COMMENTARY

Assessing carbon cycle projections from complex and simple models under SSP scenarios

Irina Melnikova[1,2] · Philippe Ciais[1] · Olivier Boucher[3] · Katsumasa Tanaka[1,2]

Abstract

Both full-fledged Earth system models (ESMs) and simple climate models (SCMs) have been used to investigate climate change for future representative CO_2 concentration pathways under the sixth phase of the Coupled Model Intercomparison Project. Here, we explore to what extent complex and simple models are consistent in their carbon cycle response in concentration-driven simulations. Although ESMs and SCMs exhibit similar compatible fossil fuel CO_2 emissions, ESMs systematically estimate a lower ocean carbon uptake than SCMs in the historical period and future scenarios. The ESM and SCM differences are especially large under low-concentration and overshoot scenarios. Furthermore, ESMs and SCMs deviate in their land carbon uptake estimates, but the differences are scenario-dependent. These differences are partly driven by a few model outliers (ESMs and SCMs) and the procedure of observational constraining that is present in the majority of SCMs but not applied in ESMs. The differences in land uptake arise from the difference in the way land-use change (LUC) emissions are calculated and different assumptions on how the carbon cycle feedbacks are defined, possibly reflecting the treatment of nitrogen limitation of biomass growth and historical calibration of SCMs. The differences in ocean uptake, which are especially large in overshoot scenarios, may arise from the faster mixing of carbon from the surface to the deep ocean in SCMs than in ESMs. We also discuss the inconsistencies that arise when converting CO_2 emissions from integrated assessment models (IAMs) to CO_2 concentrations inputs for ESMs, which typically rely on a single SCM. We further highlight the discrepancies in LUC emission estimates between models of different complexity, particularly ESMs and IAMs, and encourage climate modeling groups to address these potential areas for model improvement.

✉ Irina Melnikova
  melnikova.irina@nies.go.jp

✉ Katsumasa Tanaka
  katsumasa.tanaka@lsce.ipsl.fr

1   Laboratoire des Sciences du Climat et de l'Environnement (LSCE), IPSL, CEA/CNRS/UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France

2   Earth System Division, National Institute for Environmental Studies (NIES), Tsukuba, Japan

3   Institut Pierre-Simon Laplace (IPSL), Sorbonne Université / CNRS, Paris, France

1 Introduction

A hierarchy of models, including ESMs and SCMs, have been used during the preparation of the Intergovernmental Panel on Climate Change Sixth Assessment Report (IPCC AR6) (Masson-Delmotte et al. 2021; Shukla et al. 2022). SCMs benefit from computational efficiency, while ESMs have a more detailed representation of complex Earth system processes. SCMs allow extending simulations by ESMs for a larger set of scenarios, but their fidelity to the more complex models is a key to this modeling strategy. SCMs can be calibrated against either observations or ESMs, or both. Phases 1 and 2 of the Reduced Complexity Model Intercomparison Project (RCMIP) provide an overview of the approaches used to constrain SCMs against observations to reproduce responses of the Earth system variables, e.g., temperature and ocean heat uptake (Nicholls et al. 2020, 2021). They show that SCMs constrained to observations against key benchmarks perform well in estimating these benchmarks (Nicholls et al. 2021). This leads, particularly, to lower estimates of future increases in global surface air temperature (GSAT) compared to the sixth phase of the Coupled Model Intercomparison Project (CMIP6) ESMs that are not constrained (Eyring et al. 2016). At the same time, applying observationally based warming constraints to the CMIP6 ESMs reduces the ensemble mean of GSAT estimates (Tokarska et al. 2020). Calibrating SCMs against a single or a range of ESMs, i.e., using them as climate model emulators, enables replicating and interpolating the climate responses of complex ESMs in a large set of scenarios (Nicholls et al. 2020; Forster et al. 2021). Yet, SCMs can diverge from ESMs because they lack the model structure needed to represent some particular climate process or climate feedback.

Previous studies, to some extent, examined the consistency between SCMs and ESMs (Joos et al. 2013; Nicholls et al. 2020, 2021; Liddicoat et al. 2021). Among CMIP6 generation model-based studies, RCMIP evaluated the SCMs with a focus on the response of climate variables (Nicholls et al. 2020, 2021). Yet, its comparison to ESMs remained limited, because the carbon cycle was examined only via the transient climate response to cumulative carbon emissions (TCRE). Liddicoat et al. (2021) focused on evaluating CMIP6 ESMs using the concentration-driven Shared Socioeconomic Pathways (SSP) simulations and compared their fossil fuel (FF) emissions compatible with the prescribed CO_2 concentration pathway with the FF emissions that are generated by integrated assessment models (IAMs) and harmonized to make those SSP scenarios (Gidden et al. 2019). However, SCMs were out of the scope of their study. Joos et al. (2013) evaluated the responses of Earth system models of different complexities to a CO_2 emission pulse. However, the authors focused on idealized emission-driven simulations and provided limited details on the land and ocean carbon cycle processes that induce uncertainties in global warming and temperature change. Limited research has been performed to compare the carbon cycle responses between CMIP6 ESMs and SCMs, with Quilcaille et al. (2023) being an exception.

Concentration-driven simulations, as opposed to their emission-driven counterparts, ensure consistency between background CO_2 concentrations across models and thus enable a more consistent analysis of carbon cycle processes under the same CO_2 concentrations (Gregory et al. 2009; Arora et al. 2020). In concentration-driven simulations, compatible FF CO_2 emissions (E_{FF}) can be calculated as the difference between the prescribed atmospheric CO_2 growth rate (G_{Atm}) and the estimated net ocean- (S_{Ocean}) and land- (S_{Land}) carbon fluxes, accounting for land-use change (LUC) emissions (Liddicoat et al. 2021):

$$E_{FF} = G_{Atm} + S_{Land} + S_{Ocean} \tag{1}$$

The consistency of compatible emission estimates between SCMs and ESMs is crucial because SCMs are widely used in climate negotiations, e.g., when assessing the adequacy (or lack) of the nationally determined contributions with respect to climate objectives (Tanaka and O'Neill 2018; Shukla et al. 2022). Here, we explore whether prescribed CO_2 concentration trajectories in SCMs and ESMs result in consistent carbon cycle fluxes, including but not limited to compatible emissions. We analyze the concentration-driven outputs of temperature and land and ocean carbon cycle fluxes simulated by RCMIP SCMs (Table 1) and CMIP6 ESMs (Table 2) under the eight available SSPs. The SSPs are developed by IAMs that describe the social and economic components that can determine future climate change (Ackerman et al. 2009; van Vuuren et al. 2017) (Table 3, Fig. 1). In CMIP6, the emissions of different greenhouse gases (GHGs) from IAMs are then harmonized for the base year to their respective values in the historical inventories using the Aneris software (Gidden et al. 2019) and converted to concentrations by the SCM MAGICC version 7 (Fig. 1c). The land-use and land cover data from IAMs are converted to a readable gridded format suitable for ESMs within the harmonization of global land-use change and management version 2 (LUH2) project (Hurtt et al. 2017, 2020).

In this study, we tackle the following research questions:

1) Are the historical carbon fluxes estimated by ESMs and SCMs consistent between each other and with observations?

2) How do the future carbon fluxes estimated by ESMs and SCMs under SSPs compare against each other?

3) What are the sources of inconsistencies between carbon fluxes estimated by ESMs and SCMs?

This study concludes with a set of recommendations for improving the representation and consistency of the carbon cycle processes in complex and simple models.

2 Methods

2.1 Data

We analyzed the outputs of concentration-driven SCMs (Table 1) and ESMs (Tables 2, S3) simulations under eight SSPs of ScenarioMIP: SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP4-3.4, SSP4-6.0, SSP5-3.4-OS, and SSP5-8.5 (O'Neill et al. 2014, 2016; Riahi et al. 2017) (Table 3) that were available from the RCMIP Phases 1 and 2 archives (Nicholls et al. 2020, 2021). We used the following variables of seven SCMs: surface air temperature change, net ocean-to-atmosphere flux CO_2, net land-to-atmosphere flux CO_2 (that is "natural" land sink without accounting for LUC emissions), net primary production (NPP), and agriculture, forestry and other land use (AFOLU) emissions. In further calculations, we used the difference between net land-to-atmosphere flux CO_2 and AFOLU emissions as S_{Land} and net ocean-to-atmosphere flux CO_2 as S_{Ocean}, both positive sinks to land/ocean. SCMs were observationally constrained in the historical calibration and probabilistic set-ups (see Table 1, description papers, and Nicholls et al. (2021)).

