Implementation and Evaluation of Emission-Driven Land-Atmosphere Coupled Simulation in E3SMv2.1

Sha Feng[1], Bryce E. Harrop[1], Daniel M. Ricciuto[2], Susannah M. Burrows[1], Qing Zhu[3], Wuyin Lin[4], Nathan Collier[2], Ben Bond-Lamberty[1], Chengzhu (Jill) Zhang[5], Ryan M. Forsyth[5], Jonathan D. Wolfe[3,6], Xiaoying Shi[2], Peter E. Thornton[2], Yohei Takano[6,7], Mathew E. Maltrud[6], Balwinder Singh[1], Yilin Fang[1], Jennifer A. Holm[3], Nicole Jeffery[6], and L. Ruby Leung[1]

[1]Pacific Northwest National Laboratory, Richland, WA, USA, [2]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA, [3]Lawrence Berkeley National Laboratory, Berkeley, CA, USA, [4]Brookhaven National Laboratory, Upton, NY, USA, [5]Lawrence Livermore National Laboratory, Livermore, CA, USA, [6]Los Alamos National Laboratory, Los Alamos, NM, USA, [7]Polar Oceans Team, British Antarctic Survey, Cambridge, UK

Abstract Emissions-driven (prognostic CO_2) simulations are essential for representing two-way carbon-climate feedback in Earth System Models. We present an emissions-driven land–atmosphere coupled biogeochemistry (BGC) configuration (BGCLNDATM_progCO2) in version 2.1 of the Energy Exascale Earth System Model (E3SMv2.1). This is the first E3SM configuration that performs land-atmosphere emission-hindcasts. Here, we document its implementation, evaluate the model's performance against observations and other models, and propose a structured evaluation protocol for such emissions-driven simulations. We conducted transient historical simulations (1850–2014) with BGCLNDATM_progCO2 and compare them to reference simulations—a land-atmosphere coupled simulation without BGC and a standalone land simulation with BGC, both using prescribed CO_2 concentrations—and to observations. BGCLNDATM_progCO2 overestimates atmospheric CO_2 concentrations by 11–23 ppm yet stays within the 40-ppm spread CMIP6 emission-driven models and retains physical climate properties comparable to the reference runs. The CO_2 biases are partly attributed to underrepresented oceanic CO_2 uptake and inadequate representations of some terrestrial processes. In general, introducing prognostic CO_2 did not change physical climate metrics at the global scale but had larger regional effects, particularly over land where spatially heterogeneous CO_2 and prognostic leaf area index influenced surface energy balance. Finally, we propose a general evaluation protocol including spin-up assessment, atmospheric CO_2 benchmarking, physical climate evaluation, and land biogeochemical analysis to support scientific rigor and facilitate inter-model comparisons. The new configuration lays the groundwork for future enhancements, including improved terrestrial biogeochemical processes, integrated marine biogeochemistry, and additional human–Earth system interactions. These developments advance E3SM toward fully coupled emissions-driven simulations, enabling more accurate carbon–climate feedback projections and informing mitigation policy by providing physically consistent carbon-budget metrics for mitigation scenarios.

Plain Language Summary Understanding the impact of carbon dioxide (CO_2) emissions on climate is vital for predicting future changes and crafting effective policies. Earth System Models (ESMs) are essential tools for simulating Earth's climate and assessing various influencing factors. In this study, we extended the Energy Exascale Earth System Model (E3SM)'s capabilities so that CO_2 levels are calculated directly from human and natural emissions instead of being prescribed as a single global value. This extension allows for a more realistic representation of CO_2 exchange between the atmosphere and land. We conducted historical simulations from 1850 to 2014 using this new development and compared results with observations and other models. Our model slightly overestimates atmospheric CO_2 levels compared to measurements but is comparable to other models in capturing key climate features. To help other researchers build and test similar "emission-driven" models, we created a step-by-step evaluation framework that checks CO_2 behavior, climate variables, and land-atmosphere interactions. Our work advances E3SM modeling by accurately representing how CO_2 emissions affect Earth's systems. This enhancement lays the groundwork for modeling interactions between human-Earth interactions, thereby enabling future studies that can inform mitigation and adaption.

Data curation: Sha Feng, Bryce E. Harrop, Daniel M. Ricciuto, Qing Zhu, Nathan Collier, Chengzhu (Jill) Zhang, Ryan M. Forsyth, Jonathan D. Wolfe
Formal analysis: Sha Feng, Nathan Collier, Chengzhu (Jill) Zhang, Ryan M. Forsyth
Funding acquisition: L. Ruby Leung
Investigation: Sha Feng, Bryce E. Harrop, Daniel M. Ricciuto, Susannah M. Burrows, Qing Zhu, Wuyin Lin, Nathan Collier, Jonathan D. Wolfe, Xiaoying Shi, Peter E. Thornton, Yohei Takano, Mathew E. Maltrud, Yilin Fang, Jennifer A. Holm, Nicole Jeffery, L. Ruby Leung
Methodology: Sha Feng, Bryce E. Harrop, Daniel M. Ricciuto, Qing Zhu, Nathan Collier, Xiaoying Shi
Project administration: Susannah M. Burrows, L. Ruby Leung
Resources: Sha Feng, Qing Zhu, Wuyin Lin, Nathan Collier, Chengzhu (Jill) Zhang, Ryan M. Forsyth, Jonathan D. Wolfe, Xiaoying Shi
Software: Sha Feng, Daniel M. Ricciuto, Wuyin Lin, Nathan Collier, Chengzhu (Jill) Zhang, Ryan M. Forsyth, Xiaoying Shi, Balwinder Singh
Validation: Sha Feng, Bryce E. Harrop, Daniel M. Ricciuto, Susannah M. Burrows, Qing Zhu, Nathan Collier, Chengzhu (Jill) Zhang, Ryan M. Forsyth, Peter E. Thornton, Yohei Takano, Mathew E. Maltrud, Balwinder Singh
Visualization: Sha Feng, Bryce E. Harrop, Wuyin Lin, Nathan Collier, Chengzhu (Jill) Zhang, Ryan M. Forsyth, Yohei Takano
Writing – original draft: Sha Feng
Writing – review & editing: Sha Feng, Bryce E. Harrop, Daniel M. Ricciuto, Susannah M. Burrows, Qing Zhu, Wuyin Lin, Nathan Collier, Ben Bond-Lamberty, Chengzhu (Jill) Zhang, Ryan M. Forsyth, Jonathan D. Wolfe, Xiaoying Shi, Peter E. Thornton, Yohei Takano, Mathew E. Maltrud, Yilin Fang, Jennifer A. Holm, Nicole Jeffery, L. Ruby Leung
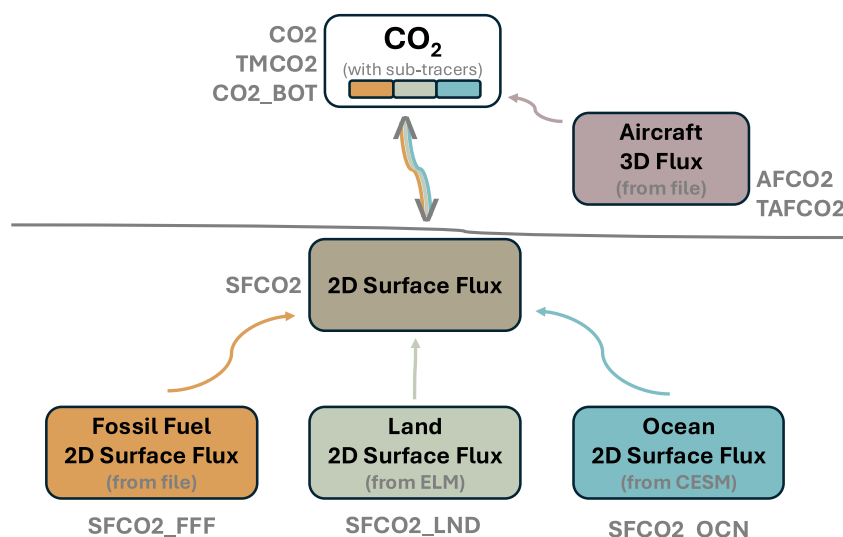
Figure 1. Schematic representation of CO_2 transport pathways among the atmospheric, land, and ocean components in BGCLNDATM_progCO2 (with the two-way arrows denoting the interactive exchanges). The diagram distinguishes between total CO_2 values (encompassing all sources and sinks) and source-specific contributions from the ocean, land, and fossil fuel emissions. "TM" denotes the total mass in the atmospheric column, and "SF" represents surface fluxes. TAFCO2 and AFCO2 indicate the total mass in the column and flux for aircraft, respectively. CO2_BOT refers to the bottom level of atmospheric CO_2 that directly communicates with the land model ELM. The gray line in the middle divides the atmospheric and surface processes. Details of the development and implementation are available in the commit history of the GitHub development branch listed in the Data and Code Availability section.

1. Introduction

Earth system modeling has become an indispensable tool in climate science, serving as a cornerstone for understanding the intricate feedback and interactions within the climate system (Bonan & Doney, 2018; Hurrell et al., 2013). These comprehensive models integrate atmospheric, oceanic, terrestrial, and cryospheric components, enabling researchers to simulate past, present, and future climate conditions with increasing fidelity (M. Collins et al., 2013; Flato, 2011). By incorporating biogeochemical cycles and human activities, Earth system models (ESMs) enhance our ability to assess the impacts of anthropogenic forcings on global climate dynamics (Eyring et al., 2016; Friedlingstein et al., 2014).

The critical role of ESMs is further underscored by their central contribution to coordinated modeling efforts such as the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al., 2016), which supplied foundational data sets for the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (IPCC, 2021). ESMs inform policymakers by projecting potential climate trajectories under varying greenhouse gas (GHG) concentration or emission pathways (O'Neill et al., 2016; Gillett et al., 2016).

However, the primary simulations produced by CMIP6 and prior Model Intercomparison Projects (MIPs) are predominantly concentration-driven: greenhouse-gas mole fractions are prescribed rather than predicted from emissions. Although prescribing greenhouse-gas concentrations greatly simplifies cross-model comparison, it has recognized drawbacks—among them the need to derive "compatible" emissions, ambiguity in closing the carbon budget, and the impossibility of diagnosing interactive carbon-cycle feedback—discussed in detail by Sanderson et al. (2024) and Hajima et al. (2025). A major complication is the discrepancy between the carbon cycle representations in the original Integrated Assessment Models (IAMs), which generate the emission scenarios, and the ESMs that implement climate simulations based on these scenarios. This mismatch leads to "compatible emissions" that are conceptually distinct from the original scenario designs (Koven et al., 2022; Liddicoat et al., 2021 and Figure 2). These differences can result in substantial variations between the IAM-derived emissions and those computed within ESMs.

A second challenge with compatible emissions is the model-dependent ambiguity in how they are calculated. Since compatible emissions are calculated as a residual after accounting for carbon exchange by land ecosystems, oceans, and the accumulation in the atmosphere, it is essential that all models output the necessary fields to close

the carbon budget. However, inconsistencies persist among CMIP6 models regarding output conventions, units, and definitions of component-level carbon fluxes, complicating analyses and hindering efforts to reconcile results across different models (Friedlingstein et al., 2020; Jones et al., 2016). Additionally, ESMs dynamically simulate land use, land-use change, and forestry (LULUCF) emissions based on evolving land-use patterns, which can markedly differ from the original LULUCF fluxes computed in IAMs. Different LULUCF methodologies can cause the intended fossil fuel emissions signal to become mixed with other factors (Quesada et al., 2018; Wilkenskjeld et al., 2014).

Recognizing these issues, CMIP7 has recommended a shift toward emissions-driven or "hybrid" experiments in which CO_2 and other species are calculated from specified emissions (Hajima et al., 2024; Sanderson et al., 2023). The increasing sophistication and stability of emissions-driven ESM configurations relevant for simulating GHG cycles suggest that this approach warrants reassessment and greater use (Koven et al., 2022). Emissions-driven experiments can provide critical, process-based information on the mitigation effectiveness of fossil fuel emission reductions, carbon dioxide removal strategies, and land use policies, and provide a more stringent test for ESMs when replicating the observational record (Hoffman et al., 2014). This shift toward emissions-driven modeling is thus essential for advancing our understanding of coupled carbon–climate dynamics and for providing the process-based information required by mitigation pathways.

To enable emissions-driven capabilities in E3SM, a necessary step is the implementation of land-atmosphere interactive biogeochemistry. Here, we present the first land-atmosphere coupled biogeochemistry simulations with prognostic CO_2 in E3SM version 2.1 (E3SMv2.1), referred to as BGCLNDATM_progCO2, while retaining prescribed marine-surface conditions. This intermediate set-up isolates land–atmosphere carbon feedback from ocean feedback and avoids ocean-circulation drift while the full ocean biogeochemistry option is under development. By explicitly simulating carbon fluxes between terrestrial ecosystems and the atmosphere, the configuration improves the physical consistency of climate–carbon projections and provides a testbed for mitigation studies. Throughout this paper we therefore use the term "emission-driven" to denote this land–atmosphere implementation, which is a stepping-stone toward the hybrid experiments recommended for CMIP7 (Sanderson et al., 2023) that will also include prognostic aerosols and interactive ocean carbon. Earlier generations of ESMs pioneered emission-driven experiments (Cox et al., 2000) and subsequent multi-model assessments have quantified their behavior (P. Friedlingstein et al., 2006, 2014; Gier et al., 2020; Hajima et al., 2024; Sanderson et al., 2024). What is new here is the extension of that methodology to the E3SM framework, the first public description of the technical steps required to enable prognostic CO_2 in E3SM, and a transferable evaluation protocol that can be adopted by other ESMs.

This paper has three objectives. First, we describe the implementation of emissions-driven land-atmosphere coupled biogeochemistry in E3SMv2.1. Second, we evaluate the performance of the land-atmosphere coupled biogeochemistry against observations and other models, including previous versions of E3SM with prescribed CO_2 concentrations. This evaluation highlights improvements achieved through the new implementation and identifying areas needing further refinement. Third, we describe a structured evaluation protocol for emissions-driven simulations to ensure scientific quality, which can be adopted by the broader ESM community, including initiatives like CMIP.

2. Methodology

2.1. Model Description

In this study we introduce BGCLNDATM_progCO2—the emissions-driven land–atmosphere biogeochemistry configuration for E3SM v2.1—which adds the tracer and coupling modifications required for interactive CO_2 while otherwise retaining the default physical-climate set-up described by Golaz et al. (2022).

Compared to E3SMv2 in Golaz et al. (2022), E3SMv2.1 has minor updates to the ocean and sea ice models, but otherwise no significant changes. Given that BGCLNDATM_progCO2 uses prescribed sea surface temperatures (SSTs) and sea ice, these changes are not anticipated to alter the overall performance of the atmospheric model EAMv2 and the land model ELMv2 used in this work.

The active components in BGCLNDATM_progCO2 are EAMv2 and ELMv2, documented by Golaz et al. (2022). Briefly, EAMv2, the atmospheric component, uses the hydrostatic reformulation of the non-hydrostatic dynamical core (M. A. Taylor et al., 2020) and separate grids for dynamics and physics, with average

spacings of 110 and 165 km, respectively. Details on the grid implementation and its impacts are described by Hannah et al. (2021). Subgrid turbulence and cloud macrophysics are parameterized with the Cloud Layers Unified By Binomials (CLUBB; Golaz et al., 2002; Larson, 2022). Deep convection is parameterized following G. J. Zhang and McFarlane (1995), with additional modifications related to the triggering of convection (Golaz et al., 2022). Microphysics is parameterized using Morrison-Gettelman version 2 (Gettelman & Morrison, 2015). A simplified form of atmospheric chemistry predicts stratospheric and tropospheric ozone concentrations (O3v2; Tang et al., 2021). Aerosol treatment employs the four-mode version of the Modal Aerosol Module (MAM4; Wang et al., 2020).

ELMv2 simulates biogeophysical and biogeochemical processes critical for emissions-driven simulations, including interactive energy, water, carbon, nitrogen, and phosphorus processes that control the fluxes of CO_2, energy, and water between the land and atmosphere (Ricciuto et al., 2018; P. E. Thornton et al., 2009; Yang et al., 2023). Key updates in ELMv2 include allowing transient crops on separate land units and supporting transient crops without the need for a prognostic crop model (Sinha, Bond-Lamberty et al., 2023). Additionally, soil erosion and sediment yield are now represented (Tan et al., 2018). Relative to E3SMv1.1-BGC (Burrows et al., 2020), multiple issues with the water budget and soil nutrient concentration calculations have been fixed. The introduction of a global carbon budget and corrections to the carbon isotope flux calculation further refine the model's capacity to simulate carbon dynamics accurately. A fix was implemented to correct the carbon mass non-conservation bug affecting atmospheric CO_2 concentrations reported by Burrows et al. (2020). These bug fixes can be found at E3SM version release notes: https://github.com/E3SM-Project/E3SM/releases.

2.2. The Implementation of Land-Atmosphere Interactive BGC Configuration

2.2.1. Conservation of CO_2 Mass

In emissions-driven simulations, maintaining the conservation of CO_2 mass throughout the model components is crucial. We addressed potential sources of mass imbalance, particularly within the atmospheric model EAMv2, by ensuring consistent treatment of CO_2 tracers and their tendencies. While EAMv2 generally uses "dry" mixing ratios for non-water tracers like CO_2, some parameterizations, such as CLUBB, require "wet" mixing ratios. We implemented appropriate conversions between wet and dry mixing ratios within CLUBB to maintain mass conservation and preserve the integrity of CO_2 distribution in the atmosphere.

2.2.2. Connectivity of CO_2 Transport Between Components

Accurate exchange of CO_2 between the atmosphere, land, and ocean components is essential for the closure of carbon budgets. We established seamless communication of CO_2 fluxes and concentrations among these components (focusing on atmosphere and land in this work), enabling the model to effectively simulate interactions among carbon sources and sinks.

In the atmospheric component, prognostic CO_2 tracers represent total and source-specific concentrations and fluxes, allowing detailed tracking of CO_2 dynamics. The land model ELMv2 receives atmospheric CO_2 concentrations, influencing processes such as photosynthesis and stomatal conductance, and returns land CO2 flux as net ecosystem exchange (NEE) to the atmosphere. This bidirectional exchange captures the feedback between terrestrial ecosystems and atmospheric CO_2 concentrations.

Although ocean biogeochemistry is not active in BGCLNDATM_progCO2, we implemented the framework for prognostic CO_2 transport within the ocean component to ensure compatibility with future fully coupled carbon cycle simulations involving the ocean. This preparation, though not used in BGCLNDATM_progCO2, allows for future integration when marine (ocean + sea ice) biogeochemistry becomes active.

In Figure 1, we illustrate the CO_2 transport pathways between the atmospheric, land, and ocean components, highlighting the exchange processes critical for emissions-driven simulations.

