Investigating Carbon and Nitrogen Conservation in Reported CMIP6 Earth System Model Data

Gang Tang[1,2], Zebedee Nicholls[1,3,4], Chris Jones[5,6], Thomas Gasser[3], Alexander Norton[7], Tilo Ziehn[8], Alejandro Romero-Prieto[9], Malte Meinshausen[1,4]

[1]School of Geography, Earth and Atmospheric Sciences, The University of Melbourne, Melbourne, Australia
[2]Max Planck Institute for Biogeochemistry, Jena, Germany
[3]International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
[4]Climate Resource, Fitzroy North, Australia
[5]Met Office Hadley Centre, Exeter, United Kingdom
[6]School of Geographical Sciences, University of Bristol, Bristol, United Kingdom
[7]Research School of Biology, Australian National University, Canberra, Australia
[8]CSIRO Environment, Aspendale, Australia
[9]School of Earth and Environment, University of Leeds, Leeds, United Kingdom

Correspondence to: Gang Tang (gang.tang.au@gmail.com; gang.tang@student.unimelb.edu.au; tgang@bgc-jena.mpg.de)

Abstract. Reliable, robust, and consistent data are essential foundations for analysis of carbon cycle feedbacks. Here, we consider the data from multiple Earth System Models (ESMs) participating in the Coupled Model Intercomparison Project Phase 6 (CMIP6). We identify a mass conservation issue in the reported carbon and nitrogen data, with a few exceptions for specific models and reporting levels. The accumulated mass imbalance in the reported data can amount to hundreds of gigatons of carbon or nitrogen by the end of the simulated period, largely exceeding the total carbon/nitrogen pool size changes over the same period. Nitrogen mass imbalance is evident across all reported organic and inorganic pools, with mineral nitrogen exhibiting the most significant cumulative mass imbalance. Due to a lack of details in the reported data, we cannot uniquely identify the cause of this imbalance. However, we postulate that the carbon mass imbalance primarily arises from missing fluxes in the reported data and inconsistencies between these data and the definitions provided by the C4MIP protocol (e.g., land use and fire emissions), rather than from an underlying mass conservation issue in the models themselves. Our findings suggest that future CMIP reporting protocols should consider incorporating mass conservation into their data validation processes so that such issues are caught before users have to deal with them, rather than forcing all users to handle this issue in their own way. In addition, attention from model groups to the detailed diagnostic request and definitions, along with their own quality control will also help to avoid such issues in future. Given that no additional CMIP6 data is currently being published and none is expected in the future, we recommend that data users that rely on a closed carbon/nitrogen cycle address potential flux imbalances by using the workarounds provided in this study.

1 Introduction

The Coupled Model Intercomparison Project Phase 6 (CMIP6, the latest phase at the time of writing) is a collaborative effort among international climate research communities and plays a fundamental role in advancing our understanding of past, present, and future climate (Meehl et al., 2020; Tokarska et al., 2020; Zelinka et al., 2020). CMIP protocols coordinate and standardize climate model experiments, establish consistent requirements for model outputs, and define protocols for model evaluations (Eyring et al., 2016). The Earth System Model (ESM) outputs serve as a primary model data resource for climate scientists worldwide, supporting a wide range of analyses (Tang et al., 2024a; Nicholls et al., 2022; Turnock et al., 2020; Stouffer et al., 2017) and forming the scientific basis for assessments such as those conducted by the Intergovernmental Panel on Climate Change (IPCC) to inform global climate policy and decision-making (Ipcc, 2021b; Arias et al., 2021). Ensuring the reliability, reproducibility, and robustness of the published data is of paramount importance for advancing our understanding of the climate system.

The Coupled Climate-Carbon Cycle Model Intercomparison Project (C4MIP) is one of the CMIP6-Endorsed MIPs focusing on the design, documentation, and analysis of carbon cycle feedbacks and interactions in climate simulations (Friedlingstein et al., 2006; Jones et al., 2016). The interactions between the carbon cycle and atmospheric CO_2 concentration, as well as physical climate, are some of the most important feedbacks in the earth system (Friedlingstein et al., 2006; Arora et al., 2013; Arora et al., 2020). The nitrogen cycle plays a crucial role in carbon cycle feedbacks. For example, the carbon-climate feedback can become negative when the nitrogen cycle is incorporated into models (Zaehle and Dalmonech, 2011; Zaehle et al., 2015). Observational and modeling studies suggest that nitrogen limitation reduces net primary productivity, a key carbon flux that removes CO_2 from the atmosphere (Lebauer and Treseder, 2008; Schulte-Uebbing and De Vries, 2018). Accurate and robust estimates of carbon cycle feedbacks rely on careful analysis of model output. The model output is one line of evidence, but is generally of high importance for the question of carbon cycle feedbacks given that there is, to date, little observational evidence with which to constrain these nonlinear feedback processes. Accurate and robust estimates of carbon cycle feedbacks are also highly desirable for developing climate mitigation and adaptation strategies. The need for robust analysis is why C4MIP made considerable effort to define precise reporting instructions and variable definitions in a way that can be applied to all models (Fig. 1). Tier-1 variables are expected from all models, as is carbon conservation in the reported output (Jones et al., 2016). However, the conservation of carbon and nitrogen in the reported output has not, to date, been assessed.

Figure 1. Land carbon-nitrogen pools and fluxes requested by the C4MIP protocol. The boxes represent pools, while the arrows indicate fluxes and their directions. Details of variable descriptions are provided in Table A1. This figure is reproduced from the C4MIP protocol (Jones et al., 2016) under the Creative Commons Attribution 3.0 License.

The "data request" in CMIP defines specific requirements for model outputs (variables) when submitting experiment results (Juckes et al., 2020). As one of the most complex elements in the infrastructure, the CMIP6 data request encompasses thousands of variables from the various CMIP experiments (Eyring et al., 2016). Despite detailed standardizations and requirements in variable definitions and output formats, the large number of variables present challenges for data quality control for both modeling groups and Earth System Grid Federation (ESGF) (Petrie et al., 2021). The data quality control issues lead to challenges for data processing and analyses for research communities who use the data, such as end-users who require adjustment of spatiotemporal resolution for various analytical purposes (e.g., estimating carbon cycle feedbacks). Where issues are present, users will either use data that has problems they are unaware of or, rather than being able to use the data directly, have to apply a number of fixes before performing their own analysis.

In this study, we focus on the reported land carbon and nitrogen pools and fluxes. Sections 2 and 3 outline our data collection and processing approach, as well as the method for validating mass conservation. Sections 4 and 5 address mass conservation for land carbon and nitrogen, respectively, including their subpools. Each section begins with an introduction to the theoretical mass conservation equations based on the C4MIP protocol (Fig. 1), followed by an analysis of diagnosed mass conservation issues in the reported CMIP6 carbon-nitrogen cycle data. Section 6 discusses additional issues arising from the CMIP6 data request and C4MIP, which contribute to potential confusions. In Section 7, we examine possible causes for mass conservation discrepancies and their implications. Based on the results, we provide suggestions in Section 8 for the CMIP data request team, ESM groups, the C4MIP and related MIPs, and CMIP6 data users. We hope this manuscript will help shed light on these issues, aiding the community in achieving a unified understanding of definitions, reporting protocols, diagnostic recipes and ultimately providing a cleaner, easier to use set of data in CMIP7 to ensure that the

remarkable progress made by modeling teams in recent decades is not lost in data reporting issues (Arora et al., 2020; Jones et al., 2016; Arora et al., 2013; Friedlingstein et al., 2006).

2 Data collection and processing

For the purpose of mass conservation validation, the selection of ESMs and experiments is based on the following criteria: 1) The model must provide a relatively complete set of carbon cycle variables requested by C4MIP. This completeness primarily considers the availability of carbon pool sizes, key fluxes such as 'net primary production' (npp) and 'heterotrophic respiration' (rh), and cross-verification fluxes, for instance, 'net biosphere production' (nbp) and netAtmosToLandCO2Flux. 2) The model must have completed three representative experiments: 1pctCO2, historical, and ssp585. Based on these criteria, this study includes fifteen sets of model runs from twelve ESMs across nine institutes. Details on the models and experiments are provided in Table A1.

The C4MIP variables and related data (Fig. 1 and Table A1) were obtained from ESGF (https://esgf.llnl.gov/, last access 25 June 2023; data availability varies across different ESMs and experiments, Fig. A1). To ensure consistent data post-processing, the most frequently used monthly data with the model's native grid (with "gn" as "grid_label", see CMIP6 global attributes details at https://goo.gl/v1drZl, last access 25 June 2023) was collected.

The data was processed into global-mean, annual-mean values. As a first step, we used the model-specific grid area ('areacella') and land area fraction ('sftlf') to determine global land pool sizes and fluxes. To calculate their annual means, we accounted for differences in model-specific calendars by weighting the monthly global pool sizes/fluxes according to the number of days in each month. While the translation from monthly to annual values is not strictly necessary for the mass conservation test, it largely aids in clarity and visualization.

3 Methods

In this paper, mass conservation is defined as follows: For any single pool - whether the total land carbon/nitrogen or their respective subpools - mass is conserved when the net flux (the difference between fluxes entering and leaving the pool) equals the change in pool size over time (Fig. 1). This definition applies at all time points, meaning conservation must hold at each individual time step as well as over longer periods.

To verify mass conservation, we performed the following calculations: First, we reconstructed the net flux by summing the individual influxes and outfluxes for each pool (Fig. 1). Second, we determined the change in the pool size over time. While the first calculation is straightforward, the second one requires consideration. Both the pool size data and flux data are temporal mean values (a requirement from the CMIP6 data request[2]). However, this reporting convention means that there is only one pool size in each month, rather than information about the size of the pool at the beginning and end of the month

(which is what you would need to unambiguously calculate the change in the pool size over the month of interest). As a result, blindly calculating the difference in pool size between each reported data point will inevitably result in a net flux that differs from the flux reported in the time series. Considering this, we employed a gradient method (Fornberg, 1988), which calculates gradients using second order accurate central differences in the interior points and first order accurate one-sides differences at the boundaries, to derive the change in the pool size from the reported pool size data. It should be noted that this approach is simply a method to obtain a better idea of the actual differential (example provided in Section A1 and Fig. A2). It does not fundamentally resolve the discrepancy between differencing discrete data and differentiating continuous data. However, given the relatively long time series of monthly data (e.g., 165 years for the historical period), the numerical errors introduced by the reporting convention and the method used to calculate changes in pool size are expected to be minimal. In other words, while the discrepancy persists, the magnitudes of dy/dt derived from both methods should closely align.

4 Carbon mass conservation from reported data

To facilitate comparison and formulation, the reconstructed net flux derived from component fluxes and the reconstructed pool size calculated from subpool sizes are denoted with an asterisk ("*") for both carbon and nitrogen. The mass imbalance discussed in the following sections, pertaining to both carbon and nitrogen, refers to the discrepancy between the fluxes and the changes in pool sizes within the reported data. The terms "overestimation" and "underestimation" in this paper specifically denote cases where the net flux exceeds or is smaller than the pool size change in the reported data, rather than implying a comparison with observational data or any external benchmark.

4.1 The top-level closure for land carbon: Equations and results

Based on the C4MIP protocol and CMIP6 data request (Fig. 1), the Tier-1 land carbon pools and directional composite fluxes should have the following relationships.

$$cLand^* = cVeg + cLitter + cSoil + cProduct \tag{1}$$

$$\frac{dcLand^*}{dt} = netAtmosLandCO2Flux - fCLandToOcean \tag{2}$$

Note that: a) we use the sum of the subpools (cLand*) rather than directly using reported cLand in Eq. 2 to ensure consistency in carbon conservation analysis between total land carbon and subpool carbon (cLand and cLand* differ in some ESMs; see details in Section 6.1); b) netAtmosLandCO2Flux and nbp (introduced in CMIP5) are interchangeable based on their definitions (which may also be part of the confusion, see details in Section 6.2); c) fCLeach and fLandToOcean are in C4MIP protocol but not in the CMIP6 data request, thus not included in Eq. 2 (see details in Section 6.3); and d) fCLandToOcean is provided only by CESM2-WACCM (Fig. A1) with negligible values (on the order of 1e-9 GtC/yr, not

shown). Therefore, the following results compare nbp, netAtmosLandCO2Flux, and pool size changes derived from pool size data (dcLand*/dt).

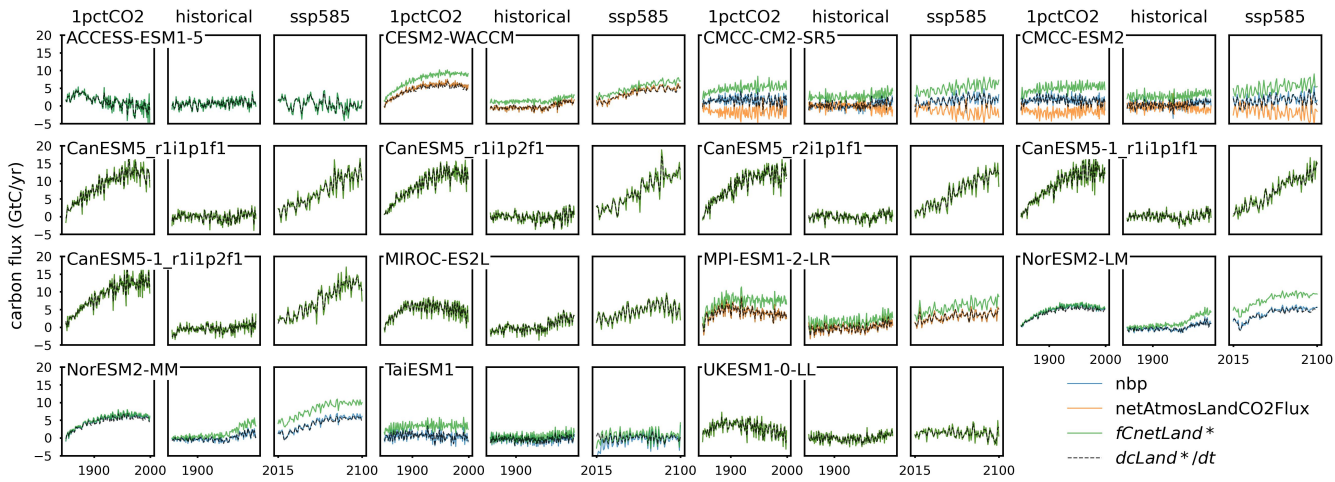



Figure 2. Comparison of 'net biosphere production' (nbp), netAtmosLandCO2Flux, net land carbon flux (fCnetLand*, reconstructed from flux data), and the change of land carbon pool size over time (dcLand*/dt, derived from pool size data) from CMIP6 ESMs across different scenarios

The nbp and dcLand*/dt exhibit nearly identical patterns across all the studied ESMs and experiments (Fig. 2 and Fig. A3), with the small differences likely resulting from the differentiation method applied (see details in Section 3).

The netAtmosLandCO2Flux and dcLand*/dt are either opposite to each other (the two CMCC models seem to report with a sign issue, Fig. 2) or exhibit the same trend (all the other models with netAtmosLandCO2Flux provided, Fig. 2). According to its definition, the netAtmosLandCO2Flux should be identical to nbp, as demonstrated in all ESMs except for the two CMCC models (Fig. A4). The opposite trends in CMCC models are because of the conflicting description in the C4MIP protocol and the CMIP6 data request (see details in Section 6.1). Nonetheless, the consistency among nbp, netAtmosLandCO2Flux (ignoring the sign difference for the CMCC models), and dcLand*/dt indicates three things:

 a) The models themselves should be carbon mass conserved.

 b) The reliability and robustness of the pool size data and aggregated fluxes (nbp and netAtmosLandCO2Flux).

 c) The modeling teams can report data that conserves mass; hence the root of the issue is very likely in confusion about how the reporting protocol is meant to be applied.

4.2 The closure of land carbon from sum of fluxes: Equations and results

According to the C4MIP protocol and the CMIP6 data requirements (Fig. 1), the Tier-1 land carbon pools and component fluxes for the net land flux (fCnetLand*) should have the following relationships.

$$fCnetLand^* = npp - rh - fAnthDisturb - fProductDecomp - fFireNat - fCLandToOcean \qquad (3)$$

$$\frac{dcLand^*}{dt} = fCnetLand^* \qquad (4)$$

Note that many ESMs do not report fAnthDisturb (Fig. A1). In such cases, the Tier-2 fluxes fHarvestToAtmos and fDeforestToAtmos (Fig. 1) are combined to reconstruct it.

