SPECIAL
Process-Oriented Model Diagnostics
COLLECTION

Evolution of Uncertainty in Terrestrial Carbon Storage in Earth System Models from CMIP5 to CMIP6

NING WEI,[a,b] JIANYANG XIA,[a] JIAN ZHOU,[a] LIFEN JIANG,[b] ERQIAN CUI,[a] JIAYE PING,[a] AND YIQI LUO[b]

[a] *Research Center for Global Change and Complex Ecosystems, State Key Laboratory of Estuarine and Coastal Research, School of Ecological and Environmental Sciences, East China Normal University, Shanghai, China*
[b] *Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, Arizona*

(Manuscript received 29 September 2021, in final form 10 April 2022)

ABSTRACT: The spatial and temporal variations in terrestrial carbon storage play a pivotal role in regulating future climate change. However, Earth system models (ESMs), which have coupled the terrestrial biosphere and atmosphere, show great uncertainty in simulating the global land carbon storage. Here, based on multiple global datasets and a traceability analysis, we diagnosed the uncertainty source of terrestrial carbon storage in 22 ESMs that participated in phases 5 and 6 of the Coupled Model Intercomparison Project (CMIP5 and CMIP6). The modeled global terrestrial carbon storage has converged among ESMs from CMIP5 (1936.9 ± 739.3 PgC) to CMIP6 (1774.4 ± 439.0 PgC) but is persistently lower than the observation-based estimates (2285 ± 669 PgC). By further decomposing terrestrial carbon storage into net primary production (NPP) and ecosystem carbon residence time (τ_E), we found that the decreased intermodel spread in land carbon storage primarily resulted from more accurate simulations on NPP among ESMs from CMIP5 to CMIP6. The persistent underestimation of land carbon storage was caused by the biased τ_E. In CMIP5 and CMIP6, the modeled τ_E was far shorter than the observation-based estimates. The potential reasons for the biased τ_E could be the lack of or incomplete representation of nutrient limitation, vertical soil biogeochemistry, and the permafrost carbon cycle. Moreover, the modeled τ_E became the key driver for the intermodel spread in global land carbon storage in CMIP6. Overall, our study indicates that CMIP6 models have greatly improved the terrestrial carbon cycle, with a decreased model spread in global terrestrial carbon storage and less uncertain productivity. However, more efforts are needed to understand and reduce the persistent data–model disagreement on carbon storage and residence time in the terrestrial biosphere.

KEYWORDS: Ecological models; Model comparison; Model evaluation/performance

1. Introduction

Earth's land surface stores more than 3 times the amount of carbon (C) in the atmosphere (Houghton 2007). Based on the updated global C budget (Friedlingstein et al. 2019), the terrestrial ecosystem currently acts as a C sink and has cumulatively removed 31% of anthropogenic C emissions since industrial times. However, the vast terrestrial C storage is vulnerable to climate change (Nottingham et al. 2020; Schuur et al. 2015; Xia et al. 2014) and could turn into sources of C emission in a warming future, further exacerbating climate change (Arneth et al. 2010). For example, the substantial frozen C stored in high latitudes can be thawed, decomposed, and released into the atmosphere as the world warms (Koven et al. 2017; Ahlström et al. 2013). Climate warming can also trigger unexpectedly large CO_2 releases from the tropical forest soil (Nottingham et al. 2020). Thus, understanding and quantifying the vast but vulnerable C storage on the land have profound implications for predicting how the terrestrial C cycle responds to climate change (Nottingham et al. 2020; Schuur et al. 2015; Todd-Brown et al. 2013).

Earth system models (ESMs) are an essential tool for understanding and predicting climate change (Bonan and Doney 2018; Flato 2011). However, the uncertainty stemming from the terrestrial C cycle component strongly influences the trajectory of future climate change (Bodman et al. 2013; Booth et al. 2012; Cox et al. 2000; Friedlingstein et al. 2006, 2013; Xia et al. 2020). Phase 5 of the Coupled Model Intercomparison Project (CMIP5) (Taylor et al. 2011), for example, has recognized uncertainty in modeled terrestrial C cycle as a dominant factor for the spread of model ensemble in future climate prediction (Anav et al. 2013; Ciais et al. 2014; Friedlingstein et al. 2013). To confidently predict future climate–C feedback, ESMs need to improve the estimate of the current state of terrestrial C stocks (Todd-Brown et al. 2013). However,

Wei's ORCID: 0000-0003-2864-9568.
Xia's ORCID: 0000-0001-5923-6665.
Zhou's ORCID: 0000-0001-9276-6086.
Cui's ORCID: 0000-0002-2639-5069.
Ping's ORCID: 0000-0003-3427-4328.
Luo's ORCID: 0000-0002-4556-0218.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-21-0763.s1.

Corresponding author: Jianyang Xia, jyxia@des.ecnu.edu.cn

DOI: 10.1175/JCLI-D-21-0763.1

TABLE 1. List of CMIP5 ESMs selected in this study. The ESM names in boldface indicate models participating in two phases of CMIP but in different versions. (Expansions of many acronyms are available online at http://www.ametsoc.org/PubsAcronymList.)

ESM	Land model	N cycle	Number of C pools			Reference(s)
			Plant	Litter	Soil	
BCC-CSM1-1m	BCC_AVIM1.0	No	3	2	6	Wu et al. (2013)
CanESM2	CTEM	No	3	1	1	Arora et al. (2009)
CCSM4	CLM4.0	Yes	18	4	3	Gent et al. (2011), Lindsay et al. (2014)
HadGEM2-ES	MOSES/TRIFFID	No	3	—	4	Collins et al. (2011)
IPSL-CM5A-MR	ORCHIDEE	No	8	4	3	Dufresne et al. (2013), Krinner et al. (2005)
MIROC-ESM	SEIB-DGVM	No	4	1	2	Watanabe et al. (2011), Sato et al. (2007)
MPI-ESM-MR	JSBACH	No	3	4	1	Giorgetta et al. (2013), Schneck et al. (2013)
NorESM1-M	CLM4.0	Yes	18	4	3	Tjiputra et al. (2013)
BNU-ESM	CoLM	No	4	1	1	Ji et al. (2014), Sitch et al. (2003)
GFDL-ESM2G	LM3.0	No	5	—	2	Dunne et al. (2012)
MRI-ESM1	HAL	No	3	1	1	Yukimoto et al. (2011)

CMIP5 ESMs vary widely in simulating both the magnitude and the spatial distribution of global terrestrial C storage (Anav et al. 2013; Jones et al. 2016; Todd-Brown et al. 2013; Wieder et al. 2015).

Over the past few years, numerous studies have evaluated the terrestrial C cycle in the CMIP5 models. Most of the CMIP5 models have overestimated leaf area (Anav et al. 2013) and net primary productivity (Anav et al. 2013; Wieder et al. 2015) but underestimated ecosystem C turnover time (Carvalhais et al. 2014; Todd-Brown et al. 2013; Wang et al. 2019). Koven et al. (2015) have separated ecosystem C storage change into four categories based on productivity and C turnover time. They found that productivity determines the intermodel spread in C stocks changes, but C turnover time dominates model disagreement in the initial state. Those model evaluation analyses together recommend the implementation of many critical processes into ESMs, such as nutrient cycles (Koven et al. 2015; Wieder et al. 2015), vertical soil profiles (Koven et al. 2013), permafrost dynamics (Schuur et al. 2015), and disturbance regimes (Ciais et al. 2014).

In CMIP6, models have significantly improved the process representation for the terrestrial C cycle (Eyring et al. 2016; Jones et al. 2016). For example, an increasing number of models have explicitly included the nutrient limitation on terrestrial C processes (Davies-Barnard et al. 2020); some models have added vertical representation of soil biogeochemistry (Lawrence et al. 2019; Seland et al. 2020); land use and land cover changes are also explicitly involved in modeling terrestrial C cycle in several models (Lawrence et al. 2019; Wu et al. 2019). Even with those substantial developments, models in CMIP6 still show considerable uncertainty in terrestrial C stocks (Akihiko et al. 2020). A critical step to reducing such model uncertainty is understanding the causes of model spread and the model deviations from the observations.

This study evaluates whether and how the intermodel spread and model–data difference in terrestrial C storage change from CMIP5 to CMIP6. Recently, a growing body of observation-derived data products at high spatiotemporal resolutions has become publicly available (Fan et al. 2020; Hengl et al. 2017; Kolby Smith et al. 2016; Spawn et al. 2020), such as satellite-derived terrestrial productivity maps (Kolby Smith et al. 2016; Running et al. 2015), global maps of aboveground and belowground biomass C (Spawn et al. 2020), and observation-based estimates on soil organic C density (Hengl et al. 2017). These valuable data pave the way to assess the current state of the terrestrial C cycle and can be used to benchmark model performance. This study first evaluated the intermodel spread and data–model difference in terrestrial C storage in CMIP5 and CMIP6. Then, we conducted data–model comparisons on specific processes to identify the potential reasons for model bias in terrestrial C storage. Last, we applied a traceability analysis to diagnose why models performed differently within and between the two CMIP ensembles.

2. Methods

a. *Model selection*

We conducted intermodel and data–model comparisons on the near-present land C storage in CMIP5 (Taylor et al. 2011) and CMIP6 (Eyring et al. 2016). In the historical experiment, ESMs were coordinately forced by reconstructed forcing data, including atmospheric composition and land-use change (Eyring et al. 2016; Hoesly et al. 2018; Taylor et al. 2011). Thus, the historical model outputs included the impacts of land-use and land-cover change (LULCC) on terrestrial C cycle (Lawrence et al. 2016; Akihiko et al. 2020). Given that the covered historical periods were different in two CMIP ensembles (CMIP5: 1850–2005; CMIP6: 1850–2014), the near-present land C storage was calculated as the 2001–05 mean for consistency. ESMs were selected based on whether key variables were provided, including Vegetation C (cVeg), Litter C (cLitter), Soil C (cSoil), net primary production (NPP), gross primary production (GPP), near-surface temperature (tas), precipitation (pr), grid cell area (areacella), and land area fraction (sftlf). For ESMs from the same modeling center, we only selected one model as the representative because ESMs within a center tended to produce very similar results (Todd-Brown et al. 2014; Todd-Brown et al. 2013). Eventually, we selected 11 ESMs from CMIP5 (Table 1) and 11 ESMs selected from CMIP6

TABLE 2. List of CMIP6 ESMs selected in this study. The ESM names in boldface indicate models participating in two phases of CMIP but in different versions.