2.2.3. Incorporation of CO_2 Emissions Data

For emissions-driven simulations, we incorporated surface and aircraft CO_2 historical emissions (Hoesly et al., 2018). These emissions provide the necessary inputs for fossil fuel combustion and aviation-related CO_2 sources. We processed the emissions data to ensure conservation of mass once ingested within E3SM. This

included conservative remapping to the horizontal physics grid used by EAMv2 and an adjustment of input data timestamps to conserve mass of emissions after online linear interpolation in time. Aircraft emissions are interpolated online onto EAMv2's hybrid vertical grid following the method of Jöckel (2006).

Given that BGCLNDATM_progCO2 uses a data ocean model that prescribes SST and sea ice, the oceanic CO_2 fluxes must also be prescribed. For this study, we chose to use CESM2 simulated oceanic CO_2 fluxes from the CMIP6 historical period (O'Neill et al., 2016), obtained from the Earth System Grid Federation (ESGF) CMIP6 (McBride et al., 2021) submission archive. We selected CESM2 simulations because CESM2 and E3SMv2 use the same ocean biogeochemistry component, the Marine Biogeochemistry Library (MARBL; Long et al., 2021), allowing us to incorporate consistent oceanic CO_2 fluxes that are compatible with the physical and biogeo-chemical processes represented in E3SMv2.

2.2.4. Output Variables and Diagnostics

The model is configured to output key CO_2 variables (see Figure 1), including concentrations and fluxes for total and source-specific components. This allows for detailed tracking of CO_2 transport and budget assessments, which is critical for evaluating model performance and understanding carbon cycle dynamics. A comprehensive monthly carbon budget table is generated online to monitor the conservation of carbon throughout the simulations.

2.3. Experiment Design and Description of Simulations

To validate the implementation of the emissions-driven land-atmosphere coupled biogeochemistry configuration (BGCLNDATM_progCO2) in E3SMv2.1 and ensure that the coupling does not introduce significant changes to the physical climate, we conducted a transient simulation using BGCLNDATM_progCO2. We compared the results with those from two reference simulations:

1. LNDATM_presCO2: A concentration-driven land-atmosphere coupled physical climate configuration without active BGC in land following the Atmospheric Model Intercomparison Project (AMIP) protocol (K. E. Taylor et al., 2000). This simulation serves as a reference for the atmospheric component (EAM).
2. BGCLND: A land-only simulation driven by the Global Soil Wetness Project Phase 3 (GSWP3) meteoro-logical reanalysis data (Dirmeyer, 2011). This simulation serves as a reference for the land component (ELMv2).

Comparing BGCLNDATM_progCO2 with LNDATM_presCO2 and BGCLND allowed us to assess the con-sistency of EAMv2 and ELMv2 between the emissions-driven and reference simulations, ensuring that the implementation of land-atmosphere interaction and prognostic CO_2 does not significantly alter climate results.

2.3.1. Simulation Configurations

2.3.1.1. BGCLNDATM_progCO2 Simulation

The BGCLNDATM_progCO2 simulation employed a tri-grid setup using different grid spacing for atmosphere, land, and ocean/ice components. In this study, EAMv2.1 used approximately 1° horizontal resolution for its dynamics and approximately 1.5° horizontal resolution for the physics parameterizations, both with 72 vertical layers. ELMv2.1 operated at a 0.5° horizontal resolution. For the marine component, we used prescribed SSTs and sea ice concentrations, consistent with the AMIP protocol. Our configuration differed from the standard AMIP setup by including prognostic CO_2, activating BGC processes in the land component, and prescribing oceanic CO_2 fluxes to enable emission-driven simulations.

2.3.1.2. LNDATM_presCO2 Simulation

LNDATM_presCO2 is essentially an AMIP simulation from E3SMv2-DECK (Golaz et al., 2022) that serves as a reference for assessing the atmospheric component's performance. It shares the same atmospheric and land configurations as BGCLNDATM_progCO2 but differs in two key aspects:

- Inactive land BGC: LNDATM_presCO2 uses satellite phenology to provide seasonal changes in vegetation based on climatology, and its land BGC remains inactive. In contrast, BGCLNDATM_progCO2 activates land BGC processes to simulate interactive carbon and nutrient cycles.

Table 1
Summary of Simulation Configurations

Simulations	Atmosphere	Land	Marine	Prognostic CO_2	Active BGC	Purpose
BGCLNDATM_progCO2	EAMv2.1	ELMv2.1	Prescribed SST and sea ice	Yes	Yes	Emission-driven
LNDATM_presCO2[a]	EAMv2	ELMv2	Prescribed SST and sea ice	No	No[b]	Reference case for EAMv2
BGCLND	Prescribed GSWP3 forcing	ELMv2	Null	No	Yes	Reference case for ELMv2

[a]LNDATM_PresCO2 is equivalent to typical AMIP simulations. [b]Satellite phenology.

- Prescribed CO_2 concentrations: LNDATM_presCO2 utilizes prescribed annual global mean time series of CO_2 concentrations for both the land and atmosphere components, rather than prognostic CO_2 as in BGCLNDATM_progCO2. This means that LNDATM_presCO2 does not simulate the dynamic exchange of CO_2 based on emissions and sinks within the simulation.

LNDATM_presCO2 serves as a reference for assessing the atmospheric component's performance. By comparing BGCLNDATM_progCO2 with LNDATM_presCO2, we can isolate and understand the effects of the emissions-driven approach and active land biogeochemistry on atmospheric processes and climate variables.

2.3.1.3. BGCLND Simulation

BGCLND is a land-only simulation using ELMv2 at $0.5° \times 0.5°$ resolution, consistent with BGCLNDATM_progCO2. This simulation is driven by GSWP3 meteorological forcing data (Dirmeyer, 2011), providing observed atmospheric conditions to the land surface. BGCLND has active biogeochemical processes enabled, simulating interactive carbon and nutrient cycles within the land component. Unlike BGCLNDATM_progCO2, which use prognostic CO_2 concentration, BGCLND employs prescribed annual global mean CO_2 concentrations for its biogeochemical processes, similar to LNDATM_presCO2. This configuration allows us to isolate and evaluate the performance of the land biogeochemistry in ELMv2.1 without the influence of coupling with the atmospheric model, serving as a reference for assessing the terrestrial component's performance. Importantly, this design isolates one feedback pathway at a time while holding the others, as well as the boundary conditions (ocean and sea ice), fixed.

2.3.1.4. Summary of Simulations

The configurations of the three simulations are summarized in Table 1.

Figure 2. Global mean time series from 1959 to 2013, showing the oceanic CO_2 fluxes and fossil fuel emissions prescribed in E3SM (dashed lines) in comparison with the Global Carbon Project (solid lines) reported by Friedlingstein et al. (2021), which includes two components: "FFF" represents fossil fuel CO_2 emissions, while "OCN" denotes the oceanic carbon budget.

2.3.2. Spin-Up Processes

We followed a multi-stage spin-up process similar to that of Burrows et al. (2020), beginning with accelerated decomposition and vegetation mortality to quickly reach equilibrium states for land carbon, and nitrogen pools, followed by simulations with normal rates to achieve dynamic equilibrium.

2.3.2.1. Land-Only Spin-Up

The spin-up began with a land-only simulation where accelerated decomposition and vegetation mortality were applied for 200 simulation years to quickly reach an initial equilibrium state for carbon and nitrogen pools under nitrogen limitation using the repeated 1901–1920 climate forcings from GSWP3 (Dirmeyer et al., 2006). During this phase, phosphorus was considered fully available. Subsequently, soil phosphorus was initialized from global maps, phosphorus limitation was enabled, and the land-only simulation continued with normal decomposition and mortality rates for an additional 600 simulation years, allowing all pools to reach dynamic equilibrium.
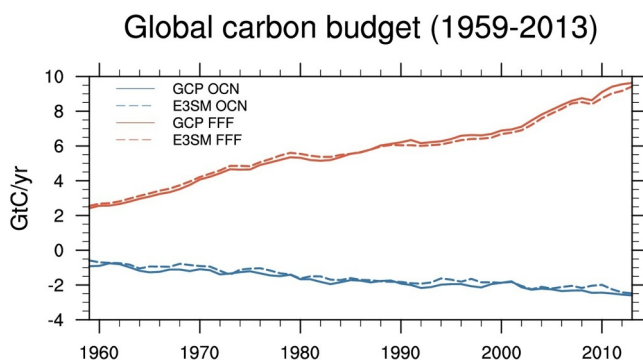
2.3.2.2. Coupled Spin-Up

Following the land-only spin-up, we conducted a coupled BGCLNDATM_progCO2 spin-up simulation for 500 years. This allowed the terrestrial carbon pools to adjust to land-atmosphere feedback under the prognostic CO_2 configuration. Because the observational SST data set extends only from 1870 to 2100, the coupled spin-up is driven by the 1870 SST and sea-ice fields recycled annually; inter-annual ocean variability is therefore absent during spin-up. Atmospheric boundary conditions during the land-only spin-up phases were prescribed using repeated 3-hourly outputs from 25 years of the E3SMv2-DECK preindustrial control (piControl) experiment (Golaz et al., 2022). The extensively evaluated climate of the piControl simulation provided reliable forcing for the spin-up process.

2.3.3. Historical Simulation and Prescribed CO_2 Fluxes

The BGCLNDATM_progCO2 historical simulation covers the period from 1850 to 2014, initialized from the end of the coupled spin-up simulation. The simulation incorporates prescribed surface fossil fuel emissions, aircraft emissions, and oceanic CO_2 fluxes described in Section 2.2.3.

Figure 2 compares the prescribed CO_2 fluxes used in the BGCLNDATM_progCO2 simulation with the global carbon budget estimates documented by (Friedlingstein et al., 2021). The prescribed fluxes align well with observational estimates, providing confidence in the emissions inputs used for the simulation.

2.3.4. Rationale for Model Evaluation

We compared the BGCLNDATM_progCO2 simulation results with those from LNDATM_presCO2 and BGCLND to assess the consistency of the atmospheric model (EAM) between BGCLNDATM_progCO2 and LNDATM_presCO2, and the land model (ELM) between BGCLNDATM_progCO2 and BGCLND. The main difference between BGCLNDATM_progCO2 and the reference simulations is the treatment of CO_2—prognostic versus prescribed. If the coupling is successful and the model dynamics are preserved, implementing prognostic CO_2 should not substantially alter the global-mean climate results compared to the reference simulations. We also conducted benchmarking against observations and other CMIP6 models to evaluate the performance of BGCLNDATM_progCO2 within the context of the broader modeling community.

In general, we evaluated atmospheric CO_2 levels, both because prognostic CO_2 is the key advance being tested here and because of CO_2's role as primary GHG in the climate system; and surface air temperature because it provides an immediate sanity check that the prognostic-CO_2 wiring does not upset the large-scale physical climate of E3SMv2. Burned area, leaf area index (LAI), and spatial CO_2 variability are analyzed only to the extent required to explain the atmospheric-CO_2 bias and the land–surface temperature response. A full process-level evaluation of each flux pathway is better suited to follow-on studies.

2.4. Evaluation Metrics and Tools

We assessed the performance of BGCLNDATM_progCO2 using a range of evaluation metrics and diagnostic tools. Our evaluations include both the model spin-up phase and the transient historical period (1850–2014), highlighting key aspects of the model's behavior.

2.4.1. Spin-Up Evaluation Criteria

We evaluated the 500-year spin-up of the land biogeochemical state using several criteria to confirm that the model reached equilibrium and that the carbon pools stabilized:

1. Balanced Land-Atmosphere Carbon Exchange: The NEE between the land and atmosphere should be near zero with minimal drift, indicating that the terrestrial carbon pools have equilibrated. In the coupled spin-up simulation, the mean NEE over the 500-year constant-forcing (control) experiment was 0.05 Pg C yr^{-1}, well within the ± 0.1 Pg C yr^{-1} threshold recommended by the Coupled Climate–Carbon Cycle MIP (C4MIP; C. D. Jones et al., 2016).
2. Temporal Trends Analysis: We analyzed the time series of global mean total vegetation carbon, soil carbon, gross primary productivity (GPP), net primary productivity (NPP), and nitrogen and phosphorus
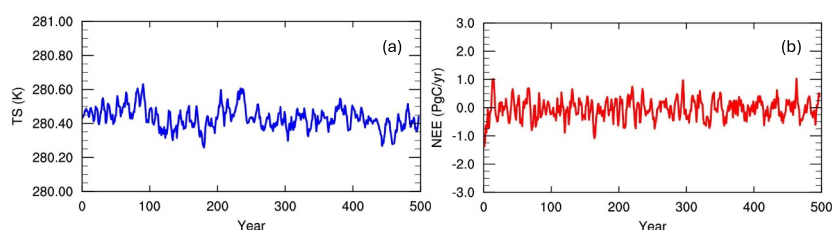
Figure 3. Time series from the coupled BGCLNDATM_progCO2 spin-up simulations showing (a) global mean surface air temperature (K) and (b) net ecosystem exchange (NEE; Pg C yr^{-1}).

mineralization rates. Minimal trends in these variables over time indicate that the model has reached a steady state, although no formal threshold criteria were applied to these trends. Details can be seen in Figure A1.

3. Qualitative Assessment of Spatial Distributions: We qualitatively assessed the spatial distributions of NEE, global total vegetation biomass, and global total soil carbon to ensure that the model produces realistic patterns consistent with observational understanding.

For the spin-up of the *coupled BGCLNDATM_progCO2 simulations*, the evaluation focused on global mean surface air temperature and NEE (Figure 3). Given that prescribed SSTs and sea ice concentrations constrain the climate, demonstrating the equilibrium of atmospheric radiative flux at the top of the atmosphere, as shown in Burrows et al. (2020), is not necessary for this study.

Figure 3 illustrates the 500-year land–atmosphere spin-up. The global-mean surface-air temperature oscillates about a constant value, confirming thermal equilibrium under prescribed SSTs. The global-mean NEE settles to fluctuations around zero after ~100 years, indicating that terrestrial carbon sources and sinks are balanced. This carbon balance equilibrium is further supported by the biogeochemical variables shown in Figure A1.

2.4.2. Evaluation of the Transient Historical Period (1850–2014)

To evaluate the performance of BGCLNDATM_progCO2 during the transient historical period, we used three open-source diagnostic tools that streamline analysis and benchmarking:

E3SM Diagnostics Package (E3SM Diags; C. Zhang et al., 2022) is a Python-based tool designed to assess E3SM outputs by comparing them with observations and historical benchmarks. It evaluates core aspects of the mean physical climate and includes process-oriented diagnostics. To streamline the evaluation processes, we used *zppy* (https://e3sm.org/resources/tools/end-to-end-processing/zppy/), a performant, Python-based post-processing
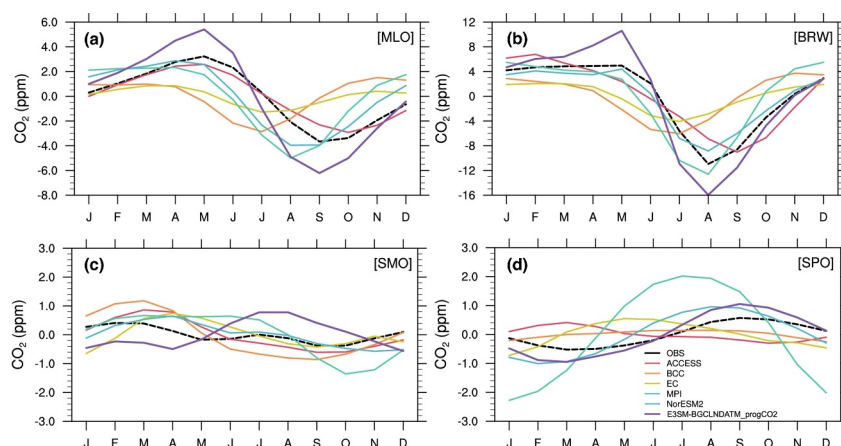
Figure 4. Comparison of atmospheric CO$_2$ seasonal variability and trend (the insert figure) among the BGCLNDATM_progCO2 simulation, other CMIP6 emission-driven simulations, and average observations (black line) at four monitoring sites from the NOAA ObsPack framework (Masarie et al., 2014) conducted by GDESS (Kaufman et al., 2022). The four observation sites are (a) Mauna Loa, Hawaii (MLO); (b) Barrow, Alaska (BRW); (c) Samoa (SMO); and (d) South Pole (SPO).
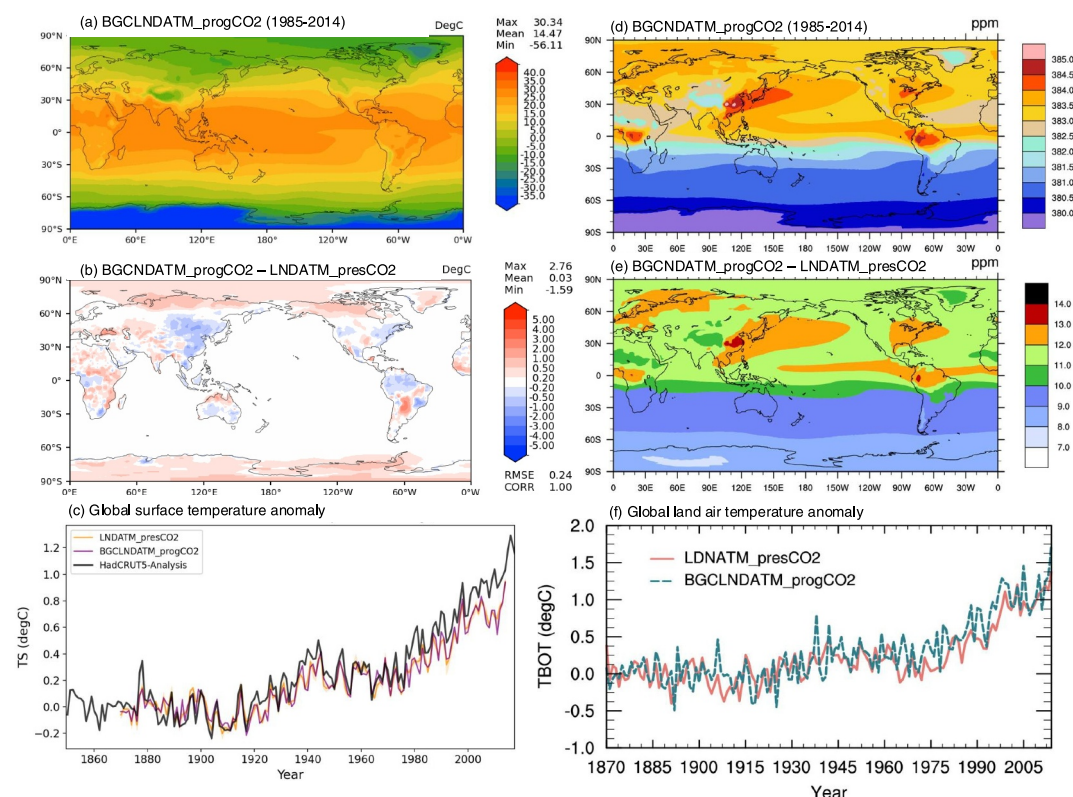
Figure 5. Comparison of the annual global mean surface temperature between the BGCLNDATM_progCO2 and LNDATM_presCO2 experiments. (a) The global mean temperature for BGCLNDATM_progCO2 from 1985 to 2014. (b) The spatial distribution of the surface air temperature difference between BGCLNDATM_progCO2 and LNDATM_presCO2 (BGCLNDATM_progCO2 minus LNDATM_presCO2) over the same period. (c) The time evolution of the anomaly relative to 1850–1899, with observations from the HadCRUT5 Analysis. (d) The atmospheric CO_2 concentration for BGCLNDATM_progCO2 averaged over 1985–2014. (e) The spatial distribution of the atmospheric CO_2 concentration difference between BGCLNDATM_progCO2 and LNDATM_presCO2 over the same period. (f) The time evolution of the global mean land surface air temperature anomaly relative to 1870–1899 for both simulations. Note that LNDATM_presCO2 is driven by prescribed annual global mean $[CO_2]$ time series, whereas BGCLNDATM_progCO2 simulates prognostic $[CO_2]$ that dynamically interact with the climate system. Note that the analyses for (a, b, d, e) were conducted by E3SM_digas (C. Zhang et al., 2022).

toolchain to automate routine tasks such as generating climatology and time series files, creating diagnostic plots with E3SM Diags (i.e., Figures 5a–5d and 6), and producing global time series plots (Figure 3). The development team is actively enhancing zppy by adding more land biogeochemical diagnostics, such as *International Land Model Benchmarking* (ILAMB; Collier et al., 2018), and ocean diagnostics, which will further extend its utility for evaluating comprehensive ESM simulations. This automation allowed efficient evaluation of the overall physical climate features of BGCLNDATM_progCO2, ensuring consistency with established climate patterns.

