https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


# Consistency of global carbon budget between concentration- and emission-driven historical experiments simulated by CMIP6 Earth system models and suggestion for improved simulation of $CO_2$ concentration

Tomohiro Hajima[1], Michio Kawamiya[1], Akihiko Ito[2,1], Kaoru Tachiiri[1], Chris D. Jones[3,4], Vivek Arora[5], Victor Brovkin[6,7], Roland Séférian[8], Spencer Liddicoat[3], Pierre Friedlingstein[9], Elena Shevliakova[10]

1 Research Institute for Global Change, Japan Agency for Marine-Earth Science and Technology, Yokohama 236-0001, Japan
2 Graduate School of Life and Agricultural Sciences, the University of Tokyo, Tokyo, Japan
3 Met Office Hadley Centre, Exeter, UK
4 Department of Geographical Sciences, University of Bristol, Bristol, UK
5 Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, University of Victoria, Victoria, BC, Canada
6 Max-Planck-Institute for Meteorology, Hamburg, Germany
7 Center for Earth System Research and Sustainability, University of Hamburg, Germany
8 CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France
9 Faculty of Environment, Science and Economy, University of Exeter, Exeter, UK
10 NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey, USA

*Correspondence to*: Tomohiro Hajima (hajima@jamstec.go.jp)

**Abstract.** Anthropogenically emitted $CO_2$ from fossil fuel use and land use change is partly absorbed by terrestrial ecosystems and the ocean, while the remainder retained in the atmosphere adds to the ongoing increase in atmospheric $CO_2$ concentration. Earth system models (ESMs) can simulate such dynamics of the global carbon cycle and consider its interaction with the physical climate system. The ESMs that participated in the Coupled Model Intercomparison Project phase 6 (CMIP6) performed historical simulations to reproduce past climate–carbon cycle dynamics. This study investigated the cause of $CO_2$ concentration biases in ESMs and identified how they might be reduced. First, we compared simulated historical carbon budgets in two types of experiments: one with prescribed $CO_2$ emissions (the emission-driven experiment, "E-HIST") and the other with prescribed $CO_2$ concentration (the concentration-driven experiment, "C-HIST"). As CMIP7 design is being considered it is important to explore any differences or implications in what these variations can tell us. The findings confirmed that the multi-model means of the carbon budgets simulated by one type of experiment generally showed good agreement with those simulated by the other. However, the multi-model average of cumulative compatible fossil fuel emission diagnosed from the C-HIST experiment was lower by 35 PgC than that used as the prescribed input data to drive the E-HIST experiment; the multi-model average of simulated $CO_2$ concentration for 2014 in E-HIST was higher by 7 ppmv

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


than that used to drive C-HIST. Second, we investigated the potential linkages of two types of carbon cycle indices: simulated $CO_2$ concentration in E-HIST and compatible fossil fuel emission in C-HIST. It was confirmed quantitatively that the two indices are reasonable indicators of overall model performance in the context of carbon cycle feedbacks, although most models cannot accurately reproduce the cumulative compatible fossil fuel emission and thus cannot reproduce the $CO_2$ concentration precisely. Third, analysis of the atmospheric $CO_2$ concentration in five historical eras enabled identification of periods that caused the concentration bias in individual models. Further analysis based on a combination of four types of historical experiments suggested non-negligible impacts of non-$CO_2$ effects on the carbon cycle, implying their potential importance for future projections. It is suggested that this non-$CO_2$ effect is the reason why the magnitude of the natural land carbon sink in historical simulations is difficult to explain based on analysis of idealized experiments. Finally, accurate reproduction of land use change emission is critical for better reproduction of the global carbon budget and $CO_2$ concentration. The magnitude of simulated land use change emission not only affects the level of net land carbon uptake but also determines the magnitude of the ocean carbon sink in the emission-driven experiment.

## 1 Background and Objective

The observed increase in atmospheric $CO_2$ has been caused by anthropogenically emitted carbon. In the Global Carbon Budget report 2021 (GCB2021; Friedlingstein et al., 2021), the cumulative anthropogenic-related emissions of $CO_2$ from fossil fuel use and land use change during 1850–2021 are estimated at 465 ± 25 and 205 ± 65 PgC, respectively. Approximately half of the emitted carbon has been absorbed by the land and the ocean, both of which exhibit a similar level of carbon sink capacity in terms of their cumulative uptakes (i.e., 200 ± 45 PgC for land, 170 ± 35 PgC for the ocean). For the period 1850–2014, which corresponds to the "historical" period of the Coupled Model Intercomparison Project phase 6 (CMIP6; Eyring et al., 2015), GCB2021 states that those cumulative values are 400 ± 20 PgC for fossil fuel emission, 195 ± 60 PgC for land use change emission, 180 ± 40 PgC for land carbon uptake, and 150 ± 30 PgC for ocean carbon uptake. The numbers for land and ocean carbon uptake are slightly updated in GCB2022 (Friedlingstein et al., 2022) to be 185 ± 40 and 155 ± 30 PgC, respectively.

The carbon sink capacity of both the land and the ocean alters in response to environmental changes, and one of the major influencing processes is caused by atmospheric $CO_2$ concentration. Atmospheric $CO_2$ increase stimulates plant photosynthesis, which leads to accumulation of carbon as organic matter in plants and soils. Atmospheric $CO_2$ increase also drives the ocean carbon sink by accelerating $CO_2$ dissolution into the surface water, a certain amount of which is transported to the deeper ocean via oceanic circulation and biological processes. Consequently, these processes occurring over land and in the ocean buffer the rate of increase in atmospheric $CO_2$ concentration triggered by external forcing (e.g., anthropogenic emissions), and thus yield a negative feedback loop between atmospheric $CO_2$ and land/ocean carbon, named "$CO_2$–carbon feedback" or "concentration–carbon feedback" (Arora et al., 2020; Hajima et al., 2014b; Boer and Arora, 2009; Gregory et

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


al., 2009). There exists another type of carbon cycle feedback, named "climate–carbon feedback" (Friedlingstein et al., 2006; Boer and Arora, 2009; Gregory et al., 2009), which quantifies the response of the carbon cycle to climatic changes expressed in terms of temperature change. Warming of the surface air and soil accelerates land ecosystem respiration, leading to the loss of carbon from terrestrial ecosystems to the atmosphere. Similarly, warming of the upper ocean reduces $CO_2$ dissolution into the seawater, and global warming also prevents effective transport of dissolved carbon to the deeper ocean owing to greater oceanic stratification. Because this feedback process likely reduces the amount of carbon stored in the land and the ocean, it is regarded as a positive feedback loop between the climate system and the carbon cycle (Arora et al., 2020).

Historical change in the global carbon budget has been investigated by examining the component fluxes using both observations and stand-alone land, ocean, and bookkeeping models, which provide independent estimates of anthropogenic emissions of fossil fuel and land use change, natural sinks of the land and the ocean, and the rate of increase in atmospheric carbon (LeQuéré et al., 2018; Friedlingstein et al., 2021). These component fluxes and the changes in their magnitudes induced by external forcing and feedbacks can be simulated explicitly and consistently using Earth system models (ESMs), which integrate physical climate models (i.e., coupled atmosphere–ocean general circulation models) with models for land and ocean biogeochemistry (Hajima et al., 2014a; Kawamiya et al., 2020). Such models, with prescribed fossil fuel $CO_2$ emissions and scenarios of land use land cover change, simulated explicitly historical changes in atmospheric $CO_2$ and underlying land use change emissions, natural carbon land and ocean sinks, including their interaction with the physical climate. Because such a simulation is driven by prescribed fossil fuel $CO_2$ emissions, it is called an "emission-driven" (hereafter, "E-driven") experiment (Friedlingstein et al., 2014; Jones et al., 2016a). The historical E-driven experiment (named "esm-hist") comprises one of the core experiments in CMIP6 (Eyring et al., 2015).

Another type of historical experiment (named "historical") was conducted under the auspices of CMIP6. Such a simulation used as an input a prescribed $CO_2$ concentration pathway, and the configuration is called a concentration-driven (hereafter, "C-driven") experiment (Jones et al., 2013, 2016b; Liddicoat et al., 2021). The C-driven setting is necessary to drive conventional climate models that do not include land and ocean biogeochemistry components and therefore cannot predict $CO_2$ concentration prognostically. Additionally, the C-driven setting is necessary and favored for use in ESM simulations for several reasons. First, the $CO_2$ concentration in some idealized experiments is preferentially prescribed (e.g., abrupt $CO_2$ quadrupling experiment; 1% per year $CO_2$ increase experiments) such that such experiments can be performed with conventional climate models. Second, a C-driven experiment facilitates assessment of the effects of different forces (e.g. GHGs, short-lived climate forcers, land use land cover change) and feedback processes of climate–carbon cycle systems. One example is the evaluation of land use change emission, which is sometimes assessed by comparing two types of C-driven experiment, i.e., a normal historical experiment and a special historical experiment in which fractional coverage of land cover is fixed at the preindustrial level ("hist-noLu"; Lawrence et al., 2016). Such a fixed land use change experiment can also be conducted using the E-driven mode, enabling assessment of combined biophysical and biogeochemical

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


contribution of land use cover change to coupled climate-carbon cycle system (e.g., Shevliakova et al., 2013). However, in that case, land use change emissions in the E-driven simulation are diminished or completely stopped, leading to lower $CO_2$ concentration and land/ocean carbon uptake that make comparison to the normal E-driven historical experiment difficult. In this regard, various types of CMIP6 experiment have been designed to be run in the C-driven mode to assess the impact of external forcing and feedbacks (e.g., Eyring et al., 2015; Jones et al., 2016a; Lawrence et al., 2016; Gillett et al., 2016; Keller et al., 2018). Third, the $CO_2$ concentrations simulated by ESMs remain biased. Gier et al. (2020) compared CMIP6 ESMs' column-averaged $CO_2$ concentrations and found that the models have a concentration bias of −15 to +20 ppmv in comparison with that of satellite-derived observations for 2014. This large bias might prevent consistent comparison of the simulated climate between atmosphere–ocean general circulation models forced by $CO_2$-concentrations and ESMs forced by $CO_2$ emissions because the simulated atmospheric $CO_2$ concentration in experiments run in the E-driven mode is different compared to that used in experiments run in the C-driven mode. Finally, the results of a C-driven experiment allow a posteriori diagnosis of fossil fuel $CO_2$ emission, i.e., the "compatible fossil fuel $CO_2$ emission" (Jones et al., 2013; Liddicoat et al., 2021). These emissions are diagnosed from the simulated $CO_2$ fluxes and prescribed $CO_2$ concentration data used in C-driven experiments; such analysis helps elucidate future anthropogenic emission pathways to achieve specific global warming targets such as the +1.5°/+2.0° C goals of the Paris Agreement.

The compatible fossil fuel emissions in C-driven experiments are diagnosed to be consistent with the prescribed $CO_2$ concentration, and thus a model with stronger (weaker) natural ocean and land carbon sinks yields larger (smaller) compatible fossil fuel emissions. Compatible fossil fuel emissions therefore integrate the carbon cycle response to external forcings, and are an indicator that characterizes the total strength of the climate and carbon cycle feedbacks in models. E-driven experiments, however, project $CO_2$ concentration based on prescribed fossil fuel emissions, and a model with stronger (weaker) ocean and land carbon sinks yields a lower (higher) simulated $CO_2$ concentration, thereby making the simulated $CO_2$ concentration an indicator of model feedbacks. Thus, a model that can produce realistic anthropogenic emissions in a C-driven experiment is expected to reproduce atmospheric $CO_2$ concentrations in an E-driven experiment. However, this expectation is not always met when analyzing simulation results from CMIP6 ESMs. For example, Liddicoat et al. (2021) compared the cumulative value of the compatible fossil fuel emissions from the C-driven historical experiment. They showed that the result from one ESM (i.e., MIROC-ES2L; Hajima et al., 2020) was among the closest to the values reported in the Global Carbon Budget report 2019 (Friedlingstein et al., 2019); however, this model simulated a lower $CO_2$ concentration (by more than 10 ppmv) in the E-driven historical experiment (Gier et al., 2020). ACCESS-ESM1.5 (Ziehn et al., 2020), evaluated in Gier et al. (2020), realistically reproduced the $CO_2$ concentration in the E-driven historical experiment, but its cumulative fossil fuel emissions in the C-driven experiment were over 500 PgC (Liddicoat et al., 2021), i.e., 25% higher than that stated in the Global Carbon Budget report 2019. These examples illustrate that models that realistically reproduce compatible fossil fuel emissions in C-driven experiments do not necessarily reproduce the historical $CO_2$ concentration in E-driven experiments with the same consistency. The simulation results of C-driven and E-driven

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


historical experiments have been analyzed separately in a number of previous studies (Friedlingstein et al., 2014; Jones et al., 2013; Gier et al., 2020; Liddicoat et al., 2021), whereas only a limited number of studies (Friedlingstein et al., 2014) have conducted quantitative comparisons of multiple ESMs in terms of the level of consistency between the two types of historical experiments.

Historical experiments performed in the E-driven mode have the advantage of yielding simulations that capture the chain of the climate–carbon cycle processes that occur in the real world, whereas many CMIP6 experiments are designed to be performed with the C-driven setting for many of the reasons mentioned above. One of the advantages of C-driven experiments is the ability to identify the effect of individual drivers of global change in the coupled climate–carbon cycle systems that arise from external forcings on and feedbacks of the Earth system. Thus, analysis of C-driven experiments is useful for identifying the carbon cycle processes that are inadequately simulated in each model, which also could hint at potential improvements in model performance in E-driven experiments. Specifically, bias in the simulated $CO_2$ concentration in E-driven experiments over the historical period is suggested to become amplified in future climate projections (Friedlingstein et al., 2014; Hoffman et al., 2014); therefore, reduction of the bias in the simulated $CO_2$ concentration is recognized as one of the most urgent objectives for improving ESM performance. As Earth System models evolve, and the science they enable becomes more relevant it is important to fully understand implications of experimental design choices. It is clear that E-driven simulations enable a fuller sampling of the range of uncertainty (Lee et al. 2021) in the evaluation of historical runs, and Sanderson et al. (2023) recommend a fundamental move of modelling towards E-driven as default. Here we assess implications of such a choice and make some recommendations to enable full optimization of the resulting simulations. In this study, we first compared the global carbon budget of C-driven and E-driven historical experiments simulated by CMIP6 ESMs to confirm the level of consistency between these two types of experiment. Then, the linkages between the results of the C-driven and E-driven experiments were further investigated, focusing on the extent to which the C-driven simulations could explain the results of the E-driven experiments. Finally, we investigated how the $CO_2$ concentration simulated in ESMs could be improved. The models, simulations, and analysis methods used in the study are described in Sect. 2. The main results of the analysis and a discussion are presented in Sect. 3. Suggestions for improved simulation of $CO_2$ concentration are summarized in Sect. 4. Finally, a summary and our conclusions are presented in Sect. 5.

## 2 Methods

### 2.1 CMIP6 Experiments

Details of the CMIP6 experiments analyzed in this study are summarized in Table 1. Historical simulations obtained with the E-driven mode “esm-hist” (hereafter, E-HIST) were used for analysis (12 models in total) after correcting for the drift found in the preindustrial control experiment “esm-piControl.” The correction was made by simply subtracting the drift found in “esm-piControl.” Similarly, historical simulation results obtained with the C-driven mode “historical” (hereafter, C-HIST)

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


were also used for analysis (14 models) after correction of the drift found in the corresponding C-driven preindustrial control experiment "piControl." These drift corrections were applied to the variables of $CO_2$ concentration, cumulative nbp and fgco2.

To decompose the historical response of the carbon cycle to external forcings, the results of other types of C-driven experiments were also analyzed in this study. First, "hist-noLu" (hereafter, C-HIST-NOLU), which uses the preindustrial land use state throughout the entire simulated historical period, was used to diagnose land use change emissions that include foregone sink, i.e. loss of additional sink capacity due to historical land use land cover change (Ciais et al., 2022). The "hist-bgc" experiment (hereafter, C-HIST-BGC), an experiment that comprises C4MIP simulations (Jones et al., 2016a), was used to analyze carbon cycle feedbacks in the historical simulation. In this experiment, the radiation processes "see" a constant $CO_2$ concentration fixed at its preindustrial level but the carbon cycle processes "see" the changes in $CO_2$ over the historical period. Thus, because $CO_2$-induced climate change is suppressed, taking the difference between C-HIST-BGC and C-HIST enables quantification of the climate–carbon feedback in the models (Friedlingstein et al., 2006). Results from only two models were available for this experiment (Table 2). Third, "hist-CO2" (hereafter, C-HIST-CO2) was also used for the analysis. This historical experiment was proposed as part of the Detection and Attribution Model Intercomparison Project (DAMIP; Gillett et al., 2016) and designed to run in the C-driven mode. External forcings other than $CO_2$ (including non-$CO_2$ GHG concentrations, aerosol emissions, land use change, and nitrogen deposition) were fixed at their preindustrial level, and the prescribed $CO_2$ concentration pathway was identical to that used for C-HIST. This experiment is useful for separating the responses of the climate and the carbon cycle to $CO_2$ alone compared to those induced by non-$CO_2$ forcings. Results from only two models were available for this experiment. Finally, an idealized experiment ("1pctCO2"; hereafter, C-1PCT) was also used in this study. In this experiment, $CO_2$ concentration was increased by 1% annually and all other external forcings were fixed at their preindustrial level. This experiment was used to investigate the linkages between this idealized simulation and more realistic historical simulations.

