AGU ADVANCING EARTH AND SPACE SCIENCES

# AGU Advances

**Key Points:**
- Regional carbon cycle assessment enables a quality control on future projections
- Here we compare global model projections with regional assessment of carbon fluxes and stocks: Global multi-model mean is broadly consistent with regional assessments
- Regional analysis yields important insights for future model developments

**Correspondence to:**
C. D. Jones,
chris.d.jones@metoffice.gov.uk

**Author Contributions:**
**Conceptualization:** Chris D. Jones, Tilo Ziehn, T. Luke Smallman
**Formal analysis:** Chris D. Jones, Anatoly Shvidenko, T. Luke Smallman

# RECCAP2 Future Component: Consistency and Potential for Regional Assessment to Constrain Global Projections

Chris D. Jones[1], Tilo Ziehn[2], Jatin Anand[3], Ana Bastos[4], Eleanor Burke[1], Josep G. Canadell[5], Manoel Cardoso[6], Yolandi Ernst[7], Atul K. Jain[3], Sujong Jeong[8], Elizabeth D. Keller[9,10], Masayuki Kondo[11], Ronny Lauerwald[12], Tzu-Shun Lin[3,13], Guillermo Murray-Tortarolo[14], Gert-Jan Nabuurs[15], Mike O'Sullivan[16], Ben Poulter[17], Xiaoyu Qin[18,19], Celso von Randow[6], Marcos Sanches[6], Dmitry Schepaschenko[20], Anatoly Shvidenko[20], T. Luke Smallman[21,22], Hanqin Tian[23], Yohanna Villalobos[5], Xuhui Wang[24], and Jeongmin Yun[25]

[1]Met Office Hadley Centre, Exeter, UK, [2]CSIRO Environment, Aspendale, VIC, Australia, [3]Department of Atmospheric Sciences, University of Illinois, Urbana, IL, USA, [4]Department Biogeochemical Integratio, Max-Planck-Institut für Biogeochemie, Jena, Germany, [5]Global Carbon Project, CSIRO Environment, Canberra, ACT, Australia, [6]National Institute for Space Research, São José dos Campos, Brazil, [7]Global Change Institute, University of the Witwatersrand, Johannesburg, South Africa, [8]Department of Environmental Planning, Graduate School of Environmental Studies, Seoul National University, Seoul, Republic of Korea, [9]GNS Science, Lower Hutt, New Zealand, [10]Victoria University of Wellington, Wellington, New Zealand, [11]Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Japan, [12]INRAE/AgroParisTech-UMR Ecosys, Université Paris-Saclay, Palaiseau, France, [13]National Center for Atmospheric Research, Boulder, CO, USA, [14]Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México, Morelia, México, [15]Wageningen Environmental Research, Wageningen, The Netherlands, [16]Faculty of Environment, Science and Economy, University of Exeter, Exeter, UK, [17]Biospheric Sciences Lab, NASA Goddard Space Flight Center, Greenbelt, MD, USA, [18]State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, China, [19]University of Chinese Academy of Sciences, Beijing, China, [20]International Institute for Applied Systems Analysis, Laxenburg, Austria, [21]School of GeoSciences, University of Edinburgh, Edinburgh, UK, [22]National Centre for Earth Observations, University of Edinburgh, Edinburgh, UK, [23]Department of Earth and Environmental Sciences, Schiller Institute for Integrated Science and Society, Boston College, Chestnut Hill, MA, USA, [24]College of Urban and Environmental Sciences, Institute of Carbon Neutrality, Sino-French Institute of Earth System Sciences, Peking University, Beijing, China, [25]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

**Abstract** Projections of future carbon sinks and stocks are important because they show how the world's ecosystems will respond to elevated $CO_2$ and changes in climate. Moreover, they are crucial to inform policy decisions around emissions reductions to stay within the global warming levels identified by the Paris Agreement. However, Earth System Models from the 6th Coupled Model Intercomparison Project (CMIP6) show substantial spread in future projections—especially of the terrestrial carbon cycle, leading to a large uncertainty in our knowledge of any remaining carbon budget (RCB). Here we evaluate the global terrestrial carbon cycle projections on a region-by-region basis and compare the global models with regional assessments made by the REgional Carbon Cycle Assessment and Processes, Phase 2 activity. Results show that for each region, the CMIP6 multi-model mean is generally consistent with the regional assessment, but substantial cross-model spread exists. Nonetheless, all models perform well in some regions and no region is without some well performing models. This gives confidence that the CMIP6 models can be used to look at future changes in carbon stocks on a regional basis with appropriate model assessment and benchmarking. We find that most regions of the world remain cumulative net sources of $CO_2$ between now and 2100 when considering the balance of fossil-fuels and natural sinks, even under aggressive mitigation scenarios. This paper identifies strengths and weaknesses for each model in terms of its performance over a particular region including how process representation might impact those results and sets the agenda for applying stricter constraints at regional scales to reduce the uncertainty in global projections.

**Plain Language Summary** Future climate change, and higher levels of $CO_2$ in the atmosphere will have significant effects on ecosystems around the world. It is important that we can model these effects both because the ecosystems are important in their own right and because the carbon they store affects how much $CO_2$ accumulates in the atmosphere, and thus, how strongly we must reduce human emissions in order to meet internationally agreed climate goals. However, the models often give different results, leading to a large spread

**Investigation:** Chris D. Jones
**Methodology:** Chris D. Jones, Tilo Ziehn, Dmitry Schepaschenko, Anatoly Shvidenko, T. Luke Smallman
**Resources:** Dmitry Schepaschenko
**Software:** Chris D. Jones
**Visualization:** Chris D. Jones
**Writing – original draft:** Chris D. Jones, Tilo Ziehn, Jatin Anand, Ana Bastos, Eleanor Burke, Josep G. Canadell, Manoel Cardoso, Yolandi Ernst, Atul K. Jain, Sujong Jeong, Elizabeth D. Keller, Masayuki Kondo, Ronny Lauerwald, Tzu-Shun Lin, Guillermo Murray-Tortarolo, Gert-Jan Nabuurs, Mike O'Sullivan, Ben Poulter, Xiaoyu Qin, Celso von Randow, Marcos Sanches, Hanqin Tian, Yohanna Villalobos, Xuhui Wang, Jeongmin Yun
**Writing – review & editing:** Chris D. Jones, Tilo Ziehn, Jatin Anand, Ana Bastos, Eleanor Burke, Josep G. Canadell, Manoel Cardoso, Yolandi Ernst, Atul K. Jain, Sujong Jeong, Elizabeth D. Keller, Masayuki Kondo, Ronny Lauerwald, Tzu-Shun Lin, Guillermo Murray-Tortarolo, Gert-Jan Nabuurs, Mike O'Sullivan, Ben Poulter, Xiaoyu Qin, Celso von Randow, Marcos Sanches, Dmitry Schepaschenko, Anatoly Shvidenko, T. Luke Smallman, Hanqin Tian, Yohanna Villalobos, Xuhui Wang, Jeongmin Yun

in our predictions of how ecosystems will respond in the future and a large uncertainty in any remaining carbon budget. We use expert knowledge and assessment of the carbon budget and its components for different regions. By comparing these assessments against the global models we see how good their future projections are, both as an average across models and also teaching us how individual models and processes affect prediction skill. For each region, the average of CMIP6 models is generally consistent with the regional assessments giving us confidence that CMIP6 global models can be used to look at future changes at the regional scale.

## 1. Introduction

Earth System Models (ESMs) are used to project future climate and carbon sinks and inform the remaining carbon budgets (RCBs) which allow us to stay within target warming levels. These models remain the most comprehensive and complex tools available for making such projections, and IPCC assessments of RCBs have relied on these since IPCC AR5 (IPCC, 2014). The techniques for determining future carbon budgets have matured since AR5 and can now be broken down into different components (IPCC, 2018; Rogelj et al., 2019). In parallel, IPCC AR6 drew on modeling results from a new generation of ESMs which contributed output to the 6th Coupled Model Intercomparison Project (CMIP6, Eyring et al., 2016). Assessments of the RCB are centered on the CMIP multi-model mean as is standard when using projections without further lines of evidence to act as constraint, but the model spread is substantial and leads to a wide range of possible future carbon budgets. In AR5 the assessed carbon budget for 2°C had a spread of 570 PgC for the 33%–66% range and in AR6 this range was almost unchanged at 550 PgC. Thus, the large spread in model results reported in AR5 (Ciais et al., 2013) has not been reduced, and the dominant cause of this uncertainty remains the land carbon cycle and its response to environmental changes (Jones & Friedlingstein, 2020). This uncertainty therefore hinders the usefulness of the RCB concept, and it is an outstanding research priority to try to constrain our knowledge of future carbon cycle changes.

In parallel, stand-alone (or "offline") land-surface models are used to simulate the historical and contemporary carbon cycle for both global assessments (Friedlingstein, O'Sullivan, et al., 2022) and regional ones (Sitch et al., 2015). These offline land models are typically the same, or very similar, to the land-surface components of ESMs, but there is surprisingly little direct use made of offline simulation results to help inform or reduce uncertainty in results from coupled models. Analyses of the contemporary period also draw on many other lines of evidence in addition to the land models—such as atmospheric inversion modeling, forest inventories, national reporting of anthropogenic emissions and remote sensing products, all combined with data assimilation approaches (e.g., Bloom et al., 2016). So while offline models have a large hand in assessing the carbon budget to date, and coupled models in quantifying the RCB for the future, we believe there is an activity required to help join these communities to fully utilize available data and constraints to improve consistency over the historical, contemporary and future time periods.

Past attempts to evaluate ESMs can be broadly categorized along three orthogonal axes of focus: analysis along different timescales; analysis of different processes; and analysis at different regional and spatial scales. For example, Cox et al. (2013) and Bastos et al. (2016) explore the carbon cycle on interannual timescales, and Wenzel et al. (2016) and Graven et al. (2013) down to seasonal scales. Burke et al. (2013) and Koven, Riley, and Stern (2013) analyze permafrost representation; Kloster et al. (2011) and Burton et al. (2019) fire; and Erb et al. (2017) and Pongratz et al. (2018) discuss land-use and management. Davies-Barnard et al. (2020) assess representation of the nitrogen cycle and Goll et al. (2012) the phosphorus cycle. Across regions, Anav et al. (2013) perform a thorough evaluation of CMIP5 models, but at global and large latitude band scale, while activities such as MsTMIP (Huntzinger et al., 2013) look at single continental regions, or biomes (Ahlström et al., 2015), and some studies perform analysis at single sites such as FACE experiments (Walker et al., 2015) and forest droughting (Rowland et al., 2018). Here we extend the analysis of continental-scale regions to look systematically across CMIP6 projections compared with the REgional Carbon Cycle Assessment and Processes, Phase 2 (RECCAP2) assessments.

While there are papers looking at global constraints on climate and carbon cycle sensitivity (e.g., Cox et al., 2013; Wenzel et al., 2016), no activity to date has attempted to use directly the wealth of knowledge we have on the contemporary carbon cycle to reduce the wide spread of future projections. This is partly because there is a disconnect between the tools used for each job—although ESMs are often based around the same land-surface

models used in the Global Carbon Project budget updates and TRENDY (Friedlingstein, Jones, et al., 2022), the method of driving offline land surface models differs sufficiently that these do not provide an easy opportunity to constrain projections from coupled ESMs. The first RECCAP assessment (e.g., Sitch et al., 2015) brought together land models and regional-specific data and expertise to quantify regional carbon budgets. This activity is repeated in RECCAP2 (Poulter et al., 2022; https://www.globalcarbonproject.org/reccap/index.htm) enabling new insights into the drivers and changes in regional scale carbon cycle processes. This brings an opportunity to make use of this vast assembly of information to improve our ability to predict changes and constrain future projections of carbon stocks and fluxes.

Much activity to reduce spread in future projections comes from optimizing and constraining model parameters and processes (i.e., trying to improve the models before we run them). This is a valuable activity and leads to improved models with reduced errors (Eyring et al., 2019). However, as is the case with climate projections and quantities such as climate sensitivity, the approach of improving models in itself has not proved successful in terms of reducing uncertainty. A second approach—to constrain results in a posterior sense may prove more fruitful. A range of papers have attempted to do this for climate sensitivity based on observations (Sherwood et al., 2020); aerosol forcing (Bellouin et al., 2020) and rates of global warming based on recent trends (Y. Liang et al., 2020; Ribes et al., 2021; Tokarska et al., 2020). For the carbon cycle, the use of posterior methods has an advantage that different constraints can be utilized for different regions. For example, interannual variability might be best suited for constraining future tropical land carbon changes (Cox et al., 2013) while changes in the seasonal cycle reflect mid- to high-latitude changes (Wenzel et al., 2016).

This paper is an initial attempt to develop techniques to derive greater benefit from present day assessments in identifying short-comings in global model projections and eventually constraining future behavior of the carbon cycle simulated by ESMs. It is not possible in a single study to perform and combine all the possible constraints which could be inferred from such a wealth of understanding of regional carbon cycle processes, but we hope it provides guidance for more detailed and complete model evaluation and sparks research into extracting greater value from present day analysis to constrain future projections. We highlight regions specifically where global ESMs are in good agreement with each other and/or with expert assessment. For regions where there is pronounced disagreement, we discuss potential for both *a prior* improvement via model development and *posterior* constraint via regional assessments such as RECCAP2.

## 2. Methods

We present here global and regional carbon cycle projections from CMIP6. As per standard usage of such ensembles we present the multi-model mean, but we also delve into individual model performance for simulation of carbon fluxes and carbon storage on regional scales and attempt to link this to the process completeness of each model.

The justification of using the unweighted multi-model mean is that there remains no better agreed way of combining outcomes from an unbalanced "ensemble of opportunity" like CMIP6. Members of the ensemble differ in complexity, skill, independence from each other and in number of ensemble members submitted for analysis. Multiple ways exist to try to select, filter or weight outputs using either or both model skill or independence but with the exception of global temperature no technique has been proven to add skill to the resulting mean (Lee et al., 2021, Box 4.1). Knutti et al. (2017) argue against the traditional "one model one vote" democracy, but show that success of weighting will be very application dependent. Quantities more closely related to global temperature (such as sea-ice) may be more amenable to a model weighting approach. Sanderson et al. (2017) do apply a regional weighting to projections from global models and discuss merits of univariate or multi-variate constraints showing that weighting can be done based on both skill and uniqueness, but Abramowitz et al. (2019) are clear that weighting schemes must be very carefully tested with out-of-sample data in order to avoid overfitting due to the generally limited sample size of the ensemble. They conclude that model dependence is unlikely to be able to be defined in a universally unambiguous way. Weigel et al. (2010) go further and caution against weighting unless uncertainty in the constraint can be properly accounted for. They show that the multi-model mean often outperforms even the best model in an ensemble such as CMIP6 and that poorly chosen constraint metrics can actually be counter-productive. This is particularly true when internal variability is large (which is the case for terrestrial carbon fluxes)—in this case more information may be lost by inappropriate weighting than gained. An example
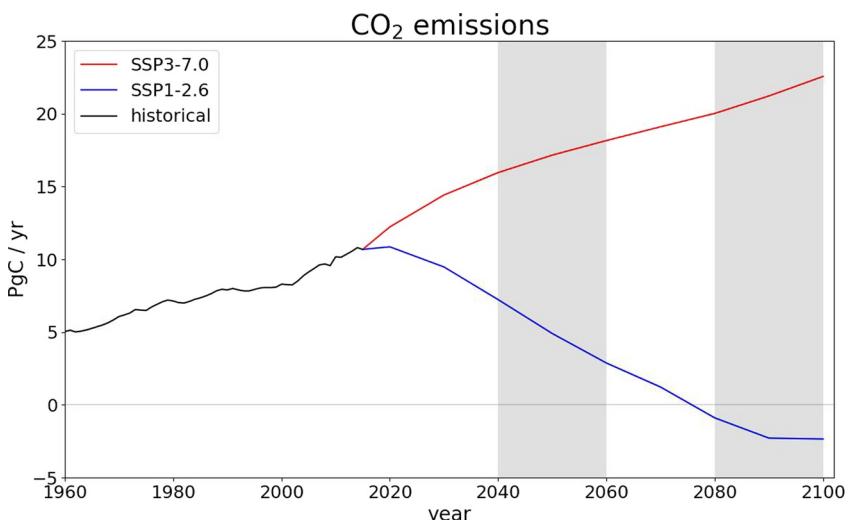
**Figure 1.** Historical and future anthropogenic $CO_2$ emissions (from fossil fuel and land-use change) following two future scenario pathways (SSP1-2.6, blue and SSP3-7.0, red) and showing the periods of focus from 2041 to 2060 and 2081-2100.

is the model selection used in AR5 (Collins et al., 2013) for sea-ice projections which was subsequently shown to be overly restrictive due to internal variability (Notz, 2015; Stroeve & Notz, 2015).

In the case of the carbon cycle we know that model differences are dominated by land rather than ocean carbon uptake. Hewitt et al. (2016) showed that model differences for land carbon outweighed internal variability and choice of scenarios. This was confirmed by Lovenduski and Bonan (2017) who showed very limited ability to constrain future projections even if they applied an impractical level of precision to the observational constraint. Both Hoffman et al. (2014) and Friedlingstein et al. (2014) found correlations between present day $CO_2$ biases in emissions-driven simulations suggesting a possible route to constraint, but Booth et al. (2017) warn against over-constraining the results due to unaccounted for uncertainty in land-use emissions. Further to this, Jones et al. (2013) found different future biases from the same models but for different scenarios in CMIP5 simulations implying that the balance of feedbacks may be scenario dependent and cannot be done in an absolute sense at present day. Given that the terrestrial carbon cycle has large interannual variability (Braswell et al., 1997; Jones & Cox, 2001) we conclude that it is as yet premature to attempt to constrain future projections using a model weighting or selection approach.

In this paper we analyze results from 16 CMIP6 models (see Table S1 in Supporting Information S1). RECCAP2 has identified a focus study period of the decade 2010–2019, and 10 land regions which we use for our comparison with CMIP6. These regions are: North and South America, Central-, East-, South-, and South East Asia, Europe, Africa, Australasia, and Russia (Figures 2 and 11 in Poulter et al. (2022)). Additionally, we compare results for the permafrost region which cuts across some of the defined region but carries particular interest for future carbon cycle response to climate change.

For consistency with IPCC projections (Lee et al., 2021), we assess future carbon sinks under a high and low future emission scenario (SSP1-2.6, SSP3-7.0) and focus on two future time horizons (mid-term: 2041–2060, and the long-term: 2081–2100: Figure 1). For slow-changing quantities such as carbon storage we show results for the years 2050 and 2100, while for quantities such as carbon fluxes with substantial year-to-year variations we show means over these 20-year periods.

We perform a filtering for each carbon cycle variable in order to assess how many models perform well for each region and how many regions are well simulated for each model. We calculate the mean over the 2010–2019 period from each CMIP6 model and compare with the best estimate and uncertainty from the RECCAP2 assessment. This enables us to count models consistent with RECCAP2 for each region and variable. We then show projections into the future as both the mean and range of the full CMIP6 ensemble and also the sub-ensemble of models passing the filtering at present day. We note there is inevitably some circularity in this comparison where RECCAP2 assessments draw on very similar land models used in CMIP6 ESMs. This is unavoidable and is one reason discussed later we do not proceed to make quantitative constraints from this evaluation, but
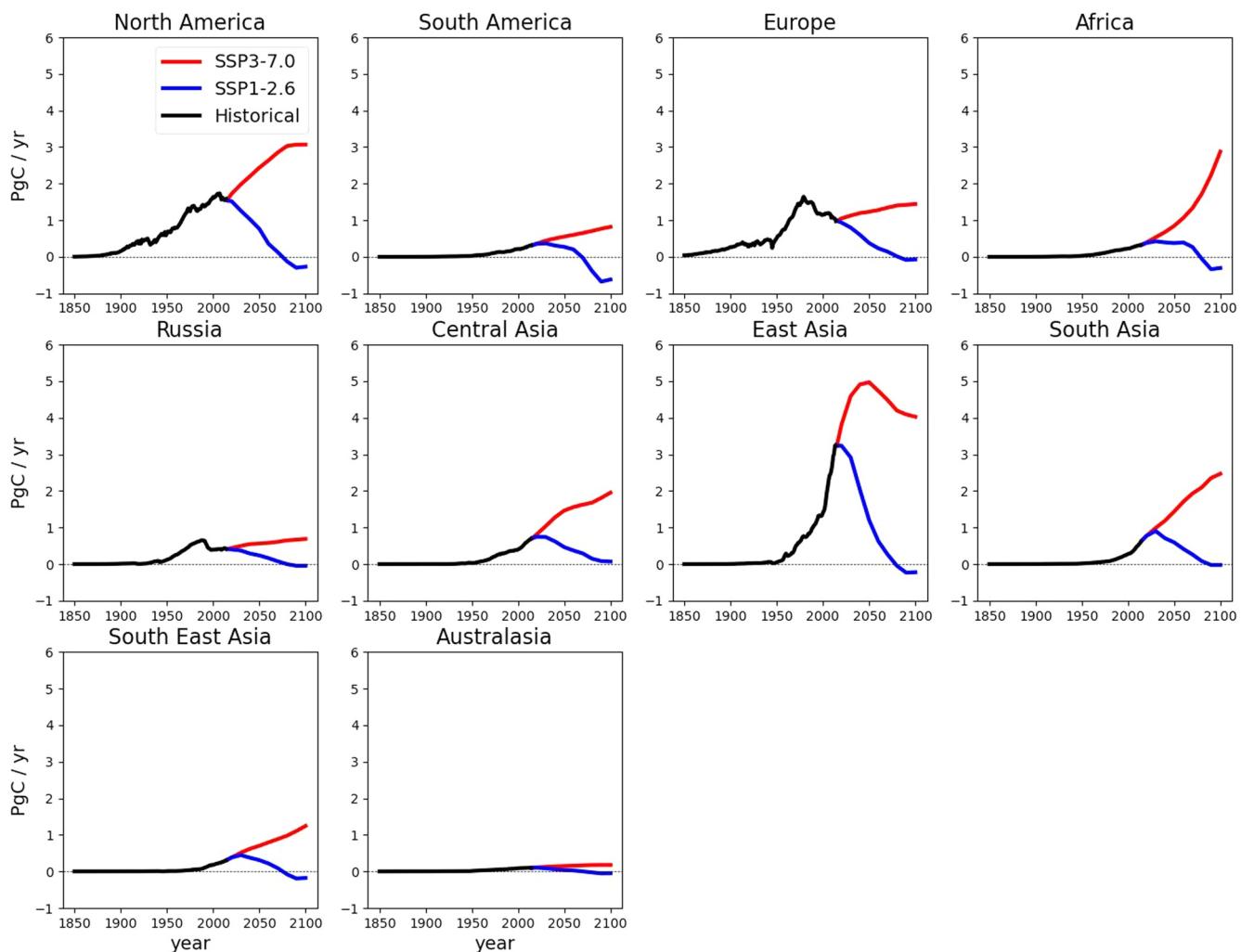
**Figure 2.** Regional fossil fuel emissions for the historical period (black) and future scenarios SSP1-2.6 (blue) and SSP3-7.0 (red). In SSP1-2.6 there are large quantities of bioenergy with carbon capture and storage (BECCS) which create negative emissions in the energy sector, leading to some regions exhibiting net-negative anthropogenic emissions by 2100.

the comparison is valid because the TRENDY models are constrained by use of observed climate data and the RECCAP2 assessments bring in other, regionally specific datasets and expert assessment.