The ILAMB framework quantitatively assesses land surface models against a wide range of observational data sets, focusing on land biogeochemistry, biogeophysics, and hydrology. We employed ILAMB to evaluate key land biogeochemical processes in BGCLNDATM_progCO2, including energy, water, carbon and nutrient cycles (see Figure 7). By benchmarking model outputs against in situ and remote sensing observations spanning from 1960 to 2014, we identified strengths and areas for improvement in the land component of the model.

Greenhouse gas Diagnostics for Earth System Simulations (GDESS; Kaufman et al., 2022) evaluates ESMs by comparing simulated atmospheric CO_2 concentrations with observations. It processes data sets from the NOAA Observation Package (ObsPack) and generates statistical metrics and visualizations. We used GDESS to assess how well BGCLNDATM_progCO2 reproduces observed atmospheric CO_2 patterns (see Figure 4), aiming to identify discrepancies in simulating global CO_2 sources and sinks and to improve the representation of atmospheric CO_2 dynamics.
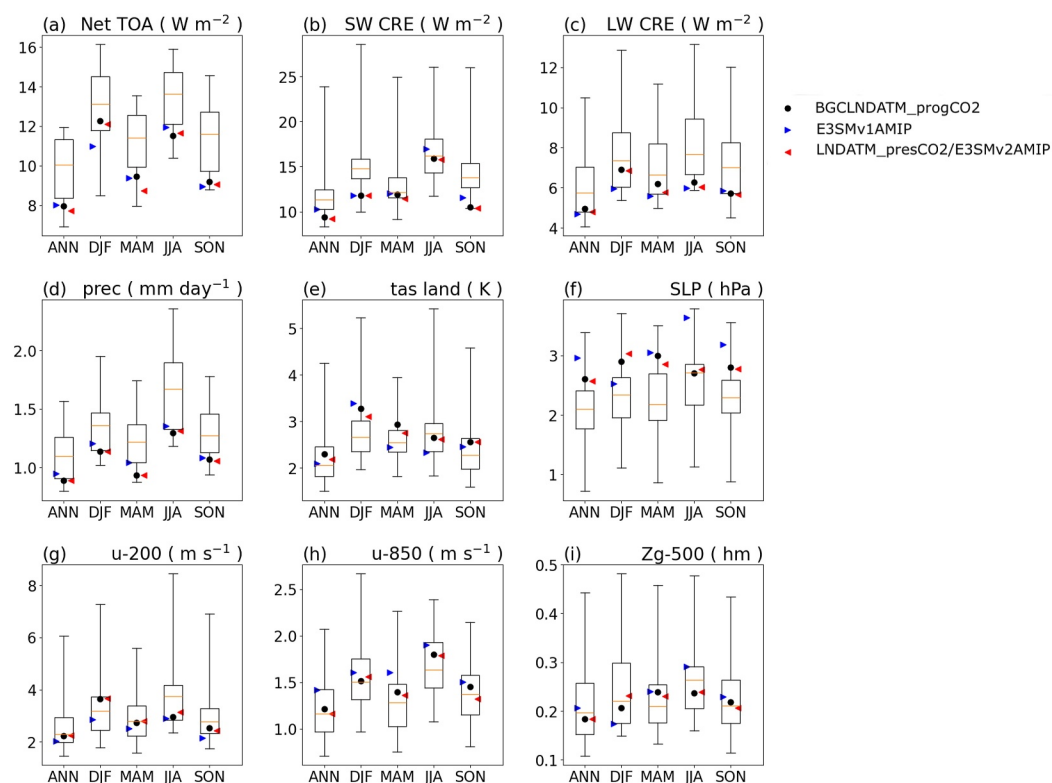
Figure 6. Comparison of RMSEs for the period 1985–2014 among an ensemble of 52 CMIP6 models (using the first historical members r1i1p1f1) and the first historical members of E3SMv1AMIP (blue triangles) and LNDATM_presCO2/ E3SMv2MIP (red triangles), including the BGCLNDATM_progCO2 simulation (black circles). The CMIP6 ensemble is represented by box-and-whisker plots showing the minimum, 25th percentile, median, 75th percentile, and maximum RMSE values. Spatial RMSEs relative to observations are calculated for annual and seasonal averages using the E3SM_Diags (C. Zhang et al., 2022). The fields evaluated include: (a) top-of-atmosphere (TOA) net radiation, (b) TOA shortwave (SW) cloud radiative effects (CRE), (c) TOA longwave (LW) CRE, (d) precipitation, (e) surface air temperature over land (tas), (f) sea-level pressure (SLP), (g) zonal wind (u) at 200 hPa, (h) zonal wind at 850 hPa, and (i) geopotential height (Zg) at 500 hPa. The mean climatologies of the reference observational and reanalysis data sets are derived from CERES-EBAF Edition 4.1 (Loeb et al., 2018; 2001–2014) for panels (a–c), GPCP Version 2.3 (Adler et al., 2018; 1985–2014) for panel (d), and ecosystem respiration (Hersbach et al., 2020 A5; 1985–2014) for panels (e–i). Seasonal abbreviations are DJF (December–February), MAM (March-May), JJA (June–August), and SON (September–November). The full E3SM_Diags diagnostic results are available at https://web.lcrc.anl.gov/public/e3sm/public_html/diagnostic_output/ac.sfeng1/paper/CBGCv2/ e3sm_diags/.

3. Evaluation of Historical Simulations

3.1. Evaluation of Prognostic Historical Atmospheric CO_2 Concentrations

Figure 4 presents the seasonal variability of atmospheric CO_2 concentrations ([CO_2]) at four representative monitoring sites: Mauna Loa, Hawaii (MLO); Barrow, Alaska (BRW); Samoa (SMO); and the South Pole (SPO). These sites were selected to capture diverse climatic regions, encompassing the global background atmosphere (MLO), high-latitude tundra environments (BRW), tropical marine conditions (SMO), and polar regions (SPO). The BGCLNDATM_progCO2 simulation consistently overestimates [CO_2] at all sites compared to observational records, with positive biases ranging from 11 to 23 ppm throughout the simulation period. To contextualize these overestimates, the CMIP6 emission-driven historical (esm-hist) simulations participating in the Coupled Climate–Carbon Cycle MIP (C4MIP; C. D. Jones et al., 2016) exhibit a wide range of atmospheric CO_2 concentrations. Specifically, most models (10 out of 13) fall within a CO_2 concentration range of approximately 40 ppm (Figure B1), representing some 20 years of historical emissions. The positive biases in our simulation are therefore within the spread of the CMIP6 models, suggesting that our model's overestimates are comparable to those observed in other emission-driven ESMs. This indicates that such biases are common and underscores the need for greater attention to the calibration of historical CO_2 concentrations in emissions-driven simulations.
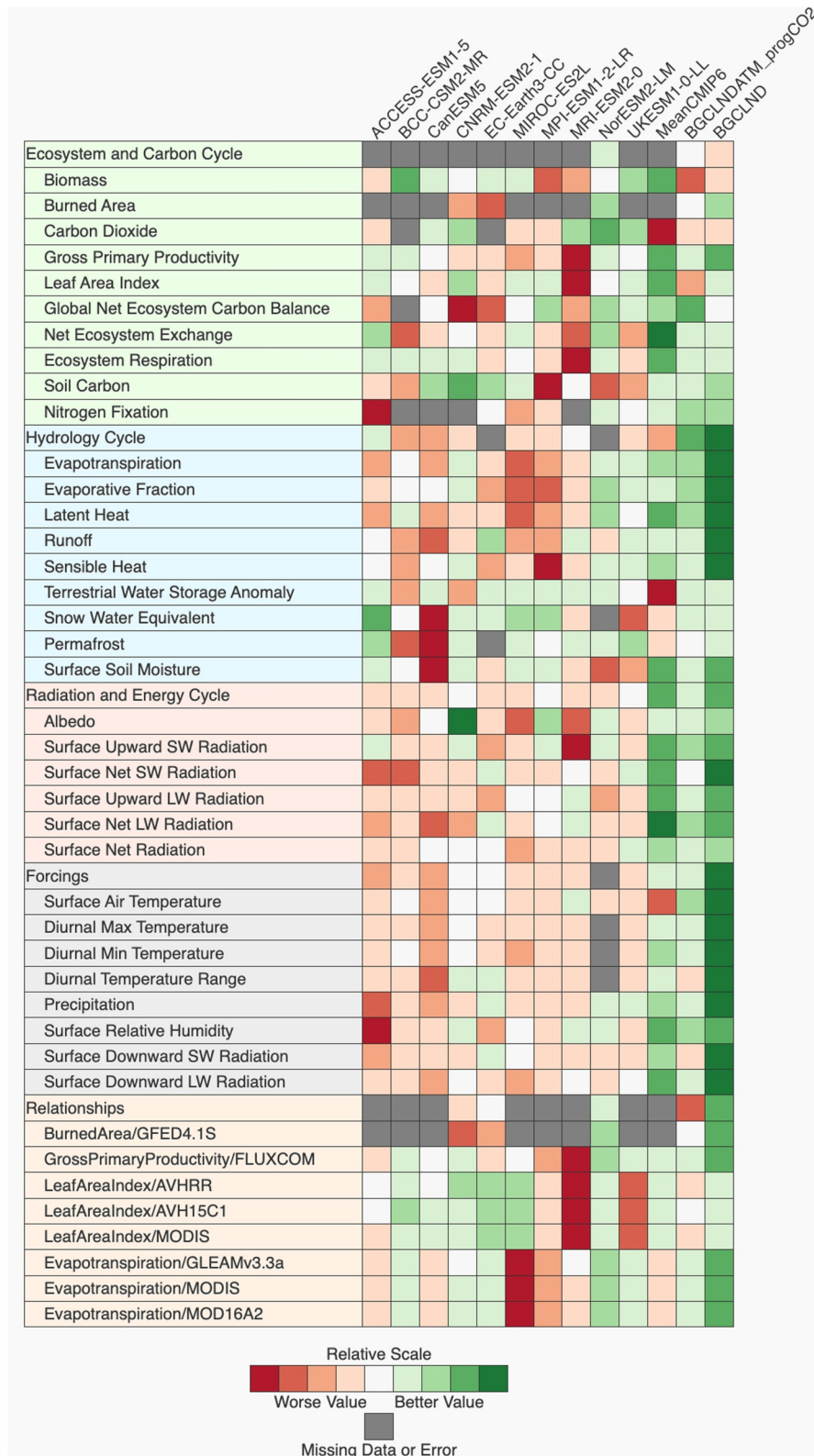
Figure 7.

Beyond the long-term trends, we examined the seasonal variability of $[CO_2]$ at these sites. The BGCLNDATM_progCO2 simulation demonstrates consistently better representation of seasonal cycles compared to other emission-driven models, particularly in terms of phase accuracy, although it tends to overestimate the amplitude of peak values. This overestimation is more pronounced at MLO and BRW than at SMO and SPO, suggesting regional differences in the model's simulation of seasonal carbon exchanges.

Further investigation reveals that the overestimation of both the trend and the magnitude of the seasonal cycle is partially due to the weak uptake of prescribed ocean CO_2 fluxes in the model (Figure C1 and C2). The prescribed oceanic CO_2 fluxes, based on CESM2 historical simulations, may underestimate the strength of oceanic carbon sinks during the simulation period, contributing to higher $[CO_2]$ and amplified seasonal amplitudes in the simulation. ILAMB diagnostics further show that the terrestrial sink is also too weak: both GPP and ecosystem respiration (ER) are biased low, leaving NEE less negative than the observation-based estimate (Figure D1). Such land-flux uncertainty is known to exceed ocean-sink uncertainty and is a primary driver of the seasonal CO_2 cycle (Piao et al., 2018). Hence, the combination of an under-simulated ocean sink and a muted land uptake explains much of the positive CO_2 bias. Remaining discrepancies may also reflect uncertainties in prescribed fossil-fuel emissions or limitations in the model's atmospheric transport.

3.2. Evaluation of Physical Climate Variables

Given the consistent positive biases in $[CO_2]$ shown in Figure 4, one might expect BGCLNDATM_progCO2 to have a warmer climate compared to the LNDATM_presCO2 simulation. When examining the global mean surface air temperature, we find that there is minimal difference between the two simulations (Figure 5c), consistent with Burrows et al. (2020). This minimal difference is likely due to both simulations being strongly constrained by the prescribed SSTs, as the ocean covers approximately two-thirds of the Earth's surface. Over the period 1985–2014, the average difference in $[CO_2]$ between the two simulations is approximately 12.5 ppm, which corresponds to only a 0.03°C difference in the global mean surface air temperature. However, this temperature difference is amplified in over the land ranging from −1.59 to 2.76°C over the 30-year period at the regional scale. Because the prognostic set-up explicitly resolves spatial CO_2 gradients—rather than imposing a single global-mean value—the climatic impact of elevated CO_2 can emerge heterogeneously, particularly over land areas, potentially due to interactions with regional climate processes and feedback. Elevated or heterogeneous CO_2 distributions are not, however, the only possible source of the temperature differences observed here. Internal variability within the climate system could also contribute to these regional temperature differences. Comparisons with additional ensemble members in LNDATM_presCO2 (Golaz et al., 2022) indicate that typical internal variability leads to 30-year temperature differences ranging from −0.5 to 0.5°C. Since the differences between BGCLNDATM_progCO2 and LNDATM_presCO2 exceed this range, it suggests that they may be influenced by the higher $[CO_2]$. Nevertheless, without multiple ensemble runs for BGCLNDATM_progCO2, we cannot conclusively separate the effects of internal variability from the forced response due to CO_2 differences. Additionally, the change from satellite phenology (used in LNDATM_presCO2) to prognostic LAI (used in BGCLNDATM_progCO2) could alter the land surface sufficiently to affect the surface energy balance and, hence, the surface temperature distribution. Disentangling these (and potentially other) factors is beyond the scope of the current work, and we leave testing these hypotheses to future research efforts.

The spatial distribution of the CO_2 concentration difference between the two simulations over 1985–2014 shows that BGCLNDATM_progCO2 consistently has higher $[CO_2]$ than the prescribed global mean CO_2 time series used in LNDATM_presCO2 (Figure 5e). This difference is more pronounced over the Northern Hemisphere, with hotspots located in major industrial regions such as eastern China, the eastern United States, and the Amazon (due to fire emissions), as well as downwind areas influenced by atmospheric transport.

Figure 7. The overall performance of land model configurations of BGCLNDATM_progCO2 compared with BGCLND and a subset of emission-driven CMIP6 simulations conducted by ILAMB (Collier et al., 2018). In the ILAMB scorecard, net biome production is labeled as "Global net-ecosystem carbon balance." The CMIP6 subset includes models that (i) completed the full 1850–2014 emissions-driven historical simulation and (ii) provided the complete set of monthly land–biogeochemistry variables required by ILAMB—such as gross primary productivity, net ecosystem exchange, biomass, leaf area index, soil carbon, evapotranspiration, and surface meteorology—on the Earth System Grid Federation archive. These criteria ensured a consistent and comprehensive basis for evaluating BGCv2LNDATM against other models. The full ILAMB results are available at https://doi.org/10.25584/2525948.

Figure 6 summarizes how the prognostic simulation performs for a suite of large-scale climate fields. Spatial root-mean-square errors (RMSEs) for annual and seasonal means were calculated with the E3SM Diagnostics Package (C. Zhang et al., 2022) relative to observations or reanalyzes. For every metric except sea-level pressure, the RMSE of BGCLNDATM_progCO2 (prognostic CO_2) falls within—or below—the inter-quartile range of the CMIP6 ensemble and is comparable to that of the reference run LNDATM_presCO2 (prescribed CO_2) (Golaz et al., 2022). This close agreement shows that activating prognostic CO_2 leaves the core physical-climate skill of E3SMv2 unchanged and, for several fields, yields performance that is as good as or better than the median CMIP6 model.

The consistency of the global-mean physical climate properties between BGCLNDATM_progCO2 and LNDATM_presCO2 lends confidence in the coupling implementation. At the same time, the differences in land surface air temperature show that spatially varying, prognostic CO_2 produces a measurable regional signal. Thus, the new configuration preserves the large-scale energy balance while revealing regional carbon-climate feedback that cannot be captured with a globally prescribed CO_2 field.

Further research is needed to quantify how CO_2 spatial variability impacts regional surface air temperature responses and the underlying biogeochemical processes. Because the global-mean climate baseline is preserved, the differences in regional land-surface patterns shown in Figure 5 can be attributed to the spatially heterogeneous CO_2 forcing rather than to energy-budget drift, giving the confidence that subsequent sensitivity studies will target the intended biogeochemical mechanisms rather than reflecting artifacts of the coupling.

3.3. Observational Benchmarking of Land Biogeochemical and Physical Variables

In this section, we assess the performance of the land component of BGCLNDATM_progCO2 against observational data sets and compare it with the land stand-alone simulation (BGCLND) and other emission-driven CMIP6 simulations. A high-level summary of this assessment is presented in Figure 7, with detailed geographic and temporal comparisons available online (https://doi.org/10.25584/2525948).

As expected, due to the climate constraints from using prescribed ocean data, BGCLNDATM_progCO2 is expected to outperform the coupled models in the hydrological cycle, radiation and energy cycles, and forcings categories. Therefore, we focus the land process evaluation on the ecosystem and carbon cycle categories and functional responses of the carbon cycle. It is important to note that the ILAMB framework infers $[CO_2]$ primarily from land carbon fluxes, while observational benchmarks include $[CO_2]$ contributions from other sources, that is, anthropogenic activities and the ocean. Therefore, we recommend directly comparing the simulated total $[CO_2]$ with the observational benchmark, as illustrated in Section 3.1.