Some of the simulation results mentioned above have already been considered in previous studies. For example, analysis of the multi-model simulated $CO_2$ concentration from E-HIST experiments has already been presented in Gier et al. (2020); the compatible fossil fuel emission and global carbon budgets from C-HIST experiments have been investigated by Liddicoat et al. (2021); the carbon cycle response to land use change scenarios was analyzed by both Liddicoat et al. (2021) and Ito et al. (2020); and the results from an idealized experiment C-1PCT have been utilized in analysis of carbon cycle feedbacks by Arora et al. (2020). In recognition of those previous studies, this study focused mainly on examining the linkages of the global carbon budgets between these multiple experiments. For this purpose, the simulated variables were reanalyzed in this study using a different detrending method, analysis period, and target models.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


Table 1. CMIP6 experiments used in this study, and the combinations of experiments used to decompose the global carbon budgets into the changes induced by four types of drivers: (1) $CO_2$–carbon feedback ("CO2-BGC"), (2) climate–carbon feedback caused by $CO_2$ increase ("CO2-CLIM"), (3) land use change ("LUC"), and (4) non-$CO_2$ agents ("NONCO2"). Note that all simulation results shown hereafter were corrected by removing the trend found in the preindustrial control run ("piControl" for the C-driven mode, "esm-hist" for the E-driven mode).

| Configuration of $CO_2$ | Experiment type | Formal name in CMIP6 | Abbreviated name in this study | Drivers to change carbon cycle | | | | References for multi-model analysis |
|---|---|---|---|---|---|---|---|---|
| | | | | CO2-BGC | CO2-CLIM | LUC | NONCO2 | |
| Emission-driven | (PI control) | (esm-piControl (DECK[1])) | (E-PI) | – | – | – | – | |
| | Historical | esm-historical (DECK[1]) | E-HIST | ○ | ○ | ○ | ○ | Gier et al. (2020) |
| | | | | | | | | |
| Concentration-driven | (PI control) | (piControl (DECK[1])) | (C-PI) | – | – | – | – | |
| | Historical | historical (DECK[1]) | C-HIST (**A**) | ○ | ○ | ○ | ○ | Liddicoat et al. (2021) |
| | | hist-noLu (LUMIP[2]) | C-HIST-NOLU (**B**) | ○ | ○ | – | ○ | Ito et al. (2020) |
| | | hist-bgc (C4MIP[3]) | C-HIST-BGC (**C**) | ○ | – | ○ | ○ | |
| | | hist-CO2 (DAMIP[4]) | C-HIST-CO2 (**D**) | ○ | ○ | – | – | |
| | | | | | | | | |
| | Idealized | 1pctCO2 (DECK[1]) | C-1PCT | ○ | ○ | – | – | Arora et al. (2020) |
| | | | | | | | | |
| Concentration-driven | Combination of experiments | hist-CO2 – historical + hist-bgc | **D–(A-C)** | ○ | | | | |
| | | historical – hist-bgc | **A-C** | | ○ | | | |
| | | historical – hist-noLu | **A-B** | | | ○ | | Liddicoat et al. (2021) |
| | | hist-noLu – hist-CO2 | **B-D** | | | | ○ | |

1. "Diagnostic, Evaluation, and Characterization of Klima," Eyring et al. (2016)
2. "Land Use Model Intercomparison Project," Lawrence et al. (2016)
3. "Coupled Climate–Carbon Cycle Model Intercomparison Project," Jones et al. (2016a)
4. "Detection and Attribution Model Intercomparison Project," Gillett et al. (2016)

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


## 2.2 Models

This study analyzed 12 CMIP6 ESMs for which simulation results of both C-HIST and E-HIST are available (Table 2). Furthermore, because land use change emission, which can be diagnosed from C-HIST-NOLU experiment, is an important component of the global carbon budget, two ESMs (i.e., CMCC-ESM2 and IPSL-CM6A-LR) were added to the list of target models, although corresponding E-HIST results are unavailable.

Table 2. List of participating CMIP6 ESMs and the experiment outputs used in this study. Circle symbols and "NA" represent model data availability and unavailability, respectively, for the experiments. The "–" symbol represents a model/experiment for which data are available but were not analyzed in this study; C-HIST-BGC results from 9 ESMs are available but only 2 were used for analysis because the analysis needed both C-HIST-BGC and C-HIST-CO2 results; C-1PCT results are available for all ESMs but this study used 10 models that also provided C-HIST-NOLU results; C-HIST-NOLU results of NorESM2-LM were not used in this study because of the quality of land carbon flux data in the experiment.

| | E-driven (emission-driven) | | C-driven (concentration-driven) | | | | | C-driven, idealized | |
|---|---|---|---|---|---|---|---|---|---|
| | E-PI (esm-piControl) | E-HIST (esm-historical) | C-PI (piControl) | C-HIST (historical) | C-HIST-NOLU (hist-noLu) | C-HIST-BGC (hist-bgc) | C-HIST-CO2 (hist-CO2) | C-1PCT (1pctCO2) | References |
| ACCESS-ESM1-5 | ○ | ○ | ○ | ○ | ○ | – | NA | ○ | Ziehn et al. 2020 |
| CESM2 | ○ | ○ | ○ | ○ | ○ | – | NA | ○ | Danabasoglu et al. 2020 |
| CMCC-ESM2 | NA | NA | ○ | ○ | ○ | NA | NA | ○ | Lovato et al. 2022 |
| CNRM-ESM2-1 | ○ | ○ | ○ | ○ | ○ | – | NA | ○ | Séférian et al. 2019 |
| CanESM5 | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Swart et al. 2019 |
| CanESM5-CanOE | ○ | ○ | ○ | ○ | NA | NA | NA | – | Swart et al. 2019 |
| EC-Earth3-CC | ○ | ○ | ○ | ○ | NA | NA | NA | – | Döscher et al. 2021 |
| GFDL-ESM4 | ○ | ○ | ○ | ○ | ○ | – | NA | ○ | Dunne et al. 2020 |
| IPSL-CM6A-LR | NA | NA | ○ | ○ | ○ | – | NA | ○ | Boucher et al. 2020 |
| MIROC-ES2L | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Hajima et al. 2020 |
| MPI-ESM1-2-LR | ○ | ○ | ○ | ○ | ○ | – | NA | ○ | Mauritsen et al. 2019 |
| MRI-ESM2-0 | ○ | ○ | ○ | ○ | NA | NA | NA | – | Yukimoto et al. 2019 |
| NorESM2-LM | ○ | ○ | ○ | ○ | – | NA | NA | – | Seland et al. 2020 |
| UKESM1-0-LL | ○ | ○ | ○ | ○ | ○ | – | NA | ○ | Sellar et al. 2019 |
| | | | | | | | | | |
| Number of models | 12 | 12 | 14 | 14 | 10 | 2 | 2 | 10 | |

## 2.3 Definition of analyzed variables and global carbon budget equations

The global carbon budget can be expressed using five terms:

$$E_{FF}(t) + E_{LUC}(t) = C_A(t) + C_O(t) + C_{LN}(t),$$

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


where $E_{FF}(t)$ is the cumulative emission from fossil fuels from 1850 to $t$, and $E_{LUC}(t)$ is that from net land use change (i.e., carbon emission derived from vegetation disturbances (e.g., deforestation and crop harvesting) minus carbon uptake by plant regrowth after the disturbances), and $C_A(t)$, $C_O(t)$, and $C_{LN}(t)$ represent the change in carbon amount in the atmosphere, ocean, and natural land ecosystem, respectively. In this expression, $C_{LN}(t)$ is equivalent to the cumulative carbon uptake by land where land-use status is fixed at the preindustrial condition. Calculations of the cumulative values start from 1850. Hereafter, for concise expression, we drop the expression "($t$)" from the above equation:

$$E_{FF} + E_{LUC} = C_A + C_O + C_{LN}. \quad (1a)$$

Using a term of land carbon change that includes land use change impact ($C_L$), this equation can be rewritten as follows:

$$E_{FF} = C_A + C_O + (C_{LN} - E_{LUC}) = C_A + C_O + C_L, \text{ where } C_L = (C_{LN} - E_{LUC}). \quad (1b)$$

In this expression, land use change emission $E_{LUC}$ becomes implicit and is incorporated into $C_L$. In most cases, ESMs simulate land carbon fluxes in this way. See appendix A for more details regarding the derivation of Eq. 1b.

On the basis of Eq. (1b), the method of simulation in the E-driven mode (i.e., the change in atmospheric $CO_2$ concentration is simulated explicitly) can be summarized as follows:

$$C_A^{E\text{-}HIST} = E_{FF}^{CMIP6F} - (C_O^{E\text{-}HIST} + C_L^{E\text{-}HIST}), \quad (2)$$

where superscript *E-HIST* represents the historical experiment with the E-driven mode, and *CMIP6F* implies the forcing prescribed by CMIP6. Using this prescribed fossil fuel emission rate ($E_{FF}^{CMIP6F}$) and the prescribed land cover change, models simulate the change in atmospheric $CO_2$ concentration (presented here as atmospheric carbon burden change, $C_A^{E\text{-}HIST}$) based on simulation of the land and ocean fluxes ($C_O^{E\text{-}HIST}$ and $C_L^{E\text{-}HIST}$, respectively) that are affected by the carbon cycle and other feedbacks.

The calculation of compatible fossil fuel emission in the C-driven historical experiment (C-HIST) can be expressed as follows:

$$E_{FF}^{C\text{-}HIST} = C_A^{CMIP6F} + C_O^{C\text{-}HIST} + C_L^{C\text{-}HIST}. \quad (3a)$$

By prescribing the $CO_2$ concentration (presented here as atmospheric carbon change, $C_A^{CMIP6F}$), models can simulate the ocean and land carbon fluxes ($C_O^{C\text{-}HIST}$ and $C_L^{C\text{-}HIST}$, respectively) that reflect both climate and carbon cycle feedbacks and the impacts from other external forcing. Through a posteriori summation of $C_A^{CMIP6F}$, $C_O^{C\text{-}HIST}$, and $C_L^{C\text{-}HIST}$, the cumulative value of the compatible fossil fuel emission $E_{FF}^{C\text{-}HIST}$ can be diagnosed.

When the natural carbon uptake by land ($C_{LN}^{C\text{-}HIST}$) and land use change emission ($E_{LUC}^{C\text{-}HIS}$) can be assessed through combination of other historical experiment (i.e., "hist-noLu"), this expression can be rewritten as follows:

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


$$E_{FF}^{C\text{-}HIST} = C_A^{CMIP6F} + C_O^{C\text{-}HIST} + (C_{LN}^{C\text{-}HIST} - E_{LUC}^{C\text{-}HIST}). \quad (3b)$$

The analysis performed in this study was based on Eqs. 1–3, and three types of variables were used (Table 3). The first variable was atmospheric $CO_2$ concentration, which is the three-dimensional atmospheric $CO_2$ concentration named "*co2*" in CMIP6. In this study, the globally averaged concentration (hereafter, "*CO2*") was analyzed. Using this variable, $C_A$ was calculated as $C_A = \Delta CO2 \times 2.124$, where 2.124 is the ppmv–PgC conversion factor (Prather et al., 2012). If the "*co2*" variable was unavailable, the global mass of atmospheric carbon "*co2mass*" was used instead in the analysis. The second and third variables, named "*nbp*" and "*fgco2*" in CMIP6, represent the rate of $CO_2$ exchange between the atmosphere and the land biosphere and the ocean, respectively; in this analysis, these fluxes were analyzed after being converted to cumulative values, i.e., $\Delta C_L = \int_{1850}^{t} nbp\, dt$ and $\Delta C_O = \int_{1850}^{t} fgco2\, dt$., respectively. As mentioned above, the drift corrections were applied to these cumulative variables because models are not necessarily fully-equilibrated in piControl run (Appendix B).

Although the scope of this study focused primarily on global carbon cycle processes, it was considered valuable to compare the simulated global mean surface air temperature (hereafter, "*GSAT*") between C-HIST and E-HIST. Therefore, GSAT was calculated from the "*tas*" variable in CMIP6, and drift correction was performed by evaluating the GSAT trend linearly in the preindustrial control experiment and subtracting the trend from the simulated GSAT in C-HIST and E-HIST (Fig. S1).

Table 3 Definition of variables.

| Variable Name | Description | Unit | Original variable name in CMIP6 |
|---|---|---|---|
| *CO2* | Global mean $CO_2$ concentration | [ppmv] | *Amon/co2, co2mass* |
| $E$ | Cumulative Anthropogenic Carbon Emission | [PgC] | – |
| $E_{FF}$ | Cumulative (Compatible) Fossil Fuel Carbon Emission | [PgC] | – |
| $E_{LUC}$ | Cumulative Land-Use Change Carbon Emission | [PgC] | – |
| $C_A$ | Change in Atmospheric Carbon | [PgC] | – |
| $C_{LN}$ | Change in Land Carbon (without Land-use change impact) | [PgC] | – |
| $C_L$ | Change in Land Carbon (with Land-use change impact) | [PgC] | *Lmon/nbp* |
| $C_O$ | Change in Ocean Carbon | [PgC] | *Omon/fgco2* |
| *GSAT* | Global-mean Surface Air Temperature | [deg.C] | *Amon/tas* |

### 2.4 Analysis procedures and variables

The analysis procedures adopted are summarized in this section and illustrated in Figs. 1 and 2. Figure 1 summarizes the first stage of the multi-model comparison, the purpose of which was to confirm the level of consistency between the C-HIST and E-HIST experiments regarding fundamental terms of the global carbon budget. This stage consists of the following steps:

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


(1) comparison of the prescribed fossil fuel emissions used for E-HIST ($E_{FF}^{CMIP6F}$) with compatible emissions obtained from the C-HIST experiment ($E_{FF}^{C\text{-}HIST}$);

(2) comparison of the prescribed $CO_2$ concentration used for C-HIST ($CO2^{CMIP6F}$) with the simulated concentration in E-HIST ($CO2^{E\text{-}HIST}$);

(3) comparison of simulated ocean and land carbon uptake in C-HIST ($C_O^{C\text{-}HIST}$ and $C_L^{C\text{-}HIST}$, respectively) with those of E-HIST ($C_O^{E\text{-}HIST}$ and $C_L^{E\text{-}HIST}$, respectively); and

(4) comparison of GSAT between C-HIST and E-HIST ($GSAT^{C\text{-}HIST}$ and $GSAT^{E\text{-}HIST}$, respectively).

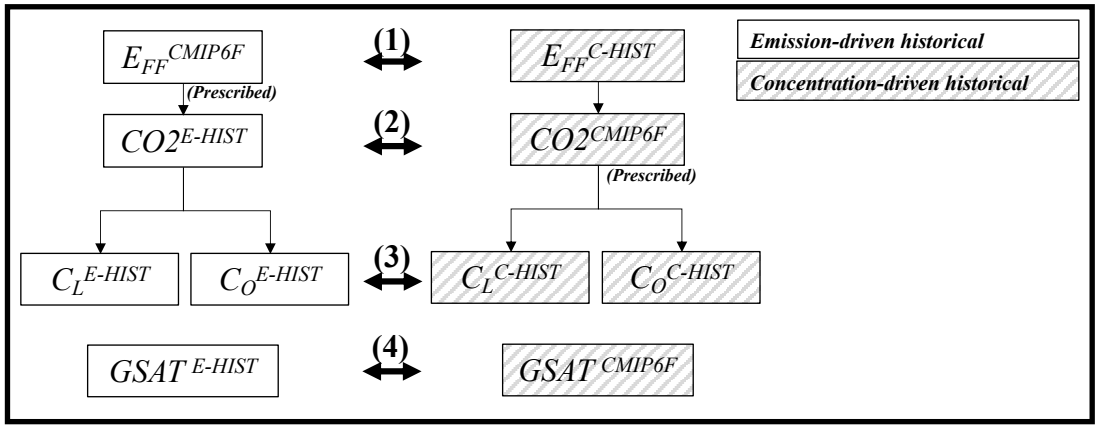



Figure 1 Analysis flow chart 1. The purpose of this series of analyses was to confirm the level of consistency between the historical E-driven experiments (E-HIST, left) and the C-driven experiments (C-HIST, right). Definitions of the variables and experiments can be found in Tables 1, 2, and 3. Solid arrows represent the decomposition of the global carbon budget, and the bold arrows with numbers represent the comparison steps. All analyses were performed after removing the trend found in the preindustrial control experiments (C-PI and E-PI).

The second stage of the multi-model comparison, which consists of three steps, was designed to investigate the linkages of E-driven historical experiments with other C-driven experiments, as summarized in Fig. 2.

(1) A potential linkage between atmospheric $CO_2$ concentration simulated in the E-HIST experiment ($CO2^{E\text{-}HIST}$) and compatible anthropogenic emissions of C-HIST ($E_{FF}^{C\text{-}HIST}$, see Eq. 3a and Table 3) was investigated because both are important quantities in summarizing the carbon cycle processes in each experimental configuration.

(2) Compatible fossil fuel emission ($E_{FF}^{C\text{-}HIST}$) was compared with the diagnosed land use change emission ($E_{LUC}^{C\text{-}HIST}$) in C-HIST to investigate the potential relationship between them in C-HIST because their negative correlation was identified in

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


a previous study (Liddicoat et al. 2021). In this study, $E_{LUC}^{C\text{-}HIST}$ was estimated by taking the difference between the results of C-HIST (normal historical simulations) and C-HIST-NOLU (historical simulation without land use change), as presented in Table 1 and Eqs. 3a and 3b. We note that the diagnosis method of land-use change emission is under debate and that the $E_{LUC}^{C\text{-}HIST}$ diagnosed in this study may yield a different magnitude of the cumulative emission compared with other approaches (Obermeier et al., 2021; Ciais et al., 2022).

(3) $E_{FF}^{C\text{-}HIST}$ depends on the magnitude of natural carbon sinks of the land and the ocean (Eq. 3b), which are affected by carbon cycle feedbacks. The carbon cycle feedbacks in models have been widely measured using the idealized experiment C-1PCT. Thus, in this study, the magnitude of the land and ocean natural sinks were compared between the realistic historical simulation (C-HIST) and the idealized experiment (C-1PCT). For this analysis, historical land carbon change without land use change impact ($C_{LN}^{C\text{-}HIST}$) was required because land carbon change is simulated in the idealized experiment in that way.

On the basis of the analysis steps, the linkages between the C-HIST and E-HIST historical experiments were investigated. In the analysis, it was assumed that the major difference between the two types of experimental configuration reflects whether the simulated carbon fluxes change the $CO_2$ concentration (E-HIST) or not (C-HIST). We note, however, that there could be other reasons that might cause systematic differences in the carbon cycle behavior between the two types of experiment. First, the $CO_2$ concentration in E-HIST is usually simulated using a three-dimensional field with sub-daily time steps, while that in C-HIST might be spatially homogeneous or longitudinally averaged with annual or seasonal time steps. This might affect the geographical and seasonal pattern of natural carbon sinks. Second, because of the difference in the spatial distribution of $CO_2$ concentration, the radiative forcing that arises from $CO_2$ might also be different between the two types of experiment, affecting the meteorological conditions over the land and ocean surfaces and altering carbon cycle and/or biophysical feedbacks. Finally, because of the differences mentioned above, the spin-up procedure is usually performed separately for each experiment and the spin-up duration might be different. This can cause different initial states in the climate and the carbon cycle system between C-HIST and E-HIST, and thus might affect the historical change in the global carbon budget.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024

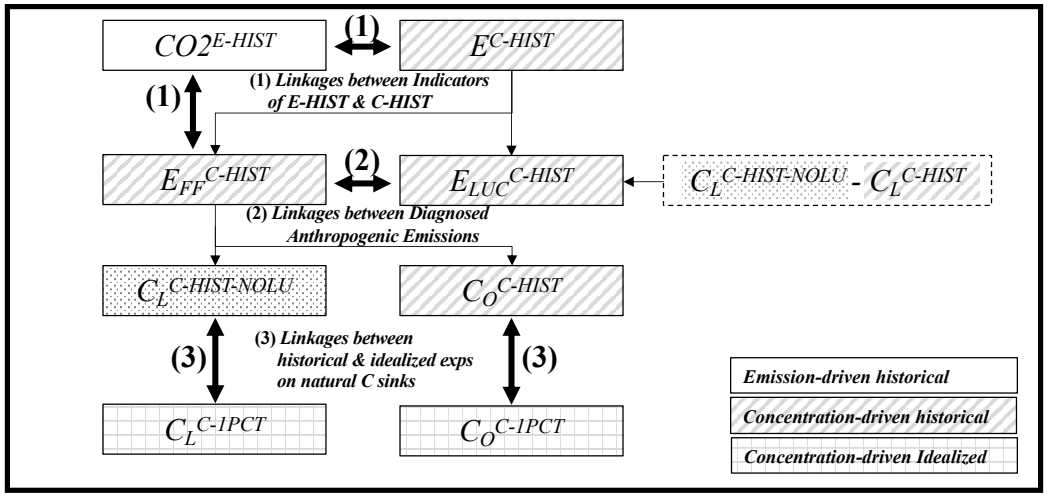




Figure 2 Analysis flow chart 2. The purpose of this series of analyses was to investigate the linkages between E-driven historical experiments and other C-driven experiments. Solid arrows represent the decomposition of the global carbon budget, and the bold arrows with numbers represent the comparison steps. Definitions of the variables and experiments can be found in Tables 1, 2, and 3. All analyses were performed after removing the trend found in the preindustrial control experiments (C-PI and E-PI).

