In the following, reviewers' comments are in black, whilst our responses are in red. The text added in the revised version of our manuscript is in italics.

Anonymous Referee #1

The manuscript describes the new version of Global Terrestrial Ecosystem Model OR-CHIDEE with Nitrogen interactions integrated into the trunk. The paper is very well written and very clearly presents the description of the nitrogen cycle and its interaction with the photosynthesis and carbon/nitrogen allocation, and shows the results of the model validation and sensitivity analysis. I think the manuscript describes significant contribution to the field of modeling of carbon and nitrogen interactions in the terrestrial biosphere, and deserves to be promptly published.

We thank referee #1 for their opinion on our manuscript.

I do, however, have few minor remarks and questions, outlined below.

Page 3, line 1: typo, "pionneering" should be "pioneering"
This will be corrected in the revised manuscript.

Page 4, line 34 "Nitrogen inputs in the soil/plant system...(ii) biological nitrogen fixation and nitrogen fertilisation over managed grasslands and croplands..." I think this phrase needs some disambiguation, does it mean biological nitrogen fixation everywhere and fertilization over managed grass/crop lands, or both over managed grass/crop lands?
The sentence has been rephrased as follows:
"Nitrogen inputs in the soil-plant system are related to (i) atmospheric nitrogen deposition under the form of NHx and NOy components, (ii) biological nitrogen fixation on any land category and (iii) nitrogen fertilisation over managed grasslands and croplands."

Page 5, lines 6-7: "Furthermore, the present study considers biological nitrogen fixation rates invariant in time and computed them as a function of evapotranspiration..." if the nitrogen fixation rates are a function of evapotranspiration, and in experiments with elevated CO2 transpiration drops, does it mean that the nitrogen fixation drops as well? Based on observational evidence, is there a reason to believe that this effect is real and nitrogen fixation will drop in the elevated CO2 world? Naively, I would think that the opposite is true : more available carbon in CO2-rich world may lead to the plants being able to spent more on symbionts, increasing fixation. How does

this fixation treatment effect the differences between CNfix and CNdyn experiments presented in this paper?
Thanks for raising this issue and giving us the opportunity of clarifying what has been done. Indeed, we use a single climatology of evapotranspiration based on one reference ORCHIDEE global simulation for present-day conditions to compute the biological nitrogen fixation, for any model configuration and model simulation. Thus, biological nitrogen fixation (BNF) rates remain invariant in time but also from one model configuration to another. Because there are currently large uncertainties associated with the estimates of BNF, we think it was more suitable to assume it constant across model configurations and to analyse the modelled GPP from the different model configurations in relation with how Carbon-Nitrogen interactions are considered in each of the model configuration but not with changes on BNF. Note also that we choose to compute BNF as a function of evapotranspiration following the study of Cleveland et al. (1999) as we are still lacking for a "process-based" description of atmospheric nitrogen fixation that could be implemented in a global land surface model.
The sentence has been rephrased as follows:
"Furthermore, in the present study, the BNF rates are computed as a function of evapotranspiration following the approach of Cleveland et al. (1999). For this purpose, a single climatology of evapotranspiration, based on a global ORCHIDEE simulation for present-day conditions, is used in all simulations performed in this study. As a consequence, the differences in modelled GPP by the different model configurations (see below) cannot be attributed to changes on BNF, an approach we consider reasonable due to the large uncertainties associated with the estimates of BNF (Zheng et al., 2019)*."*

Page 9, lines 13-16: What is the length of the in-situ meteorological data? Is it enough to sample representative interannual variability? Under-sampling climate variability might lead to biases in the base state of the vegetation, and perhaps also to the biases in the responses to model treatment.