ESM	Land model	N cycle	Number of C pools			Reference(s)
			Plant	Litter	Soil	
BCC-CSM2-MR	BCC-AVIM2	No	3	2	6	Wu et al. (2019)
CanESM5	CLASS-CTEM	No	3	1	1	Swart et al. (2019)
CESM2	CLM5.0	Yes	18	4 × 20	3 × 20	Lawrence et al. (2019)
UKESM1-0-LL	JULES-CN;TRIFFID	Yes	3	—	4	Best et al. (2011), Clark et al. (2011)
IPSL-CM6A-LR	ORCHIDEE	Yes	8	4	3	Vuichard et al. (2019)
MIROC-ES2L	VISIT-e	Yes	3	1	1	Hajima et al. (2019), Ito and Oikawa (2002)
MPI-ESM1-2-LR	JSBACH	Yes	3	2	1	Goll et al. (2012), Mauritsen et al. (2019)
NorESM2-LM	CLM5.0	Yes	18	4 × 20	3 × 20	Seland et al. (2020)
ACCESS-ESM1-5	CABLE	Yes	3	3	3	Law et al. (2017)
CNRM-ESM2-1	ISBA	No	6	4	3	Séférian et al. (2019)
EC-Earth3-Veg	LPJ-GUESS	Yes	3	8	3	Smith et al. (2014)

(Table 2). Eight models from eight modeling centers participated in both CMIP5 and CMIP6 (Tables 1 and 2).

b. *Data–model comparison on the near-present terrestrial C storage*

The modeled ecosystem C storage was calculated as the sum of cVeg, cLitter, and cSoil. In HadGEM2-ES (Collins et al. 2011), GFDL-ESM2G (Dunne et al. 2012), and UKESM1-0-LL (Clark et al. 2011), litterfall from vegetation is directly transformed into soil C, based on local litterfall rates and the decomposable fraction of the litter input (Clark et al. 2011). For these models, the combination of cVeg and cSoil thus represents the estimation of ecosystem C storage. We used a recently published dataset product of cVeg (Spawn et al. 2020) and multiple observation-derived datasets of cSoil for data–model comparison. The dataset of cVeg represented biomass C density for the year 2010, including both aboveground and belowground amounts. Here, we assumed that biomass C stocks over 2001–05 could approximate that in 2010. The dataset products of cSoil included the regridded Harmonized World Soil Database (HWSD) v1.2 (Wieder 2014), the SoilGrids (Hengl et al. 2017), the LandGIS (https://zenodo.org/record/2536040#.Xzt1yOgzZjX), and the Northern Circumpolar Soil Carbon Database version 2 (NCSCDv2) (Hugelius et al. 2013). The covered soil depth varied from 1 m (HWSD v1.2) to 3 m (NCSCD). For SoilGrids and LandGIS, the covered depth was 2 m. For consistency, we used the 0–1-m layer of cSoil. We further combined the observation-based estimates on cVeg and cSoil at 0–1 m to derive the near-present terrestrial ecosystem C storage. Some previous studies have detected significant differences among these cSoil datasets (Fan et al. 2020; Todd-Brown et al. 2013). Given a lack of consensus as to which dataset was better, we merged these cSoil datasets and estimated terrestrial ecosystem C storage through factorial combinations of the cVeg and cSoil datasets. The uncertainty of dataset products was provided by the range of different estimates. Furthermore, many studies have suggested that the CMIP5 models show large intermodel and data–model disagreement over the northern high latitudes (Carvalhais et al. 2014; Todd-Brown et al. 2013; Koven et al. 2015). To investigate whether model performance over this region improved from CMIP5 to CMIP6, we separated global terrestrial C stocks into that over the circumpolar and non-circumpolar areas based on the map of the NCSCD (Hugelius et al. 2013). Dataset products are summarized in Table 3.

TABLE 3. Observation-based estimates of circumpolar, non-circumpolar, and global terrestrial C storage. Plant biomass C was from Spawn et al. (2020). The values of soil C stock above 1 m estimated from four datasets were extracted from Fan et al. (2020) (based on their Table 2). Ecosystem C storage was calculated based on factorial combinations of the reported plant and soil C estimates. The classification of circumpolar and non-circumpolar regions was based on the map of the Northern Circumpolar Soil Carbon Database (NCSCD; Hugelius et al. 2013).

C storage (Pg C)	Circumpolar	Non-circumpolar	Global	Reference[a]
Plant C ± SD	45.5 ± 21	366.8 ± 207	412.3 ± 227.9	Spawn et al. (2020)
Soil C	796	1399	2195	SoilGrids
	787	1305	2091	LandGIS
	568	764	1332	HWSD
	567			NCSCD
Ecosystem C (minimum–maximum range)	725 (591.6–862.5)	1522.8 (923.8–1972.7)	2285 (1516.4–2835.2)	Factorial combinations of plant C and soil C (0–1 m) estimations

[a] For more information on the references, see https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1763 (Spawn et al. 2020); https://www.isric.org/explore/soilgrids (SoilGrids); https://zenodo.org/record/2536040#.XhxHRBf0kUF (LandGIS); http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/ (HWSD); and https://bolin.su.se/data/ncscd/ (NCSCD).

None of the ESMs in CMIP5 explicitly represented the vertical distributions of soil C. From CMIP5 to CMIP6, only CESM2 and NorESM2-LM from NCAR and the Norwegian Climate Centre, respectively, had evolved to incorporate the vertically resolved soil C and litter C to a depth of 8.03 m (20 layers) (Lawrence et al. 2019). To simplify intermodel and data–model comparison, we assumed models with a single soil layer could simulate cSoil within the 1-m depth (Todd-Brown et al. 2013). The 0–1-m cSoil and cLitter layers were used for the two models with multiple soil layers to calculate the modeled ecosystem C storage. The spatial resolution of the used dataset products varied from 250 m (SoilGrids and LandGIS) to 0.5° × 0.5° (NCSCD). Both model outputs and dataset products were regridded into the 0.5° × 0.5° resolution for intermodel and data–model comparisons.

c. Data–model comparison on τ_E and NPP

The capability of a terrestrial ecosystem to store C is determined by the rate of C into the ecosystem (through photosynthesis) and how long this C input can reside in the ecosystem (Carvalhais et al. 2014; Keenan and Williams 2018) (ecosystem C residence time or ecosystem C transit time τ_E). When an ecosystem approaches the steady state, the C influx equals efflux, and C reservoirs reach their maximum C storage. The product of C influx (or efflux) and τ_E defines the C storage capacity (X_C) when the ecosystem is at a steady state (Koven et al. 2015; Luo et al. 2017; Luo and Weng 2011; Xia et al. 2013):

$$X_C = \tau_E \times \text{flux}. \tag{1}$$

Physiologically, a large proportion of photosynthetically fixed C passes quickly back into the atmosphere via respiration (Bradford and Crowther 2013). On the yearly to decadal time scale, this study considered NPP as the C influx to an ecosystem, assuming C losses via autotrophic respiration happened instantaneously in the process of photosynthesis. Under the assumption of steady state, land C storage (X) equals to X_C, and the X in both observations and models could be decomposed into NPP and τ_E:

$$X = \tau_E \times \text{NPP}. \tag{2}$$

Then τ_E can be estimated as the ratio of X to NPP:

$$\tau_E = \frac{X}{\text{NPP}}. \tag{3}$$

We then conducted a model–data comparison on NPP and τ_E to investigate the sources of model bias in the modeled land C storage in two CMIP ensembles. Three dataset products of NPP were used as the benchmark, including MOD17A2 (Running et al. 2015), GIMMS-NPP (Kolby Smith et al. 2016), and CARDAMOM NPP (Bloom et al. 2016). The GIMMS-NPP was derived based on the MODIS NPP algorithm but driven by different satellite products, while the CARDAMOM NPP was retrieved using data–model fusion analysis. Similarly, we considered the range of three NPP dataset products as the uncertainty range of the benchmarks. The mean and uncertainty range of the ensemble data products were used to evaluate the two CMIP model ensembles. Based on the dataset products of ecosystem C storage and NPP, we further calculated τ_E in the manner as shown in Eq. (3). The uncertainty range of τ_E is estimated based on different combinations of ecosystem C storage data and NPP data. Note that we only considered cSoil within the top 1 m when comparing modeled τ_E against the data.

d. Traceability analysis on the modeled land C storage

A traceability framework was further applied to address how sources of model differences changed across two CMIP ensembles. The traceability analysis has been developed for tracing intermodel spread in the terrestrial C cycle (Huang et al. 2018; Jiang et al. 2017; Xia et al. 2013; Zhou et al. 2018). In the same manner, as shown in Eq. (2), the modeled terrestrial C storage was first decomposed into NPP and τ_E. NPP is jointly determined by gross primary productivity (GPP) and C use efficiency (CUE) (Xia et al. 2017):

$$\text{NPP} = \text{GPP} \times \text{CUE}, \tag{4}$$

where CUE is a physiological parameter that defines the proportion of assimilated C from photosynthesis allocating to plant biomass (Bradford and Crowther 2013). We also conducted a data–model comparison on GPP and CUE to diagnose the sources of model biases in NPP. Three dataset products of GPP were used as benchmarks, including MODIS (MOD17A2; Running et al. 2015), FLUXCOM (Jung et al. 2017), and VPM (Zhang et al. 2017). The CUE dataset product was from CARDAMOM (Bloom et al. 2016).

The τ_E is an important ecosystem property involving multiple processes (Carvalhais et al. 2014; Xia et al. 2013; Luo et al. 2017), including C allocation (term B), C transferring network (term A), decomposition processes (term K), and regulations from environmental factors (term ξ). These processes can be characterized by an analytical solution (Xia et al. 2013; Luo et al. 2017) as

$$\tau_E = B\left[AK\xi(t)\right]^{-1}. \tag{5}$$

To reduce complexity in the traceability analysis, we merged the influences of B, A, and K on τ_E and refer to it as the baseline C residence time (τ_E'). Thus the τ_E can be decomposed into τ_E' and the influences from environmental factors (Xia et al. 2013):

$$\tau_E = \tau_E'\xi^{-1}, \tag{6}$$

where the ξ is a scalar representing modifications of environmental factors on τ_E'. In this study, we only considered the modifications of temperature and precipitation, that is, the temperature scalar (ξ_T) and the precipitation scalar (ξ_W):

$$\xi = \xi_T\xi_W, \tag{7}$$

$$\tau_E = \tau_E'(\xi_T\xi_W)^{-1}. \tag{8}$$

Based on Zhou et al. (2018), mean annual temperature (T) and annual total precipitation (W) were used to estimate ξ_T and ξ_W. For ξ_T, it is expressed as a function of T, parameter Q_{10},

and the reference temperature T_0. For ξ_W, it is expressed as a function of W and the reference precipitation W_0:

$$\xi_T = Q_{10}^{[(T-T_0)/10]}, \tag{9}$$

$$\xi_W = \frac{W}{W_0}, \tag{10}$$

where Q_{10} is a parameter associated with the temperature sensitivity of ecosystem respiration; T_0 and W_0 were set to be the maximum values of annual temperature and precipitation, respectively, across the simulation period for each ESM (Todd-Brown et al. 2013; Zhou et al. 2018). The same optimization method as in Zhou et al. (2018) was applied to estimate two unknown parameters, Q_{10} and τ_E'. The τ_E calculated from model outputs [Eq. (3)] was used to calibrate the estimated τ_E from the baseline residence time and environmental scalars [Eq. (8)]. The coefficient of determination R^2 and the root-mean-square error (RMSE) are used as indicators for the goodness of the calibration, and the objective is to maximize R^2 and minimize RMSE. The comparison between the τ_E calculated from model outputs based on Eq. (3) and that reproduced by using the optimization method is shown in Fig. S22.

e. *Variance partitioning*

Based on the traceability framework, we first decomposed the modeled terrestrial C storage into its two determinants (NPP and τ_E), and then traced the influential factors for both NPP and τ_E. After the decomposition processes, we applied the hierarchical partitioning method (Chevan and Sutherland 1991; Murray and Conner 2009) to detect independent contributions of different variables to the total variance of land C storage across models. The hierarchical partitioning method is briefly described below.

For a given dependent variable Y and for k explanatory variables ($x_1, x_2, \ldots, x_k$), there would be 2^k different fits to Y with different combinations of explanatory variables. Taking $k = 3$ as an example, there are 2^3 combinations, namely X_0, X_1, X_2, X_3, X_{12}, X_{13}, X_{23}, and X_{123}. The subscripts indicate explanatory variables in that combination; X_{12} indicates the combination of x_1 and x_2, while X_0 is a null number. The independent effect of a specific explanatory variable x_l (I_{x_l}, $l = 1, 2, 3, \ldots, k$) to the variance in Y was calculated by comparing the fit of all models including X_i to that when omitting X_i from the model, based on a hierarchical ordering (Chevan and Sutherland 1991; Murray and Conner 2009):

$$I_{x_l} = \sum_{i=0}^{k-1} \frac{\sum (r^2_{y,\,x_l,x_h} - r^2_{y,x_h}) \Big/ \binom{k-1}{i}}{k}, \tag{11}$$

where x_h is any combination of explanatory variables without x_l (Chevan and Sutherland 1991; Murray and Conner 2009). Its relative independent contribution ($\mathrm{re}I_{x_l}$) is calculated as

$$I = \sum_{l=1}^{k} I_{x_l}, \tag{12}$$

$$\mathrm{re}I_{x_l} = \frac{I_{x_l}}{I}. \tag{13}$$

We conducted the hierarchical partitioning analysis in R by using the "hier.part" package. Linear regression and the R^2 goodness-of-fit measure were used. Formulations were logarithmically transformed if the formulation was not expressed as the sum of variables.

