

Re-review of bg-2019-513

Nitrogen Cycling in CMIP6 Land Surface Models: Progress and Limitations

I thank the authors for making their manuscript easier to follow. The manuscript is in much better form but I am afraid there are still some errors and seemingly incorrect misinterpretations of the data that need to be addressed before it can be accepted for publication. These are mostly minor. I also have several suggestions (some of which are personal choices on how a sentence may be phrased) that I have marked on the manuscript itself whose scanned version is attached.

Major comments

1. Please report the key numbers in the abstract including that the average +CO₂ response of the models is X% compared to observations (Y%), and similarly for +N response.
2. One key analysis that is missing seems to be the comparison of late 20th century sink. It should be pretty straightforward to compare the time series of net atmosphere-land CO₂ flux from the five models with estimates from the latest Global Carbon Project (GCP) numbers for the decades of 1960s, 70s, 80s, 90s, and 2000s (<https://essd.copernicus.org/articles/11/1783/2019/>, their table 5). The range and average sink over the period 1960-2010, from the model, can also be reported in the abstract (since this is also a key number) and compared with the GCP's estimate.
3. Right now a large fraction of the Conclusions section seems like part of the Discussion since it discusses the performance of the individual models just like in the Discussion section. I think, it would be helpful if the Conclusion section is more generalized.
4. On page 9, lines 279-289. These lines discuss Figure 7a (+CO₂ response) but the text (line 279) says they discuss Figure 7b (the +N response).
5. Page 10, lines 300-303 read "Since the BNF in JULES is directly related to NPP, so the reduction in NUE indicates excess N in the system from mineralisation, possibly related to soil warming, in boreal regions with +CO₂, leading to decreased N uptake." This sentence attempts to explain the decrease in NUE of the JULES model for the +CO₂ scenario in Figure 7c. This appears to be an incorrect explanation since, I am wondering, how can the soils warm in this offline experiment which is driven with specified meteorological data, compared to the Control run. It seems there has to be some other explanation.

Minor comments

6. The colour scheme for the five models can be better. I find it hard to differentiate between CLM 4.5 and CLM5, and CLM5 and JSBACH. Also, the yellow colour LPJ-GUESS is not readable at all on the grey boxes in Figure 1.
7. Figure 2 in SI. Do the results show model minus observations, or observations minus model?

8. In context of comparing observations to model results, the manuscript doesn't explain what does "upscaling" of observations means and how it is done.

9. The phrase "dynamic vegetation" (e.g. on line 343), I think, is meant to imply competition between different plant functional (or vegetation) types. If yes, say this explicitly since prognostic LAI, for example, is also an example of vegetation dynamics.

10. Page 4, lines 118-120 read "the net ecosystem balance of N, which determines the change in the N capital available for plant growth and soil organic matter decay". This sentence doesn't read properly and, I think, is incorrect in saying that N balance is the "N capital available for plant growth". Clearly, we know that net N balance is given by

$$\Delta N = N_{dep} + BNF + N_{loss} = \Delta N_{veg} + \Delta N_{soil+litter} + \Delta N_{mineral_pools}$$

so not all of the N balance is the "capital available for plant growth".

11. Equation (1) in the manuscript, and the analysis in the paper, discusses the inputs (N_{dep} , BNF) and outputs (N_{loss}) but not the changes in pool sizes (ΔN_{veg} , $\Delta N_{soil+litter}$, $\Delta N_{mineral_pools}$). This would have been helpful in investigating how the N balance is split across the organic and inorganic pools in different models but the paper is okay without these too.

12. Page 11, lines 342 and 343 read "direct control of NPP by N availability, whereas photosynthetic C uptake (GPP) is not directly affected by N" in context of JULES and JSBACH showing little productivity response to increased N availability. Since $NPP = GPP - Ra$, I am struggling to figure, how can NPP be controlled by N availability but not GPP. The only way this can happen is Ra is controlled by N (through N content of vegetation components) in which case can this be made more clear.

Nitrogen Cycling in CMIP6 Land Surface Models: Progress and Limitations

5 Taraka Davies-Barnard^{1,2}, Johannes Meyerholt², Sönke Zaehle², Pierre Friedlingstein^{1,3}, Victor Brovkin⁴, Yuanchao Fan^{5,6}, Rosie A. Fisher^{7,8}, Chris D. Jones⁹, Hanna Lee⁵, Daniele Peano¹⁰, Benjamin Smith^{11,12}, David Wårlind^{11,12}, and Andy Wiltshire⁹

¹University of Exeter, Exeter, UK

²Max Planck Institute for Biogeochemistry, Jena, Germany

³Laboratoire de Meteorologie Dynamique, Institut Pierre-Simon Laplace, CNRS-ENS-UPMC-X, Departement de Geosciences, Ecole Normale Supérieure, 24 rue Lhomond, 75005 Paris, France

10 ⁴Max Planck Institute for Meteorology, Hamburg, Germany

⁵NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Bergen, Norway

⁶Harvard University, Cambridge, USA

⁷National Center for Atmospheric Research, Boulder, Colorado, USA

15 ⁸Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique, Toulouse, France

⁹Met Office Hadley Centre, Exeter, UK

¹⁰Fondazione Centro euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy

¹¹Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

¹²Hawkesbury Institute for the Environment, Western Sydney University, Richmond, Australia

20 *Correspondence to:* T. Davies-Barnard (t.davies-barnard@exeter.ac.uk)

Abstract. The nitrogen cycle and its effect on carbon uptake in the terrestrial biosphere is a recent progression in earth system models. As with any new component of a model, it is important to understand the behaviour, strengths, and limitations of the various process representations. Here we assess and compare five models with nitrogen cycles that are used as the terrestrial components of some of the earth system models in CMIP6. We use a historical control simulation and two perturbations to assess the models' nitrogen-related performance: a simulation with atmospheric carbon dioxide 200 ppm higher, and one with nitrogen deposition increased by 50 kg N ha⁻¹ yr⁻¹. There is generally greater variability in productivity response across models to increased nitrogen than to carbon dioxide. Compared to observations, two ~~models~~ of the models considered here have low productivity response to nitrogen, and another one a low response to elevated atmospheric carbon dioxide. In all five models individual grid cells tend to exhibit ~~toward bimodality~~, with either a strong response to increased nitrogen or atmospheric carbon dioxide, but rarely to both to an equal extent. However, this local effect does not scale to either the regional or global level. The global and tropical responses are generally better represented than boreal, tundra, or other high latitude areas. These results are due to divergent though valid choices in the representation of key nitrogen cycle processes. They show the need for better understanding and ~~more provision~~ ^{availability} of observational constraints of nitrogen processes, especially nitrogen-use efficiency and biological nitrogen fixation.