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


## 3 Results and Discussion

### 3.1 Result of Analysis 1: Consistency between emission- and concentration-driven historical simulations

The results of analysis 1 (Fig. 1), i.e., the comparison between C-HIST and E-HIST experiments with regard to the basic components of the global carbon budget, are shown in Table 4 and Fig. 3. In Table 4, the global carbon budgets are presented as cumulative values (except for $CO_2$ concentration) during 1850–2014. Inspection of Fig. 3 confirms that the multi-model averages of C-HIST and E-HIST generally show reasonable agreement for the temporal changes in fossil fuel and land use emission, $CO_2$ concentration, and carbon uptake by the ocean and by the land. However, several discrepancies exist between the two experiments. First, the multi-model average of compatible fossil fuel emission is 374 PgC, which is smaller by 35 PgC than that used as the prescribed emission for E-HIST. Second, during 1900–1950, the carbon budget components of fossil fuel emission, change in $CO_2$ concentration, and carbon uptake by the land and by the ocean in E-HIST are slightly smaller than those in C-HIST. This suggests that it is difficult for E-HIST to reproduce the $CO_2$ concentration plateau observed in ice core measurements, which is discussed later. Third, the multi-model average of simulated $CO_2$ concentration at the end of E-HIST is 405 ppmv, which is larger than that of the prescribed concentration used for C-HIST by approximately 7 ppmv. Fourth, because of the variation of simulated $CO_2$ concentration between the models (405 ± 14.4 ppmv in 2014), E-HIST exhibits larger spread of GSAT (by 30%) than that of C-HIST, for the case when the models use a common $CO_2$ concentration pathway.

In comparison with the numbers reported in the best estimates of the global carbon budget, i.e., GCB2021, the multi-model average of compatible fossil fuel emission diagnosed from C-HIST is smaller by approximately 26 PgC. Additionally, the prescribed fossil fuel emission used for E-HIST has minor discrepancies with GCB2021. In CMIP6, the emissions were based on a Community Emissions Data System approach (Hoesly et al., 2018) that produces $CO_2$ emissions consistent with all the other species. Fundamentally, the Community Emissions Data System is sector-based and not fuel-based; consequently, the cumulative CMIP6 emissions are higher than the GCB emissions by approximately 10 PgC (Table 4; Andrew, 2020), mainly in the period 1950–1999. Moreover, the total anthropogenic $CO_2$ emissions in CMIP6 consisted of the sum of all sectors in the two-dimensional files plus the three-dimensional emissions in the aircraft $CO_2$ emissions files. Although most modeling groups likely used this summation to force emission-driven experiments, some might have neglected the aviation emissions, and the corresponding discrepancy in the simulated atmospheric $CO_2$ concentration by 2014 could be of the order of several parts per million.

The multi-model average of land use change emission, which was diagnosed by taking the difference between C-HIST and C-HIST-NOLU, is 129 PgC. This is much smaller (by approximately 65 PgC) than that of the GCB2021 estimation. This large discrepancy between the CMIP6 ESMs and GCB2021 might arise from the different assumptions, definitions, or approaches adopted for the land use change emission. In the CMIP6 simulations, the land use change emission is

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


interactively computed in the transient historical simulation. Thus, the emission calculation is subject to the effect of environmental changes, e.g., carbon emission from deforestation and carbon uptake through forest regrowth are simulated under time-varying $CO_2$ concentration and climate change. In addition, the models analyzed here for land-use change emissions (Table 4), except for three models of GFDL-ESM4, MIROC-ES2L, and MPI-ESM1-2-LR, consider net land-use changes ( i.e. concurrent, bidirectional transformations between land-use types within a grid cell are not considered; Ito et al., 2020), which may lead to an underestimation of the magnitude of land-use change emission as pointed out previously (e.g., Ciais et al., 2022; Friedlingstein et al., 2021). Meanwhile, GCB2021 adopts a bookkeeping method for estimating the land use change emission with gross transition of land-use changes. In addition, the method usually assumes constant biomass throughout the historical period for the emission calculation; because contemporary biomass is used for the calculation, the emission from deforested biomass and the loss of additional carbon sink could be larger than that of the ESM simulations (Obermeier et al., 2021; Friedlingstein et al., 2021).

The ocean carbon sink is simulated to be 137 ± 10 PgC in C-HIST and 145 ± 17 PgC in E-HIST, both of which are slightly lower than that of the GCB2021 estimate (150 ± 30 PgC). The natural land carbon sink is simulated by the CMIP6 ESMs to be 148 ± 31 PgC (C-HIST-NOLU), whereas GCB2021 has a value of 180 ± 40 PgC. We note that simulated climate variability could change the magnitudes of simulated $CO_2$ uptakes, causing different magnitudes of cumulative land and ocean uptake in different ensemble members of the historical experiment. However, examination using multiple ensemble members of C-HIST, performed using MIROC-ES2L (30 members) and UKESM-1-0-LL (12 members), reveals that the impact of internal climate variability on these cumulative quantities is small (Table S1). The multi-ensemble spread of cumulative terrestrial carbon uptake is ±4.1 PgC for MIROC-ES2L and ±4.1 PgC for UKESM, which is only approximately 6% of the multi-model spread (±74.2 PgC; Table 4); for the ocean, the multi-ensemble spread is confirmed to be ±0.8 PgC for MIROC-ES2L and ±1.4 PgC for UKESM, i.e., both substantially smaller than the multi-model spread of ±10.7 PgC.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


Table 4 Global carbon budget simulated in concentration- and emission-driven historical simulations (C-HIST and E-HIST, respectively) performed by CMIP6 ESMs and the difference ("Diff."). Results of global mean surface air temperature (GSAT) are also presented. Note that these results are based on a single simulation of each C- and E-driven historical experiment. The analysis corresponds to steps (1), (2), and (3) in Fig. 1. The simulated period is 1850–2014, and the results of the global carbon budgets are quantified as cumulative values and evaluated at the end of the historical simulation. GSAT is presented by the anomalies of GSAT (2005–2014 average). Definitions of the variables and the experiments can be found in Tables 1, 2, and 3. Corresponding GCB2021 values are also presented at the bottom of the table. Unavailability of data is represented by the "–" symbol.

| | Atmosphere | | | Ocean | | | Land | | | | | Emissions | | | Imbalance | GSAT | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $CO_2$ [ppmv] | | | $C_O$ [PgC] | | | $C_L$ [PgC] | | | $C_{LN}$ [PgC] | $E_{LUC}$ [PgC] | $E_{FF}$ [PgC]* | | | IB [PgC]† | [°C]‡ | |
| | C-HIST (Prescribed) | E-HIST | Diff. | C-HIST | E-HIST | Diff. | C-HIST | E-HIST | Diff. | C-HIST -NOLU | C-HIST-NOLU minus C-HIST | C-HIST | E-HIST (Prescribed) | Diff. | Evaluated by E-HIST only | C-HIST | E-HIST |
| ACCESS-ESM1-5 | | 387.9 | -9.6 | 154.0 | 117.8 | -36.2 | 107.0 | 81.0 | -26.1 | 133.0 | 26.0 | 501.6 (501.5) | | -92.3 | 0.1 | 0.80 | 0.53 |
| CESM2 | | 413.5 | 15.9 | 135.9 | 154.1 | 18.3 | -30.2 | -18.1 | 12.2 | 146.4 | 176.7 | 349.8 (346.2) | | 63.0 | 3.7 | 0.90 | 1.18 |
| CMCC-ESM2 | | – | – | 129.6 | – | – | 7.4 | – | – | 111.1 | 103.7 | – (377.5) | | 31.7 | – | 1.13 | – |
| CNRM-ESM2-1 | | 401.3 | 3.8 | 129.1 | 141.2 | 12.1 | 149.7 | 156.8 | 7.1 | 173.2 | 23.5 | 382.4 (519.3) | | -110.2 | -137.0 | 0.92 | 0.65 |
| CanESM5 | | 406.8 | 9.3 | 133.9 | 135.6 | 1.7 | -14.8 | 11.2 | 26.0 | 122.9 | 137.7 | 362.8 (359.6) | | 49.6 | 3.2 | 1.37 | 1.51 |
| CanESM5-CanOE | | 409.8 | 12.2 | 127.5 | 128.8 | 1.4 | -18.5 | 13.0 | 31.5 | – | – | 353.4 (349.5) | | 59.7 | 3.9 | 1.49 | 1.46 |
| EC-Earth3-CC | 397.6 | 436.1 | 38.6 | 128.3 | 174.9 | 46.7 | -130.3 | -93.2 | 37.1 | – | – | 245.2 (238.4) | 409.2 | 170.8 | 6.7 | 1.43 | 1.45 |
| GFDL-ESM4 | | 419.9 | 22.3 | 148.2 | 179.6 | 31.4 | -120.4 | -84.8 | 35.6 | 121.0 | 241.5 | 291.8 (268.2) | | 141.0 | 23.6 | 0.52 | 1.17 |
| IPSL-CM6A-LR | | – | – | 123.4 | – | – | 60.1 | – | – | 159.1 | 99.0 | – (423.9) | | -14.8 | – | 0.93 | – |
| MIROC-ES2L | | 384.3 | -13.3 | 133.0 | 130.9 | -2.1 | 49.7 | 55.0 | 5.2 | 209.1 | 159.4 | 433.2 (423.2) | | -14.0 | 10.0 | 0.72 | 0.62 |
| MPI-ESM1-2-LR | | 404.3 | 6.8 | 136.3 | 144.8 | 8.5 | -14.3 | 2.3 | 16.5 | 186.5 | 200.8 | 364.8 (362.5) | | 46.7 | 2.3 | 0.89 | 1.01 |
| MRI-ESM2-0 | | 383.5 | -14.1 | 163.3 | 142.5 | -20.8 | 63.0 | 59.6 | -3.3 | – | – | 470.8 (466.7) | | -57.5 | 4.1 | 0.84 | 0.62 |
| NorESM2-LM | | 405.6 | 8.1 | 135.5 | 148.7 | 13.2 | -18.9 | 0.3 | 19.1 | – | – | 357.4 (357.2) | | 52.0 | 0.2 | 0.73 | 0.46 |
| UKESM1-0-LL | | 406.4 | 8.9 | 136.7 | 146.1 | 9.4 | -5.8 | -0.7 | 5.1 | 119.4 | 125.2 | 375.2 (371.4) | | 37.8 | 3.8 | 0.85 | 0.77 |
| Average | 397.6 | 404.9 | 7.4 | 136.7 | 145.4 | 7.0 | 6.0 | 15.2 | 13.8 | 148.2 | 129.3 | 374.0 (383.2) | 409.2 | 26.0 | -6.3 | 0.97 | 0.95 |
| S.D. | – | 14.4 | 14.4 | 10.7 | 17.1 | 20.7 | 74.2 | 65.5 | 17.3 | 31.2 | 66.6 | 67.1 (76.2) | – | 76.2 | 39.9 | 0.28 | 0.37 |
| GCP2021 | 397.6 (+110.6±2.4 ppmv during 1850-2014) | | | 150±30 | | | | | | 180±40 | 195±60 | 400±20 | | | 30 | – | – |

(*) Compatible fossil fuel emission $E_{FF}^{C\text{-}HIST}$ is modified by the *IB* term, i.e., $E_{FF}^{C\text{-}HIST} = (397.6 - 284.3) \times 2.124 + C_O^{C\text{-}HIST} + C_L^{C\text{-}HIST} + IB$, and the $E_{FF}^{C\text{-}HIST}$ before the *IB* correction is shown in parentheses.
(†) The imbalance term *IB* is evaluated using the E-HIST result as $IB = 409.2 - \Delta CO_2^{E\text{-}HIST} \times 2.124 - C_O^{E\text{-}HIST} - C_L^{E\text{-}HIST}$
(‡) Corrected by the GSAT drift that is found in the corresponding preindustrial control simulations (Fig. S1)

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024

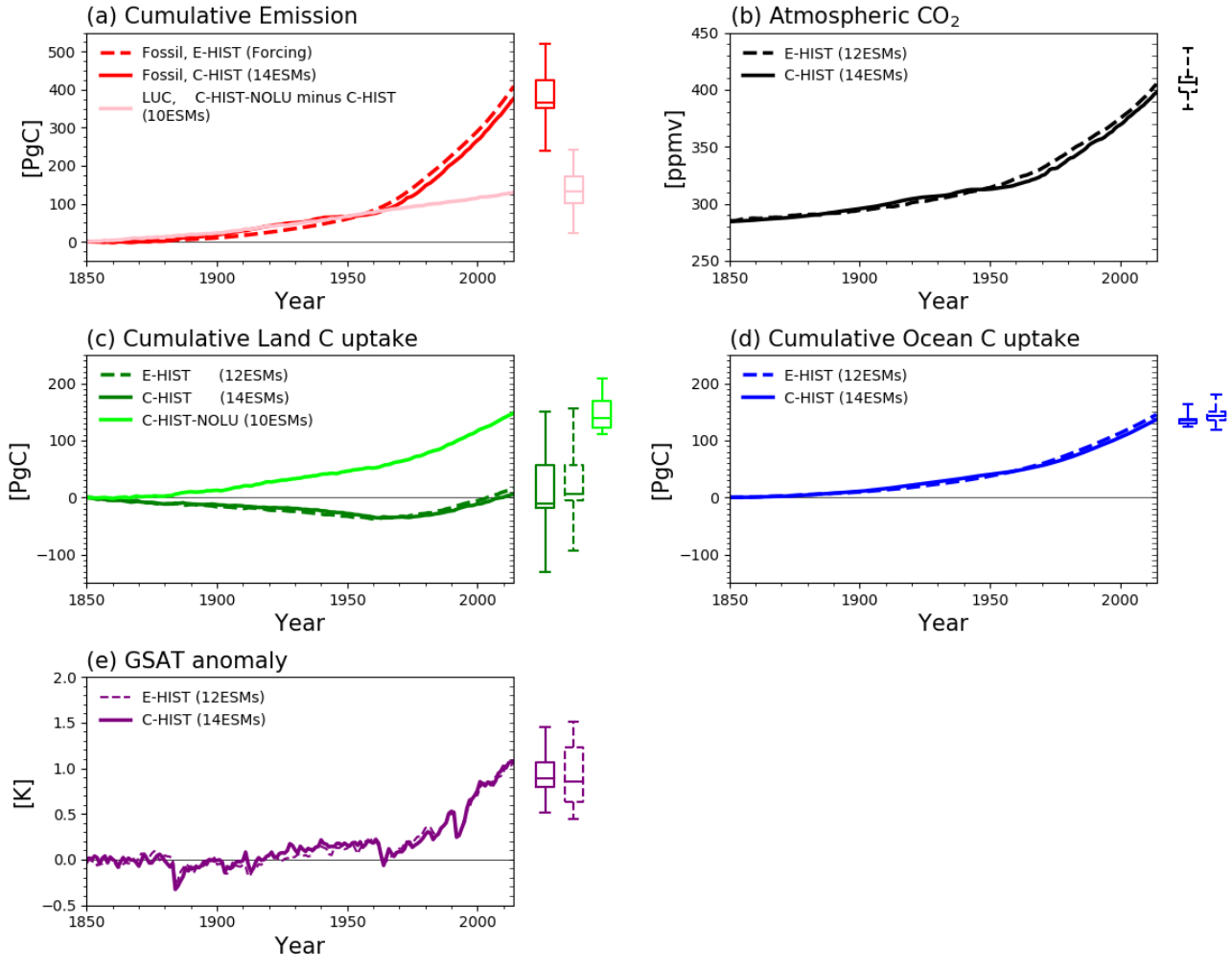




Figure 3 Consistency of global carbon budget and anomaly of global mean surface air temperature (GSAT) between C-driven and E-driven CMIP6 historical (1850–2014) simulations (i.e., C-HIST and E-HIST, respectively), shown by the multi-model means of the ESMs. This analysis corresponds to steps (1), (2), (3), and (4) in Fig. 1. Simulated results of C-HIST and E-HIST are represented by solid lines and dashed lines, respectively, in panels of (a) cumulative anthropogenic emission $E_{FF}$, which is before the imbalance correction, (b) atmospheric $CO_2$ concentration *CO2*, (c) cumulative land C uptake with ($C_L$, dark green) and without ($C_{LN}$, light green) consideration of the impact of land use change, (d) cumulative ocean carbon uptake $C_O$, and (e) anomaly of GSAT. In panels (a) and (b), fossil fuel emission for the E-driven run and $CO_2$ concentration for the C-driven run, respectively, are obtained from the prescribed forcing datasets used for CMIP6. In the boxplots, the bottom and top caps represent the minimum and maximum model results (no outlier), respectively, and the bottom and top ends of the boxes represent the 25th and 75th percentiles, respectively; horizontal lines in the boxes represent the median (50th percentile); the numbers for 2014 are used in the boxplots in panels (a)–(d), and the 2005–2014 average is used in panel (e).

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


Although the multi-model averages have been confirmed to show general agreement between C-HIST and E-HIST, the simulation results of each individual model sometimes show large differences between the two types of experiment (Fig. 4). First, although land C uptake in each model shows similar magnitudes between the two experiments ($R^2 = 0.97$; Fig. 4), weak correlation was found for ocean carbon uptake ($R^2 = 0.01$; Fig. 4b), suggesting that ocean carbon uptake of E-HIST cannot be well explained by that of C-HIST in some models. This is likely due to differences in the experimental configurations. In E-HIST, the land carbon flux, the magnitude of which is estimated very differently by the models and forced to change by the prescribed land-cover dataset, changes the simulated atmospheric $CO_2$ concentration. Thus, models with land carbon uptake that is too strong likely simulate lower $CO_2$ concentration, leading to a weaker ocean carbon sink through the $CO_2$–oceanic carbon feedback process. This mechanism makes the ocean carbon uptake dependent on the land C uptake ($R^2 = 0.56$; Fig. 4d), yielding a different magnitude of ocean carbon uptake between E-HIST and C-HIST. Meanwhile, in C-HIST, the land and ocean carbon fluxes do not change the $CO_2$ concentration, and thus the land carbon flux in the models does not affect the ocean carbon sink, resulting in independent behavior of the land and ocean carbon fluxes ($R^2 = 0.01$; Fig. 4c). In addition, the spatial distribution and the seasonality of the prescribed $CO_2$ concentration used for C-HIST are lost or latitudinally fixed, while the concentration field in E-HIST is freely simulated by models. This may also be an additional reason to yield a different magnitude of ocean carbon sink between C-HIST and E-HIST to some extent (Halloran, 2012).

