

Response to reviewer 1

We thank the reviewer for their time in reviewing this paper. The reviewer comments are in *blue italics* and our responses are in plain black text.

Overview

The work by Davies-Barnard and colleagues evaluates the Biological Nitrogen Fixation (BNF) simulated by CMIP6 models against observation. The authors also assess the influence of BNF on projected changes in the terrestrial carbon storage simulated by the CMIP6 models under the SSP370 scenario. This work presents the principal abilities and limitations of state-of-the-art earth system models in representing BNF fostering the future improvement of BNF in earth system models. At this stage, the manuscript requires some missing information and deeper discussion before being ready for publication, as described in the following sections.

General comments

The representation of BNF is divided between direct use of Cleveland et al. (1999) parameterization and models with more complex techniques. The first case, then, represents a well-established description of BNF, while the latter depicts the effort in improving the process based representation of BNF. Are the models departing from the Cleveland et al. (1999) parameterization improving the BNF computation? Similarly, does the separate modelling of symbiotic BNF and free-living BNF improve the simulated BNF? A deeper discussion of these points would emphasize the effort needed to improve BNF representation in the current Earth System Models at global and local scales.

We deliberately avoid making a direct judgement on these issues, as while the physical process basis of the models using Cleveland et al. (1999) is weak, the more complex models are not necessarily performing better (e.g. ACCESS), particularly given the increased computational cost.

However, we see your point that there is a need to address this, and have added more direct comparisons of model performance and discussion of these issues to the Discussion and Conclusions.

In the local comparison between model and observation, are there any differences in land cover type between model and observations, and among models? In fact, models with vegetation closer to the observed one could reach better results thanks to the land cover representation despite the BNF parameterization.

While the models might be able to simulate BNF better if the vegetation closer to the observed, the reason we chose not to pursue this approach was that the difference in results of BNF between the LSMs driven by observations (including land cover) and their corresponding ESMS is small. Therefore, we would not expect the models to do significantly better with the correct land cover. The simple representation of BNF in most of these models is based on superseded results. Since the relationship between, say NPP and BNF, is weak, we would not expect any significant improvement with an improvement in vegetation type, except by chance. The main results are from the global models and global upscaled observations, and while interesting, the local comparisons are supporting evidence.

We cover this limitation in the Discussion, starting on line 229.

The authors should also present the land-only and future scenario simulations in the methodology section (section 2.1).

Added.

Moreover, the authors should explain the choice of ssp370 as the future scenario. Why don't you analyze other future scenarios?

We chose SSP3-7.0 because it had the greatest number of model simulations available, and the other SSPs had fewer model simulations available. Given that the number of ensemble members is already small, a cross-comparison might lead to biased results.

Nevertheless, it appears unlikely that there is a strong non-linearity of the response to climate change for BNF in the models that would lead to strongly alter conclusions about the uncertainty in BNF projections in the current CMIP6 ensemble. Therefore, we judge that one future scenario will provide the most important conclusions without cluttering the paper with unnecessary results that show gradations of the same results with fewer models.

Specific comments

Line 21: should be Eyring et al., 2016 instead of Taylor et al., 2012.

Done.

Line 120: SI Table 1 is not available in the supplement link of the manuscript.

Apologies, this should have been added to the SI during the upload process. It is now included.

Figure 1: "T The global" remove the initial "T"

Done.

Figure 1: add information about the panels in the caption.

Done.

Figure 1: the second row has the name of the model truncated due to the figure caption. Reshape the figure.

Done.

Lines 139-140: "which [...] tropical forest BNF". Rephrase.

Done.

SI Figure 2: second line in the caption "CM5" should be "CLM5"

Done.

SI Figure 2: Why do you compute the values over the 2005-2014 period, while the reference period is 1980-2014 in the paper?

Apologies, this is a typo and has been corrected.

SI Figure 3: explain the grey lines of panel a in the caption.

Done.

Figure 3, line 484: "reported by the paper". Which paper?

Clarified that this means the paper the observational data came from.

Figure 4: panels c, d and e are missing.

Apologies, in the conversion from draft to Biogeosciences formatting these panels were missed off. They have been re-added.

Lines 261-262: "Some models [...] nitrogen cycle". Rephrase.

Done.

Response to reviewer 2

We thank the reviewer for their time in reviewing this paper. The reviewer comments are in *blue italics* and our responses are in plain black text.

I can only make few editorial comments on this manuscript:

135-136: confusion in the numbering of figures 1b, 1c, 1d.

Fixed.

144-145: in figure 3 the units of measurement are in kgN, and in the text - in kg.

Clarified to kgN.

152-153: the marker "b" is missing in Figure 3.

Done.

187-200: figures 4c, 4d, 4e are missing in the manuscript.

In the conversion from draft to Biogeosciences formatting this part of the figure was missed off. Our apologies: it is now reattached.

470-...: we see BNF for the periods 1980-2014 (Figure 1), 1980-2015 (Figure 2), 2005-2014 (Figure SI 2). May be it would be better to average BNF estimates for the period of 1980-2014.

Apologies, this is a typo and has been corrected.

I think this is a very good student-level technical study. I do not see new scientific results in this manuscript. For these reasons, I consider this article to be inappropriate for Biogeosciences status.

We disagree with the notion that this study does not contain new scientific results and would therefore be inappropriate for Biogeosciences. The inclusion of the nitrogen cycle is one of the key innovations of CMIP6 compared to CMIP5. BNF is one of the most important factors determining the long-term trajectory of N availability and therefore the N limitation of the simulated carbon cycle. While individual assessments of individual model components have been published elsewhere, This paper is, to our knowledge, the first evaluation of the performance of these new CMIP6 models using a new synthesis of field observations and upscaled observational data.

Evaluation of climate and Earth System Models is a well-known line of scientific enquiry and has been for decades. However, routine benchmarking tools do not account for any N cycle specific database, and our manuscript thereby addresses a key gap in the literature. We hope that the demonstration of the current state of N cycle modelling in Earth System Models contributes to improved consideration of the ecological complexity of BNF in these models.