The length of the in-situ meteorological data varies from 1 to 16 years depending on the site. The mean length of the meteorological data is ~5 years and we agree with the referee that this is a relative short period. The relative short period for which meteorological data are available made us decide not to evaluate the model's capacity to simulate interannual variability. Hence our focus on seasonal variability, and long-term mean. Nevertheless, as raised by Referee #1, base state of the vegetation might be biased due to under-sampling of climate variability. In order to check if under-sampling may explain model bias, we plot the RMSE of the modelled daily GPP fluxes of each site against the length (in years) of measurement period (figure below). There is no clear relationship between RMSE and the length of the in-situ meteorological

data. We will address this issue in the manuscript by adding on page 9, line 16: *“The mean length of the meteorological data was 5 years and ranged between 1 to 16 years.”*
And Page 14 line 7: *“RMSE was plotted against the length of the in-situ meteorological data (not shown) to check whether under-sampling of the climate variability explained part of the bias. The data did not support such a relationship (correlation -0.218, CI 95% -0.448 – 0.004) “.*

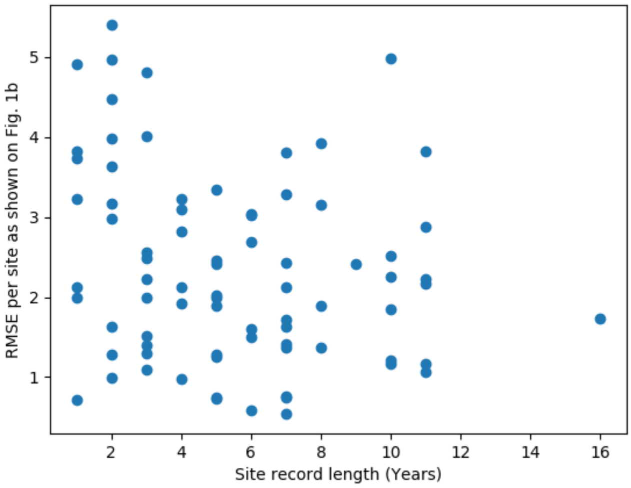


Figure R1-1 RMSE of the modelled daily GPP (gC m^{-2} day^{-1}) for each site against length of the site record in years.

Page 16, lines 1-4, Figures 5 and 6: The enhanced interannual variability on BoENF sites in CNfix simulations (and lack of this variability in respective CNdyn) looks very interesting, especially what looks like long-term oscillations in CNfix output. What can be the cause of that, in the system with less degrees of freedom than CNdyn configuration?
Thanks for raising this issue that we did not investigate in the manuscript. We should first detail how the figures 5 to 7 showing annual mean difference between the EXP-CNdyn and pd-CNdyn (respectively EXP-CNfix and pd-CNfix) simulations have been built, EXP being either 1%CO2 or 2xCO2. The EXP-CNdyn (resp. EXP-CNfix) simulations start from the pd-CNdyn (resp. pd-CNfix) simulation and run for 100 years, with only the CO2 varying, the other forcing remaining constants (as in the pd-CNdyn simulation). By simplicity, we mention that the plotted differences are between EXP-CNdyn and pd-CNdyn (resp. EXP-CNfix and pd-CNfix). However, because the pd-CNdyn (resp. pd-CNfix) are not simulations at equilibrium, and in order to avoid attributing differences which are not due to the CO2 treatment, the pd-CNdyn (resp. pd-CNfix) simulations have not been simply duplicated to reach a 100-year length. Instead, the pd-CNdyn (resp. pd-CNfix) simulations have been prolonged, with all forcings fixed as in the pd-CNdyn (resp. pd-CNfix) simulations. In this way, we ensured that the differences between the EXP-CNdyn and the prolonged

pd-CNdyn (resp. EXP-CNfix and prolonged pd-CNfix) are due only to the CO2 treatment.

The long-term oscillations that the referee #1 mentions (longer than the meteorological forcing length that we recycle) are present at three sites over the eight BoENF sites; and over two sites, they are only present in the CNfix-time configuration and mainly for the control simulation (present-day CO2 concentration). Over these sites for this specific simulation (prolonged pd-CNfix-time simulations), there are periodic oscillations over periods longer than the meteorological forcing length or nearly pseudo-chaotic oscillations. Based on our model understanding, the oscillations are due to a feedback loop between GPP, water stress and leaf age (which directly impacts the maximum photosynthetic capacity of leaves and thus GPP). Sometime, there are bifurcations points where GPP starts to be lower than “expected” (based on the recycling of the meteorological forcing) due to very small water limitation. The small GPP drop leads to increase leaf age, which in turn will tend to decrease GPP and etc. This tendency may last several years (more than the meteorological forcing length) up to a point where the negative feedbacks between GPP – leaf age – water stress stops, allowing GPP to recover (See Figure R1-2). The reason why this behaviour is only exhibited in this specific simulation pd-CNfix-time appears unclear but it cannot be attributed to difference in model version or simulation setup because we use an unique model version and the same setup for all simulations (pd-CNfix-time, 2xCO2-CNfix-time, pd-CNvar-Depvar, 2xCO2-CNvar-Depvar). It is likely linked to small differences between the two configurations in the feedback between GPP, leaf age and water stress.