In addition, the spatial distribution and the seasonality of $CO_2$ concentration in C-HIST is fixed or stylized in the prescribed data, while the concentration in E-HIST is simulated by models. Thus, it may yield a different magnitude of ocean carbon uptake between C-HIST and E-HIST to some extent.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024

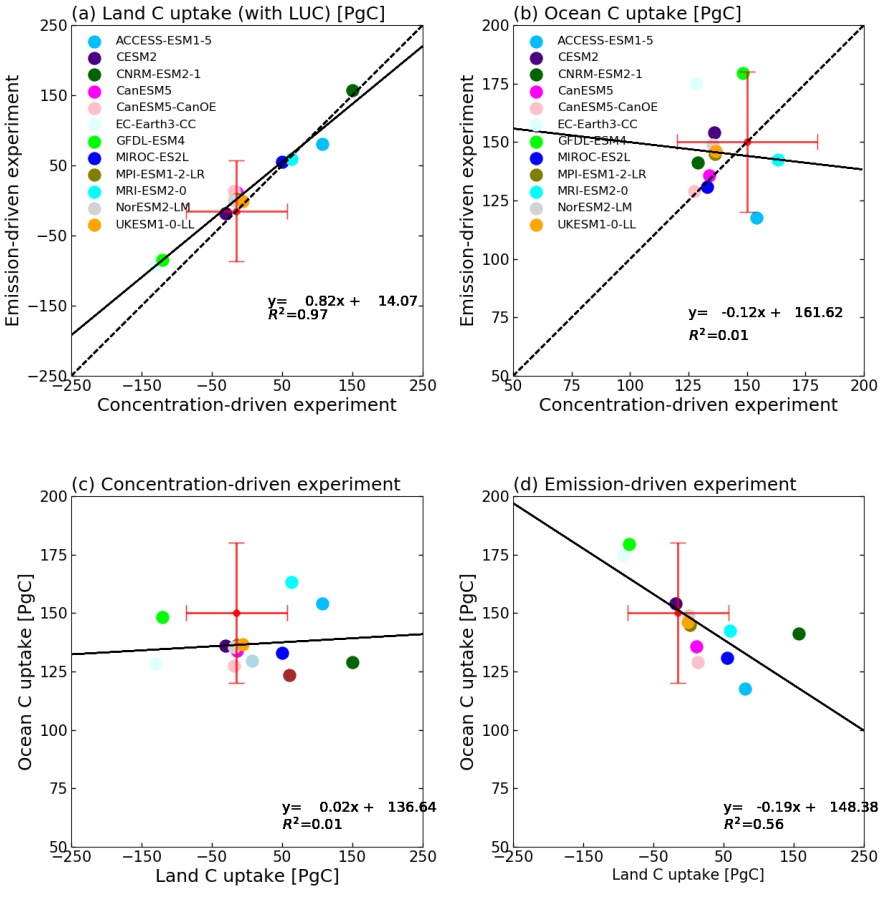




Figure 4 Comparison of cumulative C uptakes by land and ocean between the C-driven and E-driven historical (1850–2014) experiments (i.e., C-HIST and E-HIST, respectively) simulated by CMIP6 ESMs. This analysis corresponds to step (3) in Fig. 1, and the comparison is made at the end of the historical simulations. The two types of historical simulation are compared regarding (a) cumulative land carbon uptake $C_L$ and (b) cumulative ocean carbon uptake $C_O$. The dependency between the carbon uptake of the land and of the ocean in each experiment is shown in (d) for C-HIST and (e) for E-HIST. Solid and dashed lines represent the regression line and the 1:1 line, respectively. Red bars represent the range of uncertainty obtained from GCB2021 (Friedlingstein et al., 2021); it should be noted that GCB2021 does not report land carbon uptake with consideration of the impact of land use change ($C_L$), and thus it is estimated here as $C_L = C_{LN} - E_{LUC}$ and $\sigma_{CL}^2 = \sigma_{CLN}^2 + \sigma_{ELUC}^2$.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


### 3.2 Result of Analysis 2: Linkages of $CO_2$ concentration in E-driven experiment with that in other C-driven experiments

In the analysis in Sect. 3.1, we compared the fundamental terms of the global carbon budget between C-HIST and E-HIST. In the following, based on the analytical procedure shown in Fig. 2, the linkages between E-HIST and several types of C-driven experiment are investigated.

In Fig. 5, the simulated $CO_2$ concentration for 2014 in the E-HIST ($CO2^{E\text{-}HIST}$) simulation is plotted against the cumulative anthropogenic $CO_2$ emissions ($E_{FF}^{C\text{-}HIST}$ and $E_{LUC}^{C\text{-}HIST}$) diagnosed from C-HIST, following step (1) in Fig. 2. As mentioned in Sects. 1 and 2, $E_{FF}^{C\text{-}HIST}$ can be considered an indicator that aggregates carbon cycle feedbacks and the response to environmental change, while $CO2^{E\text{-}HIST}$ is a quantity that summarizes the performance of carbon cycle processes in E-HIST. It should be noted that we used the compatible fossil fuel emission $E_{FF}^{C\text{-}HIST}$ that is corrected by the carbon budget imbalance found in each model ("*IB*" column in Table 4), because $CO2^{E\text{-}HIST}$ and $E_{FF}^{C\text{-}HIST}$ should be compared with the equivalent quality (Appendix B).

In Fig. 5a, the compatible fossil fuel plus the simulated land use change emission ($E_{FF}^{C\text{-}HIST} + E_{LUC}^{C\text{-}HIST}$) is used to explain $CO2^{E\text{-}HIST}$; in Fig. 5b, the compatible fossil fuel emission ($E_{FF}^{C\text{-}HIST}$) alone is used as the explanatory variable. In the analysis, $E_{FF}^{C\text{-}HIST} + E_{LUC}^{C\text{-}HIST}$ does not well explain $CO2^{E\text{-}HIST}$ ($R^2 = 0.05$), whereas using $E_{FF}^{C\text{-}HIST}$ alone shows strong correlation ($R^2 = 0.91$). $E_{FF}^{C\text{-}HIST}$ is originally defined to be determined by the prescribed atmospheric carbon change, ocean sink, and net land carbon uptake including land use change flux (Eq. (3a)); thus, the additional inclusion of $E_{LUC}^{C\text{-}HIST}$ into the explanatory variable likely reduces the correlation coefficient.

$CO2^{E\text{-}HIST}$ and $E_{FF}^{C\text{-}HIST}$ are anticorrelated (Fig. 5b), and the slope of the regression line is −0.20 ppmv/PgC, which is equivalent to a cumulative airborne fraction of −0.47 PgC/PgC. Thus, models with larger compatible fossil fuel emission in C-HIST have lower simulated $CO_2$ concentration in E-HIST. Interestingly, the observed concentration and the fossil fuel emission reported in GCB2021 almost plot on the regression line, providing an important indication that models with an adequate cumulative compatible fossil fuel emission value (400 PgC) will simulate a $CO_2$ concentration that is sufficiently close to the observed value (397.6 ppmv). However, there is no model that can simulate the compatible fossil fuel emission within the range of the GCB2021 values, except for CNRM-ESM2-1 (382.4 PgC). Consequently, most models overestimate or underestimate the concentration by more than 5 ppmv (Fig. 5b and Table 4).

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024

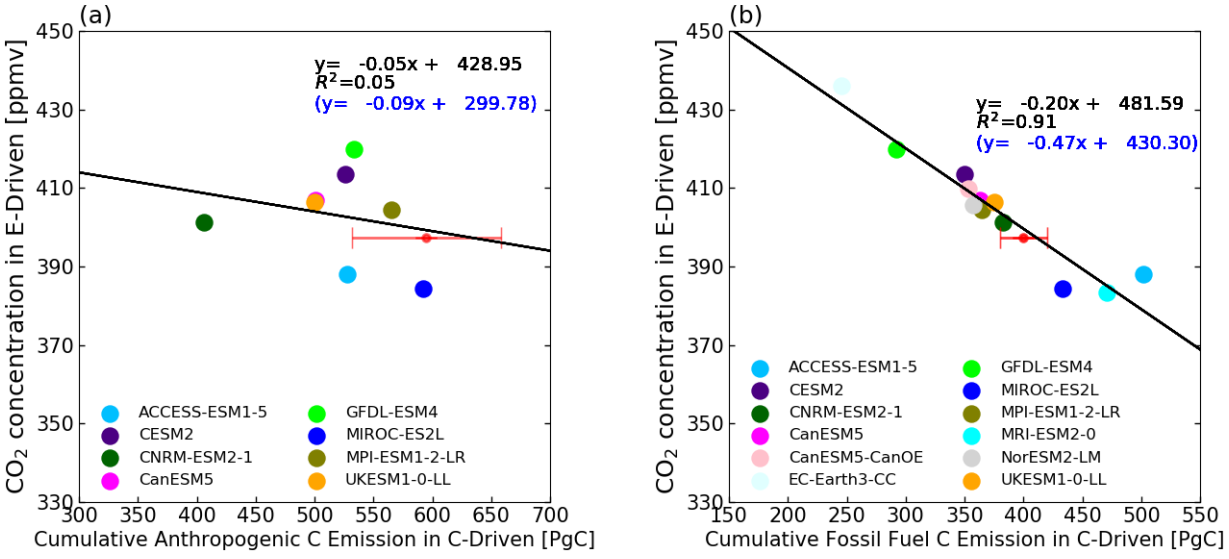




Figure 5 Comparison of $CO_2$ concentration in the E-driven historical run $CO2^{E\text{-}HIST}$ with cumulative values of (a) total anthropogenic (i.e., compatible fossil fuel plus land use change, $E_{FF}{}^{C\text{-}HIST} + E_{LUC}$) emission and (b) compatible fossil fuel emission $E_{FF}{}^{C\text{-}HIST}$ alone. This analysis corresponds to step (1) in Fig. 2, and the comparison is made at the end of the historical simulations for 2014. The solid line represents the regression line (the equation in black represents the regression using atmospheric $CO_2$ concentration, and the equation in blue denotes the regression result that uses atmospheric carbon loading as the dependent variable). The red bars represent the range of uncertainty obtained from GCB2021 (Friedlingstein et al., 2021). The number of models reflects the availability of simulation results necessary for the analysis. Note that the compatible emissions are corrected using the imbalance term found in each model, and the plot before the correction is shown in Fig. S2.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


In the next step, we compare two types of anthropogenic emission, i.e., fossil fuel ($E_{FF}^{C\text{-}HIST}$) emission and land use change ($E_{LUC}^{C\text{-}HIST}$) emission (step (2) in Fig. 2) in C-HIST, and the result is shown in Fig. 6. As reported by Liddicoat et al. (2021), the two variables in C-HIST show an anticorrelation relationship ($R^2 = 0.56$). This is because, by replacing $C_A^{CMIP6F} + C_O^{C\text{-}HIST} + C_{LN}^{C\text{-}HIST}$ *by* $S$ in Eq. (3b), the equation for the global carbon budget can be rewritten as $E_{FF}^{C\text{-}HIST} = S - E_{LUC}^{C\text{-}HIST}$. This equation suggests an underlying mechanism for the creation of the negative correlation between $E_{FF}^{C\text{-}HIST}$ and $E_{LUC}^{C\text{-}HIST}$, i.e., a model with higher land use change emission will yield lower fossil fuel emission to achieve the same $CO_2$ increase over the historical period (we note that $C_O^{C\text{-}HIST}$, $C_{LN}^{C\text{-}HIST}$, and $E_{LUC}^{C\text{-}HIST}$ are confirmed as independent of each other). Because the term $S$ is not constant and differs among the various models, the individual models do not lie directly on the regression line. Comparison with GCB2021 reveals that no model can reproduce values of $E_{FF}^{C\text{-}HIST}$ and $E_{LUC}^{C\text{-}HIST}$ that are simultaneously within the range suggested by GCB2021.

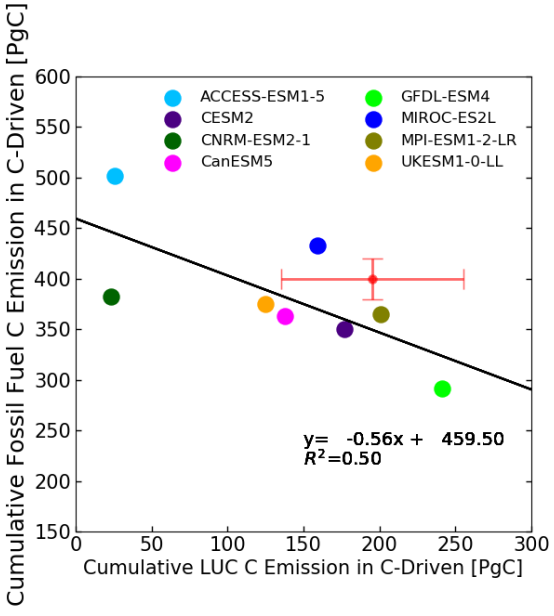



Figure 6 Comparison of cumulative land use change emission $E_{LUC}^{C\text{-}HIST}$ and compatible fossil fuel emission $E_{FF}^{C\text{-}HIST}$ in the C-driven historical run; land use change emission is diagnosed from a combination of C-HIST and C-HIST-NOLU. This analysis corresponds to step (2) in Fig. 2, and the comparison is made at the end of the historical simulations. The solid line represents the regression line, and the red bars represent the range of uncertainty obtained from GCB2021 (Friedlingstein et al., 2021). Note that the compatible emissions are corrected using the imbalance term found in each model, and the result before the correction is shown in Fig. S2.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


This analysis confirms, as expected, that diagnosed compatible fossil fuel emissions in the C-HIST experiment are affected by the magnitude of the land use change emissions simulated by the models. Additionally, the magnitude of the compatible fossil fuel emissions is also controlled by the magnitude of the simulated natural carbon sinks ($C_{LN}^{C\text{-}HIST}$ for land and $C_O^{C\text{-}HIST}$ for ocean, Eq. (3b)), both of which are strongly affected by the magnitude of the $CO_2$–carbon and climate–carbon feedbacks. Here, we compare the simulated natural sinks over land ($C_{LN}^{C\text{-}HIST}$) and ocean ($C_O^{C\text{-}HIST}$) in the C-HIST experiment with those simulated in the idealized C-1PCT experiment, as illustrated in step (3) of Fig. 2. This comparison examines the linkages between the realistic historical experiment and the idealized experiment, where the latter is configured to analyze the carbon cycle feedbacks of the ESMs.

The comparison is made at the end of C-HIST simulation (397.6 ppmv) and the 34th year of C-1PCT (398.8 ppmv). The results illustrated in Fig. 7b show that the natural ocean carbon sink has positive correlation between C-HIST and C-1PCT ($R^2 = 0.81$). This suggests that the magnitude of the natural ocean sink in C-HIST could be approximated from that in C-1PCT, although the magnitudes of the sinks differ owing to the different rate of $CO_2$ increase assumed in the scenarios. Conversely, the historical natural land carbon sink, which is obtained from C-HIST-NOLU, shows weak correlation with the idealized experiment ($R^2 = 0.07$, Fig. 7a), although a systematic trend that models with higher $C_L^{C\text{-}1PCT}$ tend to have larger $C_{LN}^{C\text{-}HIST}$ is confirmed. This suggests that the C-1PCT results, which have been widely used to investigate the carbon cycle feedbacks of the models, cannot be extrapolated to the C-HIST results in terms of the natural land carbon sink.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024

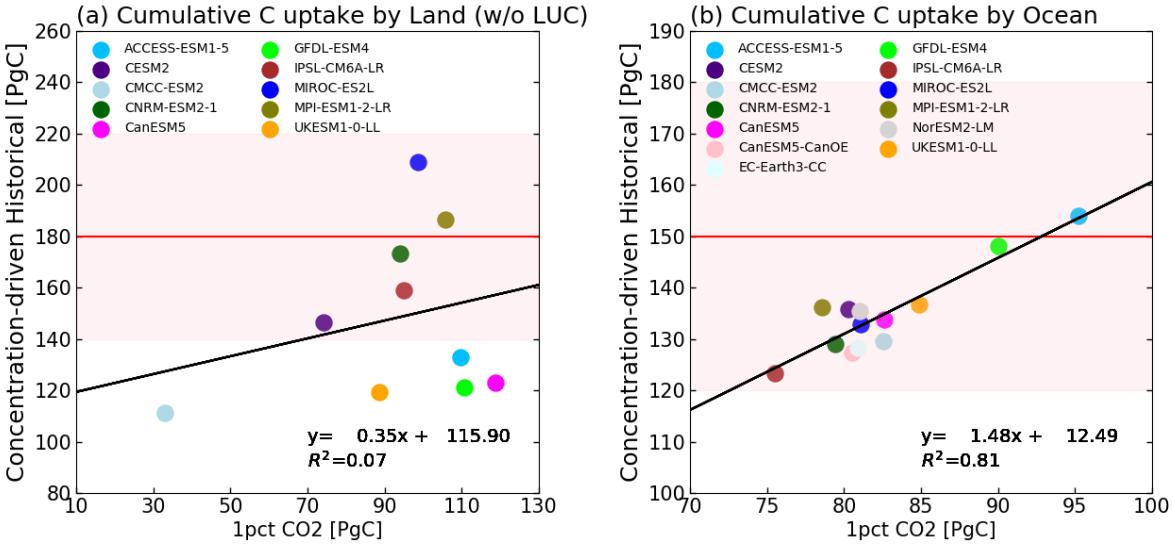




Figure 7 Comparison between the idealized experiment with 1% $CO_2$ increase (C-1PCT) and the C-driven historical experiment (C-HIST): (a) cumulative land C uptake without consideration of land use change ($C_{LN}^{C\text{-}HIST\text{-}NOLU}$ versus $C_L^{C\text{-}1PCT}$) and (b) cumulative ocean carbon uptake ($C_O^{C\text{-}HIST}$ versus $C_O^{C\text{-}1PCT}$). This analysis corresponds to step (3) in Fig. 2. The comparison is made at the end of C-HIST simulation (397.6 ppmv) and the 34th year of C-1PCT (398.8 ppmv). The solid lines represent the regression lines. The horizontal red line identifies the corresponding GCB2021 values, and the red shaded area shows the range of uncertainty.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


## 4. Suggestions for improved simulation of $CO_2$ concentration

We compared the basic terms of the global carbon budgets for C-driven and E-driven historical experiments by following the analysis flow illustrated in Fig. 1, and confirmed that general agreements between the two types of experiment can be found in the multi-model averages but not necessarily in individual model results. Additionally, we examined how E-driven experiments are linked to various types of C-driven experiments by following the analysis flow illustrated in Fig. 2. To obtain further insight into how best to improve the reproducibility of $CO_2$ concentration by ESMs, we performed four types of additional assessment (Sects. 4.1, 4.2, 4.3, and 4.4).

### 4.1 Comparison with GCB budgets in terms of simulated $CO_2$ concentration and compatible fossil fuel emission

We quantitatively confirmed that adequate reproduction of $E_{FF}^{C\text{-}HIST}$ likely leads to accurate $CO2^{E\text{-}HIST}$ (Fig. 5b), both of which have been confirmed good indicators for characterizing the carbon cycle response to external forcing in models. Here, we first discuss how the global carbon budget in each model should be modified in the context of the $E_{FF}^{C\text{-}HIST}$ versus $CO2^{E\text{-}HIST}$ relationship through comparison with the global carbon budgets reported in GCB2021. The result is shown in Fig. 8, which is based on Fig. 5b. It summarizes the extent to which each term of the global carbon budget in each model should be improved to realize a result closer to the best estimate of GCB2021, thereby reproducing an accurate $CO_2$ concentration in E-HIST.