In addition to simulated changes in land carbon stock, we also assemble regional fossil fuel emissions as specified in the SSP1-2.6 and SSP3-7.0 scenarios for context. We note though that these are not the output from the ESMs, and that the behavior of global and local climate and the carbon cycle is dependent on global-scale rather than regional-scale emissions. These anthropogenic emissions by region are shown in Figure 2. All regions exhibit similar trends, with a marked increase over the present day for SSP3-7.0 and a decrease for SSP1-2.6. Some regions, such as Europe and North America, have recent decreases in emissions before these scenarios diverge while most others have yet to peak. Globally, SSP1-2.6 has net negative emissions after about 2070, but again this is regionally differentiated as discussed in IPCC AR6 WG3 Summary for Policy Makers (IPCC, 2022). Figure S1 in Supporting Information S1 shows the land-use change in the two scenarios by RECCAP2 region.

## 3. Earth System Model Projections From CMIP6

### 3.1. Global

Projected global climate changes from the SSP1-2.6 and SSP3-7.0 scenarios are presented in detail in the IPCC AR6 report (Lee et al., 2021); a brief overview is shown here in Figure 3. Compared to the recent past
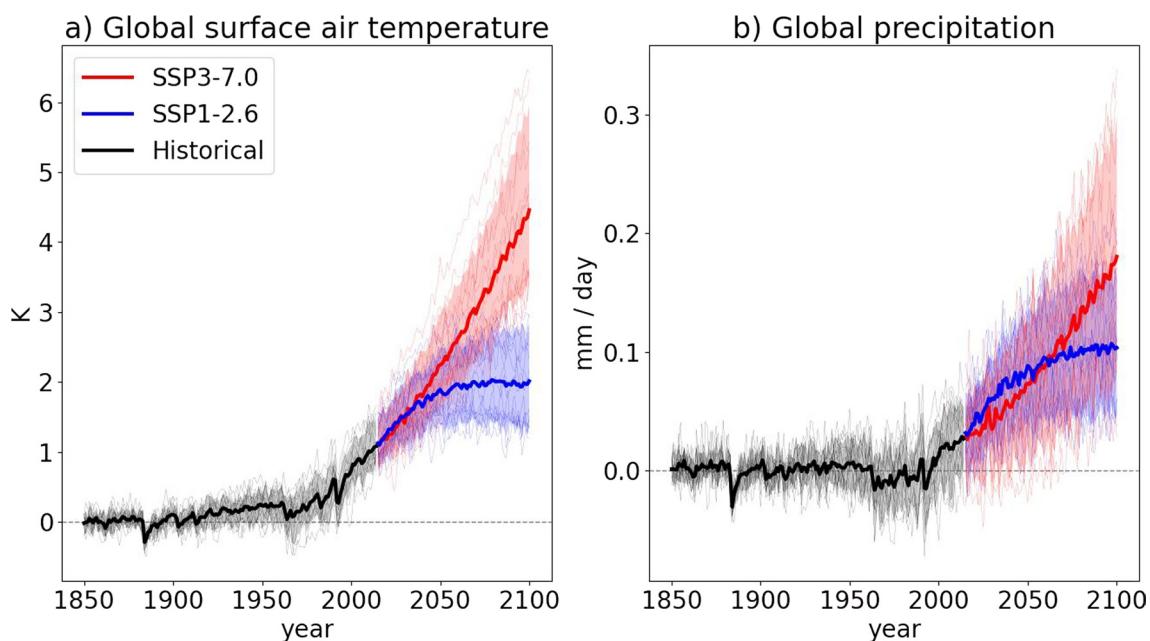
**Figure 3.** Change, since 1850, in (a) global temperature and (b) precipitation projections from CMIP6.

(1995–2014), global surface air temperature is assessed to increase by 0.5–1.3°C and 0.9–1.7°C for the period 2041–2060, respectively for scenarios SSP1-2.6 and SSP3-7.0; and increasing to 0.5–1.5°C and 2.0–3.7°C for the period 2081–2100. Global precipitation over land is projected to increase under all scenarios with increases of 2.8% and 2.5% for the period 2041–2060, respectively for scenarios SSP1-2.6 and SSP3-7.0; and increasing to 3.3% and 5.8% for the period 2081–2100. Uncertainty in these projections is dominated by the uncertainty in climate sensitivity (i.e., the response of the global climate system to a doubling of $CO_2$ in the atmosphere), although regional changes in precipitation are highly variable between climate models. The near-term divergence between these scenarios may be driven by aerosols (SSP3-7.0 has higher aerosol emissions than any other SSP) and is particularly evident over Asia (Figure S3 in Supporting Information S1).

Corresponding changes in the global carbon cycle are shown in Figure 4. As with CMIP5 projections shown in AR5 (Ciais et al., 2013), projected carbon cycle changes show substantial uncertainty dominated by the land (Jones & Friedlingstein, 2020). Most ESMs agree that vegetation carbon declined during the early part of the twentieth century due to land-use change but increased toward the present day and that this increase continues into the future. ESMs are generally consistent with assessed present-day sinks but for both land and ocean are slightly lower than those estimated by the Global Carbon Budget (Figure 4) of $3.1 \pm 0.6$ for the land and of $2.9 \pm 0.4$ PgC yr$^{-1}$ for the ocean during the decade 2012–2021 (Friedlingstein, O'Sullivan, et al., 2022). During the 21st century, land and ocean sinks show similar responses, but the land sink has greater variability and spread between ESMs. Under the higher emissions and $CO_2$ level of SSP3-7.0, sinks grow initially due to increasing atmospheric concentrations of $CO_2$ but are subsequently limited by emerging carbon–climate feedbacks as the climate warms: both land and ocean sinks are projected to stop growing from the second part of the 21st century even as $CO_2$ continues to increase. Under the lower emissions of SSP1-2.6, the weakening growth rate of carbon sinks is a response to the declining atmospheric $CO_2$ concentrations toward the end of the century. However, while future projected ocean carbon sinks are clearly differentiated under different future scenarios (in common again with CMIP5: Hewitt et al., 2016), the spread for future land fluxes is bigger between models than across these very different scenarios.

Aggregating across models allows us to see that the future land and ocean carbon sinks remain of comparable magnitude under both high and low future emissions and for mid and long-term time horizons (Figure 5). This is in agreement with the assessed fractional uptake by land and ocean shown in IPCC AR6 Summary for Policy Makers (IPCC, 2021). For the mid-term, the two scenarios have not diverged markedly, and both land and ocean have taken up approximately 40 PgC over the 35 years from the present day (taken here as 2015) to 2050. By
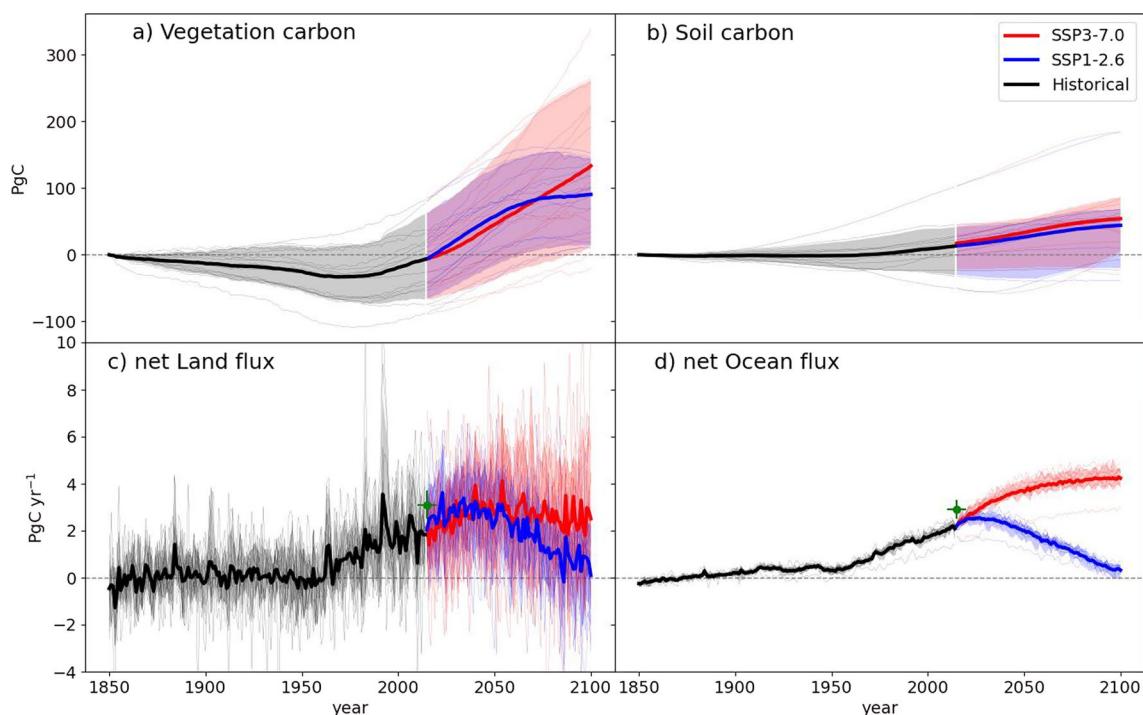
**Figure 4.** CMIP6 simulation of historical (black) and projections of future carbon cycle variables for SSP1-2.6 (blue) and SSP3-7.0 (red). Stocks are shown as changes since 1850 for: (a) global vegetation carbon; (b) global soil carbon; fluxes are shown as (c) net land flux (nbp); (d) net ocean flux (fgco2). Global Carbon Project estimates for recent land and ocean sinks ("$S_{LAND}$" and "$S_{OCEAN}$") are shown as green dots and error bars from Friedlingstein, O'Sullivan, et al. (2022).

2100 though the higher $CO_2$ in SSP3-7.0 has driven larger sinks (albeit a smaller fraction of human emissions) of between 100 and 150 PgC compared with around 60 PgC in SSP1-2.6.

## 3.2. Regional

In order to enable direct comparison with the findings of RECCAP2, CMIP6 data is assessed on the RECCAP2 regions. A summary of climate and carbon cycle changes is summarized here before we provide direct comparison with RECCAP2 assessment in Section 4.

Regional climate change is not uniform, and these regional differences may drive different responses of the carbon cycle. Supporting Information S1 shows CMIP6 projected changes in temperature and precipitation over the RECCAP2 regions (Figures S2 and S3 in Supporting Information S1). The expected signal of SSP3-7.0 warming more rapidly than SSP1-2.6 is clear for all regions by 2100, although at mid-century interannual variability obscures the difference between high and low emissions pathways. The picture for rainfall is more nuanced with most regions exhibiting substantial interannual variability. Some regions, such as North America or Russia, show a clear model signal toward getting wetter, especially under the high emissions scenario. In contrast, some, such as South America, show a weak trend or even drying.

Simulated changes in terrestrial carbon stocks are shown in Figure 6 (vegetation carbon) and Figure 7 (soil carbon). As with the global total some signals are clear—such as a reduction in biomass due to land use change simulated in many regions followed by more recent increases, which are stronger under SSP3-7.0. Equally, the spread between models is large and the signal of different future scenarios is small compared to the uncertainty between-models. This is especially true for soil carbon (Anav et al., 2013) which is not well represented by models in terms of total carbon stocks due to larger difference in turnover times (Carvalhais et al., 2014). Not all models include treatment of land-use change and this drives large inter-model spread in regions where this is an important forcing such as over Africa under SSP3-70. Here we show changes in vegetation and soil carbon stocks so that all models can be included from a zero baseline at 1850 and the plumes show differences in simulated gains and losses. In the next section, we describe region-by-region absolute amounts of carbon stocks and fluxes.
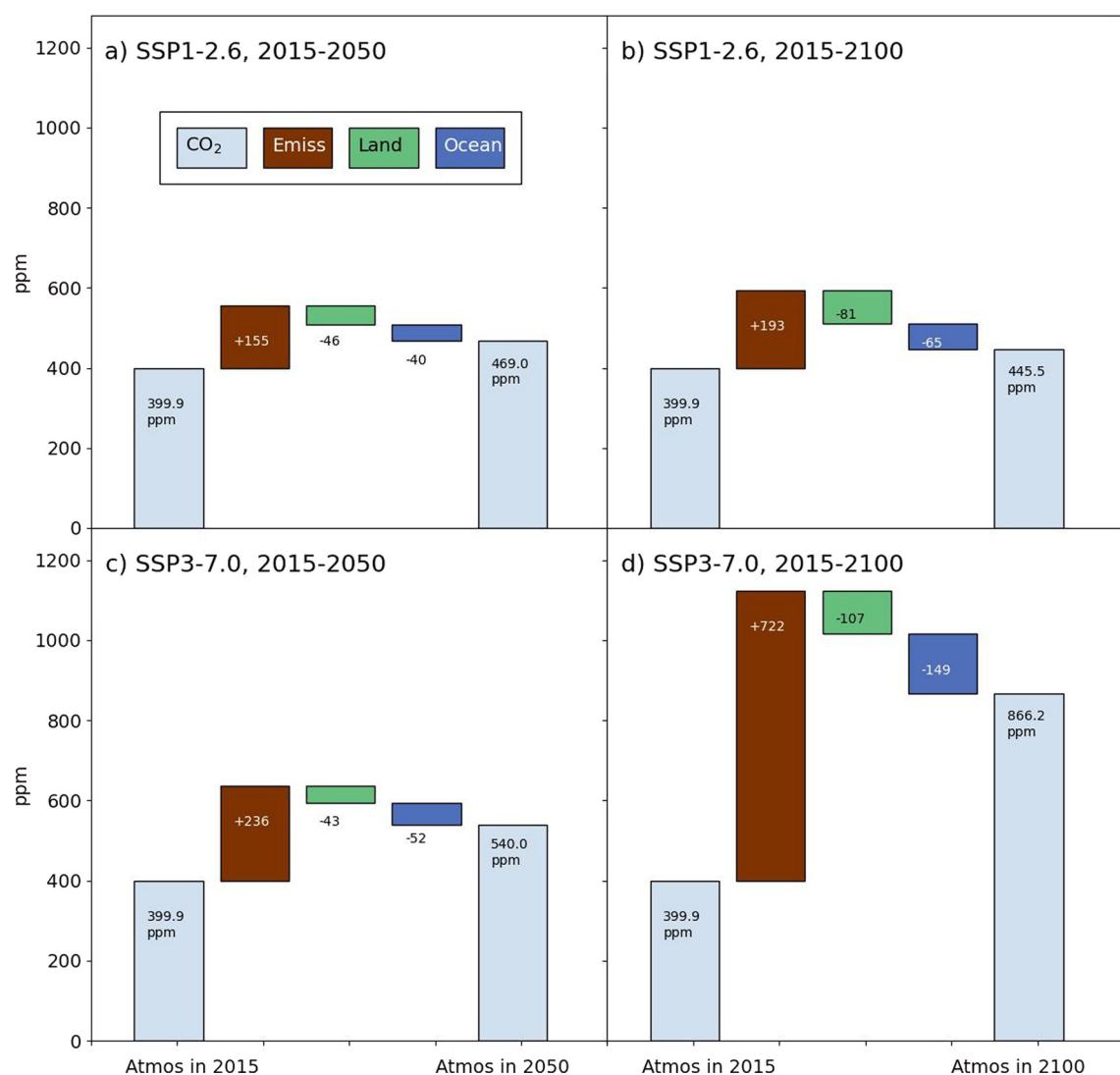
**Figure 5.** CMIP6-simulated changes in the global carbon budget for the two scenarios SSP1-2.6 (top row) and SSP3-7.0 (bottom row) and the two focus periods up to 2050 (left-hand column) and up to 2100 (right-hand column).

## 4. Regional Assessment of Carbon Cycle Projections

In this section we present a region-by-region comparison of the carbon stocks and fluxes simulated by CMIP6 and assessed by the RECCAP2 chapter teams. The data and its derivation are presented in detail in the regional chapters of the RECCAP2 assessment (special collection across multiple journals of the American Geophysical Union). Here we assemble top-level summary statistics of carbon stocks and fluxes for each RECCAP2 region in a common table to enable a direct comparison with that same region as simulated by CMIP6 ESMs. Figures show how the historical simulations from CMIP6 compare with the present-day RECCAP2 assessed values before showing projections into the future under the low and high emissions scenarios.

In this section, we discuss specific responses for each region and options to constrain the projection in future work. In the next section we synthesize strengths and weaknesses of each ESM and analyze systematic issues associated with regions or process inclusion in the models. Where there is poor agreement between CMIP and RECCAP we discuss why this may be. For each region we structure the analysis the same way and compare ESM simulated fluxes and stocks with RECCAP2 assessed values. Briefly, we compare gross and net primary productivity (GPP, NPP) and also two measures of net flux: net ecosystem productivity, NEP (defined as the difference between NPP and heterotrophic respiration) and net biome productivity, NBP (defined as the total flux in or out
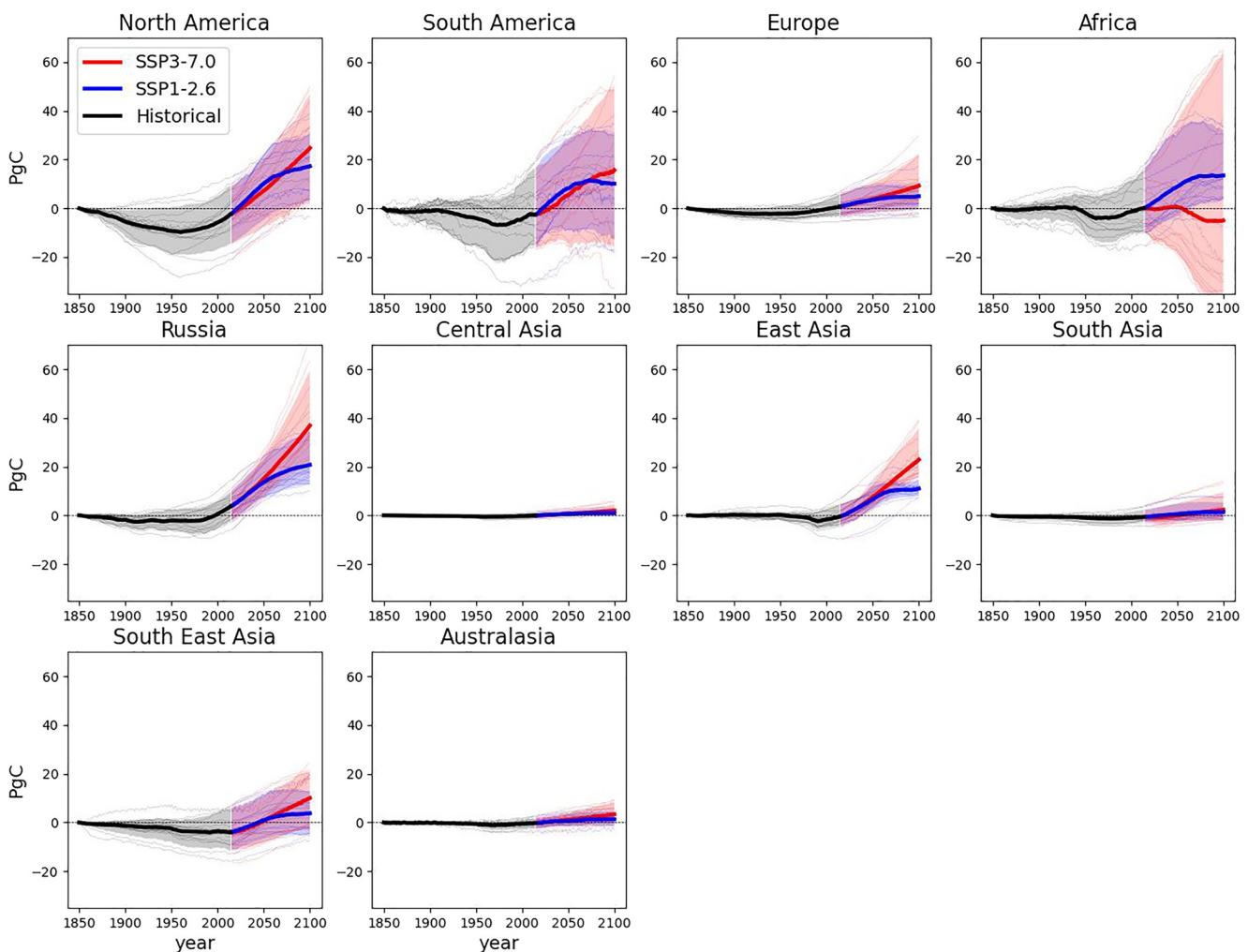
**Figure 6.** CMIP6 projected changes in vegetation carbon for SSP1-2.6 (blue) and SSP3-7.0 (red) for the REgional Carbon Cycle Assessment and Processes, Phase 2 regions.

of the land carbon stock). NBP is the same as net ecosystem exchange as defined by Ciais et al. (2022), but here taken with a sign convention that positive reflects a *sink* into the land carbon stock. The reason for comparing these two measures of net flux is that most CMIP6 ESMs lack some of the relevant processes (such as fire, harvest and lateral transport) that would be required to enable a direct comparison with regional estimates of the total flux, NBP. Therefore, although NBP may be more relevant in terms of capturing the full carbon balance of a region, NEP is more easily defined and compared directly between modeled and observed estimates. We also compare carbon stocks in vegetation and soils (taking the sum of CMIP variables "cSoil + cLitter") and the total terrestrial carbon stock. For brevity, we show the simulated values of GPP, NBP and total terrestrial stock up to 2100 for the two scenarios (SSP1-2.6 and SSP3-7.0). As described in Section 2, we show examples of how a possible filtering of CMIP6 outputs could be used to constrain future projections. The science behind such filtering is not well developed and it is not possible to use such a simple technique to provide a robust constraint. Hence we use it to illustrate how a comparison with regional assessments may influence future projections, but we do not quantitatively assemble these constrained values into a new prediction of future carbon changes. Further justification of this approach is discussed in Section 5.