The reconstructed fCnetLand* and dcLand*/dt are well-aligned in ACCESS-ESM1-5, CanESM5-1, CanESM5, MIROC-ES2L, and UKESM1-0-LL (Fig. 2), with the difference fluctuating around zero and showing no discernible trends (Fig. A5), suggesting that these differences are due to numerical processing issues alone. In contrast, the reconstructed fCnetLand* is much higher than the change in the pools in all the other studied ESMs (with the green line above the black dashed line, Fig. 2). In these models, the disparity between dcLand*/dt and fCnetLand* ranges from 2 to 6 GtC/yr, with an increasing trend in flux differences, particularly in the 1pctCO2 and ssp585 scenarios (Fig. 2 and Fig. A5). The biased imbalance flux indicates that some component fluxes (outfluxes) in Eq. 3 may be missing from the reported data or incorrectly reported/double counted.

4.3 The significance of cumulative carbon imbalance

Comparing nbp with dcLand*/dt reveals that, although the differences are relatively small and fluctuate around zero (Fig. 2 and Fig. A3), the accumulated mass imbalance is still substantial in certain experiments from some ESMs (Fig. 3). For example, in the 1pctCO2 runs of CESM2-WACCM, NorESM2-LM, and NorESM2-MM, there is an overestimation of approximately 60 GtC. In contrast, during the historical experiment, nbp and dcLand*/dt are well-aligned, resulting in a negligible accumulated imbalance. The mass imbalance in the ssp585 run is also small except for the TaiESM1, where the integration of (nbp - dcLand*/dt) reveals an underestimation of 53.8 GtC of carbon pool size change, comparable in magnitude to the overestimation resulting from the (fCnetLand* - dcLand*/dt) integration (48.3 GtC). This indicates that both nbp and fCnetLand* in TaiESM1's ssp585 run are inconsistent with its reported pool size data.

The imbalance flux from the reconstructed fCnetLand* results in an accumulated mass imbalance ranging from 93.7 to 531.8 GtC, 115.1 to 412.9 GtC, and 48.3 to 351.2 GtC for the 1pctCO2, historical, and ssp585 runs, respectively, in models where carbon conservation is not strictly maintained. These values are substantially higher than the cumulative imbalance from nbp - dcLand*/dt (with a maximum of approximately 60 GtC). By contrast, models that maintain relative carbon conservation (ACCESS-ESM1-5, CanESM5-1, CanESM5, MIROC-ES2L, and UKESM1-0-LL) exhibit much smaller imbalances (<5.5 GtC), likely reflecting unavoidable numerical errors arising from the differentiation of discrete pool size time series.

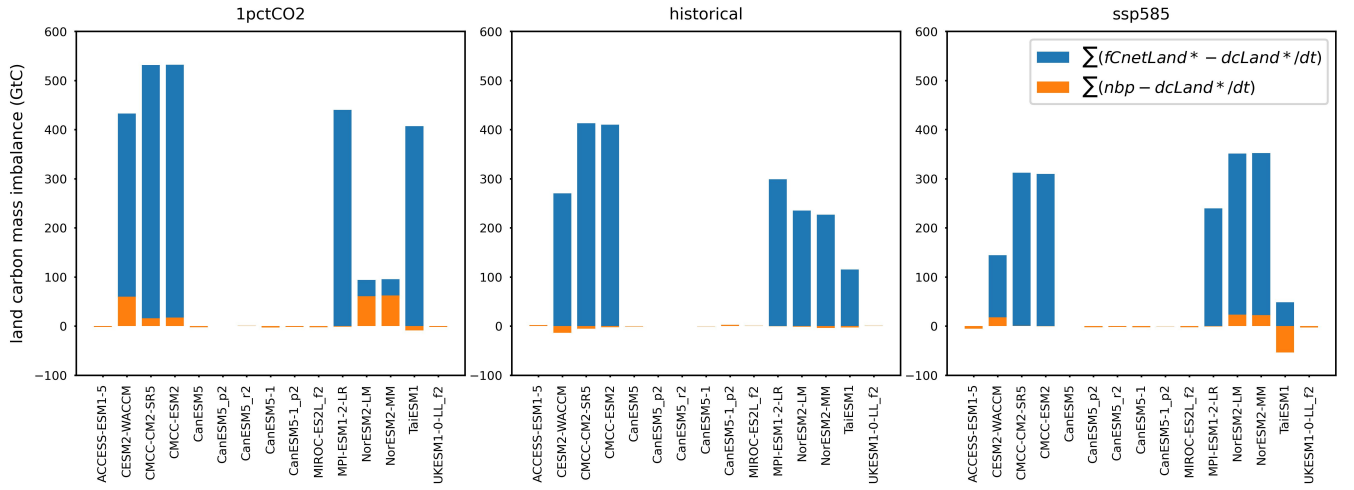



Figure 3. The carbon mass imbalance from mismatch of net land carbon flux (fCnetLand*, reconstructed from flux data) or 'net biosphere production' (nbp) and the change in land carbon pool size over time (dcLand*/dt, derived from pool size data) for CMIP6 ESMs across different scenarios.

From the historical period to ssp585, the land carbon storage trends based on pool size data show a decrease from 1850 to 1975, followed by either continued increase or stabilization until 2100, depending on the ESM (Fig. A6). During the historical period, most ESMs exhibit a decrease in land carbon pool size ranging from 6.8 GtC (CanESM5-1) to 62.6 GtC (TaiESM1). In contrast, ACCESS-ESM1-5, CMCC-CM2-SR5, CMCC-ESM2, and MIROC-ES2L show accumulation of land carbon ranging from 39.0 to 109.1 GtC. Integration of fCnetLand* indicates a maximum carbon accumulation of 465.4 GtC during the historical period, significantly exceeding the changes observed in land carbon pool sizes (Fig. A6). Additionally, for ESMs exhibiting mass imbalance, none capture the decreasing trend in land carbon before 1975 through integration of fCnetLand*. Under the ssp585 scenario, land carbon pool sizes increase from 2.4 GtC (TaiESM1) to 600.4 GtC (CanESM5-1). ESMs with mass imbalance show integration of fCnetLand* at least 1.5 times higher than their corresponding changes in pool sizes. The large discrepancy suggests that ESMs with different fCnetLand* and dcLand*/dt may lead to significant issues in carbon budget (here meaning the internal flows within the carbon cycle, not our remaining carbon budget) assessments. Some of these conservation issues may be solved by considering variables outside the C4MIP protocol (see discussions in Section 7.2).

4.4 The closure of land carbon in subpools: Equations and results

The subpools within the land carbon pool should also maintain mass conservation. Note that the mass conservation of subpools involves some Tier-2 fluxes, such as rhLitter and rhSoil (subfluxes for rh, Fig. 1), which are unreported for many ESMs (Fig. A1). Therefore, we combine the litter and soil pools when writing the mass conservation equations for subpools.

$$fCnetVeg^* = npp - fDeforestToProduct - fHarvestToProduct - fAnthDisturb - fVegLitter - fVegSoil - fVegFire \quad (5)$$

$$\frac{dcVeg^*}{dt} = fCnetVeg^* \tag{6}$$

$$fCnetLitterSoil^* = fVegLitter + fVegSoil - fLitterFire - rh - fCLandToOcean \tag{7}$$

$$\frac{dcLitterSoil^*}{dt} = fCnetLitterSoil^* \tag{8}$$

$$fCnetProduct^* = fDeforestToProduct + fHarvestToProduct - fProductDecomp \tag{9}$$

$$\frac{dcProduct^*}{dt} = fCnetProduct^* \tag{10}$$

Note that the data availability of Tier-2 variables hinders the mass conservation analysis of subpools (Fig. A1). We thus slightly adjust the calculation of the above equations (see details in Section 6.4).

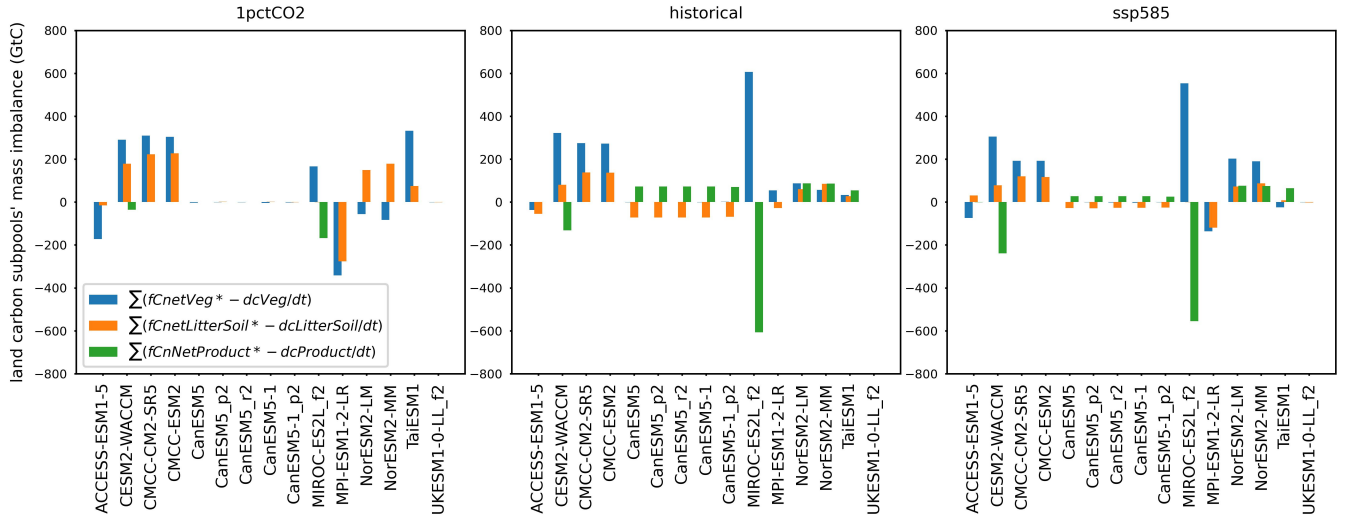



Figure 4. The carbon mass imbalance from mismatch of net carbon fluxes for vegetation, litter+soil, and product pools (fCnetVeg*, fCnetLitterSoil*, and fCnetProduct*, reconstructed from flux data) and the change in carbon pool size over time for vegetation, litter+soil, and product pools (dcVeg/dt, dcLitterSoil/dt, and dcProduct/dt, derived from pool size data) for CMIP6 ESMs across different scenarios.

When comparing the reconstructed net flux with the pool size change, it is clear that net flux overestimation or underestimation occurs across all subpools for nearly all the models (Fig. 4, details in Section A2 and Fig. A7). Notably, all CanESM models and UKESM1-0-LL show conserved carbon in the vegetation pool. The cumulative imbalance, calculated from the integration of imbalance flux, shows that the vegetation carbon pool exhibits the highest imbalance in most cases, with cumulative mass imbalances ranging from ~200 to 600 GtC. Exceptions are found in the 1pctCO2 runs of the two NorESM2 models, where the imbalance from the litter+soil pools surpasses that of the vegetation pools. MIROC-ES2L, despite achieving overall mass conservation in the total land carbon pool (Fig. 3), shows an overestimation of the vegetation carbon pool and an underestimation of the product carbon pool across all experiments (Fig. 4). This discrepancy arises from

the absence of anthropogenic fluxes fDeforestToProduct and fHarvestToProduct from cVeg to cProduct in MIROC (Fig. 1 and Fig. A1). The CanESM models also conserve total land carbon (Fig. 3); however, their net flux integration reveals an underestimation of the litter+soil carbon pool and an overestimation of the product carbon pool in both the historical and ssp585 runs (Fig. 4). The overestimation in the product pool is due to the absence of fProductDecomp and cProduct in these
235 experiments (Fig. A1), with only the influx fDeforestToProduct reported.

5 Nitrogen mass conservation from reported data

There is no reference composite flux like nbp or netAtmosLandCO2Flux for the land nitrogen cycle in the C4MIP protocol or CMIP6 data request (Fig. 1). Thus, we start by directly analysing the land nitrogen mass conservation based on the sum of fluxes. We also analyze the mass conservation from the organic and inorganic nitrogen pools. However, because of a lack of
240 available data (Fig. A1), the further investigation of mass conservation of the organic nitrogen subpools (nVeg, nLitter, nSoil, and nProduct) is not included in this study. Besides, the CanESM series models do not include a nitrogen cycle and are therefore excluded from the following figures.

5.1 The closure of land nitrogen from sum of fluxes: Equations and results

The Tier-1 land nitrogen pools and component fluxes should have the following relationships (based on the C4MIP protocol,
245 Fig. 1).

$$nLand^* = nVeg + nLitter + nSoil + nProduct + nMineral \tag{11}$$

$$fNnetLand^* = fBNF - fNAnthDisturb + fNdep + fNfert - fNloss \tag{12}$$

$$\frac{dnLand^*}{dt} = fNnetLand^* \tag{13}$$

The organic and inorganic nitrogen pools should maintain their respective mass conservation as follows.

250

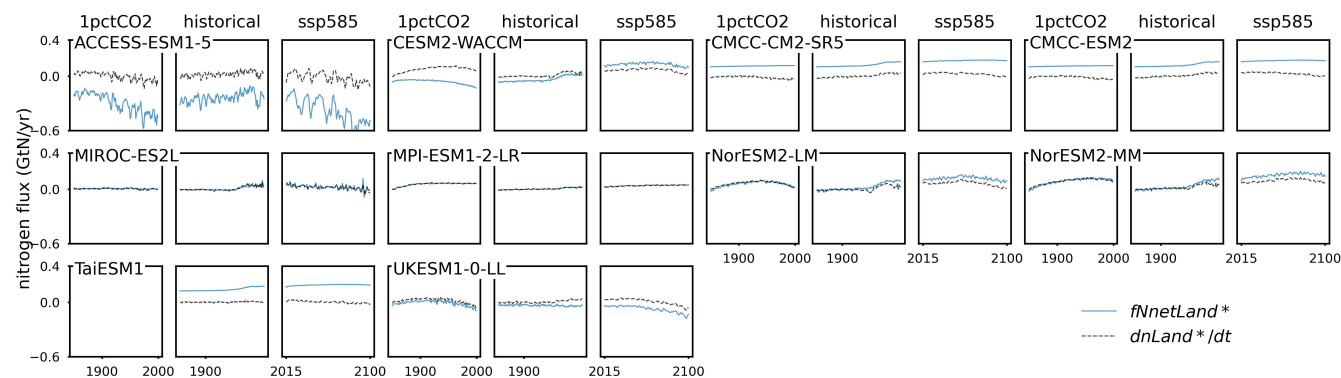
$$nOrganic^* = nVeg + nLitter + nSoil + nProduct \tag{14}$$

$$fNnetOrganic^* = fBNF - fNAnthDisturb + fNup - fNnetmin \tag{15}$$

$$\frac{dnOrganic^*}{dt} = fNnetOrganic^* \tag{16}$$

$$fNnetMineral^* = fNdep + fNfert + fNnetmin - fNup - fNloss \tag{17}$$

$$\frac{dnMineral}{dt} = fNnetMineral^* \tag{18}$$

255 Note that: a) similar to the carbon cycle data, we use the sum of subpools (nLand*) rather than the reported nLand for mass conservation analysis (nLand and nLand* differ in some ESMs; see details in Section 6.1); b) because many ESMs do not report fNloss (Fig. A1), we substitute it with Tier-2 variables, 'gaseous nitrogen loss' (fNgas) and 'nitrogen leaching' (fNleach), where necessary.

260 **Figure 5.** Comparison of net land nitrogen flux (fNnetLand*, reconstructed from flux data) and the change of land nitrogen pool size over time (dnLand*/dt, derived from pool size data) from CMIP6 ESMs across different scenarios.