We used the following variables of nine ESMs (one ensemble member for each ESM): surface air temperature (*tas*), net biome production, NBP (*nbp*), gas exchange flux of CO_2 (*fgco2*), net carbon mass flux into atmosphere due to land-use change (*fLuc*), gross primary

Table 1 List of RCMIP Phases 1 and 2 models analyzed in this study and their characteristics

Model	Reference	Configuration	Setup description	Carbon cycle component	Historical constraining	Short name
SCMs of RCMIP Phase 1						
ACC2	(Tanaka and O'Neill 2018)	Historical calibration	ECS* = 3K, all other parameters optimized based on prior information of historical data and parameter estimates	Four-box model for the ocean–atmosphere system and four-box model for land	Uncertain parameters in the carbon cycle, atmospheric chemistry, and physical climate system are optimized based on an inverse estimation method that considers historical observations and their uncertainties (Tanaka et al. 2009)	S1
Hector	(Hartin et al. 2015; Vega-Westhoff et al. 2019)	Default	ECS = 3K	Three-box land, and four-box ocean	CO_2 concentration, GSAT anomaly, total ocean and land carbon fluxes, and total radiative forcing are constrained with observations	S2
		Histcalib	Probabilistic setup, calibrated against historical observations			S9
MCE-v.1-1	(Tsutsui 2020)	Default	CMIP5 25-model mean	IRF for land and ocean carbon cycle	Carbon cycle parameters are tuned to historical observations and mean CMIP5 ESMs	S10
		ECS3K	ECS = 3K			S11
		Prob-50th	Probabilistic setup, representative TCRE of 1.65 K per 1000 GtC			S12

Table 1 (continued)

Model	Reference	Configuration	Setup description	Carbon cycle component	Historical constraining	Short name
OSCARv3.1	(Gasser et al. 2017)		ECS = 3K	Mixed-layer response function for ocean, five biomes and five regions in three-box model for land carbon cycle	Ocean carbon-cycle module: the stratification effect calibrated on CMIP5 models, terrestrial carbon-cycle calibrated on many TRENDY and CMIP5 models	S6
WASP-v2	(Goodwin 2018)	Prob-ECS3	ECS = 3K	Eight-box model: one for atmosphere, ocean surface mixed layer, four ocean boxes, vegetation, and soil boxes		S13
		Prob-hist	Probabilistic setup, calibrated against historical observations		GSAT anomaly, ocean heat uptake, and ocean carbon uptake posteriors are observationally constrained	S14
SCMs of RCMIP Phase 2						
MAGICCv7.5.1	(Meinshausen et al. 2011a, 2011b)		Probabilistic setup, climate feedback is calibrated against CMIP6 ESMs	Globally integrated three-box land model and IRF-based ocean carbon cycle	Constrained by C4MIP carbon cycle responses, GSAT anomaly, ocean heat uptake, radiative forcing observationally constrained	S3
MCE-v1-2	(Tsutsui 2020)		Probabilistic setup, calibrated against CMIP ESMs		Ocean carbon flux, parameters for land CO_2 fertilization and climate sensitivity of respiration tuned to CMIP5, CMIP6 ESMs	S5

Table 1 (continued)

Model	Reference	Configuration	Setup description	Carbon cycle component	Historical constraining	Short name
OSCARv3.1	(Quilcaille et al. 2023)		Probabilistic setup, weights applied by using carbon fluxes simulated in emissions-driven historical experiment		Terrestrial carbon-cycle calibrated on many TRENDY and CMIP5 models	S6
SCM4OPTv2.1	(Su et al. 2018)		Probabilistic setup, constrained by the historical GSAT	Three-box land and ocean box model, based on Hector	GSAT anomaly tuned to observations and CMIP5 ESMs	S7

*ECS equilibrium climate sensitivity

Table 2 List of CMIP6 ESMs analyzed in this study and their characteristics

Model	Reference	Nitrogen cycle	Land carbon cycle	Ocean carbon cycle	Short name
ACCESS-ESM1-5	(Ziehn et al. 2020)	Yes	CASA-CNP	WOMBAT	E1
CanESM5	(Swart et al. 2019)	No	CLASS-CTEM	CMOC	E2
CESM2-WACCM	(Danabasoglu et al. 2020)	Yes	CLM5	MARBL-BEC	E3
CNRM-ESM2-1	(Séférian et al. 2019)	No (implicit)	ISBA-CTRIP	PISCESv2-gas	E4
IPSL-CM6A-LR	(Boucher et al. 2020)	No (but downregulation of photosynthesis)	ORCHIDEE, br.2.0	PISCES-v2	E5
MIROC-ES2L	(Hajima et al. 2020)	Yes	VISIT-e	OECO2	E6
MPI-ESM1-2-LR	(Müller et al. 2018)	Yes	JSBACH3.2	HAMOCC6	E7
NorESM2-LM	(Seland et al. 2020)	Yes	CLM5	iHAMOCC	E9
UKESM1-0-LL	(Sellar et al. 2019)	Yes	JULES-ES-1.0	MEDUSA-2.1	E8

Table 3 SSP scenarios and their characteristics

| SSP	IAM	Citation	Short description
SSP1-1.9	IMAGE	(van Vuuren et al. 2017)	CO_2 emissions down to net zero around 2050 (increase in forest and bioenergy crops)
SSP1-2.6	IMAGE	(van Vuuren et al. 2017)	CO_2 emissions down to net zero around 2075 (increase in forest and bioenergy crops)
SSP2-4.5	MESSAGE-GLOBIOM	(Fricko et al. 2017)	A low stabilization scenario, an initial forest decrease from 2020 to 2050, with a subsequent increase from 2050 to 2100
SSP3-7.0	AIM/CGE	(Fujimori et al. 2017)	Slow economic development and intensive consumption in industrialized but not developing countries. Land use change is hardly regulated with large-scale deforestation.
SSP4-3.4	GCAM4	(Calvin et al. 2014)	Forest increases in high- and medium-income regions but decreases in low-income regions as the result of agricultural land expansion
SSP4-6.0	GCAM4	(Calvin et al. 2014)	Environmental policies, including afforestation programs in high- and medium-income regions, only
SSP5-3.4-OS	REMIND-MAGPIE	(Kriegler et al. 2017)	Delayed aggressive mitigation via BECCS after 2040.
SSP5-8.5	REMIND-MAGPIE	(Kriegler et al. 2017)	Very high levels of fossil fuel use, up to a doubling of global food demand, and up to a tripling of GHG emissions, the strong expansion of global cropland into pasture and forest land
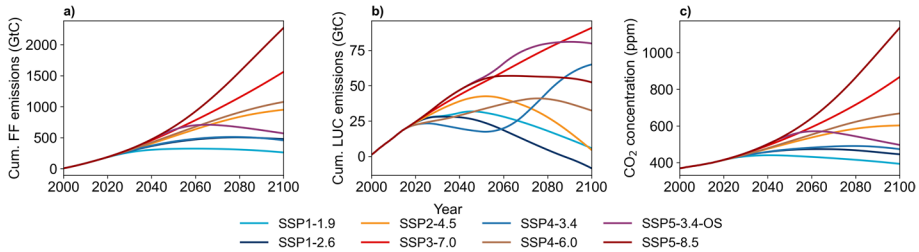
Fig. 1 Global cumulative **a** FF and **b** LUC CO_2 emissions (GtC, relative to the year 2000) generated by the IAMs that provided the CO_2 emission pathways corresponding to each SSP scenario and **c** atmospheric CO_2 mixing ratio calculated by MAGICC7 and used as an input for concentration-driven ESM simulations

production (*gpp*), autotrophic respiration (*ra*), and heterotrophic respiration (*rh*). In further calculations, we used *nbp* and *fgco2* as S_{Land} and S_{Ocean}. The *fLuc* variable of ESMs provides incomplete quantification of LUC emissions that excludes forest regrowth and legacy soil carbon decay or gains (Ciais et al. 2022). Among SCMs, only OSCAR of RCMIP Phase 1 internally calculated LUC emissions.

For ESMs, we used anomalies relative to the long-term mean *piControl* values, removing any residual trend in the *piControl* experiment. For SCMs, we estimated the NBP as the difference between their net land-to-atmosphere flux CO_2 and LUC emissions from IAMs provided in the SSP database (except for OSCAR, which provided both land carbon sink and LUC emission outputs). When ESMs provided NBP and *fLuc* estimates, we calculated the "natural land carbon sink" (excluding LUC emissions) by adding *fLuc* (positive to atmosphere) to NBP. We used the CO_2 concentration by Meinshausen et al. (2020) to estimate G_{Atm} for Eq. 1.

To evaluate GSAT estimates of the models, we used the mean of the following observationally based datasets: Cowtan and Way v2 (Cowtan and Way 2014), GISTEMP v4 (NASA Goddard and Institute for Space Studies 2020), CRU TS 4.00 (Harris et al. 2014), University of Delaware 4.01 (Lawrimore et al. 2011), and reanalyses-observation hybrids — CRU-NCEP and Princeton University (Sheffield et al. 2006). To evaluate historical carbon fluxes, LUC, and FF emissions with associated uncertainties, we used the Global Carbon Budget 2021 (GCB2021) dataset (Friedlingstein et al. 2021), historical data from Gruber et al. (2019), Khatiwala et al. (2009), and Li et al. (2016). GCB2021 is a synthesized dataset of major global carbon budget components, including FF emissions based on energy statistics and cement production data, LUC emissions based on land use and land-use change data and bookkeeping models, ocean carbon uptake based on estimates of global ocean models and observationally-based data products, and land carbon sink based on dynamic global vegetation models (Friedlingstein et al. 2021). Historical data from Gruber et al. (2019) include ocean carbon flux based on observations from US Global Ocean Carbon and Repeat Hydrography Program. Historical data from Khatiwala et al. (2009) include reconstruction of anthropogenic ocean carbon flux using tracer observations. Flux estimates of Gruber et al. (2019) and Khatiwala et al. (2009) include FF emissions from Boden et al. (2009) and LUC emissions based on bookkeeping models. For the two datasets, Khatiwala et al. (2009) and Boden et al. (2009), we calculate the land carbon flux using Equation 1. Finally, Li et al. (2016) provide estimates of land and ocean carbon uptakes, as well as LUC and FF emissions. These estimates are constrained with the Bayesian fusion approach and averaged over 5-year periods between 1980 and 2014. The constraining process of Li et al. (2016) estimates involves all the datasets described

above (although earlier version of GCB). We used the posterior values and uncertainties from Table S1 of Li et al. (2016) to estimate cumulative values of two 15-year periods (1980–1994 and 2000–2014) and evaluate models (Fig. S1).