### 4.1. North America

For North America there is broad agreement between CMIP6 models and RECCAP2 best estimates for all C fluxes and pools (Table 1). The multi-model mean is close to the assessed value for all of GPP, NBP and total
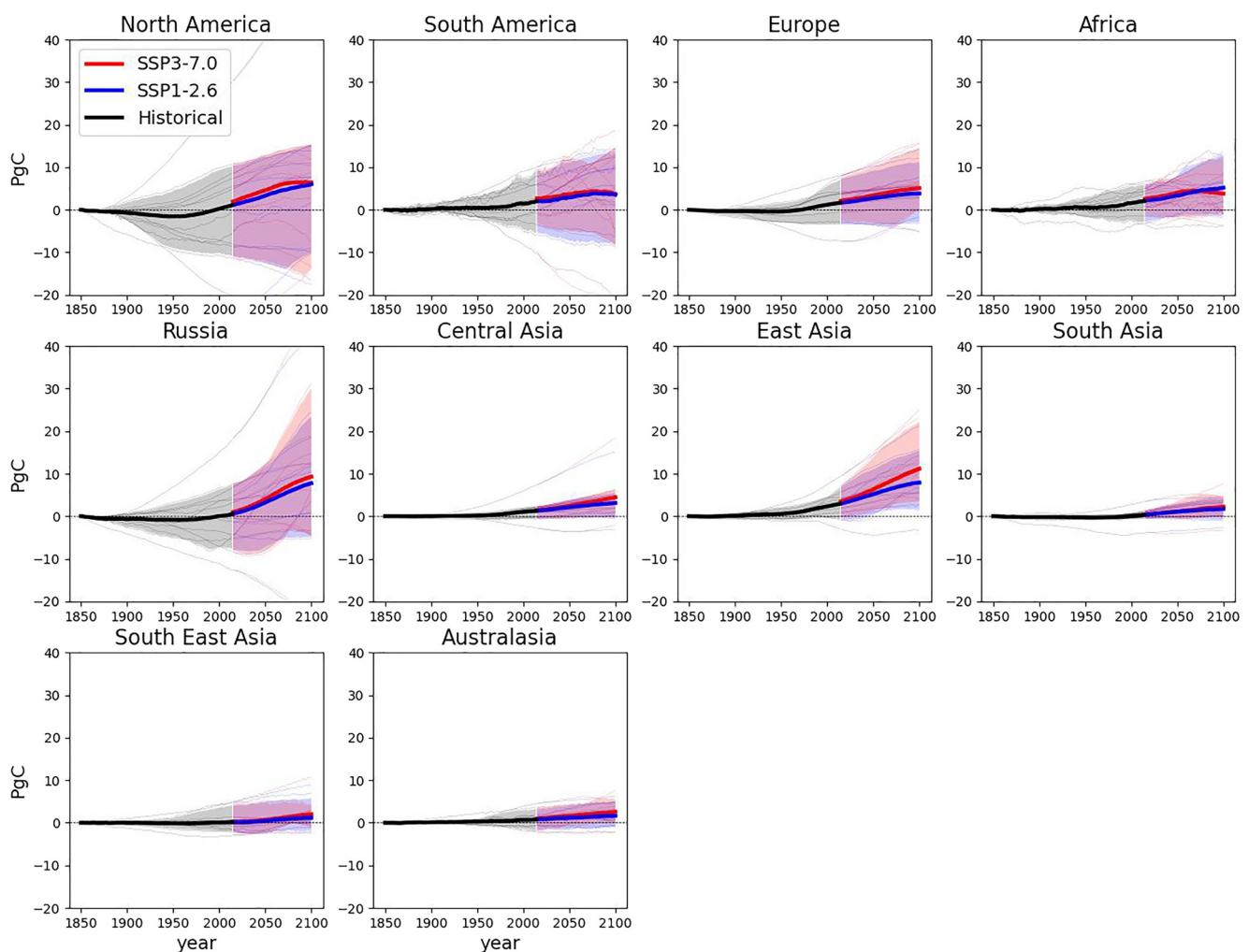
**Figure 7.** CMIP6 projected changes in soil carbon for SSP1-2.6 (blue) and SSP3-7.0 (red) for the REgional Carbon Cycle Assessment and Processes, Phase 2 regions.

terrestrial carbon ($C_{tot}$), with 9, 11 and 8 models passing the filtering respectively. Without fossil-fuel emissions, the region is consistently recognized as a C sink over last decade with an estimated magnitude of $0.35 \pm 0.32$ PgCyr$^{-1}$ in the CMIP6 models and $0.41 \pm 0.23$ PgCyr$^{-1}$ in RECCAP2, which is also consistent with other

**Table 1**
*Comparison of CMIP6 Simulated and REgional Carbon Cycle Assessment and Processes, Phase 2 Assessed Carbon Stocks and Fluxes for N. America*

|  |  | CMIP6 (mean ± 1 sigma) | RECCAP2 (best estimate ± error estimate) |
|---|---|---|---|
| Fluxes PgC/yr mean over 2010–2019 inclusive | GPP | 17.46 ± 2.49 | 16.79 ± 2.10 |
|  | NPP | 8.41 ± 1.69 | 8.87 ± 1.04 |
|  | NEP | 1.02 ± 2.42 | 2.77 ± 0.76 |
|  | NBP | 0.35 ± 0.19 | 0.41 ± 0.23 |
| Stocks PgC mean over 2010–2019 inclusive | Vegetation carbon | 73.64 ± 26.86 | 53.3 ± 18.5 |
|  | Soil carbon | 313.11 ± 207.03 | 301.6 ± 206.8 |
|  | Product carbon | 1.20 ± 0.99 | 10.4 |
|  | Total terrest. carbon | 404.64 ± 204.49 | 382.6 ± 208.1 |

*Note.* Values are shown as the average over a 10 year period from 2010 to 2019, and ranges are one standard deviation unless stated otherwise. For model outputs this reflects spread across different models, while for assessed values this can include other means of assessing uncertainty. Units are PgC or PgC yr$^{-1}$.
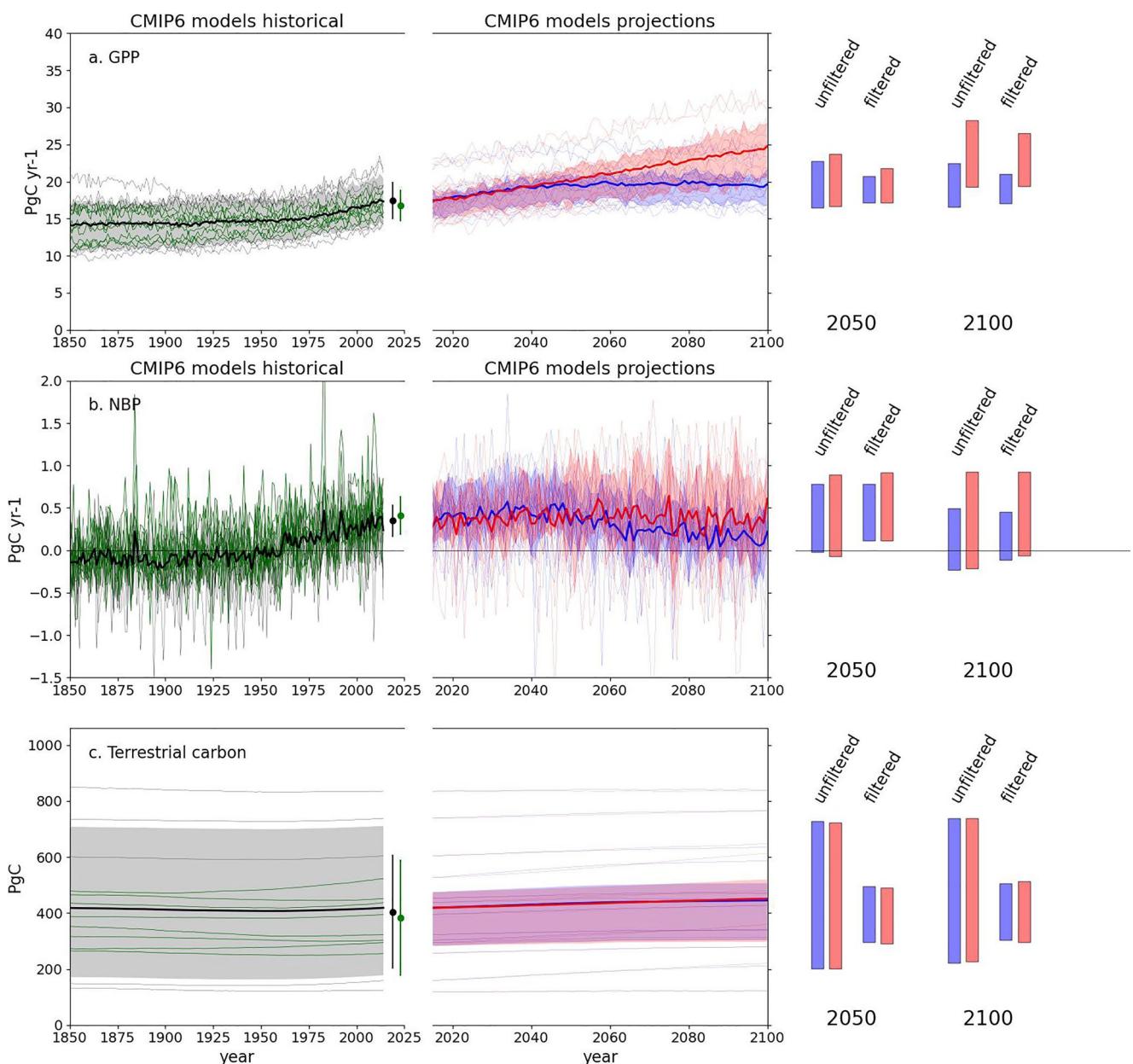
**Figure 8.** CMIP6 simulated GPP (top), NBP (middle) and total terrestrial carbon stock (bottom) for the historical period (1850–2014, left hand side) and future projections (to 2100, right hand side) for N. America. The panels show individual CMIP6 models in faint lines. For the historical period, individual models are shown in gray and multi model mean in thick black. CMIP6 mean and standard deviation are shown by the black dot and error bar, and REgional Carbon Cycle Assessment and Processes, Phase 2 (RECCAP2) assessed best estimate and uncertainty by the green dot and error bar. Models which fall within the RECCAP2 assessed range are shown in pale green. Future projections are for SSP1-2.6 (blue) and SSP3-7.0 (red), with single models in faint lines and multi-model mean in thick lines. Filtering just those models which match the RECCAP2 assessed range gives the red and blue shaded region. Mean values for 2050 and 2100 before and after filtering are shown to the right of the plots.

estimates (see e.g., Cavallaro et al., 2018; Murray-Tortarolo et al., 2022). As a result, the projected total future land C in the region changed very little in both scenarios after applying the filtering, while nonetheless reducing their uncertainties.

Both GPP and $C_{tot}$ show good agreement between CMIP6 outputs and RECCAP2 assessment (Figure 8). There is a strong overlap in the estimates with a close alignment of both the central estimates and uncertainty ranges for GPP and $C_{tot}$. RECCAP2 assessed values are slightly lower leading to a slight reduction in values and narrowing of the range after filtering. The results show an ongoing increase in GPP under SSP3-7.0 while GPP does not

**Table 2**
*Comparison of CMIP6 Simulated and REgional Carbon Cycle Assessment and Processes, Phase 2 Assessed Carbon Stocks and Fluxes for S. America*

| | | CMIP6 (mean ± 1 sigma) | RECCAP2 (best estimate ± error estimate) |
|---|---|---|---|
| Fluxes PgC/yr mean over 2010–2019 inclusive | GPP | 33.72 ± 5.81 | 33.02 ± 5.23 |
| | NPP | 14.14 ± 3.53 | 15.38 ± 3.47 |
| | NEP | 1.19 ± 4.76 | 1.51 ± 5.56 |
| | NBP | 0.21 ± 0.41 | Bottom-up: +0.7 ± 3.4 |
| | | | Top-down: +0.083 ± 0.55 |
| Stocks PgC mean over 2010–2019 inclusive | Vegetation carbon | 140.38 ± 29.43 | 94.1 ± 37.3[a] |
| | Soil carbon | 158.28 ± 74.34 | 220.6 ± 66.2[b] |
| | Product carbon | 1.12 ± 1.00 | |
| | Total terrest. carbon | 315.11 ± 81.78 | 314.7 ± 76.0 |

*Note.* Values are shown as the average over a 10 year period from 2010 to 2019, and ranges are one standard deviation unless stated otherwise. For model outputs this reflects spread across different models, while for assessed values this can include other means of assessing uncertainty. Units are PgC or PgC yr$^{-1}$.
[a]vegetation carbon estimated from Spawn et al. (2020). [b]Soil carbon estimated from Poggio et al. (2021).

increase under SSP1-2.6 after mid 21st century. This leads to the two scenarios diverging significantly by 2100. However, this does not translate into similar increases in carbon stocks with both scenarios showing very similar changes (slight increases) on both 2050 and 2100 time horizons.

Despite this agreement, some key processes of the C cycle of the region are not necessarily well characterized either in the models or in our best estimations in the RECCAP2 project. For example, mountain pine beetle outbreaks have been shown to drive large pulses of C to the atmosphere, via increased tree mortality and reduced C uptake; a process widely missing from both estimations (Kurz et al., 2008). Similarly, disturbance—particularly fire and drought—seems to be exerting increasingly larger control on the interannual variability of the C uptake (Murray-Tortarolo et al., 2022; J. A. Wang et al., 2021), processes likely to strengthen in the future, and most of which are not represented in most ESMs. Finally, both estimates lack the inclusion of lateral fluxes, such as wood and crop trade, or the C transport to the ocean. Thus, future efforts to constrain future C projections in the region should focus on further benchmarking model results based on the inclusion of these fluxes.

### 4.2. South America

CMIP6 simulated, and RECCAP2 assessed values of South American carbon fluxes and stocks are summarized in Table 2. The multi-model mean is close to the assessed value for GPP and biased low for NBP but within the error bars and close for $C_{tot}$, with 10, 16 and 11 models passing the filtering respectively. With extensive tropical forests, GPP and NPP are high compared to other regions, and the CMIP6 simulations are in good agreement with these estimates. However, the strong productivity of the land biosphere is balanced with high emission fluxes from deforestation and wildfires (Assis et al., 2020; Harris et al., 2021) and the resulting carbon budget is particularly difficult to quantify, due to challenging methodological difficulties associated with sampling such complex ecosystem processes. Fire emissions alone, based on the Global Fire Emissions Database version 4 (GFED4s, Van Der Werf et al., 2017), amount to 0.285 PgC yr$^{-1}$ in 2010–2019, mostly due to fires in Brazil (68%) and Bolivia (15%), followed by Venezuela, Argentina, Paraguay and Colombia (15%).

The estimates for NBP from RECCAP2 that combine process-based models (TRENDY v9) and a land data assimilation system (CARDAMOM; Bloom et al., 2016) referred here as "bottom-up" estimate, and an ensemble of atmospheric inversion models, referred here as "top-down" estimate, both show small regional sinks, but are not distinct from carbon neutrality given their large uncertainty ranges. For alternative data sources for carbon stocks in the regions, assessment from remote sensing products estimate a range of 94–120 PgC stored in the vegetation carbon (Gloor et al., 2012) and around 220 PgC stored in soils (Hengl et al., 2017). The ensemble of CMIP6 models tend to estimate higher values for the vegetation carbon stock and lower for soil carbon, but both within the large uncertainty bounds.

Filtering of GPP leads to a lower magnitude and range than from the full set of CMIP6 models (Figure 9). This is because, despite general good agreement between CMIP6 and RECCAP2 mean and uncertainty values,
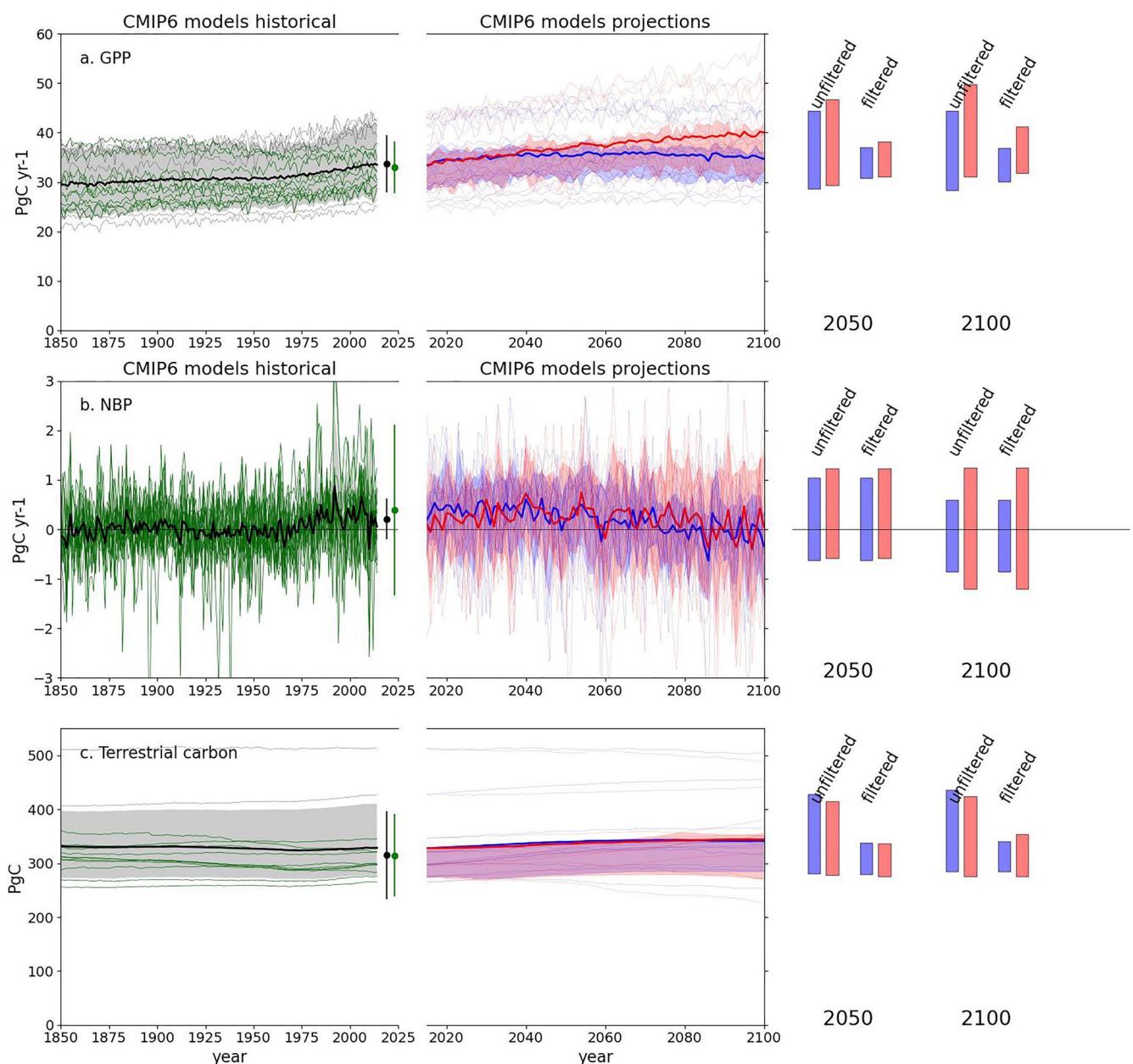
**Figure 9.** (top) CMIP6 simulated GPP, (middle) NBP, and (bottom) total terrestrial carbon stock for the historical period (1850–2014, left hand side) and future projections (to 2100, right hand side) for S. America. The panels show individual CMIP6 models in faint lines. For the historical period, individual models are shown in gray and multi model mean in thick black. CMIP6 mean and standard deviation are shown by the black dot and error bar, and REgional Carbon Cycle Assessment and Processes, Phase 2 (RECCAP2) assessed best estimate and uncertainty by the green dot and error bar. Models which fall within the RECCAP2 assessed range are shown in pale green. Future projections are for SSP1-2.6 (blue) and SSP3-7.0 (red), with single models in faint lines and multi-model mean in thick lines. Filtering just those models which match the RECCAP2 assessed range gives the red and blue shaded region. Mean values for 2050 and 2100 before and after filtering are shown to the right of the plots.

the filtering removes a few of the models at the top end of ESM results thus reducing future projections of GPP. The results show a small, but ongoing increase in GPP under SSP3-7.0 while GPP stops its increase under SSP1-2.6 after mid 21st century. It is likely that the larger warming in this region under SSP3-7.0 reduces future increases in GPP, and thus lead to the two scenarios diverging less significantly by 2100 than in temperature or higher-latitude regions. A similar trend is seen for $C_{tot}$, but in this case the very close agreement between present day values is a compensation of CMIP6 models have significantly greater biomass and smaller soil carbon than regional assessments—again highlighting the dangers of a simplistic filtering approach.