35

Compared to observations.

1

It is hard to say if model processes are valid without renewing them in detail.

Of the ESMS that contributed results to CMIP5 only two -----

1 Introduction

The terrestrial carbon (C) cycle currently removes around a third of anthropogenic carbon emissions from the atmosphere (Friedlingstein et al., 2019; Le Quéré et al., 2018). Changes in this uptake will affect the allowable emissions for targets such as limiting warming to 1.5°C (Millar et al., 2017; Müller et al., 2016). Nitrogen (N) is required to synthesise new plant tissue (biomass) out of plant-assimilated C, in differing ratios across biomes and tissue types (McGroddy et al., 2004). Therefore, future projections of terrestrial C uptake and allowable emissions are dependent on N availability, particularly under high atmospheric carbon dioxide (CO₂) conditions (Arora et al., 2019; Meyerholt et al., 2020; Wieder et al., 2015b; Zaehle et al., 2014b). A key tool for projections of allowable emissions are Earth System Models (ESMs), which project the responses of the coupled earth system (Anav et al., 2013; Arora et al., 2013; Friedlingstein et al., 2006; Jones et al., 2013). The Fifth Phase of the Coupled Model Intercomparison Project (CMIP5, Taylor et al., 2012) had numerous ESMs with a global C cycle but only two, based on the same land component, with terrestrial N cycling (Thornton et al., 2009). A number of studies with stand-alone terrestrial biosphere models (Sokolov et al., 2008; Wårlind et al., 2014; Zaehle et al., 2010; Zhang et al., 2013) as well as post-hoc assessments of CMIP5 projections suggest that predictions of terrestrial C storage would decrease by 37 – 58% if ESMs accounted for N constraints (Wieder et al., 2015b; Zaehle et al., 2014b).

Amasy The latest generation of models in CMIP6 (Eyring et al., 2016) ~~has~~ ^{to} at least six ESMs ~~that~~ ^{that contribute results to} incorporate the N cycle (Arora et al., 2019). These models employ a range of assumptions and process formulations, reflecting divergent theory and significant knowledge gaps (Zaehle and Dalmonech, 2011). Since N availability is an important source of uncertainty for the C cycle, (Meyerholt et al., 2020) ^{an} assessment of the sensitivity of the N cycle ^{in these models} to changes in atmospheric CO₂ and N inputs is required. Because of the tight coupling of C and N dynamics, a direct evaluation of the N effects on simulated C cycle dynamics using conventional model benchmarking approaches (Collier et al., 2018; Luo et al., 2012) is challenging. More insights into the magnitude of a N effect can be gained by comparing model simulations against perturbation experiments that provide evidence for the responses of terrestrial ecosystems to changes in the C and N availability (Thomas et al., 2013; Wieder et al., 2019; Zaehle et al., 2010).

In this study, we test five land surface models (LSMs) with ^{N cycle} N cycles employed in the latest generation of ESMs used in European Earth System modelling centres that contribute to CMIP6, we use a set of standardised model forcing and protocol to simulate historical changes in the C and N balance, as well as the response to N and C perturbations. The perturbation experiments (described in the methods) are designed to approximate field experiments undertaken to understand the effects of elevated CO₂ or N. These simulations reveal the overall pattern of response of the model to these forcings. We use a range of upscaled meta-analyses of observations, satellite observations, and model-to-model comparisons to assess the behaviour and performance of the models. Comparisons between models alone can also provide useful insight into the models' behaviour. The approach of assessing ESM N cycles via their corresponding offline LSMs, driven by a standardised set of model forcing, has the advantage of making model projections directly comparable while giving a representative view of the latest N cycle developments.

what does 'upscaled' means in this context?

2 Methods

70 2.1 Models

We ran simulations with five LSMs that are the land components of five different European ESMs taking part in CMIP6. The key N process formulations are summarized in Table 1. A brief description of each model follows.

The Community Land Model version 4.5 (CLM4.5; Koven et al., 2013; Oleson et al., 2010) is used in the Euro-Mediterranean Centre on Climate Change coupled climate model (CMCC-CM2; Cherchi et al., 2019). The N component is described in Koven et al., (2013) and was the first N model for ESMs, used in CMIP5 (Thornton et al., 2007, 2009). While the N cycling component of CLM4.5 is similar to CLM4 features of CLM4.5, such as leaf physiological traits (Bonan et al., 2012), were modified and there is a vertically resolved soil biogeochemistry scheme (Koven et al., 2013) as opposed to the single-layer box modelling scheme for CN in CLM4. *C and N ?* *also mention CESM2*

80 The Community Land Model version 5 (CLM5; Lawrence et al., 2019) is used in the Norwegian Earth System Model version 2 (NorESM2; Seland et al., 2020). CLM5 is the latest version of CLM and represents a suite of developments on top of CLM4.5. The N component is described in Fisher et al., (2010); and Shi et al., (2016). The key difference for the N cycle compared to CLM4 is the implementation of a C cost basis for acquiring N, derived from the Fixation and Uptake of Nitrogen (FUN) approach (Fisher et al., 2010).

85 JSBACH version 3.20 model (Goll et al., 2017) is used in the Max Planck Earth System Model version 1.2 (MPI-ESM; Mauritsen et al., 2019). The N component is described in Goll et al., (2017).

The Joint UK Land Environment Simulator version 5.4 (JULES-ES; Best et al., 2011; Clark et al., 2011) is used in the UK Earth System Model (UKESM1; Sellar et al., 2020.). The N component is described in Wiltshire et al. (forthcoming) and Sellar et al., (2020).

90 The Lund-Potsdam-Jena General Ecosystem Simulator version 4.0 (LPJ-GUESS; Olin et al., 2015; Smith et al., 2014) is used in the European community Earth-System Model (EC-Earth; Hazeleger et al., 2012). The N component is described in Smith et al., (2014).