For example, ACCESS-ESM1.5 should increase land use change emission by +169 PgC (brown arrow in Fig. 5, which corresponds to the difference of $E_{LUC}^{C\text{-}HIST}$ between GCB2021 and ACCESS-ESM1.5 in Table 4) because this model was diagnosed to have extremely low land use change emission. This is the primary reason why this model shows the larger compatible fossil fuel emission despite the relatively realistic $CO_2$ concentration in E-HIST, as mentioned in the background section. The improvement of $E_{LUC}^{C\text{-}HIST}$ would lead to displacement of the position of the model in this plotting space toward the upper-left corner because an increase in land use change emission accompanies both reduction in $E_{FF}^{C\text{-}HIST}$ (Fig. 6) and increase in $CO2^{E\text{-}HIST}$. Additionally, the model is evaluated to have a smaller carbon uptake by natural land, which should be strengthened by +47 PgC (green arrow). This improvement is depicted as movement toward the lower-right corner in the plotting space.

It should be noted that the $CO_2$ concentration simulated by the models does not necessarily become closer to the observed concentration, even if the simulated natural sinks and land use change emission become closer to the GCB2021 values. This is because the GCB2021 global carbon budget has a budget imbalance term "*IB*" ($=E_{FF} + E_{LUC} - (C_A + C_O + C_{LN}) = 30$ PgC), and thus the atmospheric $CO_2$ concentration calculated as the residual of other budgets should become 409 ppmv ($C_A = E_{FF}$

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


590 + $E_{LUC}$ − $C_O$ − $C_{LN}$ = 400 + 195 − 180 − 150 = 265 PgC, using the conversion factor of 2.12 PgC/ppmv of Prather et al. (2012)). Therefore, the simulated $CO_2$ concentration is likely to converge to 409 ppmv if the models are driven by GCB-based fossil fuel emission and they perfectly reproduce the natural carbon sink and land use change emission to match the GCB2021 values. Furthermore, because CMIP6 experiments used the prescribed fossil fuel emission (409 PgC) that is larger than the GCB2021 value (400 ± 20 PgC), this overestimation of simulated $CO_2$ concentration would be further enlarged by 595 several parts per million. For the same reason, the compatible fossil fuel emission in C-HIST should converge to 370 PgC under the same condition ($E_{FF}$ = $C_A$ + $C_O$ + $C_{LN}$ − $E_{LUC}$ = 235 + 180 + 150 − 195 = 370 PgC; atmospheric C burden, 235 PgC, is calculated as (397.6 ppmv − 284.3 ppmv)/2.12)). This might explain why more than half of the models are concentrated at the position around ($E_{FF}^{C\text{-}HIST}$, $CO2^{E\text{-}HIST}$) = (370 PgC, 410 ppmv) in Fig. 5b.

600 The existence of the imbalance term is recognized as an important issue in GCB2021, and this problem also affects the interpretation of the results simulated by ESMs. If all the imbalance of the global carbon budget (30 PgC) was attributed to fossil fuel emission, and if the models were run with reduced fossil fuel emission, the simulated $CO_2$ concentration should also be reduced, as depicted by the vertical gray arrows in Fig. 8. In this case, the reference values used for evaluating fossil fuel emission should be revised as well, as depicted in the lower-right panel in Fig. 8. If the imbalance was resolved by 605 strengthening the natural carbon sink or by reduction in the land use change emission, the simulated $CO_2$ concentration should be reduced in conjunction with an increase of 30 PgC in the compatible fossil fuel emission (shown by gray arrow pointing toward the lower-right of the plot). The existence of the budget imbalance term cannot be ignored in attempts to improve the accuracy of ESM simulations. For example, MPI-ESM1.2-LR overestimates the $CO_2$ concentration in E-HIST by 7 ppmv, and much of the overestimation should be attributed to the imbalance problem because the magnitudes of the 610 simulated natural carbon sink and the land use change emission are already sufficiently close to the GCB2021 values.

Even after consideration of the budget imbalance problem, unresolved concentration biases remain in the models, as shown by the deficits indicated between the arrowheads of the gray lines and the best estimate of GCB2021. An example is MIROC-ES2L, for which the arrowheads of the gray lines are lower than the observed $CO_2$ concentration. This suggests that 615 this model likely reproduces lower $CO_2$ concentration, even though this model can reproduce the magnitudes of natural carbon sinks and land use change emissions realistically. This bias should be attributed to other problems, e.g., different behavior in the climate and carbon cycle processes between C-HIST and E-HIST. This likely explains why MIROC-ES2L reproduces a lower $CO_2$ concentration (by more than 10 ppmv) despite the reasonable diagnosis of the compatible fossil fuel emission, as mentioned in the background section.

620

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024

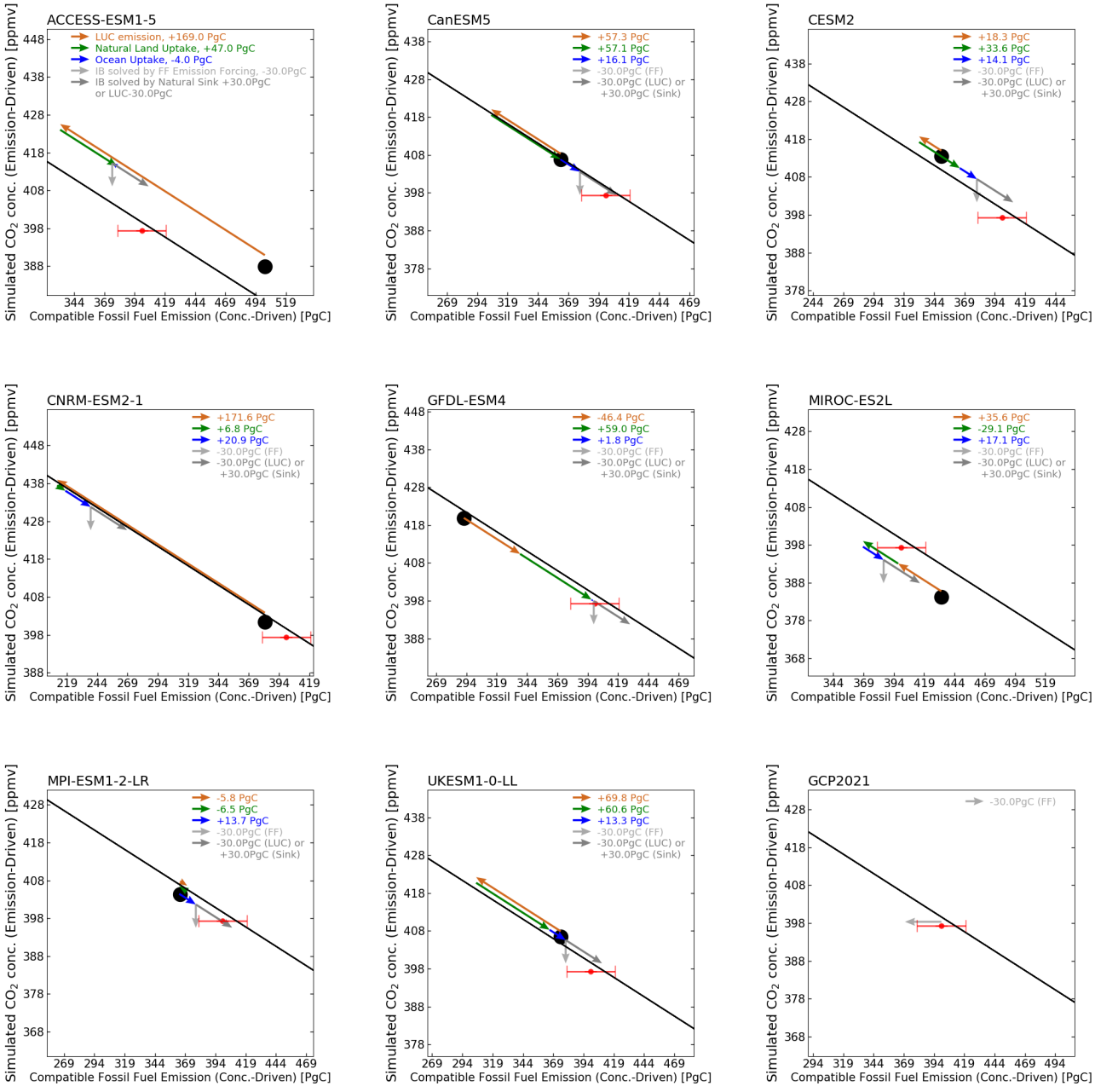




Figure 8 Atmospheric $CO_2$ concentration in E-driven experiments ($CO2^{E\text{-}HIST}$) and compatible fossil fuel emission in C-driven experiments ($C_{FF}^{C\text{-}HIST}$). $C_{FF}^{C\text{-}HIST}$ used here is corrected by the carbon budget imbalance found in each model (Appendix B). Circles represent the targeted model in each panel, and corresponding values of GCB2021 are shown in red. The scatter plots and regression lines are the same as in Fig. 5, and the carbon budget improvements toward the best estimate

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


of GCB2021 are shown by arrows; improvements of land use change emission (brown), natural land carbon sink (green), ocean carbon sink (blue). Gray arrows show the possible change when the carbon budget imbalance of GCB2021 ($IB = E_{FF} + E_{LUC} - (C_L + C_O + C_A)$ = +30 PgC) is resolved by reduction in fossil fuel emission (light gray vertical arrow) or by other means (dark gray). The solid black line represents the regression line depicted in Fig. 5, and the inclination of the arrows (except for the light gray arrow) is assumed to follow the slope of the regression line (−0.21 ppmv/PgC). Among the panels, the x- and y-axis limits are not fixed and are different for visualization reason, but the relative sizes of the x- and y-axis and the intervals remain the same; thus, the length of the arrows can be compared directly among the panels. This analysis relies on the availability of "historical," "esm-hist," and "hist-noLu" simulation results.

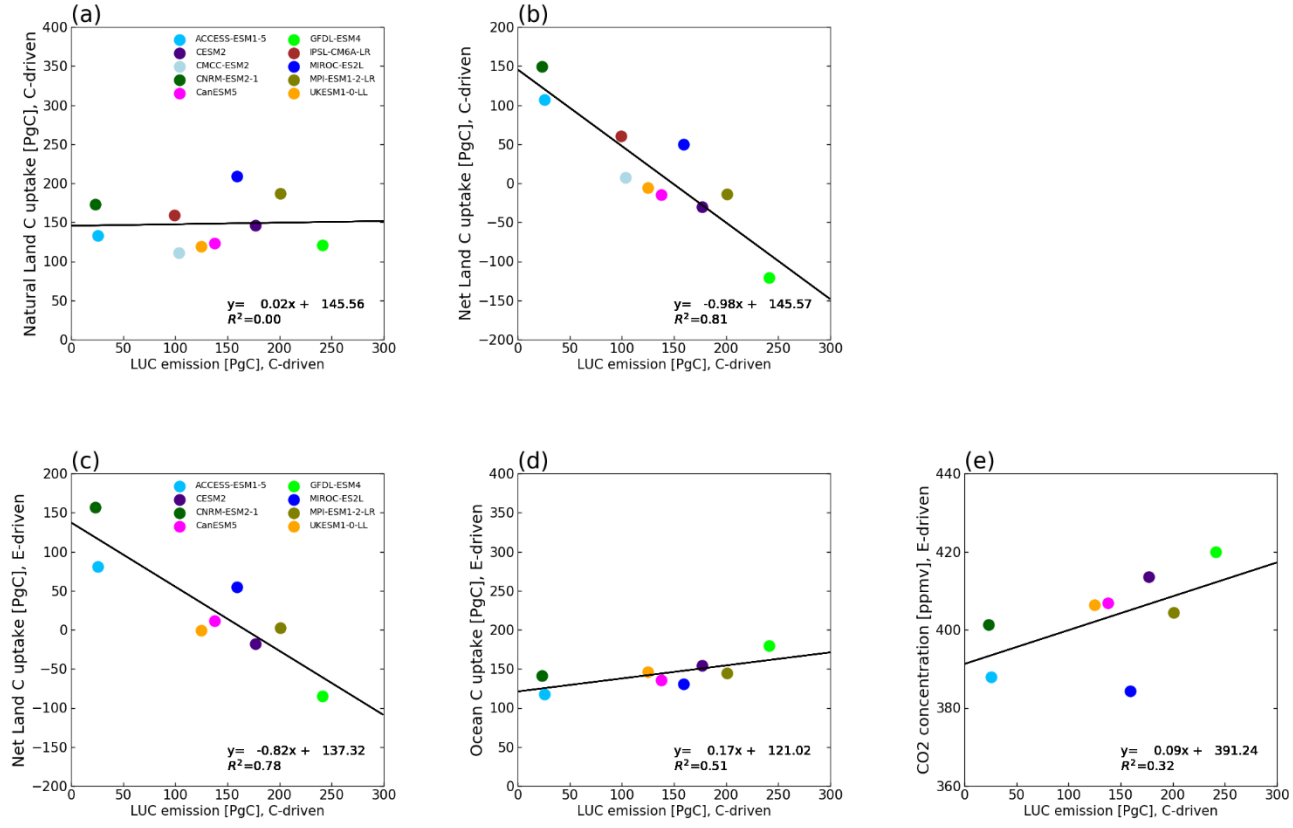
https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


### 4.2 Linkage of land use change emission with other terms of global carbon budgets

The land use change processes and the subsequent carbon emissions are modeled to be forced by changes in land cover that are read externally. In the analysis, the land use change emissions have the largest spread among the models ($E_{LUC}^{C\text{-}HIST}$ = 129.3 ± 66.6 PgC; Table 4), and the standard deviation is more than twice than that of the natural land sink ($C_{LN}^{C\text{-}HIST}$ = 148.2 ± 31.2 PgC). The magnitude of land use change emission is independent of the natural carbon sink on land (Fig. 9a), and the net land carbon uptake ($C_L^{C\text{-}HIST} = C_{LN}^{C\text{-}HIST} - E_{LUC}^{C\text{-}HIST}$) is strongly controlled by the level of $E_{LUC}^{C\text{-}HIST}$ simulated by the models ($R^2 = 0.81$; Fig. 9b). The same is true of the net land carbon uptake in the emission-driven historical experiment ($C_L^{E\text{-}HIST}$, $R^2 = 0.78$; Fig. 9c). Furthermore, in the emission-driven experiment, the level of ocean carbon uptake ($Co^{E\text{-}HIST}$) is also determined by $E_{LUC}^{C\text{-}HIST}$ (Fig. 9d) because a model with a higher (lower) land use change emission would lead to a higher (lower) $CO_2$ concentration, promoting (reducing) ocean carbon uptake in the emission-driven simulation. Consequently, $Co^{E\text{-}HIST}$ in the current generation of ESMs is correlated more with $E_{LUC}^{C\text{-}HIST}$ ($R^2 = 0.51$; Fig. 9d)) than with $Co^{E\text{-}HIST}$ ($R^2 = 0.01$; Fig. 4b), suggesting that greater attention should be paid to the magnitude of simulated land use change emission when examining the absolute magnitude of ocean carbon uptake in E-HIST. Finally, the magnitude of $E_{LUC}^{C\text{-}HIST}$ also determines the simulated atmospheric $CO_2$ concentration ($R^2 = 0.32$; Fig. 9e).

This study confirmed that emissions from land use change showed the biggest uncertainty among the terms of the global carbon budget. Several reasons for this can be considered. First, the largest uncertainty might partly arise from the relatively small number of available models that are necessary for the diagnosis (10 models in this study). Second, vegetation carbon is different among the models, and this might explain the different magnitudes of $E_{LUC}^{C\text{-}HIST}$, which originates from stored carbon on land. The investigation, however, confirmed no correlation between the amount of initial vegetation carbon and the magnitude of $E_{LUC}^{C\text{-}HIST}$ (Fig. S4). Therefore, difference in the magnitude of $E_{LUC}^{C\text{-}HIST}$ among the models probably arises from differences in the definition, structure, and parameters of land use processes. Indeed, there are multiple definitions of emissions of "land use, land use change, and forestry" (Grassi et al., 2021), which are sometimes inconsistent between the estimation methods, and they are linked not only to $CO_2$ but across all GHGs (Lamb et al., 2021). This analysis result stresses the urgent necessity for more realistic simulation of those land use change processes for better reproduction of $CO_2$ concentration by ESMs.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


665 Figure 9 Relationships of cumulative land use change emission diagnosed from concentration-driven historical experiments ($E_{LUC}^{C\text{-}HIST}$) versus (a) cumulative land carbon uptake without consideration of land use change ($C_{LN}^{C\text{-}HIST\text{-}NOLU}$) and (b) cumulative net land carbon uptake ($C_L^{C\text{-}HIST} = C_{LN}^{C\text{-}HIST\text{-}NOLU} - E_{LUC}^{C\text{-}HIST}$) in concentration-driven historical experiments. The lower panels represent the scatter plots, presenting $E_{LUC}^{C\text{-}HIST}$ versus (c) cumulative net land carbon uptake ($C_L^{E\text{-}HIST}$), (d) cumulative ocean carbon uptake ($C_O^{E\text{-}HIST}$), and (e) simulated $CO_2$ concentration ($CO_2^{E\text{-}HIST}$) in the emission-driven historical
670 experiments. The number of analyzed models differ between the concentration-driven (upper panels) and emission-driven (lower panels) experiments because of data availability.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


### 4.3 Growth of atmospheric $CO_2$ in five qualitatively divided eras

The analysis in Sect. 4.1 provided perspectives on how to improve the diagnosed compatible fossil fuel emissions and the simulated $CO_2$ concentration in individual models, considering the budget imbalance term found in GCB2021. This analysis, however, used cumulative values of the carbon fluxes at the end of the historical simulation, and thus it is difficult to specify the period in which the biases of simulated $CO_2$ concentration arise. To determine this period, the historical changes in the simulated $CO_2$ concentrations of the models were analyzed.

First, the simulated $CO_2$ concentrations of E-HIST and the prescribed concentration used in C-HIST are compared in Fig. 10; the absolute concentration is shown in Fig. 10a, and the annual $CO_2$ growth rate is shown in Fig. 10b. As shown in Fig. 3, the multi-model mean of the $CO_2$ concentration well agreed with the concentration pathways of the reference data, except for the period of 1940–1960, during which the atmospheric growth rate in the reference data remained almost zero.