In principle, nitrogen mass should be conserved in the organic nitrogen pool, the mineral nitrogen pool, and the total land nitrogen pool (Eqs. 11-18). However, there is a significant discrepancy between the reconstructed net flux and pool size change over time across nearly all land nitrogen pools in all ESMs and experiments (Fig. 5 and Fig. A8). Notably, only

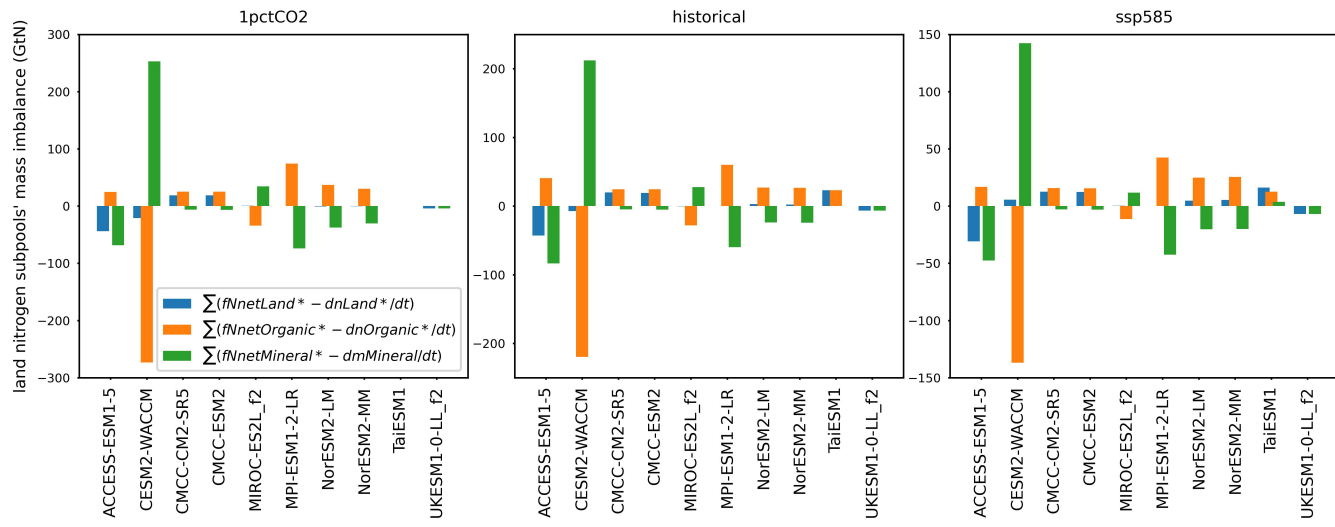
265 MIROC-ES2L and MPI-ESM1-2-LR show mass conservation from the reconstructed net land nitrogen flux (fNnetLand*) in all three experiments (Fig. 5). The two NorESM models show fNnetLand* matching the land nitrogen pool size change (dnLand*/dt) in their 1pctCO2 runs.

The gap between the reconstructed net flux and pool size change for the land organic and inorganic nitrogen pools are significantly larger than that for the total land nitrogen pool (Fig. A8, note the larger y-axis scale). For the organic nitrogen

270 pool, the fNnetOrganic* and dnOrganic*/dt only match well in the three experiments from UKESM1-0-LL. For the inorganic nitrogen pool, while in some models, like the two CMCC models and UKESM1-0-LL, the fNnetMineral* and dnMineral*/dt time series are very close to each other (Fig. A8), the cumulative imbalance is still significant considering the small mineral nitrogen pool size (Fig. A9).

5.2 The significance of cumulative nitrogen imbalance

275 Regarding the cumulative nitrogen imbalance, in ACCESS-ESM1-5, the fluxes underestimate the change in the total land nitrogen and mineral nitrogen pools, while the fluxes overestimate the change in the land organic nitrogen pool, with a mass imbalance ranging from 20 to 80 GtN (Fig. 6). In CESM2-WACCM, the reconstructed fNnetLand* underestimates the change in the land nitrogen pool size (dnLand*/dt) for the 1pctCO2 and historical experiments by 20.9 GtN and 7.3 GtN,

respectively, while overestimating it by 5.5 GtN in the ssp585 run. The total land nitrogen changes in CESM2-WACCM for these experimental periods are +11.7, +1.7, and +5.7 GtN (Fig. A9), respectively. This significant imbalance shows that the change in total land nitrogen varies depending on whether it is assessed using pool size or flux data.

Examining the mass conservation of the organic and mineral land nitrogen subpools, the net flux in CESM2-WACCM significantly underestimates the organic nitrogen pool size change by 136.9 to 273.6 GtN while overestimating the mineral nitrogen pool size change by 142.4 to 252.7 GtN (Fig. 6). CMCC-CM2-SR5 and CMCC-ESM2 both overestimate the total land nitrogen and organic nitrogen while underestimating the mineral nitrogen, with a mass imbalance of less than 25 GtN. For CMCC-CM2-SR5, the fNnetLand* consistently exceeds the land nitrogen pool size change, with overestimation ranging from 12.6 to 19.7 GtN. This imbalance accounts for at least five times the total land nitrogen change for the experimental period in this model (<2.2 GtN). MIROC-ES2L, MPI-ESM1-2-LR, NorESM2-LM, and NorESM2-MM are relatively mass-conserved in their total land nitrogen pool (mass imbalance <5 GtN), particularly in their 1pctCO2 runs (mass imbalance <0.8 GtN). However, this conservation is achieved at the expense of simultaneously underestimating the land organic nitrogen pool size while overestimating the mineral nitrogen pool size (for MIROC-ES2L) or vice versa (for the other three). Both the overestimations and underestimations are substantial, ranging from 25 to 75 GtN. The reconstructed net flux in TaiESM1 shows an overestimation of its land nitrogen pool size in the historical and ssp585 experiments (~25 GtN, with no data provided for its 1pctCO2 run), primarily due to the overestimation of the organic nitrogen pool. In UKESM1-0-LL, the reconstructed net flux overestimates the total land nitrogen by 4.1 GtN in the 1pctCO2 experiment and underestimates it by

6.7 and 7.0 GtN in the historical and ssp585 experiments, respectively. The mass imbalance for UKESM1-0-LL predominantly arises from its organic nitrogen pool.

Based on reported pool size data, the total land nitrogen changes across all studied ESMs during the 1pctCO2, historical, and ssp585 experimental periods range from -2.0 to +13.1 GtN, +0.4 to +3.8 GtN, and -1.7 to +5.5 GtN, respectively (Fig. A9).
305 Changes in the mineral nitrogen pool are substantially smaller, remaining below 0.3 GtN for all experiments and ESMs, with the exception of the 1pctCO2 run in NorESM2-LM (Fig. A9). Comparatively, the cumulative nitrogen imbalance from the reconstructed net flux (Fig. 6) is orders of magnitude larger than the total nitrogen pool size changes over the same periods.

6 Other issues with the reported CMIP6 data

6.1 The reported cLand and nLand

310 In many of the ESMs examined in this study, the reported total land carbon or nitrogen (cLand or nLand) does not match the sum of the respective subpools (i.e., the reconstructed cLand* or nLand*). The discrepancy averages between 50 and 150 GtC or 25 and 55 GtN across the three experimental periods. This inconsistency is unexpected, as pool size, being a state variable, should be simple to calculate and report. This hence indicates that modeling teams may face challenges in fully adhering to C4MIP guidelines, an issue that warrants attention as we approach CMIP7. However, without further evidence
315 (higher resolution or more comprehensive datasets), we are unable to analyze the underlying causes in the CMIP6 data.

6.2 The definition of nbp and netAtmosLandCO2Flux

The netAtmosLandCO2Flux was a new variable introduced in the CMIP6 phase of C4MIP, defined as the "Net Flux of CO_2 Between Atmosphere and Land (Positive into Land) as a Result of All Processes." Many models do not report this flux (Fig. A1). Instead, all models report 'net biosphere production' (nbp), defined as the "Carbon Mass Flux out of Atmosphere Due
320 to Net Biospheric Production on Land," a variable introduced in CMIP5. Based on their definitions, netAtmosLandCO2Flux and nbp appear interchangeable.

However, it should be noted that in the C4MIP paper (Jones et al., 2016), netAtmosLandCO2Flux is illustrated as directed from land to atmosphere (Fig. 1) and described as the "total flux of CO_2 from land to atmosphere," which conflicts with its definition in the CMIP6 data request and may be misleading (e.g., the opposite signs for netAtmosLandCO2Flux and nbp
325 observed in the CMCC models, Fig. 2). Resolving these issues may help reduce confusion and improve the consistency of reported data.

6.3 The changes in fCLeach and fLandToOcean from C4MIP protocol to CMIP6 data request

The original C4MIP design includes a 'leaching carbon flux' (fCLeach) as a Tier-1 variable (Fig. 1). However, this flux is not in the CMIP6 data request (and therefore ESGF) and is thus excluded from Eq. 2. Additionally, fLandToOcean in the

330   C4MIP protocol is renamed to fCLandToOcean in the CMIP6 data request. These changes introduce potential confusion for data users and suggest that a living variable definition and protocol may be essential for providing groups with clarity, given the slow moving nature of paper publication (which makes papers a source that quickly becomes outdated). Living data/protocol papers may provide another solution which could be updated quickly enough to avoid becoming quickly out of date.

6.4 The adjustments for conservation calculations due to limited data availability

For carbon conservation in vegetation and litter-soil pools, consideration of the Tier-2 component fluxes (fVegFire and fLitterFire) for fFireNat is unavoidable (Eq. 5 and Eq. 7). However, fFireNat itself is reported only in TaiESM1 and the two NorESM models (Fig. A1), and component fluxes are even rarer (only CESM2-WACCM reports fVegFire, and none of the studied models report fLitterFire, Fig. A1). Consequently, in calculating fCnetVeg* and fCnetLitterSoil, we replace
340   fVegFire in Eq. 5 with fFireNat and omit fLitterFire from Eq. 7. While this adjustment is not ideal, it is the only feasible approach given the limited data availability.

It is common for ESMs to report Tier-2 variables while omitting Tier-1 variables for certain fluxes (e.g., fHarvestToAtmos and fDeforestToAtmos vs. fAnthDisturb; fNgas and fNleach vs. fNloss, Fig. 1 and Fig. A1). This does not pose a major issue, as we can reconstruct the composite variables by combining the component variables, as done in the conservation
345   calculations. However, this suggests that the C4MIP or CMIP6 data request could consider upgrading the priority of some component variables and potentially removing certain composite variables.

7 Discussions

7. 1 Top level implications

First, it is important to note that we cannot directly determine whether there is a true "mass conservation issue" in the
350   physics of the ESMs based on the published data. From a modeller's perspective, such an issue seems unlikely. Furthermore, we cannot directly verify discrepancies that may arise from the post-processing of the original ESM output to meet the CMIP6 data requirements. It is worth noting that CMIP6 requested variables, the most common of which is monthly gridded data, do not match the original temporal solving frequency in ESMs (Thornton et al., 2007; Sokolov et al., 2008; Wiltshire et al., 2021; Lawrence et al., 2019). This means that post-processing steps like temporal averaging and spatial
355   summing/regridding are inevitable. Numerical errors or other issues may occur during this data processing. Without thoroughly examining the ESM groups' post-processing steps for CMIP6 data submission, it is impossible to pinpoint the source of the mass imbalance. Therefore, the following discussion of the mass conservation issue refers solely to the notable mismatch between the flux data and pool size data submitted to CMIP6/ESGF by ESMs - based on the C4MIP protocol.

The inconsistency between changes in pool size (derived from pool size data) and the reconstructed net flux (derived from
360   flux data, as detailed in Sections 4 and 5) is substantial and is observed across various models and experiments for both
carbon and nitrogen, including their subpools (however limited data availability for subpools precludes further discussion
here). This discrepancy warrants careful attention in future data requests and presents challenges for current data users.

The carbon mass conservation analysis reveals that the land carbon pool size data align well with both the nbp and
netAtmosLandCO2Flux (where available, and ignoring sign issues) in most cases (Fig. 2), which enhances the credibility of
365   the pool size data. The matching of nbp and pool sizes (state variables) also supports the assertion that the ESMs maintain
mass conservation in their raw data.

7.2 The inconsistent land carbon pool size change (dcLand*/dt) and net flux (fCnetLand*)

To understand the potential causes of the land carbon conservation issue, we look into the component fluxes in the net flux
(Eq. 3), which include npp, rh, fAnthDisturb, fFireNat, fProductDecomp, and fLandToOcean. For clarity, Table 1 provides a
370   summary of the land carbon mass conservation status across the models and experiments analyzed.

Table 1. Land carbon mass conservation based on balance between reported fluxes and changes in pool sizes (Equations 3 and 4)

| Models | Land carbon mass conservation status | | | | | | | | |
| | use C4MIP variables only | | | add fLuc | | | add fFire | | |
	1pctCO2	historical	ssp585	1pctCO2	historical	ssp585	1pctCO2	historical	ssp585
ACCESS-ESM1-5	Y	Y	Y	/	/	/	/	/	/
CESM2-WACCM	N	N	N	N (n.d.)	N (-)	N (-)	N (-)	N (+)	N (+)
CMCC-CM2-SR5	N	N	N	N (n.d.)	N (-)	N (-)	N (n.d.)	N (n.d.)	N (n.d.)
CMCC-ESM2	N	N	N	N (n.d.)	N (-)	N (-)	N (-)	N (-)	N (-)
CanESM5	Y	Y	Y	/	/	/	/	/	/
CanESM5_p2	Y	Y	Y	/	/	/	/	/	/
CanESM5_r2	Y	Y	Y	/	/	/	/	/	/
CanESM5-1	Y	Y	Y	/	/	/	/	/	/
CanESM5-1_p2	Y	Y	Y	/	/	/	/	/	/
MIROC-ES2L_f2	Y	Y	Y	/	/	/	/	/	/
MPI-ESM1-2-LR	N	N	N	N (-)	N (-)	N (-)	Y	N (-)	N (+)
NorESM2-LM	N	N	N	N (-)	N	N (-)	N (n.d.)	N (n.d.)	N (n.d.)
NorESM2-MM	N	N	N	N (n.d.)	Y	N (-)	N (n.d.)	/	N (n.d.)
TaiESM1	N	N	N	N (n.d.)	Y	N (+)	Y	/	N (n.d.)
UKESM1-0-LL_f2	Y	Y	Y	/	/	/	/	/	/

Y: mass conserved (cumulative imbalance <= 5GtC)
N: mass not conserved (cumulative imbalance > 5GtC)
N (-): mass not conserved, but including this variable decreased mass imbalance (make things better)

375 N (+): mass not conserved, and including this variable increased mass imbalance (make things worse)

N (n.d.): mass not conserved, no data for this variable

/: mass conserved with C4MIP variables so no further processing

7.2.1 fAnthDisturb vs. fLuc: The land use emissions reported outside C4MIP

The fAnthDisturb variable, comprising fDeforestToAtmos and fHarvestToAtmos (Fig. 1), is related to land use, land-use
380 change, and forestry (LULUCF) emissions (Houghton et al., 2012; Gasser and Ciais, 2013; Stocker and Joos, 2015; Lawrence et al., 2016; Hurtt et al., 2020). Among the studied ESMs, only UKESM1-0-LL - showing negligible mass imbalance (Fig. 2 and Fig. 3) - reports an increasing fHarvestToAtmos from 1 to approximately 5 GtC/yr during the historical to ssp585 periods. The historical fAnthDisturb in UKESM1-0-LL aligns with modelled LULUCF emissions from various dynamic global vegetation models (DGVMs, process-based) and bookkeeping models (e.g., OSCAR, used in the
385 Global Carbon Budget program) (Gasser et al., 2020).

Except for UKESM1-0-LL, CanESM5, CanESM5-1 (both mass-conserved in the land carbon pool), and CESM2-WACCM (not mass-conserved) have reported fDeforestToAtmos (0 - 1.5 GtC/yr). The other ESMs have not provided any data for fAnthDisturb, fDeforestToAtmos, or fHarvestToAtmos (Fig. A1). The definitions of these fluxes are similar to those of gross land-use emissions. Their absence in the historical and SSP585 experiments appears unrealistic and may contribute to the
390 diagnosed carbon mass conservation issue. This is partially corroborated by the mass conservation analysis of the land carbon subpools (Section A2 and Fig. A7), where the vegetation carbon pool - which also has fAnthDisturb (or fDeforestToAtmos, and fHarvestToAtmos) as an outflux (like the total land carbon pool, Fig. 1) - accounts for most of the land carbon mass imbalance in several ESMs (Fig. 4).

To further investigate the land use fluxes, we analyze the fLuc ("Net Carbon Mass Flux into the Atmosphere Due to Land-
395 Use Change") variable from the CMIP5 data request. With the exception of ACCESS-ESM1-5, CanESM5-1, CanESM5, and MIROC-ES2L (which are already land carbon mass-conserved, Fig. 3), all other ESMs report fLuc (Fig. A1), with values ~4 GtC/yr for the two NorESM models and <1.5 GtC/yr for most of the others. Since fLuc is not included in the C4MIP protocol (Fig. 1), it does not fit into the equations used for mass conservation calculations (see Section 3). Given its definition, we assume fLuc as an additional outflux from the land carbon pool and recalculate the imbalance accordingly.