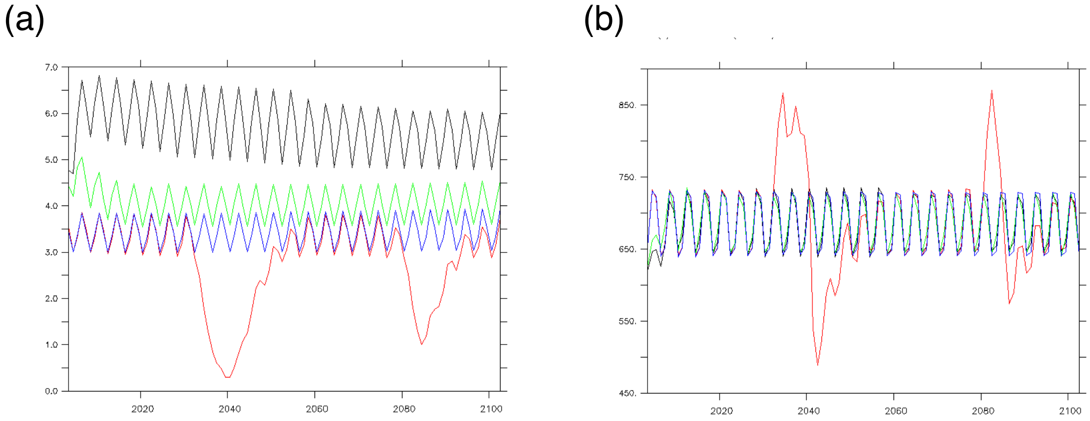
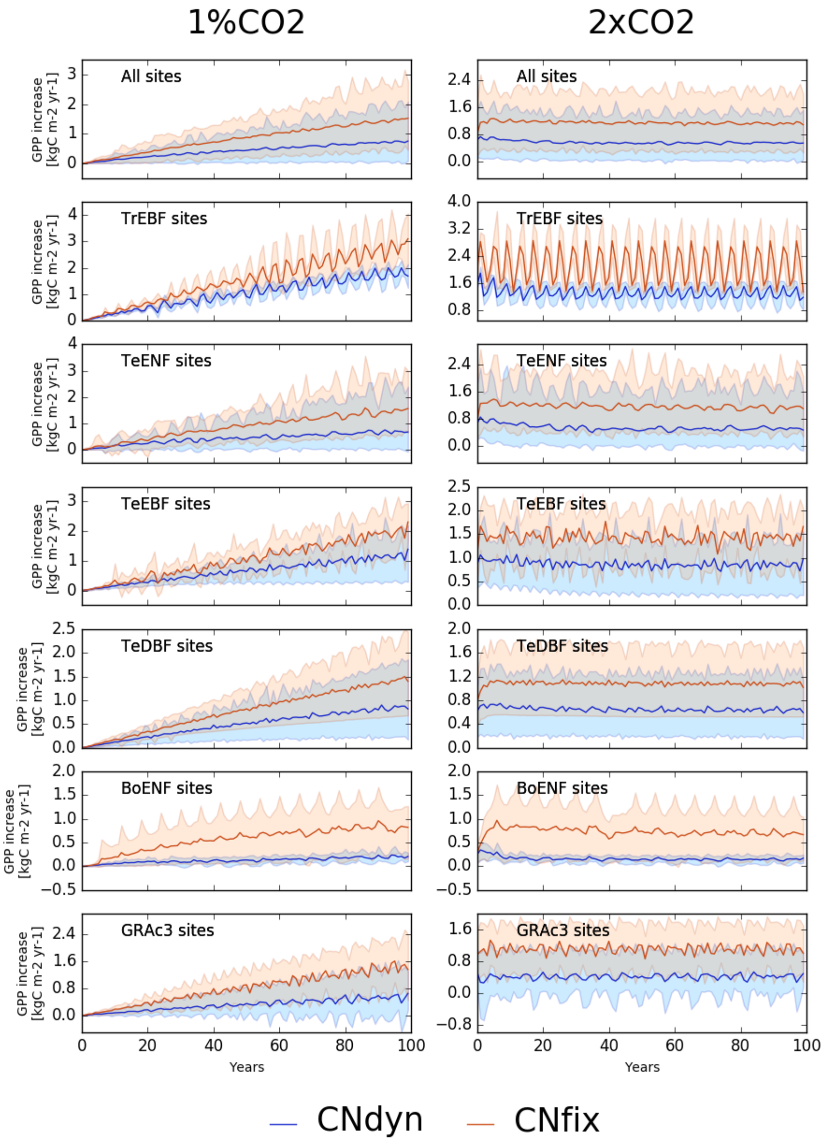