3. Results

a. *Converging but still underestimated near-present land C storage*

We first compared intermodel spread in the near-present terrestrial C storage between CMIP5 and CMIP6. The across-model variation in the global terrestrial C storage decreased from CMIP5 to CMIP6, and the model ensemble average became smaller in CMIP6 (Figs. 1a,d). At the end of the historical period, the global terrestrial C storage was 1774.4 ± 439.0 PgC (mean ± σ) for CMIP6 and 1936.9 ± 739.3 PgC for CMIP5 (Table 3; see also Table S2 in the online supplemental material). From CMIP5 to CMIP6, across-model divergence (indicated by σ) in global terrestrial C storage decreased by 40.6%. We analyzed a subset of models from eight modeling centers that participated in both CMIP5 and CMIP6 (Tables 1 and 2) to make it more comparable. For the eight CMIP6 models, their ensemble model spread in terrestrial C storage decreased by 46.4% compared to their predecessors in CMIP5. While the decreased model spread was detected over many regions in CMIP6 (Figs. 1a,d), northern high-latitude regions showed relatively larger model divergence than tropical regions (Figs. 1a,d). By combining dataset products of biomass (Spawn et al. 2020) and soil C stocks (Table 3), the terrestrial ecosystem was estimated to store 2285 PgC, with the uncertainty ranging between 1516 and 2835 PgC (Table 3). Compared with the ensemble mean value derived from the dataset products, the two CMIP model ensembles underestimated the global terrestrial C storage (Figs. 1b,e). At the regional scale, ecosystem C stocks at northern high latitudes were considerably underestimated by both CMIPs (Figs. 1c,f).

From CMIP5 to CMIP6, model performance in simulating circumpolar-region C storage did not show significant improvement (Fig. 2). In general, biomass C stocks simulated from CMIP6 models (41.9 ± 16.9 PgC) converged more and were closer to the range of observation-based estimates (45.5 ± 21 PgC) than CMIP5 (48.4 ± 45.2 PgC; Fig. 2). However, most CMIP6 models underestimated soil C at the top 1-m depth (Fig. 2), even for the two CMIP6 models (CESM2 and NorESM2-LM) representing vertically resolved soil C (Table S2). Circumpolar-region soil C stocks in the top 1-m layer simulated by CESM2 and NorESM2-LM were 316.8 and 317.8 PgC, respectively. Despite the tremendous increases compared to their predecessors (CCSM4, 56.9 PgC; NorESM1-M, 60.9 PgC), the two models were still far smaller than the observation-based estimation, 679.5 PgC (ranging from 567 to 796 PgC; Table 3). Outside of the circumpolar region, the model ensemble means of C stocks in two CMIP ensembles fell within

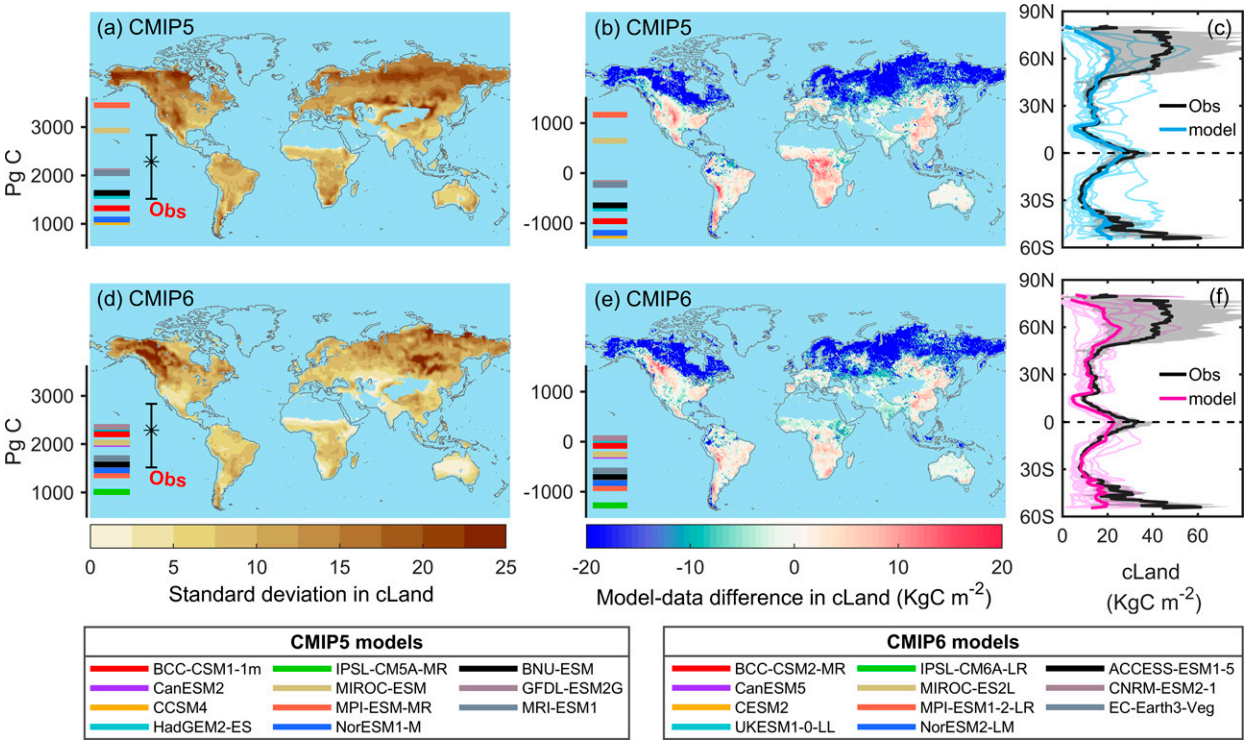

FIG. 1. Intermodel variation and model–data difference in the near-present land C storage (2001–05 mean). (left) The spatial patterns of standard deviation in land C storage for (a) CMIP5 and (d) CMIP6 model ensembles. The inserted panels show intermodel spread in global land C storage for two CMIP ensembles and the range of observation-based estimates (values shown in Table 3). Asterisks represent the mean and the edges of the bar represent the minimum and maximum. (center) The spatial patterns of model–data difference in land C storage for (b) CMIP5 and (c) CMIP6, computed as the difference between the model ensemble mean and the observation-derived mean. The inserted panels show the model–data difference in global land C storage (in PgC) for individual models. (right) Zonal mean plots of land C storage for (c) CMIP5 and (f) CMIP6, in comparison with dataset products. The thin lines represent simulations from individual models, while the thick lines show ensemble-mean values for models and observation-based estimates. The gray shaded area shows the uncertainty range of observation-based estimates from maximum to minimum. The spatial patterns of land C storage for individual models and dataset products are displayed in Figs. S1–S3 in the online supplemental material.

the ranges of observation-based estimates, and the across-model variations decreased from CMIP5 to CMIP6 (see supplemental Figs. S4d,e).

b. *Model bias stemming from NPP and τ_E*

We then conducted a model–data comparison on NPP and τ_E to gain a deeper insight into the persistently underestimated global terrestrial C storage in two CMIP ensembles. The analyses detected dramatic reductions in model bias and intermodel variation in global terrestrial NPP from CMIP5 to CMIP6 (Fig. 3). However, the global terrestrial τ_E was persistently underestimated by two CMIP ensembles, and the intermodel spread did not significantly change between the two CMIP ensembles (Fig. 3). The persistent underestimation of τ_E led to the persistently underestimated global terrestrial C storage in two CMIP ensembles (Fig. 3). In CMIP6, the RMSE of modeled land C storage decreased by 16.2% compared to CMIP5, even though the bias score of land C storage in CMIP6 was slightly larger than that in CMIP5. The primary reason for the lower underestimation bias in CMIP5 was the high NPP in CMIP5 (Fig. 3a). Global terrestrial NPP in CMIP5 was 67.7 ± 16.3 PgC yr^{-1}, which overestimated the observation-derived value by 30.9% on average (Fig. 3a). From CMIP5 to CMIP6 (58.2 ± 11.3 PgC yr^{-1}), both the overestimation bias and RMSE in modeled global terrestrial NPP decreased (Fig. 3).

The decreased uncertainty in NPP from CMIP5 to CMIP6 was widely detected, particularly over tropical regions where NPP was vastly overestimated by the CMIP5 models (Figs. 4a–d). Model performance in the African tropical forests was greatly improved (Fig. 4c), while the improvement for the tropics over South America was relatively small (Fig. 4c). In both the circumpolar and non-circumpolar areas, the overall bias and RMSE of modeled NPP decreased from CMIP5 to CMIP6 (Figs. 4f,g). Observation-based estimates on circumpolar-region NPP ranged from 3.8 to 5.1 PgC yr^{-1}. The modeled NPP over the circumpolar region was 5.4 ± 2.5 PgC yr^{-1} in CMIP5 and decreased to 4.5 ± 1.1 PgC yr^{-1} in CMIP6. Over the non-circumpolar region, the observation-based estimates ranged from 40.7 to 53.4 PgC yr^{-1}. There were four CMIP5 models within the observation-derived NPP range,

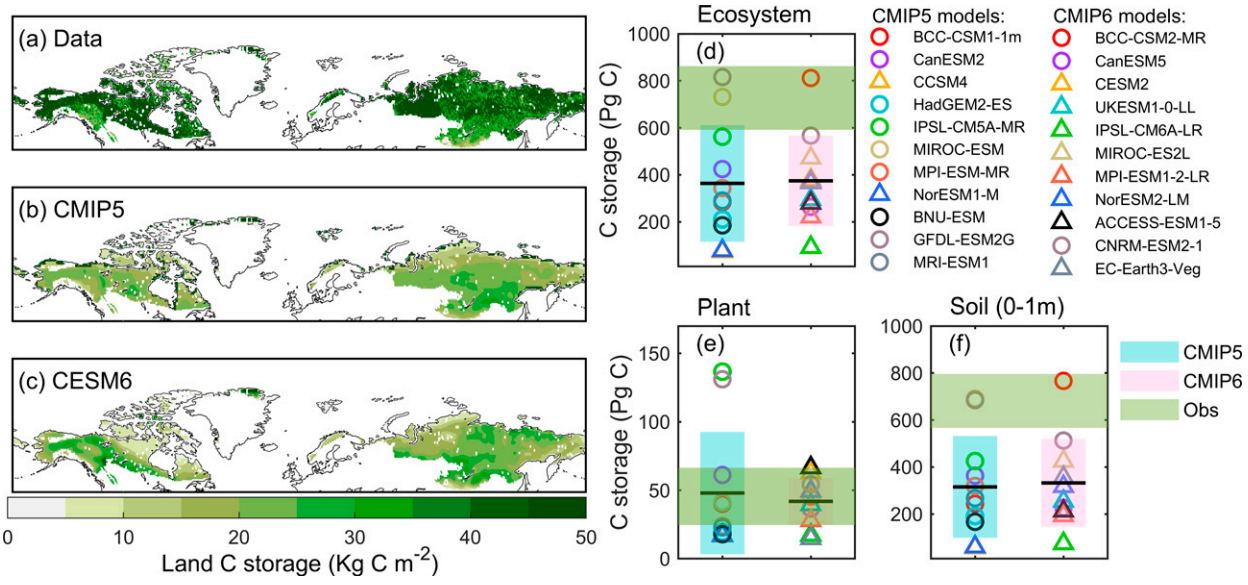

FIG. 2. Comparison between model-based and observation-based estimations on the land C storage over the circumpolar region. (left) Distributions of ecosystem C storage over the circumpolar region estimated from (a) dataset products, (b) CMIP5, and (c) CMIP6. The classification of circumpolar and non-circumpolar regions was based on the map of the Northern Circumpolar Soil Carbon Database (NCSCD; Hugelius et al. 2013). (right) Comparison of (d) ecosystem (involving plant C and soil C above 1 m), (e) plant, and (f) soil (0–1 m) C storage in the circumpolar region between models and dataset products. The green shaded area shows the range (minimum–maximum) of observation-based estimates (Table 3). Circles represent models without nutrient cycle. Triangles represent models explicitly incorporating nutrient cycle with terrestrial C cycle. Solid dark lines of boxplots show multimodel means. The edges of boxplots show standard deviation.

namely BCC-CSM1-1m, CCSM4, NorESM1-M, and BNU-ESM. In CMIP6, the number increased to seven: BCC-CSM2-MR, CESM2, IPSL-CM6A-LR, NorESM2-LM, ACCESS-ESM1-5, CNRM-ESM2-1, and EC-Earth3-Veg.