2.2 Evaluation of the models and model configurations

We selected ESMs and SCMs that provide carbon flux estimates for at least one SSP at the time of the analysis. For three ESMs, CanESM5, IPSL-CM6A-LR, and MIROC-ES2L, we estimate internal uncertainty, i.e., the range of ensemble runs of individual ESMs, based on the 5–50 ensemble members (depending on availability) that differ in their initial conditions and physical processes (only CanESM5). Further, because some SCMs were available in multiple versions and calibrations, we selected/configurations based on an evaluation of their performance during the historical period and future SSP scenarios to ensure an equal weight to each model in estimating the model-ensemble mean. To measure the similarity between time series produced by models, we used the figure of merit in time (FMT), also known as Ruzicka similarity, which is defined as

$$FMT = \frac{\sum_{t=1}^{tmax} \min (X_t, Y_t)}{\sum_{t=1}^{tmax} \max (X_t, Y_t)}, \tag{1}$$

where X_t and Y_t are values of two sets of time series at a fixed time location t. FMT ranges from 0 to 1 with lower values indicating less similarity. A phase shift in time series causes a decrease in FMT, and ESM simulation outputs with large interannual variability are expected to have lower FMTs than SCMs (Figs. S2–S8). Besides, lower FMTs during historical period than future scenarios may indicate dominance of long-term trends over interannual variability.

When there were multiple versions or calibrations of a single model (e.g., Hector, MCE, WASP), we selected one version among the multiple versions or calibrations. Specifically, we used Hector default calibration (and not Histcalib). Although two calibrations led to similar results compared to other models, the DEFAULT calibration provided data for more scenarios. We used the newer version of MCE (v.1-2 from RCMIP Phase 2) rather than MCE v.1-1 (from RCMIP Phase 1). Finally, we used the historical calibration of WASP so that all considered SCMs are in historical calibration or probabilistic setup. Although OSCAR provided outputs for both phases of RCMIP, we utilized phase 1 outputs because they had explicit LUC emission estimates (including loss of additional sink capacity) obtained using a bookkeeping model (Gasser et al. 2020). ACC2 did not provide carbon cycle outputs for the concentration-driven SSPs in RCMIP. Thus, we additionally performed CO_2 concentration-driven simulations for ACC2.

RCMIP Phase 2 focused on the evaluation of the probabilistic climate based on SCMs. However, it did not discuss the carbon cycle in detail. Although we mainly used the central (50th percentile value) estimates of SCMs for this study when simulations were performed in a probabilistic setup, here we also discuss the assessment range of SCMs driven in a probabilistic setup (Table 1). We compare the SCM model spread and the probabilistic SCM distribution against ESMs estimate. We also compare inter-model SCM uncertainty and their probability ranges to the internal uncertainties of selected ESMs (ranges of ensemble runs of individual ESMs). Although we do not use all SCM versions/calibrations for computing ensemble means, we show the separate SCM estimates in Figs. S11–S18.

3 Results

We evaluated carbon fluxes from ESMs and SCMs in historical and future periods. In the historical period, the differences between estimates by complex and simple models could be attributed to the procedure of observational constraining that is present in the majority of SCMs (Table 1) but not applied in ESMs. In the future period, the differences could be partly explained by a few model outliers (both ESMs and SCMs). Besides, the differences in future scenarios arose from the differences in the way how LUC emissions were calculated and the assumptions on carbon-concentration and carbon-climate feedbacks. In this section, we discuss these sources of the differences in projections between complex and simple models.

3.1 Performance of models in the historical period

The larger warming estimated by CMIP6 ESMs than by SCMs over the historical period was discussed by Liddicoat et al. (2021) and Nicholls et al. (2021). It can be attributed to (i) a higher climate sensitivity of CMIP6 ESMs (on average) and (ii) the fact that SCMs are constrained to historical observations (Fig. 2a). The estimates of total (FF and LUC) compatible CO_2 emissions by CMIP6 ESMs are in their majority within the uncertainty range of GCB2021. The corresponding estimates by SCMs are lower and outside the uncertainty range of the GCB2021 historical CO_2 emissions (Fig. 2c). CMIP6 ESMs tend to estimate higher compatible FF emissions than SCMs and higher cumulative compatible FF emissions than observationally based FF emissions over the last three decades (Fig. 2f, h–m). Consistent with GCB2021 estimates of total compatible emissions and higher than GCB2021 estimates of compatible FF emissions by ESMs imply their underestimation of LUC emissions, discussed by Melnikova et al. (2022) (Fig. 2g).

In the concentration-driven model simulations, the estimates of compatible FF emissions are directly linked to the land and ocean carbon uptakes. Both ESMs and SCMs underestimate decadal ocean carbon uptake for the current period relative to estimates by Li et al. (2016), who reduced the carbon flux uncertainty using a Bayesian fusion approach. But the cumulative ocean carbon uptake estimates by both ESMs and SCMs are within the uncertainty range of historical estimates by Gruber et al. (2019) and GCB2021. ESMs, with a few exceptions, estimate slightly higher NBP than historical observationally-based estimates and SCMs over the historical period (Fig. 2b, h–m). The higher estimates of cumulative land carbon uptake by ESMs are not fully compensated by the lower estimates of cumulative ocean carbon uptake (relative to historical observationally based datasets), which leads to higher compatible FF emissions (Fig. 2h–l). In contrast, the lower estimates of cumulative land carbon uptake (NBP, Natural Land Sink) by ~20% of SCMs lead to lower compatible FF emissions estimates.

3.2 Emergent constraints

We evaluated the ESMs and SCMs' response to the future forcing using emergent constraints (ECs). EC approaches were broadly used in CMIP5 and CMIP6 communities on future warming (Tokarska et al. 2020; Schlund et al. 2020) and carbon cycle (Cox et al. 2013; Varney et al. 2020; Wenzel et al. 2014). While existing studies offer ECs on specific aspects of the carbon cycle, such as tropical carbon sensitivity to warming (Cox et al. 2013; Wenzel et al. 2014) or soil carbon turnover (Varney et al. 2020), we attempted to
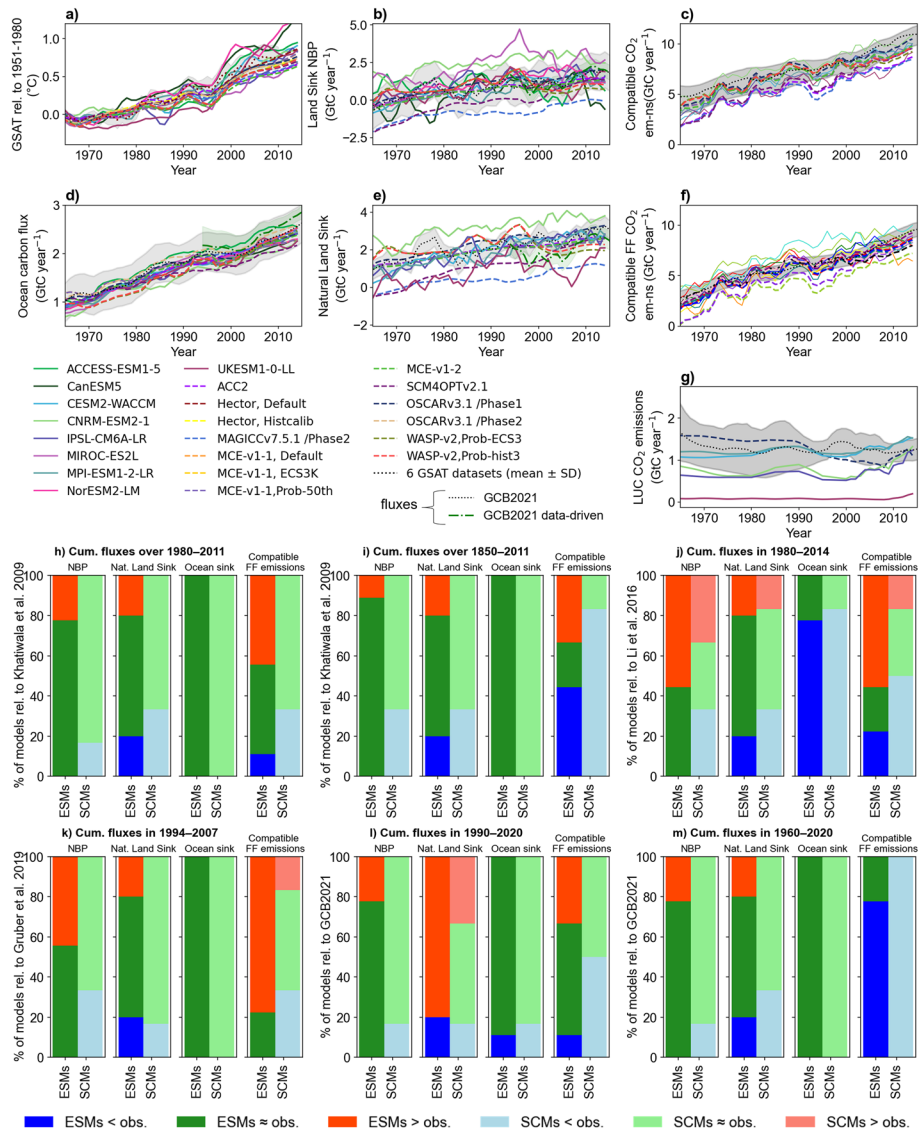
Fig. 2 **a** Simulated GSAT change (in °C, relative to 1850–1899) evaluated against the mean of six observational datasets for the 1965–2015 period. Same as **a** for **b** NBP (land sink with LUC emissions), **c** compatible FF and LUC CO_2 emissions NBP, **d** ocean carbon sink, **e** "natural" land sink without LUC emissions, **f** compatible FF CO_2 emissions, and **g** LUC emissions, evaluated against GCB2021 (black) and GCB2021 data-driven ocean sink with residual land sink (green) in GtC year^{-1}. Compatible FF CO_2 emissions of models are compared to GCB2021 FF CO_2 emissions. The 5-year moving averages of model-ensemble means are shown. Gray shading indicates uncertainty from the SD of observational datasets for GSAT and the uncertainty provided by GCB2021. The percentage of ESMs and SCMs that have consistent (within the uncertainty range provided with the data), higher, or lower estimates of decadal and cumulative global NBP, natural land sink, ocean carbon flux, and compatible FF CO_2 emissions with historical observationally-based datasets (h) over 1980–2011 and **i** 1850–2011 periods by Khatiwala et al. (2009), **j** over 1980–2014 period by Li et al. (2016), **k** over 1994–2007 period by Gruber et al. (2019), and **l** over 1990–2020 and **m** 1960–2020 period by GCB2021 and. ESMs and SCMs are shown in bright and pastel colors, respectively. Note that figures show the inter-model spreads of median projections from each model, without accounting for uncertainties within each model projections