Figure R1-2 (a) Annual mean GPP (gC m^{-2} day^{-1}) for the pd-CNfix-time (red), 2xCO2-CNfix-time (black), pd-CNdyn (blue) and 2xCO2-CNdyn (green) over 100 years and (b) Leaf age (days) in the pd-CNfix-time (red), 2xCO2-CNfix-time (black), pd-CNdyn (blue) and 2xCO2-CNdyn (green) over 100 years at one Boreal Evergreen Needleleaf site

We believe that discussing this unexpected behaviour – although very interesting - goes beyond the scope of our manuscript. In order to not present differences which are not attributed to the CO2 treatment, we propose to modify the figure 5, 6 and 7, replacing only for these 3 sites, the prolonged pd-CNfix-time time series by a time series that duplicates the pd-CNfix-time time-series for 100 years. By doing so, we avoid this spurious effect that is fortunately not present in the pd-CNfix-time simulations. As an example, the modified figure 5 is here below:



Modified Figure 5 -

Page 16, lines 26-29. What is the reason for the large GPP biases of different signs in two tropical forest regions (Africa and Amazonia)?

The GPP biases over tropical forest regions are driven by different leaf C/N ratios across regions. The figure below represents grid-cell GPP of broadleaf evergreen tropical forests against leaf C/N ratio and shows that GPP is highly correlated with C/N ratio. However, it is still unclear what is/are the primary drivers of the spread of the leaf C/N ratio and consequently of GPP. First analysis seems indicating that the GPP spread is partly explained by different NOx deposition rates (see figure below). Although there is certainly a combination of additional drivers, which have not been yet identified, there is no such relationship between tropical GPP and BNF, nor between GPP and NHx deposition.