**Table 3**
*Comparison of CMIP6 Simulated and REgional Carbon Cycle Assessment and Processes, Phase 2 Assessed Carbon Stocks and Fluxes for Europe*

|  |  | CMIP6 (mean ± 1 sigma) | RECCAP2 (best estimate ± error estimate) |
|---|---|---|---|
| Fluxes PgC/yr mean over 2010–2019 inclusive | GPP | 5.04 ± 0.89 | 5.48 ± 0.5 |
|  | NPP | 2.57 ± 0.64 | Not available from RECCAP2 assessment |
|  | NEP | 0.40 ± 1.00 | 0.86 ± 0.43 |
|  | NBP | 0.09 ± 0.08 | 0.10 ± 0.05 |
| Stocks PgC mean over 2010–2019 inclusive | Vegetation carbon | 16.61 ± 9.15 | Not available from RECCAP2 assessment |
|  | Soil carbon | 61.13 ± 42.34 | Not available from RECCAP2 assessment |
|  | Product carbon | 0.50 ± 0.48 | Not available from RECCAP2 assessment |
|  | Total terrest. carbon | 83.38 ± 46.30 | Not available from RECCAP2 assessment |

*Note.* Values are shown as the average over a 10 year period from 2010 to 2019, and ranges are one standard deviation unless stated otherwise. For model outputs this reflects spread across different models, while for assessed values this can include other means of assessing uncertainty. Units are PgC or PgC yr$^{-1}$.

### 4.3. Europe

CMIP6 simulated, and RECCAP2 assessed values of European carbon fluxes and stocks are summarized in (Table 3). The multi-model mean is close to the assessed for both GPP and NBP. Two and seven models pass the filtering for GPP and NBP respectively, whilst RECCAP2 did not assess carbon stocks.

While we do not attempt to evaluate CMIP6 trends in fluxes, we note the different contributions between managed and natural land in driving changes in GPP for Europe over the last decade. GPP in Europe has an increasing trend over 2000–2018 by 18.2 TgC yr$^{-2}$ (S. Liang et al., 2021), which when separated into natural ecosystems versus agricultural vegetation (cropland and pasture), using Hilda + data (Winkler et al., 2021) as mask, both contribute about half of the total absolute value of GPP. However, natural land drives the upwards trend, primarily due to an increase in areal GPP rates (+5.2 gC m$^{-2}$ yr$^{-2}$), and secondarily due to an increase in areal extent of natural vegetation at the expense of agricultural area (+5 tsd. Km$^2$/yr). Agricultural GPP does not show any significant trend during this period. From regional inversions, RECCAP2 finds a tendency to predict $CO_2$ sources in SW and E-Europe, but stronger sinks in rest of Europe (mainly in Central, Northern and SE Europe), while TRENDYv10 models simulate a net land sink in most of Europe. This indicates that increase in natural vegetation GPP is the likely main cause of net-C sink, stressing the need to better understand—and include in land models and CMIP ESMs—fuller description of land-use and management (Pongratz et al., 2018).

For Europe, RECCAP2 assessed GPP is slightly higher than the CMIP6 mean and the narrow uncertainty range results in the filtering leading to a marked decrease in future projection spread, but little change in mean value (Figure 10). In either case, though, both filtered and unfiltered results show an ongoing increase in GPP under SSP3-7.0 while GPP stops increasing under SSP1-2.6 after mid 21st century. This leads to the two scenarios diverging significantly by 2100.

### 4.4. Africa

CMIP6 simulated, and RECCAP2 assessed values of African carbon fluxes and stocks are summarized in Table 4. The multi-model mean matches the assessed value very well for GPP while being low but within the error bars for NBP. It is biased high but again within the error bars for C$_{tot}$, largely due to CMIP6 simulating greater soil carbon amounts. Six, 12 and 5 models pass the filtering respectively.

For both GPP and C$_{tot}$, but especially for carbon stocks, RECCAP2 assesses lower values than simulated by CMIP6 ESMs (Figure 11). Filtering for GPP leads to narrowed projected ranges but centrally based within the unfiltered range. This leads to SSP3-7.0 diverging more markedly above SSP1-2.6 future projection of GPP.

For C$_{tot}$ only about the lower half of CMIP6 models fall within the present-day range of RECCAP2 and so filtered results much lower and narrower than unfiltered. A notable feature unique to Africa is that future NBP is lower for SSP3-7.0 thank SSP1-2.6 despite substantially higher GPP. This is explained not by a climate-driven effect on turnover, but by regional land-use changes. Figure S1 in Supporting Information S1 shows Africa has the highest increase in land-use (defined here as crop and pasture fraction) of any region under SSP3-7.0. Thus, when looking at carbon stocks, SSP3-7.0 actually sees reduced carbon storage on land by 2100.
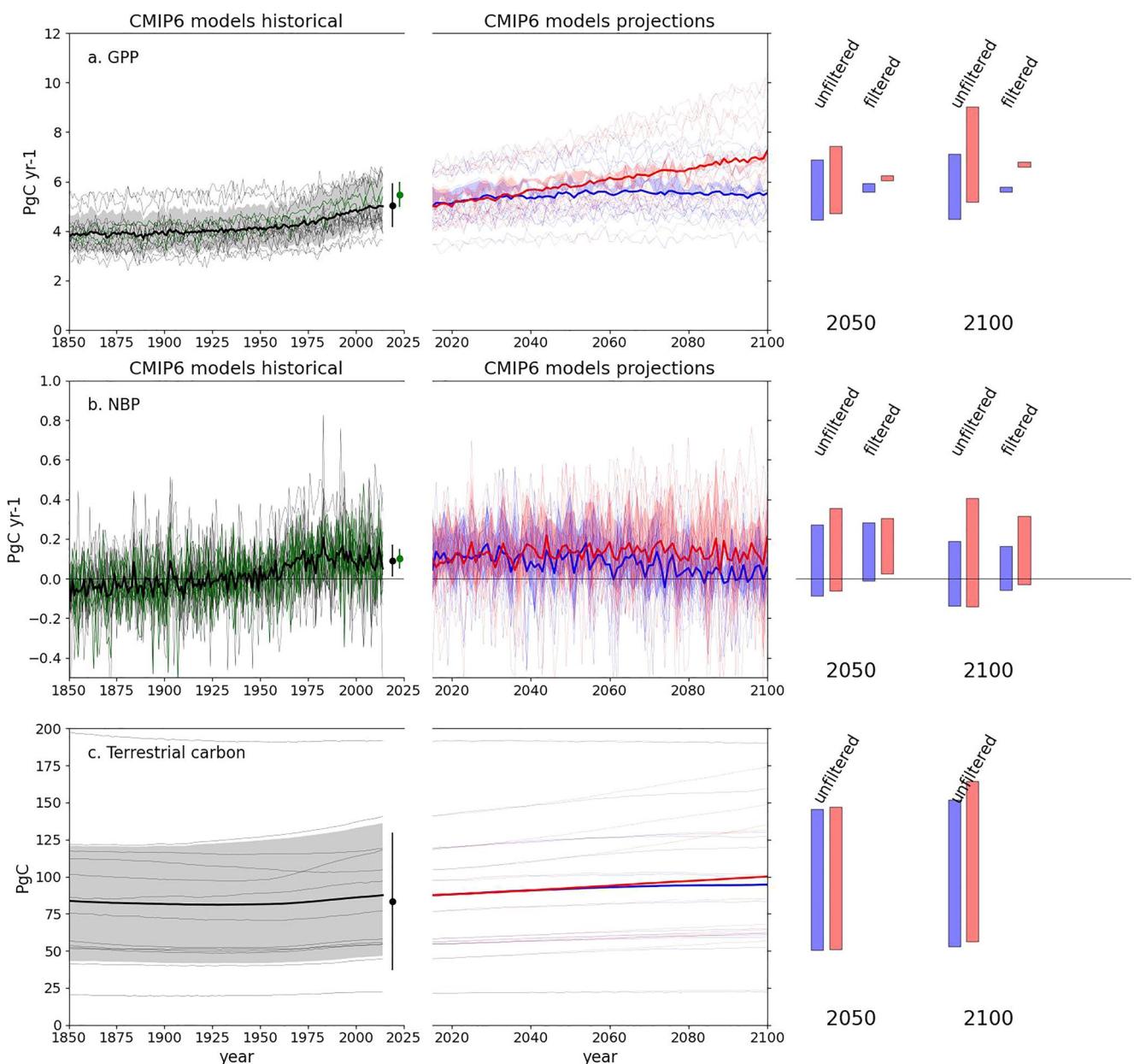
**Figure 10.** (top) CMIP6 simulated GPP, (middle) NBP, and (bottom) total terrestrial carbon stock for the historical period (1850–2014, left hand side) and future projections (to 2100, right hand side) for Europe. The panels show individual CMIP6 models in faint lines. For the historical period, individual models are shown in gray and multi model mean in thick black. CMIP6 mean and standard deviation are shown by the black dot and error bar, and REgional Carbon Cycle Assessment and Processes, Phase 2 (RECCAP2) assessed best estimate and uncertainty by the green dot and error bar. Models which fall within the RECCAP2 assessed range are shown in pale green. Future projections are for SSP1-2.6 (blue) and SSP3-7.0 (red), with single models in faint lines and multi-model mean in thick lines. Filtering just those models which match the RECCAP2 assessed range gives the red and blue shaded region. Mean values for 2050 and 2100 before and after filtering are shown to the right of the plots.

### 4.5. Russia

CMIP6 simulated, and RECCAP2 assessed values of Russian carbon fluxes and stocks are summarized in Table 5. There is strong agreement between the simulated results of CMIP6 models and those of RECCAP2 regarding overall carbon fluxes and stocks (Table 5). CMIP6 models show slightly lower estimates of biomass and slightly higher estimates of soil carbon but GPP, NPP, NBP are extremely close, with 12 and 7 models passing the filtering for GPP and NBP. The carbon stores also appear close but are outside the RECCAP2 assessed range for all models (i.e., none pass the filtering) due to the low uncertainty stated in RECCAP2.

**Table 4**
*Comparison of CMIP6 Simulated and REgional Carbon Cycle Assessment and Processes, Phase 2 Assessed Carbon Stocks and Fluxes for Africa*

|  |  | CMIP6 (mean ± 1 sigma) | RECCAP2 (best estimate ± error estimate) |
|---|---|---|---|
| Fluxes PgC/yr mean over 2010–2019 inclusive | GPP | 26.74 ± 6 | 27.95 ± 3.62 (TRENDY) |
|  |  |  | 24.67 ± 2.46 (satellite) |
|  |  |  | Best estimate: 26.3 ± 3.1 |
|  | NPP | 12.13 ± 4 | 12.8 ± 2.8 |
|  | NEP | 1.25 ± 5 | 0.65 ± 4.8 |
|  | NBP | 0.24 ± 0.4 | 0.432 ± 0.484 |
| Stocks PgC mean over 2010–2019 inclusive | Vegetation carbon | 85.05 ± 25 | 84 ± 12 |
|  | Soil carbon | 127.02 ± 55 | 87.7 ± 11 |
|  | Product carbon | 0.90 ± 0.9 | Not available from RECCAP2 assessment |
|  | Total terrest. carbon | 220.2 ± 75 | 171.7 ± 16 |

*Note.* Values are shown as the average over a 10 year period from 2010 to 2019, and ranges are one standard deviation unless stated otherwise. For model outputs this reflects spread across different models, while for assessed values this can include other means of assessing uncertainty. Units are PgC or PgC yr$^{-1}$.

Future projections of fluxes for this region (Figure 12) show a strong dependence on scenario, with GPP under SSP3-7.0 increasing strongly due to both elevated $CO_2$ and warmer temperature sin this high-latitude region. It is the region where GPP differs most under the two scenarios by 2100 although this is not reflected in NBP increases where SSP3-7.0 is only slightly higher than SSP1-2.6 due to increased turnover as well as increased productivity. It is also, however, a region where carbon stocks vary extremely widely across CMIP6 ESMs indicating an urgent need for models here to be confronted with data. A factor of almost 10 between the highest and lowest values of total carbon stock show the huge discrepancy between models. Part of the reason for this is that this region covers a large amount of frozen ground. Permafrost carbon is not represented well in CMIP6 models and this is discussed more in Section 4.11. The RECCAP2 assessment here is for the top 1 m of mineral and organic soils of Russia (Schepaschenko et al., 2013).

### 4.6. Central Asia

CMIP6 simulated, and RECCAP2 assessed values of Central Asian carbon fluxes and stocks are summarized in Table 6. The carbon flux and carbon pool values for Central Asia RECCAP2 assessment were based on TRENDY model all-combined simulations (version 10) (Friedlingstein, Jones, et al., 2022). The ensemble mean of 17 terrestrial biosphere models (TRENDY v10 models) was considered as the best estimate for Central Asia region, and the standard deviation between model outputs was used as uncertainties.

There is a general agreement between the simulated results of CMIP6 models and those of RECCAP2 regarding overall carbon fluxes and stocks (Table 6). CMIP6 models show slightly lower estimates with broader ranges than RECCAP2 for GPP, NEP and NBP. In contrast, they suggest slightly lower estimates with narrower ranges for NPP, soil carbon stock, and biological products carbon pool. The CMIP6 multi-model mean is above the RECCAP2 assessment for vegetation carbon stock but within the uncertainty range. For NBP CMIP6 model mean and RECCAP2 assessment suggest a weak sink. CMIP6 mean agrees well with RECCAP2 for GPP, NBP and C$_{tot}$. Thirteen, 1, and 9 models pass the filtering respectively.

The results show (Figure 13) an ongoing increase in GPP under SSP3-7.0 while GPP stops its increase under SSP1-2.6 after mid 21st century. While the multi-model mean NBP values are in close agreement, the narrow range of uncertainty in the RECCAP2 assessment means only one CMIP6 ESM passes the filtering, although the filtered projections remain central within the unfiltered CMIP6 range. The two scenarios diverging significantly by 2100. As seen for other regions, this does not translate into similar increases in carbon stocks with both scenarios showing very similar changes (slight increases) on both 2050 and 2100 time horizons. It may further indicate that the components of terrestrial ecosystem C cycle in Central Asia are vulnerable to and strongly influenced by climate changes.
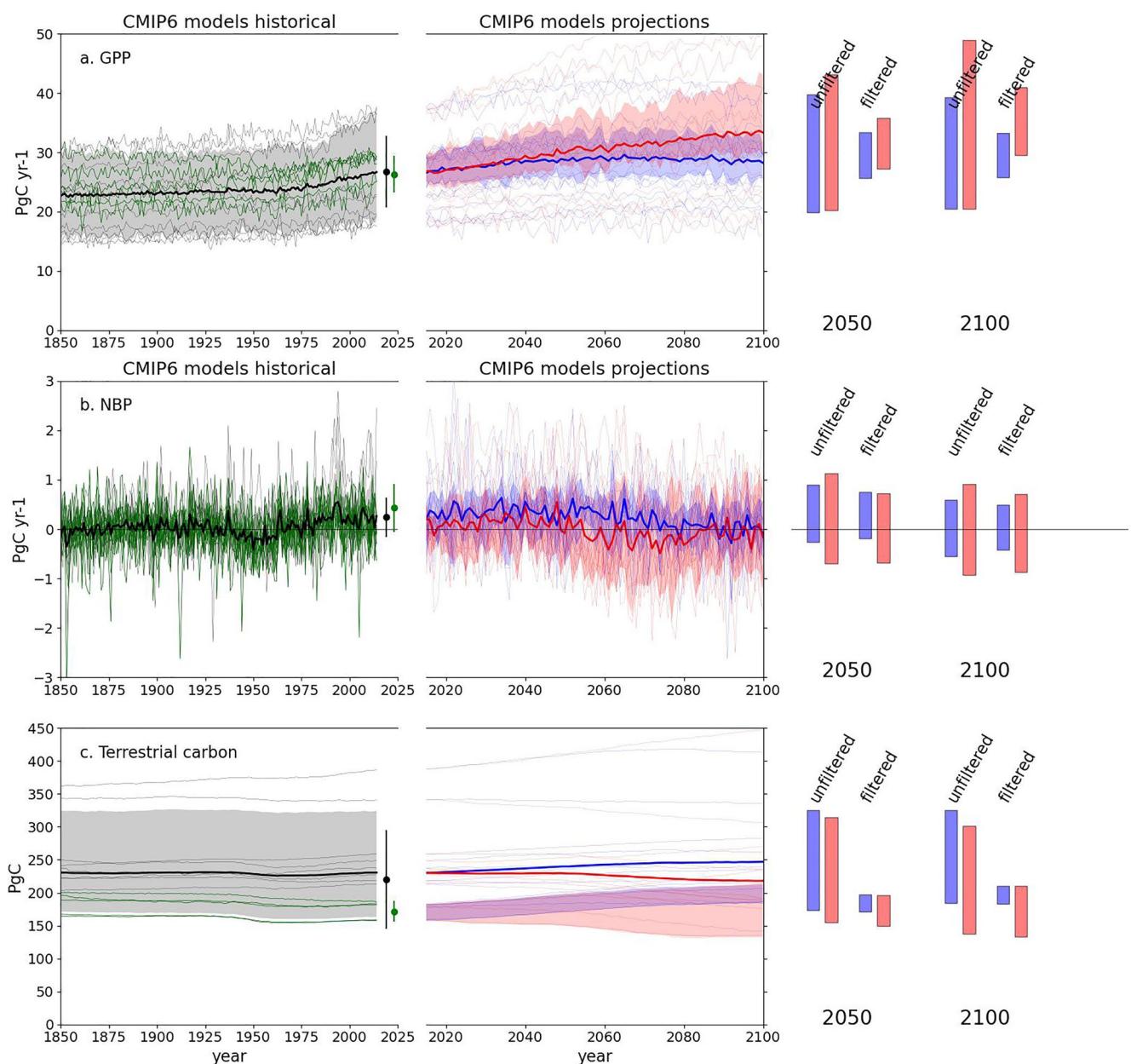
**Figure 11.** CMIP6 simulated GPP (top), NBP (middle) and total terrestrial carbon stock (bottom) for the historical period (1850–2014, left hand side) and future projections (to 2100, right hand side) for Africa. The panels show individual CMIP6 models in faint lines. For the historical period, individual models are shown in gray and multi model mean in thick black. CMIP6 mean and standard deviation are shown by the black dot and error bar, and REgional Carbon Cycle Assessment and Processes, Phase 2 (RECCAP2) assessed best estimate and uncertainty by the green dot and error bar. Models which fall within the RECCAP2 assessed range are shown in pale green. Future projections are for SSP1-2.6 (blue) and SSP3-7.0 (red), with single models in faint lines and multi-model mean in thick lines. Filtering just those models which match the RECCAP2 assessed range gives the red and blue shaded region. Mean values for 2050 and 2100 before and after filtering are shown to the right of the plots.

## 4.7. East Asia

For East Asia, CMIP6 model predictions generally agree with the RECCAP2 evaluation results for 2010–2019 (Table 7). The multi-model mean is close to the assessed value for GPP, NBP and $C_{tot}$, with 8, 1 and 8 models passing the filtering respectively. The ensemble mean of CMIP6 models lies within the uncertain bounds of RECCAP2 best estimates for all terrestrial carbon fluxes and stocks except NBP (i.e., net carbon flux). The RECCAP2 reveals the region is a net carbon sink by combining a bottom-up approach utilizing inventory, simple ESM (OSCAR, Gasser et al., 2020), statistical models based on satellite and field measurements (0.370 ± 0.042

**Table 5**
*Comparison of CMIP6 Simulated and REgional Carbon Cycle Assessment and Processes, Phase 2 Assessed Carbon Stocks and Fluxes for Russia*

| | | CMIP6 (mean ± 1 sigma) | RECCAP2 (best estimate ± error estimate) |
|---|---|---|---|
| Fluxes PgC/yr mean over 2010–2019 inclusive | GPP | 9.65 ± 1.08 | 9.41 ± 1.12 |
| | NPP | 5.09 ± 1.06 | 5.33 ± 0.37 |
| | NEP | 0.58 ± 1.21 | 1.19 ± 0.14 |
| | NBP | 0.34 ± 0.16 | 0.33 ± 0.09 |
| Stocks PgC mean over 2010–2019 inclusive | Vegetation carbon | 48.85 ± 17.18 | 59.50 ± 4.85 |
| | Soil carbon | 368.32 ± 283.34 | 317.10 ± 15.86 |
| | Product carbon | 0.57 ± 0.52 | 0.65 ± 0.33 |
| | Total terrest. carbon | 438.90 ± 283.05 | 377.25 ± 15.90 |

*Note.* Values are shown as the average over a 10 year period from 2010 to 2019, and ranges are one standard deviation unless stated otherwise. For model outputs this reflects spread across different models, while for assessed values this can include other means of assessing uncertainty. Units are PgC or PgC yr$^{-1}$.

PgC/yr) and top-down inverse modeling approach (0.341 ± 0.125 PgC/yr). The two different approaches show great agreement even though the uncertainty bounds derived from inverse modeling are wider than that from the bottom-up approach due to the lack of atmospheric $CO_2$ measurements to constrain the carbon flux over the region (Y. Wang et al., 2022). CMIP6 models slightly underestimate the net carbon flux with a value of 0.20 ± 0.23 PgC/yr, consistent with the previous finding showing that process-based models tend to underestimate the carbon flux in the region compared to inverse modeling (Kondo et al., 2020). The underestimation of CMIP6 models may come from the not well-implemented processes (e.g., land use change and forest demography) in most CMIP6 models to accurately simulate the several decades of afforestation effects in the region (Piao et al., 2012; J. Wang et al., 2020).