Second, we divided the historical period of 1850–2014 into five eras based on qualitative characteristics, and then the cumulative $CO_2$ growth in each era was evaluated (Fig. 10c). The definitions of the five eras are follows:

- Era 1 (1850–1899): A period corresponding to preindustrial conditions; land use change emission is the dominant source of anthropogenic $CO_2$ emission
- Era 2 (1900–1939): A period after the start of the industrial revolution, in which fossil fuel emission becomes comparable with land use change emission
- Era 3 (1940–1959): A period when observed $CO_2$ concentration determined from ice core measurement suggests a concentration plateau
- Era 4 (1960–1999): A period when direct measurement of atmospheric $CO_2$ started; fossil fuel becomes the dominant source of anthropogenic emission, and agriculture (cropland expansion and application of nitrogen fertilizer) rapidly grows; nitrogen deposition increases accompanied by worsening air quality
- Era 5 (2000–2014): A recent period before the Paris Agreement; growth of fossil fuel emission continues; more observation datasets, including satellite measurements, become available for evaluating carbon cycle processes in models; nitrogen deposition reduces because of air pollution regulations

In Fig. 10d, the simulated concentration bias at the end of each era was visualized by normalizing the concentration bias as follows:

$$NB\,(i,j) = \frac{\Delta CO2(i,j) - \Delta CO2^{CMIP6F}(i)}{\Delta CO2^{CMIP6F}(i)};\ i = 1,\ \ldots,5;\ j = 1,\ \ldots,12, \quad (5)$$

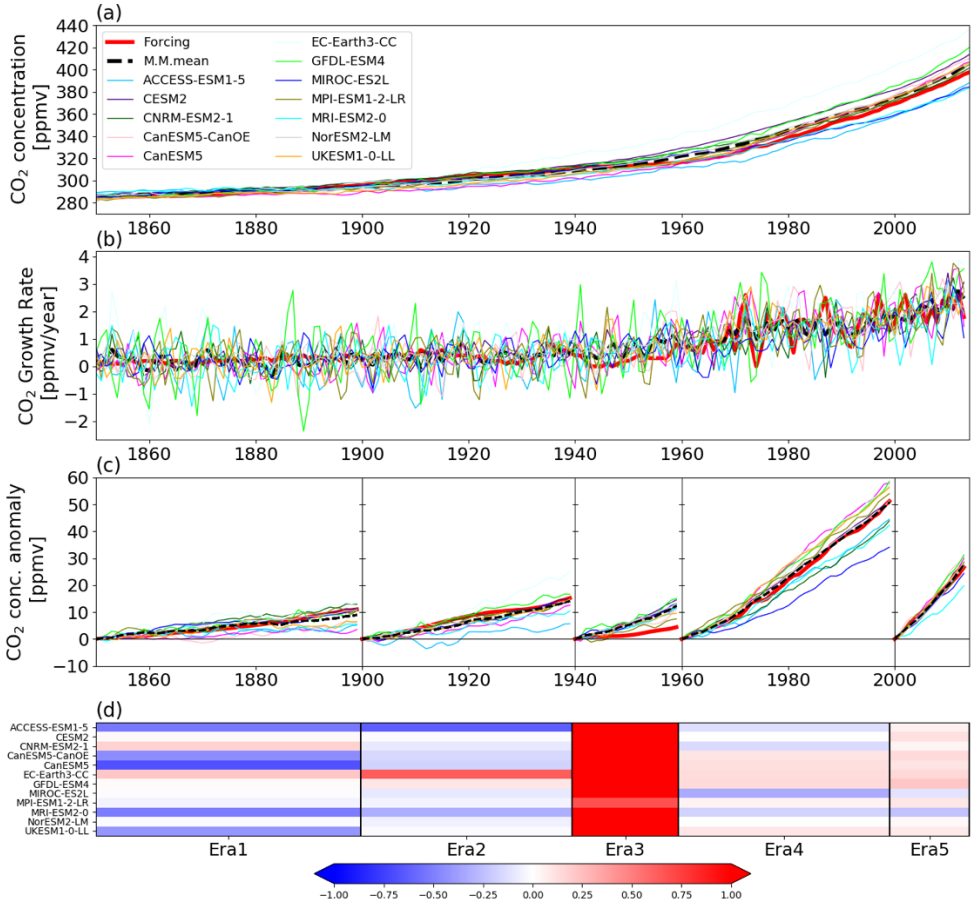
https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


where $NB(i, j)$ represents the normalized bias at the end of each era ($i$ = 1 to 5) in each model ($j$ =1 to 12), and $\Delta CO2(i)^{CMIP6F}$ and $\Delta CO2(i, j)$ are the change in $CO_2$ concentration in the reference data and the change in the simulated concentration, respectively.

In Era 1, half of the models underestimate the $CO_2$ concentration by a few parts per million by volume at the end of this period, while other models well reproduce the $CO_2$ concentration (except EC-Earth-CC, which overestimates the concentration). In this era, land use change was the dominant source of anthropogenic emission; therefore, the simulated $CO_2$ concentration bias is most likely caused by biases in land use change processes in the models. Era 2 represents the period after the start of the industrial revolution, but land use change emissions continued to increase. Most models with positive (negative) concentration bias in Era 2 have positive (negative) concentration bias in Era 1 as well. At the end of Era 3, all models overestimate the $CO_2$ concentration by approximately 5 ppmv on average. The possible reasons for such overestimation are numerous: (1) $CO_2$ emission used for E-HIST might be larger than that expected from the observed $CO_2$ concentration. The Law Dome ice core record (Etheridge et al., 1996) used for historical $CO_2$ concentration shows almost no increase during 1940s (<1.5 ppm), which is inconsistent with 14 PgC emissions during this decade; (2) ocean and/or land might have strengthened carbon sinks during this period attributable to internal climate variability (Joos et al., 1999), which cannot be captured by the freely evolving climate simulations; and (3) the land use change emissions simulated by the ESMs for this period are perhaps larger than observed, because the land-abandonment during World War II might have led to reduced $CO_2$ emissions from land use change (Bastos et al., 2016). In Eras 4 and 5, various processes such as land use change, agriculture, and nitrogen deposition, as well as carbon cycle feedbacks, become increasingly important. However, the normalized concentration bias shown in Fig. 10d is smaller than that in the other periods likely because land use change emission in Eras 4 and 5, which is the most uncertain term of the global carbon budget, becomes weaker in this period. More than six models reproduce the positive $CO_2$ biases in Eras 4 and 5, while the other models largely underestimate the rate of growth in $CO_2$ concentration.

Assessment of $CO_2$ over the individual eras is essentially the same as found in the examination of the annual growth rate of $CO_2$ shown in Fig. 10b. However, this method has three advantages over the comparison of the annual $CO_2$ growth rate: (1) the discrepancies among the models are emphasized and clearly visualized, (2) the sum of the concentration biases in each era is identical to the concentration bias at the end of the historical experiment, which makes it easy to find linkages between the concentration bias over the entire simulation period and each of the divided periods, and (3) the qualitatively characterized period can draw the attention of modelers to the dominant processes specific to each period, suggesting that consideration of such processes should be improved.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


740 Figure 10 Temporal variation of global $CO_2$ concentration simulated by the CMIP6 ESMs and the analysis: (a) annual mean $CO_2$ concentration, (b) annual $CO_2$ growth rate, (c) $CO_2$ concentration anomaly in five eras, and (d) the normalized bias in each era. In (a) and (b), the thick red line represents the $CO_2$ concentration used as prescribed data for CMIP6, the $CO_2$ concentration simulated by each model is represented by the thin colored lines, and the multi-model mean (12 ESMs) is shown by the thick dashed line. In (c), the historical period is divided into five eras, and the $CO_2$ concentration anomaly from
745 the beginning of each era (1850, 1900, 1940, 1960, and 2000) is presented. The heatmap in (d) shows the normalized $CO_2$ concentration bias in each era, with a stronger positive (negative) bias shown by the denser red (blue) color. The calculation of normalized concentration bias follows Eq. (5).

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


### 4.4 Analysis of feedback in the five qualitatively divided eras

In Sect. 4.3, both the $CO_2$ concentration and the bias in E-HIST were assessed for five eras. To obtain further insights into the cause of the concentration biases, we propose an analysis that enables assessment of the forced responses and feedbacks working on the carbon cycle processes in the five eras.

Here, the forced responses and feedbacks were decomposed by combining four types of historical experiment: (A) C-HIST, (B) C-HIST-NOLU, (C) C-HIST-BGC, and (D) C-HIST-CO2 (Table 1).

As summarized in Table 1, the cumulative carbon uptake by land and ocean can be decomposed into four types of drivers:

(1) $CO_2$–carbon feedback (CO2-BGC), calculated by (D) − ((A) − (C))

(2) climate–carbon feedback caused by $CO_2$-induced warming (CO2-CLIM), calculated by (A) − (C)

(3) land use change impact, calculated by (A) − (B)

(4) non-$CO_2$ effects (NONCO2), calculated by (B) − (D). This term should include the warming/cooling effects from non-$CO_2$ greenhouse gases and aerosols on natural carbon sinks as previously examined in Jones et al., 2003; additionally, the direct stimulation of non-$CO_2$ on land biogeochemistry, i.e., nitrogen deposition, is also included in this term.

We note that only two models (CanESM5 and MIROC-ES2L) were available for the analysis (Tables 1 and 2), depending on the data availability of the C-HIST-CO2 result. The two have similar structure in terms of carbon cycle models and the spatial resolution on the land and in the ocean (Arora et al., 2020); however, one of the distinct differences between the two is the terrestrial nitrogen cycle, i.e., CanESM5 does not include the explicit nitrogen cycle, whereas MIROC-ES2L does. The terrestrial nitrogen cycle regulates plant growth and resultant terrestrial carbon uptake through soil inorganic nitrogen availability. A nitrogen deficit could down-regulate the carbon uptake, whereas nitrogen fertilizer and deposition could enhance plant growth and the total carbon accumulation on land. Moreover, these two models are distinct in the plotting space of Fig. 7a, i.e., CanESM5 is the model that shows the largest land carbon accumulation in C-1PCT, and MIROC-ES2L is the model that shows the largest land carbon sink in C-HIST. Thus, detailed analysis of the two models would be informative in exploring the reason why less correlation was found between C-HIST and C-1PCT with land carbon uptake.

The results of the assessment are shown in Fig. 11, where the prescribed fossil fuel emissions, simulated natural sinks, and simulated land use change emission in each of the five eras are depicted explicitly. The simulated natural carbon sink is decomposed and depicted by the four types of drivers (CO2-BGC, CO2-CLIM, LUC, and NONCO2). Additionally, the atmospheric carbon ($CO_2$ concentration) change is also displayed as the residual of the other carbon budgets. In the previous figure, CanESM5 showed distinct underestimation of $CO_2$ in Era 1, which is considered

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


induced by lower land use change emission in this period. Indeed, Fig. 11 shows that this model simulates a cumulative land use change emission of 12 PgC (tan arrow) in the corresponding era, which is smaller than that of MIROC-ES2L (35 PgC). In Fig. 10, MIROC-ES2L shows a distinctly low rate of growth of $CO_2$ in Era 4, and it is revealed that this model accumulates carbon over land by 30 PgC (i.e., atmospheric carbon is removed by 30 PgC) by the NONCO2 effect (purple arrow), while CanESM5 reduces land carbon by 30 PgC (i.e., 30 PgC of carbon is released to the atmosphere). As assumed above, this era (1960–1999) saw worsening air quality in some regions and thus increased nitrogen deposition on the land surface, suggesting that the response of the carbon–nitrogen processes of MIROC-ES2L is likely sensitive to growing nitrogen deposition. Additionally, the land use change emissions in Era 4 might also be small in MIROC-ES2L.

The magnitude of the ocean carbon sink is similar in both models, and the sink is diagnosed to be mainly controlled by $CO_2$–carbon feedback throughout the entire simulation period (blue arrows), as confirmed previously in the C-1PCT experiment (Arora et al., 2020). We note, however, that it is suggested that regional oceans experience strong impact from non-$CO_2$ forcings, particularly the nutrient input via atmospheric deposition and river discharge, and such local non-$CO_2$ forcings can increase global net primary productivity and carbon uptake in the ocean (Yamamoto et al., 2022).

When evaluating the four types of carbon cycle drivers for the entire historical period (Table 5), the most distinct difference between the two models exists in the NONCO2 of land. MIROC-ES2L accumulates terrestrial carbon by 37.2 PgC in response to NONCO2, whereas CanESM5 reduces the carbon by 78.4 PgC, leading to the largest difference of 115.6 PgC among the four types of drivers. As mentioned above, land carbon uptake simulated by MIROC-ES2L is likely sensitive to growing nitrogen deposition (Era 4), and this is likely true for the entire historical period, yielding the positive response of land carbon to NONCO2 in the model. Conversely, CanESM5 does not change the terrestrial carbon uptake in response to nitrogen deposition owing to the lack of the terrestrial nitrogen cycle; thus, the reduction of land carbon uptake should be explained by warming and/or cooling effects induced by non-$CO_2$ GHGs and aerosols.

The NONCO2 effect works in C-HIST but not in the C-1PCT simulation because external forcing other than $CO_2$ concentration is fixed at the preindustrial level in the latter (Table 2). Thus, this effect likely explains the distinct difference in land carbon uptake of the two models in the plotting space of Fig. 7a. CanESM5 has a largely negative effect of NONCO2 (i.e., $CO_2$ release from land to the atmosphere) in the historical simulation, and thus the land carbon uptake in C-HIST became weaker in comparison. However, MIROC-ES2L shows a positive response to NONCO2 (i.e., carbon accumulation on land), and therefore the land carbon uptake in C-HIST became larger than that expected from C-1PCT. Although only two models are available in this study, the analysis strongly suggests that the NONCO2 effect plays an important role in causing the land carbon uptake in C-HIST to be different from that expected from C-1PCT.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


815

Table 5 Changes in natural carbon fluxes integrated over the entire historical period (1850–2014), presented by four drivers (unit: PgC). Positive numbers represent C uptake by land/ocean.

| | Land | | | | Ocean | | | |
|---|---|---|---|---|---|---|---|---|
| | CO2-BGC | CO2-CLIM | NONCO2 | LUC | CO2-BGC | CO2-CLIM | NONCO2 | LUC |
| CanESM5 | 229.9 | -28.6 | -78.4 | -137.7 | 137.9 | -6.6 | 1.2 | 1.4 |
| MIROC-ES2L | 202.5 | -30.6 | 37.2 | -159.4 | 130.9 | -3.8 | 5.6 | 0.3 |
| Difference | 27.4 | 2.0 | -115.6 | 21.7 | 7.0 | -2.8 | -4.4 | 1.1 |

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024

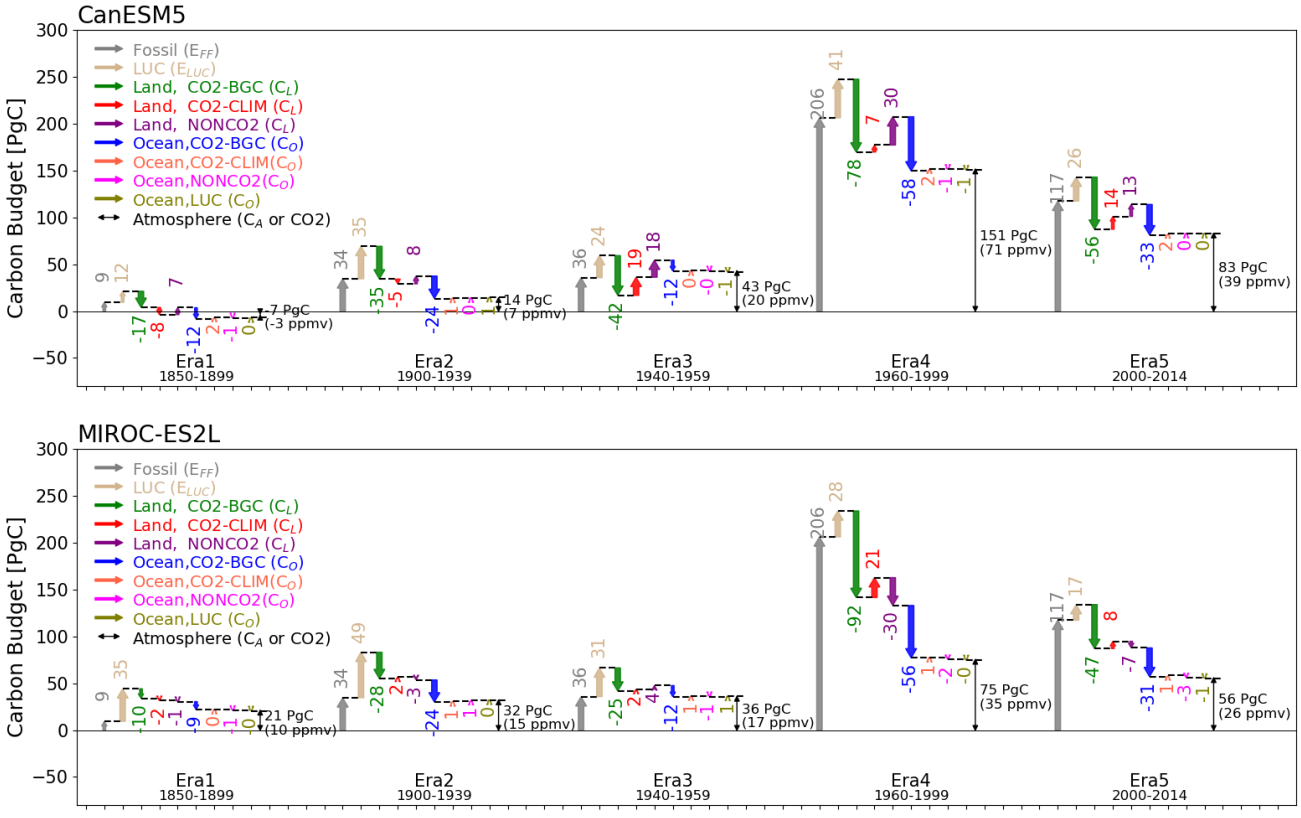




Figure 11 Water flow diagram of global carbon budget simulated by two CMIP6 ESMs (upper: CanESM5, lower: MIROC-ES2L). This analysis depends on the availability of four types of C-driven experiment (Tables 1 and 2). The historical period (1850–2014) was divided into the same five eras as used in Fig. 10, and the flux components in the cumulative values are shown by arrows and numbers (positive means carbon input into the atmosphere). Fossil fuel emission (gray) was obtained from the CMIP6 prescribed data. Changes in natural carbon fluxes were further decomposed into four drivers: (1) $CO_2$–carbon feedback (CO2-BGC, green and blue), (2) climate–carbon feedback induced by $CO_2$ (CO2-CLIM, red and magenta), (3) land use change effect (LUC, tan and olive), and (4) non-$CO_2$ effects (NONCO2, purple and pink), as in Table 1. Black arrows show the increase in atmospheric carbon (PgC) with the concentration change in parentheses (ppmv), calculated as the residual of the other carbon budgets in the C-driven experiments.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


## 5. Summary and Conclusions

In this study, with the objective of acquiring insights into how to improve the accuracy of simulated atmospheric $CO_2$ concentration by ESMs, we first examined the agreement between the C-driven and E-driven historical experiments by CMIP6 ESMs, focusing on the fundamental terms of global carbon budgets (fossil fuel and land use change emissions, land and ocean carbon uptakes, and atmospheric $CO_2$ concentration). The multi-model means of the two types of experiment generally showed good agreement with each other, but some discrepancies were found. The cumulative compatible fossil fuel emission diagnosed from C-driven experiments was lower by approximately 35 PgC than the prescribed emission used to drive the E-driven historical experiment. The simulated $CO_2$ concentration at the end of the historical simulation is higher than that of the observed value by 7 ppmv. Although the reason for these discrepancies between the C-driven and E-driven historical experiments is unclear, the overestimation of simulated $CO_2$ concentration is sufficiently small to make the multi-model averages of GSAT between the E-driven and C-driven simulations negligible. The spread of GSAT among the models, however, becomes larger in the E-driven experiment (±0.37 °C) than that of the C-driven experiment (±0.28 °C) because of the large variation in the simulated $CO_2$ concentrations (±14 ppmv), suggesting that the E-driven setting would bring additional uncertainty into the simulated GSAT owing to the bias in the simulated $CO_2$ concentration. This small additional spread in historical simulations is compensated by the advantage of being able to much more fully span the uncertainty in future projections.