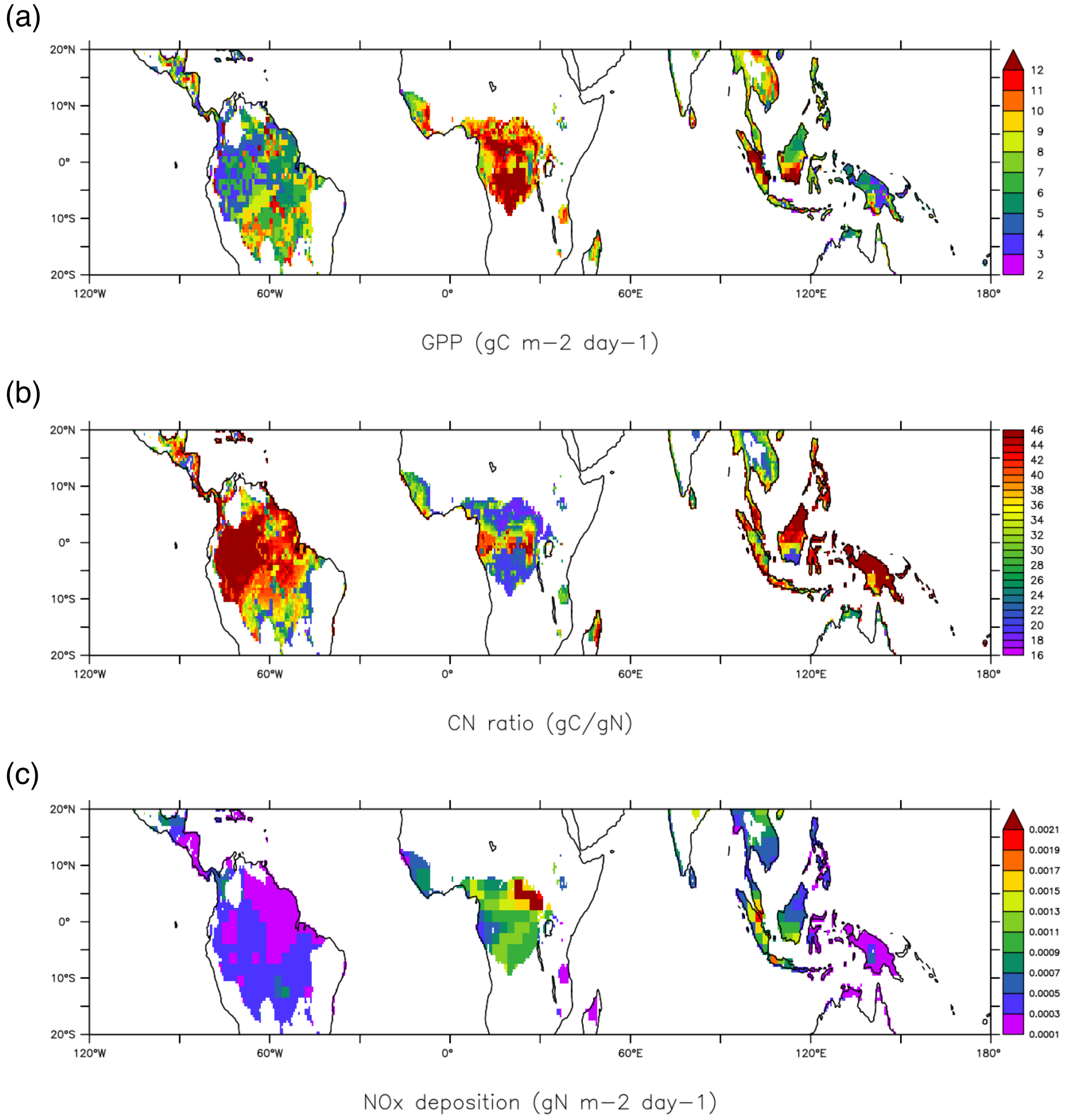


Figure R1-3 Mean annual GPP (a), mean annual CN ratio (b) and mean annual NOx deposition (c) for the Tropical Broadleaf Evergreen forests in regions where annual precipitation is higher than 1300 mm per year for the period 2011-2016.