Both GPP and $C_{tot}$ agree well with the CMIP6 mean estimate, but the RECCAP2 assessment has narrower uncertainty bounds leading to reduced spread in filtered projections (Figure 14). The results show an ongoing increase in GPP under SSP3-7.0 while GPP stops increasing under SSP1-2.6 after mid 21st century. For SSP3-7.0 this increase is quite marked and leads to the two scenarios diverging significantly by 2100. This leads to diverging NBP trends. After an increase until the mid 21st century, NBP remains constant for SSP3-7.0 but rapidly decrease for SSP1-2.6. This trend is also reflected somewhat in carbon stocks with SSP3-7.0 showing a larger increase by 2100, which is especially notable in the filtered results as being above SSP1-2.6.

### 4.8. South Asia

CMIP6 simulated, and RECCAP2 assessed values of South Asian carbon fluxes and stocks are summarized in Table 8. The multi-model mean is below the assessed value for GPP but the error bars overlap and close for NBP. It is also biased low but for $C_{tot}$ but again the error bars overlap, with 4, 11 and 5 models passing the filtering respectively.

In this region, the RECCAP2 assessment of the magnitude of NPP and NBP and stock values (vegetation and total terrestrial carbons) are similar to those simulated in CMIP6, but for some other variable differences are quite large. For example, in the case of flux values, CMIP6 GPP is 27% too low, whereas, CMIP total carbon ($C_{tot}$) is 30% underestimated compared to RECCAP2 assessment (Table 8 and Figure 15). These values result in filtered ranges being narrower and higher than the raw unfiltered results for 2050 and 2100 cases. This outcome is especially true for $C_{tot}$ where the filtering removes a cluster of ESMs from the bottom of the range. The results show an ongoing increase in GPP under SSP3-7.0 while GPP stops increasing under SSP1-2.6 after mid 21st century. This leads to the two scenarios diverging significantly by 2100. However, the divergence does not translate into similar increases in carbon stocks, with both scenarios showing similar changes (slight increases) for both 2050 and 2100 time horizons.

### 4.9. South East Asia

CMIP6 simulated, and RECCAP2 assessed values of South East Asian carbon fluxes and stocks are summarized in Table 9. The multi-model mean is below the assessed value for GPP and for NBP CMIP6 mean simulates a
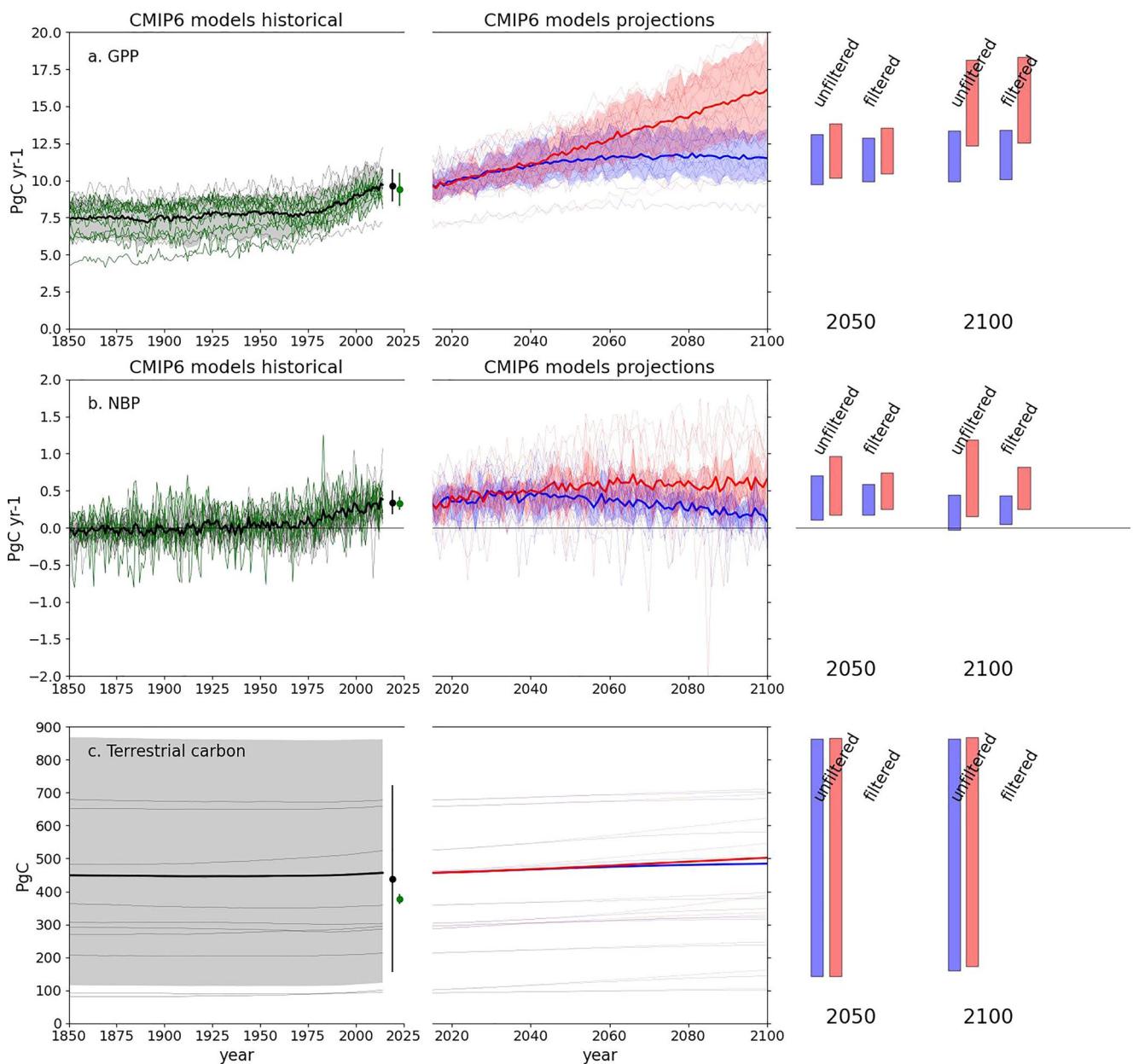
**Figure 12.** (top) CMIP6 simulated GPP, (middle) NBP, and (bottom) total terrestrial carbon stock for the historical period (1850–2014, left hand side) and future projections (to 2100, right hand side) for Russia. The panels show individual CMIP6 models in faint lines. For the historical period, individual models are shown in gray and multi model mean in thick black. CMIP6 mean and standard deviation are shown by the black dot and error bar. Future projections are for SSP1-2.6 (blue) and SSP3-7.0 (red), with single models in faint lines and multi-model mean in thick lines. Mean values for 2050 and 2100 are shown to the right of the plots.

weak sink while RECCAP2 assesses a source. The error bars do not overlap in this region for NBP, which is the only case we found for all regions and variables assessed. However, the CMIP6 mean agrees well for $C_{tot}$. Overall 7, none and 5 models pass the filtering respectively.

GPP, NPP, NEP, and carbon stocks of Southeast Asia were estimated by an ensemble of DGVM simulations from TRENDY v10 (Friedlingstein, Jones, et al., 2022), and the carbon budget of the region was estimated by averaging estimates from three independent approaches: bottom-up, top-down, and carbon stock change ($\Delta$C) approaches (Ciais et al., 2022; Kondo et al., 2020, 2022). The bottom-up approach integrates all major sources and sinks of $CO_2$ in the region, such as GPP, RE, $CO_2$ emissions from land-use and land-cover changes, emissions from peat drainage and peat fire, $CO_2$ evasion from rivers and lakes, coastal carbon fluxes, and lateral transport of carbon

**Table 6**
*Comparison of CMIP6 Simulated and REgional Carbon Cycle Assessment and Processes, Phase 2 Assessed Carbon Stocks and Fluxes for Central Asia*

| | | CMIP6 (mean $\pm$ 1 sigma) | RECCAP2 (mean $\pm$ 1 sigma) |
|---|---|---|---|
| Fluxes PgC/yr mean over 2010–2019 inclusive | GPP | $2.24 \pm 1.16$ | $2.33 \pm 1.08$ |
| | NPP | $1.10 \pm 0.49$ | $1.22 \pm 0.59$ |
| | NEP | $0.13 \pm 0.60$ | $0.19 \pm 0.19$ |
| | NBP | $0.05 \pm 0.05$ | $0.06 \pm 0.02$ |
| Stocks PgC mean over 2010–2019 inclusive | Vegetation carbon | $3.78 \pm 2.27$ | $2.83 \pm 1.09$ |
| | Soil carbon | $29.33 \pm 18.36$ | $33.14 \pm 20.96$ |
| | Product carbon | $0.06 \pm 0.09$ | $0.13 \pm 0.33$ |
| | Total terrest. carbon | $35.32 \pm 20.20$ | $36.07 \pm 21.62$ |

*Note.* Values are shown as the average over a 10 year period from 2010 to 2019, and ranges are one standard deviation unless stated otherwise. For model outputs this reflects spread across different models, while for assessed values this can include other means of assessing uncertainty. Units are PgC or PgC yr$^{-1}$.

through river carbon export. Top-down and $\Delta$C estimates provide a partial net carbon balance, the former only represents the vertical component of the net balance, and the latter only represents the aboveground component of the net balance. Missing components in top-down and $\Delta$C approaches were supplemented by independent data so that all three approaches fully account for necessary components to represent the net carbon balance for South East Asia.

CMIP6- and RECCAP2-estimated mean GPP, RE, and NEP values were similar in magnitude, and consistent in having positive NEP values suggesting natural vegetation uptake of $CO_2$ in the region. However, a discrepancy between the methods appeared in NBP, where CMIP6 indicates carbon neutrality, but all three approaches in RECCAP2 estimated a net $CO_2$ source. This discrepancy is likely attributed to differences in the land-use change representation between the methods, and maybe lack of inclusion of fire in many ESMs. RECCAP2 explicitly considers the key components of land-use change fluxes in Southeast Asia, not only $CO_2$ fluxes from deforestation and forest regrowth, but also emissions from peat drainage and peat fires. However, CMIP6 models are not comprehensive enough to represent such specific land-use change fluxes, thus resulting in underestimated $CO_2$ emissions in Southeast Asia.

For South East Asia, the filtering had similar results to South Asia for GPP, with RECCAP2 values being consistent but slightly above CMIP6 simulated values (Figure 16). However, for $C_{tot}$ RECCAP2 had marked tighter uncertainty bounds and therefore the filtering led to strongly reduced spread as models at both the top and bottom of the CMIP6 range were filtered out. The results show an ongoing increase in GPP under SSP3-7.0 while GPP stops its increase under SSP1-2.6 after mid 21st century. This leads to the two scenarios diverging significantly by 2100. This difference can also be seen in carbon stocks, with SSP3-7.0 leading to slightly greater increases in carbon storage by 2100.

### 4.10. Australasia

Major carbon flux components (GPP, NPP, NEP) for Australasia RECCAP2 assessment were calculated by combining three different model simulations: CABLE-POP model (Australia), Biome-BGC MuSO (NZ), and CenW (NZ) simulations. RECCAP2 Carbon fluxes for New Zealand were simulated with the Biome-BGC MuSO model (Hidy et al., 2016, 2022) and CenW model (Kirschbaum & Watt, 2011). Biome-BGC MuSO was used to represent grasslands and native forest (evergreen broadleaf forest), whereas CenW was used to model non-native forests (Pinus radiata/plantation forest) and shrub (~60% Manuka/Kanuka) landcover classifications. Two different models were used for New Zealand because Biome-BGC MuSO was calibrated specifically for New Zealand grasslands (including grazed pasture), whereas CenW was calibrated for plantation forest and Manuka/Kanuka. No calibration currently exists for NZ native forest, so the default evergreen broadleaf forest biome in Biome-BGC MuSO was used. Although New Zealand contributes a small component of the total Australasia carbon budget, it is worth noting that the Biome-BGC MuSO and CenW models are not currently used in any of the CMIP models, and so they provide an independent assessment. It also explicitly includes the effect of grazing on carbon fluxes, which affects about half of all land in New Zealand. We expect that CMIP6 models that do not include grazing show larger disagreement.

Australian carbon flux components were derived from CABLE-POP in the set-up of the BIOS3 environment (referred to as CABLE-BIOS3) (Haverd et al., 2018) simulations. CABLE model same as BGC MuSO and CenW models, are not used in CMIP and provide an independent assessment; both represent the best knowledge of Australia
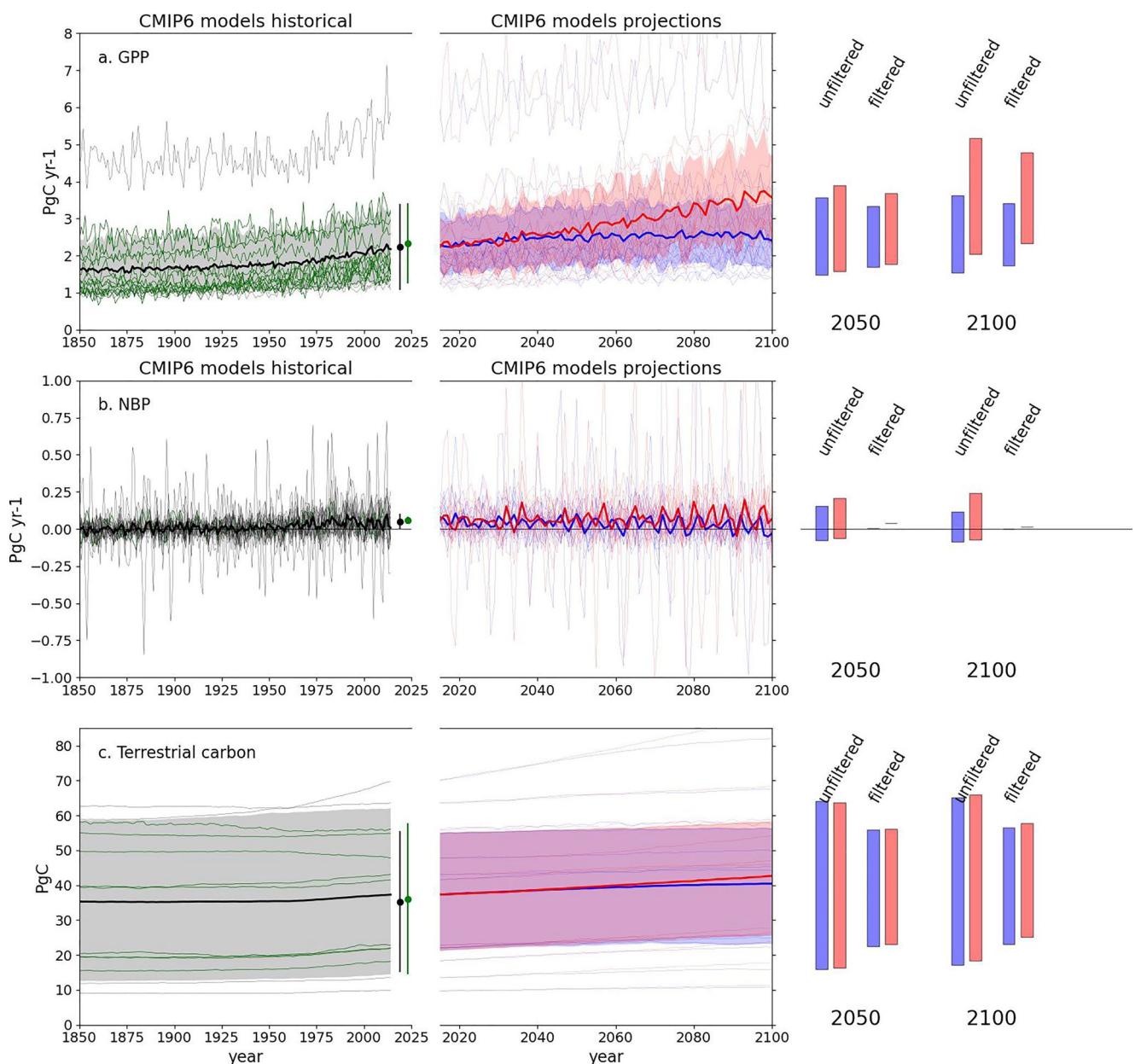
**Figure 13.** (top) CMIP6 simulated GPP, (middle) NBP, and (bottom) total terrestrial carbon stock for the historical period (1850–2014, left hand side) and future projections (to 2100, right hand side) for Central Asia. The panels show individual CMIP6 models in faint lines. For the historical period, individual models are shown in gray and multi model mean in thick black. CMIP6 mean and standard deviation are shown by the black dot and error bar, and REgional Carbon Cycle Assessment and Processes, Phase 2 (RECCAP2) assessed best estimate and uncertainty by the green dot and error bar. Models which fall within the RECCAP2 assessed range are shown in pale green. Future projections are for SSP1-2.6 (blue) and SSP3-7.0 (red), with single models in faint lines and multi-model mean in thick lines. Filtering just those models which match the RECCAP2 assessed range gives the red and blue shaded region. Mean values for 2050 and 2100 before and after filtering are shown to the right of the plots.

and New Zealand's carbon cycle because they were forced with regional driver and observations. NPP from CABLE simulation were also corrected for fire disturbances, and fire fluxes for New Zealand were assumed to negligible.

The spread (the standard deviation) between Australian and New Zealand model output (GPP, NPP NEP), and ensemble mean of 16 global dynamic vegetation models (DVGMs) (TRENDY models (version 10) (Friedlingstein, Jones, et al., 2022) was used as uncertainties. We considered ensemble mean of TRENDY models as a baseline for Australasia, because it offers good representation of fluxes across the region. Australasia RECCAP2 net carbon exchange (NBP) was built considering the terrestrial carbon fluxes described above and other carbon flux components: land-to-ocean aquatic continuum (LOAC) system (water $CO_2$ emissions from lakes and reservoirs),

**Table 7**
*Comparison of CMIP6 Simulated and REgional Carbon Cycle Assessment and Processes, Phase 2 Assessed Carbon Stocks and Fluxes for East Asia*

| | | CMIP6 (mean ± 1 sigma) | RECCAP2 (best estimate ± error estimate) |
|---|---|---|---|
| Fluxes PgC/yr mean over 2010–2019 inclusive | GPP | 9.44 ± 1.74 | 9.32 ± 1.39 |
| | NPP | 4.68 ± 0.91 | 4.72 ± 0.79 |
| | NEP | 0.54 ± 1.26 | 0.68 ± 0.50 |
| | NBP | 0.20 ± 0.17 | Bottom-up: 0.370 ± 0.042 |
| | | | Top-down: 0.341 ± 0.125 |
| Stocks PgC mean over 2010–2019 inclusive | Vegetation carbon | 32.12 ± 7.63 | 33.34 ± 14.26 |
| | Soil carbon | 113.06 ± 61.27 | 115.81 ± 36.24 |
| | Product carbon | 0.89 ± 0.66 | 0.97 ± 0.63 |
| | Total terrest. carbon | 154.19 ± 63.01 | 150.12 ± 38.95 |

*Note.* Values are shown as the average over a 10 year period from 2010 to 2019, and ranges are one standard deviation unless stated otherwise. For model outputs this reflects spread across different models, while for assessed values this can include other means of assessing uncertainty. Units are PgC or PgC yr$^{-1}$.

estuaries and coastal ecosystems, as well as lateral fluxes associated with harvested crop, wood products and livestock and fossil fuel emissions.

CMIP6 results agree with the Australasia RECCAP2 assessment within the uncertainty estimate (Table 10). The multi-model mean is close to the assessed value for GPP and NBP and biased low but within the error bars for $C_{tot}$, with 14, 9, and 12 models passing the filtering respectively. Results filtered for present-day model skill show only a very small reduction in the range of predicted land carbon stocks (Figure 17). A more thorough analysis of predicted carbon fluxes and stocks for the Australasian region will be covered elsewhere, including additional criteria to filter CMIP6 model results.

Australasia NBP (based on RECCAP2 of the region) is very close to carbon neutral (0.0004 ± 0.08 PgC yr$^{-1}$) compared to CMIP6 (0.04 ± 0.07 PgC yr$^{-1}$). The sink in CMIP6 does not include carbon sources from estuaries or sinks from coastal ecosystems (e.g., mangroves or seagrasses) or correction for lateral fluxes (e.g., river carbon export) in the simulations.

Given that the carbon in soil is one of the most uncertain estimates in the ESMs, we also compared the soil carbon estimates from RECCAP2 derived from TRENDY against the Harmonized World Soil Database (HWSD, version 1.2). We found that the RECCAP2 estimate of 44.74 ± 24.01 is in closer agreement with the HWSD of 45.66 PgC yr$^{-1}$, when compared with the CMIP6 estimate of 30.74 ± 15.22 PgC yr$^{-1}$, which suggest a possible reason for the disagreement between the two approaches.

### 4.11. Permafrost

RECCAP2 also assesses the carbon cycle of the permafrost region, which cuts across the North America and Russia RECCAP2 regions described above. That is, the permafrost region overlaps with others, and therefore, is not additional to the others, and as such, values presented here are not additive to the totals given for the other regions. The permafrost region is of particular interest given that high latitudes are warming substantially faster than the global land average (Chylek et al., 2022) and the large amount of carbon stored there (Schuur et al., 2022). This region provides a potentially important but under-studied feedback on future climate projections, but it is at best poorly represented (or more often not at all) in CMIP models (Burke et al., 2013). Important feedbacks also exist with methane (Gedney et al., 2019) and frozen nitrogen (Burke et al., 2022).

Estimates of present-day carbon balance for the permafrost region differ in sign with site based extrapolation indicating the possibility of a net source, while top-down inversion estimates suggesting the region to be a net sink of $CO_2$. Process-based models also simulate a net sink for terrestrial ecosystems, but disturbances such as fire and abrupt thaw, which are likely to be a source of carbon, along with aquatic ecosystems are not typically included. A model-data fusion based on the CARDAMOM framework (Bloom et al., 2016) has a large uncertainty in the carbon budget with both the NEP and NBP spanning the source-sink boundary.

