT: Thanks. We have cited the related papers that can direct the readers to the original data source. We have now clarified this in the data availability section.

Line 364-363: "as results from experiments without the nitrogen effect are unavailable". Is this a wish for future experiments? How useful would these be? How could they aid in work similar to this? I think this might be worth revisiting in the outlook section.

GT: Thanks for the comments. This is a wish from the emulator's perspective. The only way to get a robust emulated nitrogen effect is to calibrate the CMIP6 EMSs 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. We have discussed this in the limitation part Section 5.3 The disentangled climate feedback and nitrogen effect from emulation, as quoted below:

However, the absence of nitrogen-off simulations from the CMIP6 ESMs presents challenges for direct verification. Given the computational expense of running all scenarios in nitrogen-off mode, it is recommended that ESMs perform nitrogen-off simulations for select idealized scenarios (e.g., 1pctCO2 or flat10) for diagnostic purposes.

Line 369-370: I suspect the reference is not in the right place here "calibration in this paper" refers to the current paper I am reviewing while "the original paper" is Meyerholt et al 2020. Please move it if my assumption is correct.

GT: Thanks for the reminder. Yes it should be moved to "the original paper" and now it is moved.

Line 381-384: This long listing of models, land models and references are perfect for a table. They are considerable less perfect for this sentence format.

GT: Thanks for the suggestion. We agree with it and now a table (Table 1) is added to the manuscript.

Line 392-417: This section left me with a lot of questions: 1. What experiments are used for the calibration? All? Or do you reserve some for testing? Which parameters are you fitting and which variables are you fitting to (I'd like a table or two for that in the supplement)? How do you weight the different variables when you fit? Do you fit entire timeseries? Again how do you measure the fitness of a run? Do you scale them? You say offline calibration, please define? In fact, is anything in this paper online? I could not reproduce your procedure from this text.

GT: We are sorry about the confusion of the calibration section. We have rewritten the whole section for clarifying the details. Short answers for the questions here: All experiments are used for calibration; the variables (targets) include NPP, heterotrophic respiration, nitrogen plant uptake, and carbon and nitrogen pool sizes - the entire time series; All the parameters are fitted; we defined cost function (with normalization of errors for different variables) for the optimization; The offline calibration means the CNit is not connected to the climate module or atmosphere module (i.e., temperature and CO_2 concentration are prescribed). There is nothing online in this paper. Thanks.

Line 403: I assume "imputed" should be "inputed"

GT: Thanks. We have checked and revised the typo.

Line 412-416: These three sentences read to me like a three step process. Numbering it as such might aid the understanding.

GT: Thanks. We have added "first, next, and finally" to indicate it is a step-by-step process.

Line 419- 463: I am very confused by this section. It reads mainly as a discussion of differences between CABLE and OCN with occasional references to MAGICC (which MAGICC, BTW? the offline CCNC I presume, but I would bet on it...) being able to capture them. If this is calibration results maybe this makes sense, but then the headline should reflect that. Also some summary table of how well the fits are doing would make sense to have here. It is also not clear to me whether you've used the same calibration for all experiments (I think so, but I shouldn't have to figure that out from gathering scattered clues in the text...)

Lines 472- 528: This section too seems to be about calibration results rather than the detailing the calibration. I also expect some sort of error or performance summaries. I am also still confused about what MAGICC we are talking about when and whether the results are all "offline" (whatever that means) and if so why no "online" results are included. Maybe they wouldn't change anything? Maybe it's not ready for that? I'm fine with either, but make it clear.

GT: Thanks for the comments. Both sections are indeed the calibration results - calibrating CNit to CABLE/OCN/CMIP6 ESMs. We are sorry for the confusion. We have now clarified this in the subtitles as well as in the section overview. Now the "MAGICC" is removed following our rules that MAGICC refers only to the full model structure coupled with climate, land, ocean, etc (i.e., the online model).

The new subtitles:

3.4 Calibrating CNit to CABLE and OCN: Results and comparison

3.5 Calibrating CNit to CMIP6 ESMs: Results and comparison

All the results are from the offline calibration, which is clarified in the revised calibration section. The calibration performance is presented at the beginning of each section - we quantified the error of each target variable by RMSE and normalized RMSE.

There are no online results provided. From our perspective those results would be part of a different study, which is not ready yet.

Lines 485-500: This discussion on mismatching is interesting, but I'd wish you'd take the discussion a bit further, to raise questions like: Does this mean you are perhaps fitting to the "wrong" variables? Are the data/observations/models to uncertain for the type of exercise you've done here (that doesn't mean what has been done isn't very useful, in fact in my opinion it might make it all the more interesting in fact)? Is your underlaying parametrization what has an issue? Also, some of this could have been discussed before the calibration and when you did your data selection.

GT: Thanks for the comments. When having noticeable mismatch between emulation and model output, we primarily focus on the potential causes from our formulation part (the discussion following Line 480). We also discussed the strange relationship of NPP and PU in UKESM and MIROC - higher NPP but lower PU

in 1pctCO2 (the discussion before the Line 480). However, this cannot lead to a conclusion that these models or model outputs are too uncertain because it is only found for the 1pctCO2 run.