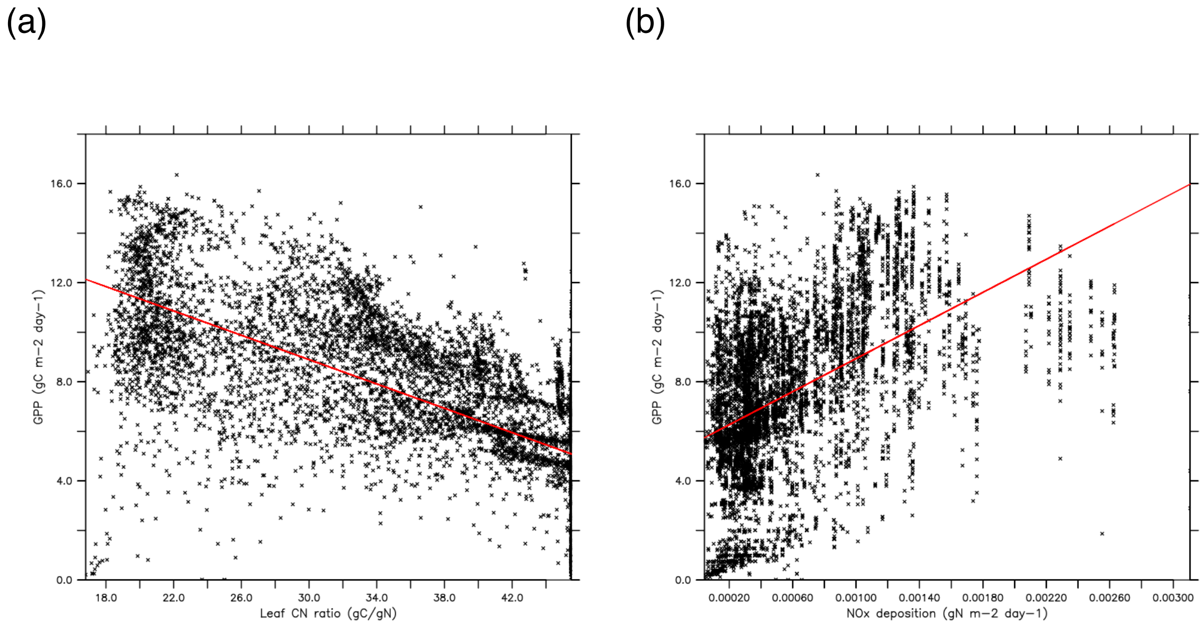


Figure R1-4 Mean annual GPP against mean annual leaf CN ratio (a) and mean annual GPP against mean annual NOx deposition (c) for the Tropical Broadleaf Evergreen forests for pixels where annual precipitation is higher than 1300 mm per year for the period 2011-2016.

We propose to add the following sentences in the discussion section to provide this information (page 19 line 5 of the initial version):
"Nevertheless, GPP appears significantly biased – against MTE-GPP - over the tropical regions, with positive biases in Central Africa and negative ones in Amazonia. One may note that these biases are not so contrasted in the original ORCHIDEE version without the nitrogen cycle (r3977, see Figure S3). Further analyses showed that the GPP biases over tropical forest regions are driven by different leaf C/N ratios across regions. However, it remains unclear what are the primary drivers of the spatial variation of the leaf C/N ratio and consequently of GPP. One of the drivers is likely to be NOx deposition, which is lower in Amazonia compared to Central Africa (not shown). There is no such relationship between GPP and BNF rate, nor between GPP and NHx deposition rate, in tropical regions. The drivers and/or processes that are responsible for turning large scale differences in NOx deposition into fine-scale differences in GPP are yet to be identified."

Page 41, line 9: typo, "S1-CNdy" should be "S1-CNdyn"
This will be corrected in the revised manuscript.

A general question: How does geographical distribution of GPP biases compare with the original ORCHIDEE model? How does it translate in the biases in other biophysical characteristics, such as biomass or LAI? I understand that the main focus of this manuscript is GPP, but I think it would be beneficial to the reader if some other results were shown too, at least from the global simulation. Unless the authors plan further publications which would address validation of the presented model version in a broader sense, of course.

Thank you for the suggestion. Although our initial intention was not to focus on comparing model versions - as the models are too different to attribute different model behaviour to newly added processes. For example, the carbon allocation scheme of the ORCHIDEE model presented in this study is very different from the allocation scheme of the ORCHIDEE version without the N-cycle. Even if the current and previous versions of the model are run without accounting for the nitrogen cycle, results may differ due to differences in the carbon allocation scheme. Hence, we thought it was more interesting to specifically investigate the role of leaf C/N dynamics within the current version with the Nitrogen cycle and the C/N interactions. We realize this focus may have been a bit to narrow and, therefore, propose to update figures 8 and 9 showing the difference between modelled and observed GPP, by adding information for the original ORCHIDEE model from which we started the development of the nitrogen cycle (revision 3977 without N cycle). This will provide a comparison of how model biases have evolved owing to all developments that were necessary to implement the N-cycle, for example, a new allocation scheme.

In the revised manuscript, we also propose to show in a supplementary figure (Figure S1) - similar to Figure 1 - the GPP model/data comparison at the site level for the former trunk version (r3977).

Although we will use this model version in future studies, we also propose to add two figures similar to figures 8 and 9 but for LAI (Figure S3 and S4). They will compare modelled LAI for S1-CNdyn simulation to the LAI provided by the GIMMS data, in terms of spatial distribution (Fig. S3) and of time evolution of the annual mean LAI for the globe and three latitudinal bands (Fig. S4). In these 2 figures, we will also add the LAI for the initial ORCHIDEE version (rev 3977).