develop a statistical relationship between global carbon fluxes. To this end, we plotted the estimates of 2015–2049 and 2065–2099 cumulative carbon fluxes against the estimates of 1980–2014 cumulative fluxes over the historical period (Figs. S9 and S10). The ESMs that estimate higher land and ocean uptakes during the historical period also give higher future carbon uptake in high CO_2 concentration scenarios. The correlations weaken (in terms of statistical significance) with time so that they are more reliable in the earlier future period and are not sustained in the low CO_2 concentration and overshoot pathways. This might be related to the more complex nature of mitigation scenarios that include ramp-up and ramp-down phases of CO_2 concentration and GSAT, as well as assumptions on implementing the land-based CO_2 removal technologies in the climate mitigation scenarios that influence the land carbon sink. A few model outliers also weaken the ECs. For example, compared to other models, WASP-v2 and CanESM5 simulate much larger increases in ocean and land carbon uptake, respectively, under high-concentration scenarios. The time series of the future carbon fluxes for each model confirmed these deviations (Figs. S13–S18). There is a weaker (less reliable) EC for SCMs due to the larger range of carbon cycle feedbacks to the changes in CO_2 and GSAT under SSPs, as well as the historical constraining of SCMs' carbon cycle feedbacks (see Section 3.4.3).

3.3 Performance of models in the future scenarios

The discrepancies in the future GSAT change estimates between CMIP6 ESMs and SCMs are consistent with those during the historical period (Fig. 3a). The GSAT increase is always higher in ESMs than in SCMs, the compatible CO_2 emissions are nearly consistent between ESMs and SCMs (Figs. 3, S11–18). Compared to SCMs, the ESMs estimate slightly lower inter-model ensemble mean cumulative emissions in high-concentration and slightly higher emissions in low-concentration SSPs. The larger GSAT increase estimated by ESMs and the consistent cumulative compatible emissions between ESMs and SCMs indicate that ESMs have higher TCRE than SCMs in the historical period and future SSP scenarios if we do not consider the contributions from non-CO_2 forcing (Fig. S19).

LUC emission data are largely inconsistent between ESMs and IAMs (used in SCMs) partly because of the discrepancies that emerge during the translation of the data from IAMs to ESMs (Melnikova et al. 2022). ESMs estimate lower positive LUC emissions (net source to the atmosphere), possibly because they do not include forestry and managed land practice. They also estimate smaller negative LUC emissions (net sink to the land) relative to emissions created by IAMs. The discrepancies may arise because LUC emissions reported by ESMs include deforestation (biomass loss during deforestation), wood harvest, and the carbon release by harvested wood products, but do not account for forest regrowth and legacy soil carbon decay or gains (Melnikova et al. 2022). Among SCMs, only one model, OSCAR, estimates LUC emissions via a bookkeeping approach.

Both ESMs and SCMs estimate higher future compatible FF emissions compared to those simulated by IAMs (Fig. 3c). This issue has been previously discussed by Liddicoat et al. (2021) and may be related to lower estimates of land and ocean carbon uptakes by MAGICC7.0, which was used with IAMs to generate the CMIP6 ScenarioMIP input CO_2 concentrations. MAGICC7.0 (slightly different version from MAGICCv7.5.1 of RCMIP phase 2) was calibrated to CMIP5 ESM carbon cycle to include permafrost CO_2 and methane feedbacks (Meinshausen et al. 2020; Nicholls et al. 2021). MAGICCv7.5.1 estimates a lower land carbon uptake than observationally based datasets during the historical period and than the model ensemble means in future scenarios (Figs. 1, S15–16). However, the
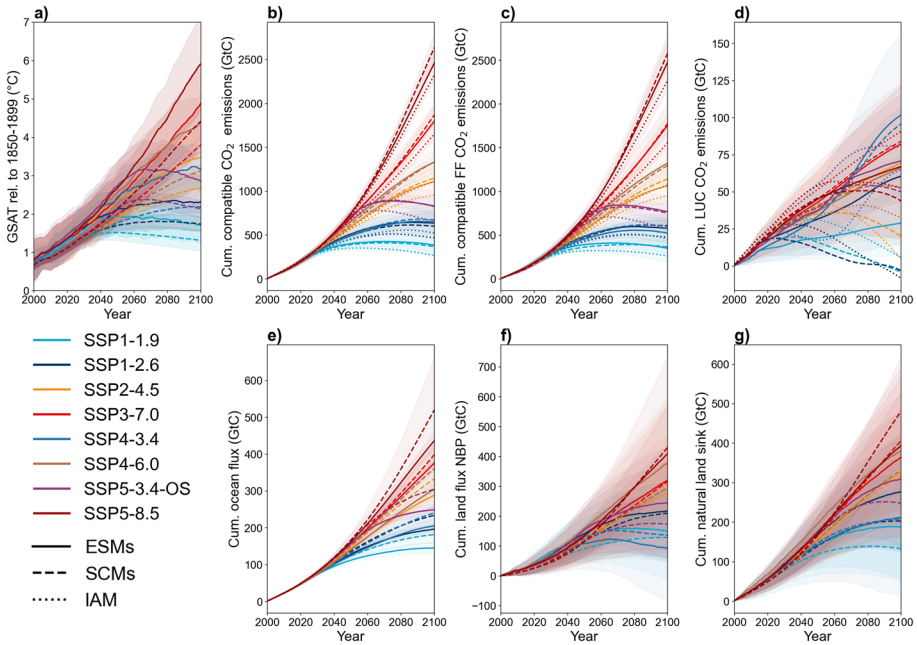
Fig. 3 **a** GSAT change (in °C, relative to 1850–1899), cumulative compatible **b** total, **c** FF and **d** LUC CO_2 emissions, **e** ocean carbon flux, **f** NBP (land sink with LUC emissions), and **g** natural land sink without LUC emissions estimated by ESMs (solid lines) and SCMs (dashed lines) under SSP scenarios over the 2000–2100 period (in GtC). FF and LUC CO_2 emissions generated by IAMs (dotted lines) corresponding to each SSP scenario are provided for reference. Shaded areas indicate the ESMs and SCMs inter-model spread for each scenario as one standard deviation (SD). Note that because LUC emissions are estimated by only one SCM, OSCAR, no spread is given. 5-year moving averages are shown. Note that figures show the inter-model spreads of median projections from each model, without accounting for uncertainties within each model projections

lower land carbon uptake estimates by MAGICCv7.5.1 are partly compensated by higher ocean carbon uptake estimates, especially in future scenarios. Thus, the total future land + ocean carbon uptake simulated by MAGICCv7.5.1 is lower than those by other models. Such deviation of the carbon cycle behavior of MAGICC from other models has broader implications because MAGICC is widely used for future projections informing policies and for translating the IAM emissions to concentrations used by ESMs.

Despite the general agreement of the estimates of cumulative compatible FF emissions between ESMs and SCMs, their estimates of land and ocean carbon uptakes deviate. SCMs estimate higher ocean carbon uptake than ESMs in the historical period and future SSPs. The inter-model spread of cumulative ocean carbon flux in future scenarios is also larger in SCMs (Fig. S9). The estimates of cumulative NBP are larger in ESMs than SCMs over the historical period and all future scenarios, except for SSP4-3.4, which assumes low FF but high LUC emissions relative to other SSPs (Fig. 1, Table 3).