400 NorESM2-MM and TaiESM1 are mass-conserved in their historical runs when fLuc is considered, with carbon mass imbalances dropping from 226.5 and 115.1 GtC to 4.7 and 3.1 GtC, respectively (Table 1 and Fig. A10). The mass imbalance in the NorESM2-LM's ssp585 run also significantly decreases from 351.2 GtC to 22.0 GtC (Fig. A10). Other ESMs with fLuc considered show varying degrees of imbalance reduction (Fig. A10), although the imbalances remain substantial (>55 GtC). The fLuc results further highlight the importance of including land use emissions for mass
405 conservation. However, the discrepancy between the C4MIP protocol and ESM data reporting obscures this issue and may cause confusion.

7.2.2 fFireNat vs. fFire: The fire emissions reported outside the C4MIP

NorESM2-LM, NorESM2-MM, and TaiESM1 are the three models that report fFireNat (Fig. A1). However, the mass imbalance in these four models remains substantial (Fig. 2 and Fig. 3). We notice there is a similar CMIP5 requested variable with fFireNat, fFire ("Carbon Mass Flux into Atmosphere Due to CO_2 Emission from Fire Excluding Land-Use Change"), which is reported by many ESMs (Fig. A1). Thus, we re-check the mass conservation by replacing fFireNat with fFire, where fFire is available.

It is found the fFire is quite high in the 1pctCO2 experiment for the models reporting this flux (e.g., in the range of 1 to 4 GtC/yr with an increasing trend). Considering fFire in the 1pctCO2 run leads to the mass conservation of MPI-ESM1-2-LR and TaiESM1, with the imbalance of 1.6 and 9.1 GtC, respectively (Table 1 and Fig. A11). The imbalance also drops from 432.7 to 59.7 GtC in CESM2-WACCM and from 531.8 to 207.0 GtC in CMCC-ESM2 (Fig. A11). In the historical and ssp585 experiments, though the consideration of fFire largely decreases the imbalance in CMCC-ESM2, it results in some negative imbalance in CESM2-WACCM and MPI-ESM1-2-LR (i.e., fFire is too large). That indicates there might be double counting when simply replacing fFire with fFireNat, which complicates the analysis.

7.2.3 The remaining component fluxes

Besides fAnthDisturb and fFireNat, there are four other fluxes determining the net land carbon fluxes (Eq. 3): npp, rh, fProductDecomp, and fCLandToOcean. Among them, npp and rh stand out as the major carbon fluxes, extensively studied and analyzed in numerous investigations (Jian et al., 2022; Wei et al., 2022; Zhu et al., 2022; Qiu et al., 2023). Observations contribute to constraining the uncertainty of modelled npp and rh (Haverd et al., 2013; Guenet et al., 2024), thereby enhancing their data credibility.

The flux fProductDecomp is reported by ACCESS-ESM1-5, CESM2-WACCM, CMCC-CM2-SR5, CMCC-ESM2, MIROC-ES2L, MPI-ESM1-2-LR, and UKESM1-0-LL (Fig. A1). Notably, four of these seven ESMs demonstrate land carbon mass conservation (Fig. 3). During the historical to SSP585 period, fProductDecomp increases significantly, reaching up to 5 GtC/yr in CESM2-WACCM and MPI-ESM1-2-LR, and as high as 7 GtC/yr in MIROC-ES2L. This substantial flux underscores its importance in closing the carbon cycle. The product pool mass imbalance observed in MIROC-ES2L highlights the importance of carbon fluxes into the product pool (Fig. 4).

Among the studied ESMs, CESM2-WACCM is the only model that provides fCLandToOcean. As most of the CMIP6 ESMs may not simulate the river transport and we have not found other similar CMIP6 variables representing the carbon from land to ocean, we do not discuss this flux here.

7.3 The complexity of carbon mass conservation for subpools

In principle, land carbon subpools should also adhere to mass conservation (Eqs. 5-10). However, the results indicate that most do not, even for those ESMs where total land carbon is mass conserved (Fig. 4 and Fig. A7). Given that fluxes for subpool mass conservation follow specific pathways - such that an outflux for one subpool serves as an influx for the next (e.g., fVegLitter is an outflux for cVeg and an influx for cLitter) - imbalanced subpools present challenges in pinpointing the potentially problematic fluxes. Therefore, it would be unreasonable to attribute the imbalance to any single flux based solely on the reported data. This complexity necessitates a more detailed examination of the underlying processes and how these fluxes are reported to identify sources of mass imbalance in the subpools (see Section 8 below).

7.4 Nitrogen mass conservation: An emerging and challenging issue

The conservation of land nitrogen mass presents a more intricate challenge due to the complexities inherent in the nitrogen cycle (e.g., the need to conserve both organic and mineral nitrogen, Fig. 1). The total land nitrogen, as well as the organic and mineral nitrogen pools, should maintain their respective mass conservation (Eqs. 11-18). However, current CMIP6 data only supports mass conservation in the total land nitrogen pool for MIROC-ES2L, MPI-ESM1-2-LR, NorESM2-LM, and NorESM2-MM, and in the organic nitrogen pool for UKESM1-0-LL (Fig. 6). The imbalance in the reported mineral nitrogen pool is also significant across all the studied CMIP6 ESMs (Fig. 6).

Unlike the carbon cycle, there is no reference nitrogen flux equivalent to nbp or netAtmosLandCO2Flux in the CMIP6 data request and C4MIP design, which complicates the direct validation of both nitrogen pool size data and flux data (Fig. 1). The current fNloss includes fluxes into both the atmosphere (fNgas) and the ocean (fNleach), making it challenging to define a clear reference flux (Fig. 1). Additionally, fNup and fNnetmin link the organic and mineral pools in two directions (Fig. 1), further complicating the attribution of the imbalances.

It is worth noting that only MPI-ESM1-2-LR and UKESM1-0-LL have reported the fNAnthDisturb flux. While MPI-ESM1-2-LR is land nitrogen mass conserved and UKESM1-0-LL is not, this suggests that attributing the total land nitrogen imbalance solely to anthropogenic disturbance may not be appropriate. The lack of a straightforward reference flux and the complex interconnections within the nitrogen cycle highlight the need for more detailed scrutiny to identify and address the sources of nitrogen mass imbalance in these models.

7.5 The quest for reporting balanced carbon and nitrogen cycles

The carbon and nitrogen cycle data, together with many climate variables, are frequently used to support analyses with various purposes (Ipcc, 2021a; Arora et al., 2020; Thornhill et al., 2021; Canadell et al., 2021; Hermans et al., 2021; Fan et al., 2020). These results are particularly crucial for reduced complexity models (emulators), decision-making, and climate policy (Meinshausen et al., 2022; Koven et al., 2022; Kikstra et al., 2022). Considering the challenges in data storage and distribution, it is reasonable that CMIP requests the "most useful" data rather than universally useful output. There is no "one

solution for all" for the CMIP data request, meaning that different research groups require varying levels of post-processing based on their specific purposes. This means that ensuring consistency between the model's original outputs and the processed data is important.

Ensuring mass conservation is crucial for data users, especially those analyzing the comprehensive behaviors of the full carbon-nitrogen cycle, such as reduced complexity modeling groups (Meinshausen et al., 2011a; Meinshausen et al., 2011b; Nicholls et al., 2021; Nicholls et al., 2020; Tang et al., 2024a). Estimating carbon cycle feedback also relies on published ESM data (Arora et al., 2020; Melnikova et al., 2021). Although the nitrogen cycle is relatively new in ESMs (Davies-Barnard et al., 2022), its significance has been widely discussed and demonstrated in both experimental results and models (Zaehle et al., 2010; Zaehle and Dalmonech, 2011; Zaehle et al., 2014; Schulte-Uebbing and De Vries, 2018). Addressing mass conservation of carbon and nitrogen data, a fundamental requirement for accurate analysis, remains a significant challenge. Achieving mass conservation in reported data would significantly simplify the life of users, facilitating more science done faster.

8 Suggestions for the path forward

Based on the preceding discussions, the following practical recommendations are provided for the CMIP data request team, modeling groups, C4MIP and related MIPs (for the future development) and CMIP6 data users, focusing on carbon-nitrogen cycle data.

8.1 For the CMIP data request and ESM groups

a) It is suggested that the next generation of the CMIP data request and the ESM groups incorporate mass conservation of reported data as a data validation routine. ESM groups may consider reporting fluxes that are consistent with the pool size changes, i.e. there may need to be a better way to represent the assumptions required to ensure consistency between the sum of the reported fluxes and the change in pool sizes. Since the process from original model outputs to reported data is not fully transparent to data users, any imbalances in the reported data, traceable to the smallest subpools the model can analyze, must be documented by the ESM groups. This documentation should include the destination of the imbalance flux (e.g., atmosphere or ocean) and the underlying causes (model process or technical issues like temporal averaging). ESGF may need to develop necessary data quality control tools considering mass conservation for data publication. Given the significant effort required to manage and share large data volumes, it is reasonable to prioritize these fundamental checks to guarantee that the data shared is both reliable and valuable to the broader community.

b) Direct communication between the team behind the CMIP data request (based on MIP protocols) and the ESM groups should be strengthened to address data reporting challenges and to inform updates to MIP protocols accordingly. Currently, incomplete data reporting (Fig. A1) or fluxes that do not adhere to the prescribed definitions (as evidenced by fLuc and

 Bridging this gap requires a collaborative effort and should be addressed promptly to improve data quality and make the substantial data sharing effort done by the ESGF worth it.

c) The CMIP data request may consider mandating complete data submission from ESM groups for MIPs (e.g. a core set of required variables like the Tier-1 variables in C4MIP) before publication. In particular, reporting Tier-1 variables should be a prerequisite for reporting Tier-2 variables to prevent confusion among data users. For variables requested by MIPs that are not simulated by ESMs, ESM groups should provide a data availability statement for the variable rather than simply not reporting them. Specifically for C4MIP, the note may indicate how to conserve the carbon and nitrogen despite the absence of this variable. Such practices would enhance mass conservation validation for both data users and ESM groups. Currently, the absence of certain variables (Fig. A1) leads to ambiguity from the perspective of data users: it is unclear whether these variables are not modelled or simply not reported.

8.2 For C4MIP and related MIPs

a) For the carbon cycle, collaborations between C4MIP, the Land Use Model Intercomparison Project (LUMIP) (Lawrence et al., 2016), the Fire Model Intercomparison Project (FireMIP) (Rabin et al., 2017), and possibly others should be considered to harmonize data requests and avoid overlap in variables with similar definitions. Currently, the coexistence of variables such as fLuc and fAnthDisturb, or fFire and fFireNat, not only places an additional burden on ESM teams for data reporting but also creates confusion for data users. From a user perspective, relying on model-specific variables to achieve mass conservation is far from ideal. Addressing this issue should be a priority in future updates to the MIP protocols.

b) It is suggested to promote fDeforestToAtmos and fHarvestToAtmos (Tier-2 fluxes for fAnthDiturb), fVegFire and fLitterFire (Tier-2 fluxes for fFireNat), and rhLitter and rhSoil (Tier-2 fluxes for rh) to Tier-1 variables as they are necessary for analysing mass conservation in key subpools.

c) In the nitrogen cycle, it is suggested to add a new Tier-1 variable to represent the net nitrogen flux between land and atmosphere (e.g., netAtmosLandNgasFlux, similar to nbp and netAtmosLandCO2Flux in the carbon cycle). If ESMs calculate this flux internally (thus, no data processing for CMIP data submission involved), it will aid ESM groups and data users in identifying whether any imbalance originates from fluxes or pool size discrepancies.

d) It is also suggested to promote fNgas and fNleach to Tier-1 variables and remove fNloss, considering their distinct destinations (atmosphere vs. ocean). This might be necessary for the future analyses on the land nitrogen cycle and its interaction with the atmosphere and ocean.

8.3 For users of existing CMIP6 data

525 a) It is recommended to prioritize pool size, 'net primary productivity' (npp), 'heterotrophic respiration' (rh), and 'net biome productivity' (nbp) data. For instance, in carbon cycle feedback calculations, utilizing consistent carbon pool size differencing, as exemplified in CMIP6 carbon cycle feedback analyses (Arora et al., 2020), while avoiding time-integrated fluxes used in CMIP5 and earlier studies (Arora et al., 2013; Friedlingstein et al., 2006), is advisable.

b) When a closed carbon-nitrogen cycle is strictly required, specifying how imbalance fluxes are calculated and managed
530 becomes crucial. Data users may consider two main approaches for carbon mass conservation: (i) Directly using npp and rh flux data and manually calculating the imbalance flux, attributing it to disturbances (land use or fire emissions, which should not be assumed as zero in many models/experiments) or land-to-ocean carbon transport flux. (ii) Using npp, rh, fLuc, and fFire (where available) and manually calculating the imbalance flux, attributing it to other disturbances or land-to-ocean transport.

535 c) For nitrogen mass conservation, prioritizing the use of pool size data, 'biological nitrogen fixation' (fBNF, closely linked to npp), 'nitrogen deposition' (fNdep, verifiable through the input4MIPs dataset), 'nitrogen uptake' (fNup), and 'net mineralization' (fNnetmin, both significant nitrogen fluxes) is recommended. Then, attributing imbalance fluxes in mineral and organic nitrogen pools to mineral nitrogen loss and anthropogenic perturbation, respectively, might be considered with minimal data adjustments.

9 Conclusions

540

The CMIP6 data archive serves as the primary data repository for various communities concerned with climate change. Building and maintaining such a vast data infrastructure is a collaborative effort, understandably fraught with challenges. In this study, we have identified a common issue of mass conservation in reported CMIP6 data from ESMs and experiments following the C4MIP protocol, resulting in accumulated mass imbalances reaching hundreds of gigatons of carbon and
545 nitrogen. The potential reasons include incomplete data submission, inconsistent definitions of fluxes with the C4MIP protocol, different definitions of reporting, particularly concerning land use and fire emissions, and numerical errors from spatio-temporal regridding (minor). The presence of mass imbalance limits the usability of the data and may exacerbate uncertainties in results/conclusions derived from the data. Therefore, we recommend that ESM groups and the next generation of CMIP, C4MIP, and related MIPs prioritize mass conservation as a fundamental data validation step and
550 document any identified mass imbalances. We think this is a reasonable minimum step to justify the effort required to process, share and analyse the data included in CMIP. We also suggest that data users that require closed carbon/nitrogen cycles prioritize carbon/nitrogen pool size data and major carbon/nitrogen flux data (ignoring other data which would conflict with the mass balance inferred from these sources), while also specifying how they handle any remaining imbalance in their analyses (for example, attributing these imbalances to anthropogenic disturbances).

555 Achieving these goals requires collaborative efforts among the CMIP data request team, related MIPs, and modeling groups. Given the substantial efforts already invested in making ESM outputs publicly accessible and analyzable, conducting mass conservation verification is valuable. This step not only facilitates data analysis and reanalysis but also reduces uncertainty in the results and conclusions.