The magnitude of simulated net land carbon uptake is determined by the strength of terrestrial carbon cycle feedbacks and the response to land use change forcing. Although the multi-model spread of the net land carbon uptake is large (6.0 ± 74.2 PgC for the C-driven and 15.2 ± 65.5 PgC for the E-driven experiments), each model showed similar magnitudes between the C-driven and E-driven experiments ($R^2$ = 0.97). In contrast, some individual models showed a distinctly different magnitude of ocean carbon uptake between the C-driven and E-driven experiments ($R^2$ = 0.01). One of the reasons is that the E-driven setting allows ocean carbon fluxes to be dependent on land carbon fluxes via $CO_2$–carbon feedback, and even though a model might have an overly strong (weak) carbon sink in the land component of the C-driven experiments, it can be partly alleviated in the E-driven setting by decreasing (increasing) the ocean carbon flux. This mechanism explains that the magnitude of the ocean carbon sink in the E-driven setting is correlated with land carbon uptake ($R^2$ = 0.56) rather than with the magnitude of the ocean carbon sink that is evaluated in the C-driven historical simulation.

These quantitative assessments between the two types of experiment were confirmed by multiple CMIP6-ESMs, and one of the most important confirmations obtained from this study was the fact that strong negative correlation was found between the simulated $CO_2$ concentration and the compatible fossil fuel emission ($R^2$ = 0.91). This suggests that reasonable reproduction of compatible fossil fuel emission in the C-driven experiments likely assures reasonable performance of the simulated $CO_2$ concentration in the E-driven experiments, although most of the current generation of ESMs analyzed in this

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


study cannot reproduce the compatible fossil fuel emission reasonably i.e., within the range of GCB2021. The multi-model average values of $E_{FF}^{C\text{-}HIST}$ and $CO2^{E\text{-}HIST}$ are 374 PgC and 405 ppmv, respectively, whereas the corresponding GCB2021 values are 400 ± 20 PgC and 397.6 ppmv, respectively. Jones et al. (2023) found that CMIP6 multi-model mean performed well for all regions and variables assessed by the RECCAP2 regional assessments for the land carbon cycle. However, no single model performed well for every region or for every variable, and there were complex signals of the role of process-inclusion in helping improve model fidelity.

We then summarized the information on the improvements required in each individual model for each term of the global carbon budget in an analysis space of the "compatible fossil fuel emission" versus the "simulated $CO_2$ concentration." In this analysis, the simulated results were compared with the estimates of GCB2021. This analysis succeeded in visualizing the budget improvements that are suggested from GCB2021, explicitly considering the budget imbalance term in the current estimation. Such summarization and visualization of the simulated carbon budget by ESMs are important for providing guidance for model improvement, even under the condition that the estimation uncertainties specified in GCB2021 remain.

We proposed a method to evaluate the simulated $CO_2$ concentration and the biases by dividing the entire historical simulation period into five eras that were characterized qualitatively. The division succeeded in clarifying the period when the $CO_2$ concentration biases of the models are produced, and thus could draw the attention of modelers to the most important processes specific to each period. One common feature confirmed among the models was overestimation of $CO_2$ concentration during the 1940–1959 period, when ice core measurements suggest a $CO_2$ concentration plateau. Although it is difficult to specify the reason from this study, the models produce an overestimate of approximately 5 ppmv on average in this era. Subsequently, the global carbon budget was split into components by combination of four types of historical simulation results, although only two models were available for this analysis. The results provided insight into the causes of the $CO_2$ concentration bias: one model underestimated the $CO_2$ concentration in 1850–1899, which was likely induced by the relatively low land use change emission in this period; the other model distinctly underestimated the growth in $CO_2$ concentration in 1960–1999, which was likely induced by terrestrial carbon uptake stimulated by non-$CO_2$ effects (nitrogen deposition), and/or induced by low land use change emission in the corresponding period. The analysis was conducted during the 1850–2014 simulation period, and extension of the historical simulation beyond 2020 and additional analysis with one more "post-Paris agreement" era might highlight implications for global warming mitigation policies.

In the analysis, the non-$CO_2$ effects on land, which include the climate–carbon feedback induced by non-$CO_2$ agents and direct stimulation of land biogeochemistry via deposition, showed the largest discrepancy between the two models, implying that other models also simulate such an impact of non-$CO_2$ agents with nonnegligible magnitude in the historical simulations. Because the non-$CO_2$ effect is working in the historical simulations and absent in an idealized experiment, this process is likely the reason for weakening of the linkage of land carbon uptake between the realistic historical simulation and the

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


idealized experiment. These results have important implications for future global warming projections because non-$CO_2$ impacts on the terrestrial carbon cycle will likely become increasingly apparent as anthropogenic $CO_2$ emission is mitigated. To obtain robust conclusions, it will be necessary for larger numbers of models to participate in all the experiments listed in Table 2. In particular, the DAMIP "hist-CO2" experiment was performed by only two ESMs, thereby precluding all others from the more detailed analysis in Section 4.3.

In this study, much of the discussion focused on comparison with GCB2021. Through such comparison, it is suggested that the natural land sink simulated by ESMs, irrespective of whether E-driven or C-driven experiments, is lower than that of GCB2021, which relies on offline land/ocean biogeochemistry models. Similarly, it was also confirmed that ESMs generally simulate smaller cumulative land use change emission than that reported by GCB2021. Therefore, the mechanism that produces such systematic discrepancies between the GCB models and ESMs should be explored in future work.

Finally, on the basis of the findings of this study, suggestions to improve the $CO_2$ concentration simulated by ESMs are summarized as follows.

1. It is likely that a model with a cumulative compatible fossil fuel emission of approximately 400 PgC during 1850–2014 would be able to adequately capture the $CO_2$ concentration level in the E-driven historical experiment. A model with a larger compatible fossil fuel emission in a C-driven run should have a lower simulated $CO_2$ concentration in an E-driven experiment, and vice versa. However, most CMIP6 ESMs cannot simulate compatible fossil fuel emission within the range of the GCB2021 estimate; consequently, they cannot reproduce an accurate $CO_2$ concentration at the end of the historical simulation.
2. We note that the best estimation of the global carbon budget has a budget imbalance term (30 PgC in GCB2021, which has been updated to 15 PgC in GCB2022). Because of the existence of the budget imbalance, models should have lower compatible fossil fuel emission, even though the other budget terms are simulated to be identical to GCB2021 values. For the same reason, the simulated $CO_2$ concentration will be also overestimated, even when the land use change emission and natural carbon sinks reproduced by the models are identical to those of GCB2021. This is because the imbalance term is imposed on the simulated $CO_2$ concentration in emission-driven historical experiments.
3. To accurately reproduce the atmospheric $CO_2$ concentration, simulating land and ocean carbon uptakes with reasonable magnitudes is necessary. We should recognize that these carbon uptakes in a model can behave differently among various types of simulations. The magnitude of ocean carbon uptakes simulated in the 1%$CO_2$ experiment well explain that in the C-driven historical simulations, likely because of the dominant role of $CO_2$–carbon feedback in the ocean; however, the ocean sink in the C-driven historical simulation can be different from that in the E-driven simulation, in which carbon uptakes by the land and the ocean interact via $CO_2$ concentration. Land carbon uptakes between the C- and E-driven historical simulations behave very similarly; however, it is difficult to approximate the magnitude of land carbon uptakes in the historical simulations from the simulation result of the idealized experiment, because land carbon

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


uptake probably receives nonnegligible impact from non-$CO_2$ effects (climate–carbon feedback induced by non-$CO_2$ agents and/or direct stimulation on land carbon uptake via nitrogen deposition), absent in the idealized experiment.

4. One of the largest estimation spreads among the models was found in the term of land use change emission, and accurate reproduction of land use change emission is critical for better reproduction of $CO_2$ concentration and other global carbon budget terms. The magnitude of simulated land use change emission not only affects the level of net land carbon uptake, but also determines the magnitude of the ocean carbon sink in the E-driven experiment. In the current generation of ESMs, the ocean carbon sink in the E-driven experiment is well explained by the magnitude of the simulated land use emissions, rather than the magnitude of the ocean carbon sink that is evaluated in the C-driven historical experiment. For CMIP7 and beyond, performing "hist-noLu" experiments by more models in both historical and future periods would allow more accurate quantification of the simulated land use change emissions, and clarification of the reasons for such variation in the land use change emission in ESMs and for the large discrepancy between ESMs and GCB models.
5. Ideas for improved model performance in terms of the global carbon budget would be inspired through evaluation of the simulated $CO_2$ concentration because $CO_2$ concentration is observed or estimated with the lowest estimation uncertainty. Analysis that divides the entire historical simulation period into five eras would be helpful to specify when the concentration bias is produced in the models. Further analysis that decomposes the feedbacks and forced response of the carbon cycle in the five eras would verify the reasons behind the production of the concentration bias. The non-$CO_2$ effects were confirmed to be nonnegligible, and mitigation of anthropogenic $CO_2$ emission in the near future would shed light more on the existence of the impact of non-$CO_2$ effects on the carbon cycle. Similar analysis with a larger number of ESMs is necessary to obtain more robust conclusions in the next phase of CMIP activity.

## 6. Appendices

### Appendix A. Derivation of the global carbon budget equations in flux and cumulative values

The global carbon budget equation can be formalized using global carbon flux variables as follows:

$$F_{FF}(t) = \frac{dC_A{}'(t)}{dt} + fgco2(t) + nbp(t),$$

where $F_{FF}(t)$ is the carbon flux of fossil fuel emission, $fgco2(t)$ and $nbp(t)$ are the net atmosphere–ocean and atmosphere–land carbon fluxes, respectively, and $C_A{}'(t)$ is the amount of carbon in the atmosphere.

By integrating both sides of the equation with time *t*, we obtain the following:

$$\int F_{FF}(t)\,dt = \Delta C_A{}'(t) + \int fgco2(t)\,dt + \int nbp(t)\,dt,$$

and this equation is equivalent to the following:

$$E_{FF}(t) = \Delta C_A{}'(t) + \Delta C_O{}'(t) + \Delta C_L{}'(t),$$

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


where $E_{FF}(t) = \int F_{FF}(t)\,dt$, $\Delta C_O'(t) = \int fgco2(t)\,dt$, and $\Delta C_L'(t) = \int nbp(t)\,dt$.

By further replacing $\Delta C_A'(t)$ with $C_A$, $\Delta C_O'(t)$ with $C_O$, and $\Delta C_L'(t)$ with $C_L$, this equation can be expressed as follows:

$E_{FF} = C_A + C_O + C_L$.

(Eq. A1, same as Eq. 1a)

**Appendix B. Compatible fossil fuel emission modified by residual sink/source terms**

In Eq. A1, we assumed in this study that $C_O$ and $C_L$ for 2014 correspond to the cumulative land– and ocean–atmosphere $CO_2$ exchange, respectively, during 1850–2014, i.e., $C_L = \int_{1850}^{2014} nbp\,dt$ and $C_O = \int_{1850}^{2014} fgco2\,dt$. However, this assumption does not always hold strictly in the ESM simulations because of the following reasons:

(1) models might have an additional source/sink of carbon attributable to external natural carbon input/loss ($IB_{natural}$), which is not represented in Eq. A1, e.g., natural carbon input by volcanic eruptions (Hajima et al. 2020) or a "missing C sink" to partly offset the external carbon input from rivers (Seferian et al., 2019);

(2) models sometimes have an additional source/sink of carbon that arises from imperfect mass conservation in advection schemes or other artifacts in models ($IB_{artifact}$).

By explicitly expressing these additional sink/source terms, Eq. A1 can be expressed as follows:

$E_{FF} = C_A + C_O + C_L + IB$,

(Eq. B1)

where $IB = IB_{natural} + IB_{artifact}$.

The $IB$ in each model can be assessed by applying the E-HIST result to Eq. B1, as follows:

$IB^{E\text{-}HIST}$

$= E_{FF}{}^{CMIP6} - C_A{}^{E\text{-}HIST} - C_O{}^{E\text{-}HIST} - C_L{}^{E\text{-}HIST}$

$= E_{FF}{}^{CMIP6} - C_A{}^{E\text{-}HIST} - \int_{1850}^{2014} fgco2^{E-HIST}\,dt - \int_{1850}^{2014} nbp^{E-HIST}\,dt$,

where $E_{FF}{}^{CMIP6}$ is the cumulative value of the emission applied to E-HIST, and the other three terms are obtained from the E-HIST simulation by each model.

We note that the atmospheric $CO_2$ concentrations in E-HIST were simulated under the influence of those additional source/sink terms, while the compatible fossil fuel emission was not. Thus, the compatible fossil fuel emission obtained from C-HIST should also reflect those additional source/sink terms in the calculation, particularly when investigating potential linkages between E-HIST and C-HIST. The compatible fossil fuel emission modified by the residual sink/source terms can be written as follows:

$E_{FF}{}^{C\text{-}HIST} = C_A{}^{CMIP6} + \int_{1850}^{2014} fgco2^{C-HIST}\,dt + \int_{1850}^{2014} nbp^{C-HIST}\,dt + IB^{C\text{-}HIST}$

$= C_A{}^{CMIP6} + \int_{1850}^{2014} fgco2^{C-HIST}\,dt + \int_{1850}^{2014} nbp^{C-HIST}\,dt + IB^{E\text{-}HIST}$,

assuming $IB^{C\text{-}HIST} = IB^{E\text{-}HIST}$.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


## Acknowledgments

We acknowledge the various climate modeling centers/groups for producing and making available their model output via the Earth System Grid Federation (https://esgf-node.llnl.gov/search/cmip6/). TH, MK, KT, and AI were supported by the MEXT-Program for advanced studies of climate change projection (SENTAN, Grant Number JPMXD0722681344) and the Integrated Research Program for Advancing Climate Models (TOUGOU, Grant Number JPMXD0717935715). TH and AI were also supported by the Environment Research and Technology Development Fund (JPMEERF21S20820) of the Environmental Restoration and Conservation Agency provided by the Ministry of Environment of Japan. RS, CDJ, and VB acknowledge supports from the European Union's Horizon 2020 Research and Innovation programme as part of the ESM2025 project (grant number 101003536). CDJ and SKL were supported by the Joint UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101). We thank James Buxton MSc, from Edanz (https://jp.edanz.com/ac), for editing a draft of this manuscript.

## Author contribution

TH conceptualized this study, wrote the draft, and performed the analyses. All co-authors contributed to the analysis and writing of the paper regarding the whole climate-carbon cycle system; MK and RS contributed to the ocean part; AI and ES contributed the land-use change part; PF contributed to the part of comparison with GCP; other co-authors contributed to the land part.

## Code/Data availability

The model outputs analyzed in this study (Table 2) are distributed and made freely available through the Earth System Grid Federation (ESGF). Details on the ESGF can be found on the website of the CMIP Panel (https://www.wcrp-climate.org/wgcm-cmip/wgcm-cmip6, last access: 21 January 2024).

## Competing interests

The authors declare that they have no conflict of interest except that one co-author of this paper is one of the members of Associate editors of this journal.

## References

Andrew, R. M.: A comparison of estimates of global carbon dioxide emissions from fossil carbon sources, Earth Syst. Sci. Data, 12, 1437–1465, 2020.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


Bastos, A., Ciais, P., Barichivich, J., Bopp, L., Brovkin, V., Gasser, T., Peng, S., Pongratz, J., Viovy, N., and Trudinger, C. M.: Re-evaluating the 1940s CO2 plateau, Biogeosciences, 13, 4877–4897, 2016.

Boer, G. J. and Arora, V.: Temperature and concentration feedbacks in the carbon cycle, Geophys. Res. Lett., 36, L02704–L02704, 2009.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, L. E., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A. K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., and Vuichard, N.: Presentation and evaluation of the IPSL-CM6A-LR climate model, J. Adv. Model. Earth Syst., 12, https://doi.org/10.1029/2019ms002010, 2020.

Ciais, P., Bastos, A., Chevallier, F., Lauerwald, R., Poulter, B., Canadell, J. G., Hugelius, G., Jackson, R. B., Jain, A., Jones, M., Kondo, M., Luijkx, I. T., Patra, P. K., Peters, W., Pongratz, J., Petrescu, A. M. R., Piao, S., Qiu, C., Von Randow, C., Regnier, P., Saunois, M., Scholes, R., Shvidenko, A., Tian, H., Yang, H., Wang, X., and Zheng, B.: Definitions and methods to estimate regional land carbon fluxes for the second phase of the REgional Carbon Cycle Assessment and Processes Project (RECCAP-2), Geosci. Model Dev., 15, 1289–1316, 2022.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The community earth system model version 2 (CESM2), J. Adv. Model. Earth Syst., 12, https://doi.org/10.1029/2019ms001916, 2020.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., Bernardello, R., Boussetta, S., Caron, L.-P., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F. J., Docquier, D., Echevarria, P., Fladrich, U., Fuentes-Franco, R., Gröger, M., v. Hardenberg, J., Hieronymus, J., Karami, M. P., Keskinen, J.-P., Koenigk, T., Makkonen, R., Massonnet, F., Ménégoz, M., Miller, P. A., Moreno-Chamarro, E., Nieradzik, L., van Noije, T., Nolan, P., O'Donnell, D., Ollinaho, P., van den Oord, G., Ortega, P., Prims, O. T., Ramos, A., Reerink, T., Rousset, C., Ruprich-Robert, Y., Le Sager, P., Schmith, T., Schrödner, R., Serva, F., Sicardi, V., Sloth Madsen, M., Smith, B., Tian, T., Tourigny, E., Uotila, P., Vancoppenolle, M., Wang, S., Wårlind, D., Willén, U., Wyser, K., Yang, S., Yepes-Arbós, X., and Zhang, Q.: The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6, Geosci. Model Dev., 15, 2973–3020, 2022.