3.4 Probablistic distributions of SCM estimates

The probabilistic ranges of single SCMs are larger than the SCMs inter-model spreads for all scenarios with a few exceptions (Fig. 4). The MCE of RCMIP Phase 1 provides
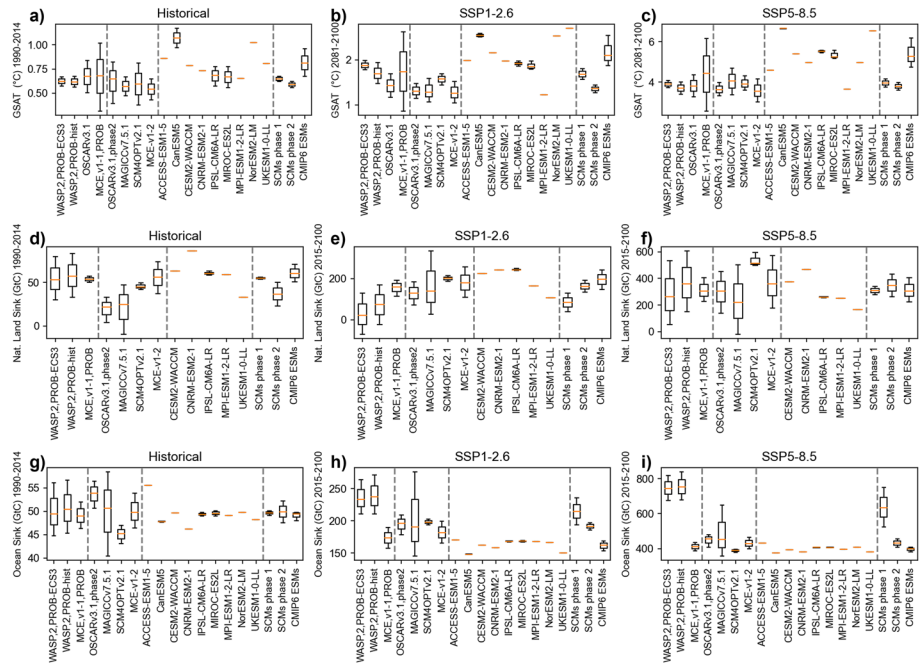
Fig. 4 **a–d** GSAT change (in °C, relative to 1850–1899), cumulative **e–h** natural land sink without LUC (GtC), and **i–l** ocean carbon flux (GtC) by ESMs and SCMs (Phases 1 and 2) under selected SSP scenarios over the 1990–2014 historical period and 2081–2100 (GSAT) and 2015–2100 (fluxes). The box plots of SCMs are the assessed probabilities (50th percentile, central box line; 33rd and 67th percentiles, lower and higher box limits; and 17th and 83rd percentiles, whiskers); the box plots of three selected ESMs are shown by the 6-member ensemble mean values in the central box line (33rd and 67th percentiles, lower and higher box limits; and 17th and 83rd percentiles, whiskers). The ESM ensemble members differ in their initial conditions. The right side of each panel displays the inter-model means and percentiles for SCMs of RCMIP Phases 1 and 2 and CMIP6 ESMs

the largest probabilistic range for GSAT estimates and MAGICC for the carbon cycle. The SCM probability ranges are comparable to the CMIP6 ESM inter-model spreads for all considered variables (apart from WASP's future ocean carbon flux estimates discussed in Section 3.5.6). Furthermore, the ranges for ocean carbon flux largely exceed the inter-model spread of CMIP6 estimates. Thus, despite the ESMs and SCMs differences in the median estimates of carbon fluxes, SCMs may be a useful tool for future probabilistic estimates of carbon cycle.

To increase the number of models, the study incorporated outputs of SCMs in different configurations, e.g., with warming calibrated to historical observations or with an equilibrium climate sensitivity (ECS) of 3K. Additionally, models from both phases of the RCMIP are combined. These differences may have affected future projections. For example, when comparing WASP-v2 calibrated with historical observations to WASP-v2 calibrated with an ECS of 3K, the former gives a lower estimate of future GSAT increase. While parameters other than ECS, such as those affecting CO_2 fertilization and temperature feedback to the carbon cycle, are calibrated with historical observations (Table 1), the GSAT differences further affect the estimates of land and ocean carbon fluxes. Consequently, WASP-v2 gives a higher median estimate of carbon sink

under future SSP scenarios when the model parameters are calibrated with historical observations.

3.5 The sources of discrepancies in the carbon cycle between ESMs and SCMs

The discrepancies between ESM- and SCM-ensemble means may originate from differences in the representation of some specific climate or carbon cycle processes and due to single or few model outliers that impact ensemble means. Besides, SCMs use observational constraints to reproduce the historical changes in climate and carbon cycle, while such constraints are not applied on ESMs. Here, we discuss how these effects could lead to the discrepancies in the carbon cycle between ESMs and SCMs.

3.5.1 Model outliers

We defined outlier models simply as the ESMs/SCMs that estimate maximum/minimum global land and ocean carbon uptake (maximum/minimum, or upper/lower ends) over 1850–2014 historical and 2015–2100 future SSP scenarios (Tables S1 and S2). The model outliers vary depending on the target carbon flux and SSP scenario, e.g., low vs. high FF or LUC emission pathways, increasing CO_2 concentration, and temperature vs. mitigation pathways. Besides, the models that provide data vary with the target scenario.

In the case of land carbon flux, CanESM5 and CNRM-ESM2-1 estimate the highest NBP during historical and future periods among ESMs. These two ESMs do not include a nitrogen cycle explicitly, a process that is shown to limit the land carbon uptake through nitrogen limitations of plant growth (Arora et al. 2020). OSCARv3.1 and ACC2 estimate the highest NBP among SCMs. The land carbon cycle of the version of ACC2 adopted in this study has limited sensitivity to GSAT increase.

In the case of ocean carbon flux, ESMs are nearly consistent with each other (Figs. S6 and S17). However, CanESM5 estimates slightly lower ocean carbon uptake than other ESMs in all future SSPs. This has been shown in an existing study (Arora et al. 2020), but the reasons remain unclear. Among SCMs, WASP-v2 and MAGICCv7.5.1 (when WASP-v2 scenario outputs are not available for the scenario) give maximum and minimum ocean carbon fluxes, respectively, in most future scenarios. MCE-v1-2, which shows the lowest future ocean uptake among SCMs under mitigation scenarios, provides an estimate of ocean carbon uptake closest to ESMs. Furthermore, the response of ocean carbon fluxes to declining CO_2 and temperature has larger hysteresis in some SCMs, such as SCM-4OPTv2.1 (Fig. S17).

Removing the ESMs and SCMs that produce the maximum/minimum cumulative fluxes over a target scenario improves the agreement of the carbon flux estimates between concentration-driven ESMs and SCMs and reduces their ensemble spreads, especially for NBP but not for all the scenarios (Fig. S20). Removing model outliers is less effective on improving the agreement between models in the scenarios that were run by few (<5) models. Besides, there is a systematic difference in the cumulative ocean uptake between ESMs and SCMs, so the presence of outliers cannot fully explain the discrepancy in the ocean carbon cycle between ESMs and SCMs. After removing outliers, the discrepancy in the future ocean carbon uptake between ESMs and SCMs, albeit of a smaller magnitude, persists in all SSPs (Fig. S20).

3.5.2 LUC emissions

The differences in the land carbon fluxes by ESMs and SCMs arise from several sources. They are partly explained by the LUC emissions. The CMIP6 ESMs provide the simulation outputs of NBP, i.e., land carbon uptake accounting for incompletely quantified LUC emissions via the "fLuc" ESM variable, thus underestimating LUC emissions (Melnikova et al. 2022) (Figs. 3 and S14). Additional simulations for each SSP scenario with a fixed land cover (like the "hist-noLu" simulation of Land-Use MIP (LUMIP)) are required to separate the "natural" land sink from the gross LUC emissions, including the foregone sink of land exposed to LUC. On the other hand, among SCMs, only OSCAR has interactions between LUC emissions and land carbon cycle. All other considered SCMs do not estimate LUC emissions but directly use the prescribed values that come from several methods and models (Le Quéré et al. 2016; Gütschow et al. 2016) in the historical period and from IAMs in future scenarios. This may lead to underestimating the impact of LUC on the land carbon uptake, e.g., by ignoring the reduced carbon turnover time in LUC-impacted ecosystems (Erb et al. 2016; Melnikova et al. 2022). Furthermore, during the historical period, the global net LUC emissions are consistently positive, i.e., directed to the atmosphere. However, in future scenarios, the LUC emissions by design may be either negative or positive. For instance, in low-concentration scenarios like SSP1-1.9 and SSP1-2.6 that assume large-scale afforestation for climate change mitigation, net LUC emissions become negative, indicating a land carbon sink (Figs. 1b and S14). Consequently, discrepancies in future net LUC emission estimates between ESMs and SCMs (via IAMs) may be both positive and negative.

3.5.3 Calibration of SCMs with observational constraints

SCMs can undergo a calibration process using historical observations (Table 1), during which the model's parameters are adjusted to accurately replicate our best-estimate observations, typically represented by median values and probabilistic distributions. The effectiveness of the calibration becomes evident as the resulting probabilistic distribution of model outputs successfully encompasses the historical range of observations (compare Fig. 2 and Fig. S21). While constraining SCMs allows for good agreement with existing observations, it is crucial to note that the range of model calibration may become too narrow when the models are applied to high warming scenarios. This narrow calibration range can lead to the misrepresentation of certain parameters within the models, thereby introducing bias into future projections. We discuss relevant examples, e.g., nitrogen limitation under high CO_2 and carbon-climate feedback, in the following section.