Lines 530-536: There is so much going on in the axis here that I had trouble initially understanding the x- and y-axis setup here. Please make that clearer also in the caption.

Lines 534-535: "Diagonal dashed lines represented points where the emulation equals the target". The use of the past tense here is confusing, maybe also rewrite overall "Diagonal dashed lines represent the line where model and emulation are the same" or something like that.

GT: Thanks for the comments. Now we have added the specific explanation for the axis and the tense is changed. The changes are quoted below.

Results are normalized to a range of 0-1 using the following transformation: x-axis = (target – target_{min}) / (target_{max} – target_{min}), y-axis = (emulation – target_{min}) / (target_{max} – target_{min}). The diagonal dashed line represents points where the emulation matches the target exactly, with positions below and above the line indicating underestimation and overestimation by the emulator, respectively.

Line 437: Is this discussion or results? What are we discussing? Please give me a few sentences of summary of what discussion are upcoming, and maybe think about whether they are really discussions or just results. (Results are fine too, but they shouldn't be called discussions)

GT: Thanks for the comments. We started by pointing out the nitrogen limitation on NPP is marginal in OCN and we concluded that this minor NPP limitation still leads to considerable change in the land carbon storage. It is a result from the OCN data. Now the revised subtitle should be clear about this.

Line 539: "remain considerably different" from what? Each other? Reality? The emulation here?

GT: Now clarified with "from each other". Thanks.

Lines 549-551: I smell a table here... Also are these model means and spread? Again, this could be more easily communicated in a table.

GT: Thanks. We have claimed at the beginning of the discussion section "If not specified, the value and spread in the discussion are expressed as mean ± one standard deviation across ESMs." Table might be unnecessary here as these values are only shown once.

Line 559: I think you should probably drop "is" here.

GT: Thanks for the comments. We did not find any grammar issues with the "is" here, but we revised the sentence structure a bit for readability.

The continuous and rapid depletion of mineral nitrogen is observed in NorESM2-LM under the 1pctCO2 scenario, coinciding with the highest accumulation of organic nitrogen (Fig. 4C).

Line 589: This would be a great please to discuss whether the model presented here is perhaps too complex. Maybe it isn't, but you've done nothing to convince me.

GT: Thanks for the comments. From my perspective the discussion here emphasizes the uncertainty and complexity of the carbon cycle and nitrogen cycle. This may indicate CNit might be too simple for emulating the system rather than too complex. In emulator development, we are trying to use the least complexity model to emulate the complex system - with the consideration of necessary biophysic processes. The latter is critical because we are not only craving model simplicity but also seeking the explanation of the processes. For instance, when designing the carbon-nitrogen coupling effect for carbon turnovers [turnover = pool size / turnover time * effect (temperature) * effect (carbon-nitrogen coupling)], we inevitably introduce new parameters. Without it, the model [turnover = pool size / turnover time * effect (temperature)] may still be calibrated, but then it effectively means we misattribute the nitrogen effect on turnover to temperature effect and/or turnover time. We would like to quantify how this kind of lack of process representation can lead to the misattribution (planning work), but that would be another study. We do agree that, overall, the question of whether the model is too complex or too simple is a valid one. We also think that, given the lack of available data, it is challenging to provide a highly confident answer. For simple models, the key point is always to strike a balance between computational simplicity and mechanistic insight.

Line 620-621: Is this emulation also consistent with this for CLM?

GT: Yes, it is (as evidenced that the NPP in CMCC models is well emulated).

Line 590-636: This whole discussion is interesting, but it is not entirely clear whether the results discussed includes information gleamed specifically from the emulation or whether it is just model comparison which you could do without it. If it is some mix of those it should be clearer what insights have actually arisen from the emulation.

GT: Sorry for the confusion. The discussion is specifically based on the emulation because no ESMs run the nitrogen-off experiments (i.e., direct nitrogen effect from the ESMs is not available). We added "based on our calibration" at the beginning of all the discussion when we started talking about the findings from the emulation. The discussion here is first to present what the nitrogen effect on NPP is and the difference between different ESMs; followed by the explanation with either our emulation results or the intrinsic differences between complex models to support the emulated nitrogen effect differences. Finally we discuss the potential limitations from our emulation. We have thought about renaming the subtitle to "4.3 The emulated nitrogen effect on NPP". However, as this section mixes the emulation and model comparison, we think the original might be more suitable.

The insight from the emulation results is as follows: Given that ESMs cannot afford comprehensive nitrogen-on/off experiments to isolate the nitrogen effect, we currently have no clear understanding of the nitrogen effect in the model. However, the emulation, which demonstrates good performance, can at least provide a quantitative sense of "what the nitrogen effect is and how much difference it can cause."

Line 691: I want a summary of what's to come... Somewhere here I also want a real and open discussion on whether this level of complexity is the right one. How important is it versus how expensive is it? How well do we believe in the tuned parameters? Are there too many tuned parameters? Are some of the parameters very constrained overall (i.e. maybe they don't need to be free)? Are some all over the place (completely unconstrained)? Also I'm not even sure if this model needs specific inputs or if it can be run online with MAGICC for something like an AR6 scenario database member? Has it been tested with online MAGICC? How fast is it? How compute intensive was the calibration? For a future online run with this for impacts, what would you recommend calibrating to? All of these are questions that you don't have to have an answer to, but I expect you to acknowledge them, and say something about how they may be addressed in the future.