BGCLND, driven with reanalysis data (see Table 1), exhibits overall better performance in most categories except for biomass, which likely reflects potential problems in ELMv2. Compared with other CMIP6 emission-driven models, BGCLNDATM_progCO2 produces consistently better results in GPP, global net ecosystem carbon balance (represented a net biome production NBP), NEE, and Soil Carbon, and Nitrogen Fixation, but underperforms biomass, burned area, LAI, and ER.

Although several models—including BGCLNDATM_progCO2—simulate global biomass that exceeds most observational estimates, the spread of both observations and models is large. Observational estimates of global biomass range from 378 Pg for 2000–2020 in XuSaatchi2021 (Xu et al., 2021) to 588 Pg for 2000–2010 in GEOCARBON (Araza et al., 2022). A comparison with other biomass products can be found in Figure D2, showing that no two observational products fall on the 1:1 line. For models, BGCLNDATM_progCO2 simulates 706 Pg while the CMIP6 models range from 367 Pg to 511 Pg, with an ensemble mean of 452 Pg. Despite the large uncertainty, both BGCLNDATM_progCO2 and BGCLND (667 Pg) produce excessive biomass, which likely arises from multiple processes. First, the ILAMB scorecard shows that burned area is underestimated, implying smaller fire carbon losses (Li et al., 2024). Second, the parameterizations that control tree mortality, and the turnover of leaves, stems, and roots determine how long carbon resides in vegetation; a residence time that is too long will by definition inflate standing biomass. ILAMB's derived metric for vegetation carbon residence time (live biomass/NPP) is 34 yr in BGCLNDATM_progCO2 versus 20 yr in the observational benchmark, confirming that carbon remains in vegetation too long and indicating that background mortality and turnover rates in ELMv2 may be too low. Third, insufficient soil nutrient limitation, particularly phosphorus limitation in tropical rainforests, could allow productivity to remain high (Du et al., 2025; Ellsworth et al., 2022; He et al., 2025; Terrer

et al., 2019). However, confirming this hypothesis would require further analysis of plant phosphorus limitation factors in BGCLNDATM_progCO2, which is beyond the scope of this study.

Due to the lack of the burned areas across CMIP6 emission-driven model output from the ESGF data archive, we limited the comparison of simulated burned areas in BGCLNDATM_progCO2 and BGCLND. While the observational benchmark from GFED4.1s (Giglio et al., 2013; Randerson et al., 2012; van der Werf et al., 2010) indicates a mean annual burned area of 0.33%, BGCLNDATM_progCO2 and BGCLND burned 0.14% and 0.23%, respectively, indicating a significant underestimation by ELMv2. Furthermore, the land-atmospheric coupling with biogeochemistry in BGCLNDATM_progCO2 reduces the burned area even further, partly due to biases in the simulated atmospheric forcing—surface air temperature, diurnal max and min temperature, precipitation, surface relative humidity, and downward SW/LW radiation, shown in the "Forcings" section of Figure 7—in BGCLNDATM_progCO2.The significant bias in burned area primarily appears in tropical African grasslands and broadleaf areas (Figure D3). A comparison of seasonal variations between the observations and models reveals that E3SMv2 significantly underestimates burned areas during the peak fire seasons (Figure D4). These discrepancies highlight the need for further investigation into the representation of fire processes in ELMv2.

Fire significantly alters land cover and carbon stocks, releasing aerosols and GHGs. Post-fire landscapes exhibit different albedo and vegetation dynamics, influencing land-atmosphere interactions and contributing to climate variability. ILAMB evaluates the functional responses between burned area and precipitation (Adler et al., 2018) and surface air temperature (Harris et al., 2014). Due to E3SM's underestimation of burned area over equatorial Africa, particularly during peak seasons, the relationships between simulated burned area and both precipitation and temperature are inconsistent with observations (Figure D5).

Leaf area index (LAI) is a crucial variable affecting climate by modulating evapotranspiration and influencing the exchange of water and energy between the land and atmosphere. The observational benchmark for global mean LAI is 1.36 (Figure D6), derived from monthly climatology of MODIS total LAI observations. Both BGCv2LANDATM and BGCLND are within a margin of 0.1 of the observational benchmarks, compared to a broader range of 1.2–2.46 for CMIP6 models. However, BGCLNDATM_progCO2 has a lower overall score for LAI due to its failure to accurately capture the spatial distribution, particularly in tropical regions where RMSEs are evident. This underperformance in LAI is present only in the coupled E3SM models, not in the standalone land experiment BGCLND, which suggests that d atmospheric forcings and/or biogeochemical feedback.

This deficiency is also apparent in the relationship between the prognostic LAI and precipitation forcing in the model (Figure D7). E3SM experiments consistently overestimate precipitation when LAI is less than ∼2.5 and underestimate precipitation when LAI exceeds ∼2.5. Further regional analysis reveals that the underestimation of precipitation becomes more pronounced when LAI is greater than 4 in tropical regions. However, caution is necessary when integrating these biases due to the large uncertainty in the satellite-based LAI estimations over tropical regions (Mannschatz et al., 2014). Presumably, this discrepancy arises from a combination of biases in water and energy fluxes in EAM compared to observations; however, pinpointing the exact contributions of these biases will require dedicated process experiments beyond the scope of this paper.

As for ER, both BGCLNDATM_progCO2 and BGCLND exhibit similar performance in terms of mean state, spatial distribution, and seasonal cycle (Figure D8). ELMv2's ER performance is within the model uncertainty bounds; the land–atmosphere coupling does not compromise ER performance. However, both E3SM simulations significantly underestimate ER in the warm seasons.

BGCLNDATM_progCO2 simulates global net ecosystem carbon balance (represented by net biome productivity) of −42.1 Pg C (Figure D9), closely aligning with the observational estimate of −44.6 Pg from the Global Carbon Project from 1959 to 2016 (Le Quéré et al., 2016), whereas other models range from −4.09 Pg to −105 Pg. The standalone land simulation BGCLND estimates an NBP of −57.3 Pg. The land-atmosphere coupling in BGCLNDATM_progCO2 somehow reduces carbon loss, which can be partially explained by reduced burned areas in the simulation.

For GPP, all the models underestimate GPP compared to the FLUXNET benchmarks, which report a value of $3.35 \, \mathrm{g \, m^{-2} \, d^{-1}}$ (Knauer et al., 2018; Lasslop et al., 2010; Reichstein et al., 2007). Both BGCLNDATM_progCO2 ($2.35 \, \mathrm{g \, m^{-2} \, d^{-1}}$) and BGCLND ($2.54 \, \mathrm{g \, m^{-2} \, d^{-1}}$) are within the uncertainty bounds of the CMIP6 emission-driven

models, which range from 2.16 to 2.87 g m^{-2} d^{-1} (Figure D10). However, a clear improvement is evident in ELMv2 for the seasonal cycle of GPP compared to other models. Similar characteristics extend to NEE.

ILAMB evaluates the functional relationships of GPP with evapotranspiration (ET), precipitation, surface downward shortwave (SW) radiation, surface net radiation, and surface air temperature. The relationship between GPP and ET (Figure D11) arises from their shared dependence on environmental drivers such as solar radiation and air temperature. BGCLNDATM_progCO2 tends to overestimate ET when GPP is less than 2 g C m^{-2} day^{-1} and underestimate ET when GPP exceeds this threshold globally. This bias is predominantly driven by discrepancies in tropical regions. Specifically, regions with high GPP values—greater than 5 g C m^{-2} day^{-1} in Equatorial Africa, 8 g C m^{-2} day^{-1} in Equatorial Asia, and the Amazon—show the most pronounced biases (Figure D12).

In terms of soil carbon stock, all the models simulated a range from 1,110 to 1,750 Pg, while the observation benchmark is 1,330 Pg (Ito et al., 2020; Wieder et al., 2014). Among the comparisons, BGCLNDATM_progCO2 and BGCLND simulated soil carbon are 1,140 and 1,200 Pg, respectively, which relatively underestimates soil carbon. However, BGCLNDATM_progCO2 shows a better spatial distribution score than BGCLND, significantly reducing the bias over the Northern Hemisphere's high-latitude needleleaf deciduous region, although a larger bias over the tropical region is evident (Figure D13).

Nutrient availability influences vegetation growth, carbon uptake, and biogeochemical feedback, which are crucial for simulating long-term carbon-climate interactions. CMIP6 emission-driven models ranging from 44.0 to 208 Tg yr^{-1} with an ensemble mean of 105 Tg yr^{-1}, while the observational benchmark is 88 Tg yr^{-1} (Davies-Barnard & Friedlingstein, 2020). ELMv2 experiments display the closet agreement with the benchmark, showing 93 Tg yr^{-1} for BGCv2LNATM and 94.7 Tg yr^{-1} for BGCLND, respectively, and outperforming in spatial distribution (Figure D14).

In conclusion, the evaluation of land biogeochemical and physical variables in BGCLNDATM_progCO2 reveals a mixed performance. While the model aligns well with observational benchmarks in simulating global net ecosystem carbon balance (or NBP), soil carbon, and nitrogen fixation, it exhibits notable biases in biomass, burned area, and LAI, particularly in tropical regions. These discrepancies suggest potential issues in the representation of fire processes, vegetation dynamics, land-atmosphere interactions, and atmospheric forcing biases within the coupled model. Addressing these challenges is crucial for improving the model's ability to simulate carbon–climate feedback.

4. Planned Improvements to Biogeochemistry, Land, and Energy Systems in E3SM

Section 4 outlines planned model extensions and is included for project-documentation purposes; readers primarily interested in the current evaluation may proceed directly to the next section. To achieve the full emissions-driven capacity for E3SM, our initial focus has been on implementing a land–atmosphere interactive configuration in E3SMv2.1 biogeochemistry. This approach allows us to isolate land–atmosphere interactions and feedback from other components, such as the ocean, facilitating a detailed examination of terrestrial processes and their impacts on atmospheric CO_2 concentrations. The success of this first step provides a solid foundation for integrating additional components and processes into the model.

Future versions of E3SM will incorporate enhanced biogeochemical processes, including the activation of ocean and sea ice biogeochemistry modules. Integrating marine biogeochemistry is a critical next step toward achieving a fully coupled emissions-driven ESM, as the ocean plays a significant role in global carbon cycling and climate regulation. The ocean absorbs approximately one-quarter of anthropogenic CO_2 emissions annually (Friedlingstein et al., 2024), and capturing this dynamics accurately is essential for reliable climate projections.

The Marine Biogeochemical Library (MARBL, Long et al., 2021) was recently ported to E3SM for version 2.1. A suite of dynamic ocean-ice simulations was conducted to validate the biogeochemical model coupling with MPAS-Ocean, the physical ocean model of E3SM; to validate 2-way coupling with the biogeochemical model (Jeffery et al., 2020) of MPAS-Seaice, the physical sea ice model of E3SM; and to understand the ecosystem model behavior under prescribed forcing conditions (Nissen et al., 2024; Takano et al., 2023). Recent advances in ocean physical processes in E3SM (Smith et al., 2024) have improved the fidelity of ocean biogeochemical processes, setting the stage for fully coupled, emissions driven, biogeochemical simulations in E3SMv3. This will enable feedback mechanisms between ocean circulation and biogeochemical cycles.

In addition to marine biogeochemistry, new land features are implemented in E3SMv3-BGC, including a new plant hydraulic scheme (Fang & Leung, 2022), a new crop module including more than 30 new crop species (Sinha, Calvin, et al., 2023), a two-way land-river coupled water management module (Zhou et al., 2020), a topographic unit representation of sub-grid heterogeneity of surface forcings (Tesfa et al., 2024), and Flow of Agricultural Nitrogen fertilization representation (Vira et al., 2022). With these new features, both natural vegetated land surface and agriculture lands will be significantly impacted, in terms of carbon and water cycles. The new development of the land model will be used to support the upcoming coupled E3SMv3 BGC simulation campaign that aims to couple land, atmosphere, ocean, and sea ice components with emission-driven configuration.

Lastly, the E3SM team has implemented a two-way online coupling between E3SM and the Global Change Analysis Model (GCAM) (Di Vittorio et al., 2025). By enabling two-way feedback between GCAM and E3SM, the impacts of land-use change, agriculture, and energy production on biogeochemical cycles and climate can respond dynamically to climate. By linking socio-economic drivers with a process-based ESM, the coupled GCAM-E3SM system enables exploratory studies of mitigation pathways and their physical feedback; further work will be required before such simulations can be used in formal policy assessments (W. D. Collins et al., 2015; A. D. Jones et al., 2018; Peter E. Thornton et al., 2017).

5. Discussion

This work delivers three advances not previously available for E3SM or documented in the broader CMIP literature. (i) It introduces BGCLNDATM_progCO2—the first E3SM configuration that couples interactive land biogeochemistry with prognostic atmospheric CO_2 while retaining prescribed sea-surface conditions, thereby enabling land–atmosphere carbon feedback in an emission-driven framework. (ii) It provides the first public set of E3SMv2.1 historical hindcasts with prognostic CO_2, creating a new benchmark for future studies. (iii) It offers the community's first end-to-end evaluation protocol tailored to emission-driven simulations, integrating GDESS, E3SM-Diags, ILAMB and zppy into a reproducible workflow that can be adopted directly by future fully coupled simulation campaigns. Together, these contributions extend E3SM capability, supply new open data, and present a transferable framework for rigorous, standardized assessment of emission-driven Earth-system simulations.

In our evaluation, the global mean surface air temperature showed minimal differences between the emissions-driven and reference simulations due to the strong constraint of prescribed SSTs. However, significant regional differences were observed in land surface air temperature. This indicates that while biogeochemical feedback onto global mean physical climate properties are marginal, they can have a measurable impact on regional climates, especially over land. Understanding these interactions is essential for projecting future climate change (Arora et al., 2020; Cox et al., 2000; Hewitt et al., 2016; C. G. Jones et al., 2024; Reichstein et al., 2013; Vichi et al., 2011; Vonk et al., 2025).

The underestimation of burned area and overestimation of biomass highlights the need to improve the representation of fire processes within ELMv2. Incorporating sophisticated fire modules that account for fire dynamics could lead to more realistic fire simulations (Li et al., 2024). In light of these findings, we have investigated the fire parameters within ELMv2 and implemented new calibrations based on our results. While reduced fire mortality is a plausible first-order cause of the biomass bias, other factors cannot be excluded at this stage. An overly strong GPP, a low representation of plant or microbial respiration, and vegetation turnover rates that are too slow would all act to increase standing biomass. ILAMB diagnostics of carbon-residence time (Section 3.3) suggest that background mortality is weak, pointing to a combination of insufficient disturbance and long turnover. Future sensitivity tests that vary GPP, respiration and turnover parameters alongside the updated fire scheme will be necessary to quantify the relative contributions of these processes. Refining these components will in turn improve simulated biomass accumulation and carbon emissions, leading to more accurate terrestrial carbon budgets (Dolman et al., 2012; Forkel et al., 2025; Randerson et al., 2006; Seiler & Crutzen, 1980).

We also note that the biases in LAI and ER, particularly in tropical regions, suggest that vegetation dynamics and land–atmosphere interactions require further refinement. Enhancing the model parameters for vegetation phenology and physiology, together with biogeochemical feedback, may improve the spatial and temporal variability of LAI and ER (Woodward & Lomas, 2004; Zhao et al., 2025). Investigating the coupling between atmospheric forcings and land surface processes is crucial for resolving these discrepancies.

Our evaluations underscore the challenges inherent in emissions-driven modeling. Accurately simulating atmospheric CO_2 concentrations, physical climate variables, and land biogeochemical processes requires precise representations of complex interactions among Earth system components. Positive biases in atmospheric CO_2 concentrations, along with regional temperature differences (over land), highlight the challenges in calibrating models to reproduce historical concentrations and observed warming trends simultaneously. Careful consideration is needed to understand how regional biases may offset each other, leading to minimal differences in global averages. Part of the warm-land signal may stem from the fact that BGCLNDATM_progCO2 uses prognostic LAI whereas the AMIP reference run (LNDATM_presCO2) prescribes satellite LAI; altered canopy forcing can modify surface energy balance independently of CO_2. A concentration-driven experiment that retains prognostic LAI is planned to quantify this contribution. This underscores the complexity of accurately simulating both global and regional climate responses in emissions-driven models. However, in C4MIP, 10 out of 13 CMIP6 models driven by historical emissions exhibited [CO_2] concentrations within a 40 ppm range (Liddicoat et al., 2021; Sanderson et al., 2024), encompassing the biases observed in our simulation.

A final challenge associated with running models in emissions-driven mode is the additional degrees of freedom involved in calibrating the coupled climate–carbon cycle system to reproduce both the historical concentrations of climate forcers and the observed warming trends (C. D. Jones et al., 2016; Sanderson et al., 2023). Accurately simulating the joint evolution of atmospheric CO_2 and global temperatures requires precise representations of carbon sinks and sources, as well as feedback mechanisms. The biases observed in our simulation underscore the complexity of modeling the coupled climate–carbon system and highlight the need for further investigation into the underlying causes of these discrepancies.

6. Toward a Protocol for Evaluating Emissions-Driven Simulations

Although many of the individual diagnostics used here are well established, they have not previously been combined into a single, end-to-end workflow tailored to the requirements of emission-driven simulations. Building upon the methodologies and evaluation results reported and discussed above, we propose a structured, and to some degree standardized, evaluation protocol for emissions-driven simulations with ESMs like E3SM. Several steps (e.g., physical-climate metrics) are common to concentration-driven runs, whereas the CO_2 observational benchmarking step is specific to emission-driven simulations. This protocol aims to ensure scientific quality, facilitate model intercomparison, and guide future modeling efforts, and consists of the following key components:

First, a robust and standardized carbon-cycle spin-up protocol is essential for stabilizing the model's carbon and nutrient pools before conducting emissions-driven simulations. To confirm that the model has reached equilibrium and starts from a stable state, we recommend verifying that NEE approaches zero with minimal drift (ideally within ± 0.1 Pg C yr^{-1}), as per C4MIP guidelines (C. D. Jones et al., 2016), ensuring balanced land-atmosphere carbon exchange. Monitoring the temporal stability of key variables—such as global mean surface air temperature, vegetation carbon, soil carbon, GPP, NPP, and nutrient mineralization rates—is important to confirm minimal trends over the spin-up period. Qualitative assessment of the spatial distributions of NEE, vegetation biomass, and soil carbon helps verify that the model produces realistic patterns consistent with observations. These checks ensure that simulations begin from a well-initialized state, reducing the likelihood of artifacts influencing the results.

Second, benchmarking of atmospheric CO_2 is unique to emission-driven and crucial in emissions-driven models, as atmospheric CO_2 is central to assessing carbon-climate feedback. We recommend benchmarking the modeled atmospheric CO_2 against observational records from key long-term monitoring sites (e.g., Mauna Loa, Barrow, Samoa, SPO) to evaluate the model's ability to capture both decadal trends and seasonal variability. Tools such as GDESS (Kaufman et al., 2022) facilitate this comparison by processing model outputs and observational data to generate statistical metrics and visualizations.