Here we present the RECCAP2-assessed and CMIP6 simulated last decade values (Table 11) and CMIP6 future projections (Figure 18). The multi-model mean is close to the assessed value for NBP, with 16 models passing the
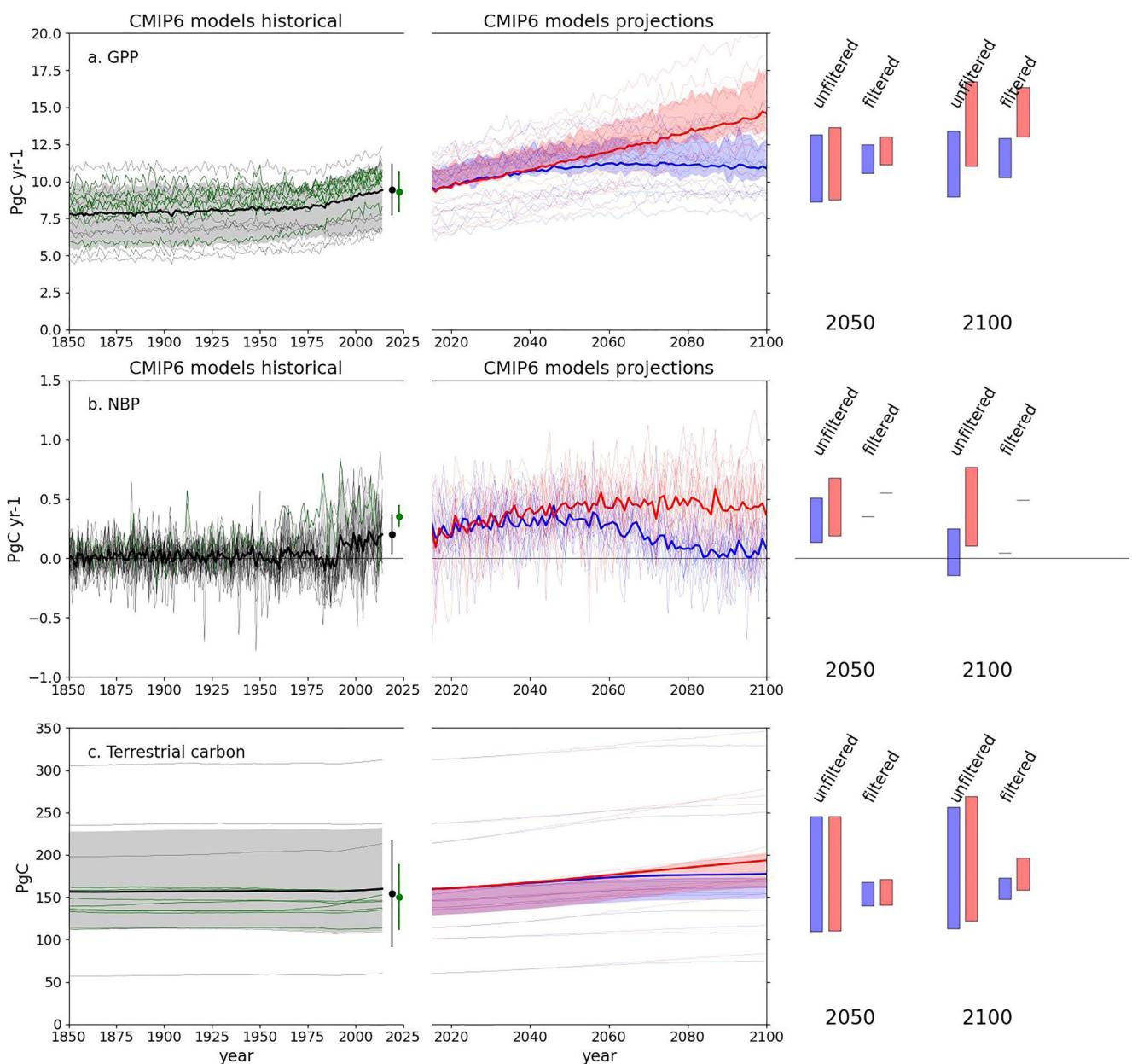
**Figure 14.** (top) CMIP6 simulated GPP, (middle) NBP, and (bottom) total terrestrial carbon stock for the historical period (1850–2014, left hand side) and future projections (to 2100, right hand side) for East Asia. The panels show individual CMIP6 models in faint lines. For the historical period, individual models are shown in gray and multi model mean in thick black. CMIP6 mean and standard deviation are shown by the black dot and error bar, and REgional Carbon Cycle Assessment and Processes, Phase 2 (RECCAP2) assessed best estimate and uncertainty by the green dot and error bar. Models which fall within the RECCAP2 assessed range are shown in pale green. Future projections are for SSP1-2.6 (blue) and SSP3-7.0 (red), with single models in faint lines and multi-model mean in thick lines. Filtering just those models which match the RECCAP2 assessed range gives the red and blue shaded region. Mean values for 2050 and 2100 before and after filtering are shown to the right of the plots.

filtering despite substantially over-estimating GPP where only two models pass the filtering. Due to the incompleteness of permafrost-related processes in CMIP6 ESMs we do not attempt a constraint based on comparison with the regional RECCAP2 contemporary assessment. We conclude that there is much work to be done in this area and that inclusion of properly evaluated permafrost biogeochemistry and hydrological processes in CMIP-class climate models is a high priority.

CMIP6 projections show that future changes in gross fluxes (e.g., GPP: Figure 18) are strongly dependent on the choice of scenario, in this case with SSP3-7.0 leading to large increases in GPP—possibly even doubling

**Table 8**
*Comparison of CMIP6 Simulated and REgional Carbon Cycle Assessment and Processes, Phase 2 Assessed Carbon Stocks and Fluxes for South Asia*

|  |  | CMIP6 (mean ± 1 sigma) | RECCAP2 (mean ± 1 sigma) |
|---|---|---|---|
| Fluxes PgC/yr mean over 2010–2019 inclusive | GPP | 3.71 ± 1.30 | 5.11 ± 1.13 |
|  | NPP | 1.74 ± 0.73 | 2.46 ± 0.63 |
|  | NEP | 0.25 ± 1.00 | 0.12 ± 0.06 |
|  | NBP | 0.05 ± 0.06 | 0.04 ± 0.06 |
| Stocks PgC mean over 2010–2019 inclusive | Vegetation carbon | 7.91 ± 5.80 | 9.29 ± 3.26 |
|  | Soil carbon | 19.53 ± 10.65 | 30.44 ± 12.3 |
|  | Product carbon | 0.36 ± 0.35 | 0.41 ± 0.26 |
|  | Total terrest. carbon | 28.02 ± 12.93 | 39.84 ± 11.21 |

*Note.* Values are shown as the average over a 10 year period from 2010 to 2019, and ranges are one standard deviation unless stated otherwise. For model outputs this reflects spread across different models, while for assessed values this can include other means of assessing uncertainty. Units are PgC or PgC yr$^{-1}$.

compared to present day, although it cannot be found from these results if this is due to elevated $CO_2$ or regional warming—both of which can enhanced vegetation productivity in high-latitude regions. However, compensating changes in residence time mean that increased productivity may not result in increased carbon storage, and so most CMIP6 models show only weak scenario-dependence in terms of future carbon stocks, with the multi-model mean increasing slightly more under SSP3-70 than under SSP1-2.6. Previous studies from CMIP5 also found a persistent sink under low and medium warming scenarios, but a suggestion of a transition to a source under the very highest scenario (RCP8.5) (Qiu et al., 2020). The inclusion of permafrost carbon, omitted from the majority of CMIP6 models, will impact this assessment of the carbon budget under future climate scenarios (Schuur et al., 2022).

## 5. Synthesis and Discussion

For each region, we explore to what extent the CMIP6 simulated values are consistent with RECCAP2 assessments for the past decade and the possible implications for future projections. Techniques do not yet exist to robustly constrain future projected values. For example, it would be possible (even tempting) to simply filter out those models which do not match RECCAP2 assessed values for each variable, and that would lead to a reduced spread of future projections. In such a case, models may perform better in some regions due to better calibration or process inclusion (such as permafrost or nutrient inclusion), and therefore there is some merit in treating the world as a mosaic and reassembling the projections from the best suited models for each region. However, we do not feel that this technique is yet mature enough for quantitative application, but rather is better suited to drawing insights model-by-model and region-by-region into the value of different processes. For illustration we have shown the results from such a filtering for each region, and this is illustrative when making expert judgments on the ability of global models contributing to CMIP6 to make regional-scale projections. But we find it is not a robust quantitative constraint to be able to simply use this filtering to reduce spread in future projections. Reasons why this is not yet robust include:

- Weighting by model skill is a debated art (Abramowitz et al., 2019; Merrifield et al., 2020) but it is generally accepted that simple binary filtering of models in/out is not optimal. More details are provided in Section 2.
- Filtering variable-by-variable and region-by-region would leave different models selected each time and therefore potential inconsistent projections if they got the right answer for the wrong reasons. For example, for South Asia, results show consistent projections of present day carbon stocks, but a significant bias in GPP, meaning there are likely compensating errors in other processes in the CMIP6 ESMs.
- Filtering on all variables (e.g., only use models with both the correct stocks and fluxes) would potentially leave very few (or none) models remaining, and therefore over-reliance on too few models whilst discarding potentially useful information from others.
- Inconsistent definitions of variables—especially for NBP it is likely that CMIP6 models lack many of the relevant components and processes which may be included in the RECCAP2 regional assessments—for example, such as fire, harvesting and grazing.
- Poorly quantified error estimates on observed values—RECCAP2 regions have taken different approaches to providing error or uncertainty estimates for different quantities, and hence a one-size-fits-all approach to model selection would not be appropriate.
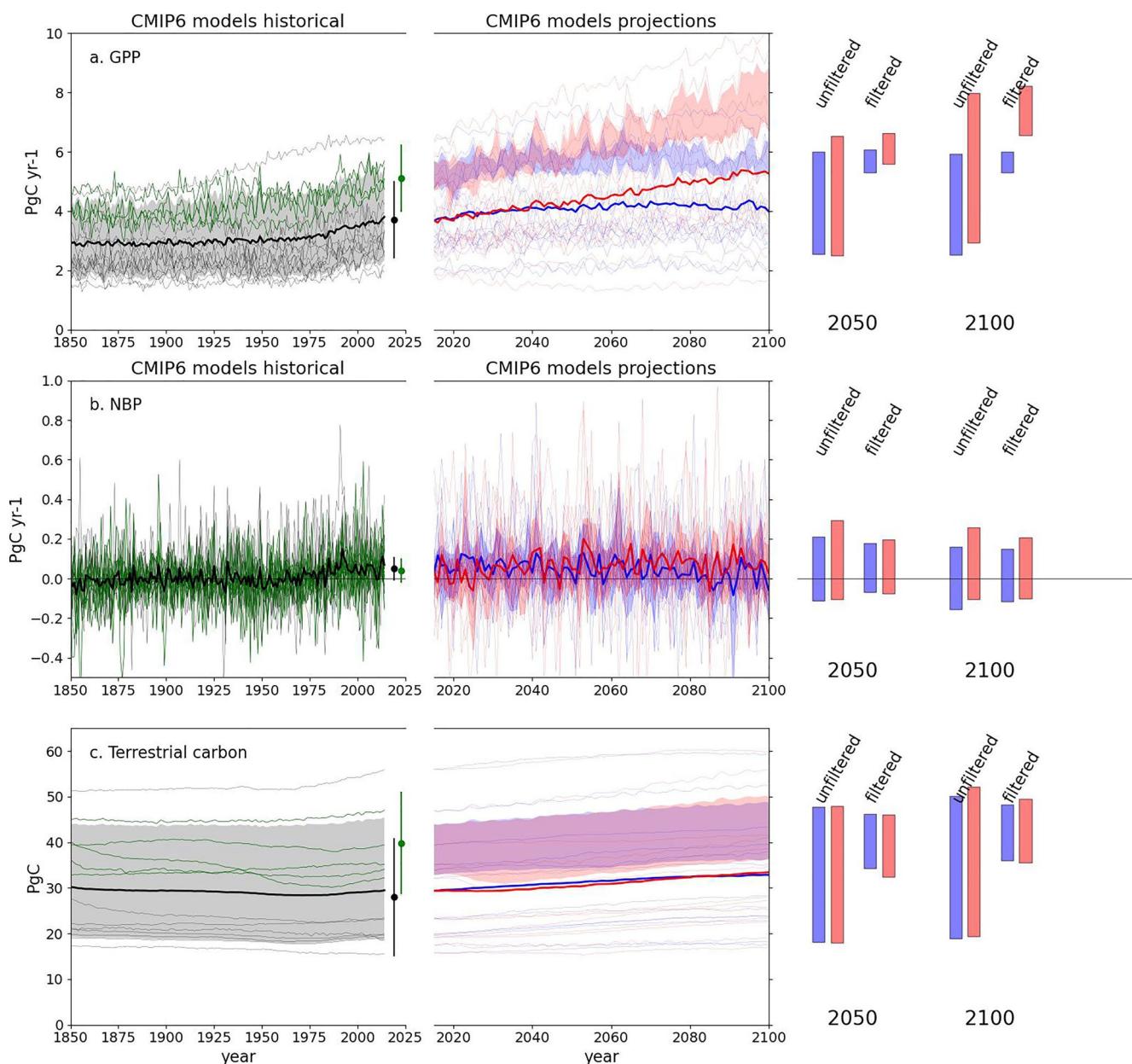
**Figure 15.** (top) CMIP6 simulated GPP, (middle) NBP, and (bottom) total terrestrial carbon stock for the historical period (1850–2014, left-hand side) and future projections (to 2100, right-hand side) for South Asia. The panels show individual CMIP6 models in faint lines. For the historical period, individual models are shown in gray and multi-model mean in thick black. CMIP6 mean and standard deviation are shown by the black dot and error bar, and REgional Carbon Cycle Assessment and Processes, Phase 2 (RECCAP2) assessed best estimate and uncertainty by the green dot and error bar. Models which fall within the RECCAP2 assessed range are shown in pale green. Future projections are for SSP1-2.6 (blue) and SSP3-7.0 (red), with single models in faint lines and multi-model mean in thick lines. Filtering just those models which match the RECCAP2 assessed range gives the red and blue shaded region. Mean values for 2050 and 2100 before and after filtering are shown to the right of the plots.

- Circularity of evidence—we know that many CMIP6 models draw on the same process representation, and often the same land-model components that also contribute to TRENDY and RECCAP2 regional assessments. This may lead to spurious constraints or over-confidence in any constrained results from CMIP6 results.

We are therefore left with a more subjective approach whereby we can make qualitative statements about regional scale projections. Nevertheless, although we stop short of making quantitative adjustments to CMIP6 projections we can dig into model performance at an individual level and then synthesize multi-model performance across variables.

**Table 9**
*Comparison of CMIP6 Simulated and REgional Carbon Cycle Assessment and Processes, Phase 2 Assessed Carbon Stocks and Fluxes for South East Asia*

| | | CMIP6 (mean ± 1 sigma) | RECCAP2 (best estimate ± error estimate) |
|---|---|---|---|
| Fluxes PgC/yr mean over 2010–2019 inclusive | GPP | 10.31 ± 2.35 | 11.67 ± 2.43 |
| | NPP | 4.19 ± 1.05 | 4.79 ± 0.92 |
| | NEP | 0.42 ± 1.10 | 0.47 ± 0.32 |
| | NBP | 0.05 ± 0.16 | Bottom -up: −0.46 ± 0.14 |
| | | | Top-down: −0.28 ± 0.12 |
| | | | ΔC: −0.44 (no std estimate) |
| | | | Best estimate: −0.39 ± 0.10 |
| Stocks PgC mean over 2010–2019 inclusive | Vegetation carbon | 44.95 ± 12.57 | 48.11 ± 13.43 |
| | Soil carbon | 37.09 ± 17.18 | 39.95 ± 15.53 |
| | Product carbon | 0.69 ± 0.68 | 1.13 ± 1.09 |
| | Total terrest. carbon | 84.31 ± 22.85 | 89.20 ± 10.02 |

*Note.* Values are shown as the average over a 10 year period from 2010 to 2019, and ranges are one standard deviation unless stated otherwise. For model outputs this reflects spread across different models, while for assessed values this can include other means of assessing uncertainty. Units are PgC or PgC yr$^{-1}$.

### 5.1. Model and Process Assessment

For GPP, NBP, and $C_{tot}$ we assess which models, ESM-by-ESM, perform well or not in terms of whether they fall within the error bars of the RECCAP2 regional assessment. The number of regions for which each ESM passes this filtering can be counted and compared to the process complexity of each model to assess whether different processes add skill.

Table 12 shows the results for GPP, NBP, and $C_{tot}$ for each ESM and each region, along with a list of process detail. For this assessment we consider whether or not each model includes: terrestrial nitrogen cycle; permafrost dynamics; dynamic vegetation and fire because these processes are routinely documented and reported components of the models (e.g., see Table 2 of Arora et al. (2020)).

#### 5.1.1. Model-By-Model Analysis

We discuss here model-by-model results as if they were all independent and different models but note that in reality some are more closely related to each other than others. For example, there are high- and low-resolution variants of NorESM2 and some of these models share land-surface schemes. Brunner et al. (2020) present a family-tree of CMIP6 models to help judge levels of independence.

In our analysis, 7 out of 16 models passed the filtering in half or more regions for GPP although all but two do so for five of the 11 regions. Ten out of 16 pass for half or more regions for NBP, and 9 of 13 passed for $C_{tot}$ noting that not all models reported soil carbon in the CMIP data archive and so cannot be included in the $C_{tot}$ comparison. Five out of 13 models passed the filtering for half or more regions for all three quantities. Some models simulated one quantity better than another—for example, IPSL-CM6A-LR performed the best in simulating GPP (in terms of the number of regions where it passed the filtering) but worst in simulating $C_{tot}$. Conversely, EC-Earth3-Veg was one of the poorest for GPP but the best for $C_{tot}$.

This analysis therefore may be useful for individual model development groups to identify specific areas for improved simulation in their models. It is not possible though in this study to explore specific reasons for these differences—be they simply not fully calibrated components, climate biases or lack of appropriate complexity. But we recommend that model evaluation for development goes beyond simple global or latitudinal metrics, as per Anav et al. (2013), and draws on regional information such as provided by RECCAP2.

#### 5.1.2. Region-By-Region Analysis

For both GPP and NBP, 6 out of 11 regions saw half or more models pass the filtering, while for $C_{tot}$, 5 out of 9 regions saw half or more models pass.

In simulating GPP, for most regions about half (circa 8 from 16) the CMIP models fell within the filtering. Well simulated regions spanned the world and were not restricted to hot, cold or dry or wet bioclimatic zones—for
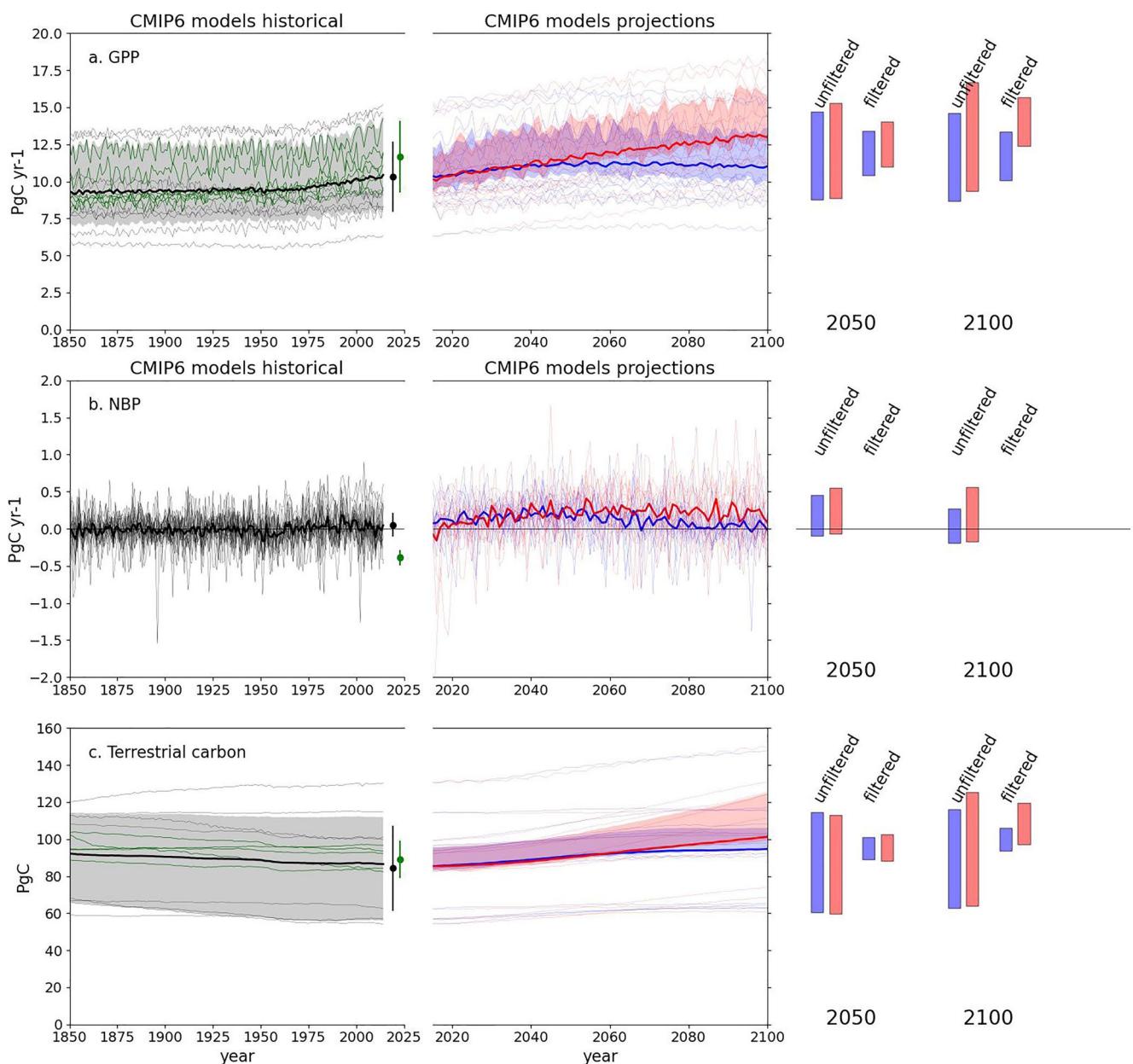
**Figure 16.** (top) CMIP6 simulated GPP, (middle) NBP, and (bottom) total terrestrial carbon stock for the historical period (1850–2014, left hand side) and future projections (to 2100, right hand side) for South East Asia. The panels show individual CMIP6 models in faint lines. For the historical period, individual models are shown in gray and multi model mean in thick black. CMIP6 mean and standard deviation are shown by the black dot and error bar, and REgional Carbon Cycle Assessment and Processes, Phase 2 (RECCAP2) assessed best estimate and uncertainty by the green dot and error bar. Models which fall within the RECCAP2 assessed range are shown in pale green. Future projections are for SSP1-2.6 (blue) and SSP3-7.0 (red), with single models in faint lines and multi-model mean in thick lines. Filtering just those models which match the RECCAP2 assessed range gives the red and blue shaded region. Mean values for 2050 and 2100 before and after filtering are shown to the right of the plots.

example, South America, Australasia, Russia and the permafrost regions were all simulated by the majority of models, while for Europe and South Asia skill was lower. Areas with fewer successful models were for different reasons—in South Asia it represented a real disparity between the CMIP6 ESMs which consistently and substantially under-estimated productivity while for Europe and Africa the CMIP models were clustered around the RECCAP2 estimate but the RECCAP2 estimated error bars are very tight and therefore remove many more models than in regions where the assessment is less confident.