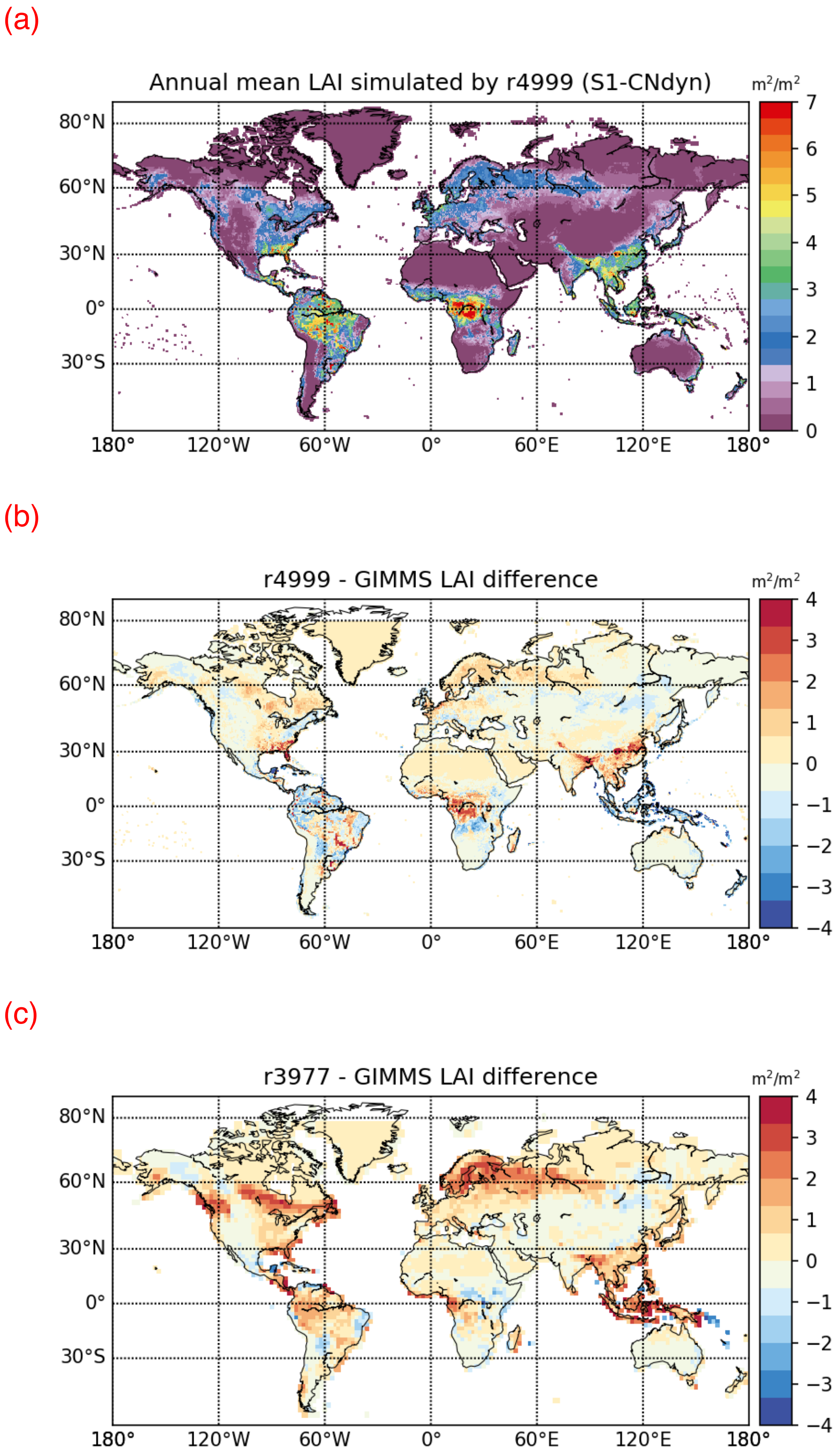


Figure S3: Global scale evaluation of ORCHIDEE against the observation-based GIMMS product. (a) Global distribution of the simulated annual mean LAI by ORCHIDEE r4999 (m^2 m^{-2}) over 2001-2010; (b) Global distribution of the difference between the simulated annual mean LAI by ORCHIDEE r4999 and the GIMMS product; (c) Global distribution of the difference between the simulated annual mean LAI by ORCHIDEE r3977 and the GIMMS product