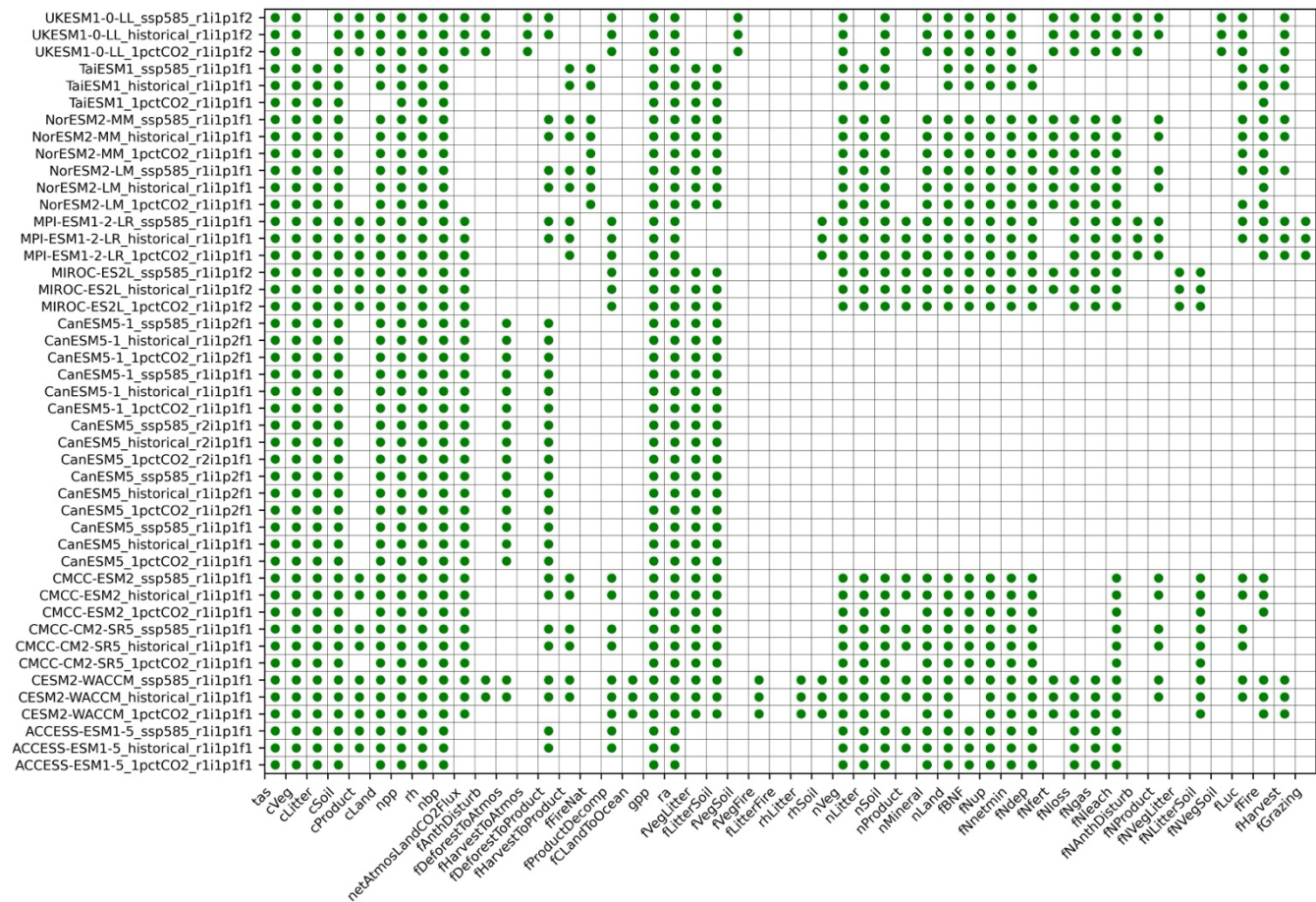



Figure A1. **Overview of data availability for the studied CMIP6 ESMs and experiments. The green dots indicate the data is available from ESGF.**

565

570

Table A1. Details of variables, experiments, and models

Variable	Long name
cVeg	Carbon Mass in Vegetation
cLitter	Carbon Mass in Litter Pool
cSoil	Carbon Mass in Model Soil Pool
cProduct	Carbon Mass in Products of Land-Use Change
cLand	Total Carbon in All Terrestrial Carbon Pools
npp	Net Primary Production on Land as Carbon Mass Flux
rh	Total Heterotrophic Respiration on Land as Carbon Mass Flux
nbp	Carbon Mass Flux out of Atmosphere Due to Net Biospheric Production on Land
netAtmosLandCO2Flux	Net Flux of CO_2 Between Atmosphere and Land (Positive into Land) as a Result of All Processes
fAnthDisturb	Carbon Mass Flux from Vegetation, Litter or Soil Pools into the Atmosphere Due to any Human Activity
fDeforestToAtmos	Deforested Biomass That Goes into Atmosphere as a Result of Anthropogenic Land-Use Change
fHarvestToAtmos	Harvested Biomass That Goes Straight into Atmosphere as Carbon Mass Flux
fDeforestToProduct	Deforested Biomass That Goes into Product Pool as a Result of Anthropogenic Land-Use Change
fHarvestToProduct	Harvested Biomass That Goes into Product Pool
fFireNat	Carbon Mass Flux into Atmosphere Due to CO_2 Emission from Natural Fire
fProductDecomp	Decomposition out of Product Pools to CO_2 in Atmosphere as Carbon Mass Flux
fCLandToOcean	Lateral Transfer of Carbon out of Grid Cell That Eventually Goes into Ocean
gpp	Carbon Mass Flux out of Atmosphere Due to Gross Primary Production on Land
ra	Carbon Mass Flux into Atmosphere Due to Autotrophic (Plant) Respiration on Land
fVegLitter	Total Carbon Mass Flux from Vegetation to Litter
fLitterSoil	Total Carbon Mass Flux from Litter to Soil
fVegSoil	Total Carbon Mass Flux from Vegetation Directly to Soil
fVegFire	Carbon Mass Flux from Vegetation into Atmosphere Due to CO_2 Emission from All Fire
fLitterFire	Carbon Mass Flux from Litter, CWD or any non-Living Pool into Atmosphere Due to CO_2 Emission from All Fire
rhLitter	Carbon Mass Flux into Atmosphere Due to Heterotrophic Respiration from Litter on Land
rhSoil	Carbon Mass Flux into Atmosphere Due to Heterotrophic Respiration from Soil on Land
nVeg	Nitrogen Mass in Vegetation
nLitter	Nitrogen Mass in Litter Pool
nSoil	Nitrogen Mass in Soil Pool
nProduct	Nitrogen Mass in Products of Land-Use Change
nMineral	Mineral Nitrogen in the Soil
nLand	Total Carbon in All Terrestrial Carbon Pools
fBNF	Biological Nitrogen Fixation
fNup	Total Plant Nitrogen Uptake (Sum of Ammonium and Nitrate) Irrespective of the Source of Nitrogen
fNnetmin	Net Nitrogen Release from Soil and Litter as the Outcome of Nitrogen Immobilisation and Gross Mineralisation
fNdep	Dry and Wet Deposition of Reactive Nitrogen onto Land
fNfert	Total Nitrogen Added for Cropland Fertilisation (Artificial and Manure)
fNloss	Total Nitrogen Lost (Including NHx, NOx, N_2O, N_2 and Leaching)
fNgas	Total Nitrogen Lost to the Atmosphere (Sum of NHx, NOx, N_2O, N_2)
fNleach	Total Nitrogen Loss to Leaching or Runoff (Sum of Ammonium, Nitrite and Nitrate)
fNAnthDisturb	Nitrogen Mass Flux out of Land Due to any Human Activity

fNProduct	Deforested or Harvested Biomass as a Result of Anthropogenic Land-Use or Change
fNVegLitter	Total Nitrogen Mass Flux from Vegetation to Litter
fNLitterSoil	Total Nitrogen Mass Flux from Litter to Soil
fNVegSoil	Total Nitrogen Mass Flux from Vegetation Directly to Soil
fLuc	Net Carbon Mass Flux into Atmosphere Due to Land-Use Change
fFire	Carbon Mass Flux into Atmosphere Due to CO_2 Emission from Fire Excluding Land-Use Change
fHarvest	Carbon Mass Flux into Atmosphere Due to Crop Harvesting
fGrazing	Carbon Mass Flux into Atmosphere Due to Grazing on Land
areacella	Grid-Cell Area for Atmospheric Grid Variables
sftlf	Percentage of the grid cell occupied by land (including lakes)

Experiment	Description
1pctCO2	CO2 prescribed to increase at 1% yr^{-1}. Simulation for at least 150 years
historical	Simulation of the recent past. Simulation from 1850-2015, 165 years
ssp585	SSP5 narrative: Fossil-fueled Development. Global markets are increasingly integrated, leading to innovations and technological progress. The social and economic development, however, is based on an intensified exploitation of fossil fuel resources with a high percentage of coal and an energy-intensive lifestyle worldwide. The world economy is growing and local environmental problems such as air pollution are being tackled successfully. SSP585: Based on SSP5 with an additional radiative forcing of 8.5 W/m² by the year 2100. Simulation from 2015-2100, 85 years

Model	Experiment label[a] and model label (if same model with different experiment labels)
ACCESS-ESM1-5	r1i1p1f1
CESM2-WACCM	r1i1p1f1
CMCC-CM2-SR5	r1i1p1f1
CMCC-ESM2	r1i1p1f1
CanESM5-1	r1i1p1f1
	r1i1p2f1, labelled CanESM5-1_p2 in the paper
CanESM5	r1i1p1f1
	r1i1p2f1, labelled CanESM5_p2 in the paper
	r2i1p1f1, labelled CanESM5_r2 in the paper
MIROC-ES2L	r1i1p1f2
MPI-ESM1-2-LR	r1i1p1f1
NorESM2-LM	r1i1p1f1
NorESM2-MM	r1i1p1f1
TaiESM1	r1i1p1f1
UKESM1-0-LL	r1i1p1f2

[a]: The experiment label is called "variant_label" in CMIP6 global attributes (https://goo.gl/v1drZl, last access 25 June 2023). It is a label constructed from four indices stored as global attributes: "r" is followed by the realization index, "i" by the initialization index, "p" by the physics index, and "f" by the forcing index

A1 Methods for calculating the change in the size of the pools

The reported pool size data represents the mean value for each year, which provides no information about the size of the pool at the beginning or the end of the year. Discrete differencing of the pool size data using [pool size(t+1) - pool size(t)] actually calculates the difference between the midpoints of year (t+1) and year (t), rather than the within-year change. In contrast, the reported flux data represents the average flux over the course of year (t). This discrepancy means that the pool size change derived from discrete differencing [pool size(t+1) - pool size(t)] does not correspond to the net flux derived from the reported flux data, potentially leading to a biased estimation of the net flux.

The following provides an example to illustrate the difference between gradient differencing and discrete differencing. Assume a constant flux of 1 GtC/yr over ten years, with the pool size starting at zero. The true pool size time series would be [0, 1, 2, ..., 10] (eleven data points, unit: GtC). However, since the reported pool size is the mean value for each year, the pool size data would be [0.5, 1.5, 2.5, ..., 9.5] (ten data points, unit: GtC, Fig. A2). Discrete differencing of this pool size data yields a flux of 0.5 GtC/yr for the first year (pool size data at t=1 minus zero start), which is lower than the actual flux of 1 GtC/yr. By applying gradient differencing, this issue can be avoided.

The same principle can be applied to more complex examples. The point here is that this is a clear issue but it is also a small one, because we have relatively long time series and relatively well behaved data (so, while we don't have information about the exact values at the start and of the year, the other information provides enough information to constrain their possible values well).

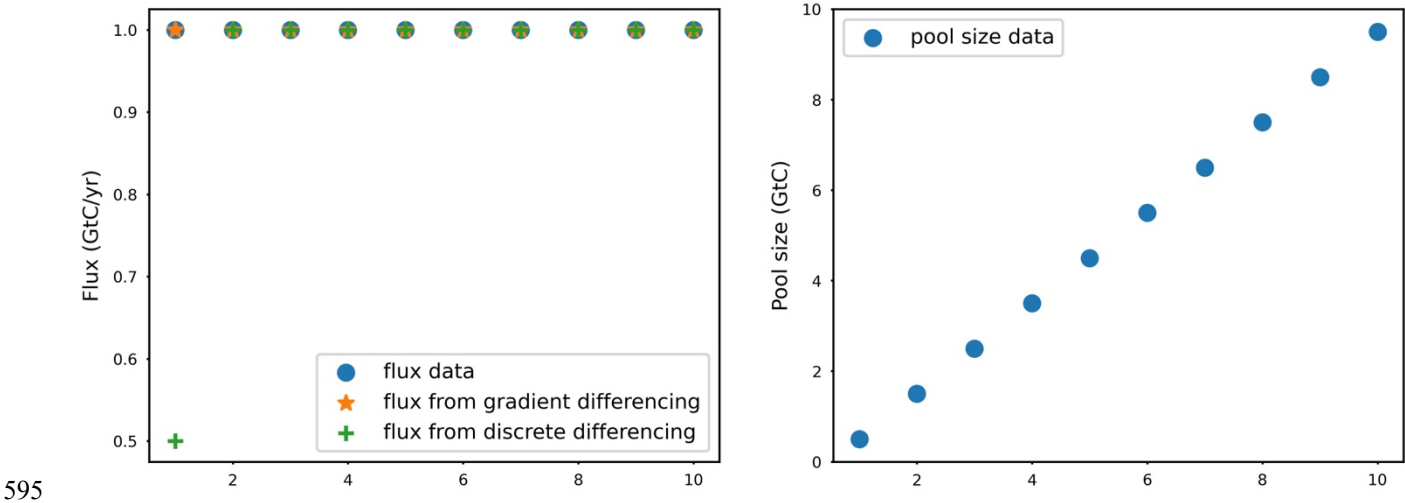



Figure A2. The comparison of differencing methods for pool size data time series

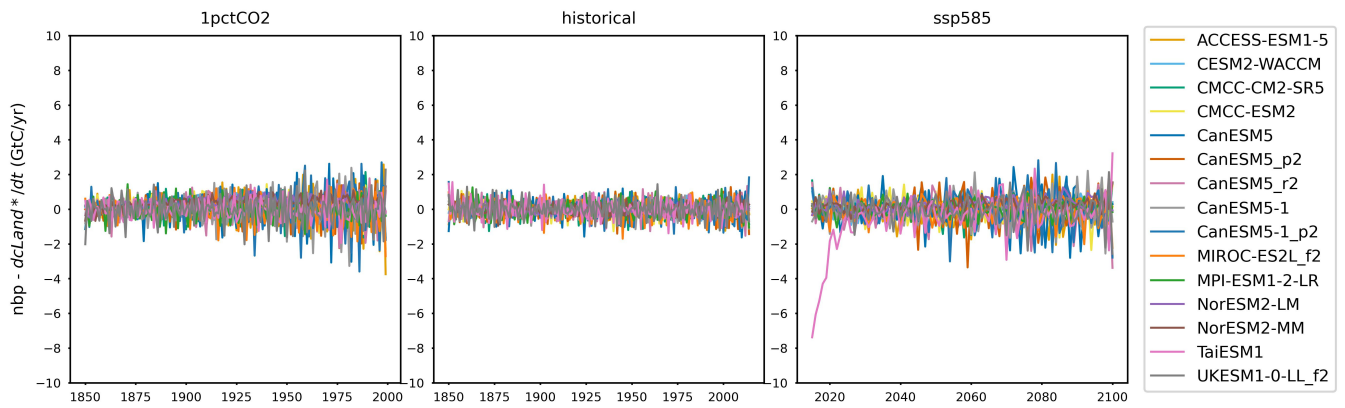
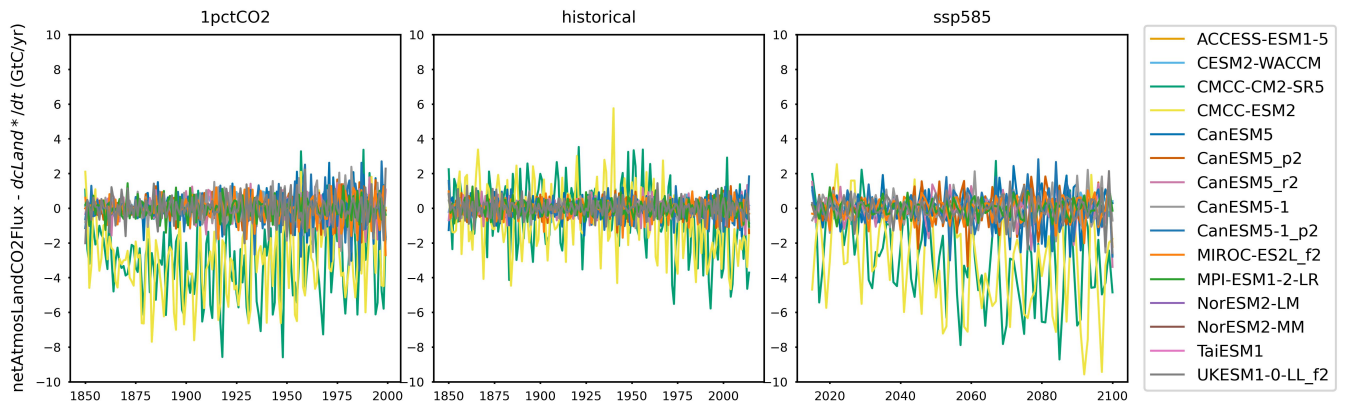



Figure A3. The difference (imbalance flux) between 'net biosphere production' (nbp) and the change of land carbon pool size over time (dcLand*/dt, derived from pool size data) from CMIP6 ESMs across different scenarios

600

 The difference (imbalance flux) between netAtmosLandCO2Flux and the change of land carbon pool size over time (dcLand*/dt, derived from pool size data) from CMIP6 ESMs across different scenarios