3.5.4 Carbon-concentration and carbon-climate feedbacks under high CO_2 concentration scenarios

Besides the discrepancy in the LUC component, part of it being due to issues with incomplete reporting of LUC fluxes by ESMs, ESMs and SCMs differ in the response of future cumulative land and ocean carbon fluxes to the CO_2 and GSAT changes. The carbon flux responses exhibit variations between high-concentration (Fig. 5) and low-concentration and overshoot (Fig. 6) SSP scenarios.
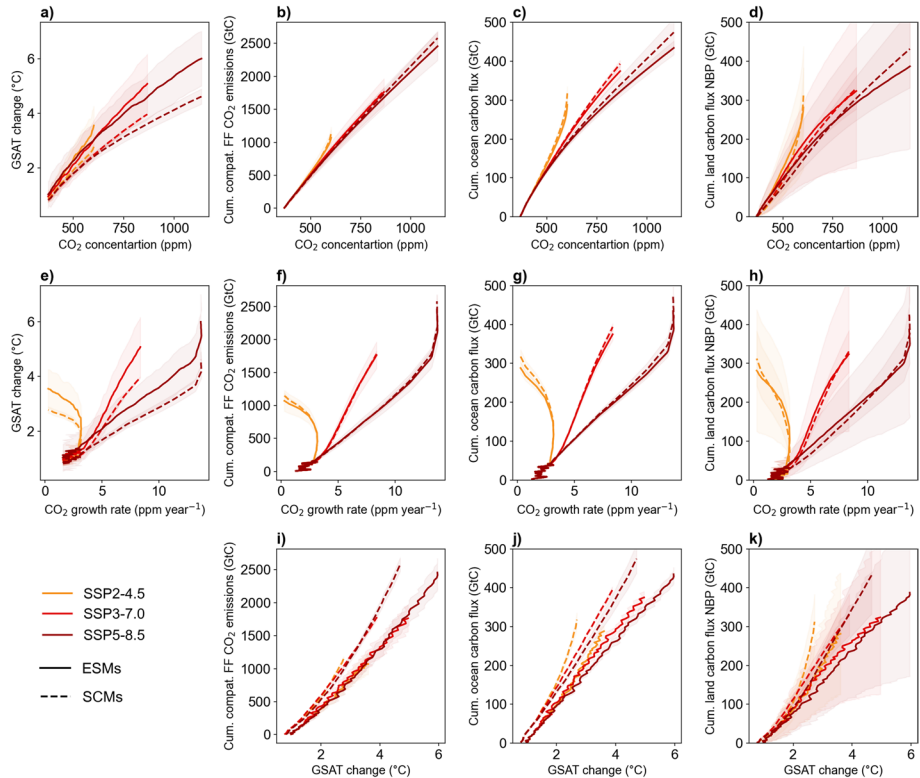
Fig. 5 Changes in **a**, **e** GSAT, global cumulative **b**, **f**, **i** compatible FF CO_2 emissions, **c**, **g**, **j** ocean carbon flux, and **d**, **h**, **k** NBP estimated by ESMs and SCMs under SSP scenarios over 2000–2100 periods plotted against changes in **a–d** CO_2 concentration (ppm), **e–h** CO_2 growth rate (ppm yr^{-1}), and **i–k** GSAT change (°C, relative to 1850–1899) under high-concentration scenarios. Ensemble means of carbon fluxes are calculated based on a suite of models, excluding outliers. The 5-year moving averages of model-ensemble means are shown. Shaded areas indicate the ESMs and SCMs inter-model spread for each scenario as one SD, when multiple models are available. Note that figures show the inter-model spreads of median projections from each model, without accounting for uncertainties within each model projections

Under high-concentration scenarios (Fig. 5), SCMs estimate larger increases of land and ocean carbon uptakes per unit CO_2 concentration and GSAT changes. ESMs estimate higher NBP per unit CO_2 increase up to CO_2 concentration level of 700–800 ppm. However, unlike SCMs, ESMs show a saturation of NBP with increasing CO_2 concentration and GSAT. The possible explanations for this discrepancy include the model differences in the land carbon-concentration and carbon-climate feedbacks.

First, a large spread of NBP estimates by ESMs reaches ca. 20 GtC $year^{-1}$ in 2100 under the high-concentration SSP5-8.5 scenario (Figs. S15h). In addition to LUC emissions, this discrepancy might be driven by the differences in the carbon-concentration (β) feedback, i.e., the impact of CO_2 concentration changes on carbon. The differences in the β feedback may be related to the inclusion or absence of the nitrogen cycle in the models (Friend et al. 2014). Most ESMs analyzed in this study include the nitrogen cycle (Table 2). Introducing the nitrogen cycle to an ESM generally weakens the β feedback on land carbon fluxes at high CO_2 concentrations (Figs. 5d, S22) because ecosystem nitrogen contents cannot keep
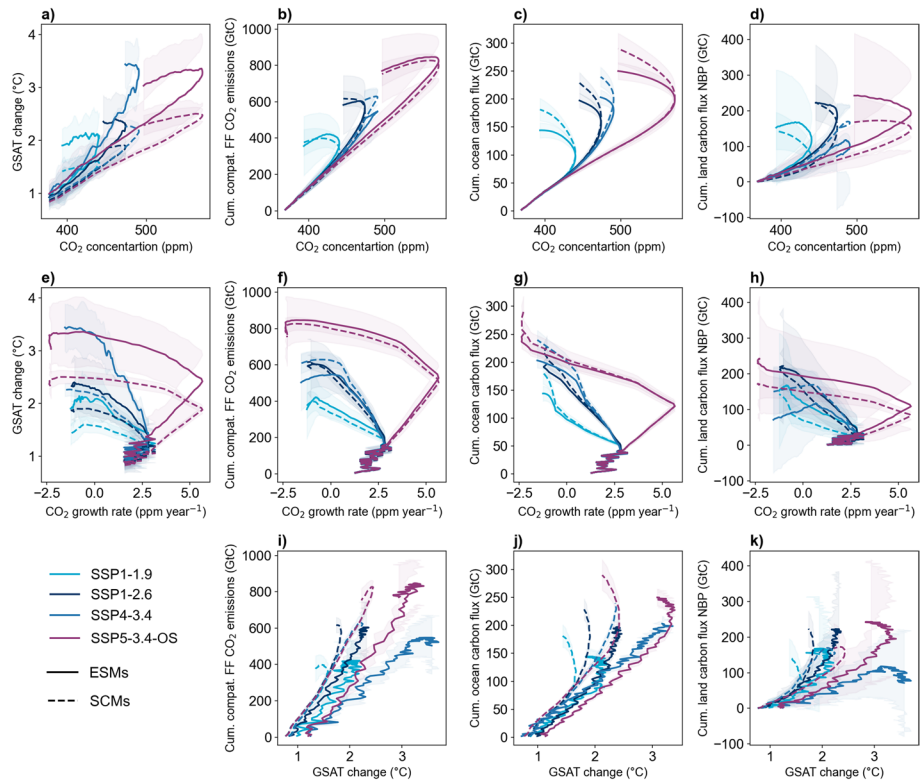
Fig. 6 Changes in **a**, **e** GSAT, global cumulative **b**, **f**, **i** compatible FF CO_2 emissions, **c**, **g**, **j** ocean carbon flux, and **d**, **h**, **k** NBP estimated by ESMs and SCMs under SSP scenarios over 2000–2100 periods plotted against changes in **a**–**d** CO_2 concentration (ppm), **e**–**h** CO_2 growth rate (ppm yr^{-1}), and **i**–**k** GSAT change (°C, relative to 1850–1899) under low-concentration and overshoot scenarios. Ensemble means of carbon fluxes are calculated based on a suite of models, excluding outliers. The 5-year moving averages of model-ensemble means are shown. Shaded areas indicate the ESMs and SCMs inter-model spread for each scenario as one SD. Note that figures do not show the internal model uncertainties and probability distributions

up with the increased photosynthetic production, thus limiting carbon assimilation (Arora et al. 2020). Among ESMs, CanESM5 and CNRM-ESM2-1 that do not have explicit nitrogen cycle module have the highest CO_2 concentration-driven carbon flux estimates at high CO_2 concentration levels (Fig. S22).

Second, SCMs may estimate higher land carbon uptake per GSAT unit change due to their insufficient carbon-climate (γ) feedback on the carbon cycle driven by historical constraining. Although historical land carbon uptake has been largely influenced by the β feedback (Tharammal et al. 2019), its might decrease with future warmer temperatures and other limitation (Figs. 5 and S22). While SCMs have been historically constrained against observations (as indicated in Table 1 and Section 3.5.3), incorporating γ feedback into these models is more challenging since there is relatively limited observational evidence of its global-scale effects.

The inter-model spread in the NBP estimates by ESMs can itself be largely explained by the uncertainty in the γ feedback in the tropics (Fig. S23). Furthermore, unlike the β feedback, which represents carbon flux responses to changes in CO_2 concentration, the γ

feedback responds to temperature changes, considering not only CO_2 but also other non-CO_2 GHGs and biogeophysical effects of land cover (Melnikova et al. 2023). Future SSP scenarios encompass diverse changes in non-CO_2 GHG concentrations that may further affect the response of the γ feedback. This adds to uncertainty in the estimates of future land carbon fluxes by ESMs.

3.5.5 Carbon flux estimates under low CO_2 concentration and overshoot scenarios

Under low-concentration (mitigation and overshoot) pathways, there is a reasonable agreement between SCMs and ESMs regarding the response of compatible FF estimates to changes in CO_2 (Fig. 6). However, this agreement between SCMs and ESMs is based on an incorrect premise. SCMs estimate a larger hysteresis in the response of ocean carbon uptake to CO_2 and GSAT (global surface air temperature) changes, while ESMs estimate a larger hysteresis in the response of NBP (net biome productivity) to CO_2 and GSAT changes.

The discrepancy in NBP is already evident in the ramp-up phase of peak and decline scenarios, due to the reasons outlined in the previous sections. The discrepancy amplifies to the extent that SCMs and ESMs exhibit opposite directions of hysteresis under the SSP4-3.4 scenario (Fig. 6d, h, k). This SSP scenario is characterized by intricate and dynamic land-cover changes (Fig. 1b, Table 3). It is worth noting that LUC emissions play an even more significant role in low-concentration scenarios due to their larger contribution to the total emissions. Moreover, many of these scenarios rely on various land-based mitigation strategies. Unfortunately, the framework of this study does not permit an in-depth exploration of the underlying reasons for the discrepancies in land carbon uptake estimated by ESMs and SCMs. Hence, we recommend that future studies thoroughly compare the carbon cycles of ESMs and SCMs under overshoot and mitigation scenarios.