Third, it is essential to verify that the implementation of prognostic CO_2 maintains the integrity of the physical climate. Evaluating global mean surface air temperature between the simulations and observations or reference runs helps identify any significant deviations, as temperature provides an immediate "sanity check" on the large-scale physical climate system. Spatial analysis of temperature and precipitation differences can reveal regional impacts of the prognostic CO_2 implementation, which as we show here can be substantial even when global

metrics are stable. Comprehensive diagnostic tools like the E3SM Diags (C. Zhang et al., 2022) allows for systematic evaluation of climate variables against observational data sets and reanalyses.

Fourth, accurate representation of land biogeochemical and physical variables, such as biomass, LAI, soil carbon, ER, GPP, and NEE, is essential for simulating carbon-climate interactions. We recommend benchmarking model outputs against observational data using the ILAMB package (Collier et al., 2018). Examining functional relationships between variables (e.g., GPP vs. evapotranspiration), in addition to simple model-observation comparisons, tests whether the model responds realistically to environmental drivers. Analyzing mean states, spatial distributions, and seasonal cycles help identify regional biases and temporal trends, guiding model improvements.

Fifth, having a standardized, reproducible evaluation process is critical. Automation frameworks like zppy streamline routine tasks such as generating climatology files, creating diagnostic plots, and producing global time series. Thorough documentation of evaluation steps, model configurations, and analysis scripts supports transparency and comparability across studies. Sharing diagnostic outputs and tools through accessible platforms facilitates collaboration and accelerates progress within the modeling community.

While the examples above use the E3SM toolchain, each step of the protocol can be executed with model-agnostic packages such as ESMValTool (Eyring et al., 2020; Righi et al., 2020), or the PCMDI Metrics Package (Gleckler et al., 2008). Adopting these community resources (Bock et al., 2020; Gier et al., 2024; Spafford & MacDougall, 2021) allows modelers or other users to apply the same evaluation logic to their own emission-driven simulations, thereby supporting consistent benchmarking across CMIP-class models.

Finally, using evaluation results to inform model development is a crucial aspect of the protocol. Identifying key biases allows for targeted refinements in model components that most significantly impact critical outcomes, such as atmospheric CO_2 concentrations and carbon fluxes. Continuously reassessing the model after modifications ensures that changes lead to improved performance without introducing new biases. This iterative approach enhances the model's fidelity in simulating emissions-driven climate dynamics.

By following this structured evaluation protocol, one can systematically assess and enhance the performance of emissions-driven simulations. The protocol emphasizes the importance of starting from a well-initialized state, accurately capturing key atmospheric and terrestrial processes, and maintaining the integrity of the physical climate.

In conclusion, this work establishes a foundational step for advancing E3SM's emissions-driven capabilities, emphasizing land-atmosphere interactions. Building on these results, future E3SM model development will incorporate ocean and sea ice biogeochemistry and extend land features, enhancing the model's representation of global carbon cycles. The integration with the GCAM will enable dynamic feedback between human and Earth systems, improving projections of land-use and energy impacts on climate. These developments will enhance E3SM's ability to simulate complex climate dynamics and provide a more robust basis for future analyses of mitigation pathways and climate-feedback.

Appendix A: Coupled Simulation Spin-Up Evaluations

A1. Equilibrium Evaluation of Biogeochemical Variables
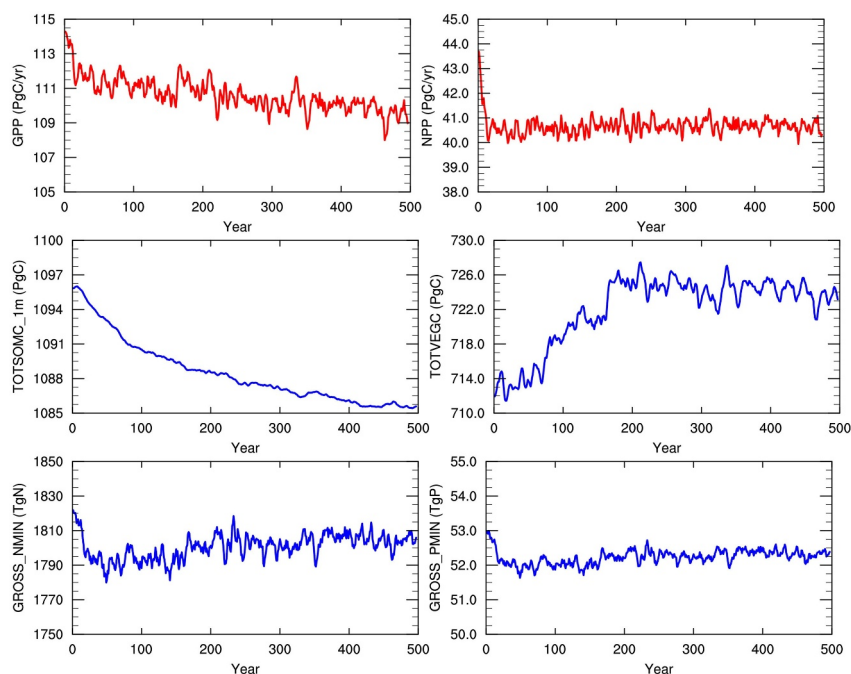
See Figure A1.

Figure A1. Time series from the coupled BGCLNDATM_progCO2 spin-up simulations showing gross primary productivity (PgC/yr), net primary productivity (PgC/yr), total soil organic carbon in the top 1 m (TOTSOMC_1m; PgC), total vegetation carbon (TOTVEGC; PgC), gross rate of nitrogen and phosphorus mineralization (GROSS_MINN and GROSS_PMIN in TgN and TgP), respectively.

A2. Time Series of Global-Mean [CO$_2$] During the Spin-Up of Coupled Emissions-Driven Simulations

Leveraging the emissions-driven configuration, we also track the global-mean atmospheric CO$_2$ concentration: it declines smoothly from the initial 285 ppm to $\approx$273 ppm and then stabilizes. This downward adjustment reflects the interactive land sink, together with the prescribed ocean sink, drawing the atmosphere toward the pre-industrial ice-core value. The subsequent plateau in [CO$_2$], combined with near-zero NEE and stabilized key biogeochemical variables, demonstrates that the coupled system has reached carbon equilibrium.

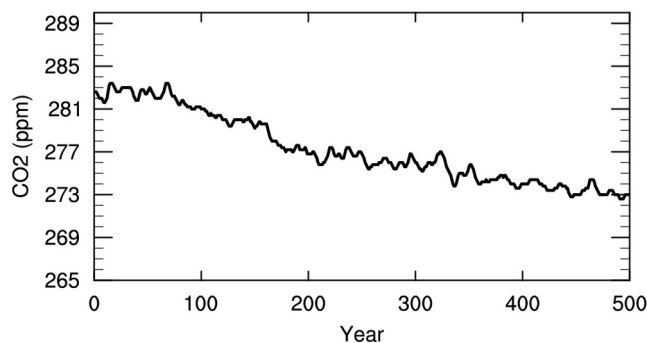**Figure A2.** Time series of global mean [CO$_2$] (ppm) from the coupled BGCLNDATM_progCO2 spin-up simulation.

Appendix B: Global Mean CO_2 Time Series in Emission-Driven Models

Figure B1 shows that BGCLNDATM_progCO2 starts 1850 at a lower CO_2 concentration (∼270 ppm) than most C4MIP models (typically 275–286 ppm). This difference reflects our spin-up procedure (Figure A2). Running with prescribed 1850 ocean CO_2 flux and the interactive land sink to pull atmospheric CO_2 toward the ice-core value. Because subsequent historical emissions match those of the other esm-hist simulations, the lower starting point produces a larger absolute rise to 2014 while the percentage increase remains typical of the ensemble. The post-1960 positive bias is dominated by an under-simulated ocean sink and residual land-process errors, not by the initial condition. Future work will repeat the spin-up with a fully prognostic ocean-biogeochemistry module and a 285-ppm target to test sensitivity to the pre-industrial reference state.
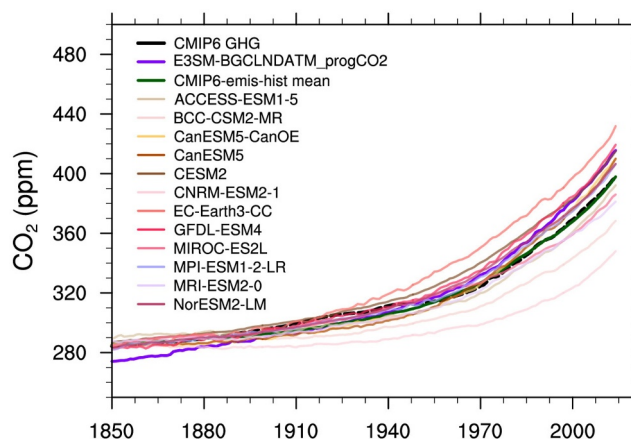

Figure B1. Global mean CO_2 time series in BGCLNDATM_progCO2 and emissions-driven historical simulations, alongside the prescribed CO_2 concentrations in CMIP6, for the period 1850–2014. CMIP6 emissions-driven historical simulations and historical GHG data are sourced from (Sanderson et al., 2023).

Appendix C: CMIP6 Modeled Ocean CO_2 Fluxes

This appendix compares air–sea carbon-flux estimates from six CMIP6 emission-driven historical simulations. These diagnostics illustrate both the long-term evolution of the global ocean sink (Figure C1) and its 1985–2014 spatial pattern (Figure C2), providing context for the prescribed ocean CO_2 fluxes used in the BGCLNDATM_progCO2 experiment.
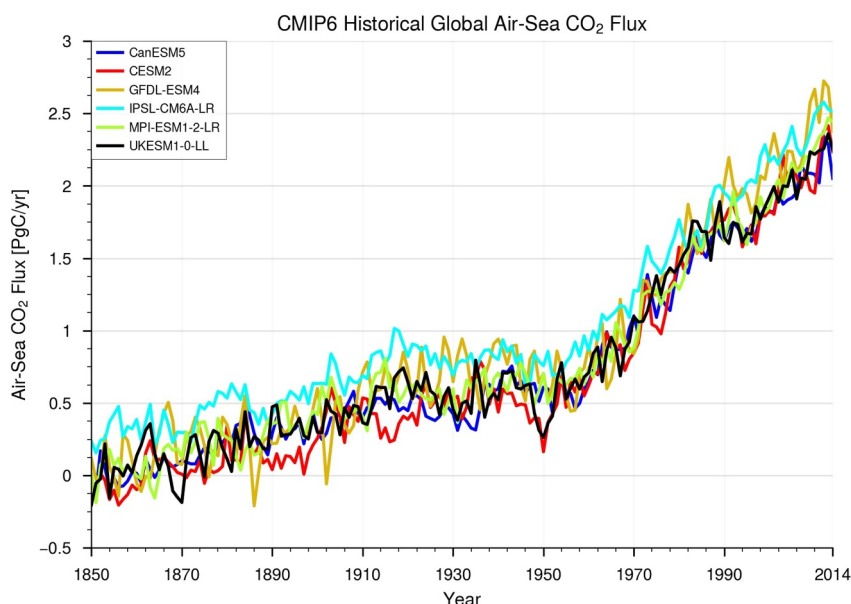
Figure C1. Globally integrated annual mean air-sea CO_2 flux from six CMIP6 historical simulations (CanESM5, CESM2, GFDL-ESM, IPSL-CM6A-LR, MPI-ESM1-2-LR, and UKESM1-0-LL) spanning 1850–2014. Positive values denote carbon uptake, whereas negative values denote outgassing.
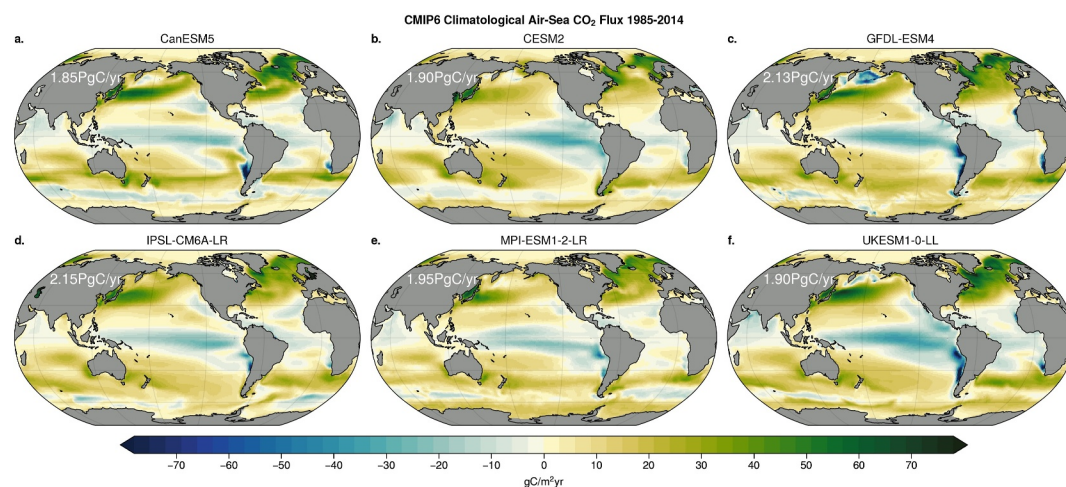

Figure C2. Map of the annual air-sea CO_2 flux climatology from six CMIP6 historical simulations: (a) CanESM5, (b) CESM2, (c) GFDL-ESM, (d) IPSL-CM6A-LR, (e) MPI-ESM1-2-LR, and (f) UKESM1-0-LL). The climatology is based on the 1985–2014 mean. Positive values denote carbon uptake, whereas negative values denote outgassing. The numbers in the top left corner of each panel represent the globally integrated annual air-sea CO_2 flux climatology.

Appendix D: ILAMB Observation Benchmarks

This appendix summarises the benchmarking diagnostics generated with ILAMB for key ecosystem and carbon-cycle variables. The figures compare the emissions-driven E3SM simulations (BGCLNDATM_progCO2 and the land-only BGCLND_presCO2) with a subset of CMIP6 esm-hist models and with state-of-the-art observational references. The collection provides context for model performance discussed in the main text and directs interested readers to the full interactive ILAMB pages, where additional maps, seasonal cycles, and statistical metrics can be explored.
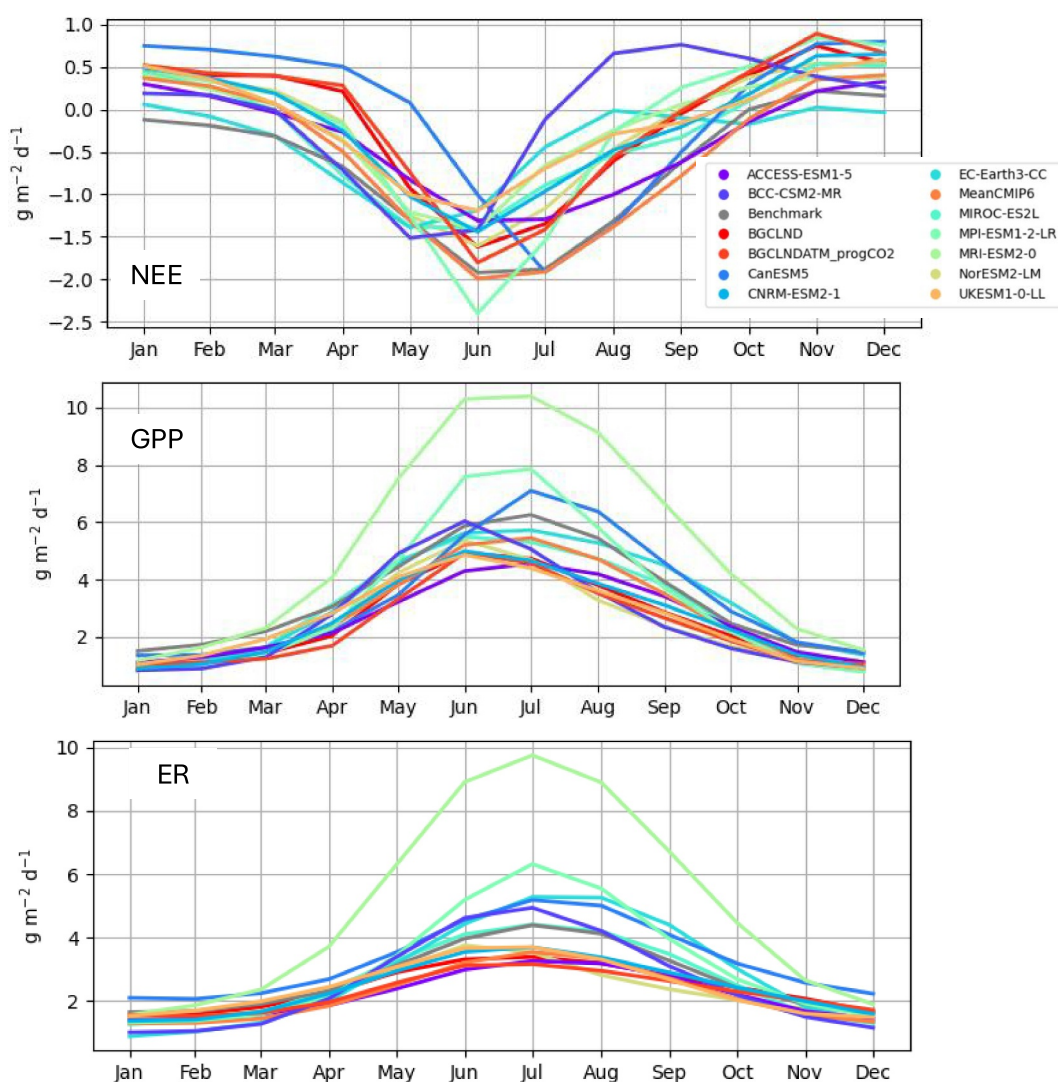
Figure D1. Climatological monthly means (1991–2015) of (top) net ecosystem exchange (NEE), (middle) gross primary productivity (GPP), and (bottom) ecosystem respiration (ER). Curves compare the emission-driven E3SM simulations (BGCLNDATM_progCO2, red; land-only BGCLND, dark red) with the observational benchmark (FLUXNET2015) and with a subset of CMIP6 esm-hist models. Positive NEE values indicate a net flux from land to atmosphere; negative values indicate net land uptake. NEE = ER − GPP. Addition analyses for NEE are available at https://www.ilamb.org/ELM/ilamb_e3sm_cbgcv2/e3sm/EcosystemandCarbonCycle/NetEcosystemExchange/FLUXNET2015/FLUXNET2015.html, GPP at https://www.ilamb.org/ELM/ilamb_e3sm_cbgcv2/e3sm/EcosystemandCarbonCycle/GrossPrimaryProductivity/FLUXNET2015/FLUXNET2015.html, ER at https://www.ilamb.org/ELM/ilamb_e3sm_cbgcv2/e3sm/EcosystemandCarbonCycle/EcosystemRespiration/FLUXNET2015/FLUXNET2015.html.
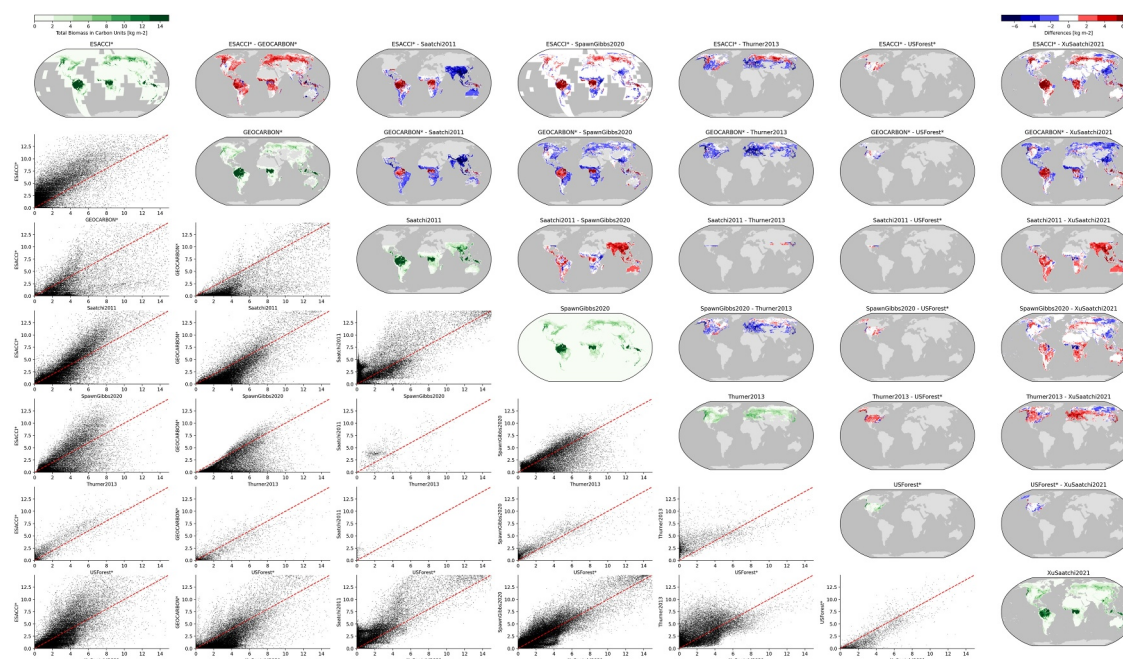
Figure D2. Comparison of **biomass** observation benchmarks in ILAMB on a one-to-one basis. Each point in the scattered plots represents the estimate from two observational benchmarks over the globe. All points deviate from the 1:1 line, indicating large uncertainty in biomass observational benchmarks. Notably, certain products report total biomass, encompassing both above and below-ground components, while others present biomass in terms of total mass rather than carbon content. Products marked with an asterisk (*) have been adjusted using a scaling factor of 0.5 to convert total mass to carbon and further adjusted by 0.7 to convert above-ground biomass estimates to total biomass estimates. A high-resolution version of this figure is available at https://web.lcrc.anl.gov/public/e3sm/diagnostic_output/ac.sfeng1/paper/CBGCv2/cVeg_adjusted.png.
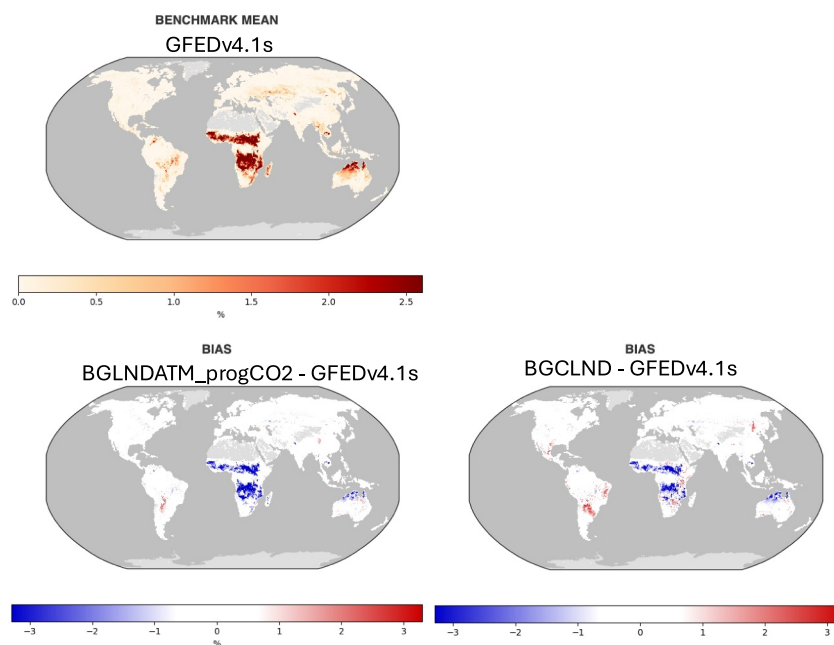