Consistent with the bias pattern of land C storage (Figs. 1b,e), model ensembles in two CMIP ensembles significantly underestimated τ_E in the northern high latitudes (Fig. 5). Over the northern circumpolar region, the τ_E simulated from the CMIP6 model ensemble was 83.7 ± 41.8 years, which was longer than CMIP5 (66.4 ± 39.8 years; Fig. 5f) but still far shorter than observation-based estimates (155.9 ± 34.2 years; Fig. 5f). The underestimation of τ_E outside of the circumpolar region was less severe (Fig. 5g). The τ_E values simulated from two CMIP ensembles were very close to the observation-derived τ_E (33 ± 6.6 years). The τ_E over the noncircumpolar region was 25.1 ± 6.7 years in CMIP5 and slightly increased to 26.5 ± 6.6 years in CMIP6.

c. *Increasing role of τ_E in causing intermodel variation in land C storage*

Our traceability analysis at a global scale showed that the role of τ_E in causing across-model variation in global land C storage increased from CMIP5 to CMIP6 (Fig. 6). NPP and τ_E contributed 49.8% and 50.2%, respectively, to the across-model variance in global land C storage in CMIP5 (Fig. 6a). In CMIP6, τ_E contributed 74% to the across-model variance in global terrestrial C storage, while the contribution of NPP decreased to 26% (Fig. 6b).

By further tracing sources of intermodel spread in τ_E into baseline residence time (τ_E') and environmental scalars (ξ), we found that τ_E' explained 98% and 67% of the across-model variance in τ_E in CMIP5 and CMIP6, respectively (Fig. 6). The contributions of environmental scalars to the across-model variance in τ_E were negligible in the CMIP5 ensembles (Fig. 6a). In CMIP6, however, the ξ_T contributed 32% to the across-model variance in τ_E. The modeled NPP was further decomposed into GPP and CUE. The latter defines the proportion of assimilated C from photosynthesis that is allocated to plant biomass growth (Bradford and Crowther 2013). In CMIP5, GPP and CUE contributed 68% and 32% to the variation of modeled NPP (Fig. 6a). In CMIP6, their relative contributions turned out to be equal (Fig. 6b).

d. *Evaluating model performance in simulating GPP and CUE*

For the global terrestrial GPP, the value simulated by CMIP6 models was 125.0 ± 13.8 PgC yr^{-1}, which was smaller and more converged than CMIP5 (152.1 ± 40.4 PgC yr^{-1}; Fig. 7a). Compared with the satellite-based estimates, which ranged from 98.1 to 136.0 PgC yr^{-1}, model uncertainty in GPP was significantly reduced from CMIP5 to CMIP6 at both global (Fig. 7a) and local scales (Fig. 8). However, model performance in representing the spatial pattern of CUE did not show significant improvement (Figs. 8d,h), although the model ensemble in CMIP6 showed slightly more convergent and less biased CUE at the global scale (Figs. 7b,c).

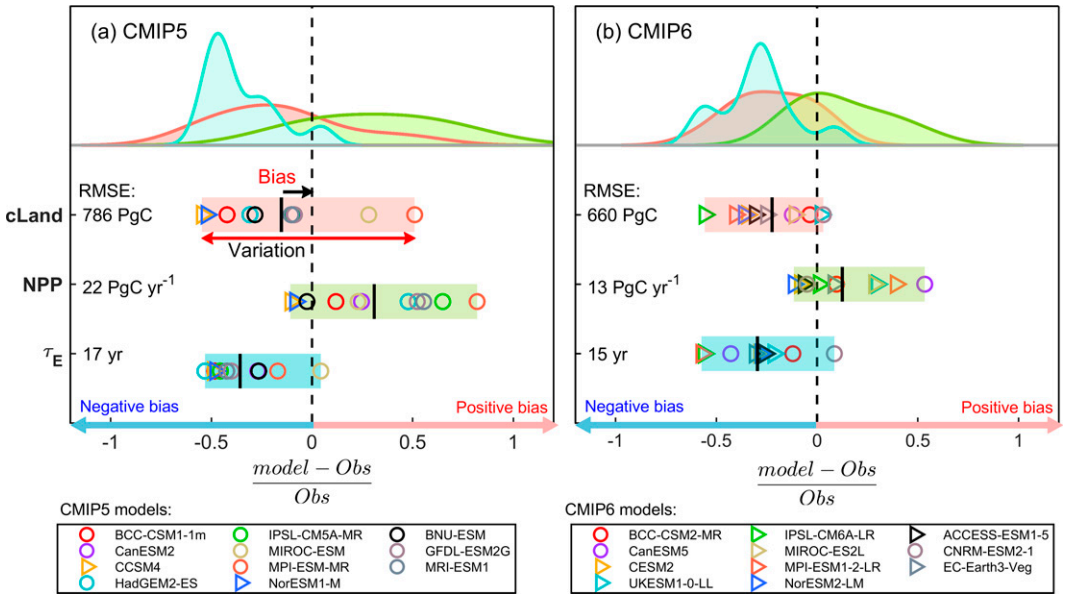

FIG. 3. Evaluating model bias score, intermodel spread, and root-mean-square error (RMSE) in global terrestrial C storage (cLand), NPP, and τ_E for (a) CMIP5 and (b) CMIP6. Here, the model bias score measures how much the model ensemble mean differs from the reference mean value. Across-model variation measures how much models differ from each other, while the RMSE measures how concentrated the model simulations are around the observation-based estimates. The above distributions show probability density plots for cLand, NPP, and τ_E. Values for individual models are shown below. Circles represent models without the nutrient cycle. Triangles represent models explicitly incorporating nutrient cycle with terrestrial C cycle. Solid dark lines of boxplots show multimodel means. The edges of boxplots show across-model variation from the maximum to the minimum.

4. Discussion

a. *The persistent underestimation of global land C storage in two CMIP ensembles*

By integrating the newly published datasets of above- and belowground C stocks, we evaluate how model performance in presenting the near-present land C storage, including both the magnitude and spatial distributions, changed between CMIP5 and CMIP6. We have identified persistent underestimation of global land C storage in two CMIP ensembles (Fig. 1), especially over the northern high latitudes (Fig. 2). Although we apply the data product of biomass C storage in the year 2010 (412.3 ± 227.9 PgC; Spawn et al. 2020) to approximate that over 2001–05, the persistent underestimation of global land C storage is still unchanged after accounting for associated uncertainties. A recent study suggests that the global C storage in vegetation biomass varies from 377 to 385 PgC over 2000–19 (Xu et al. 2021). Although different approaches are used in various studies (Ruesch and Gibbs 2008; Xu et al. 2021) to map the live biomass C, these observation-based estimates all fall within the range (412.3 ± 227.9 PgC) used in our data–model comparisons. To avoid compensation error from mixing vegetation and soil C stocks, we also conduct data–model comparisons of plant and soil C storage separately (Fig. 2; see also Fig. S4). We find that the underestimation of global terrestrial C storage primarily originates from the soil component, particularly over the circumpolar region (Fig. 2).

By decomposing terrestrial C storage into NPP and τ_E based on the steady-state assumption (Fig. 3), we further diagnose the reason behind the persistent underestimation of global terrestrial C storage. The global terrestrial τ_E in the two CMIP ensembles are far shorter than the observation-derived estimate, which caused the persistently underestimated global terrestrial C storage (Fig. 3). We also use GPP as the flux term in the estimation of τ_E to verify the results, given that assimilated C from photosynthesis is not respired instantaneously (Muhr et al. 2013; Sierra et al. 2022). The underestimation of global terrestrial τ_E is not affected when considering GPP as the flux term (Figs. S13 and S14). At the regional scale, the bias patterns of τ_E in two CMIP ensembles (Figs. 5a,c) are similar to the bias patterns of terrestrial C storage (Figs. 1b,e), which show significant underestimation in the northern high latitudes (Fig. 5). Many previous studies have identified the underestimation of τ_E at high latitudes in CMIP5 (Carvalhais et al. 2014; Todd-Brown et al. 2013) and have attributed it to the neglected representation of permafrost processes. In CMIP6, many models have implemented some key permafrost processes, such as freezing/thawing processes (Lawrence et al. 2019; Wu et al. 2019), snow insulation with multiple snow layers (Burke et al. 2017; Wu et al. 2019), and cryoturbation of soil organic C (Bowring et al. 2019; Koven et al. 2009). Although incorporating these processes leads to prolonged τ_E at northern high latitudes in CMIP6 (Fig. 5f), the modeled τ_E is still shorter than the observation-based estimates (Fig. 5f and Fig. S11).

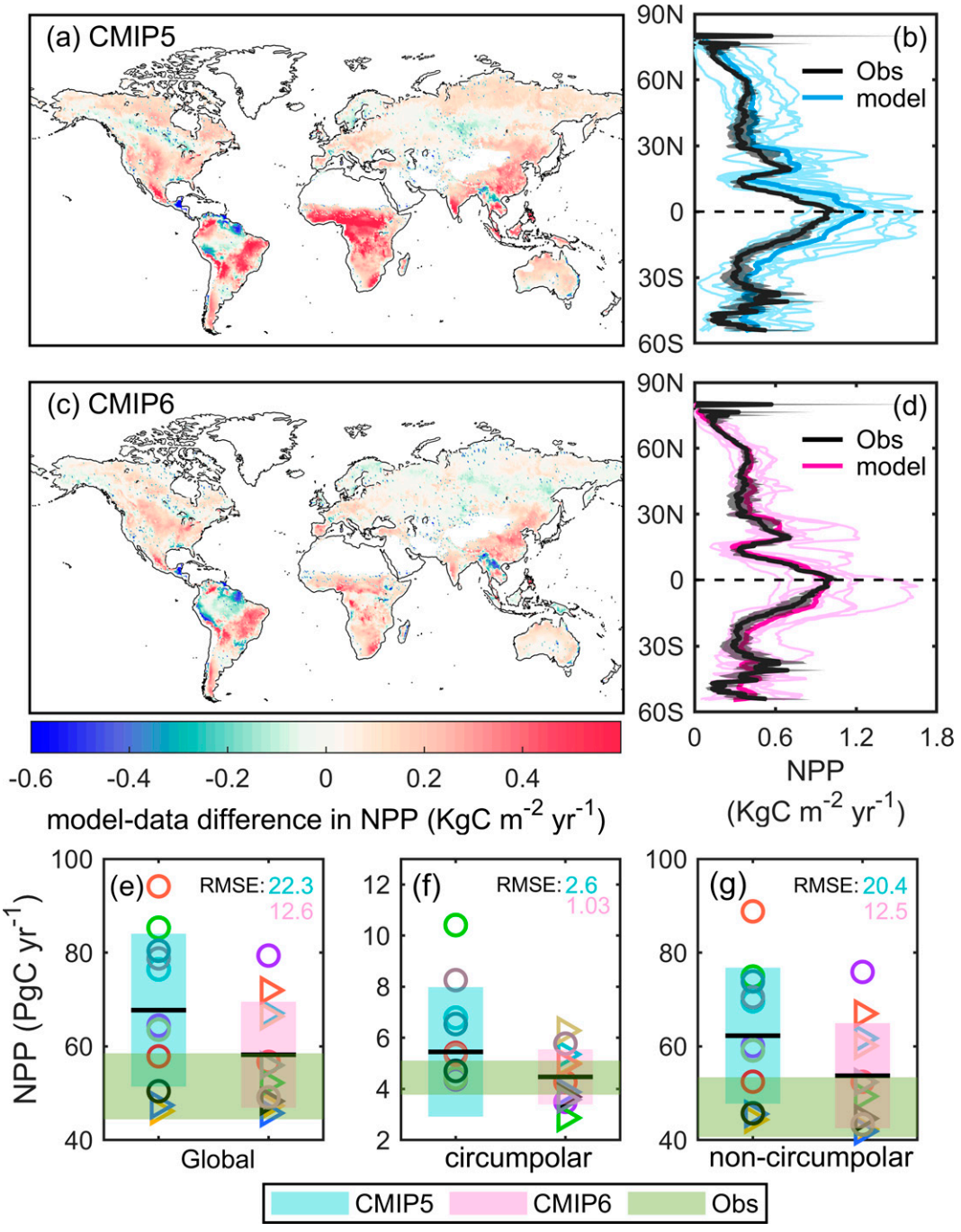

FIG. 4. Model performance in simulating the near-present (2001–05 mean) net primary productivity (NPP). (a),(c) Model bias in NPP for CMIP5 and CMIP6 model ensembles, respectively, computed as the difference between multimodel mean NPP and observation-derived estimate. (b),(d) Zonal mean plots of NPP for CMIP5 and CMIP6, respectively, in comparison with that from three dataset products (see methods). The thin lines represent simulations from each individual model, while the thick lines show the means. Also shown is the data–model comparison on NPP at the (e) global, (f) circumpolar, and (g) non-circumpolar scales. CMIP5 and CMIP6 models are color-coded as in Figs. 2 and 3. RMSE values for CMIP5 and CMIP6 are marked in different colors. The green shaded area shows the range (minimum–maximum) of observation-based estimates. Circles represent models without the nutrient cycle. Triangles represent models explicitly incorporating nutrient cycle with terrestrial C cycle. Solid dark lines of boxplots show multimodel means. The edges of boxplots show standard deviations. The spatial patterns of NPP for individual models and dataset products are displayed in Figs. S6–S8 in the online supplemental material.