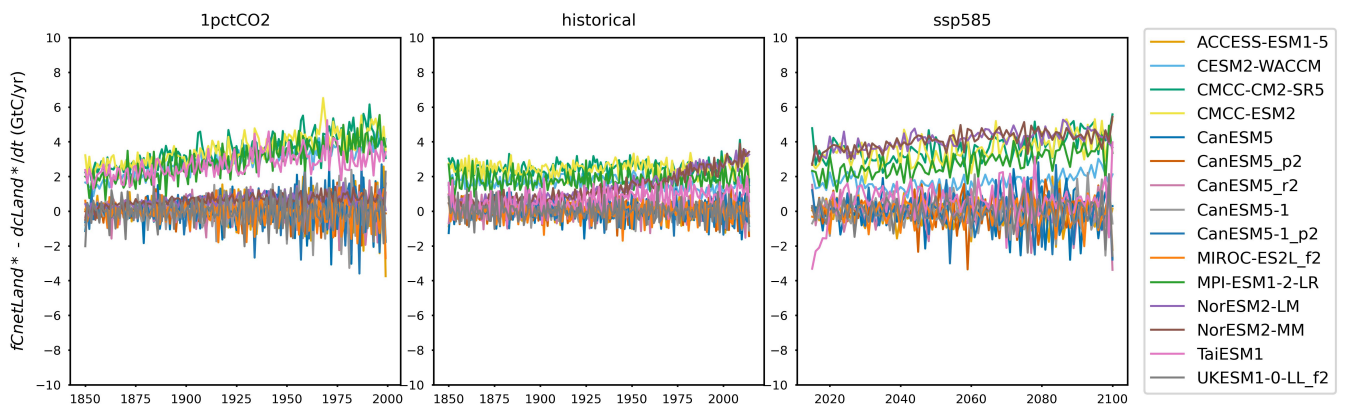



Figure A5. The difference (imbalance flux) between 'net land carbon flux' (fCnetLand*, reconstructed from flux data) and the
605 **change of land carbon pool size over time (dcLand*/dt, derived from pool size data) from CMIP6 ESMs across different scenarios**

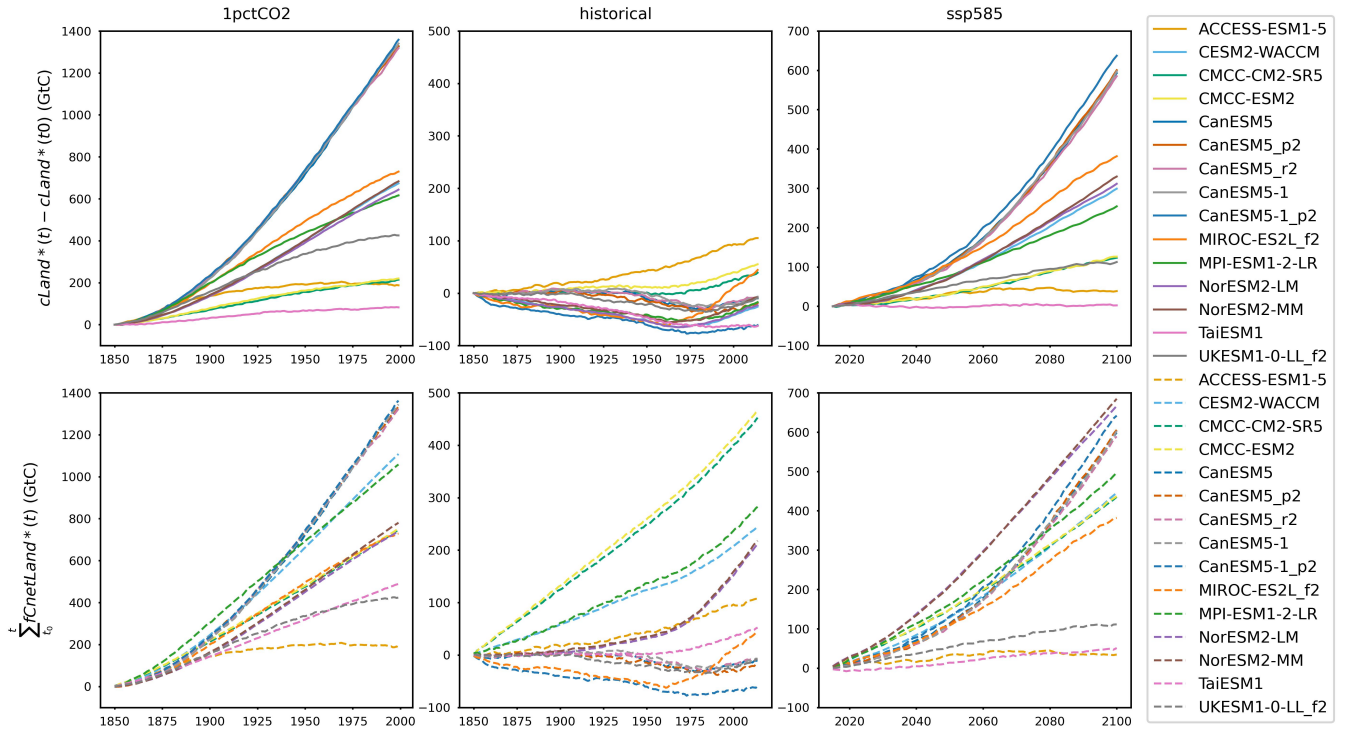



Figure A6. The mismatch of changes in land carbon pool size derived from pool size data [cLand*(t) - cLand*(t0)] (solid lines) and those retrieved from the integration of net land carbon flux reconstructed from flux data (fCnetLand*, dashed lines) in CMIP6 ESMs across different scenarios.

A2 The carbon mass imbalance from the land carbon subpools

For the vegetation carbon pool, the reconstructed net flux (fCnetVeg*) consistently overestimates the vegetation carbon pool size change (dcVeg/dt) in all experiments for CESM2-WACCM, CMCC-CM2-SR5, CMCC-ESM2, MIROC-ES2L, and TaiESM1 (Fig. A7). The fCnetVeg* overestimates the pool size change in the historical and ssp585 runs of both NorESM models, while underestimating the change in the 1pctCO2 run. All CanESM models and UKESM1-0-LL show the matching fCnetVeg* and dcVeg/dt.

The mismatch between net flux and pool size change is less pronounced for the litter+soil carbon pool (fCnetLitterSoil* and dcLitterSoil/dt) compared to the vegetation pool (Fig. A7). Nonetheless, CESM2-WACCM, CMCC-CM2-SR5, and CMCC-ESM2 show a significant overestimation by the reconstructed net flux (fCnetLitterSoil* > dcLitterSoil/dt) in all experiments. The reconstructed fCnetLitterSoil* also overestimates the litter+soil pool size change in all experiments from NorESM2-LM and NorESM2-MM, though the historical run shows a smaller overestimation compared to their 1pctCO2 and ssp585 runs

(Fig. A7). All CanESM models, MIROC-ES2L, TaiESM1, and UKESM1-0-LL exhibit closer alignment between fCnetLitterSoil* and dcLitterSoil/dt, although cumulative imbalances remain substantial in certain cases (Fig. 4).

625 The product pool presents a smaller pool size change compared to the vegetation and litter+soil pools (Fig. A7, around zero where the product pool size is reported). The comparison of the reconstructed net flux (fCnetProduct*) and the derived pool size change over time (dcProduct/dt) shows a significant underestimation (fCnetProduct* < dcProduct/dt) for CESM2-WACCM and MIROC-ES2L. In CESM2-WACCM, this underestimation arises because the influxes (fDeforestToProduct and fHarvestToProduct) are much smaller than the outflux (fProductDecomp) - for example, 0.03 and 0.84 GtC/yr versus 3.6

630 GtC/yr during ssp585, on average. For MIROC-ES2L, the mismatch results from a non-zero fProductDecomp while no fDeforestToProduct or fHarvestToProduct fluxes are reported (Fig. A1). ACCESS-ESM1-5, CMCC-CM2-SR5, CMCC-ESM2, MPI-ESM1-2-LR, and UKESM1-0-LL show nearly identical fCnetProduct* and dcProduct/dt in all experiments. Other ESMs do not report product carbon pool sizes but do report certain fluxes into or out of the product pool (Fig. A1), resulting in a non-zero net flux (Fig. A7).

635

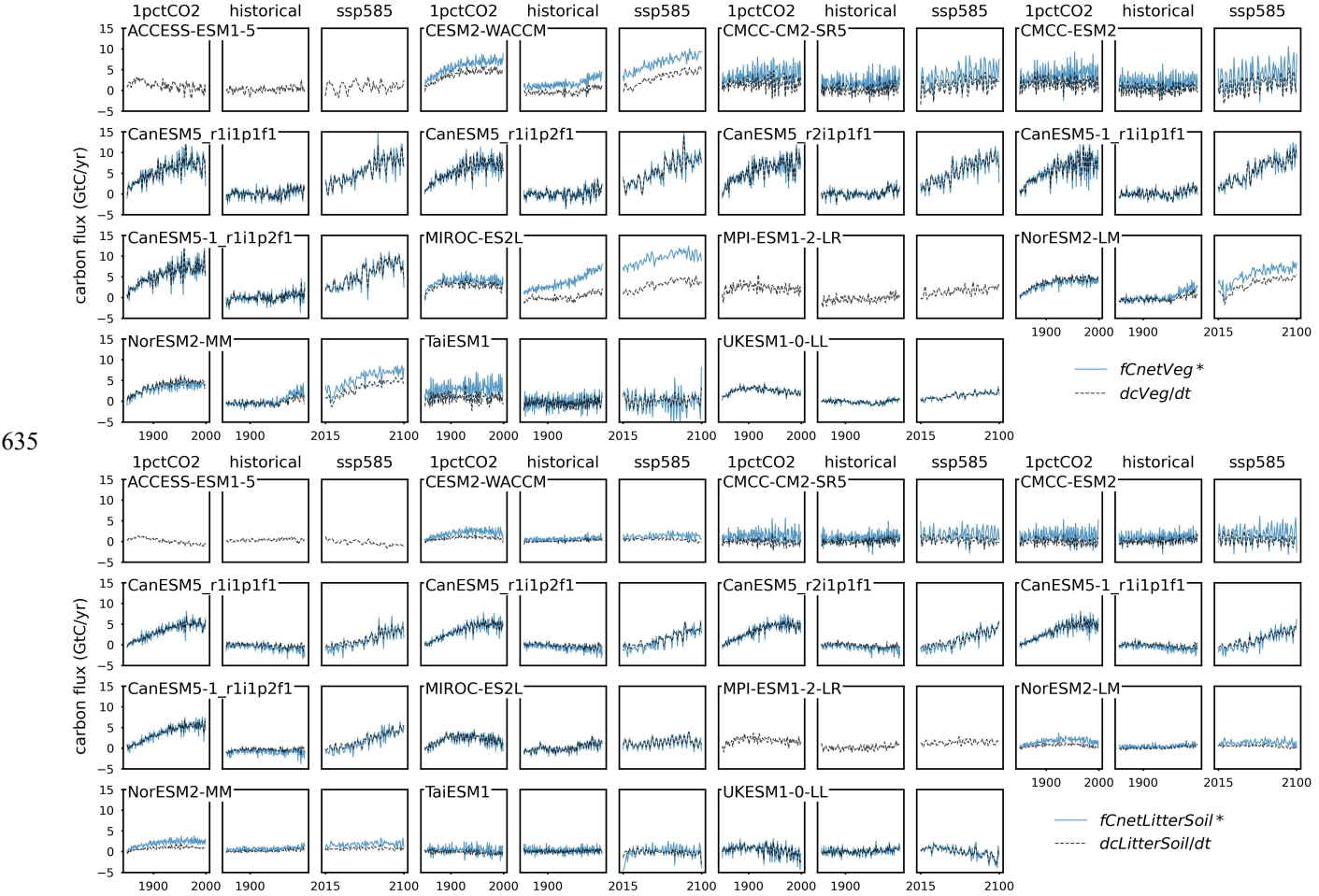
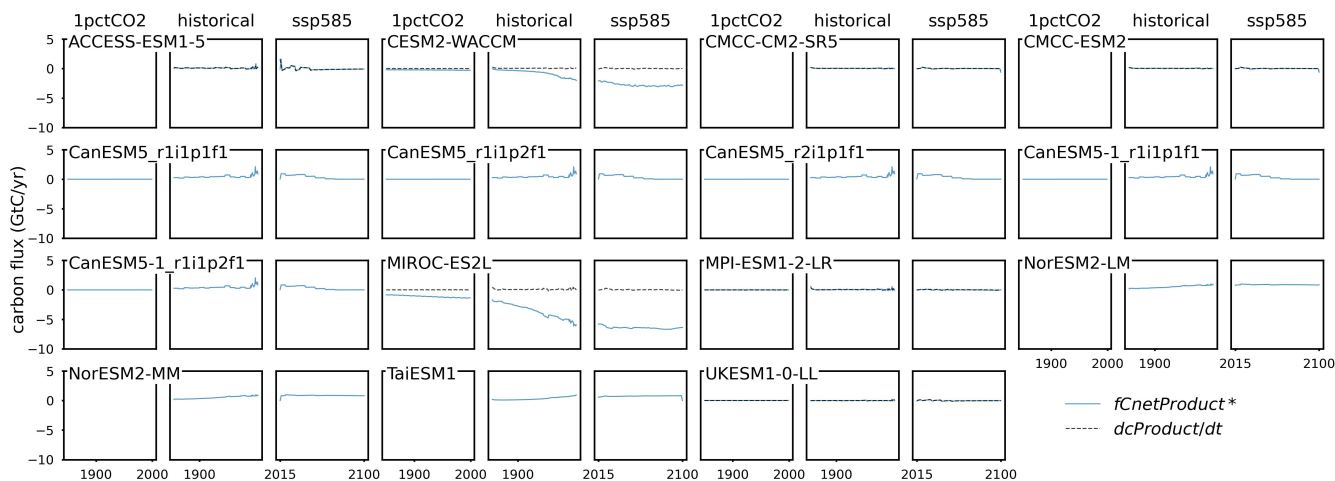



Figure A7. Comparison of net carbon flux for vegetation, litter+soil, and product pools (fCnetVeg*, fCnetLitterSoil*, and fCnetProduct*, reconstructed from flux data) and the change of carbon pool size over time for vegetation, litter+soil, and product pools (dcVeg/dt, dcLitterSoil/dt, and dcProduct/dt, derived from pool size data) from CMIP6 ESMs across different scenarios.