2.2 Forcing Data and Model Initialisation

All models' pools were spun-up to equilibrium forced by pre-industrial conditions. This comprised of a constant atmospheric CO₂ concentration of 287.14 ppm, *using* ~~cycling~~ global climate data at 0.5° x 0.5° resolution for the years 1901-1930 from the CRU-NCEP dataset version 7.0 (New et al., 2000), *using* ~~assuming~~ constant 1860 land cover from the Hurtt et al., (2020) database, and 1860s nitrogen deposition from the Atmospheric Chemistry and Climate Model Intercomparison Project (Lamarque et al., 2013). Next, transient historical runs were performed for the 1861-1900 period with the same climate forcing as the spin-up, but *with time-* ~~now including~~ varying atmospheric CO₂ concentrations from synthesized ice core and National Oceanic and Atmospheric Administration (NOAA) measurements, as well as annually varying land-use from Hurtt et al., (2020). The N *repeatedly*

100 deposition is taken from the Atmospheric Chemistry and Climate Model Intercomparison Project (Lamarque et al., 2013).
The simulations were then continued for 1901-2015 ^{period} under fully dynamic forcing including climate.
The models ^{used} applied their individual soil and vegetation spin-ups ^{using all time-varying} according to their ^{protocols} respective conventions. The goal of the spin-up procedure is to obtain quasi-steady states of the ecosystem pools in relation to climate, avoiding drifting pool sizes due to lack of equilibrium, especially for slow-turnover soil organic matter pools. Because of differences among the models,
105 pool sizes after spin-up are not identical.

2.3 Model Experiments

In addition to the historical run described above (referred to hereafter as the Control), two ^{additional} experiments were ^{performed} run for the period 1996-2015: increased CO₂ (+CO₂) and increased N (+N). These two experimental runs are compared to the corresponding 1996-2015 simulations from the unperturbed Control runs. SI Table 1 provides a summary of the experiments.
110 For the increased CO₂ experiment (+CO₂) the atmospheric CO₂ concentration was abruptly increased to constant 550 ppm. This is almost twice the pre-industrial atmospheric CO₂ of 280ppm or a 200ppm increase compared to the 1996 atmospheric CO₂ of ~350 ppm, similar to free-air CO₂ enrichment experiments performed in the 1990s (Norby et al., 2005).
For the increased N experiment (+N) N deposition was abruptly increased by 50 kg N ha⁻¹ yr⁻¹, which is roughly equivalent to a number of forest N fertilisation trials (Thomas et al., 2013) and around 5 – 10 times higher than typical background N
115 deposition (Zak et al., 2017). ^{↓ expected to be} ^{→ what has been used in}

2.4 Analytical Framework

The response of the terrestrial productivity (and with it terrestrial C storage) to changes in the N cycle is in principle controlled by two components: (i) the net ecosystem balance of N, which determines the change in the N capital available for plant growth and soil organic matter decay, and (ii) the ratio of carbon production per unit N availability, which can be most
120 effectively be described as the N-use efficiency of growth. ^{→ this sentence is not entirely correct}

Because the individual processes and pools considered varies between the five models (Table 1), we use a simplified N budget to assess the annual change in the terrestrial N store (ΔN , including soil and plants):

$$125 \Delta N = N_{dep} + BNF - N_{loss} = \Delta veg_N + \Delta mineral\ pools\ (1) + \Delta\ litter + soil\ organic\ matter$$

where N_{dep} is the N deposition, BNF is the biological N fixation, and N_{loss} is the N lost from gaseous, leaching, and other pathways, as declared by the models. This paradigm assumes that increased ecosystem N input from deposition or fixation enters the soil and then becomes available for plant uptake. In a similar way, plant N uptake (N_{up}) could led to reduced N losses, which would (assuming constant N inputs) result in an apparent increase in the ecosystem N capital. Note that crop
130 fertilisation is not included here, as it is assumed to remain constant between the 3 simulations. ^{the same?}

$$\text{Units of NUE} = \frac{gC\ m^{-2}\ \text{day}^{-1}}{gN\ m^{-2}\ \text{day}^{-1}} = \frac{gC}{gN}$$

Whether and how this change in N capital affects plant growth is dependent on the magnitude of the change in plant N uptake, as well as relationship between N_{up} and NPP (whole-plant nitrogen-use efficiency; NUE; (Zaehle et al., 2014a))

$$NUE = \frac{NPP}{N_{up}} \quad (2)$$

135

where N_{up} includes plant uptake of soil inorganic N of any origin, i.e. atmospheric deposition, fertilization, decomposition of plant litter, or biological nitrogen fixation (BNF). NUE is the outcome of the product of tissue stoichiometry and fractional allocation of NPP to different tissue types, and therefore varies with changes in the allocation fractions and tissue C:N.

2.5 Observations for Comparison

We compare the models against

140 ~~We utilise~~ a range of observation-based metrics ~~for comparison to the models~~ at global and regional scales, detailed in Table 2. Most of these are based on small numbers of field studies upscaled or averaged to give an approximate global value with confidence intervals. While these upscaled values need to be interpreted with proportional caution, in the absence of more robust comparators they are useful benchmarks that can provide real-world context in addition to field scale comparisons and inter-model comparisons.

What does 'upscaling' means? Who did it?

145 3 Results

3.1 Control Run Global C and N budgets

A range of pools and fluxes from the models compared to the closest comparable observation-based data show a good performance overall and emphasises similarities between the models (Fig. 1). For GPP, all the models compare well to the MTE data (Jung et al., 2011) and when the directly comparable time period is used (see SI Fig. 2) the models are all within

150 the MTE range. The global GPP value is underlain by some regional variations between models (SI Fig. 2 and 3).

The total respiration term is similar across all the models and within a range of estimates based of the statistical upscaling of field measurements ($102 - 128\ \text{Pg C yr}^{-1}$) (Bond-Lamberty and Thomson, 2010; Bowden et al., 1993; Luysaert et al., 2007; Piao et al., 2010) but the partitioning between the autotrophic and heterotrophic respiration differs. Autotrophic respiration is overestimated by up to ~50% in all the models (Luysaert et al., 2007; Piao et al., 2010), while heterotrophic respiration is underestimated by as much as ~20% (Bond-Lamberty and Thomson, 2010). The value from Bond-Lamberty and Thomson, (2010) was reduced by 33% to account for root respiration in line with Bowden et al., (1993), and without this adjustment the discrepancy would be larger.

is this $r_h + r_a$?

155 Despite similarities in GPP, ~~N inputs~~ differ strongly between the models ~~because of widely varying biological nitrogen fixation (BNE Fig. 1)~~. N deposition is a prescribed input with small variations resulting from differences in the land-sea mask of the individual models and does not reflect uncertainties in the simulated efficiency of ecosystems to capture nitrogen

160

5 *suggest to remove this it's confusing*

at the global scale

BNF

rh

the

Not sure what does "includes NPP" means

deposition. BNF on the other hand has a wide range among models. An upscaled meta-analysis of BNF covering the period of approximately 1990 – 2019 (Davies-Barnard and Friedlingstein, 2020) has a range of 52 – 130 Tg N yr⁻¹ and only one model is outside of that range. The three models with the highest BNF (JSBACH, CLM5, and JULES-ES) use an NPP based function. While CLM5's process based function includes NPP, JULES-ES and JSBACH use an empirical large-scale correlation with NPP (Cleveland et al., (1999). LPJ-GUESS, the lowest BNF model, also uses an empirical correlation from Cleveland et al., (1999), based on evapotranspiration rather than NPP. Thus, even BNF functions from the same source (Cleveland et al., 1999) can have very different results (Wieder et al., 2015a), due to the large range of BNF functions within the source and differences in how they are implemented (Meyerholt et al., 2016). BNF dominates N input variability both because of lack of process understanding to constrain model structures and the continued uncertainty in available observations.

parameterization

this biological process

But they are for ANPP and change in woody biomass.