**Table 10**
*Comparison of CMIP6 Simulated and REgional Carbon Cycle Assessment and Processes, Phase 2 Assessed Carbon Stocks and Fluxes for Australasia*

| | | CMIP6 (mean $\pm$ 1 sigma) | RECCAP2 (best estimate $\pm$ error estimate) |
|---|---|---|---|
| Fluxes PgC/yr mean over 2010–2019 inclusive | GPP | 4.85 $\pm$ 1.69 | 4.61 $\pm$ 2.62 |
| | NPP | 2.19 $\pm$ 0.84 | 2.12 $\pm$ 0.59 |
| | NEP | 0.18 $\pm$ 0.97 | 0.15 $\pm$ 0.07 |
| | NBP | 0.06 $\pm$ 0.07 | 0.0004 $\pm$ 0.08 |
| Stocks PgC mean over 2010–2019 inclusive | Vegetation carbon | 10.50 $\pm$ 4.34 | 11.88 $\pm$ 6.93 |
| | Soil carbon | 30.74 $\pm$ 15.22 | 44.74 $\pm$ 24.01 |
| | Product carbon | 0.09 $\pm$ 0.07 | 0.24 $\pm$ 0.22 |
| | Total terrest. carbon | 43.74 $\pm$ 17.12 | 53.86 $\pm$ 25.00 |

*Note.* Values are shown as the average over a 10 year period from 2010 to 2019, and ranges are one standard deviation unless stated otherwise. For model outputs this reflects spread across different models, while for assessed values this can include other means of assessing uncertainty. Units are PgC or PgC yr$^{-1}$.

For NBP regions were split into well simulated or poorly simulated. South America and Africa were well captured by the CMIP6 ensemble, but hardly at all for Central, East and South East Asia where only one, one, and none of the models passed the filtering respectively. In Central Asia the multi-model mean was only slightly smaller than the assessed range, but individual models fell outside this tightly constrained range, while in East Asia RECCAP2 assessed a higher NBP estimate than the CMIP6 mean. In South East Asia however there was a much bigger difference with RECCAP2 assessed NBP of $-0.39$ PgC yr$^{-1}$ from a range of estimates, while CMIP6 models consistently simulated a small sink of 0.05 PgC yr$^{-1}$. Land-use is especially dominant in this heavily managed region and this is not well captured by the CMIP6 ESMs in the detail required to recreate the observed carbon balance.

When simulating carbon stocks, there was a slightly lower level of success for most regions, with typically fewer than half passing the filtering. This may be due to the greater importance of land-use and management in controlling biomass stocks, or the well documented lack of agreement between CMIP models in simulating stocks compared to fluxes (Anav et al., 2013). This is especially true for soil carbon and residence times (Carvalhais et al., 2014; Todd-Brown et al., 2014) for all of South Asia, South-East Asia and Africa, $C_{tot}$ was less well simulated as these regions are heavily affected by land use (Figure S1 in Supporting Information S1). For North America, the reason for poorer carbon stock simulation may be due to bias in simulated residence time in soil. Other areas in high latitudes lack the carbon stock information to back this up.

### 5.1.3. Process-Analysis

Models with a nitrogen cycle consistently appear to score better than models without for both NBP and $C_{tot}$, but we note that this is not necessarily causal—caution is needed because there are conflating issues around model complexity and performance skill. For example, some ESMs make use of the same underlying land model (e.g., CESM2 and NorESM2 both use the CLM5 land model) and so these results are not made up of fully independent data points. A more nuanced consideration is that some models have larger research groups, or greater resource for model development and calibration leading to models with higher complexity also being better calibrated. That does not necessarily mean that the greater skill comes from the added complexity. Documentation and testing for each model may reveal the extent to which adding a process improves the representation of the carbon cycle (e.g., Wiltshire et al. (2021) for UKESM-CN). To fully isolate the impact of process addition to model skill a "with" and "without" pair of model simulations would be needed, and these are not available for CMIP6 simulations. Nonetheless there is reasonable evidence that suggests that nitrogen-cycle models better capture the productivity and storage of the terrestrial carbon cycle.

Dynamic vegetation shows little discernible signal in determining filtering skill at regional level. For GPP, DVGMs are comparable in skill having the same range of number of well simulated regions. NBP has slightly lower and $C_{tot}$ slightly higher number of well simulated regions. This apparent lack of contribution to skill is perhaps not unexpected for this process which captures very slow processes and does not have a marked impact on the carbon cycle up to present day but has been shown to be important in future projections (Davies-Barnard et al., 2015; Pugh et al., 2018).
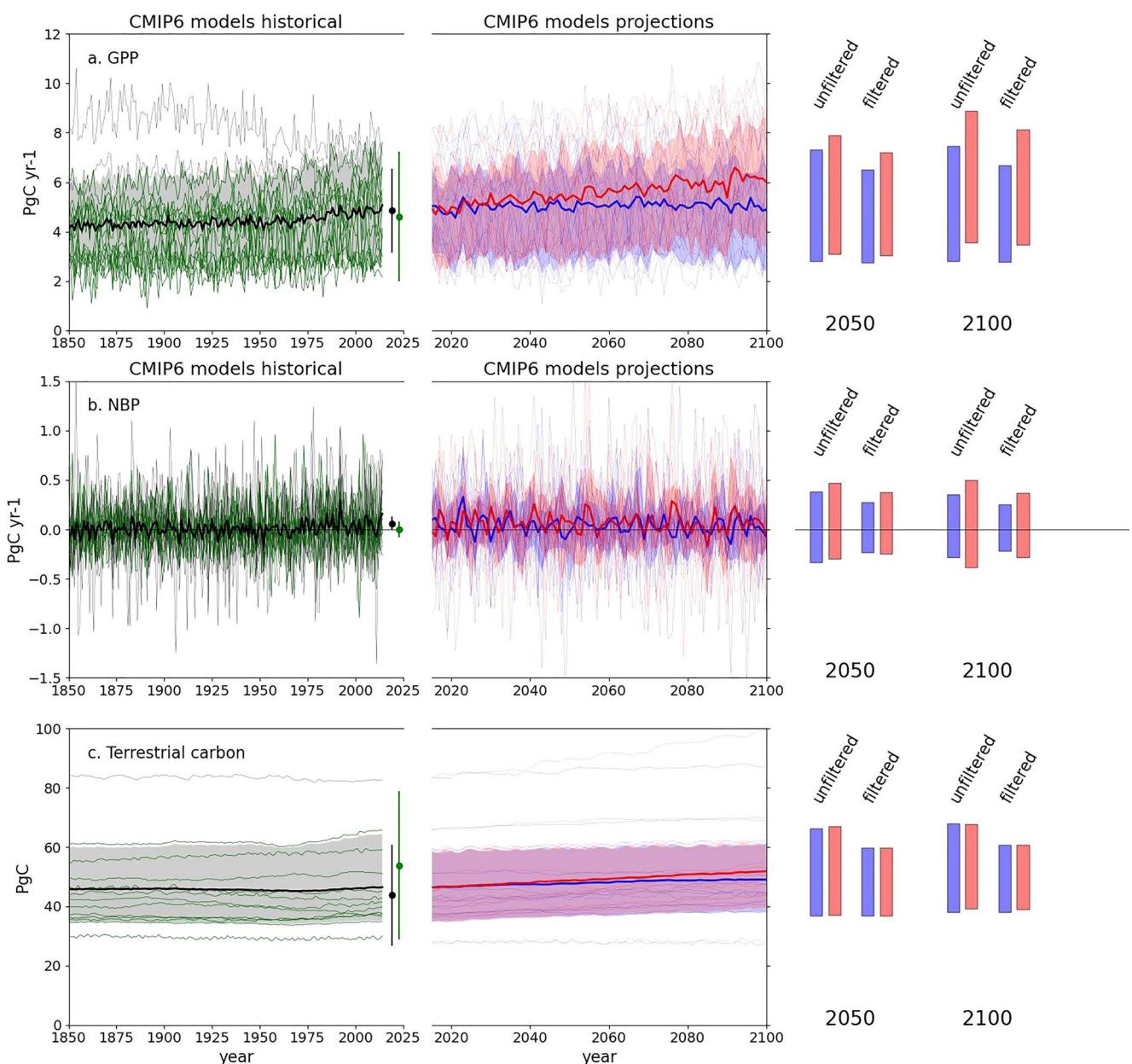
**Figure 17.** (top) CMIP6 simulated GPP, (middle) NBP, and (bottom) total terrestrial carbon stock for the historical period (1850–2014, left hand side) and future projections (to 2100, right hand side) for Australasia. The panels show individual CMIP6 models in faint lines. For the historical period, individual models are shown in gray and multi model mean in thick black. CMIP6 mean and standard deviation are shown by the black dot and error bar, and REgional Carbon Cycle Assessment and Processes, Phase 2 (RECCAP2) assessed best estimate and uncertainty by the green dot and error bar. Models which fall within the RECCAP2 assessed range are shown in pale green. Future projections are for SSP1-2.6 (blue) and SSP3-7.0 (red), with single models in faint lines and multi-model mean in thick lines. Filtering just those models which match the RECCAP2 assessed range gives the red and blue shaded region. Mean values for 2050 and 2100 before and after filtering are shown to the right of the plots.

Fire as a process, similar to nitrogen cycle but not as clear-cut, appears to slightly improve simulation skill. It increases the number of regions well simulated for NBP and $C_{tot}$. Fire is important to enable models to simulate the right vegetation cover for a given ecosystem and also the carbon residence time (and hence carbon storage) of land carbon, but the results are not clear-cut enough to conclude the impact of fire on model skill.

Very few models include permafrost carbon. In fact, only one land model (CLM5) does so, so the two ESMs with permafrost (CESM and NorESM) are effectively double counting. By sight the inclusion of permafrost does not improve GPP but does improve NBP and carbon storage. The former is not surprising as permafrost processes

**Table 11**
*Comparison of CMIP6 Simulated and REgional Carbon Cycle Assessment and Processes, Phase 2 Assessed Carbon Fluxes for the Permafrost Region*

|  |  | CMIP6 (mean ± 1 sigma) | RECCAP2 (best estimate ± error estimate) |
|---|---|---|---|
| Fluxes PgC/yr mean over 2010–2019 inclusive | GPP | 9.11 ± 1.79 | 7.0 ± 0.5 |
|  | NPP | 4.76 ± 1.28 | 4.3 ± 0.6 |
|  | NEP | 0.47 ± 1.76 | 0.9 ± 1.1 |
|  | NBP | 0.34 ± 0.17 | 0.65 ± 1.15 |
| Stocks PgC mean over 2010–2019 inclusive | Vegetation carbon | 49.41 ± 19.61 |  |
|  | Soil carbon | 440.40 ± 386.45 |  |
|  | Product carbon | 0.31 ± 0.29 |  |
|  | Total terrest. carbon | 512.78 ± 383.94 |  |

*Note.* Values are shown as the average over a 10 year period from 2010 to 2019, and ranges are one standard deviation unless stated otherwise. For model outputs this reflects spread across different models, while for assessed values this can include other means of assessing uncertainty. Units are PgC or PgC yr$^{-1}$.

would not be expected to have a big impact on GPP. The latter cannot be seen to be robust either because the regions which are better simulated are not permafrost regions, and so this is not an evaluation of permafrost carbon, but clearly an artifact that this particular land model performs well for carbon storage, but not as a result of including permafrost. It may, however, be related to the inclusion of vertically resolved carbon (Koven, Riley, Subin, et al., 2013) in the soil, which is a requirement to simulate permafrost carbon and may improve simulations of soil carbon and moisture elsewhere.

When we combine these processes to assess the role of the full degree of complexity the picture is mixed. In general models with high degrees of complexity (3 out of 4 of the focus processes), perform better than those with none of them. But there are cases of low-complexity models performing well and high-complexity models performing poorly. Given the high number of degrees of freedom of this analysis (number of regions and number of processes) and the few number of ESMs (not all of which are independent) it is not surprising that robust relationships cannot be found. However, the following broad statements can be made which are supported by the table:

- Nitrogen-cycle is beneficial. There is strong evidence that without considering the role of nutrients, CMIP ESMs over-estimate future potential for carbon sequestration (Hungate et al., 2003; Zaehle et al., 2015). The inclusion of nitrogen cycling in land models has therefore begun to address this issue reducing feedback strength (Arora et al., 2020) and limiting future carbon sinks. Much work is required to ensure process-realism of this development, but coordinated development and testing is showing encouraging signs (Davies-Barnard et al., 2020).

- Vegetation dynamics are needed to simulate long term changes in vegetation structure but are not well constrained by present day observations used here and have little influence on model skill of these metrics. Inclusion of this process though is important as future changes this century may be equally significant from natural dynamics as human land-use (Davies-Barnard et al., 2015) and even larger still beyond 2100 (Pugh et al., 2018). Existing vegetation dynamics models differ greatly in their own level of complexity, but recent literature suggests that a representation of tree demography (either by age or size class) is important (Argles et al., 2022).

- Fire as a process that disturbs vegetation and controls carbon storage in many ecosystems and is missing in many ESMs—especially in conjunction with vegetation dynamics. There is very limited evidence here that it helps evaluation against regional carbon balance assessment, but it is a vital process to represent in models as it could be a major mechanism for tipping points and abrupt ecosystem changes (Parry et al., 2022). Other disturbances, currently missing in both observation-based assessments and models might also be important, at least for some regions.

- Permafrost carbon is not yet represented in many models but is a commonly quoted process important in future carbon cycle changes (Chadburn et al., 2017; MacDougall et al., 2015). In order to simulate permafrost carbon, models are developing more realistic and extensive treatment of vertically resolved soils which have benefit to simulations outside permafrost areas too such as tropical forests where rooting depths can extend many meters. Evaluation of soil carbon in particular would benefit from process understanding of residence times, such as can be estimated from soil warming experiments (Van Gestel et al., 2018).
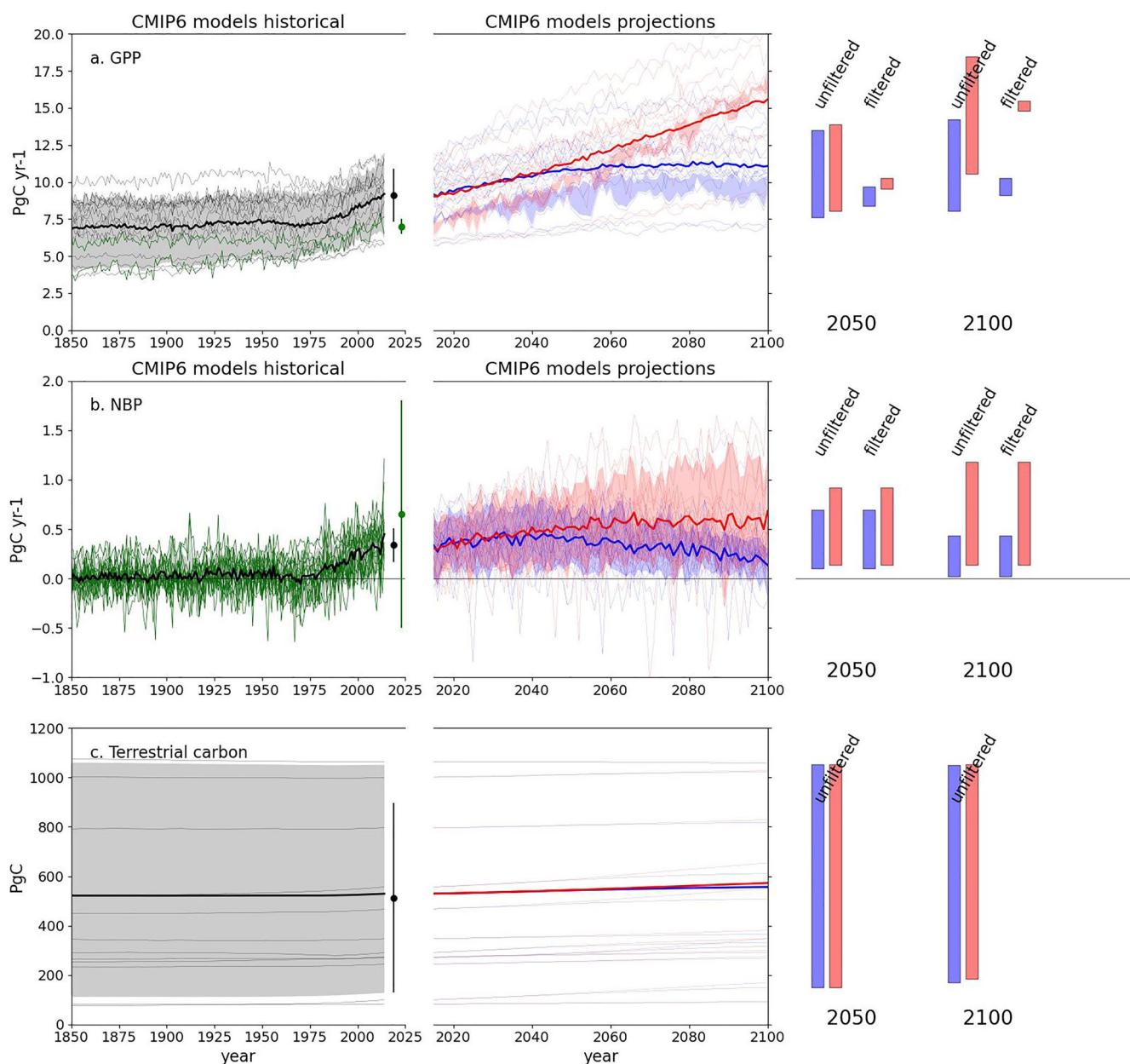
**Figure 18.** (top) CMIP6 simulated GPP, (middle) NBP, and (bottom) total terrestrial carbon stock for the historical period (1850–2014, left hand side) and future projections (to 2100, right hand side) for the permafrost region. The panels show individual CMIP6 models in faint lines. For the historical period, individual models are shown in gray and multi model mean in thick black. CMIP6 mean and standard deviation are shown by the black dot and error bar, and REgional Carbon Cycle Assessment and Processes, Phase 2 (RECCAP2) assessed best estimate and uncertainty by the green dot and error bar. Models which fall within the RECCAP2 assessed range are shown in pale green. Future projections are for SSP1-2.6 (blue) and SSP3-7.0 (red), with single models in faint lines and multi-model mean in thick lines. Filtering just those models which match the RECCAP2 assessed range gives the red and blue shaded region. Mean values for 2050 and 2100 before and after filtering are shown to the right of the plots.

More generally, routine evaluation and process-based sensitivity will enable models to develop. It is important that the push toward increasing complexity should not be to the detriment of improving existing parametrizations. Especially the response of the land carbon-cycle to elevated $CO_2$, and sensitivity to environmental changes.