Figure D3. Global burned area evaluation from ILAMB. The observation benchmark is GFED v4.1s. Additional diagnostics for burned area are available at https://www.ilamb.org/ELM/ilamb_e3sm_cbgcv2/e3sm/EcosystemandCarbonCycle/BurnedArea/GFED4.1S/GFED4.1S.html.
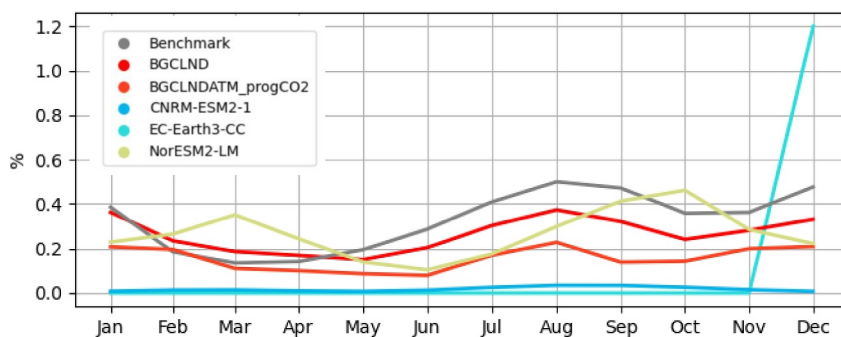
Figure D4. Comparison of climatological monthly means (1997–2017) of burned areas among BGCLNDATM_progCO2, BGCLND, a subset of CMIP6 esm-hist models, and the observational benchmark. Addition analyses for burned area are available at https://www.ilamb.org/ELM/ilamb_e3sm_cbgcv2/e3sm/EcosystemandCarbonCycle/BurnedArea/GFED4.1S/GFED4.1S.html.
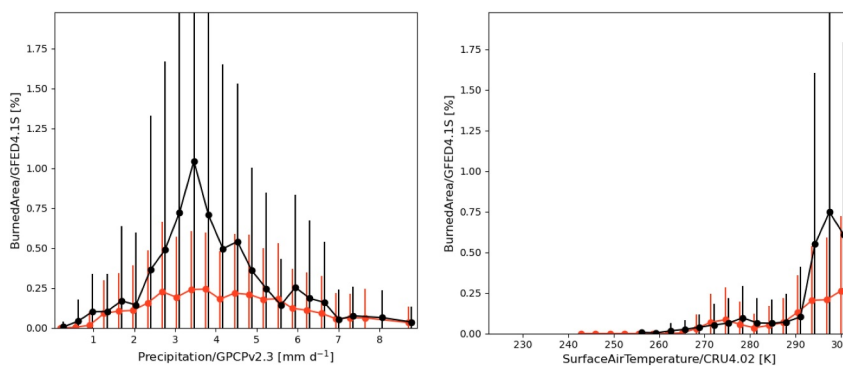

Figure D5. Functional responses of for burned areas and precipitation (left) and surface temperature (right) from ILAMB. More diagnostics in addition to reference distribution and model distribution are available at https://web.lcrc.anl.gov/public/e3sm/diagnostic_output/ac.sfeng1/paper/CBGCv2/e3sm-ilamb/EcosystemandCarbonCycle/BurnedArea/GFED4.1S/GFED4.1S.html#Relationships.

	Download Data	Period Mean (original grids) [1]	Model Period Mean (intersection) [1]	Benchmark Period Mean (intersection) [1]	Model Period Mean (complement) [1]	Benchmark Period Mean (complement) [1]	Bias [1]	RMSE [1]	Phase Shift [months]	Bias Score [1]	RMSE Score [1]	Seasonal Cycle Score [1]	Spatial Distribution Score [1]	Overall Score [1]
Benchmark	[-]	1.36												
ACCESS-ESM1-5	[-]	1.17	1.31	1.36	0.239	1.79	-0.0801	0.742	1.55	0.266	0.357	0.747	0.924	0.531
BCC-CSM2-MR	[-]	1.50	1.64	1.36	0.507	1.61	0.272	0.838	1.81	0.284	0.297	0.699	0.899	0.495
BGCLND	[-]	1.29	1.43	1.36	0.459	1.01	0.0734	0.726	1.89	0.299	0.344	0.632	0.897	0.503
BGCLNDATM_progCO2	[-]	1.41	1.59	1.36	0.407	1.01	0.228	0.956	2.03	0.231	0.323	0.628	0.774	0.456
CanESM5	[-]	1.35	1.46	1.36	0.664		0.141	0.854	1.92	0.235	0.317	0.719	0.879	0.493
CNRM-ESM2-1	[-]	1.23	1.33	1.36	0.533	1.65	0.000766	0.699	1.71	0.280	0.336	0.711	0.937	0.520
EC-Earth3-CC	[-]	1.59	1.98	1.36	0.111		0.613	1.20	1.70	0.208	0.326	0.675	0.830	0.473
MeanCMIP6	[-]	1.60	1.75	1.36	0.637		0.421	0.709	1.78	0.296	0.394	0.747	0.954	0.557
MIROC-ES2L	[-]	1.23	1.31	1.36	0.711	1.94	-0.00356	0.791	2.06	0.243	0.380	0.613	0.910	0.505
MPI-ESM1-2-LR	[-]	1.42	1.56	1.32	0.481	1.89	0.233	0.848	1.71	0.266	0.316	0.696	0.932	0.505
MRI-ESM2-0	[-]	2.33	2.53	1.36	0.932	1.98	1.21	1.71	1.99	0.183	0.307	0.574	0.517	0.378
NorESM2-LM	[-]	1.39	1.49	1.36	0.678		0.169	0.908	2.29	0.275	0.330	0.614	0.893	0.489
UKESM1-0-LL	[-]	2.46	2.43	1.38	2.97	0.472	1.10	1.30	1.83	0.179	0.420	0.731	0.864	0.523

Figure D6. Global leaf area index (LAI) evaluation from ILAMB. The observation benchmark is MODIS. Additional diagnostics for LAI are available at https://www.ilamb.org/ELM/ilamb_e3sm_cbgcv2/e3sm/EcosystemandCarbonCycle/LeafAreaIndex/MODIS/MODIS.html.
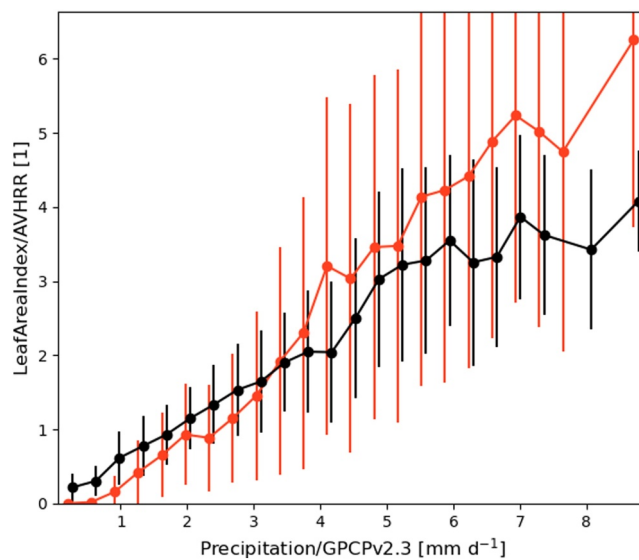

Figure D7. Same as Figure D5 but of leaf area index and precipitation from ILAMB, the reference in black and the model in navy blue. More diagnostics in addition to reference distribution and model distribution are available at https://web.lcrc.anl.gov/public/e3sm/diagnostic_output/ac.sfeng1/paper/CBGCv2/e3sm-ilamb/EcosystemandCarbonCycle/LeafAreaIndex/AVHRR/AVHRR.html#Relationships.

		Download Data Period Mean (original grids) [g m-2 d-1]	Bias [g m-2 d-1]	RMSE [g m-2 d-1]	Phase Shift [months]	Bias Score [1]	RMSE Score [1]	Seasonal Cycle Score [1]	Overall Score [1]
Benchmark	[-]	2.71							
ACCESS-ESM1-5	[-]	2.19	-0.521	1.90	0.982	0.529	0.455	0.847	0.572
BCC-CSM2-MR	[-]	2.56	-0.141	1.84	1.10	0.581	0.430	0.856	0.574
BGCLND	[-]	2.42	-0.279	1.89	0.802	0.524	0.444	0.896	0.577
BGCLNDATM_progCO2	[-]	2.27	-0.427	1.90	0.760	0.533	0.437	0.892	0.575
CanESM5	[-]	3.30	0.592	2.04	0.884	0.532	0.458	0.899	0.587
CNRM-ESM2-1	[-]	2.44	-0.269	1.82	1.14	0.559	0.420	0.845	0.561
EC-Earth3-CC	[-]	2.85	0.138	2.42	1.06	0.486	0.386	0.849	0.527
MeanCMIP6	[-]	2.19	-0.519	1.62	0.778	0.573	0.482	0.900	0.609
MIROC-ES2L	[-]	2.64	-0.0732	2.04	1.40	0.518	0.439	0.796	0.548
MPI-ESM1-2-LR	[-]	3.10	0.406	2.41	0.734	0.509	0.368	0.914	0.540
MRI-ESM2-0	[-]	4.95	2.24	4.59	0.916	0.300	0.195	0.896	0.397
NorESM2-LM	[-]	2.33	-0.384	1.98	0.982	0.514	0.418	0.885	0.559
UKESM1-0-LL	[-]	2.44	-0.260	2.01	1.12	0.540	0.404	0.857	0.551

Figure D8. Same as Figure D6 but for ecosystem respiration (ER). The observation benchmark is FLUXNET2015 (1991–2015). Additional diagnostics for ER are available at https://www.ilamb.org/ELM/ilamb_e3sm_cbgcv2/e3sm/EcosystemandCarbonCycle/EcosystemRespiration/FLUXNET2015/FLUXNET2015.html.

	Download Data	nbp(2016) [Pg]	nbp(2015) [Pg]	diff(2015) [Pg]	diff(2016) [Pg]	Difference Score [1]	Trajectory Score [1]	Temporal Distribution Score [1]	Overall Score [1]
Benchmark	[-]	-44.6							
ACCESS-ESM1-5	[-]		-58.9	-14.9		0.673	0.970	0.600	0.748
BGCLND	[-]		-57.3	-13.3		0.702	0.999	0.522	0.741
BGCLNDATM_progCO2	[-]		-42.1	1.92		0.950	1.00	0.726	0.892
CanESM5	[-]		-5.61	38.4		0.361	0.970	0.571	0.634
CNRM-ESM2-1	[-]		-97.7	-53.7		0.240	0.884	0.536	0.554
EC-Earth3-CC	[-]		5.19	49.2		0.271	0.893	0.533	0.566
MeanCMIP6	[-]	-34.2			10.3	0.760	1.00	0.536	0.765
MIROC-ES2L	[-]		-106.	-61.5		0.195	0.894	0.534	0.541
MPI-ESM1-2-LR	[-]		-28.7	15.3		0.666	0.984	0.673	0.774
MRI-ESM2-0	[-]		-79.5	-35.5		0.389	0.982	0.616	0.663
NorESM2-LM	[-]		-37.2	6.80		0.835	1.00	0.550	0.795
UKESM1-0-LL	[-]		-23.4	20.6		0.579	1.00	0.696	0.758

Figure D9. Global net ecosystem carbon balance (represented by net biome production (NBP)) from ILAMB. The observation benchmark is GCP for the period of 1959–2016. Additional diagnostics for NBP are available at https://www.ilamb.org/ELM/ilamb_e3sm_cbgcv2/e3sm/EcosystemandCarbonCycle/GlobalNetEcosystemCarbonBalance/GCP/GCP.html.