Figure A8. **Comparison of net nitrogen fluxes for organic and mineral pools (fNnetOrganic* and fNnetMineral*, reconstructed from flux data) and the change of nitrogen pool size over time for organic and mineral pools (dnOrganic*/dt and dnMineral/dt, derived from pool size data) from CMIP6 ESMs across different scenarios.**

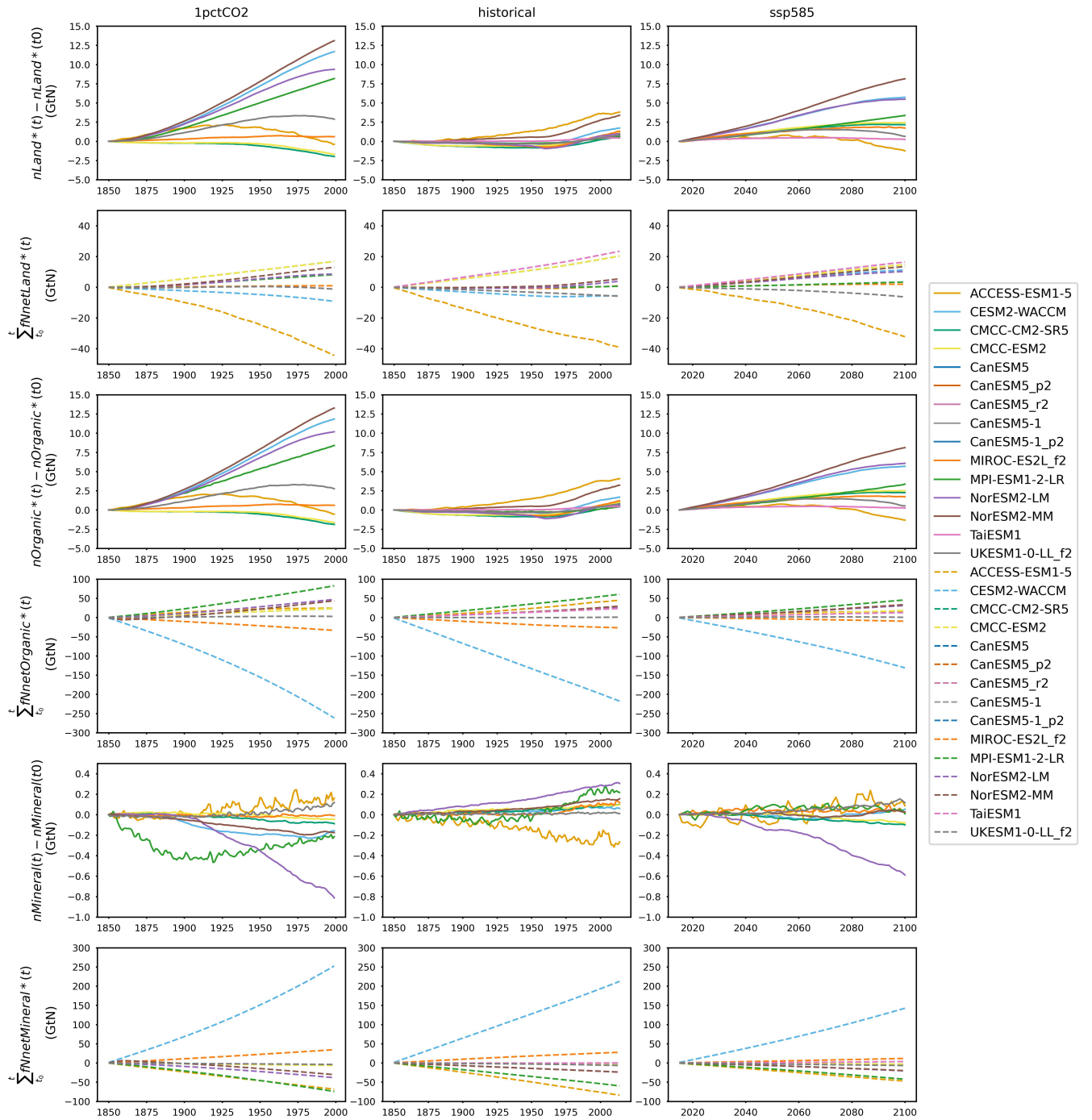
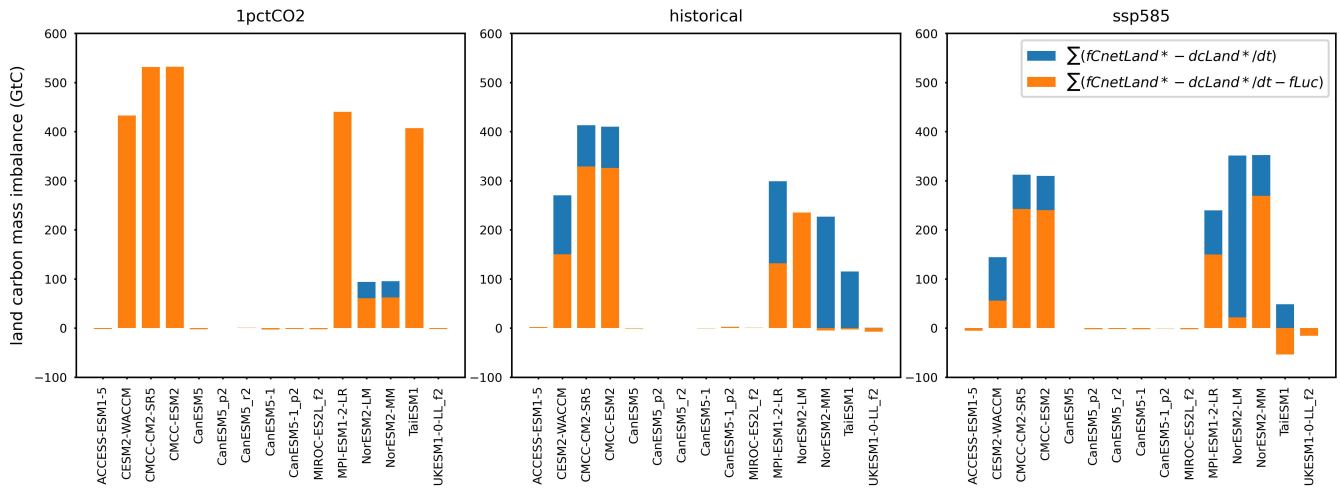
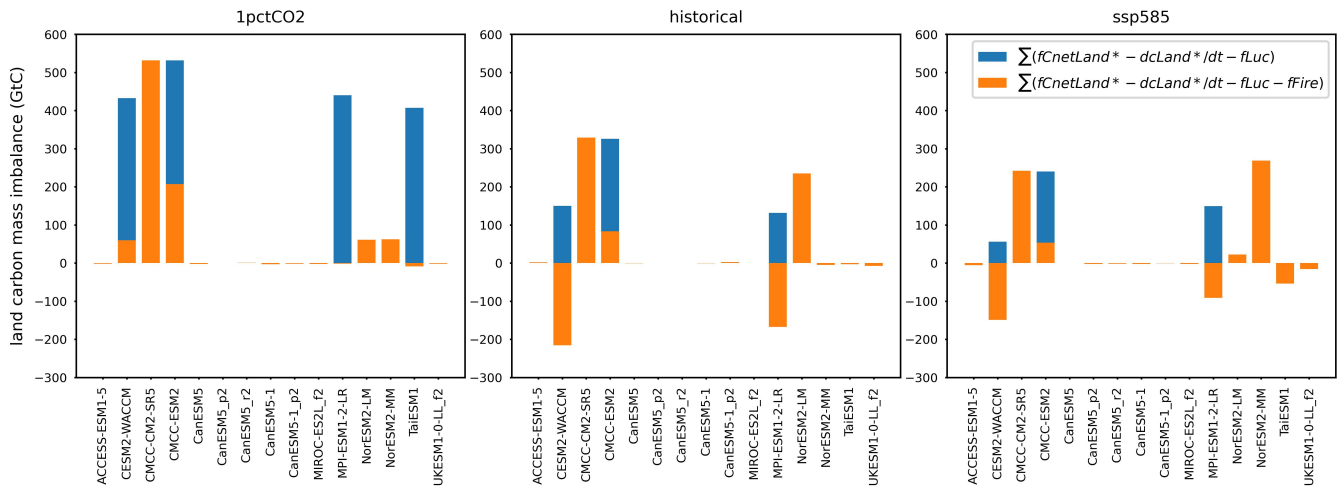



Figure A9. The mismatch of changes in land nitrogen pool size derived from pool size data [land nitrogen: nLand*(t) - nLand*(t0); organic nitrogen: nOrganic*(t) - nOrganic*(t0), and mineral nitrogen: nMineral(t) - nMineral(t0)] (solid lines) and those retrieved from the integration of net nitrogen flux reconstructed from flux data (land nitrogen: fNnetLand*, organic nitrogen: fNnetOrganic*, and mineral nitrogen: fNnetMineral*, dashed lines) in CMIP6 ESMs across different scenarios. Note there is no nitrogen cycle in CanESM.

650

655 **Figure A10. The carbon mass imbalance from mismatch of net land carbon flux (fCnetLand*, with or without the consideration of fLuc, reconstructed from flux data) and the change in land carbon pool size over time (dcLand*/dt, derived from pool size data) for CMIP6 ESMs across different scenarios.**

660 **Figure A11. The carbon mass imbalance from mismatch of net land carbon flux (fCnetLand*, with or without the consideration of fFire, reconstructed from flux data) and the change in land carbon pool size over time (dcLand*/dt, derived from pool size data) for CMIP6 ESMs across different scenarios.**

Code and data availability

The CMIP6 Earth System Model (ESM) output data were initially collected from the Earth System Grid Federation (ESGF, https://esgf-node.llnl.gov/projects/cmip6/, last accessed on 25 June 2023). The processed global annual mean data, along with the associated calculation and visualization code, are available at https://doi.org/10.5281/zenodo.14060168 (Tang et al., 2024b).

Author contribution

GT conceptualized the approach for analyzing carbon-nitrogen mass conservation in the published CMIP6 Earth System Model data. GT collected, processed, and analyzed the CMIP6 data. CJ suggested the analysis of mass conservation for carbon and nitrogen subpools, which was subsequently conducted by GT. ZN, CJ, TG, TZ, AN, MM, and ARP contributed to the methodology and analysis, while MM, CJ, ZN, and AN supervised the study. GT analyzed and interpreted the results and prepared the first manuscript draft, with CJ, MM, ZN, and TG providing suggestions on structuring the narrative. All authors contributed to the writing and revision of the manuscript.

Competing interests

One author is a member of the editorial board of Geoscientific Model Development (GMD).

Acknowledgements

This work is supported by the National Environmental Science Program (NESP, Climate Systems Hub, Malte Meinshausen and Tilo Ziehn), funded by the Australian Government Department of Climate Change, Energy, the Environment, and Water; the Horizon 2020 Research and Innovation Funding Programme (No. 101003536, Earth System Models for the Future, ESM2025, Zebedee Nicholls, Thomas Gasser, and Chris Jones), funded by the European Union; the Met Office Hadley Centre Climate Programme (Chris Jones), funded by the Government of the United Kingdom Department for Science, Innovation and Technology (DIST); and the Leeds-York-Hull Natural Environment Research Council (NERC) Doctoral Training Partnership (DTP) Panorama (No. NE/S007458/1, Alejandro Romero-Prieto).

We express our sincere gratitude to the World Climate Research Programme (WCRP), the Coupled Model Intercomparison Project Phase 6 (CMIP6), the Earth System Model (ESM) groups, the Coupled Climate-Carbon Cycle Model Intercomparison Project (C4MIP) and related MIPs, and the Earth System Grid Federation (ESGF) for their invaluable collaborative efforts in making ESM outputs available and accessible.

References

Arias, P. A., Bellouin, N., Coppola, E., Jones, R. G., Krinner, G., Marotzke, J., Naik, V., Palmer, M. D., Plattner, G. K., Rogelj, J., Rojas, M., Sillmann, J., Storelvmo, T., Thorne, P. W., Trewin, B., Achuta Rao, K., Adhikary, B., Allan, R. P., Armour, K., Bala, G., Barimalala, R., Berger, S., Canadell, J. G., Cassou, C., Cherchi, A., Collins, W., Collins, W. D., Connors, S. L., Corti, S., Cruz, F., Dentener, F. J., Dereczynski, C., Di Luca, A., Diongue Niang, A., Doblas-Reyes, F. J., Dosio, A., Douville, H., Engelbrecht, F., Eyring, V., Fischer, E., Forster, P., Fox-Kemper, B., Fuglestvedt, J. S., Fyfe, J. C., Gillett, N. P., Goldfarb, L., Gorodetskaya, I., Gutierrez, J. M., Hamdi, R., Hawkins, E., Hewitt, H. T., Hope, P., Islam, A. S., Jones, C., Kaufman, D. S., Kopp, R. E., Kosaka, Y., Kossin, J., Krakovska, S., Lee, J. Y., Li, J., Mauritsen, T., Maycock, T. K., Meinshausen, M., Min, S. K., Monteiro, P. M. S., Ngo-Duc, T., Otto, F., Pinto, I., Pirani, A., Raghavan, K., Ranasinghe, R., Ruane, A. C., Ruiz, L., Sallée, J. B., Samset, B. H., Sathyendranath, S., Seneviratne, S. I., Sörensson, A. A., Szopa, S., Takayabu, I., Tréguier, A. M., van den Hurk, B., Vautard, R., von Schuckmann, K., Zaehle, S., Zhang, X., and Zickfeld, K.: Technical Summary, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 33−144, https://doi.org/10.1017/9781009157896.002, 2021.

Arora, V. K., Boer, G. J., Friedlingstein, P., Eby, M., Jones, C. D., Christian, J. R., Bonan, G., Bopp, L., Brovkin, V., Cadule, P., Hajima, T., Ilyina, T., Lindsay, K., Tjiputra, J. F., and Wu, T.: Carbon–Concentration and Carbon–Climate Feedbacks in CMIP5 Earth System Models, Journal of Climate, 26, 5289-5314, https://doi.org/10.1175/JCLI-D-12-00494.1, 2013.

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., Schwinger, J., Bopp, L., Boucher, O., Cadule, P., Chamberlain, M. A., Christian, J. R., Delire, C., Fisher, R. A., Hajima, T., Ilyina, T., Joetzjer, E., Kawamiya, M., Koven, C. D., Krasting, J. P., Law, R. M., Lawrence, D. M., Lenton, A., Lindsay, K., Pongratz, J., Raddatz, T., Séférian, R., Tachiiri, K., Tjiputra, J. F., Wiltshire, A., Wu, T., and Ziehn, T.: Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models, Biogeosciences, 17, 4173-4222, https://doi.org/10.5194/bg-17-4173-2020, 2020.

Canadell, J. G., Monteiro, P. M. S., Costa, M. H., Cotrim da Cunha, L., Cox, P. M., Eliseev, A. V., Henson, S., Ishii, M., Jaccard, S., Koven, C., Lohila, A., Patra, P. K., Piao, S., Rogelj, J., Syampungani, S., Zaehle, S., and Zickfeld, K.: Global Carbon and other Biogeochemical Cycles and Feedbacks, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 673–816, https://doi.org/10.1017/9781009157896.007, 2021.

Davies-Barnard, T., Zaehle, S., and Friedlingstein, P.: Assessment of the impacts of biological nitrogen fixation structural uncertainty in CMIP6 earth system models, Biogeosciences, 19, 3491-3503, https://doi.org/10.5194/bg-19-3491-2022, 2022.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937-1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Fan, X., Duan, Q., Shen, C., Wu, Y., and Xing, C.: Global surface air temperatures in CMIP6: historical performance and future changes, Environmental Research Letters, 15, 104056, https://doi.org/10.1088/1748-9326/abb051, 2020.

Fornberg, B.: Generation of finite difference formulas on arbitrarily spaced grids, Mathematics of computation, 51, 699-706, https://doi.org/10.1090/S0025-5718-1988-0935077-0, 1988.

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K.-G., Schnur, R., Strassmann, K., Weaver, A. J., Yoshikawa, C., and Zeng, N.: Climate–Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison, Journal of Climate, 19, 3337-3353, https://doi.org/10.1175/jcli3800.1, 2006.

Gasser, T. and Ciais, P.: A theoretical framework for the net land-to-atmosphere CO_2 flux and its implications in the definition of "emissions from land-use change", Earth Syst. Dynam., 4, 171-186, https://doi.org/10.5194/esd-4-171-2013, 2013.

Gasser, T., Crepin, L., Quilcaille, Y., Houghton, R. A., Ciais, P., and Obersteiner, M.: Historical CO2 emissions from land use and land cover change and their uncertainty, Biogeosciences, 17, 4075-4101, https://doi.org/10.5194/bg-17-4075-2020, 2020.

Guenet, B., Orliac, J., Cécillon, L., Torres, O., Sereni, L., Martin, P. A., Barré, P., and Bopp, L.: Spatial biases reduce the ability of Earth system models to simulate soil heterotrophic respiration fluxes, Biogeosciences, 21, 657-669, https://doi.org/10.5194/bg-21-657-2024, 2024.

Haverd, V., Raupach, M. R., Briggs, P. R., Canadell, J. G., Isaac, P., Pickett-Heaps, C., Roxburgh, S. H., van Gorsel, E., Viscarra Rossel, R. A., and Wang, Z.: Multiple observation types reduce uncertainty in Australia's terrestrial carbon and water cycles, Biogeosciences, 10, 2011-2040, https://doi.org/10.5194/bg-10-2011-2013, 2013.

Hermans, T. H. J., Gregory, J. M., Palmer, M. D., Ringer, M. A., Katsman, C. A., and Slangen, A. B. A.: Projecting Global Mean Sea-Level Change Using CMIP6 Models, Geophysical Research Letters, 48, e2020GL092064, https://doi.org/10.1029/2020GL092064, 2021.

Houghton, R. A., House, J. I., Pongratz, J., van der Werf, G. R., DeFries, R. S., Hansen, M. C., Le Quéré, C., and Ramankutty, N.: Carbon emissions from land use and land-cover change, Biogeosciences, 9, 5125-5142, https://doi.org/10.5194/bg-9-5125-2012, 2012.

Hurtt, G. C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J. C., Fisk, J., Fujimori, S., Klein Goldewijk, K., Hasegawa, T., Havlik, P., Heinimann, A., Humpenöder, F., Jungclaus, J., Kaplan, J. O., Kennedy, J., Krisztin, T., Lawrence, D., Lawrence, P., Ma, L., Mertz, O., Pongratz, J., Popp, A., Poulter, B., Riahi, K., Shevliakova, E., Stehfest, E., Thornton, P., Tubiello, F. N., van Vuuren, D. P., and Zhang, X.: Harmonization of global land use change and management for the period 850–2100 (LUH2) for CMIP6, Geosci. Model Dev., 13, 5425-5464, https://doi.org/10.5194/gmd-13-5425-2020, 2020.