The estimates of ocean carbon uptake by SCMs and ESMs exhibit reasonable agreement during the ramp-up phase of overshoot scenarios. However, in the ramp-down phases of peak and decline scenarios, SCMs consistently display a larger hysteresis in the response of ocean carbon uptake to changes in CO_2 and GSAT compared to ESMs. This discrepancy may be attributed to the faster carbon mixing in the ocean, as discussed in the following section.

3.5.6 Mixing of carbon in the ocean

The differences between ESMs and SCMs also emerge in the response of ocean carbon flux to the CO_2 growth rate, CO_2 concentration, and GSAT (Fig. 4). The increase in cumulative ocean carbon uptake with increasing CO_2 growth rate, CO_2 concentration, and GSAT diminishes in the ESMs but not SCMs. The discrepancies in the ocean carbon flux may be attributed to the nonlinearities of the carbon-concentration and carbon-climate feedbacks that are the changes in the carbon storage in response to the changes in CO_2 concentration and GSAT, respectively (Gregory et al. 2009; Schwinger and Tjiputra 2018; Melnikova et al. 2021). The response of the β feedback in the ocean is complex, with change in CO_2 growth rate dominating the flux variability on year-to-year timescales (i.e., system response to the forcing rate of change) and change in CO_2 concentration dominating the variability on decadal timescales (i.e., system response to the forcing magnitude) (Schwinger and Tjiputra 2018; Melnikova et al. 2021). Thus, while the ocean carbon storage response to the forcing rate of change is nearly equal among the two types of models, the storage

response to the forcing magnitude is delayed in SCMs. We speculate that the discrepancy may be rooted in a faster mixing of carbon from the surface to the deep ocean in SCMs. Schwinger and Tjiputra (2018) showed that the carbon uptake of water masses of different ages exhibits varying degrees of hysteresis in response to changes in CO_2 concentrations. The younger water masses show considerably less hysteresis in the carbon uptake than the deep ocean water masses. Similarly, the ocean carbon uptake increases less with increasing GSAT and starts to decrease sooner with decreasing GSAT under climate change mitigation overshoot-like scenarios when estimated by ESMs compared to SCMs. We recommend that the discrepancies in the response of ocean carbon uptake by ESMs and SCMs are further investigated using the set of idealized experiments, including those that have abrupt CO_2 increases and overshoot (ramp-up and ramp-down) scenarios.

4 Conclusion

This study investigates the differences in the carbon cycle projections calculated by ESMs and SCMs during the historical period and under future SSPs. First, we evaluate models' estimates of land and ocean carbon fluxes during the historical period. Second, we analyze the discrepancy in the future land and ocean carbon uptake estimated by ESMs and SCMs that emerges due to structural differences, as well as model outliers that impact ensemble means. Although existing evidence did not allow to scrutinize thoroughly the reasons for discrepancies in carbon cycle responses between ESMs and SCMs, we propose some likely explanations. To better align the carbon cycle projections between ESMs and SCMs, we put forward a set of recommendations regarding features of models that can be further developed: (1) mixing of carbon from the surface to the deep ocean; (2) future carbon-concentration feedback, influenced by nitrogen limitation of photosynthesis, and carbon-climate feedback; (3) representation of LUC emissions in the models; and (4) historical calibration of SCMs. Many carbon removal technologies are land-based and require land use in one form or the other. In a world with decreasing FF emissions, how land is used becomes increasingly important, and it is crucial to improve the consistency between ESMs and SCMs with respect to LUC emissions.

Improving the carbon cycle features of SCMs will improve applicability of SCMs as ESM emulators for both climate and carbon cycle projections and thus advance the developments of AR6. Particularly, we highlight the deviations in the estimates of the carbon cycle dynamics by SCM MAGICC, which is widely used to convert the IAM emissions to concentrations that are used by ESMs. Furthermore, we draw attention to the inconsistencies between the LUC emission estimates reported by ESMs and IAMs. These inconsistencies arise during the data conversion from IAMs to ESMs, as well as from the differences in the definitions of LUC emissions. These definitions do not always account for the same processes and components. The differences in carbon flux estimates between SCMs and ESMs are particularly large for low CO_2 concentration and overshoot scenarios. This has important implications as these scenarios are relevant to policy analyses that are often supported by SCM simulations. We call for carrying out more SCMs–ESMs (and potentially IAMs) joint studies and inter-model comparison exercises, to explore future biogeochemical feedbacks related to land and ocean-based mitigation options, pursue more consistent reporting of various carbon fluxes, and seek a higher consistency between models used to generate scenarios.

Data availability The RCMIP Phase 1 and 2 data were downloaded from https://zenodo.org/record/4016613#.YioYXJYo9PY and https://zenodo.org/record/4624566#.YioYdpYo9PY. The ESM data were downloaded from the CMIP6 archive (https://esgf-node.llnl.gov/search/cmip6). The IAM database maintained at International Institute for Applied Systems Analysis (IIASA) is available at https://tntcat.iiasa.ac.at/SspDb/dsd?Action=htmlpage&page=welcome. The global CO_2 concentration data were downloaded from the supplement of Meinshausen et al. (2020). Processed model outputs, evaluation data, and the Jupyter notebook that were used to make the figures in the article are available at https://zenodo.org/records/10162686.

Declarations

Consent to participate All subjects gave their consent to participate in the study.

Consent for publication All subjects gave their consent to the publication of the data.

Conflict of interest K.T. is an Associate Deputy Editor of *Climatic Change*. His status has no bearing on editorial consideration. The other authors declare no competing interests.

References

Ackerman F, DeCanio SJ, Howarth RB, Sheeran K (2009) Limitations of integrated assessment models of climate change. Clim Chang 95:297–315. https://doi.org/10.1007/s10584-009-9570-x

Arora VK, Katavouta A, Williams RG et al (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

Boden TA, Marland G, Andres RJ (2009) Global, regional, and national fossil-fuel CO2 emissions, 1751-2006 (published 2009). Environmental System Science Data Infrastructure for a Virtual Ecosystem …

Boucher O, Servonnat J, Albright AL et al (2020) Presentation and evaluation of the IPSL-CM6A-LR climate model. J Adv Model Earth Syst 12:e2019MS002010. https://doi.org/10.1029/2019MS002010

Calvin K, Wise M, Kyle P et al (2014) Trade-offs of different land and bioenergy policies on the path to achieving climate targets. Clim Chang 123:691–704. https://doi.org/10.1007/s10584-013-0897-y

Ciais P, Bastos A, Chevallier F et al (2022) Definitions and methods to estimate regional land carbon fluxes for the second phase of the REgional Carbon Cycle Assessment and Processes Project (RECCAP-2). Geosci Model Dev 15:1289–1316. https://doi.org/10.5194/gmd-15-1289-2022

Cowtan K, Way RG (2014) Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. Q J R Meteorol Soc 140:1935–1944. https://doi.org/10.1002/qj.2297

Cox PM, Pearson D, Booth BB et al (2013) Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. Nature 494:341–344. https://doi.org/10.1038/nature11882

Danabasoglu G, Lamarque J-F, Bacmeister J et al (2020) The community earth system model version 2 (CESM2). J Adv Model Earth Syst 12:e2019MS001916. https://doi.org/10.1029/2019MS001916

Erb K-H, Fetzel T, Plutzar C et al (2016) Biomass turnover time in terrestrial ecosystems halved by land use. Nat Geosci 9:674–678. https://doi.org/10.1038/ngeo2782

Eyring V, Bony S, Meehl GA et al (2016) Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. Geosci Model Dev 9:1937. https://doi.org/10.5194/gmd-9-1937-2016

Forster P, Storelvmo T, Armour K et al (2021) Chapter 7: the earth's energy budget, climate feedbacks, and climate sensitivity. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press

Fricko O, Havlik P, Rogelj J et al (2017) The marker quantification of the shared socioeconomic pathway 2: a middle-of-the-road scenario for the 21st century. Glob Environ Chang 42:251–267. https://doi.org/10.1016/j.gloenvcha.2016.06.004

Friedlingstein P, Jones MW, O'Sullivan M et al (2021) Global carbon budget 2021. Earth Syst Sci Data Discuss 2021:1–191. https://doi.org/10.5194/essd-2021-386

Friend AD, Lucht W, Rademacher TT et al (2014) Carbon residence time dominates uncertainty in terrestrial vegetation responses to future climate and atmospheric CO2. Proc Natl Acad Sci 111:3280–3285. https://doi.org/10.1073/pnas.1222477110

Fujimori S, Hasegawa T, Masui T et al (2017) SSP3: AIM implementation of Shared Socioeconomic Pathways. Glob Environ Chang 42:268–283. https://doi.org/10.1016/j.gloenvcha.2016.06.009

Gasser T, Ciais P, Boucher O et al (2017) The compact earth system model OSCAR v2.2: description and first results. Geosci Model Dev 10:271–319. https://doi.org/10.5194/gmd-10-271-2017

Gasser T, Crepin L, Quilcaille Y et al (2020) Historical CO2 emissions from land use and land cover change and their uncertainty. Biogeosciences 17:4075–4101. https://doi.org/10.5194/bg-17-4075-2020

Gidden MJ, Riahi K, Smith SJ et al (2019) Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century. Geosci Model Dev 12:1443–1475. https://doi.org/10.5194/gmd-12-1443-2019