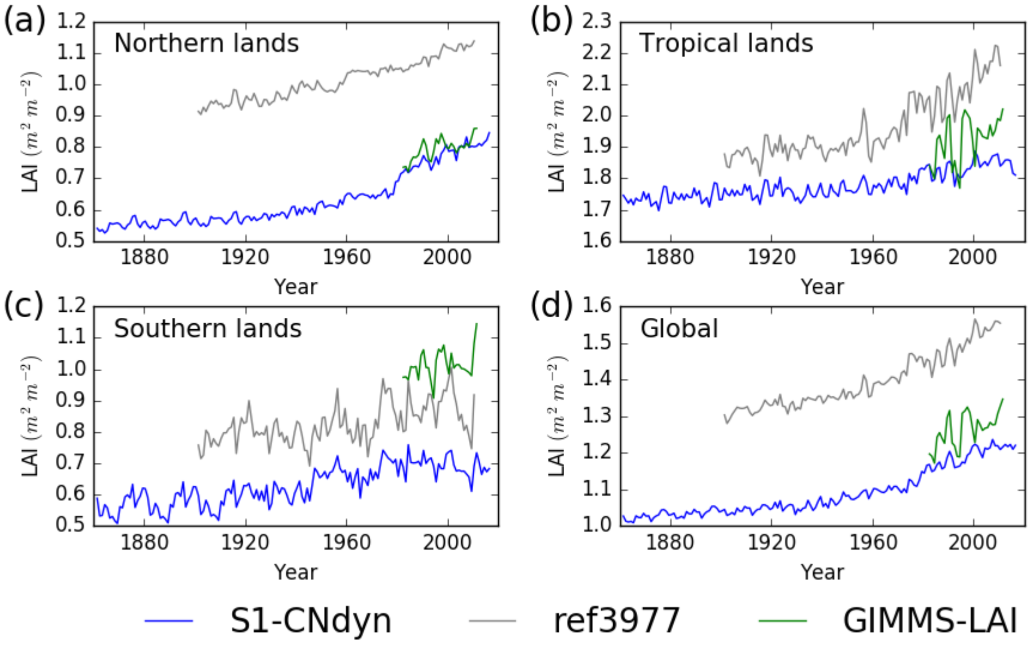


Figure S4: Evaluation of LAI from ORCHIDEE against the observation-based GIMMS product for four regions. Time evolution of the annual mean LAI ($m^2\ m^{-2}$) estimated by ORCHIDEE r4999 (in blue) and ORCHIDEE r3977 (in grey) and by the observation-based GIMMS product (in green) for (a) Northern lands (>25°N), (b) Tropical lands (<25°N and >25°S), (c) Southern lands (<25°S) and (d) all lands.

The following sentences will be added on page 17 line 16 of the initial version to present the results about the LAI global distribution and mean annual values averaged per latitudinal regions:
"Similarities between the simulated global distributions and biases in GPP and LAI (compare Fig. S3a to Fig. 8a, and Fig. S3b to Fig. 8b) suggest that the bias in GPP originates from the bias in LAI rather than from more fundamental issues with the calculation of GPP. The model/data agreement for LAI when averaged per latitudinal band is comparable to the one for GPP, with a good agreement for the Northern and Tropical lands and model underestimation in the Southern lands."

The following sentences will be added on page 17 line 16 of the initial version to compare performances of the rev3977 and rev4999 at simulating LAI and GPP at global scale.
"The agreement between the modelled and observed annual mean LAI and GPP summed over three latitudinal bands as well as at the global scale was

higher for r4999 (ie S1-CNdyn simulation) compared to r3977 without the nitrogen cycle (see Fig. 9 and S4). r3977 systematically overestimated LAI and GPP for any region, except for the Southern lands where r3977 provided similar values than the GIMMS and MTE-GPP products, respectively. Compared to GIMMS and MTE-GPP products, gridded annual mean LAI and GPP values simulated by r3977 were overestimated in the Northern lands with biases exceeding those found in r4999. On the opposite, biases of the r4999 were higher than those of r3977 in the tropical regions, in particular in Central Africa (see Fig. 8 and S3).”

References:

Zheng, M., Zhou, Z., Luo, Y., Zhao, P. and Mo, J.: Global pattern and controls of biological nitrogen fixation under nutrient enrichment: A meta-analysis, Glob. Chang. Biol., gcb.14705, doi:10.1111/gcb.14705, 2019.