3.2 Modelled NPP Responses to +CO₂ Experiment

A meta-analysis of NPP responses to +200 ppm CO₂ suggests a positive response of 15.6 ± 12.8% (Song et al., 2019) and all the models are within this range (Table 3.). Other meta-analyses of productivity changes with elevated CO₂ give higher ranges of response (Table 3.) and suggest a lower limit of around 12%. Therefore, the fact that the models fall within the uncertainty bounds of the observations is equally indicative of the biases and lack of precision in the observational estimates and their upscaling as the fidelity with which the models can predict local and global response to elevated CO₂.

pls reword

CLM4.5 has a notably lower NPP response to +CO₂ than the other models, despite areas where the absolute values of NPP are low and therefore the proportional changes are large (Fig. 2). This lower response can also be seen in the absolute changes (SI Fig. 4), where the changes are consistently less than the other four models. The low response in CLM4.5 is due to a lack of mechanisms to ameliorate N limitation when C supply increases, for instance via variable C:N ratios or increased BNF (as is the case for CLM5) (Fisher et al., 2018; Wieder et al., 2019).

Despite the seeming agreement of the NPP response to +CO₂ at the global scale, the regional patterns in response vary considerably for key biomes (Fig. 2). In high latitude tundra areas, the +CO₂ response ranges between near zero (JULES-ES), very low in CLM4.5, JSBACH and LPJ-Guess to high (CLM5). In most models, this region shows sparse vegetation cover and nitrogen availability, allowing for only little increase in response to elevated CO₂, whereas the increased BNF in CLM5 facilitates a response to increasing CO₂ levels. With the exception of JULES-ES, most models predict a large +CO₂ response in very dry ecosystems with marginal productivity.

The NPP response of the equatorial region overall (SI Table 2 and SI Fig. 2) to +CO₂ ranges from 4% for CLM4.5 to 18% for CLM5 and JSBACH. Looking at latitudinal averages (SI Fig. 4) we can see the overall trends are consistent across most models, and while the percent change varies a lot, the absolute change in NPP shows considerable agreement between models, with the exception of CLM4.5. Model responses of NPP to +CO₂ in greater Amazonia however, do not reach a consensus. Comparing the response in the Amazonia region with that of coastal regions of northern South America, the JSBACH response is lower, CLM5 and LPJ-GUESS higher, and JULES-ES and CLM4.5 are approximately the same.

Patterns

This seems like a strange comparison I'm struggling to imagine what constitutes coastal regions of northern South America. Do you mean all around the northern half of the continent?

JSBACH's dip in +CO₂ NPP response at the equator (compared to surrounding areas) can also be seen in the absolute values averaged by latitude (SI Fig. 4). The process responsible for this spatial pattern is currently unclear, but may be associated with the strongly enhanced GPP simulated by the model for this region compared to observation-derived estimates (SI Fig. 2).

3.3 Modelled NPP Responses to +N Experiment

The response to +N in the models shows a binary distribution, with models either having a high (>17%) or low (<3%) level of response (Fig. 3).

A meta-analysis of NPP responses to +50 kg N ha⁻¹ yr⁻¹ suggests a positive response of 3 – 10.5% (Song et al., 2019) but none of the models are within this range (Table 3.). Other meta-analyses of productivity changes with increased N give higher ranges of response (11 – 39.8%), encompassing three of the five models (Table 3). As both a percent change and absolute change (see SI Fig. 5) JULES and JSBACH show much lower +N NPP response than the other models considered here. CLM4.5 has the highest response (23%), on account of its high initial N limitation (Koven et al., 2013).

The tundra biome response is high in CLM5 and JULES-ES, and lower but present in LPJ-GUESS and CLM4.5 (Fig. 3 and SI Fig. 5). If low NPP is excluded then the tundra mean response across models is 2 – 9% (SI Table 2), much lower than the average of observations compiled by LeBauer and Treseder, (2008) of 35% (95% confidence interval 12 – 64%). There is a high response to +N in Africa & Australia in CLM4.5, CLM5, and LPJ-GUESS, despite aridity likely limiting increase in NPP in absolute, if not relative, terms, but insufficient observations to make meaningful comparisons. One area of agreement between the models is the lack of +N response of the Amazonian region (Fig. 3) which is consistent with observations which show just a 5% non-significant +N response in tropical forests (Schulte-Uebbing and Vries, 2018). However, when other tropical regions are included the +N NPP response rise to 18 – 27% in LPJ-GUESS, CLM4.5 and CLM5, with JULES-ES and JSBACH remaining low (SI Table 2).

3.4 Comparison of NPP +N and +CO₂ Responses

It might be anticipated that there would be a relationship between the +N and +CO₂ responses, as an ecosystem (model) that is less N limited could respond more strongly to increased atmospheric CO₂ (Meyerholt et al., 2020). Since a lack of response could indicate sufficient supply or saturation of either N or CO₂, this could enable increased NPP if the area were limited by the other (C or N) nutrient. This is the case in the models at small model scales, but does not scale to either the regional or global level. The prevalent grid cell level spatial trend is bimodal, with grid cells either having a strong sensitivity to +N or +CO₂, but not both (see Fig. 4). Comparing percent change emphasises the dichotomy of +N and +CO₂ effects, with most values clustered near either zero for +N or zero for +CO₂, but SI Fig. 6 shows that there is no positive relationship or heterogeneous distribution in the absolute values either. The bias toward +CO₂ is clear for JSBACH and

Remove this sentence, perhaps.

*80% of current vegetation is C₃ which benefits from elevated CO₂
So 80% of vegetation is CO₂ limited.*

CO₂ is not a nutrient.

↑ exhibiting

Fig 5. Please make it clear that the x-axis is no longer NPP response to +N but rather VegC response to make model output comparable to observations.

JULES-ES, with most values varying in +CO₂ sensitivity but not +N (this can also be seen in the absolute anomalies, see SI Fig. 6). A slight tendency towards the reverse is true for CLM4.5, CLM5, and LPJ-GUESS, with more points having a strong +N response and a weaker +CO₂ response (Fig. 4). Altogether, LPJ-GUESS and CLM5 show the most areas with both +N and +CO₂ sensitivity. Wieder et al., (2019) found that there was a trade-off between +N impact and +CO₂ impact in CLM4, CLM4.5 and CLM5, and this seems to be true for our ensemble of models too.