		Download Data Period Mean (original grids) [g m-2 d-1]	Bias [g m-2 d-1]	RMSE [g m-2 d-1]	Phase Shift [months]	Bias Score [1]	RMSE Score [1]	Seasonal Cycle Score [1]	Overall Score [1]
Benchmark	[-]	3.35							
ACCESS-ESM1-5	[-]	2.57	-0.793	2.64	1.08	0.574	0.459	0.867	0.590
BCC-CSM2-MR	[-]	2.62	-0.719	2.66	1.16	0.587	0.430	0.880	0.582
BGCLND	[-]	2.54	-0.800	2.73	1.01	0.540	0.450	0.875	0.579
BGCLNDATM_progCO2	[-]	2.35	-0.989	2.79	0.938	0.531	0.452	0.889	0.581
CanESM5	[-]	3.33	-0.0208	2.75	1.22	0.624	0.434	0.850	0.586
CNRM-ESM2-1	[-]	2.60	-0.754	2.57	1.16	0.589	0.441	0.849	0.580
EC-Earth3-CC	[-]	3.23	-0.126	2.95	1.14	0.590	0.431	0.852	0.576
MeanCMIP6	[-]	2.74	-0.612	2.33	0.965	0.622	0.492	0.879	0.621
MIROC-ES2L	[-]	2.93	-0.423	2.89	1.62	0.559	0.440	0.807	0.561
MPI-ESM1-2-LR	[-]	3.29	-0.0434	3.03	0.936	0.602	0.408	0.889	0.576
MRI-ESM2-0	[-]	5.12	1.77	4.73	0.942	0.419	0.291	0.873	0.469
NorESM2-LM	[-]	2.51	-0.841	2.76	1.17	0.540	0.456	0.867	0.580
UKESM1-0-LL	[-]	2.58	-0.761	2.79	1.13	0.554	0.429	0.878	0.573

Figure D10. Same as Figure D6 but for gross primary productivity (GPP). The observation benchmark is FLUXNET2015 (1991–2015). Additional diagnostics for GPP are available at https://www.ilamb.org/ELM/ilamb_e3sm_cbgcv2/e3sm/EcosystemandCarbonCycle/GrossPrimaryProductivity/FLUXNET2015/FLUXNET2015.html.
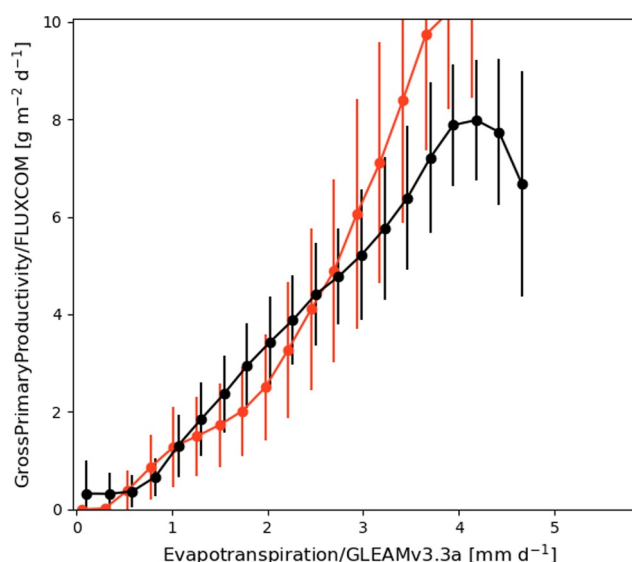
Figure D11. Same as Figure D5 but for gross primary productivity and ET from ILAMB. More diagnostics in addition to reference distribution and model distribution are available at https://web.lcrc.anl.gov/public/e3sm/diagnostic_output/ac.sfeng1/paper/CBGCv2/e3sm-ilamb/EcosystemandCarbonCycle/GrossPrimaryProductivity/FLUXCOM/FLUXCOM.html#Relationships.
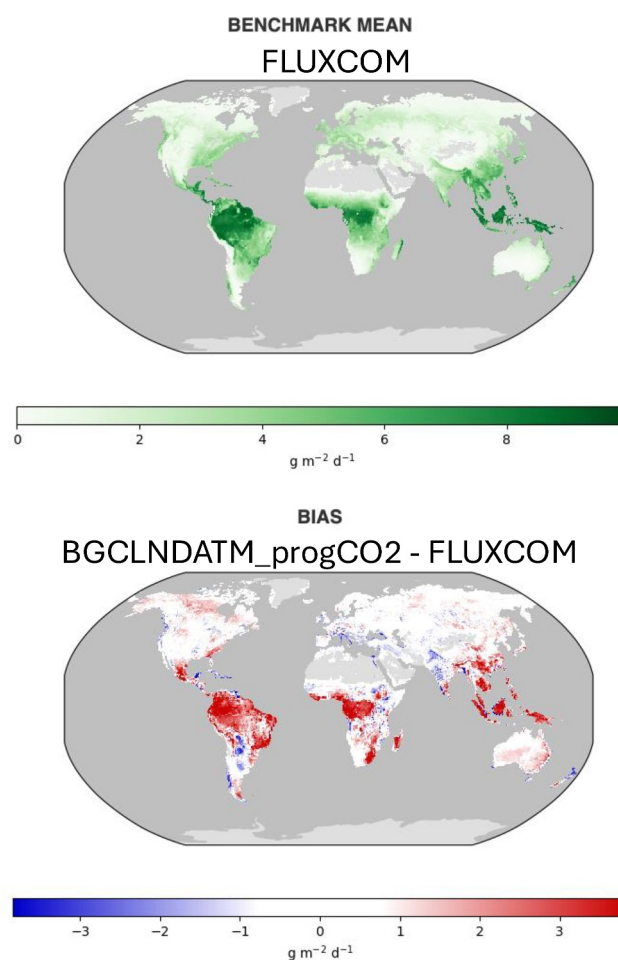
Figure D12. Global gross primary productivity (GPP) evaluation from ILAMB. The observation benchmark is FLUXCOM for the period of 1980–2014. Additional diagnostics for GPP are available at https://www.ilamb.org/ELM/ilamb_e3sm_cbgcv2/e3sm/EcosystemandCarbonCycle/GrossPrimaryProductivity/FLUXCOM/FLUXCOM.html.

		Download Data Period Mean (original grids) [Pg]	Model Period Mean (intersection) [Pg]	Benchmark Period Mean (intersection) [Pg]	Model Period Mean (complement) [Pg]	Benchmark Period Mean (complement) [Pg]	Bias [kg m-2]	Bias Score [1]	Spatial Distribution Score [1]	Overall Score [1]
Benchmark	[-]	1.33e+03								
ACCESS-ESM1-5	[-]	864.	857.	1.32e+03	6.75	18.5	-3.08	0.610	0.773	0.691
BCC-CSM2-MR	[-]	1.53e+03	1.51e+03	1.31e+03	16.6	20.9	1.83	0.535	0.444	0.489
BGCLND	[-]	1.20e+03	1.20e+03	1.31e+03	1.36	21.7	-0.812	0.596	0.747	0.671
BGCLNDATM_progCO2	[-]	1.14e+03	1.14e+03	1.31e+03	1.33	21.7	-1.26	0.576	0.844	0.710
CanESM5	[-]	1.36e+03	1.31e+03	1.33e+03	33.1		1.11	0.613	0.869	0.741
CNRM-ESM2-1	[-]	1.56e+03	1.53e+03	1.33e+03	22.3	4.45	2.39	0.614	0.848	0.731
EC-Earth3-CC	[-]	1.31e+03	1.31e+03	1.33e+03	2.75		-0.181	0.596	0.781	0.688
MeanCMIP6	[-]	1.43e+03	1.41e+03	1.33e+03	23.2		1.39	0.610	0.864	0.737
MIROC-ES2L	[-]	1.17e+03	1.12e+03	1.33e+03	30.9	3.30	-0.879	0.556	0.729	0.643
MPI-ESM1-2-LR	[-]	708.	695.	1.19e+03	10.2	147.	-3.82	0.626	0.581	0.604
MRI-ESM2-0	[-]	1.07e+03	1.05e+03	1.31e+03	12.8	19.3	-1.23	0.570	0.839	0.704
NorESM2-LM	[-]	939.	879.	1.33e+03	49.2		-2.21	0.593	0.320	0.456
UKESM1-0-LL	[-]	1.76e+03	1.73e+03	1.31e+03	25.7	23.0	4.33	0.497	0.713	0.605

Figure D13. Global soil carbon evaluation from ILAMB. The observation benchmark is HWSD for the period of 2000–2001. Additional diagnostics for soil carbon are available at https://www.ilamb.org/ELM/ilamb_e3sm_cbgcv2/e3sm/EcosystemandCarbonCycle/SoilCarbon/HWSD/HWSD.html.

Figure D14. Nitrogen Fixation evaluation from ILAMB. The observation benchmark is Davies-Barnard for the period of 1980–2020. Additional diagnostics for soil are available at https://www.ilamb.org/ELM/ilamb_e3sm_cbgcv2/e3sm/EcosystemandCarbonCycle/NitrogenFixation/Davies-Barnard/Davies-Barnard.html.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

All model codes may be accessed on the GitHub repository at https://github.com/E3SM-Project/E3SM (E3SM Project, DOE, 2023). A maintenance branch (maint-2.1; https://github.com/E3SM-Project/E3SM/tree/v2.1.0; E3SM Project, DOE, 2023) has been specifically created to reproduce these simulations. All model-source modifications required to implement BGCLNDATM_progCO2 on top of the E3SMv1.1 (Burrows et al., 2020) are available on GitHub at https://github.com/E3SM-Project/E3SM/compare/master…sfeng/cpl/BPRP-lnd-atm. Complete native model output for BGCLNDATM_progCO2 and BGCLND is accessible directly on NERSC HPSS at https://docs.e3sm.org/e3sm_data_docs/_build/html/v2.1/BGC, LNDATM_presCO2 is at https://docs.e3sm.org/e3sm_data_docs/_build/html/v2.1/WaterCycle/simulation_data/simulation_table.html.

Input4MIPs project: https://pcmdi.llnl.gov/mips/input4MIPs/.

ESGF: https://esgf-node.ipsl.upmc.fr/projects/cmip6-ipsl/.

Global Soil Wetness Project Phase 3 (GSWP3) reanalysis data: https://www.isimip.org/gettingstarted/input-data-bias-adjustment/details/4/ (Dirmeyer et al., 2006).

Zppy: https://github.com/E3SM-Project/zppy.

E3SM Diags: https://github.com/E3SM-Project/e3sm_diags (Zhang et al., 2022).

ILAMB: https://github.com/rubisco-sfa/ILAMB (Collier et al., 2018).

GDESS: https://github.com/E3SM-Project/gdess (Kaufman et al., 2022).

The full diagnostic results for atmosphere spin-up simulations: https://web.lcrc.anl.gov/public/e3sm/diagnostic_output/ac.sfeng1/paper/CBGCv2/global_time_series/global_time_series_0001-0500_results/.

The full diagnostic results for land spin-up simulations: https://web.lcrc.anl.gov/public/e3sm/diagnostic_output/ac.sfeng1/paper/CBGCv2/lineplots_spinup/.

The full diagnostic results for the historical transition simulation from the E3SM_Diags: https://web.lcrc.anl.gov/public/e3sm/diagnostic_output/ac.sfeng1/paper/CBGCv2/e3sm_diags/.

The full, interactive analysis of the BGCLNDATM_progCO2 results using the ILAMB diagnostic tool, and comparison with selected CMIP6 models and BGCLND, is available online at https://doi.org/10.25584/2525948 (Feng et al., 2025).

References

Adler, R. F., Sapiano, M. R. P., Huffman, G. J., Wang, J.-J., Gu, G., Bolvin, D., et al. (2018). The Global Precipitation Climatology Project (GPCP) Monthly Analysis (New Version 2.3) and a Review of 2017 Global Precipitation. *Atmosphere*, *9*(4), 138. https://doi.org/10.3390/atmos9040138

Araza, A., de Bruin, S., Herold, M., Quegan, S., Labriere, N., Rodriguez-Veiga, P., et al. (2022). A comprehensive framework for assessing the accuracy and uncertainty of global above-ground biomass maps. *Remote Sensing of Environment*, *272*, 112917. https://doi.org/10.1016/j.rse.2022.112917

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., et al. (2020). Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models. *Biogeosciences*, *17*(16), 4173–4222. https://doi.org/10.5194/bg-17-4173-2020

Bock, L., Lauer, A., Schlund, M., Barreiro, M., Bellouin, N., Jones, C., et al. (2020). Quantifying progress across different CMIP phases with the ESMValTool. *Journal of Geophysical Research: Atmospheres*, *125*(21), e2019JD032321. https://doi.org/10.1029/2019JD032321

Bonan, G. B., & Doney, S. C. (2018). Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models. *Science*, *359*(6375), eaam8328. https://doi.org/10.1126/science.aam8328

Burrows, S. M., Maltrud, M., Yang, X., Zhu, Q., Jeffery, N., Shi, X., et al. (2020). The DOE E3SM v1.1 Biogeochemistry configuration: Description and simulated ecosystem-climate responses to historical changes in forcing. *Journal of Advances in Modeling Earth Systems*, *12*(9), e2019MS001766. https://doi.org/10.1029/2019MS001766

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., et al. (2018). The international land model benchmarking (ILAMB) System: Design, theory, and implementation. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2731–2754. https://doi.org/10.1029/2018MS001354

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., et al. (2013). Long-term climate change: Projections, commitments and irreversibility. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Doschung, et al. (Eds.), *Climate change 2013: The physical science Basis. Contribution of Working Group I to the fifth assessment report of the intergovernmental Panel on climate change* (pp. 1029–1136). Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.024

Collins, W. D., Craig, A. P., Truesdale, J. E., Di Vittorio, A. V., Jones, A. D., Bond-Lamberty, B., et al. (2015). The integrated Earth system model version 1: Formulation and functionality. *Geoscientific Model Development*, *8*(7), 2203–2219. https://doi.org/10.5194/gmd-8-2203-2015

Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., & Totterdell, I. J. (2000). Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, *408*(6809), 184–187. https://doi.org/10.1038/35041539

Davies-Barnard, T., & Friedlingstein, P. (2020). The global distribution of biological nitrogen fixation in terrestrial natural ecosystems. *Global Biogeochemical Cycles*, *34*(3), e2019GB006387. https://doi.org/10.1029/2019GB006387

Dirmeyer, P. A. (2011). A history and review of the Global Soil Wetness Project (GSWP). *Journal of Hydrometeorology*, *12*(5), 729–749. https://doi.org/10.1175/jhm-d-10-05010.1

Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., & Hanasaki, N. (2006). GSWP-2: Multimodel analysis and implications for our perception of the land surface. *Bulletin of the American Meteorological Society*, *87*(10), 1381–1398. https://doi.org/10.1175/BAMS-87-10-1381

Di Vittorio, A. V., Sinha, E., Hao, D., Singh, B., Calvin, K. V., Shippert, T., et al. (2025). E3SM-GCAM: A synchronously coupled human component in the E3SM Earth System model enables novel human-earth feedback research. *Journal of Advances in Modeling Earth Systems*, *17*(6), e2024MS004806. https://doi.org/10.1029/2024MS004806

Dolman, A. J., Shvidenko, A., Schepaschenko, D., Ciais, P., Tchebakova, N., Chen, T., et al. (2012). An estimate of the terrestrial carbon budget of Russia using inventory-based, eddy covariance and inversion methods. *Biogeosciences*, *9*(12), 5323–5340. https://doi.org/10.5194/bg-9-5323-2012

Du, E., de Vries, W., Collalti, A., & De Marco, A. (2025). Climate warming alters nutrient cycling and its constraint on CO_2 fertilization in global forests. *Current Climate Change Reports*, *11*(1), 3. https://doi.org/10.1007/s40641-025-00201-6

E3SM Project, DOE. (2023). *Energy exascale Earth System model v2.1.0*. Computer Software. https://doi.org/10.11578/E3SM/dc.20230110.5

Ellsworth, D. S., Crous, K. Y., De Kauwe, M. G., Verryckt, L. T., Goll, D., Zaehle, S., et al. (2022). Convergence in phosphorus constraints to photosynthesis in forests around the world. *Nature Communications*, *13*(1), 5005. https://doi.org/10.1038/s41467-022-32545-0

Eyring, V., Bock, L., Lauer, A., Righi, M., Schlund, M., Andela, B., et al. (2020). Earth System Model Evaluation Tool (ESMValTool) v2.0—An extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP. *Geoscientific Model Development*, *13*(7), 3383–3438. https://doi.org/10.5194/gmd-13-3383-2020

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Fang, Y., & Leung, L. R. (2022). Relative controls of vapor pressure deficit and soil water stress on canopy conductance in global simulations by an Earth System model. *Earth's Future*, *10*(9), e2022EF002810. https://doi.org/10.1029/2022EF002810

Feng, S., Collier, N., Zhu, Q., Ricciuto, D., Bond-Lamberty, B., & Leung, L.-Y. (2025). ILAMBv2.7 benchmarking results comparing E3SMv2.1 land-atmosphere coupled (BGCv2LNDATM) and stand alone land (ELM) simulations with CMIP6 emission driven historical simulations [Dataset]. *Pacific Northwest National Laboratory*, *2*. https://doi.org/10.25584/2525948

Flato, G. M. (2011). Earth system models: An overview. *WIREs Climate Change*, *2*(6), 783–800. https://doi.org/10.1002/wcc.148

Forkel, M., Wessollek, C., Huijnen, V., Andela, N., de Laat, A., Kinalczyk, D., et al. (2025). Burning of woody debris dominates fire emissions in the Amazon and Cerrado. *Nature Geoscience*, *18*(2), 140–147. https://doi.org/10.1038/s41561-024-01637-5

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., Bloh, W. V., Brovkin, V., et al. (2006). Climate–Carbon cycle feedback analysis: Results from the C4MIP model intercomparison. *Journal of Climate*, *19*(14), 3337–3353. https://doi.org/10.1175/JCLI3800.1

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Bakker, D. C. E., Hauck, J., et al. (2021). Global Carbon Budget 2021. *Earth System Science Data Discussions*, 1–191. https://doi.org/10.5194/essd-2021-386

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., & Knutti, R. (2014). Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate*, *27*(2), 511–526. https://doi.org/10.1175/JCLI-D-12-00579.1

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Landschützer, P., et al. (2024). Global Carbon Budget 2024. *Earth System Science Data Discussions*, 1–133. https://doi.org/10.5194/essd-2024-519

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., et al. (2020). Global Carbon Budget 2020. *Earth System Science Data*, *12*(4), 3269–3340. https://doi.org/10.5194/essd-12-3269-2020

Gettelman, A., & Morrison, H. (2015). Advanced two-moment bulk microphysics for global models. Part I: Off-line tests and comparison with other schemes. *Journal of Climate*, *28*(3), 1268–1287. https://doi.org/10.1175/JCLI-D-14-00102.1

Gier, B. K., Buchwitz, M., Reuter, M., Cox, P. M., Friedlingstein, P., & Eyring, V. (2020). Spatially resolved evaluation of Earth system models with satellite column-averaged CO_2. *Biogeosciences*, *17*(23), 6115–6144. https://doi.org/10.5194/bg-17-6115-2020

Gier, B. K., Schlund, M., Friedlingstein, P., Jones, C. D., Jones, C., Zaehle, S., & Eyring, V. (2024). Representation of the terrestrial carbon cycle in CMIP6. *Biogeosciences*, *21*(22), 5321–5360. https://doi.org/10.5194/bg-21-5321-2024

Giglio, L., Randerson, J. T., & van der Werf, G. R. (2013). Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). *Journal of Geophysical Research: Biogeosciences*, *118*(1), 317–328. https://doi.org/10.1002/jgrg.20042

Gillett, N. P., Shiogama, H., Funke, B., Hegerl, G., Knutti, R., Matthes, K., et al. (2016). The detection and attribution model Intercomparison Project (DAMIP v1.0) contribution to CMIP6. *Geoscientific Model Development*, *9*(10), 3685–3697. https://doi.org/10.5194/gmd-9-3685-2016

Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research*, *113*(D6), D06104. https://doi.org/10.1029/2007JD008972

Golaz, J.-C., Larson, V. E., & Cotton, W. R. (2002). A PDF-Based model for boundary layer clouds. *Part I: Method and Model Description*. Retrieved from https://journals.ametsoc.org/view/journals/atsc/59/24/1520-0469_2002_059_3540_apbmfb_2.0.co_2.xml

Golaz, J.-C., Van Roekel, L. P., Zheng, X., Roberts, A. F., Wolfe, J. D., Lin, W., et al. (2022). The DOE E3SM model version 2: Overview of the physical model and initial model evaluation. *Journal of Advances in Modeling Earth Systems*, *14*(12), e2022MS003156. https://doi.org/10.1029/2022MS003156

Hajima, T., Kawamiya, M., Ito, A., Tachiiri, K., Jones, C., Arora, V., et al. (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* (pp. 1–49). EGUsphere. https://doi.org/10.5194/egusphere-2024-188

Hajima, T., Kawamiya, M., Ito, A., Tachiiri, K., Jones, C. D., Arora, V., et al. (2025). Consistency of global carbon budget between concentration- and emission-driven historical experiments simulated by CMIP6 Earth system models and suggestions for improved simulation of CO2 concentration. *Biogeosciences*, *22*(5), 1447–1473. https://doi.org/10.5194/bg-22-1447-2025

Hannah, W. M., Bradley, A. M., Guba, O., Tang, Q., Golaz, J.-C., & Wolfe, J. (2021). Separating physics and dynamics grids for improved computational efficiency in spectral element Earth System models. *Journal of Advances in Modeling Earth Systems*, *13*(7), e2020MS002419. https://doi.org/10.1029/2020MS002419

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations – The CRU TS3.10 Dataset. *International Journal of Climatology*, *34*(3), 623–642. https://doi.org/10.1002/joc.3711