GT: Thanks for the comments. As explained in many of the previous replies, one key point I want to reiterate to answer your questions here is: The current model is using relatively the lowest complexity to model the key processes and effects, given that we already know there should be a "nitrogen effect" to nearly all the processes and we want to represent that in our model. The calibrations with nitrogen-off experiments provide direct constraints to the nitrogen effect. The analysis of independence of the climate effect and nitrogen effect (Section 5.3 The disentangled climate feedback and nitrogen effect from emulation) further evidenced that our parameterization is reasonable.

CNit does require further inputs like nitrogen deposition, fertilizer use and biological nitrogen fixation. We have now explicitly explained this in the model overview and calibration section (full texts quoted before so below we only quoted the key sentence).

CNit takes the land use emissions of carbon and nitrogen, 'nitrogen fertilizer application' (FT), AD, and BNF, as the inputs.

Not all the scenarios in the AR6 scenario database provided all of the inputs, which means some further assumptions or data sources will be used for the scenario exploitation - it would be another study. We have tested the CNit with MAGICC full structure for the 1pctCO2 and SSP scenarios (simply for testing the functionality so far), there is no noticeable change in speed as a result of using CNit.

We have now covered the above in the conclusion and outlook of the work.

Line 748: Will this prescribing of biological nitrogen fixation make the model less easy to run "in the wild" so to speak, with just scenario information online or offline?

GT: Thanks for the comments. Short answer to this question: yes it adds up to difficulties for the scenario use. However, our justification of not modeling it is that it already shows large differences within the same SSP scenario family, so modeling it might be challenging. In complex models (like CABLE), it is common to prescribe a constant biological nitrogen fixation. The magnitude of this flux is not that significant so even a constant assumption is an ok starting point.

We assume this is a major limitation of the current model. But as mentioned in the outlook, the modeling of biological fixation is a planned next step, which we are currently working on.

Figure A1: Maybe state explicitly that this figure is like figure 2 and what the difference between them are. Also the plots here are so close that ylabels go into each other making them hard to read.

Figure A4: Again maybe like figure 2 but with differences blahblah..

GT: Figure A1 and A4 provides the full variables calibrated to supplement Figure 2 (with only the key fluxes and pool sizes plotted). We have now revised the figure to make the y label clearer.

Lines 859-887: Text A1 - It is entirely unclear to me what this section has to do with this model and its calibration.

GT: For offline calibration the temperature is prescribed (now explicitly explained in the model calibration section). Thus, the different temperature is part of the reasons for the difference in carbon-nitrogen cycle response. We present the temperature output from CMIP6 ESMs and discuss the uncertainties in Text A1. We have now changed the title to "Text A1. The diversity of temperature output from CMIP6 ESMs", which should make it clear. Thanks.

Line 879: indented -> intended

GT: We have revised the typo. Thanks a lot for checking.

Line 904: model's or models'? (It says the former, I think maybe you mean the latter?)

GT: We have revised the typo. Thanks a lot for checking.

Line 925-928: Be more clear and explicit here, especially regarding the CABLE and OCN datasets.

GT: We have now cited the original publication for the data source. Thanks.

The model code is available at https://doi.org/10.5281/zenodo.12204422 (Tang et al., 2024). The calibration data is accessible either from the original publications [for CABLE (Fleischer et al., 2019) and OCN (Meyerholt et al., 2020)] or through the Earth System Grid Federation (ESGF, for CMIP6 ESMs), with details provided in Section 3.1 Data acquisition and processing.

I have had a look at the code, though not in great detail. In an ideal world, I'd like to see it as an importable library function, but I guess the reason why it isn't is that it will be part of a not yet publicly available MAGICC codebase. Maybe say that explicitly if that is the case. In an ideal world I'd also like more function doc-strings.

GT: Thanks for the comments, they are good questions to ask. The way we think about this is this: in this study, we are checking that the model's form and behaviour is sensible. We are testing this with a purely Python codebase for ease of exploration. Once we are happy with the setup, we will invest the time to port this to MAGICC's Fortran code base, and will move the code up to reusable, library-compatible standards at that point. We understand why having it as a library now would be helpful, but we have chosen a different order of effort for the reasons described above. Please keep an eye on the MAGICC repository to keep track of this porting activity.

The Python code provided for CNit is intended primarily to facilitate the review of its functionality. Comprehensive documentation for the code will be made available in the future, either as part of a standalone Python package or integrated with the Fortran code. We have now made this explicitly in the paper, which is quoted below:

The CNit model code is available at https://doi.org/10.5281/zenodo.12204422 (Tang et al., 2024). The Python code provided is intended primarily to facilitate the review of its functionality. Comprehensive documentation for the code will be made available in the future, either as part of a standalone Python package or integrated with the MAGICC Fortran code.