The latitudinal distribution of response shows similarities across models, with high latitudes (shown in purple in Fig. 4) generally more +N sensitive, and the mid latitudes (red to orange on Fig. 4) more +CO₂ sensitive. While negative NPP values are present in both +N and +CO₂ simulations they occur in different places, with negative NPP occurring in hot arid areas for +N and cold arid areas for +CO₂ (Fig. 2, 3, and 4). In hot arid areas +N increases simulates GPP and plant growth but also plant respiration, which then exceed the additional productivity, giving a decrease in NPP. Such model behaviour has been noted before (Meyerholt et al., 2020), however, it is not evident that such a process would occur in nature. The negative values in all models except CLM4.5 also appear to have a regional bias, with a small number of grid cells responding negatively to both +CO₂ and +N in CLM5, JSBACH, and JULES-ES in the subtropics and a larger number of negative values in the subtropics in LPJ-GUESS (Fig. 4). These arid areas appear to be sensitive to neither +N nor +CO₂, due to water limitation.

We can gain further insights by considering the relationship between responses to +CO₂ and +N (Fig. 5) by forest biome. The ideal for the models is to be in the area where the observations for +N and +CO₂ intersect. Two of the models achieve this partially, JSBACH and CLM5, by having tightly clustered forest vegetation C (VegC) response to +N and forest NPP response to +CO₂. The dichotomy between +N and +CO₂ NPP response is averaged out at this scale and the models show little of the relationship between the +N response and +CO₂ response seen at the grid cell level (Fig. 4 and 5).

According to observations from N addition experiments collated in we would expect models to have biome level variation in +N response (LeBauer and Treseder, 2008; Schulte-Uebbing and Vries, 2018). Schulte-Uebbing and Vries, (2018) show that tropical forest +N VegC response is lowest and boreal and temperate forest response higher (Fig. 5). While LPJ-GUESS and CLM4.5 capture some variation between averaged biomes, none of the models have the biome responses in the correct order (Fig. 5). However, all the models except LPJ-GUESS and CLM4.5 have tropical +N response in the correct range. LPJ-GUESS is the only model to have the boreal +N response in the correct range. It is the boreal response that seems to be the main issue, as most of the models show increased +N response ^{for} compared to the tropics for temperate regions, but dampened response for boreal regions. Therefore, although the global values of response are acceptable, the relative spatial patterns show limitations in the reliability of all the models.

perhaps reword

255

3.5 N Budget Responses to +N and +CO₂

The models' responses in different components of the N budget reflect and affect their overall N sensitivity (Fig. 6). N inputs of BNF and N deposition and loss (we only consider the sum of leaching and gaseous loss so as to be consistent between

I don't see -ve values
for Δ Gas & Leach in Fig 6b
so I'm unclear what 'reduced N losses'
refers to

models) are similar between all the models in the Control simulation (Fig 6a). The uptake of N by vegetation varies more strongly between models, reflecting differing levels of N mineralisation and assumed N requirements for growth, as also reflected by the different amounts of C and N pools depicted in Fig. 1.

Changes in the N budget components to +CO₂ and +N (Fig. 6b and 6c) are not straightforwardly related to changes to productivity (Fig. 2 and 3). For instance, the weak response of NPP to +CO₂ in CLM4.5 would suggest only small changes in uptake compared to the other models (Fig. 2 and 6), however, the +CO₂ induced changes in uptake CLM4.5 are higher than that of LPJ-GUESS (Fig. 6b). Similarly, CLM5 has the largest increase in N balance for +CO₂ (Fig. 6b) amongst the models, but this does not correspond to a larger response of NPP or uptake response to elevated CO₂. Nevertheless, Fig. 6b reveals a number of important characteristics of the N cycle response to +CO₂ underlying the NPP response presented in Section 3.2. For all models except CLM5, which shows a strong response of BNF to elevated CO₂, reduced N losses are an important reason for the increased N balance of the ecosystem, which facilitates an increase in NPP in the absence of changes in ecosystem stoichiometry. For all models except CLM5, plant N uptake under elevated CO₂ is more enhanced than this change in the N balance of the ecosystem, implying a net transfer of nitrogen from the soil to vegetation.

which change?

Conversely, the N uptake changes in JULES-ES and JSBACH reflect their sensitivity of productivity to +N and +CO₂ (Figs. 2,3, and 6). For JULES-ES we can see that this is driven by changes in loss, particularly for +N, which leads to a much smaller increase in N balance in JULES-ES than the other models. In common with all the models, in JULES-ES the N loss term is a fixed fraction of the mineralisation flux and the soil N pool size. However, JSBACH has less than half the ~~increase~~ N loss of JULES-ES in the +N simulation (Fig. 6c) and almost no change in NUE (Fig. 7d). This suggests that in both JULES-ES and JSBACH there is effectively very little unmet N demand in the Control scenario but whereas JULES-ES loses the extra N, JSBACH retains it in the soil. You haven't shown Δ SoilN

BNF responses to ^{+CO₂}+N in the models differ in magnitude (Fig 7b) ^{7a} and mostly are smaller than a meta-analysis of CO₂ manipulation suggests (Liang et al., 2016). Only JULES-ES' responses, in all regions except for boreal forests, and CLM5's boreal response are within the range of the meta-analysis of observations. CLM5 is a clear outlier, with a large increase in BNF. CLM5 takes a C cost approach to BNF, which is different to the other models (Table 1), and BNF can be acquired for a relatively fixed amount of C (Houlton et al., 2008) and thus when C availability increases under +CO₂ the BNF in CLM5 increases. Fisher et al., (2018) conducted a parameter sensitivity analysis of both +CO₂ and +N fertilization which illustrates that both responses are sensitive to the maximum fraction of C which is available for fixation (a proxy for the fraction of N fixing plants and their efficiency). However, the correct parametrisation of this fraction of C available for ^{??}fixation is not well known and further field studies are required. The BNF +CO₂ response in the other four models is determined by their simple empirical BNF equations (see Table 1) based on NPP or evapotranspiration. However, new analysis suggests that simple empirical relationships cannot well represent BNF (Davies-Barnard and Friedlingstein, 2020).

The models' BNF response to +N shows one of two responses: a small increase in JULES-ES, CLM4.5, and JSBACH; or a large decrease in CLM5 and LPJ-GUESS (Fig. 7b). The latter models capture the correct BNF sign of response to +N of a decrease according to the meta-analysis of Zheng et al., (2019), though the amplitude is too large. The former models

? what does this mean?