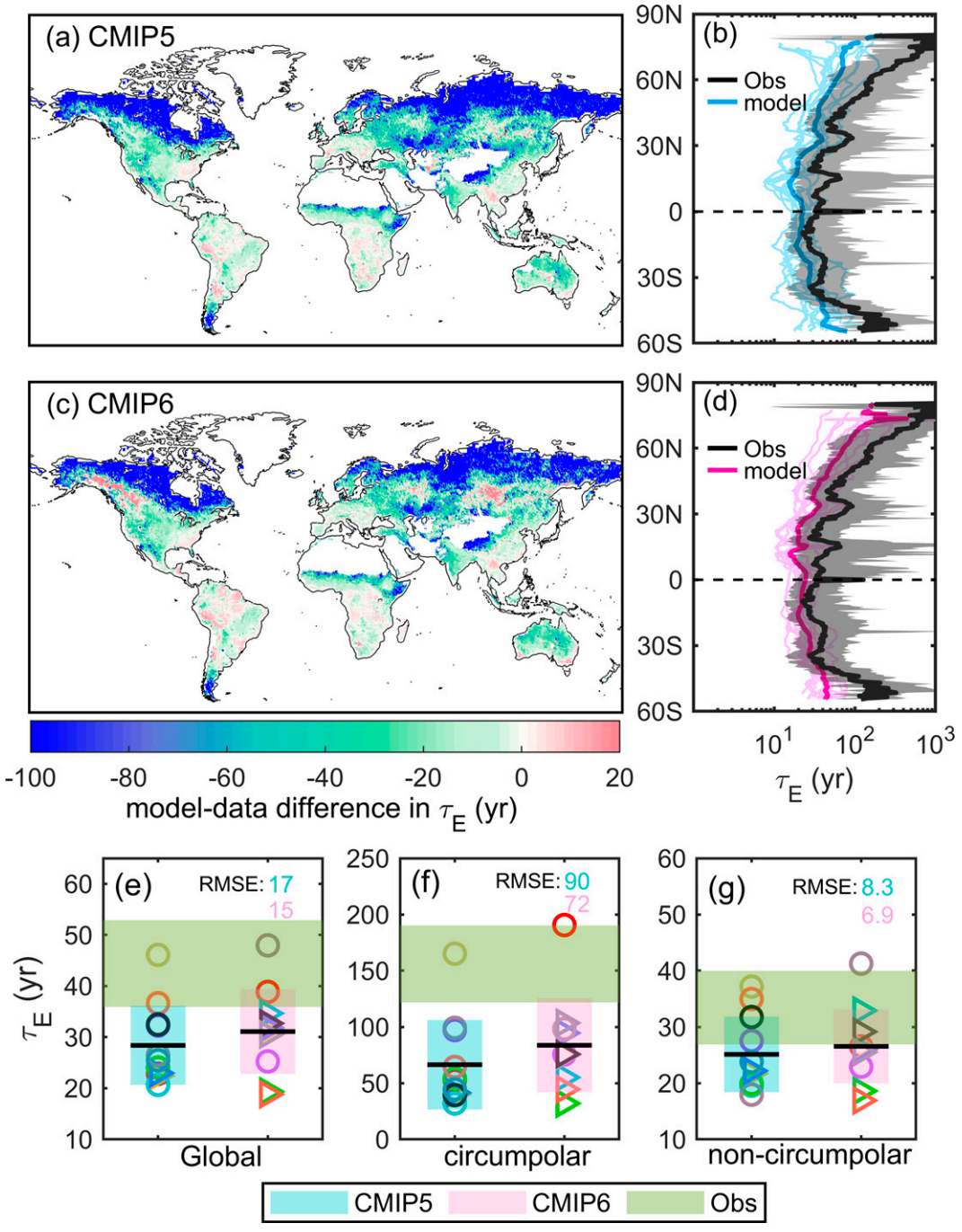

FIG. 5. As in Fig. 4, but for τ_E. The spatial patterns of τ_E for individual models and dataset products are displayed in Figs. S9–S12 in the online supplemental material.

Two CMIP6 models (CESM2 and NorESM2-LM) have applied a vertically resolved model to represent multiple layers of C distribution along with the soil profile. This new feature allows for coupling thaw depth dynamics with vertical soil C in the permafrost region and is promising to improve the permafrost C cycle in models (Koven et al. 2011, 2013). In the two models, the simulated full-depth soil C stock (Fig. S5) and the corresponding τ_E (Fig. S15) are matched with the range of observation-based estimates over the circumpolar region. Other data–model comparison studies have also highlighted the importance of vertical soil profile for C storage in the northern high latitudes (Lawrence et al. 2019; Akihiko et al. 2020). However, the two models with vertical soil biogeochemistry still underestimate the soil C stock at a depth of 1 m

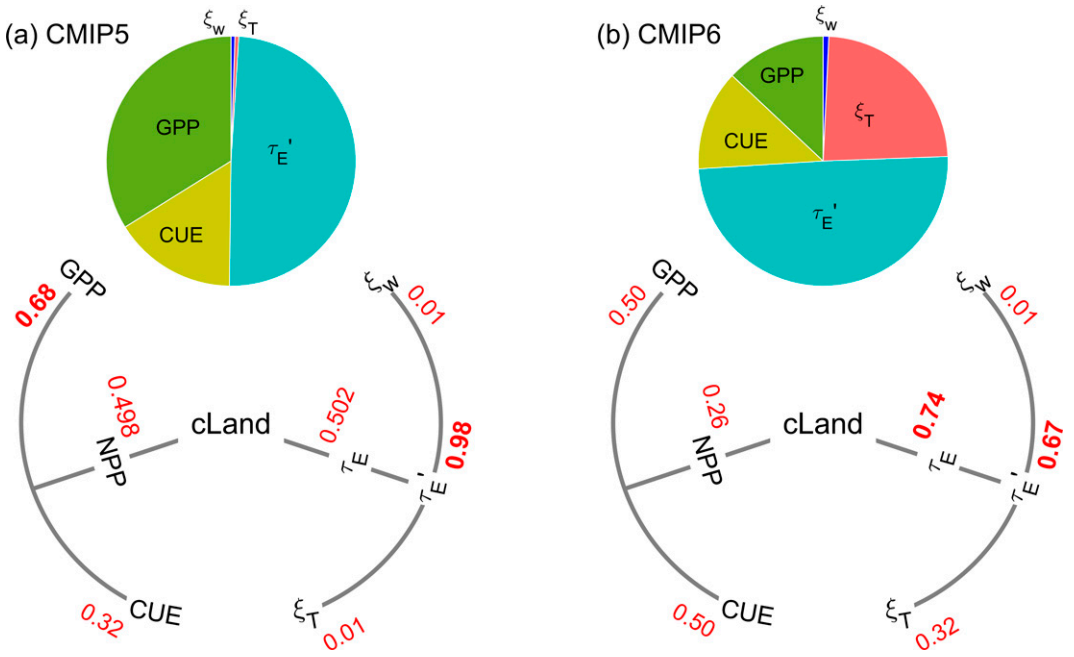

FIG. 6. Attributing the variance of global land C storage among models into different components. Results of traceability analysis on the model spread in land C storage for (a) CMIP5 and (b) CMIP6 (see section 2 on traceability analysis). The circular diagram shows the hierarchical decomposition of terrestrial C storage into its determinants. Land C storage is first decomposed into NPP and τ_E. NPP can further be traced into GPP and CUE, while τ_E can be traced into baseline residence time (τ'_E), temperature scalar (ξ_T), and precipitation scalar (ξ_W). The relative contribution of a component to its upper-level variable is shown next to the line. The above pie shows the relative contributions of GPP, CUE, τ'_E, ξ_T, and ξ_W to the variance of global land C storage among models.

over the circumpolar region. A recent study on soil age reveals that the surface soil C in current depth-resolved models cycles too fast compared to the observation-derived datasets (Shi et al. 2020). Those findings suggest that decomposition rates of surface soil are too fast over the circumpolar region in those models, which might imply an unrealistic representation of temperature sensitivity of surface soil C cycle or surface soil temperature. For example, too much snow over the circumpolar region could lead to high snow insulation effect and further cause fast C turnover at the surface soil. Therefore, a large amount of C cannot stay long at the surface soil but transfer into deeper soil horizons in these models (Shi et al. 2020). More research efforts are still needed to improve the representation of soil C processes, such as temperature sensitivity of surface soil C cycle (Koven et al. 2017; Schädel et al. 2018), vertical C transportation (Koven et al. 2013), and decomposition rates along depths (Koven et al. 2017; Shi et al. 2020).

The CMIP6 models have lower RMSEs than CMIP5 models in global terrestrial C storage and NPP (Figs. 3–5). Process developments in the terrestrial nutrient cycle might play a role in the improvement. In CMIP5, only two models explicitly represent the nutrient limitation on the terrestrial C cycle (Fig. 3a and Table 1). Most CMIP5 models showed positive bias in simulating global terrestrial NPP and land C storage (Fig. 3a). With nutrient limitation incorporated into more models (Fig. 3b and Tables 1 and 2), global terrestrial C storage and NPP simulated by CMIP6 models are smaller and more converged than CMIP5 (Fig. 3). In addition to process developments, an increasing body of data products from in situ and remote sensing is available to inform models (Collier et al. 2018; Davies-Barnard et al. 2020; Randerson et al. 2009). Therefore, the model improvement in productivity processes is expected (Davies-Barnard et al. 2020; Piao et al. 2013). Other recent studies have also revealed that models with nutrient limitations generally agree with global dataset products (Akihiko et al. 2020; Davies-Barnard et al. 2020). In this study, we further show that the improved model performance in NPP contributes to the reduced model spread in terrestrial C storage, whereas the persistent underestimation of τ_E leads to the underestimated land C storage in the two CMIP ensembles.

b. *Uneven model improvement in simulating productivity and CUE*

Consistent with NPP, model spread and bias in GPP also significantly reduced at both regional (Figs. 8c,g and Fig. S18) and global scales (Fig. 7a). The model ensemble in CMIP6 shows a more convergent and less biased CUE at the global scale (Fig. 7), but the model improvement in representing the spatial pattern of CUE is slight (Figs. 8a,e). It probably results from model differences in representing the downregulation of nutrient limitations on NPP, either by decreasing GPP or increasing autotrophic respiration, or both (Meyerholt and Zaehle 2015). Consequently, models converge in estimating