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., Krasting, J. P., Malyshev, S., Naik, V., Paulot, F., Shevliakova, E., Stock, C. A., Zadeh, N., Balaji, V., Blanton, C., Dunne, K. A., Dupuis, C., Durachta, J., Dussin, R., Gauthier, P. P. G., Griffies, S. M., Guo, H., Hallberg, R. W., Harrison, M., He, J., Hurlin, W., McHugh, C., Menzel, R., Milly, P. C. D., Nikonov, S., Paynter, D. J., Ploshay, J., Radhakrishnan, A., Rand, K., Reichl, B. G., Robinson, T., Schwarzkopf, D. M., Sentman, L. T., Underwood, S., Vahlenkamp, H., Winton, M., Wittenberg, A. T., Wyman, B., Zeng, Y., and Zhao, M.: The GFDL earth system model version 4.1 (GFDL-ESM 4.1): Overall coupled model description and simulation characteristics, J. Adv. Model. Earth Syst., 12, https://doi.org/10.1029/2019ms002015, 2020.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


Etheridge, D. M., Steele, L. P., Langenfelds, R. L., Francey, R. J., Barnola, J.-M., and Morgan, V. I.: Natural and anthropogenic changes in atmospheric CO2over the last 1000 years from air in Antarctic ice and firn, J. Geophys. Res., 101, 4115–4128, 1996.

Eyring, V., Bony, S., Meehl, G. A., Senior, C., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organisation, Geoscientific Model Development Discussions, 8, 10539–10583, 2015.

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, Geoscientific Model Development, 9, 1937–1958, 2016.

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 C 4 MIP Model Intercomparison, J. Clim., 19, 3337–3353, 2006.

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., and Knutti, R.: Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks, J. Clim., 27, 511–526, 2014.

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., DBakker, O. C. E., Canadell1, J. G., Ciais1, P., Jackson, R. B., Anthoni1, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D. S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Joetzjer, E., Kaplan, J. O., Kato, E., Goldewijk, K. K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P. C., Melton, J. R., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S. I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., Van Der Werf, G. R., Wiltshire, A. J., and Zaehle, S.: Global carbon budget 2019, Earth System Science Data, 11, 1783–1838, 2019.

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Bakker, D. C. E., Hauck, J., Le Quéré, C., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Bates, N. R., Becker, M., Bellouin, N., Bopp, L., Chau, T. T. T., Chevallier, F., Chini, L. P., Cronin, M., Currie, K. I., Decharme, B., Djeutchouang, L., Dou, X., Evans, W., Feely, R. A., Feng, L., Gasser, T., Gilfillan, D., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, Ö., Harris, I., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Luijkx, I. T., Jain, A. K., Jones, S. D., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lienert, S., Liu, J., Marland, G., McGuire, P. C., Melton, J. R., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Niwa, Y., Ono, T., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan, T. M., Schwinger, J., Schwingshackl, C., Séférian, R., Sutton, A. J., Sweeney, C., Tanhua, T., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F., van der Werf, G., Vuichard, N., Wada, C., Wanninkhof, R., Watson, A., Willis, D., Wiltshire, A. J., Yuan, W., Yue, C., Yue, X., Zaehle, S., and Zeng, J.: Global Carbon Budget 2021, Earth Syst. Sci. Data, 1–191, https://doi.org/10.5194/essd-2021-386, 2021.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Le Quéré, C., Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Alkama, R., Arneth, A., Arora, V. K., Bates, N. R., Becker, M., Bellouin, N., Bittig, H. C., Bopp, L., Chevallier, F., Chini, L. P., Cronin, M., Evans, W., Falk, S., Feely, R. A., Gasser, T., Gehlen, M., Gkritzalis, T., Gloege, L., Grassi, G., Gruber, N., Gürses, Ö., Harris, I., Hefner, M., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Jain, A. K., Jersild, A., Kadono, K., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Landschützer, P., Lefèvre, N., Lindsay, K., Liu, J., Liu, Z., Marland, G., Mayot, N., McGrath, M. J., Metzl, N., Monacci, N. M., Munro, D. R., Nakaoka, S.-I., Niwa, Y.,

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


O'Brien, K., Ono, T., Palmer, P. I., Pan, N., Pierrot, D., Pocock, K., Poulter, B., Resplandy, L., Robertson, E., Rödenbeck, C., Rodriguez, C., Rosan, T. M., Schwinger, J., Séférian, R., Shutler, J. D., Skjelvan, I., Steinhoff, T., Sun, Q., Sutton, A. J., Sweeney, C., Takao, S., Tanhua, T., Tans, P. P., Tian, X., Tian, H., Tilbrook, B., Tsujino, H., Tubiello, F., van der Werf, G., Walker, A. P., Wanninkhof, R., Whitehead, C., Willstrand Wranne, A., et al.: Global Carbon Budget 2022, , https://doi.org/10.5194/essd-2022-328, 2022.

Gier, B. K., Buchwitz, M., Reuter, M., Cox, P. M., Friedlingstein, P., and Eyring, V.: Spatially resolved evaluation of Earth system models with satellite column averaged $CO_2$, , https://doi.org/10.5194/bg-2020-170, 2020.

Gillett, N. P., Shiogama, H., Funke, B., Hegerl, G., Knutti, R., Matthes, K., Santer, B. D., Stone, D., and Tebaldi, C.: The Detection and Attribution Model Intercomparison Project (DAMIP v1.0) contribution to CMIP6, Geoscientific Model Development, 9, 3685–3697, 2016.

Grassi, G., Stehfest, E., Rogelj, J., van Vuuren, D., Cescatti, A., House, J., Nabuurs, G.-J., Rossi, S., Alkama, R., Viñas, R. A., Calvin, K., Ceccherini, G., Federici, S., Fujimori, S., Gusti, M., Hasegawa, T., Havlik, P., Humpenöder, F., Korosuo, A., Perugini, L., Tubiello, F. N., and Popp, A.: Critical adjustment of land mitigation pathways for assessing countries' climate progress, Nat. Clim. Chang., 11, 425–434, 2021.

Gregory, J. M., Jones, C. D., Cadule, P., and Friedlingstein, P.: Quantifying carbon cycle feedbacks, J. Clim., 22, 5232–5250, 2009.

Hajima, T., Kawamiya, M., Watanabe, M., Kato, E., Tachiiri, K., Sugiyama, M., Watanabe, S., Okajima, H., and Ito, A.: Modeling in Earth system science up to and beyond IPCC AR5, Progress in Earth and Planetary Science, 1, https://doi.org/10.1186/s40645-014-0029-y, 2014a.

Hajima, T., Tachiiri, K., Ito, A., and Kawamiya, M.: Uncertainty of concentration-terrestrial carbon feedback in earth system models, J. Clim., 27, https://doi.org/10.1175/JCLI-D-13-00177.1, 2014b.

Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., Ohgaito, R., Ito, A., Yamazaki, D., Okajima, H., Ito, A., Takata, K., Ogochi, K., Watanabe, S., and Kawamiya, M.: Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks, Geosci. Model Dev., 13, 2197–2244, 2020.

Halloran, P. R.: Does atmospheric $CO_2$ seasonality play an important role in governing the air-sea flux of $CO_2$?, Biogeosciences, 9, 2311–2323, 2012.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-maenhout, G., Dawidowski, L., Kholod, N., Kurokawa, J.-I., Li, M., Liu, L., and Lu, Z.: Historical ( 1750 – 2014 ) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System ( CEDS ), Geoscientific Model Development, 11, 369–408, 2018.

Hoffman, F. M., Randerson, J. T., Arora, V. K., Bao, Q., Cadule, P., Ji, D., Jones, C. D., Kawamiya, M., Khatiwala, S., Lindsay, K., Obata, A., Shevliakova, E., Six, K. D., Tjiputra, J. F., Volodin, E. M., and Wu, T.: Causes and implications of persistent atmospheric carbon dioxide biases in Earth System Models, Journal of Geophysical Research: Biogeosciences, 119, 141–162, 2014.

Ito, A., Hajima, T., Lawrence, D. M., Brovkin, V., Delire, C., Guenet, B., Jones, C. D., Malyshev, S., Materia, S., McDermid, S. P., Peano, D., Pongratz, J., Robertson, E., Shevliakova, E., Vuichard, N., Wårlind, D., Wiltshire, A., and Ziehn, T.: Soil carbon sequestration simulated in CMIP6-LUMIP models: implications for climatic mitigation, Environ. Res. Lett., https://doi.org/10.1088/1748-9326/abc912, 2020.

Jones, C., Robertson, E., Arora, V., Friedlingstein, P., Shevliakova, E., Bopp, L., Brovkin, V., Hajima, T., Kato, E., Kawamiya, M., Liddicoat, S., Lindsay, K., Reick, C. H., Roelandt, C., Segschneider, J., and Tjiputra, J.: Twenty-first-

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


century compatible co2 emissions and airborne fraction simulated by cmip5 earth system models under four representative concentration pathways, J. Clim., 26, https://doi.org/10.1175/JCLI-D-12-00554.1, 2013.

Jones, C. D.: Strong carbon cycle feedbacks in a climate model with interactive CO2and sulphate aerosols, Geophys. Res. Lett., 30, https://doi.org/10.1029/2003gl016867, 2003.

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, 2016a.

Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., Graven, H., Hoffman, F., Ilyina, T., John, J. G., and Jung, M.: The C4MIP experimental protocol for CMIP6, 1–52, 2016b.

Jones, C. D., Ziehn, T., Anand, J., Bastos, A., Burke, E., Canadell, J. G., Cardoso, M., Ernst, Y., Jain, A. K., Jeong, S., Keller, E. D., Kondo, M., Lauerwald, R., Lin, T.-S., Murray-Tortarolo, G., Nabuurs, G.-J., O'Sullivan, M., Poulter, B., Qin, X., von Randow, C., Sanches, M., Schepaschenko, D., Shvidenko, A., Smallman, T. L., Tian, H., Villalobos, Y., Wang, X., and Yun, J.: RECCAP2 future component: Consistency and potential for regional assessment to constrain global projections, AGU Advances, 4, https://doi.org/10.1029/2023av001024, 2023.

Joos, F., Meyer, R., Bruno, M., and Leuenberger, M.: The variability in the carbon sinks as reconstructed for the last 1000 years, Geophys. Res. Lett., 26, 1437–1440, 1999.

K. Arora, V., 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, A. 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, 2020.

Kawamiya, M., Hajima, T., Tachiiri, K., Watanabe, S., and Yokohata, T.: Two decades of Earth system modeling with an emphasis on Model for Interdisciplinary Research on Climate ( MIROC ), 5, 2020.

Keller, D. P., Lenton, A., Scott, V., Vaughan, N. E., Bauer, N., Ji, D., and Jones, C. D.: The Carbon Dioxide Removal Model Intercomparison Project ( CDRMIP ): rationale and experimental protocol for CMIP6, 1133–1160, 2018.

Lamb, W. F., Wiedmann, T., Pongratz, J., Andrew, R., Crippa, M., Olivier, J. G. J., Wiedenhofer, D., Mattioli, G., Khourdajie, A. A., House, J., Pachauri, S., Figueroa, M., Saheb, Y., Slade, R., Hubacek, K., Sun, L., Ribeiro, S. K., Khennas, S., de la Rue du Can, S., Chapungu, L., Davis, S. J., Bashmakov, I., Dai, H., Dhakal, S., Tan, X., Geng, Y., Gu, B., and Minx, J.: A review of trends and drivers of greenhouse gas emissions by sector from 1990 to 2018, Environ. Res. Lett., 16, 073005, 2021.

Lawrence, D. M., Hurtt, G. C., Arneth, A., Brovkin, V., Calvin, K. V., Jones, A. D., Jones, C. D., Lawrence, P. J., Noblet-Ducoudré, N. D., Pongratz, J., Seneviratne, S. I., and Shevliakova, E.: The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: Rationale and experimental design, Geoscientific Model Development, 9, 2973–2998, 2016.

Lee, J.-Y., J. Marotzke, G. Bala, L. Cao, S. Corti, J.P. Dunne, F. Engelbrecht, E. Fischer, J.C. Fyfe, C. Jones, A. Maycock, J. Mutemi, O. Ndiaye, S. Panickal, and T. Zhou: Future Global Climate: Scenario-based Projections and Near-term Information, in: Climate Change 2021: The Physical Science Basis, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 553–672, 2021.

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


LeQuéré, C., Andrew, R., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., Pickers, P., Ivar Korsbakken, J., Peters, G., Canadell, J., Arneth, A., Arora, V., Barbero, L., Bastos, A., Bopp, L., Ciais, P., Chini, L., Ciais, P., Doney, S., Gkritzalis, T., Goll, D., Harris, I., Haverd, V., Hoffman, F., Hoppema, M., Houghton, R., Hurtt, G., Ilyina, T., Jain, A., Johannessen, T., Jones, C., Kato, E., Keeling, R., Klein Goldewijk, K., Landschützer, P., Lefèvre, N., Lienert, S., Liu, Z., Lombardozzi, D., Metzl, N., Munro, D., Nabel, J., Nakaoka, S. I., Neill, C., Olsen, A., Ono, T., Patra, P., Peregon, A., Peters, W., Peylin, P., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rocher, M., Rödenbeck, C., Schuster, U., Skjelvan, I., Séférian, R., Skjelvan, I., Steinhoff, T., Sutton, A., Tans, P., Tian, H., Tilbrook, B., Tubiello, F., Van Der Laan-Luijkx, I., Van Der Werf, G., Viovy, N., Walker, A., Wiltshire, A., Wright, R., Zaehle, S., and Zheng, B.: Global Carbon Budget 2018, Earth System Science Data, https://doi.org/10.5194/essd-10-2141-2018, 2018.

Liddicoat, S. K., Wiltshire, A. J., Jones, C. D., Arora, V. K., Brovkin, V., Cadule, P., Hajima, T., Lawrence, D. M., Pongratz, J., Schwinger, J., Séférian, R., Tjiputra, J. F., and Ziehn, T.: Compatible fossil fuel CO2 emissions in the CMIP6 earth system models' historical and shared socioeconomic pathway experiments of the twenty-first century, J. Clim., 34, 2853–2875, 2021.

Lovato, T., Peano, D., and Butenschön, M.: CMIP6 Simulations With the CMCC Earth System Model (CMCC-ESM2), Journal of Advances, 2022.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Crueger, T., Esch, M., Fast, I., Fiedler, S., Fläschner, D., Hagemann, S., Hedemann, C., Hohenegger, C., and Ilyina, T.: Developments in the MPI-M Earth System Model version, Journal of Advances in Modeling Earth Systems, 11, 998–1038, 2019.

Obermeier, W. A., Nabel, J. E. M. S., Loughran, T., Hartung, K., Bastos, A., Havermann, F., Anthoni, P., Arneth, A., Goll, D. S., Lienert, S., Lombardozzi, D., Luyssaert, S., McGuire, P. C., Melton, J. R., Poulter, B., Sitch, S., Sullivan, M. O., Tian, H., Walker, A. P., Wiltshire, A. J., Zaehle, S., and Pongratz, J.: Modelled land use and land cover change emissions – a spatio-temporal comparison of different approaches, Earth Syst. Dyn., 12, 635–670, 2021.

Prather, M. J., Holmes, C. D., and Hsu, J.: Reactive greenhouse gas scenarios: Systematic exploration of uncertainties and the role of atmospheric chemistry, Geophys. Res. Lett., 39, 6–10, 2012.

Sanderson, B. M., Booth, B. B. B., Dunne, J., Eyring, V., Fisher, R. A., Friedlingstein, P., Gidden, M. J., Hajima, T., Jones, C. D., Jones, C., King, A., Koven, C. D., Lawrence, D. M., Lowe, J., Mengis, N., Peters, G. P., Rogelj, J., Smith, C., Snyder, A. C., Simpson, I. R., Swann, A. L. S., Tebaldi, C., Ilyina, T., Schleussner, C.-F., Seferian, R., Samset, B. H., van Vuuren, D., and Zaehle, S.: The need for carbon emissions-driven climate projections in CMIP7, , https://doi.org/10.5194/egusphere-2023-2127, 2023.

Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., Decharme, B., Delire, C., Berthet, S., Chevallier, M., Sénési, S., Franchisteguy, L., Vial, J., Mallet, M., Joetzjer, E., Geoffroy, O., Guérémy, J.-F., Moine, M.-P., Msadek, R., Ribes, A., Rocher, M., Roehrig, R., Salas-y-Mélia, D., Sanchez, E., Terray, L., Valcke, S., Waldman, R., Aumont, O., Bopp, L., Deshayes, J., Éthé, C., and Madec, G.: Evaluation of CNRM earth system model, CNRM-ESM2-1: Role of earth system processes in present-day and future climate, J. Adv. Model. Earth Syst., 11, 4182–4227, 2019.

Seiji, Y., Hideaki, K., Tsuyoshi, K., Naga, O., Kohei, Y., Shogo, U., Hiroyuki, T., Makoto, D., Taichu, T., Masahiro, H., Shokichi, Y., Hiromasa, Y., Eiki, S., Ryo, M., Atsushi, O., Yukimasa, A., and Masayoshi, I.: The Meteorological Research Institute Earth System Model Version 2.0, MRI-ESM2.0: Description and Basic Evaluation of the Physical Component, 気象集誌. 第 2 輯, advpub, 2019–2051, 2019.

Seland, Ø., Bentsen, M., Olivié, D., Toniazzo, T., Gjermundsen, A., Graff, L. S., Debernard, J. B., Gupta, A. K., He, Y.-C., Kirkevåg, A., Schwinger, J., Tjiputra, J., Aas, K. S., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., Karset, I. H. H., Landgren, O., Liakka, J., Moseid, K. O., Nummelin, A., Spensberger, C., Tang, H., Zhang, Z., Heinze, C., Iversen,

https://doi.org/10.5194/egusphere-2024-188
Preprint. Discussion started: 9 February 2024


T., and Schulz, M.: Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations, Geosci. Model Dev., 13, 6165–6200, 2020.

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F. M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., de Mora, L., Kuhlbrodt, T., Rumbold, S. T., Kelley, D. I., Ellis, R., Johnson, C. E., Walton, J., Abraham, N. L., Andrews, M. B., Andrews, T., Archibald, A. T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi, M., Edwards, J., Folberth, G. A., Gedney, N., Griffiths, P. T., Harper, A. B., Hendry, M. A., Hewitt, A. J., Johnson, B., Jones, A., Jones, C. D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R. J., Predoi, V., Robertson, E., Siahaan, A., Smith, R. S., Swaminathan, R., Woodhouse, M. T., Zeng, G., and Zerroukat, M.: UKESM1: Description and Evaluation of the U.K. Earth System Model, Journal of Advances in Modeling Earth Systems, 11, 4513–4558, 2019.

Shevliakova, E., Stouffer, R. J., Malyshev, S., Krasting, J. P., Hurtt, G. C., and Pacala, S. W.: Historical warming reduced due to enhanced land carbon uptake, Proc. Natl. Acad. Sci. U. S. A., 110, 16730–16735, 2013.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., von Salzen, K., Yang, D., and Winter, B.: The Canadian Earth System Model version 5 (CanESM5.0.3), Geosci. Model Dev., 12, 4823–4873, 2019.

Yamamoto, A., Hajima, T., Yamazaki, D., Noguchi Aita, M., Ito, A., and Kawamiya, M.: Competing and accelerating effects of anthropogenic nutrient inputs on climate-driven changes in ocean carbon and oxygen cycles, Sci. Adv., 8, eabl9207, 2022.

Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., Stevens, L., Wang, Y.-P., and Srbinovsky, J.: The Australian Earth System Model: ACCESS-ESM1.5, JSHESS, 70, 193–214, 2020.