It seems you are talking about Fig 7a, yet you mention Fig 7b and BNF response to +N.

→ These are offline simulations. How can there be soil warming in +CO₂ run?
Decrease in NUE implies less C per unit N being taken up in response to +CO₂, so some other process is responsible.

→ pls consider rewording hard to follow

295 estimate BNF as a function of NPP resulting in increased BNF whatever the source of the additional NPP is. Thus even when there is sufficient N more BNF is produced in these models when NPP increases, despite observational evidence (Zheng et al., 2019) showing this is wrong response and that facultative BNF reduces and obligate BNF is out-competed in these circumstances (Menge et al., 2009). Overall, there is little evidence for any of the BNF functions performing well, primarily due to a deficit of robust model parameterisations and parameter values.

300 The NUE responses allow comparison between models, though comparisons with observations are limited by a lack of field studies. With the exception of JULES-ES in the boreal region (Fig. 7c) all models have an increase in NUE with +CO₂ in line with the current theory of Walker et al., (2015). Since the BNF in JULES is directly related to NPP, so the reduction in NUE indicates excess N in the system from mineralisation, possibly related to soil warming, in boreal regions with +CO₂, leading to decreased N uptake. CLM4.5 has low NUE response to +CO₂ due to fixed C:N ratios, which allow little change in NUE. The other models allow either more allocation to wood or flexible C:N that results in the larger increases of NUE. CLM5 has large changes in NUE, and like JULES' boreal region this indicates a source of N other than BNF.

305 There is regional variation in models' NUE response to +N between biomes but all the models in our ensemble reduce NUE in response to +N. CLM5 and LPJ-GUESS are distinct in their larger NUE response to +N compared to the other models, but do not share the same geographical spread of response. There is little consistency between models as to which regions have the largest change in NUE. CLM5 has the largest NUE change in the temperate region, whereas in JULES it occurs in the boreal region. No empirical measurements are currently available for NUE response to +N. On the basis that scarcity encourages more frugal use of scarce resource a hypothesis could be that NUE could decrease with increased N availability, as the models show. However, water-use efficiency suggests an alternative hypothesis, as it tends to reduce during drought (Yu et al., 2017). The large variations in signal and sign of BNF and NUE responses to +N treatment between models suggest there is considerable uncertainty in our understanding.

4 Discussion

315 In this paper, we investigated the performance of five nitrogen-enabled land surface models that are part of current generation Earth System Models used in the framework of CMIP6 (Eyring et al., 2016). These new N-enabled land surface models in CMIP6 reproduce key global carbon cycle metrics. Nevertheless, despite the importance of N availability for regional productivity, there is large and unconstrained uncertainty in the magnitude of the global and regional N fluxes (Fig. 1).

320 We have focused on three general components of N-enabled models that affect the plant N uptake and eventual productivity: N inputs via BNF; NUE; and the N losses. We find that all three show considerable heterogeneity of response between models. Previous studies suggest that stoichiometric controls and the processing of soil organic matter are important for a realistic +CO₂ response (Zaehle et al., 2014a). These are essentially contributory factors to NUE, where we find large variation between models (Fig. 7). The lack of well constrained observations for global and biome-level NUE and N loss

→ what do u mean "another source". The models follow the protocol.

→ land models with representation of N cycle

→ It would have been helpful to see Δveg_N & $\Delta soil_N$

GPP & NPP typically respond very quickly compared to ΔV_{gC} so this is not needed.

imply that

pathways

325 responses ~~make~~ these areas ~~that~~ need more work. N loss is particularly challenging, as there are multiple ~~paths~~ (leaching, flooding, gaseous loss, fire, land use change, etc.) and forms (N_2O , N_2 , etc.) of loss and each model represents these in different ways. More observational studies and syntheses of existing observations are needed to quantify the nitrogen cycle in different biomes. In particular, better constraints are needed for the N cycle response to perturbations.

All the models show a global average productivity response to increased atmospheric CO_2 commensurate with those recorded in field studies. However, the regional responses and mechanisms behind this response vary widely, resulting from the interaction of the instantaneous physiological response to elevated CO_2 (e.g. Ainsworth and Long, (2005), which is embedded in all five models (but see Rogers et al., (2017)), with limitations imposed by temperature, water, light, and nitrogen, ~~as well as the response-time of vegetation dynamics~~. This large regional variance highlights the need for a more comprehensive observational data ~~base~~ to constrain responses to elevated CO_2 , particularly in under-sampled regions such as

335 the high arctic and tropical semi-arid regions (Song et al., 2019). Tundra and arctic responses ~~vary widely~~ and are associated with the representation of BNF. In LPJ-Guess and CLM5 the responses in semi-arid tropical ecosystems ~~is~~ smaller than that ~~of~~ temperate ecosystems and ~~the~~ other models, suggesting a combined effect of water- and nitrogen-limitation of soil organic matter decomposition and thus nitrogen availability that is not compensated for by changes in BNF.

in

to increased CO_2 ?

are

Hard to follow

The growth response to N addition across models is more varied. Two of the five models (JULES-ES and JSBACH) have little productivity response to increased N availability, indicating that they do not have any significant limitation of the carbon cycle by N availability (Fig. 3). There are four substantial similarities between these two models (Table 1): (i) the use of NPP to determine BNF; (ii) a direct control of NPP by N availability, whereas photosynthetic C uptake (GPP) is not directly affected by N; (iii) the use of dynamic (as oppose to prescribed) vegetation; and (iv) the assumption that N availability in pre-industrial times was sufficient to sustain the carbon cycle everywhere on land, and that observed present-day N limitation ~~was~~ a result of anthropogenic changes, most notably increased CO_2 (Goll et al., 2017).

how can this be?

345 The hypothesis ~~behind the assumption~~ of no ~~pre-industrial~~ N limitation ~~is that~~ prior to industrial times, ~~the conditions of~~ natural terrestrial ecosystems were stable for sufficient time to permit any lack of N availability to be filled by biological nitrogen fixation (Thomas et al., 2015). Consequently, JULES-ES and JSBACH first simulate a reference run without N interactions, mimicking the dynamics of the C-cycle only version of these models. After a C equilibrium has been reached

that

is based on the assumption

350 these models add a second spin-up simulation, in which C and N dynamics are coupled. While there is evidence for widespread (co-) limitation of NPP in recent decades (LeBauer and Treseder, 2008; Song et al., 2019; Vitousek and Howarth, 1991), there is insufficient data to test the hypothesis of no pre-industrial N limitation. A summary by Thomas et al., (2015) suggests reasons that pre-industrial productivity of terrestrial ecosystems was affected by ecosystem N availability, e.g. the presence of unavoidable losses to denitrification, or the competitive exclusion of nitrogen fixing species as ecosystems mature. The inability of JULES-ES and JSBACH to simulate observed N addition responses comparable to models without this assumption suggests that models assuming that pre-industrial N availability does not limit vegetation growth may be missing an important component of the nitrogen cycle constraint on the global carbon cycle. No pre-industrial N limitation

355

Do you mean competition between PFTs?

write here the observed value please

also drives other model decisions (such as N limitation not being incorporated into the GPP equation, see Table 1), which may further contribute to the models being under-sensitive to N compared to observations.