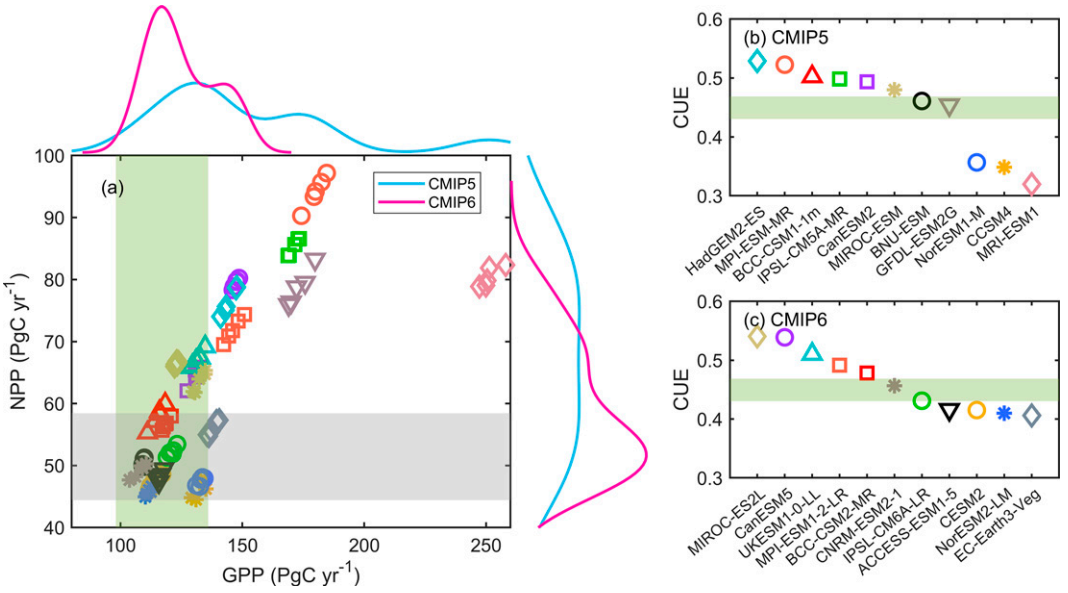

FIG. 7. Model–data comparisons on global terrestrial NPP, GPP, and CUE for two CMIP ensembles. (a) The global terrestrial GPP from 2001 to 2005 was plotted against the corresponding NPP for models in CMIP5 and CMIP6. Distributions along the axis show the probability density function of model simulations in the same CMIP. (b),(c) Temporal average (2001–05 mean) of CUE for individual models in CMIP5 and CMIP6, respectively. The shaded areas in the three panels show the range (from the maximum to the minimum) of observation-based estimates.

productivity but diverge in CUE. On the other hand, the observation-based estimations of CUE are also uncertain due to the difficulty of accurately accounting for autotrophic respiration (Kolby Smith et al. 2016).

A recent study has evaluated nitrogen (N)-related performances of the C–N coupled models in CMIP6 (Davies-Barnard et al. 2020). These C–N models do well at reproducing global GPP and NEP, but significant uncertainty still exists in modeling autotrophic and heterotrophic respiration. As N availability exerts stoichiometric controls on allocation and decomposition processes, the uncertainty of respiration in current C–N coupled models might suggest that the N cycle and C–N coupling schemes are highly uncertain (Davies-Barnard et al. 2020; Du et al. 2018). More process understanding and dataset products are in need to constrain relevant processes. For example, soil N availability regulates many C processes, such as plant C uptake, plant respiration, allocation of NPP, and decomposition of litter and soil organic matter. However, the N cycle itself is relatively open, in which multiple pathways of inputs and outputs determine the N availability (Wei et al. 2019). Davies-Barnard et al. (2020) has found that biological nitrogen fixation drives the significant variation of N inputs among models, while N losses differ strongly among current C–N coupled models. A lack of knowledge leaves space for the variety of hypotheses embedded in models, leading to various model representations of the N cycle and C–N coupling (Du et al. 2018; Meyerholt et al. 2020).

c. *Increasing importance of τ_E in causing model variance in land C storage*

The traceability framework offers us a common yet systematic way to discern the critical process governing across-model variation in terrestrial C storage, even though models had myriad different process representations of the terrestrial C cycle. By decomposing the modeled terrestrial C storage into its determinants in a hierarchical way, we have found the increasing importance of τ_E in causing across-model variation in global land C storage. The increase in fractional variance in τ_E implies two aspects: 1) NPP in CMIP6 has become less uncertain and 2) model spread in τ_E becomes the primary controller of model variation in global land C storage. Further, the variation in τ_E is dominated by τ_E'. The τ_E' is the ecosystem C residence time at a reference environment condition. It usually represents intrinsic properties of vegetation and soil processes (Carvalhais et al. 2014; Cui et al. 2019; Xia et al. 2013), such as allocation of photosynthate, C transfer from plant to soil, and decomposition of soil organic C [Eq. (5)]. The term ξ represents the modification of environmental factors on τ_E' via the temperature scalar (ξ_T) and precipitation scalar (ξ_W). Although the contribution of ξ_T to the across-model variation in τ_E increases from CMIP5 to CMIP6, τ_E' still dominates the variation of τ_E in two CMIP model ensembles. The dominant role of τ_E' implies that the parameterization and model structure of C turnover both differ significantly among models.

As models evolve from CMIP5 to CMIP6, the model structure representing the terrestrial C cycle has become more complex. All ESMs in CMIP5 simulated topsoil C dynamic based on a network system of multiple C pools. The pool numbers ranged from two in CanESM2, BNU-ESM, and MRI-ESM1 to eight in CCSM4 and NorESM1-M (Table 1). Among all the CMIP5 models, only two of them explicitly included terrestrial C–N interactions. In CMIP6, two models (CESM2 and NorESM2-LM) have applied a vertically resolved

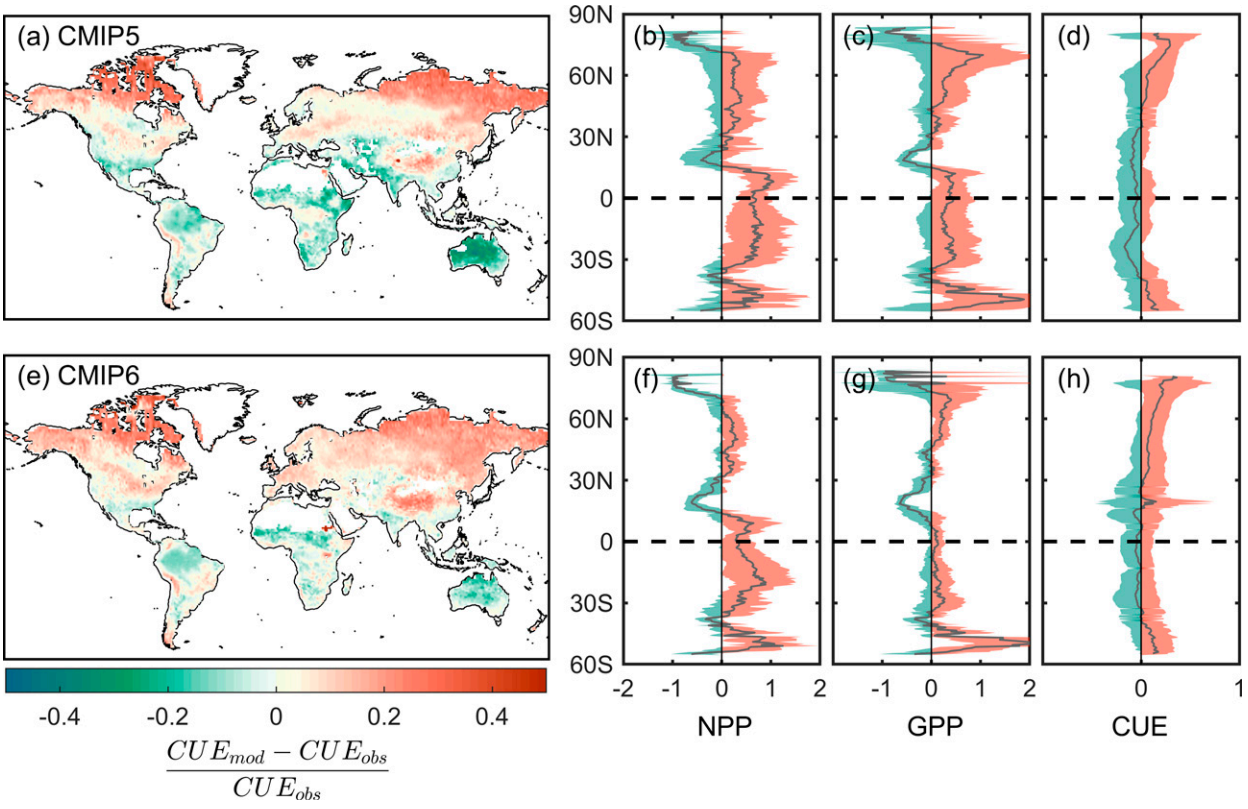

FIG. 8. Model bias and intermodel spread in NPP, GPP, and CUE. (a),(e) The spatial patterns of model bias score in CUE for CMIP5 and CMIP6 model ensembles, respectively, were computed by subtracting the multimodel mean from the observed mean value and then dividing by the observed mean value. (b),(f) Latitudinal patterns of model bias score in NPP for CMIP5 and CMIP6, respectively. (c),(g) As in (b) and (f), but for GPP. (d),(h) As in (b) and (f), but for CUE.

model to represent multiple layers of C distribution along with the soil profile. They have 140 soil and litter C pools (Table 2). Eight of the 11 CMIP6 models selected in our study have incorporated the N cycle. Furthermore, mechanism-rich models are usually "data hungry" (DeAngelis and Mooij 2003) and lack data to constrain relevant processes. Thus, a critical challenge for future CMIP models is how to improve the traceability of uncertainty in the terrestrial C cycle associated with the rapid increase in model complexity.

5. Conclusions

Overall, the converging but persistently underestimated global terrestrial C storage in CMIP6 models results from uneven model improvement in process representation for productivity and residence time. The models in CMIP6 are adequate in capturing the near-present terrestrial productivity, while the modeled τ_E is lower than the dataset products. Thus, τ_E becomes the essential source for model ensemble bias and intermodel spread in terrestrial C storage. From CMIP5 to CMIP6, the more accurate simulation of productivity and persistently biased τ_E could further influence model uncertainty in future projections of the terrestrial C feedback to climate change (Arora et al. 2020; Friedlingstein et al. 2006). Moreover, current observation-derived products show considerable disagreement, calling for more efforts to reduce data uncertainty. The more convergent but biased model simulations also emphasize the need to treat model spread in a more sophisticated way. Many recent studies have paid great attention to reducing model spread (Lovenduski and Bonan 2017; Varney et al. 2020). However, it should be noted that the reduced model spread does not necessarily lead to more reliable model simulations (as shown in Fig. 3) or projections (Bonan et al. 2019; Lovenduski and Bonan 2017). Overall, our findings show an improvement in the terrestrial C cycle in ESMs from CMIP5 to CMIP6 and call for more model evaluations based on dataset products and experimental data in the future.

Acknowledgments. We acknowledge the World Climate Research Programme's working group on Coupled Modelling, which coordinated and promoted CMIPs. We thank the climate modeling groups for producing and making their model outputs available, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies that support CMIP6 and ESGF. We thank T. Wu and J. Yan for sharing BCC-CSM1-1m outputs, and W. Kolby Smith for helping with the interpretation of GIMMS3g NPP GPP data. This work is financially supported

by the National Key R&D Program of China (2017YFA0604600) and the National Natural Science Foundation of China (31722009).

Data availability statement. CMIP5 and CMIP6 model outputs are downloaded from the Earth System Grid Federation (ESGF) data archive (https://esgf-node.llnl.gov/projects/esgf-llnl/). All datasets used in the study are publicly available, and their sources are provided in the text. References to observation-derived data of C stocks in vegetation and soil are summarized in Table 3. All codes and metadata used in this study can be accessed by https://github.com/NingWei-227/Evolution-of-Uncertainty-in-Terrestrial-Carbon-Storage-in-Earth-System-Models-from-CMIP5-to-CMP6.

REFERENCES

Ahlström, A., B. Smith, J. Lindström, M. Rummukainen, and C. B. Uvo, 2013: GCM characteristics explain the majority of uncertainty in projected 21st century terrestrial ecosystem carbon balance. *Biogeosciences*, **10**, 1517–1528, https://doi.org/10.5194/bg-10-1517-2013.

Akihiko, I., and Coauthors, 2020: Soil carbon sequestration simulated in CMIP6-LUMIP models: Implications for climatic mitigation. *Environ. Res. Lett.*, **10**, 1748–9326, https://doi.org/10.1088/1748-9326/abc912.

