General comments

This paper fully describes the newly developed coupled carbon-nitrogen cycle to be incorporated into MAGICC, a leading methodology in the reduced-complexity climate model (RCM) category. MAGICC is one of the standard tool for climate assessment of emissions scenarios, and the new component is expected to enhance the tool's functionality and improve the quality of climate assessment. RCMs deal with the global aggregate effects of Earth system responses to given forcing changes based on complex Earth system models. Among them, the nitrogen cycle has not been adequately addressed in RCMs, and this study is the first attempt of its full-scale modeling and coupling with the carbon cycle. Despite limited base data from model experiments and relevant observations, this study conducted calibrations to adjust a number of model parameters to each of target models and validated the performance of emulations.

This study also compares and discusses the behaviors of the target models, considering underlying literature, through calibrated parameters in terms of their evolutions and inter-parameter relationships. This is an interesting analytical examination enabled by the emulator method. The findings are worth feeding back to studies on Earth system modeling, supporting observations, and process understanding.

Thus, the paper is well suitable for publication in GMD. Having said that, the manuscript may need minor revisions for further clarity and usefulness. The followings are my concerns and suggestions to be considered as appropriate.

Specific comments

Main text

L55–56. Wording of 'smaller feedback' is ambiguous to me. Does it adequately represent the effect of the nitrogen cycle mentioned in the preceding sentence?

L58. This is the first appearance of JSBACH. A brief description should be given to readers unfamiliar with this abbreviation.

L89. Balancing simplicity and performance is one important factor to consider in design. It would also be useful to indicate the extent to which the coupled carbon-nitrogen cycle would involve an increase in computational load and whether the increased parameters would cause any calibration difficulties.

L141. CO2ref definition is redundant because already defined on L137.

L342. Grassi et al. (2023, https://doi.org/10.5194/essd-15-1093-2023) may also be cited on this issue.

Figure 1. Is 'Plant P' correct? I think it is 'Plant C'. Flux partition labeling related to LU is a bit confusing because LU flux directions to the atmosphere are not consistent with those inferred from labeling, which reads 'to plant', 'to litter', or 'to soil' although the text describes the meaning in the end of 2.3. Are there any differences between '2S' between '2S_N'?

L373–391. This paragraph describes the model selection and data processes very well. Is there anything to be added about normalization to eliminate model drift or some biases in the preindustrial control?

L400–402. It seems that the extended period to 2300 applies only to SSP126 and SSP585 of MIROC-ES2L. Do the calibration results depend on the period selection? This concern arises from large differences between the model outputs and emulations in 1pctCO2.

L403. Is ‘imputed’ a typo?

L407. A paper in preparation is cited.

L411–416. Are all scenario data simultaneously used without weighting in the calibration for each model? This kind of information would be useful.

L438. Probably ‘leads to’.

L574. Citing AR6 Chapter 1, specifically Section 1.5.3, is suitable here.

L586–589. It needs a reference of the online calibration. Are Hajima et al. (2020) and Lawrence et al. (2019) suitable references in this context?

L732. It may need ‘in low SSP scenarios’ after ‘NorESM2-LM’.

L741. Citing Meyerholt et al. (2016) is more suitable at the previous sentence.

L763–764. I don’t understand how this sampling is enabled from the set of single parameter value for each model. The MCMC sampling may need supporting information.

L796. ‘flat10’ needs definition.

Appendix

Table A1. Missing values in UKESM1-0-LL need explanation.

L897–899. For the land organic nitrogen pool size, the differences between the models are too large to identify the trend from the figure.

Figure A2. It seems that the left panel shows four cases although the legend contains seven cases.

Text A1. This text does not necessarily support the discussion on Figure A5 and may be omitted. I understand that the magnitude of inter-model spread is consistent with the magnitude of forcing changes, and I don’t think that 1pctCO2 is special.

L904. Is the description about the initial condition appropriate? I think that it is an issue of ESM spin-up rather than internal variability.

Figure A7. Trivial one values may be omitted for simplicity. Missing values in UKESM1-0-LL need explanation.

Code and data availability

To ensure reproducibility, it is recommended that the processed CMIP6 outputs described in 3.1 be included in the data, and that the calibration procedures described in 3.2 be included in the code.

The calibrated data is provided in a Python pickle file, but reading the pickled object seems to require associated modules not provided.