360 The models mostly represent changes in productivity from +N in high latitude northern hemisphere regions less well than other parts of the world as a percentage, as covered in the results section 3.3, Fig. 5, and SI Table 2. While the low NPP of these regions make them more likely to have high percentage increases, when these are excluded the mean +N response across the models is 2 - 9%, well below the range of a meta-analysis of observations. Thus the model response is either too low or too high. High latitude tundra is a critical but difficult to model biome because of the potential for release of methane (Nauta et al., 2015), permafrost C and N release (Anisimov, 2007; Burke et al., 2012; O'Connor et al., 2010), and albedo changes with vegetation expansion (Myers-Smith et al., 2011) and the difficulty in representing large amounts of C stored in soil. This complexity in C and N cycle is not always well understood or represented in models and therefore could limit the ability of models to provide accurate responses to perturbation. A fully integrated model that accounts correctly for all of these is not yet possible but is necessary to reduce uncertainties.

compared to

critical

Please try to reward to bring the message out more clearly

370 The greater Amazon basin is a critical area of interest for the future of the terrestrial carbon balance under climate change. Our simulations show that for most models, NPP in this area increases with +CO2, but all the models find a small or no change in NPP with +N. The small +N response is consistent with the idea that large rates of BNF in tropical rain forests make these ecosystems not strongly N limited. This result supports the idea that favourable climate conditions cause a high leaf area index (LAI) in this part of the tropics, such that there is little margin for increased NPP from +N (Fisher et al., 2018). For +CO2 there is the potential for increased NPP because of either increase in NUE or decreases in N losses, giving productivity increase without an increase in LAI. Reducing the uncertainty in NPP response to +CO2 is important, as the moist tropics represent a significant proportion of the world's aboveground biomass and therefore the size of the overall terrestrial sink will be influenced by the CO2 uptake in this biome.

critical area

Part of the uncertainty in the models comes from the reanalysis climate dataset used to drive the models. CRU-NCEP was chosen for the good spatial and temporal coverage, but some biases exist in the data compared to climatologies such as WATCH (Weedon et al., 2011). Offline simulations driven by low forcing frequency (six-hourly) CRU-NCEP data significantly overestimate evapotranspiration in regions with convective rainfall types and thereby could affect stomatal conductance and photosynthesis (Fan et al., 2019). This does not affect all the models equally, as some are known to be sensitive to the driving climatology. JSBACH, JULES-ES and LPJ-GUESS may be particularly strongly affected due to their dynamic vegetation. Lawrence et al., (2019) show that CLM5 corresponds best to benchmarks with GSWP3 forcing dataset (Hurk et al., 2016) and work with IULES shows that climate forcing is the biggest cause of variance of those considered (Ménard et al., 2015). Responses may partially be shaped by other limiting factors such as water availability, which will be handled differently between models limiting the insight on the exact processes that control model responses to change.

Model to +N and +CO2 are also affected thus

reward

As well as uncertainty in the models, the observational data also has uncertainties and limitations. Global benchmarks are approximate measures, as multi-faceted process mechanics are integrated over large domains and generalized, e.g., over climate zones that are inherently variable. Of the limited global or regional observations available, many use interpolation or

proxies such as satellite data to upscale relatively small amounts of direct observational data. In particular, the perturbed responses have uncertainties beyond the spread of the observed responses because of the small observation basis and potential biases in the geographical sampling. One of the +N global responses cited is based on 126 values from LeBauer and Treseder, (2008) but may over-estimate the global response by including high responses from young tropical soils. The NPP response to +CO₂ response for woody plants total above ground biomass (Fig. 5) is based on just 16 experiments (Baig et al., 2015), making the upscaling to biome scale less reliable than if more data were available. These meta-analyses combine measurements from a range of time periods and places, and different conditions (e.g. ~~graduated~~ ^{gradual} or instantaneous perturbations) and thus global model experiments cannot be expected to be identical. Hence statements about the marginal issues of model accuracy are unlikely to be robust as further observational constraints may alter the perspective.

unclear

unclear what are 'global model experiments'

unclear, pls reword

5 Conclusions

This is the first systematic comparison of the responses to increased N (+N) and CO₂ (+CO₂) in LSMs with terrestrial N cycle used within the CMIP6 endeavour. The five models considered here have fair overall agreement with global and tropical observations but are less robust in high latitude regions.

ad. responses to ----

yield

The models are not equally sensitive to either +CO₂ or +N, with individual grid cells tending to either be +N or +CO₂ sensitive. However, at the regional and global scale this pattern is averaged away and there is little correlation. Within this ensemble there is clear distinction between models that show strong N limitation, e.g. CLM4.5, which has a low NPP response to +CO₂, and models that show very weak N limitation, e.g. JULES-ES and JSBACH which have a low NPP response to +N. The two models with intermediate N limitation (CLM5 and LPJ-GUESS) capture the global scale response to +CO₂ and +N reasonably well. However, although CLM5 performs well by many metrics, it is an outlier compared to other models or observations in its BNF and the NUE response to CO₂ appears to be larger than supported by observations. Similarly, LPJ-GUESS captures NPP responses to +CO₂ and +N well at the global level but overestimates the vegetation C response to +N in forested tropical and temperate biomes.

this could be part of discussion

The wide range of empirical or semi-mechanistic representations for key processes such as BNF, NUE, and N loss, show how important further process understanding is for many parts of the N cycle. These parts of the models are influential, but because N cycle components are a recent addition to LSMs, fewer data are available than for carbon cycle components. Consequently, better observational constraints are required to understand whether models are working appropriately, even when the process understanding is improved.

to evaluate N cycle processes

Acknowledgements

Data used to generate the figures, plots, and tables in this paper will be archived on the University of Exeter repository upon publication, in line with the appropriate guidelines.

proxies such as satellite data to upscale relatively small amounts of direct observational data. In particular, the perturbed responses have uncertainties beyond the spread of the observed responses because of the small observation basis and potential biases in the geographical sampling. One of the +N global responses cited is based on 126 values from LeBauer and Treseder, (2008) but may over-estimate the global response by including high responses from young tropical soils. The NPP response to +CO₂ response for woody plants total above ground biomass (Fig. 5) is based on just 16 experiments (Baig et al., 2015), making the upscaling to biome scale less reliable than if more data were available. These meta-analyses combine measurements from a range of time periods and places, and different conditions (e.g. graduated or instantaneous perturbations) and thus global model experiments cannot be expected to be identical. Hence statements about the marginal issues of model accuracy are unlikely to be robust as further observational constraints may alter the perspective.