Anav, A., and Coauthors, 2013: Evaluating the land and ocean components of the global carbon cycle in the CMIP5 Earth system models. *J. Climate*, **26**, 6801–6843, https://doi.org/10.1175/JCLI-D-12-00417.1.

Arneth, A., and Coauthors, 2010: Terrestrial biogeochemical feedbacks in the climate system. *Nat. Geosci.*, **3**, 525–532, https://doi.org/10.1038/ngeo905.

Arora, V. K., and Coauthors, 2009: The effect of terrestrial photosynthesis down regulation on the twentieth-century carbon budget simulated with the CCCma Earth system model. *J. Climate*, **22**, 6066–6088, https://doi.org/10.1175/2009JCLI3037.1.

——, and Coauthors, 2020: Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models. *Biogeosciences*, **17**, 4173–4222, https://doi.org/10.5194/bg-17-4173-2020.

Best, M. J., and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), model description—Part 1: Energy and water fluxes. *Geosci. Model Dev.*, **4**, 677–699, https://doi.org/10.5194/gmd-4-677-2011.

Bloom, A. A., J.-F. Exbrayat, I. R. van der Velde, L. Feng, and M. Williams, 2016: The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proc. Natl. Acad. Sci. USA*, **113**, 1285–1290, https://doi.org/10.1073/pnas.1515160113.

Bodman, R. W., P. J. Rayner, and D. J. Karoly, 2013: Uncertainty in temperature projections reduced using carbon cycle and climate observations. *Nat. Climate Change*, **3**, 725–729, https://doi.org/10.1038/nclimate1903.

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

——, D. L. Lombardozzi, W. R. Wieder, K. W. Oleson, D. M. Lawrence, F. M. Hoffman, and N. Collier, 2019: Model structure and climate data uncertainty in historical simulations of the terrestrial carbon cycle (1850–2014). *Global Biogeochem. Cycles*, **33**, 1310–1326, https://doi.org/10.1029/2019GB006175.

Booth, B. B. B., and Coauthors, 2012: High sensitivity of future global warming to land carbon cycle processes. *Environ. Res. Lett.*, **7**, 024002, https://doi.org/10.1088/1748-9326/7/2/024002.

Bowring, S. P. K., R. Lauerwald, B. Guenet, D. Zhu, M. Guimberteau, A. Tootchi, A. Ducharne, and P. Ciais, 2019: ORCHIDEE MICT-LEAK (r5459), a global model for the production, transport, and transformation of dissolved organic carbon from Arctic permafrost regions—Part 1: Rationale, model description, and simulation protocol. *Geosci. Model Dev.*, **12**, 3503–3521, https://doi.org/10.5194/gmd-12-3503-2019.

Bradford, M. A., and T. W. Crowther, 2013: Carbon use efficiency and storage in terrestrial ecosystems. *New Phytol.*, **199**, 7–9, https://doi.org/10.1111/nph.12334.

Burke, E. J., and Coauthors, 2017: Quantifying uncertainties of permafrost carbon–climate feedbacks. *Biogeosciences*, **14**, 3051–3066, https://doi.org/10.5194/bg-14-3051-2017.

Carvalhais, N., and Coauthors, 2014: Global covariation of carbon turnover times with climate in terrestrial ecosystems. *Nature*, **514**, 213–217, https://doi.org/10.1038/nature13731.

Chevan, A., and M. Sutherland, 1991: Hierarchical partitioning. *Amer. Stat.*, **45**, 90–96, https://doi.org/10.2307/2684366.

Ciais, P., and Coauthors, 2014: Carbon and other biogeochemical cycles. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 465–570.

Clark, D. B., and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), model description—Part 2: Carbon fluxes and vegetation dynamics. *Geosci. Model Dev.*, **4**, 701–722, https://doi.org/10.5194/gmd-4-701-2011.

Collier, N., and Coauthors, 2018: The International Land Model Benchmarking (ILAMB) system: Design, theory, and implementation. *J. Adv. Model. Earth Syst.*, **10**, 2731–2754, https://doi.org/10.1029/2018MS001354.

Collins, W. J., and Coauthors, 2011: Development and evaluation of an Earth-system model—HadGEM2. *Geosci. Model Dev.*, **4**, 1051–1075, https://doi.org/10.5194/gmd-4-1051-2011.

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

Cui, E., and Coauthors, 2019: Vegetation functional properties determine uncertainty of simulated ecosystem productivity: A traceability analysis in the East Asian monsoon region. *Global Biogeochem. Cycles*, **33**, 668–689, https://doi.org/10.1029/2018GB005909.

Davies-Barnard, T., and Coauthors, 2020: Nitrogen cycling in CMIP6 land surface models: Progress and limitations. *Biogeosciences*, **17**, 5129–5148, https://doi.org/10.5194/bg-17-5129-2020.

DeAngelis, D. L., and W. M. Mooij, 2003: In praise of mechanistically rich models. *Models in Ecosystem Science*, C. D. Canham, J. J. Cole, and W. K. Lauenroth, Eds., Princeton University Press, 63–82.

Du, Z., E. Weng, L. Jiang, Y. Luo, J. Xia, and X. Zhou, 2018: Carbon–nitrogen coupling under three schemes of model representation: A traceability analysis. *Geosci. Model Dev.*, **11**, 4399–4416, https://doi.org/10.5194/gmd-11-4399-2018.

Dufresne, J. L., and Coauthors, 2013: Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5. *Climate Dyn.*, **40**, 2123–2165, https://doi.org/10.1007/s00382-012-1636-1.

Dunne, J. P., and Coauthors, 2012: GFDL's ESM2 global coupled climate–carbon Earth system models. Part II: Carbon system formulation and baseline simulation characteristics. *J. Climate*, **26**, 2247–2267, https://doi.org/10.1175/JCLI-D-12-00150.1.

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

Fan, N., and Coauthors, 2020: Apparent ecosystem carbon turnover time: Uncertainties and robust features. *Earth Syst. Sci. Data*, **12**, 2517–2536, https://doi.org/10.5194/essd-12-2517-2020.

Flato, G. M., 2011: Earth system models: An overview. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 783–800, https://doi.org/10.1002/wcc.148.

Friedlingstein, P., and Coauthors, 2006: Climate–carbon cycle feedback analysis: Results from the C^4MIP model intercomparison. *J. Climate*, **19**, 3337–3353, https://doi.org/10.1175/JCLI3800.1.

——, M. Meinshausen, V. K. Arora, C. D. Jones, A. Anav, S. K. Liddicoat, and R. Knutti, 2013: Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *J. Climate*, **27**, 511–526, https://doi.org/10.1175/JCLI-D-12-00579.1.

——, and Coauthors, 2019: Global carbon budget 2019. *Earth Syst. Sci. Data*, **11**, 1783–1838, https://doi.org/10.5194/essd-11-1783-2019.

Gent, P. R., and Coauthors, 2011: The Community Climate System Model version 4. *J. Climate*, **24**, 4973–4991, https://doi.org/10.1175/2011JCLI4083.1.

Giorgetta, M. A., and Coauthors, 2013: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *J. Adv. Model. Earth Syst.*, **5**, 572–597, https://doi.org/10.1002/jame.20038.

Goll, D. S., V. Brovkin, B. R. Parida, C. H. Reick, J. Kattge, P. B. Reich, P. M. van Bodegom, and Ü. Niinemets, 2012: Nutrient limitation reduces land carbon uptake in simulations with a model of combined carbon, nitrogen and phosphorus cycling. *Biogeosciences*, **9**, 3547–3569, https://doi.org/10.5194/bg-9-3547-2012.

Hajima, T., and Coauthors, 2019: Development of the MIROC-ES2L Earth system model and evaluation of its climate–biogeochemical processes and feedbacks. *Geosci. Model Dev.*, **13**, 2197–2244, https://doi.org/10.5194/gmd-13-2197-2020.

Hengl, T., and Coauthors, 2017: SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, **12**, e0169748, https://doi.org/10.1371/journal.pone.0169748.

Hoesly, R. M., and Coauthors, 2018: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). *Geosci. Model Dev.*, **11**, 369–408, https://doi.org/10.5194/gmd-11-369-2018.

Houghton, R. A., 2007: Balancing the global carbon budget. *Annu. Rev. Earth Planet. Sci.*, **35**, 313–347, https://doi.org/10.1146/annurev.earth.35.031306.140057.

Huang, Y., and Coauthors, 2018: Matrix approach to land carbon cycle modeling: A case study with the Community Land Model. *Global Change Biol.*, **24**, 1394–1404, https://doi.org/10.1111/gcb.13948.

Hugelius, G., and Coauthors, 2013: A new data set for estimating organic carbon storage to 3 m depth in soils of the northern circumpolar permafrost region. *Earth Syst. Sci. Data*, **5**, 393–402, https://doi.org/10.5194/essd-5-393-2013.

Ito, A., and T. Oikawa, 2002: A simulation model of the carbon cycle in land ecosystems (Sim-CYCLE): A description based on dry-matter production theory and plot-scale validation. *Ecol. Modell.*, **151**, 143–176, https://doi.org/10.1016/S0304-3800(01)00473-2.

Ji, D., and Coauthors, 2014: Description and basic evaluation of Beijing Normal University Earth System Model (BNU-ESM) version 1. *Geosci. Model Dev.*, **7**, 2039–2064, https://doi.org/10.5194/gmd-7-2039-2014.

Jiang, L., Z. Shi, J. Xia, J. Liang, X. Lu, Y. Wang, and Y. Luo, 2017: Transient traceability analysis of land carbon storage dynamics: Procedures and its application to two forest ecosystems. *J. Adv. Model. Earth Syst.*, **9**, 2822–2835, https://doi.org/10.1002/2017MS001004.

Jones, C. D., and Coauthors, 2016: C4MIP—The Coupled Climate–Carbon Cycle Model Intercomparison Project: Experimental protocol for CMIP6. *Geosci. Model Dev.*, **9**, 2853–2880, https://doi.org/10.5194/gmd-9-2853-2016.

Jung, M., and Coauthors, 2017: Compensatory water effects link yearly global land CO_2 sink changes to temperature. *Nature*, **541**, 516–520, https://doi.org/10.1038/nature20780.

Keenan, T. F., and C. A. Williams, 2018: The terrestrial carbon sink. *Annu. Rev. Environ. Resour.*, **43**, 219–243, https://doi.org/10.1146/annurev-environ-102017-030204.

Kolby Smith, W., and Coauthors, 2016: Large divergence of satellite and Earth system model estimates of global terrestrial CO_2 fertilization. *Nat. Climate Change*, **6**, 306–310, https://doi.org/10.1038/nclimate2879.

Koven, C. D., P. Friedlingstein, P. Ciais, D. Khvorostyanov, G. Krinner, and C. Tarnocai, 2009: On the formation of high-latitude soil carbon stocks: Effects of cryoturbation and insulation by organic matter in a land surface model. *Geophys. Res. Lett.*, **36**, L21501, https://doi.org/10.1029/2009GL040150.

——, B. Ringeval, P. Friedlingstein, P. Ciais, P. Cadule, D. Khvorostyanov, G. Krinner, and C. Tarnocai, 2011: Permafrost carbon-climate feedbacks accelerate global warming. *Proc. Natl. Acad. Sci. USA*, **108**, 14 769–14 774, https://doi.org/10.1073/pnas.1103910108.

——, and Coauthors, 2013: The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4. *Biogeosciences*, **10**, 7109–7131, https://doi.org/10.5194/bg-10-7109-2013.

——, and Coauthors, 2015: Controls on terrestrial carbon feedbacks by productivity versus turnover in the CMIP5 Earth system models. *Biogeosciences*, **12**, 5211–5228, https://doi.org/10.5194/bg-12-5211-2015.

——, G. Hugelius, D. M. Lawrence, and W. R. Wieder, 2017: Higher climatological temperature sensitivity of soil carbon in cold than warm climates. *Nat. Climate Change*, **7**, 817–822, https://doi.org/10.1038/nclimate3421.

Krinner, G., and Coauthors, 2005: A dynamic global vegetation model for studies of the coupled atmosphere–biosphere system. *Global Biogeochem. Cycles*, **19**, GB1015, https://doi.org/10.1029/2003GB002199.