He, X., Hui, D., Wang, F., Deng, Q., Liu, Z., Lu, H., et al. (2025). Dynamics of soil and foliar phosphorus fractions in a secondary tropical forest under altered seasonal precipitation patterns. *Plant and Soil*, *507*(1), 915–937. https://doi.org/10.1007/s11104-024-06774-9

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hewitt, A. J., Booth, B. B. B., Jones, C. D., Robertson, E. S., Wiltshire, A. J., Sansom, P. G., et al. (2016). Sources of uncertainty in future projections of the carbon cycle. *Journal of Climate*, *29*(20), 7203–7213. https://doi.org/10.1175/JCLI-D-16-0161.1

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., et al. (2018). Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). *Geoscientific Model Development*, *11*(1), 369–408. https://doi.org/10.5194/gmd-11-369-2018

Hoffman, F. M., Randerson, J. T., Arora, V. K., Bao, Q., Cadule, P., Ji, D., et al. (2014). Causes and implications of persistent atmospheric carbon dioxide biases in Earth System Models. *Journal of Geophysical Research: Biogeosciences*, *119*(2), 141–162. https://doi.org/10.1002/2013JG002381

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., et al. (2013). The community Earth System model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, *94*(9), 1339–1360. https://doi.org/10.1175/BAMS-D-12-00121.1

IPCC. (2021). Climate change 2021: The physical science basis. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, et al. (Eds.), *Contribution of Working Group I to the sixth assessment report of the intergovernmental Panel on climate change*. Cambridge University Press. https://doi.org/10.1017/9781009157896

Ito, A., Hajima, T., Lawrence, D. M., Brovkin, V., Delire, C., Guenet, B., et al. (2020). Soil carbon sequestration simulated in CMIP6-LUMIP models: Implications for climatic mitigation. *Environmental Research Letters*, *15*(12), 124061. https://doi.org/10.1088/1748-9326/abc912

Jeffery, N., Maltrud, M. E., Hunke, E. C., Wang, S., Wolfe, J., Turner, A. K., et al. (2020). Investigating controls on sea ice algal production using E3SMv1.1-BGC. *Annals of Glaciology*, *61*(82), 51–72. https://doi.org/10.1017/aog.2020.7

Jöckel, P. (2006). Technical note: Recursive rediscretisation of geo-scientific data in the Modular Earth Submodel System (MESSy). *Atmospheric Chemistry and Physics*, *6*(11), 3557–3562. https://doi.org/10.5194/acp-6-3557-2006

Jones, A. D., Calvin, K. V., Shi, X., Di Vittorio, A. V., Bond-Lamberty, B., Thornton, P. E., & Collins, W. D. (2018). Quantifying human-mediated carbon cycle feedbacks. *Geophysical Research Letters*, *45*(20), 11370–11379. https://doi.org/10.1029/2018GL079350

Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., et al. (2016). C4MIP – The coupled climate–carbon cycle model Intercomparison Project: Experimental protocol for CMIP6. *Geoscientific Model Development*, *9*(8), 2853–2880. https://doi.org/10.5194/gmd-9-2853-2016

Jones, C. G., Adloff, F., Booth, B. B. B., Cox, P. M., Eyring, V., Friedlingstein, P., et al. (2024). Bringing it all together: Science priorities for improved understanding of Earth system change and to support international climate policy. *Earth System Dynamics*, *15*(5), 1319–1351. https://doi.org/10.5194/esd-15-1319-2024

Kaufman, D. E., Feng, S., Calvin, K. V., Harrop, B. E., & Burrows, S. M. (2022). Gdess: A framework for evaluating simulated atmospheric $CO_2\sim$ in Earth System Models. *Journal of Open Source Software*, *7*(76), 4326. https://doi.org/10.21105/joss.04326

Knauer, J., Zaehle, S., Medlyn, B. E., Reichstein, M., Williams, C. A., Migliavacca, M., et al. (2018). Towards physiologically meaningful water-use efficiency estimates from eddy covariance data. *Global Change Biology*, *24*(2), 694–710. https://doi.org/10.1111/gcb.13893

Koven, C. D., Arora, V. K., Cadule, P., Fisher, R. A., Jones, C. D., Lawrence, D. M., et al. (2022). Multi-century dynamics of the climate and carbon cycle under both high and net negative emissions scenarios. *Earth System Dynamics*, *13*(2), 885–909. https://doi.org/10.5194/esd-13-885-2022

Larson, V. E. (2022). CLUBB-SILHS: A parameterization of subgrid variability in the atmosphere. *arXiv*. https://doi.org/10.48550/arXiv.1711.03675

Lasslop, G., Reichstein, M., Papale, D., Richardson, A. D., Arneth, A., Barr, A., et al. (2010). Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: Critical issues and global evaluation. *Global Change Biology*, *16*(1), 187–208. https://doi.org/10.1111/j.1365-2486.2009.02041.x

Le Quéré, C., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I., Peters, G. P., et al. (2016). Global Carbon Budget 2016. *Earth System Science Data*, *8*(2), 605–649. https://doi.org/10.5194/essd-8-605-2016

Li, F., Song, X., Harrison, S. P., Marlon, J. R., Lin, Z., Leung, L. R., et al. (2024). Evaluation of global fire simulations in CMIP6 Earth system models. *Geoscientific Model Development*, *17*(23), 8751–8771. https://doi.org/10.5194/gmd-17-8751-2024

Liddicoat, S. K., Wiltshire, A. J., Jones, C. D., Arora, V. K., Brovkin, V., Cadule, P., et al. (2021). Compatible fossil fuel CO_2 emissions in the CMIP6 Earth System models' historical and shared socioeconomic pathway experiments of the twenty-first century. *Journal of Climate*, *34*(8), 2853–2875. https://doi.org/10.1175/JCLI-D-19-0991.1

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., et al. (2018). Clouds and the Earth's radiant energy System (CERES) Energy Balanced and Filled (EBAF) top-of-atmosphere (TOA) Edition-4.0 data product. *Journal of Climate*, *31*(2), 895–918. https://doi.org/10.1175/JCLI-D-17-0208.1

Long, M. C., Moore, J. K., Lindsay, K., Levy, M., Doney, S. C., Luo, J. Y., et al. (2021). Simulations with the Marine Biogeochemistry Library (MARBL). *Journal of Advances in Modeling Earth Systems*, *13*(12), e2021MS002647. https://doi.org/10.1029/2021MS002647

Mannschatz, T., Pflug, B., Borg, E., Feger, K.-H., & Dietrich, P. (2014). Uncertainties of LAI estimation from satellite imaging due to atmospheric correction. *Remote Sensing of Environment*, *153*, 24–39. https://doi.org/10.1016/j.rse.2014.07.020

Masarie, K. A., Peters, W., Jacobson, A. R., & Tans, P. P. (2014). ObsPack: A framework for the preparation, delivery, and attribution of atmospheric greenhouse gas measurements. *Earth System Science Data*, *6*(2), 375–384. https://doi.org/10.5194/essd-6-375-2014

McBride, L. A., Hope, A. P., Canty, T. P., Bennett, B. F., Tribett, W. R., & Salawitch, R. J. (2021). Comparison of CMIP6 historical climate simulations and future projected warming to an empirical model of global climate. *Earth System Dynamics*, *12*(2), 545–579. https://doi.org/10.5194/esd-12-545-2021

Nissen, C., Lovenduski, N. S., Maltrud, M., Gray, A. R., Takano, Y., Falcinelli, K., et al. (2024). LIGHT-bgcArgo-1.0: Using synthetic float capabilities in E3SMv2 to assess spatiotemporal variability in ocean physics and biogeochemistry. *Geoscientific Model Development*, *17*(16), 6415–6435. https://doi.org/10.5194/gmd-17-6415-2024

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The scenario model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

Piao, S., Liu, Z., Wang, Y., Ciais, P., Yao, Y., Peng, S., et al. (2018). On the causes of trends in the seasonal amplitude of atmospheric CO2. *Global Change Biology*, *24*(2), 608–616. https://doi.org/10.1111/gcb.13909

Quesada, B., Arneth, A., Robertson, E., & de Noblet-Ducoudré, N. (2018). Potential strong contribution of future anthropogenic land-use and land-cover change to the terrestrial carbon cycle. *Environmental Research Letters*, *13*(6), 064023. https://doi.org/10.1088/1748-9326/aac4c3

Randerson, J. T., Chen, Y., van der Werf, G. R., Rogers, B. M., & Morton, D. C. (2012). Global burned area and biomass burning emissions from small fires. *Journal of Geophysical Research*, *117*(G4), G04012. https://doi.org/10.1029/2012JG002128

Randerson, J. T., Liu, H., Flanner, M. G., Chambers, S. D., Jin, Y., Hess, P. G., et al. (2006). The impact of boreal forest fire on climate warming. *Science*, *314*(5802), 1130–1132. https://doi.org/10.1126/science.1132075

Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M. D., Seneviratne, S. I., et al. (2013). Climate extremes and the carbon cycle. *Nature*, *500*(7462), 287–295. https://doi.org/10.1038/nature12350

Reichstein, M., Papale, D., Valentini, R., Aubinet, M., Bernhofer, C., Knohl, A., et al. (2007). Determinants of terrestrial ecosystem carbon balance inferred from European eddy covariance flux sites. *Geophysical Research Letters*, *34*(1), L01402. https://doi.org/10.1029/2006GL027880

Ricciuto, D., Sargsyan, K., & Thornton, P. (2018). The impact of parametric uncertainties on Biogeochemistry in the E3SM land model. *Journal of Advances in Modeling Earth Systems*, *10*(2), 297–319. https://doi.org/10.1002/2017MS000962

Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., Schlund, M., et al. (2020). Earth System Model Evaluation Tool (ESMValTool) v2.0—Technical overview. *Geoscientific Model Development*, *13*(3), 1179–1199. https://doi.org/10.5194/gmd-13-1179-2020

Sanderson, B. M., Booth, B. B. B., Dunne, J., Eyring, V., Fisher, R. A., Friedlingstein, P., et al. (2023). *The need for carbon emissions-driven climate projections in CMIP7* (pp. 1–51). EGUsphere. https://doi.org/10.5194/egusphere-2023-2127

Sanderson, B. M., Booth, B. B. B., Dunne, J., Eyring, V., Fisher, R. A., Friedlingstein, P., et al. (2024). The need for carbon-emissions-driven climate projections in CMIP7. *Geoscientific Model Development*, *17*(22), 8141–8172. https://doi.org/10.5194/gmd-17-8141-2024

Seiler, W., & Crutzen, P. J. (1980). Estimates of gross and net fluxes of carbon between the biosphere and the atmosphere from biomass burning. *Climatic Change*, *2*(3), 207–247. https://doi.org/10.1007/BF00137988

Sinha, E., Bond-Lamberty, B., Calvin, K. V., Drewniak, B. A., Bisht, G., Bernacchi, C., et al. (2023). The impact of crop rotation and spatially varying crop parameters in the E3SM land model (ELMv2). *Journal of Geophysical Research: Biogeosciences*, *128*(3), e2022JG007187. https://doi.org/10.1029/2022JG007187

Sinha, E., Calvin, K. V., Bond-Lamberty, B., Drewniak, B. A., Ricciuto, D. M., Sargsyan, K., et al. (2023). Modeling perennial bioenergy crops in the E3SM land model (ELMv2). *Journal of Advances in Modeling Earth Systems*, *15*(1), e2022MS003171. https://doi.org/10.1029/2022MS003171

Smith, K., Barthel, A. M., Conlon, L. M., Van Roekel, L. P., Bartoletti, A., Golez, J.-C., et al. (2024). The DOE E3SM version 2.1: Overview and assessment of the impacts of Parameterized Ocean submesoscales. *Geoscientific Model Development Discussions*, 1–38. https://doi.org/10.5194/gmd-2024-149

Spafford, L., & MacDougall, A. H. (2021). Validation of terrestrial biogeochemistry in CMIP6 Earth system models: A review. *Geoscientific Model Development*, *14*(9), 5863–5889. https://doi.org/10.5194/gmd-14-5863-2021

Takano, Y., Maltrud, M., Sinha, A., Jeffery, N., Smith, K., Conlon, L., et al. (2023). Global Ocean Carbon Cycle Simulations with the 2 E3SM version 2 (E3SMv2). https://doi.org/10.5281/zenodo.10093369

Tan, Z., Leung, L. R., Li, H.-Y., & Tesfa, T. (2018). Modeling sediment yield in land surface and Earth System models: Model comparison, development, and evaluation. *Journal of Advances in Modeling Earth Systems*, *10*(9), 2192–2213. https://doi.org/10.1029/2017MS001270

Tang, Q., Prather, M. J., Hsu, J., Ruiz, D. J., Cameron-Smith, P. J., Xie, S., & Golaz, J.-C. (2021). Evaluation of the interactive stratospheric ozone (O3v2) module in the E3SM version 1 Earth system model. *Geoscientific Model Development*, *14*(3), 1219–1236. https://doi.org/10.5194/gmd-14-1219-2021

Taylor, K. E., Williamson, D., & Zwiers, F. (2000). The Sea surface temperature and sea-ice concentration boundary conditions for AMIP II simulations. Retrieved from https://pcmdi.llnl.gov/report/ab60.html

Taylor, M. A., Guba, O., Steyer, A., Ullrich, P. A., Hall, D. M., & Eldred, C. (2020). An energy consistent discretization of the nonhydrostatic equations in primitive variables. *Journal of Advances in Modeling Earth Systems*, *12*(1), e2019MS001783. https://doi.org/10.1029/2019MS001783

Terrer, C., Jackson, R. B., Prentice, I. C., Keenan, T. F., Kaiser, C., Vicca, S., et al. (2019). Nitrogen and phosphorus constrain the CO_2 fertilization of global plant biomass. *Nature Climate Change*, *9*(9), 684–689. https://doi.org/10.1038/s41558-019-0545-2

Tesfa, T. K., Leung, L. R., Thornton, P. E., Brunke, M. A., & Duan, Z. (2024). Impacts of topography-based subgrid scheme and downscaling of atmospheric forcing on modeling land surface processes in the conterminous US. *Journal of Advances in Modeling Earth Systems*, *16*(8), e2023MS004064. https://doi.org/10.1029/2023MS004064

Thornton, P. E., Calvin, K., Jones, A. D., Di Vittorio, A. V., Bond-Lamberty, B., Chini, L., et al. (2017). Biospheric feedback effects in a synchronously coupled model of human and Earth systems. *Nature Climate Change*, *7*(7), 496–500. https://doi.org/10.1038/nclimate3310

Thornton, P. E., Doney, S. C., Lindsay, K., Moore, J. K., Mahowald, N., Randerson, J. T., et al. (2009). Carbon-nitrogen interactions regulate climate-carbon cycle feedbacks: Results from an atmosphere-ocean general circulation model. *Biogeosciences*, *6*(10), 2099–2120. https://doi.org/10.5194/bg-6-2099-2009

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., et al. (2010). Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). *Atmospheric Chemistry and Physics*, *10*(23), 11707–11735. https://doi.org/10.5194/acp-10-11707-2010

Vichi, M., Manzini, E., Fogli, P. G., Alessandri, A., Patara, L., Scoccimarro, E., et al. (2011). Global and regional ocean carbon uptake and climate change: Sensitivity to a substantial mitigation scenario. *Climate Dynamics*, *37*(9), 1929–1947. https://doi.org/10.1007/s00382-011-1079-0

Vira, J., Hess, P., Ossohou, M., & Galy-Lacaux, C. (2022). Evaluation of interactive and prescribed agricultural ammonia emissions for simulating atmospheric composition in CAM-chem. *Atmospheric Chemistry and Physics*, *22*(3), 1883–1904. https://doi.org/10.5194/acp-22-1883-2022

Vonk, J. E., Fritz, M., Speetjens, N. J., Babin, M., Bartsch, A., Basso, L. S., et al. (2025). The land–ocean Arctic carbon cycle. *Nature Reviews Earth & Environment*, *6*(2), 86–105. https://doi.org/10.1038/s43017-024-00627-w

Wang, H., Easter, R. C., Zhang, R., Ma, P.-L., Singh, B., Zhang, K., et al. (2020). Aerosols in the E3SM version 1: New developments and their impacts on radiative forcing. *Journal of Advances in Modeling Earth Systems*, *12*(1), e2019MS001851. https://doi.org/10.1029/2019MS001851

Wieder, W. R., Boehnert, J., & Bonan, G. B. (2014). Evaluating soil biogeochemistry parameterizations in Earth system models with observations. *Global Biogeochemical Cycles*, *28*(3), 211–222. https://doi.org/10.1002/2013GB004665

Wilkenskjeld, S., Kloster, S., Pongratz, J., Raddatz, T., & Reick, C. H. (2014). Comparing the influence of net and gross anthropogenic land-use and land-cover changes on the carbon cycle in the MPI-ESM. *Biogeosciences*, *11*(17), 4817–4828. https://doi.org/10.5194/bg-11-4817-2014

Woodward, F. I., & Lomas, M. R. (2004). Vegetation dynamics – Simulating responses to climatic change. *Biological Reviews*, *79*(3), 643–670. https://doi.org/10.1017/S1464793103006419

Xu, L., Saatchi, S. S., Yang, Y., Yu, Y., Pongratz, J., Bloom, A. A., et al. (2021). Dataset for "Changes in Global Terrestrial Live Biomass over the 21st Century" (version 2.0) [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.4161694

Yang, X., Thornton, P., Ricciuto, D., Wang, Y., & Hoffman, F. (2023). Global evaluation of terrestrial biogeochemistry in the Energy Exascale Earth System Model (E3SM) and the role of the phosphorus cycle in the historical terrestrial carbon balance. *Biogeosciences*, *20*(14), 2813–2836. https://doi.org/10.5194/bg-20-2813-2023

Zhang, C., Golaz, J.-C., Forsyth, R., Vo, T., Xie, S., Shaheen, Z., et al. (2022). The E3SM Diagnostics Package (E3SM Diags v2.7): A Python-based diagnostics package for Earth system model evaluation. *Geoscientific Model Development*, *15*(24), 9031–9056. https://doi.org/10.5194/gmd-15-9031-2022

Zhang, G. J., & McFarlane, N. A. (1995). Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian climate centre general circulation model. *Atmosphere-Ocean*, *33*(3), 407–446. https://doi.org/10.1080/07055900.1995.9649539

Zhao, A., Li, Z., Zou, L., Wu, J., Stan, K., & Sanchez-Azofeif, A. (2025). Evaluating dynamic Global Vegetation Models in China: Challenges in capturing trends in Leaf Area and Gross Primary Productivity, but effective seasonal variation representation. In *Earth System Dynamics Discussions* (pp. 1–32). https://doi.org/10.5194/esd-2024-44

Zhou, T., Leung, L. R., Leng, G., Voisin, N., Li, H.-Y., Craig, A. P., et al. (2020). Global irrigation characteristics and effects simulated by fully coupled land surface, River, and water management models in E3SM. *Journal of Advances in Modeling Earth Systems*, *12*(10), e2020MS002069. https://doi.org/10.1029/2020MS002069