5 Conclusions

This is the first systematic comparison of the responses to increased N (+N) and CO₂ (+CO₂) in LSMs with terrestrial N cycles used within the CMIP6 endeavour. The five models considered here have fair overall agreement with global and tropical observations but are less robust in high latitude regions.

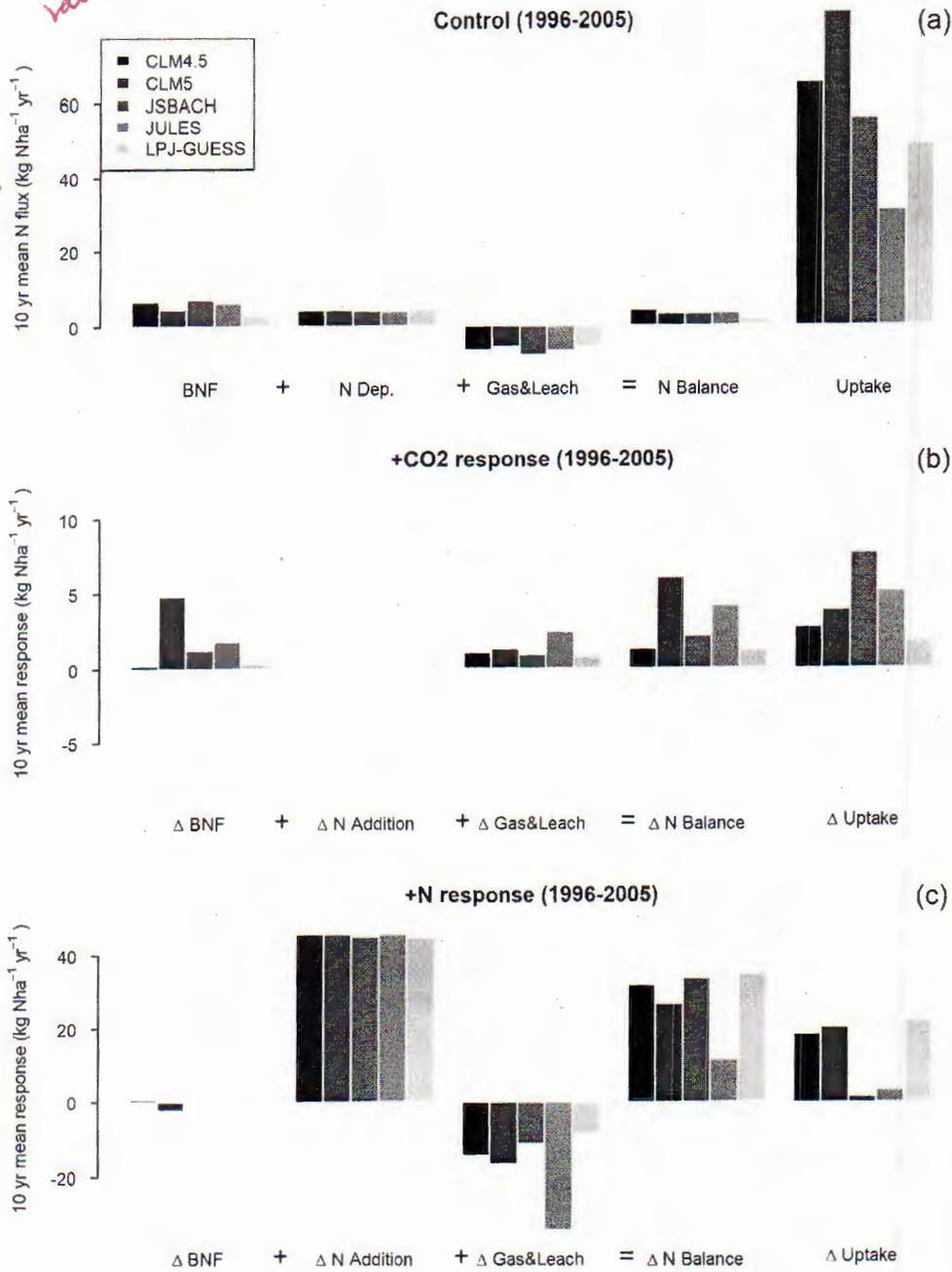
The models are not equally sensitive to either +CO₂ or +N, with individual grid cells tending to either be +N or +CO₂ sensitive. However, at the regional and global scale this pattern is averaged away and there is little correlation. Within this ensemble there is clear distinction between models that show strong N limitation, e.g. CLM4.5, which has a low NPP response to +CO₂, and models that show very weak N limitation, e.g. JULES-ES and JSBACH which have a low NPP response to +N. The two models with intermediate N limitation (CLM5 and LPJ-GUESS) capture the global scale response to +CO₂ and +N reasonably well. However, although CLM5 performs well by many metrics, it is an outlier compared to other models or observations in its BNF and the NUE response to CO₂ appears to be larger than supported by observations. Similarly, LPJ-GUESS captures NPP responses to +CO₂ and +N well at the global level but overestimates the vegetation C response to +N in forested tropical and temperate biomes.

The wide range of empirical or semi-mechanistic representations for key processes such as BNF, NUE, and N loss, show how important further process understanding is for many parts of the N cycle. These parts of the models are influential, but because N cycle components are a recent addition to LSMs, fewer data are available than for carbon cycle components. Consequently, better observational constraints are required to understand whether models are working appropriately, even when the process understanding is improved.

Acknowledgements

Data used to generate the figures, plots, and tables in this paper will be archived on the University of Exeter repository upon publication, in line with the appropriate guidelines.

Global totals rather than per unit-area values will likely be more helpful.



785

Figure 6. Global averaged 1996-2005 biological nitrogen fixation (BNF), N deposition, N loss via gaseous and leaching, the balance of those three inputs/losses, and the N uptake of the models. The top panel represents the Control scenario, and the second and third panels the response to +CO₂ and +N perturbations (see methods). Note that the y-axis scale is 4x smaller for +CO₂ response than the Control or +N response.

plant