Law, R. M., and Coauthors, 2017: The carbon cycle in the Australian Community Climate and Earth System Simulator (ACCESS-ESM1)—Part 1: Model description and pre-industrial simulation. *Geosci. Model Dev.*, **10**, 2567–2590, https://doi.org/10.5194/gmd-10-2567-2017.

Lawrence, D. M., and Coauthors, 2016: The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: Rationale and experimental design. *Geosci. Model Dev.*, **9**, 2973–2998, https://doi.org/10.5194/gmd-9-2973-2016.

——, and Coauthors, 2019: The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *J. Adv. Model. Earth Syst.*, **11**, 4245–4287, https://doi.org/10.1029/2018MS001583.

Lindsay, K., and Coauthors, 2014: Preindustrial-control and twentieth-century carbon cycle experiments with the Earth System Model CESM1(BGC). *J. Climate*, **27**, 8981–9005, https://doi.org/10.1175/JCLI-D-12-00565.1.

Lovenduski, N. S., and G. B. Bonan, 2017: Reducing uncertainty in projections of terrestrial carbon uptake. *Environ. Res. Lett.*, **12**, 044020, https://doi.org/10.1088/1748-9326/aa66b8.

Luo, Y., and E. Weng, 2011: Dynamic disequilibrium of the terrestrial carbon cycle under global change. *Trends Ecol. Evol.*, **26**, 96–104, https://doi.org/10.1016/j.tree.2010.11.003.

——, and Coauthors, 2017: Transient dynamics of terrestrial carbon storage: Mathematical foundation and its applications. *Biogeosciences*, **14**, 145–161, https://doi.org/10.5194/bg-14-145-2017.

Mauritsen, T., and Coauthors, 2019: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2. *J. Adv. Model. Earth Syst.*, **11**, 998–1038, https://doi.org/10.1029/2018MS001400.

Meyerholt, J., and S. Zaehle, 2015: The role of stoichiometric flexibility in modelling forest ecosystem responses to nitrogen fertilization. *New Phytol.*, **208**, 1042–1055, https://doi.org/10.1111/nph.13547.

——, K. Sickel, and S. Zaehle, 2020: Ensemble projections elucidate effects of uncertainty in terrestrial nitrogen limitation on future carbon uptake. *Global Change Biol.*, **26**, 3978–3996, https://doi.org/10.1111/gcb.15114.

Muhr, J., A. Angert, R. I. Negrón-Juárez, W. A. Muñoz, G. Kraemer, J. Q. Chambers, and S. E. Trumbore, 2013: Carbon dioxide emitted from live stems of tropical trees is several years old. *Tree Physiol.*, **33**, 743–752, https://doi.org/10.1093/treephys/tpt049.

Murray, K., and M. M. Conner, 2009: Methods to quantify variable importance: Implications for the analysis of noisy ecological data. *Ecology*, **90**, 348–355, https://doi.org/10.1890/07-1929.1.

Nottingham, A. T., P. Meir, E. Velasquez, and B. L. Turner, 2020: Soil carbon loss by experimental warming in a tropical forest. *Nature*, **584**, 234–237, https://doi.org/10.1038/s41586-020-2566-4.

Piao, S., and Coauthors, 2013: Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO_2 trends. *Global Change Biol.*, **19**, 2117–2132, https://doi.org/10.1111/gcb.12187.

Randerson, J. T., and Coauthors, 2009: Systematic assessment of terrestrial biogeochemistry in coupled climate–carbon models. *Global Change Biol.*, **15**, 2462–2484, https://doi.org/10.1111/j.1365-2486.2009.01912.x.

Ruesch, A., and H. K. Gibbs, 2008: New IPCC Tier-1 global biomass carbon map for the year 2000. Carbon Dioxide Information Analysis Center, https://cdiac.ess-dive.lbl.gov/epubs/ndp/global_carbon/carbon_documentation.html.

Running, S., Q. Mu, and M. Zhao, 2015: MOD17A2H MODIS/Terra gross primary productivity 8-day L4 global 500m SIN grid V006. NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MODIS/MOD17A2H.006.

Sato, H., A. Itoh, and T. Kohyama, 2007: SEIB–DGVM: A new Dynamic Global Vegetation Model using a spatially explicit individual-based approach. *Ecol. Modell.*, **200**, 279–307, https://doi.org/10.1016/j.ecolmodel.2006.09.006.

Schädel, C., and Coauthors, 2018: Divergent patterns of experimental and model-derived permafrost ecosystem carbon dynamics in response to Arctic warming. *Environ. Res. Lett.*, **13**, 105002, https://doi.org/10.1088/1748-9326/aae0ff.

Schneck, R., C. H. Reick, and T. Raddatz, 2013: Land contribution to natural CO_2 variability on time scales of centuries. *J. Adv. Model. Earth Syst.*, **5**, 354–365, https://doi.org/10.1002/jame.20029.

Schuur, E. A. G., and Coauthors, 2015: Climate change and the permafrost carbon feedback. *Nature*, **520**, 171–179, https://doi.org/10.1038/nature14338.

Séférian, R., and Coauthors, 2019: Evaluation of CNRM Earth-system model, CNRM-ESM 2-1: Role of Earth system processes in present-day and future climate. *J. Adv. Model. Earth Syst.*, **11**, 4182–4227, https://doi.org/10.1029/2019MS001791.

Seland, Ø., and Coauthors, 2020: Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations. *Geosci. Model Dev.*, **13**, 6165–6200, https://doi.org/10.5194/gmd-13-6165-2020.

Sierra, C. A., and Coauthors, 2022: Ideas and perspectives: Allocation of carbon from net primary production in models is inconsistent with observations of the age of respired carbon. EGUsphere, https://doi.org/10.5194/egusphere-2022-34, in press.

Sitch, S., and Coauthors, 2003: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biol.*, **9**, 161–185, https://doi.org/10.1046/j.1365-2486.2003.00569.x.

Shi, Z., and Coauthors, 2020: The age distribution of global soil carbon inferred from radiocarbon measurements. *Nat. Geosci.*, **13**, 555–559, https://doi.org/10.1038/s41561-020-0596-z.

Smith, B., D. Wårlind, A. Arneth, T. Hickler, P. Leadley, J. Siltberg, and S. Zaehle, 2014: Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model. *Biogeosciences*, **11**, 2027–2054, https://doi.org/10.5194/bg-11-2027-2014.

Spawn, S. A., C. C. Sullivan, T. J. Lark, and H. K. Gibbs, 2020: Harmonized global maps of above and belowground biomass carbon density in the year 2010. *Sci. Data*, **7**, 112, https://doi.org/10.1038/s41597-020-0444-4.

Swart, N. C., and Coauthors, 2019: The Canadian Earth System Model version 5 (CanESM5.0.3). *Geosci. Model Dev.*, **12**, 4823–4873, https://doi.org/10.5194/gmd-12-4823-2019.

Taylor, K. E., R. J. surname>Stouffer, and G. A. Meehl, 2011: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.

Tjiputra, J. F., C. Roelandt, M. Bentsen, D. M. Lawrence, T. Lorentzen, J. Schwinger, Ø. Seland, and C. Heinze, 2013: Evaluation of the carbon cycle components in the Norwegian Earth System Model (NorESM). *Geosci. Model Dev.*, **6**, 301–325, https://doi.org/10.5194/gmd-6-301-2013.

Todd-Brown, K. E. O., J. T. Randerson, W. M. Post, F. M. Hoffman, C. Tarnocai, E. A. G. Schuur, and S. D. Allison, 2013: Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences*, **10**, 1717–1736, https://doi.org/10.5194/bg-10-1717-2013.

——, and Coauthors, 2014: Changes in soil organic carbon storage predicted by Earth system models during the 21st century. *Biogeosciences*, **11**, 2341–2356, https://doi.org/10.5194/bg-11-2341-2014.

Varney, R. M., S. E. Chadburn, P. Friedlingstein, E. J. Burke, C. D. Koven, G. Hugelius, and P. M. Cox, 2020: A spatial emergent constraint on the sensitivity of soil carbon turnover to global warming. *Nat. Commun.*, **11**, 5544, https://doi.org/10.1038/s41467-020-19208-8.

Vuichard, N., P. Messina, S. Luyssaert, B. Guenet, S. Zaehle, J. Ghattas, V. Bastrikov, and P. Peylin, 2019: Accounting for carbon and nitrogen interactions in the global terrestrial ecosystem model ORCHIDEE (trunk version, rev 4999): Multi-scale evaluation of gross primary production. *Geosci. Model Dev.*, **12**, 4751–4779, https://doi.org/10.5194/gmd-12-4751-2019.

Wang, J., and Coauthors, 2019: Evaluating the simulated mean soil carbon transit times by Earth system models using observations. *Biogeosciences*, **16**, 917–926, https://doi.org/10.5194/bg-16-917-2019.

Watanabe, S., and Coauthors, 2011: MIROC-ESM 2010: Model description and basic results of CMIP5-20c3m experiments. *Geosci. Model Dev.*, **4**, 845–872, https://doi.org/10.5194/gmd-4-845-2011.

Wei, N., and Coauthors, 2019: Decadal stabilization of soil inorganic nitrogen as a benchmark for global land models. *J. Adv. Model. Earth Syst.*, **11**, 1088–1099, https://doi.org/10.1029/2019MS001633.

Wieder, W., 2014: Regridded Harmonized World Soil Database v1.2. Data set. Oak Ridge National Laboratory Distributed Active Archive Center, http://daac.ornl.gov.

——, C. C. Cleveland, W. K. Smith, and K. Todd-Brown, 2015: Future productivity and carbon storage limited by terrestrial nutrient availability. *Nat. Geosci.*, **8**, 441–444, https://doi.org/10.1038/ngeo2413.

Wu, T., and Coauthors, 2013: Global carbon budgets simulated by the Beijing Climate Center Climate System Model for the last century. *J. Geophys. Res. Atmos.*, **118**, 4326–4347, https://doi.org/10.1002/jgrd.50320.

——, and Coauthors, 2019: The Beijing Climate Center Climate System Model (BCC-CSM): The main progress from CMIP5 to CMIP6. *Geosci. Model Dev.*, **12**, 1573–1600, https://doi.org/10.5194/gmd-12-1573-2019.

Xia, J., Y. Luo, Y.-P. Wang, and O. Hararuk, 2013: Traceable components of terrestrial carbon storage capacity in biogeochemical models. *Global Change Biol.*, **19**, 2104–2116, https://doi.org/10.1111/gcb.12172.

——, J. Chen, S. Piao, P. Ciais, Y. Luo, and S. Wan, 2014: Terrestrial carbon cycle affected by non-uniform climate warming. *Nat. Geosci.*, **7**, 173–180, https://doi.org/10.1038/ngeo2093.

——, and Coauthors, 2017: Terrestrial ecosystem model performance in simulating productivity and its vulnerability to climate change in the northern permafrost region. *J. Geophys. Res. Biogeosci.*, **122**, 430–446, https://doi.org/10.1002/2016JG003384.

——, J. Wang, and S. Niu, 2020: Research challenges and opportunities for using big data in global change biology. *Global Change Biol.*, **26**, 6040–6061, https://doi.org/10.1111/gcb.15317.

Xu, L., and Coauthors, 2021: Changes in global terrestrial live biomass over the 21st century. *Sci. Adv.*, **7**, eabe9829, https://doi.org/10.1126/sciadv.abe9829.

Yukimoto, S., and Coauthors, 2011: Meteorological Research Institute-Earth System Model version 1 (MRI-ESM1)—Model description. Tech. Rep. 64, 96 pp., http://www.mri-jma.go.jp/Publish/Technical/DATA/VOL_64/index_en.html.

Zhang, Y., X. Xiao, X. Wu, S. Zhou, G. Zhang, Y. Qin, and J. Dong, 2017: A global moderate resolution dataset of gross primary production of vegetation for 2000–2016. *Sci. Data*, **4**, 170165. https://doi.org/10.1038/sdata.2017.165.

Zhou, S., and Coauthors, 2018: Sources of uncertainty in modeled land carbon storage within and across three MIPs: Diagnosis with three new techniques. *J. Climate*, **31**, 2833–2851, https://doi.org/10.1175/JCLI-D-17-0357.1.