GISTEMP Team (2023) GISS Surface Temperature Analysis (GISTEMP), version 4. NASA Goddard Institute for Space Studies. Dataset accessed 2023-10-01 at https://data.giss.nasa.gov/gistemp/

Goodwin P (2018) On the time evolution of climate sensitivity and future warming. Earth's Future 6:1336–1348. https://doi.org/10.1029/2018EF000889

Gregory JM, Jones CD, Cadule P, Friedlingstein P (2009) Quantifying carbon cycle feedbacks. J Climate 22:5232–5250. https://doi.org/10.1175/2009JCLI2949.1

Gruber N, Clement D, Carter BR et al (2019) The oceanic sink for anthropogenic CO2 from 1994 to 2007. Science 363:1193–1199. https://doi.org/10.1126/science.aau5153

Gütschow J, Jeffery ML, Gieseke R et al (2016) The PRIMAP-hist national historical emissions time series. Earth Syst Sci Data 8:571–603. https://doi.org/10.5194/essd-8-571-2016

Hajima T, Watanabe M, Yamamoto A et al (2020) Development of the MIROC-ES2L earth system model and the evaluation of biogeochemical processes and feedbacks. Geosci Model Dev 13:2197–2244. https://doi.org/10.5194/gmd-13-2197-2020

Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. Int J Climatol 34:623–642

Hartin CA, Patel P, Schwarber A et al (2015) A simple object-oriented and open-source model for scientific and policy analyses of the global climate system – Hector v1.0. Geosci Model Dev 8:939–955. https://doi.org/10.5194/gmd-8-939-2015

Hurtt G, Chini L, Sahajpal R et al (2017) Harmonization of global land use scenarios (LUH2). Historical v2.1h:850–2015

Hurtt GC, Chini L, Sahajpal R et al (2020) Harmonization of global land-use change and management for the period 850-2100 (LUH2) for CMIP6. Geosci Model Dev Discuss 1–65. https://doi.org/10.5194/gmd-2019-360

Joos F, Roth R, Fuglestvedt JS et al (2013) Carbon dioxide and climate impulse response functions for the computation of greenhouse gas metrics: a multi-model analysis. Atmos Chem Phys 13:2793–2825. https://doi.org/10.5194/acp-13-2793-2013

Khatiwala S, Primeau F, Hall T (2009) Reconstruction of the history of anthropogenic CO2 concentrations in the ocean. Nature 462:346–349. https://doi.org/10.1038/nature08526

Kriegler E, Bauer N, Popp A et al (2017) Fossil-fueled development (SSP5): an energy and resource intensive scenario for the 21st century. Glob Environ Chang 42:297–315. https://doi.org/10.1016/j.gloenvcha.2016.05.015

Lawrimore JH, Menne MJ, Gleason BE et al (2011) An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. J Geophys Res: Atmos 116. https://doi.org/10.1029/2011JD016187

Le Quéré C, Andrew RM, Canadell JG et al (2016) Global carbon budget 2016. Earth Syst Sci Data 8:605–649. https://doi.org/10.5194/essd-8-605-2016

Li W, Ciais P, Wang Y et al (2016) Reducing uncertainties in decadal variability of the global carbon budget with multiple datasets. Proc Natl Acad Sci 113:13104–13108. https://doi.org/10.1073/pnas.1603956113

Liddicoat SK, Wiltshire AJ, Jones CD 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. J Clim 34:2853–2875

Masson-Delmotte V, Zhai P, Pirani A et al (2021) Climate change 2021: the physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

Meinshausen M, Nicholls ZRJ, Lewis J et al (2020) The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. Geosci Model Dev 13:3571–3605. https://doi.org/10.5194/gmd-13-3571-2020

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

Meinshausen M, Wigley TML, Raper SCB (2011b) Emulating atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – Part 2: Applications. Atmos Chem Phys 11:1457–1471. https://doi.org/10.5194/acp-11-1457-2011

Melnikova I, Boucher O, Cadule P et al (2022) Impact of bioenergy crops expansion on climate-carbon cycle feedbacks in overshoot scenarios. Earth Syst Dynam 13:779–794. https://doi.org/10.5194/esd-13-779-2022

Melnikova I, Boucher O, Cadule P et al (2021) Carbon cycle response to temperature overshoot beyond 2 °C – an analysis of CMIP6 models. Earth's. Future 9:e2020EF001967. https://doi.org/10.1029/2020EF001967

Melnikova I, Ciais P, Tanaka K et al (2023) Relative benefits of allocating land to bioenergy crops and forests vary by region. Commun Earth Environ 4:230. https://doi.org/10.1038/s43247-023-00866-7

Müller WA, Jungclaus JH, Mauritsen T et al (2018) A higher-resolution version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). J Adv Model Earth Syst 10:1383–1413. https://doi.org/10.1029/2017MS001217

Nicholls Z, Meinshausen M, Lewis J et al (2021) Reduced complexity model intercomparison project phase 2: synthesizing earth system knowledge for probabilistic climate projections. Earth's. Future 9:e2020EF001900. https://doi.org/10.1029/2020EF001900

Nicholls ZRJ, Meinshausen M, Lewis J et al (2020) Reduced complexity model intercomparison project phase 1: introduction and evaluation of global-mean temperature response. Geosci Model Dev 13:5175–5190. https://doi.org/10.5194/gmd-13-5175-2020

O'Neill BC, Kriegler E, Riahi K et al (2014) A new scenario framework for climate change research: the concept of shared socioeconomic pathways. Clim Chang 122:387–400. https://doi.org/10.1007/s10584-013-0905-2

O'Neill BC, Tebaldi C, Van Vuuren DP et al (2016) The scenario model intercomparison project (ScenarioMIP) for CMIP6. Geosci Model Dev. https://doi.org/10.5194/gmd-9-3461-2016

Quilcaille Y, Gasser T, Ciais P, Boucher O (2023) CMIP6 simulations with the compact earth system model OSCAR v3.1. Geosci Model Dev 16:1129–1161. https://doi.org/10.5194/gmd-16-1129-2023

Riahi K, van Vuuren DP, Kriegler E et al (2017) The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview. Glob Environ Chang 42:153–168. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Schlund M, Lauer A, Gentine P et al (2020) Emergent constraints on equilibrium climate sensitivity in CMIP5: do they hold for CMIP6? Earth Syst Dynam 11:1233–1258. https://doi.org/10.5194/esd-11-1233-2020

Schwinger J, Tjiputra J (2018) Ocean carbon cycle feedbacks under negative emissions. Geophys Res Lett 45:5062–5070. https://doi.org/10.1029/2018GL077790

Séférian R, Nabat P, Michou M et al (2019) Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth System Processes in Present-Day and Future Climate. J Adv Model Earth Syst 11:4182–4227. https://doi.org/10.1029/2019MS001791

Seland Ø, Bentsen M, Olivié D et al (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

Sellar AA, Jones CG, Mulcahy JP et al (2019) UKESM1: description and evaluation of the U.K. Earth System Model. J Adv Model Earth Syst 11:4513–4558. https://doi.org/10.1029/2019MS001739

Sheffield J, Goteti G, Wood EF (2006) Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. J Clim 19:3088–3111

Shukla PR, Skea J, Slade R et al (2022) Climate change 2022: mitigation of climate change. In: Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK and New York, NY, USA

Su X, Shiogama H, Tanaka K et al (2018) How do climate-related uncertainties influence 2 and 1.5 °C pathways? Sustain Sci 13:291–299. https://doi.org/10.1007/s11625-017-0525-2

Swart NC, Cole JNS, Kharin VV et al (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

Tanaka K, O'Neill BC (2018) The Paris Agreement zero-emissions goal is not always consistent with the 1.5 °C and 2 °C temperature targets. Nat Clim Chang 8:319–324. https://doi.org/10.1038/s41558-018-0097-x

Tanaka K, Raddatz T, O'Neill BC, Reick CH (2009) Insufficient forcing uncertainty underestimates the risk of high climate sensitivity. Geophys Res Lett 36. https://doi.org/10.1029/2009gl039642

Tharammal T, Bala G, Devaraju N, Nemani R (2019) A review of the major drivers of the terrestrial carbon uptake: model-based assessments, consensus, and uncertainties. Environ Res Lett 14:093005. https://doi.org/10.1088/1748-9326/ab3012

Tokarska KB, Stolpe MB, Sippel S et al (2020) Past warming trend constrains future warming in CMIP6 models. Sci Adv 6:eaaz9549

Tsutsui J (2020) Diagnosing transient response to CO2 forcing in coupled atmosphere-ocean model experiments using a climate model emulator. Geophys Res Lett 47:e2019GL085844. https://doi.org/10.1029/2019GL085844

van Vuuren DP, Stehfest E, Gernaat DEHJ et al (2017) Energy, land-use and greenhouse gas emissions trajectories under a green growth paradigm. Glob Environ Chang 42:237–250. https://doi.org/10.1016/j.gloenvcha.2016.05.008

Varney RM, Chadburn SE, Friedlingstein P et al (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

Vega-Westhoff B, Sriver RL, Hartin CA et al (2019) Impacts of observational constraints related to sea level on estimates of climate sensitivity. Earth's Future 7:677–690. https://doi.org/10.1029/2018EF001082

Wenzel S, Cox PM, Eyring V, Friedlingstein P (2014) Emergent constraints on climate-carbon cycle feedbacks in the CMIP5 earth system models. J Geophys Res Biogeo 119:794–807. https://doi.org/10.1002/2013JG002591

Ziehn T, Chamberlain MA, Law RM et al (2020) The Australian Earth System Model: ACCESS-ESM1.5. JSHESS 70:193–214

Publisher's note   Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.