Ipcc: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896, 2021a.

Ipcc: Summary for Policymakers, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 3−32, https://doi.org/10.1017/9781009157896.001, 2021b.

Jian, J., Bailey, V., Dorheim, K., Konings, A. G., Hao, D., Shiklomanov, A. N., Snyder, A., Steele, M., Teramoto, M., Vargas, R., and Bond-Lamberty, B.: Historically inconsistent productivity and respiration fluxes in the global terrestrial carbon cycle, Nature Communications, 13, 1733, https://doi.org/10.1038/s41467-022-29391-5, 2022.

Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., Graven, H., Hoffman, F., Ilyina, T., John, J. G., Jung, M., Kawamiya, M., Koven, C., Pongratz, J., Raddatz, T., Randerson, J. T., and Zaehle, S.: C4MIP – The Coupled Climate–Carbon Cycle Model Intercomparison Project: experimental protocol for CMIP6, Geosci. Model Dev., 9, 2853-2880, https://doi.org/10.5194/gmd-9-2853-2016, 2016.

Juckes, M., Taylor, K. E., Durack, P. J., Lawrence, B., Mizielinski, M. S., Pamment, A., Peterschmitt, J. Y., Rixen, M., and Sénési, S.: The CMIP6 Data Request (DREQ, version 01.00.31), Geosci. Model Dev., 13, 201-224, https://doi.org/10.5194/gmd-13-201-2020, 2020.

Kikstra, J. S., Nicholls, Z. R. J., Smith, C. J., Lewis, J., Lamboll, R. D., Byers, E., Sandstad, M., Meinshausen, M., Gidden, M. J., Rogelj, J., Kriegler, E., Peters, G. P., Fuglestvedt, J. S., Skeie, R. B., Samset, B. H., Wienpahl, L., van Vuuren, D. P., van der Wijst, K. I., Al Khourdajie, A., Forster, P. M., Reisinger, A., Schaeffer, R., and Riahi, K.: The IPCC Sixth Assessment Report WGIII climate assessment of mitigation pathways: from emissions to global temperatures, Geosci. Model Dev., 15, 9075-9109, https://doi.org/10.5194/gmd-15-9075-2022, 2022.

Koven, C. D., Arora, V. K., Cadule, P., Fisher, R. A., Jones, C. D., Lawrence, D. M., Lewis, J., Lindsay, K., Mathesius, S., Meinshausen, M., Mills, M., Nicholls, Z., Sanderson, B. M., Séférian, R., Swart, N. C., Wieder, W. R., and Zickfeld, K.:

Multi-century dynamics of the climate and carbon cycle under both high and net negative emissions scenarios, Earth Syst. Dynam., 13, 885-909, https://doi.org/10.5194/esd-13-885-2022, 2022.

790 Lawrence, D. M., Hurtt, G. C., Arneth, A., Brovkin, V., Calvin, K. V., Jones, A. D., Jones, C. D., Lawrence, P. J., de Noblet-Ducoudré, N., Pongratz, J., Seneviratne, S. I., and Shevliakova, E.: The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: rationale and experimental design, Geosci. Model Dev., 9, 2973-2998, https://doi.org/10.5194/gmd-9-2973-2016, 2016.

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., Kluzek, E., Lawrence, P. J., Li, F., Li, H., Lombardozzi, D., Riley, W. J., Sacks, W. J., Shi,
795 M., Vertenstein, M., Wieder, W. R., Xu, C., Ali, A. A., Badger, A. M., Bisht, G., van den Broeke, M., Brunke, M. A., Burns, S. P., Buzan, J., Clark, M., Craig, A., Dahlin, K., Drewniak, B., Fisher, J. B., Flanner, M., Fox, A. M., Gentine, P., Hoffman, F., Keppel-Aleks, G., Knox, R., Kumar, S., Lenaerts, J., Leung, L. R., Lipscomb, W. H., Lu, Y., Pandey, A., Pelletier, J. D., Perket, J., Randerson, J. T., Ricciuto, D. M., Sanderson, B. M., Slater, A., Subin, Z. M., Tang, J., Thomas, R. Q., Val Martin, M., and Zeng, X.: The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of
800 Forcing Uncertainty, Journal of Advances in Modeling Earth Systems, 11, 4245-4287, https://doi.org/10.1029/2018MS001583, 2019.

LeBauer, D. S. and Treseder, K. K.: Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed, Ecology, 89, 371-379, https://doi.org/10.1890/06-2057.1, 2008.

Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J.-F., Stouffer, R. J., Taylor, K. E., and Schlund, M.: Context
805 for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models, Science Advances, 6, eaba1981, https://doi.org/10.1126/sciadv.aba1981, 2020.

Meinshausen, M., Raper, S. C. B., and Wigley, T. M. L.: Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – Part 1: Model description and calibration, Atmos. Chem. Phys., 11, 1417-1456, https://doi.org/10.5194/acp-11-1417-2011, 2011a.

810 Meinshausen, M., Wigley, T. M. L., and Raper, S. C. B.: Emulating atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – Part 2: Applications, Atmos. Chem. Phys., 11, 1457-1471, https://doi.org/10.5194/acp-11-1457-2011, 2011b.

Meinshausen, M., Lewis, J., McGlade, C., Gütschow, J., Nicholls, Z., Burdon, R., Cozzi, L., and Hackmann, B.: Realization of Paris Agreement pledges may limit warming just below 2 °C, Nature, 604, 304-309, https://doi.org/10.1038/s41586-022-04553-z, 2022.

815 Melnikova, I., Boucher, O., Cadule, P., Ciais, P., Gasser, T., Quilcaille, Y., Shiogama, H., Tachiiri, K., Yokohata, T., and Tanaka, K.: Carbon Cycle Response to Temperature Overshoot Beyond 2°C: An Analysis of CMIP6 Models, Earth's Future, 9, e2020EF001967, https://doi.org/10.1029/2020EF001967, 2021.

Nicholls, Z., Meinshausen, M., Lewis, J., Smith, C. J., Forster, P. M., Fuglestvedt, J. S., Rogelj, J., Kikstra, J. S., Riahi, K.,
820 and Byers, E.: Changes in IPCC Scenario Assessment Emulators Between SR1.5 and AR6 Unraveled, Geophysical Research Letters, 49, e2022GL099788, https://doi.org/10.1029/2022GL099788, 2022.

Nicholls, Z., Meinshausen, M., Lewis, J., Corradi, M. R., Dorheim, K., Gasser, T., Gieseke, R., Hope, A. P., Leach, N. J., McBride, L. A., Quilcaille, Y., Rogelj, J., Salawitch, R. J., Samset, B. H., Sandstad, M., Shiklomanov, A., Skeie, R. B., Smith, C. J., Smith, S. J., Su, X., Tsutsui, J., Vega-Westhoff, B., and Woodard, D. L.: Reduced Complexity Model
825 Intercomparison Project Phase 2: Synthesizing Earth System Knowledge for Probabilistic Climate Projections, Earth's Future, 9, e2020EF001900, https://doi.org/10.1029/2020EF001900, 2021.

Nicholls, Z. R. J., Meinshausen, M., Lewis, J., Gieseke, R., Dommenget, D., Dorheim, K., Fan, C. S., Fuglestvedt, J. S., Gasser, T., Golüke, U., Goodwin, P., Hartin, C., Hope, A. P., Kriegler, E., Leach, N. J., Marchegiani, D., McBride, L. A., Quilcaille, Y., Rogelj, J., Salawitch, R. J., Samset, B. H., Sandstad, M., Shiklomanov, A. N., Skeie, R. B., Smith, C. J.,
830 Smith, S., Tanaka, K., Tsutsui, J., and Xie, Z.: Reduced Complexity Model Intercomparison Project Phase 1: introduction and evaluation of global-mean temperature response, Geosci. Model Dev., 13, 5175-5190, https://doi.org/10.5194/gmd-13-5175-2020, 2020.

Petrie, R., Denvil, S., Ames, S., Levavasseur, G., Fiore, S., Allen, C., Antonio, F., Berger, K., Bretonnière, P. A., Cinquini, L., Dart, E., Dwarakanath, P., Druken, K., Evans, B., Franchistéguy, L., Gardoll, S., Gerbier, E., Greenslade, M., Hassell, D.,
835 Iwi, A., Juckes, M., Kindermann, S., Lacinski, L., Mirto, M., Nasser, A. B., Nassisi, P., Nienhouse, E., Nikonov, S., Nuzzo, A., Richards, C., Ridzwan, S., Rixen, M., Serradell, K., Snow, K., Stephens, A., Stockhause, M., Vahlenkamp, H., and

Wagner, R.: Coordinating an operational data distribution network for CMIP6 data, Geosci. Model Dev., 14, 629-644, https://doi.org/10.5194/gmd-14-629-2021, 2021.

840 Qiu, H., Hao, D., Zeng, Y., Zhang, X., and Chen, M.: Global and northern-high-latitude net ecosystem production in the 21st century from CMIP6 experiments, Earth Syst. Dynam., 14, 1-16, https://doi.org/10.5194/esd-14-1-2023, 2023.

Rabin, S. S., Melton, J. R., Lasslop, G., Bachelet, D., Forrest, M., Hantson, S., Kaplan, J. O., Li, F., Mangeon, S., Ward, D. S., Yue, C., Arora, V. K., Hickler, T., Kloster, S., Knorr, W., Nieradzik, L., Spessa, A., Folberth, G. A., Sheehan, T., Voulgarakis, A., Kelley, D. I., Prentice, I. C., Sitch, S., Harrison, S., and Arneth, A.: The Fire Modeling Intercomparison Project (FireMIP), phase 1: experimental and analytical protocols with detailed model descriptions, Geosci. Model Dev., 10, 845 1175-1197, https://doi.org/10.5194/gmd-10-1175-2017, 2017.

Schulte-Uebbing, L. and de Vries, W.: Global-scale impacts of nitrogen deposition on tree carbon sequestration in tropical, temperate, and boreal forests: A meta-analysis, Global Change Biology, 24, e416-e431, https://doi.org/10.1111/gcb.13862, 2018.

Sokolov, A. P., Kicklighter, D. W., Melillo, J. M., Felzer, B. S., Schlosser, C. A., and Cronin, T. W.: Consequences of 850 Considering Carbon–Nitrogen Interactions on the Feedbacks between Climate and the Terrestrial Carbon Cycle, Journal of Climate, 21, 3776-3796, https://doi.org/10.1175/2008jcli2038.1, 2008.

Stocker, B. D. and Joos, F.: Quantifying differences in land use emission estimates implied by definition discrepancies, Earth Syst. Dynam., 6, 731-744, https://doi.org/10.5194/esd-6-731-2015, 2015.

Stouffer, R. J., Eyring, V., Meehl, G. A., Bony, S., Senior, C., Stevens, B., and Taylor, K. E.: CMIP5 Scientific Gaps and 855 Recommendations for CMIP6, Bulletin of the American Meteorological Society, 98, 95-105, https://doi.org/10.1175/BAMS-D-15-00013.1, 2017.

Tang, G., Nicholls, Z., Norton, A., Zaehle, S., and Meinshausen, M.: Coupled Carbon-Nitrogen Cycle in MAGICC v1.0.0: Model Description and Calibration, EGUsphere, 2024, 1-56, 10.5194/egusphere-2024-1941, 2024a.

Tang, G., Nicholls, Z., Jones, C., Gasser, T., Norton, A., Ziehn, T., Romero-Prieto, A., and Meinshausen, M.: Code and Data 860 for "Investigating Carbon and Nitrogen Conservation in Reported CMIP6 Earth System Model Data", Zenodo [dataset], https://doi.org/10.5281/zenodo.14060168, 2024b.

Thornhill, G., Collins, W., Olivié, D., Skeie, R. B., Archibald, A., Bauer, S., Checa-Garcia, R., Fiedler, S., Folberth, G., Gjermundsen, A., Horowitz, L., Lamarque, J. F., Michou, M., Mulcahy, J., Nabat, P., Naik, V., O'Connor, F. M., Paulot, F., Schulz, M., Scott, C. E., Séférian, R., Smith, C., Takemura, T., Tilmes, S., Tsigaridis, K., and Weber, J.: Climate-driven 865 chemistry and aerosol feedbacks in CMIP6 Earth system models, Atmos. Chem. Phys., 21, 1105-1126, https://doi.org/10.5194/acp-21-1105-2021, 2021.

Thornton, P. E., Lamarque, J.-F., Rosenbloom, N. A., and Mahowald, N. M.: Influence of carbon-nitrogen cycle coupling on land model response to CO_2 fertilization and climate variability, Global Biogeochemical Cycles, 21, https://doi.org/10.1029/2006GB002868, 2007.

870 Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., and Knutti, R.: Past warming trend constrains future warming in CMIP6 models, Science Advances, 6, eaaz9549, https://doi.org/10.1126/sciadv.aaz9549, 2020.

Turnock, S. T., Allen, R. J., Andrews, M., Bauer, S. E., Deushi, M., Emmons, L., Good, P., Horowitz, L., John, J. G., Michou, M., Nabat, P., Naik, V., Neubauer, D., O'Connor, F. M., Olivié, D., Oshima, N., Schulz, M., Sellar, A., Shim, S., Takemura, T., Tilmes, S., Tsigaridis, K., Wu, T., and Zhang, J.: Historical and future changes in air pollutants from CMIP6 875 models, Atmos. Chem. Phys., 20, 14547-14579, 10.5194/acp-20-14547-2020, 2020.

Wei, N., Xia, J., Zhou, J., Jiang, L., Cui, E., Ping, J., and Luo, Y.: Evolution of Uncertainty in Terrestrial Carbon Storage in Earth System Models from CMIP5 to CMIP6, Journal of Climate, 35, 5483-5499, https://doi.org/10.1175/JCLI-D-21-0763.1, 2022.

Wiltshire, A. J., Burke, E. J., Chadburn, S. E., Jones, C. D., Cox, P. M., Davies-Barnard, T., Friedlingstein, P., Harper, A. B., 880 Liddicoat, S., Sitch, S., and Zaehle, S.: JULES-CN: a coupled terrestrial carbon–nitrogen scheme (JULES vn5.1), Geosci. Model Dev., 14, 2161-2186, https://doi.org/10.5194/gmd-14-2161-2021, 2021.

Zaehle, S. and Dalmonech, D.: Carbon–nitrogen interactions on land at global scales: current understanding in modelling climate biosphere feedbacks, Current Opinion in Environmental Sustainability, 3, 311-320, https://doi.org/10.1016/j.cosust.2011.08.008, 2011.

885 Zaehle, S., Friedlingstein, P., and Friend, A. D.: Terrestrial nitrogen feedbacks may accelerate future climate change, Geophysical Research Letters, 37, https://doi.org/10.1029/2009GL041345, 2010.

Zaehle, S., Jones, C. D., Houlton, B., Lamarque, J.-F., and Robertson, E.: Nitrogen Availability Reduces CMIP5 Projections of Twenty-First-Century Land Carbon Uptake, Journal of Climate, 28, 2494-2511, https://doi.org/10.1175/JCLI-D-13-00776.1, 2015.

890   Zaehle, S., Medlyn, B. E., De Kauwe, M. G., Walker, A. P., Dietze, M. C., Hickler, T., Luo, Y., Wang, Y.-P., El-Masri, B., Thornton, P., Jain, A., Wang, S., Warlind, D., Weng, E., Parton, W., Iversen, C. M., Gallet-Budynek, A., McCarthy, H., Finzi, A., Hanson, P. J., Prentice, I. C., Oren, R., and Norby, R. J.: Evaluation of 11 terrestrial carbon–nitrogen cycle models against observations from two temperate Free-Air CO2 Enrichment studies, New Phytologist, 202, 803-822, https://doi.org/10.1111/nph.12697, 2014.

895   Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S. A., and Taylor, K. E.: Causes of Higher Climate Sensitivity in CMIP6 Models, Geophysical Research Letters, 47, e2019GL085782, https://doi.org/10.1029/2019GL085782, 2020.

Zhu, J., Gao, X., and Zeng, X.: Response of terrestrial net primary production to climate change associated with the quadrupling CO2 forcing in CMIP6 models, Atmospheric Science Letters, 23, e1098, https://doi.org/10.1002/asl.1098, 2022.

900