In addition to this analysis, we know that the role of land-use is important, but these simulations lack enough detail to assess how land-use change and land management in CMIP6 models affects simulated outcomes. In CMIP5 some models did not include land-use at all (see Jones et al., 2013, Figure 2) and could relatively easily be excluded from projections for this reason. Pongratz et al. (2018) show large diversity of complexity level

**Table 12**
*CMIP6 Earth System Model Analysis by Region and Process*

| | | ACCESS-ESM1-5 | CanESM5 | CESM2-WACCM | CMCC-CM2-SR5 | CNRM-ESM2-1 | EC-Earth3-Veg | GFDL-ESM4 | INM-CM4-8 |
|---|---|---|---|---|---|---|---|---|---|
| | GPP | | | | | | | | |
| Region | Permafrost | | 1 | | | | 1 | | |
| | Russia | 1 | 1 | 1 | | 1 | 1 | 1 | 1 |
| | North America | 1 | 1 | | 1 | 1 | | 1 | |
| | Europe | 1 | | | | | | | |
| | Central Asia | 1 | 1 | 1 | | 1 | | 1 | 1 |
| | East Asia | | 1 | | 1 | | | | |
| | South Asia | | | | | | 1 | | 1 |
| | SE Asia | | 1 | 1 | 1 | | | | |
| | Africa | | | | | | 1 | | |
| | South America | 1 | 1 | 1 | | 1 | 1 | 1 | |
| | Australasia | 1 | 1 | 1 | 1 | 1 | | 1 | |
| | GPP Sum | 6 | 7 | 6 | 4 | 5 | 5 | 5 | 3 |
| | (% out of 11 regions) | 54.55 | 63.64 | 54.55 | 36.36 | 45.45 | 45.45 | 45.45 | 27.27 |
| | NBP | | | | | | | | |
| Region | Permafrost | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | Russia | | | | | | 1 | | 1 |
| | North America | 1 | 1 | 1 | | 1 | | | 1 |
| | Europe | | | 1 | | 1 | 1 | | |
| | Central Asia | 1 | | | | | | | |
| | East Asia | | | | | | | | |
| | South Asia | 1 | 1 | 1 | 1 | | 1 | | |
| | SE Asia | | | | | | | | |
| | Africa | 1 | 1 | 1 | 1 | 1 | 1 | | |
| | South America | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | Australasia | 1 | 1 | | 1 | 1 | | | |
| | NBP Sum | 7 | 6 | 6 | 5 | 6 | 6 | 2 | 4 |
| | (% out of 11 regions) | 63.64 | 54.55 | 54.55 | 45.45 | 54.55 | 54.55 | 18.18 | 36.36 |
| | $C_{tot}$ | | | | | | | | |
| Region | Permafrost | – | – | – | – | – | – | – | – |
| | Russia | | | | | | | | |
| | North America | 1 | 1 | | | 1 | 1 | – | – |
| | Europe | – | – | – | – | – | – | – | – |
| | Central Asia | 1 | 1 | 1 | 1 | | 1 | – | – |
| | East Asia | 1 | 1 | 1 | | | 1 | – | – |
| | South Asia | | 1 | | 1 | | 1 | – | – |
| | SE Asia | | 1 | 1 | | 1 | 1 | – | – |
| | Africa | | 1 | | | | | | |

| Model | | | | | | | | | |
| INM-CM5-0 | IPSL-CM6A-LR | MIROC-ES2L | MPI-ESM1-2-LR | NorESM2-LM | NorESM2-MM | TaiESM1 | UKESM1-0-LL | Sum | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | (% of 16 models) |
| | | | | | | | | 2 | 12.5 |
| 1 | 1 | | | 1 | 1 | | 1 | 12 | 75 |
| | 1 | 1 | | | | 1 | 1 | 9 | 56.25 |
| | 1 | | | | | | | 2 | 12.5 |
| 1 | 1 | 1 | 1 | 1 | 1 | | 1 | 13 | 81.25 |
| 1 | | 1 | 1 | 1 | 1 | | 1 | 8 | 50 |
| 1 | | | 1 | | | | | 4 | 25 |
| | | | 1 | | 1 | 1 | 1 | 7 | 43.75 |
| | 1 | 1 | 1 | | | 1 | 1 | 6 | 37.5 |
| | 1 | | | 1 | 1 | 1 | | 10 | 62.5 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 14 | 87.5 |
| 5 | 7 | 5 | 6 | 5 | 6 | 5 | 7 | | |
| 45.45 | 63.64 | 45.45 | 54.55 | 45.45 | 54.55 | 45.45 | 63.64 | | |
| | | | | | | | | | (% of 16 models) |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 16 | 100 |
| 1 | 1 | | 1 | 1 | 1 | | | 7 | 43.75 |
| 1 | 1 | | 1 | 1 | 1 | | 1 | 11 | 68.75 |
| | 1 | | 1 | 1 | 1 | | | 7 | 43.75 |
| | | | | | | | | 1 | 6.25 |
| 1 | | | | | | | | 1 | 6.25 |
| | 1 | 1 | 1 | 1 | 1 | | 1 | 11 | 68.75 |
| | | | | | | | | 0 | 0 |
| | 1 | 1 | 1 | 1 | 1 | | 1 | 12 | 75 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 16 | 100 |
| | 1 | 1 | | 1 | 1 | | 1 | 9 | 56.25 |
| 5 | 8 | 5 | 7 | 8 | 8 | 2 | 6 | | |
| 45.45 | 72.73 | 45.45 | 63.64 | 72.73 | 72.73 | 18.18 | 54.55 | | |
| | | | | | | | | | (% of 13 models) |
| – | – | – | – | – | – | – | – | – | – |
| | | | | | | | | 0 | 0 |
| – | | 1 | 1 | | 1 | | 1 | 8 | 61.5384615384615 |
| – | – | – | – | – | – | – | – | – | – |
| – | | 1 | 1 | 1 | 1 | | | 9 | 69.2307692307692 |
| – | | 1 | 1 | 1 | 1 | | | 8 | 61.5384615384615 |
| – | | 1 | | | | | 1 | 5 | 38.4615384615385 |
| – | | | | | 1 | | | 5 | 38.4615384615385 |
| | | 1 | 1 | 1 | 1 | | | 5 | 38.4615384615385 |

**Table 12**
*Continued*

| | | ACCESS-ESM1-5 | CanESM5 | CESM2-WACCM | CMCC-CM2-SR5 | CNRM-ESM2-1 | EC-Earth3-Veg | GFDL-ESM4 | INM-CM4-8 |
|---|---|---|---|---|---|---|---|---|---|
| | South America | 1 | 1 | 1 | | 1 | 1 | – | – |
| | Australasia | 1 | | 1 | 1 | 1 | 1 | – | – |
| | $C_{tot}$ Sum | 5 | 6 | 6 | 3 | 4 | 7 | – | – |
| | (% out of 9 regions) | 55.56 | 66.67 | 66.67 | 33.33 | 44.44 | 77.78 | – | – |
| Process | N-cycle | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 |
| | Dyn veg | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| | Fire | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 |
| | permafrost | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | process Sum | 1 | 0 | 3 | 2 | 1 | 3 | 2 | 0 |
| | (% out of 4 processes) | 25 | 0 | 75 | 50 | 25 | 75 | 50 | 0 |

*Note.* A "1" in a box indicates that that model passed the filtering for that region and variable, and a blank box indicates that it did not. A dash ("–") indicates that either the model did not provide that output or that the region did not assess that variable—and hence no filtering was possible. At the bottom of each column is a sum of the number of regions for which each models passed the filtering, and at the end of each row a sum of how many models passed for that region. The base of the table indicates (again with a "1") which model included each process from: nitrogen cycle; dynamic vegetation; fire and permafrost.

in ESMs for CMIP6 and plans beyond this. We therefore recommend use of the LUMIP factorial experiments (Lawrence et al., 2016) to address testing and evaluation of the components of land-use and management. It is no longer enough to simply categorize models which do or do not include land-use as the distinction is not binary.

### 5.2. Multi-Model Synthesis

Here we assemble a synthesis table of which regions are well or poorly simulated for the key variables of GPP, NBP, and vegetation and soil carbon by the CMIP6 multi-model mean. For each region we make a semi-quantitative assessment (Table 13) of whether simulated multi-model mean values are low, medium or high quality. Such a table can be used to inform judgment on the use of CMIP6 models for future projections. To do this we define:

- High quality—CMIP6 multi-model mean is within the RECCAP2 uncertainty bars, and we propose that CMIP6 outputs can be used to make useful inferences about these quantities.
- Medium quality—CMIP6 mean is outside RECCAP2 uncertainty, but error bars overlap, and we propose that CMIP6 outcomes may be useful but should be used with care and not over-interpreted as quantitatively robust.
- Low quality—no overlap in error bars between CMIP6 and RECCAP2, and we propose that CMIP6 outputs should not be used quantitatively for these variables/regions.

Table 13 shows that CMIP6 multi-model mean is remarkably good with respect to capturing the magnitude of carbon stocks and fluxes at regional scales, at least within the estimated uncertainty of the RECCAP2 assessments. While only five CMIP6 models pass the filtering for half or more regions in all three variables (GPP, NBP, and $C_{tot}$), the performance of the multi-model mean is much better. 31 out of 40 boxes are green, denoting that the CMIP6 mean is within the uncertainty range of RECCAP2, while 8 out of 40 boxes are yellow denoting overlap of the error bars. Only one box of the 40 is red which denotes values without overlap in the uncertainty ranges—here the NBP for South East Asia is poorly captured by CMIP6 models which in the mean simulate a sink (but spanning zero) compared to an assessed source from RECCAP2.

We note that there is some duplication in methodology with RECCAP2 assessments drawing on land model simulations many of which are shared with the CMIP6 ESMs. But we stress that none of these simulations are initialized at present day values and have been simulated from 1850 within climate models that exhibit their own errors and biases in meteorology and climate forcing. Nevertheless, we should be careful not to see this evaluation

| | Model | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| INM-CM5-0 | IPSL-CM6A-LR | MIROC-ES2L | MPI-ESM1-2-LR | NorESM2-LM | NorESM2-MM | TaiESM1 | UKESM1-0-LL | Sum | | |
| – | 1 | 1 | 1 | 1 | 1 | 1 | | 11 | 84.6153846153846 | |
| – | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 12 | 92.3076923076923 | |
| – | 2 | 6 | 6 | 6 | 6 | 3 | 3 | | | |
| – | 22.22 | 66.67 | 66.67 | 66.67 | 66.67 | 33.33 | 33.33 | | | |
| | | | | | | | | | (% of 16 models) | |
| 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 9 | 56.25 | |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 4 | 25 | |
| 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 8 | 50 | |
| 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 3 | 18.75 | |
| 0 | 0 | 1 | 3 | 3 | 3 | 0 | 2 | | | |
| 0 | 0 | 25 | 75 | 75 | 75 | 0 | 50 | | | |

as a "model versus observation" comparison, but rather a comparison of coupled climate models with a range of lines of evidence assessed by experts on a region-by-region basis. On the whole there are more cases of a negative bias (CMIP6 models simulating too small stocks or fluxes) than positive but there is no obvious systematic pattern of the sign of error.

Based on this assessment we think it is reasonable to present and consider future projections from the CMIP6 ESMs. But as discussed above we do not attempt to weight or select models based on how well they agree with RECCAP2 assessments. Based on numbers assembled in Section 4 we can build "waterfall" diagrams for each region and combine with regional fossil fuel emissions to synthesize the full net carbon balance of each region for the 21st century. These show how the state of the global carbon cycle has changed on two time horizons: between present day (defined here as 2015 being the start of future simulations in CMIP6) and "mid-term" (2040–2060: Figure 19) and "long-term" (2080–2100: Figure 20) under illustrative low (SSP1-2.6) and high (SSP3-7.0) emissions scenarios. Unfortunately, it is not possible from existing CMIP6 simulations to diagnose land-use fluxes distinct from other changes in simulated land stocks. Hence, we are restricted to showing the fossil fuel emissions for each region (taken directly from the scenarios) and the changes in simulated land carbon stocks, apportioned here into living carbon (cVeg) and dead carbon (the sum of cSoil, cLitter and cProduct). We plot fossil fuel emissions from 2015 to 2050 and 2090 respectively, and changes in stocks from the recent past (average from 1995 to 2014) up to the average of 2040–2060 and 2080–2100 respectively.

At 2050, the higher $CO_2$ concentration of SSP3-7.0 drives larger terrestrial sinks while the timescale of climate response means that the climate is relatively similar between high and low scenarios. Therefore, for all regions, natural sinks are larger under SSP3-7.0 than SSP1-2.6.

The analysis so far has been on the land sinks, but we can also compare with regional fossil fuel emissions from the socio-economic scenarios. At 2050

**Table 13**
*Synthesis of CMIP6 Multi-Model Mean Fidelity Recreating REgional Carbon Cycle Assessment and Processes, Phase 2 Assessed Present Day Carbon Stocks and Fluxes*

| | cVeg | cSoil | GPP | NBP |
|---|---|---|---|---|
| North America | + | = | + | − |
| South America | + | − | = | − |
| Europe | n/a | n/a | − | − |
| Africa | = | + | = | − |
| Russia | − | + | = | = |
| Central Asia | + | − | = | − |
| East Asia | = | = | = | − |
| South Asia | − | − | − | + |
| South East Asia | − | = | − | − |
| Australasia | − | − | = | + |
| Permafrost | n/a | n/a | + | − |

*Note.* Green: high quality; yellow: medium quality; red: low quality. "+" and "−" indicate if the CMIP6 is biased high or low respectively.
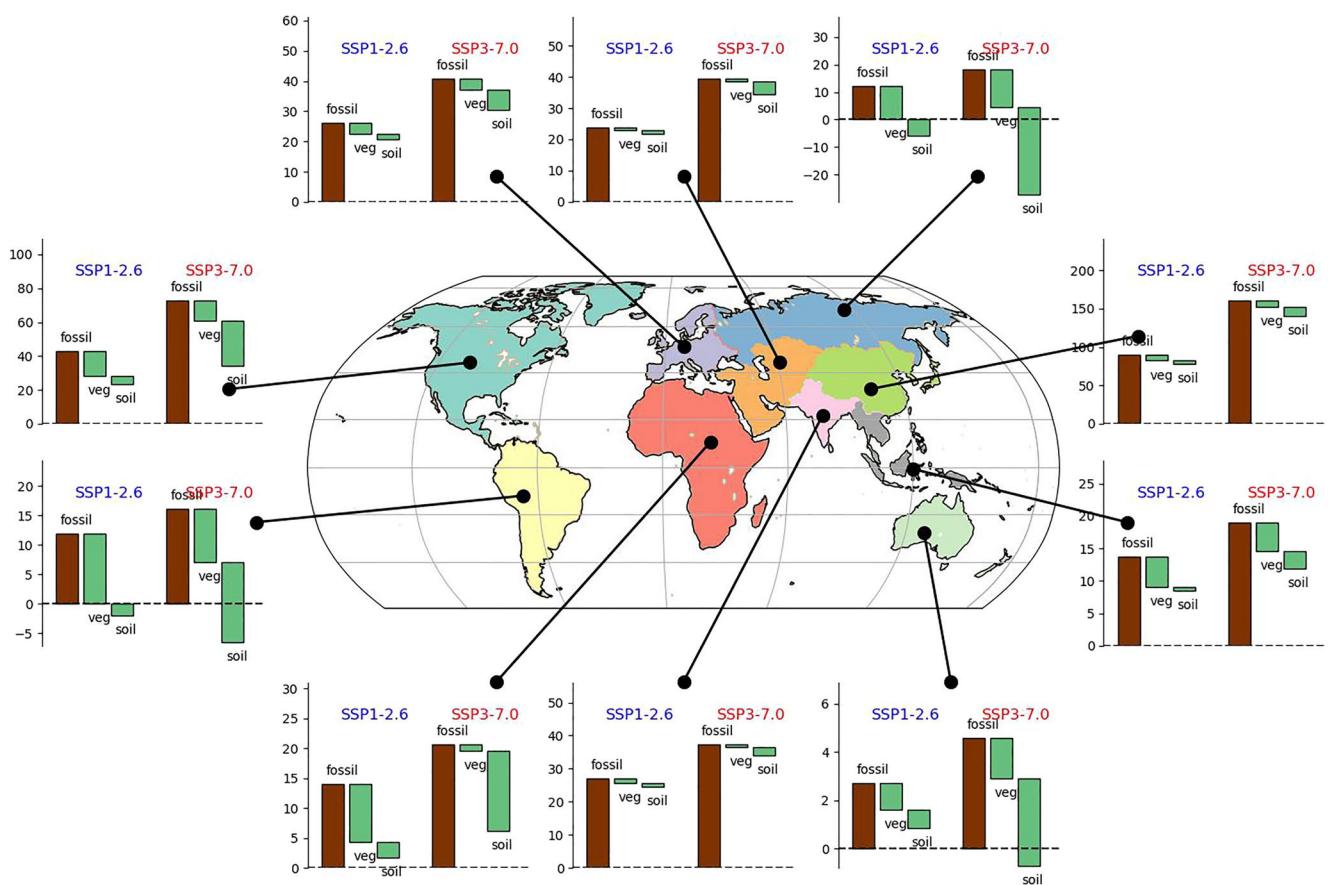
**Figure 19.** Global synthesis of regional carbon balance in 2050s under low (SSP1-2.6) and high (SSP3-7.0) scenarios.

the main fossil-fuel emitting regions will continue to be substantial sources of $CO_2$ to the atmosphere. Namely—North America, Europe and Central-, South-, South East- and East-Asia. This is true even for the low emissions scenario as it is not until after mid-century when most regions will see large emissions reductions even under ambitious mitigation.

For other regions, the net response of fossil-fuel and terrestrial sinks differs. In Russia, large natural sinks outweigh regional fossil-fuel emissions and the region is a net sink. But we note that for this region in particular a large fraction of the global permafrost stock exists and is not accounted for in these projections as we discussed in Section 4.11. Any large-scale loss of carbon from thawing permafrost has the potential to turn this and other high-latitude regions into substantial sources. Elsewhere, South America is also simulated to be a net sink but this is also highly uncertain, albeit for different reasons. The role of land-use and land-use change is central and future scenarios do not attempt to capture national-scale policies which will clearly play a crucial role in determining the carbon balance of countries with large tropical forests (e.g., Wiltshire et al., 2022). South America, and especially the Amazon forest is also an area identified as especially vulnerable to future climate change and loss of tropical forest (Armstrong McKay et al., 2022; Salazar & Nobre, 2010) and some CMIP6 models have been shown to exhibit localized dieback of the Amazon forest in future projections (Parry et al., 2022). Africa and Australasia show a relatively close balance of fossil fuel emission and terrestrial sink partly due to these regions being relatively less fossil-fuel intensive for their land-area.

The situation at 2100 is somewhat different and all regions exhibit a strong dependence on the choice of scenario. Results for SSP1-2.6 are very similar to those at 2050 due to the approximate stabilization of regional climate after mid-century (supplementary figures). But for SSP3-7.0 the very large increase in fossil-fuel emissions and larger regional climate changes mean that all regions except for Russia are now a net source of carbon. As discussed above, even Russia could become a substantial source in these models if full representation of permafrost processes was accounted for. These results are the basis for the IPCC assessment that the proportion of $CO_2$ emissions taken up by natural carbon sinks is smaller in scenarios with higher cumulative $CO_2$ emissions.
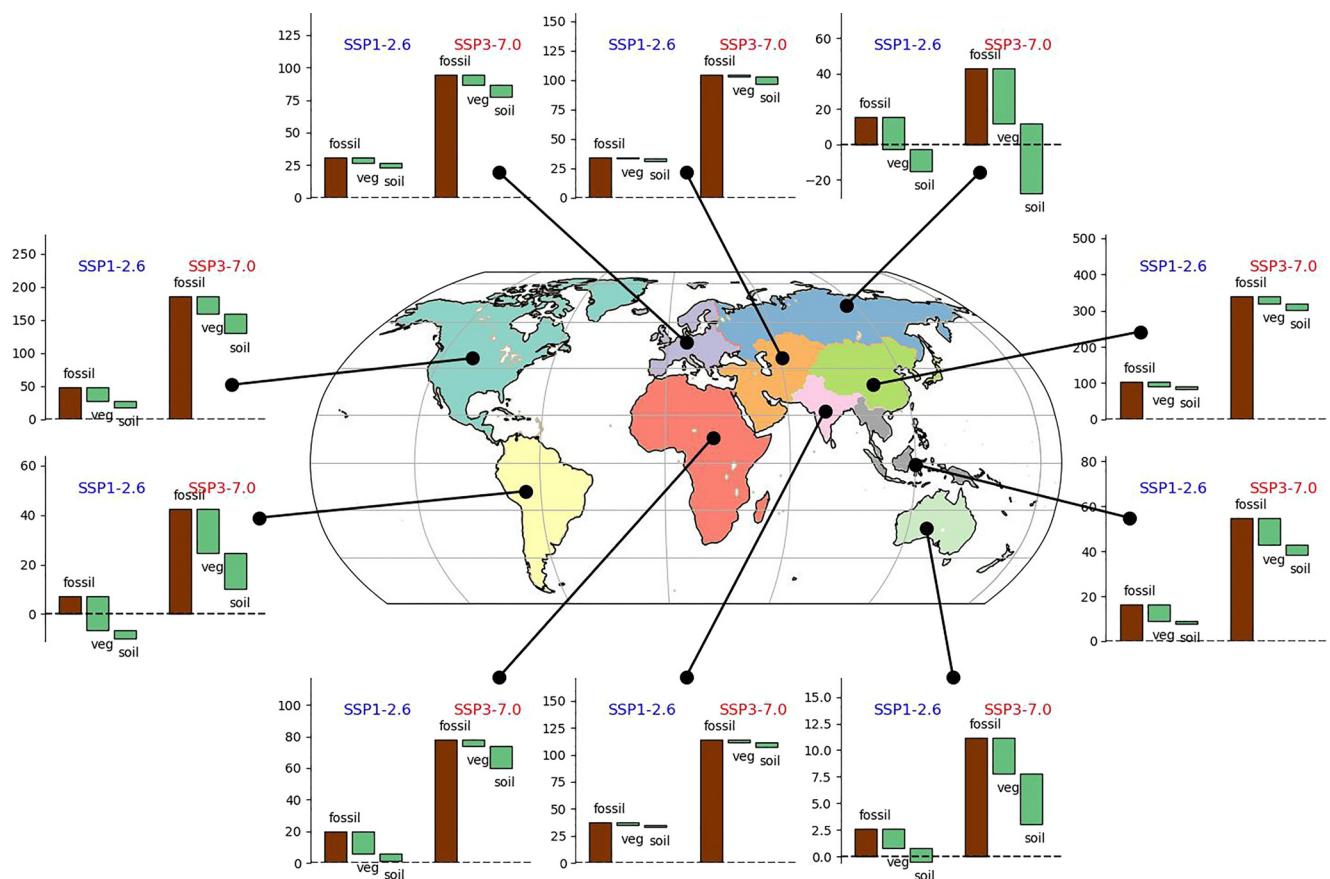
**Figure 20.** Global synthesis of regional carbon balance in 2100 under low (SSP1-2.6) and high (SSP3-7.0) scenarios.

## 6. Conclusions

CMIP6 models project future changes in carbon fluxes and stores and this is the primary resource used to quantify RCB estimates consistent with global climate goals. But these simulations are not evaluated at regional scales nor fully consistent with land-modeling used to quantify the present-day carbon budgets. Here we draw on the updated RECCAP2 assessment of regional carbon balance to assess the status of CMIP6 projections and consistency with our expert knowledge of the contemporary carbon cycle.

When evaluating the multi-model mean, we find generally good agreement with 78% of regions' fluxes and stocks simulated within the RECCAP2 assessed ranges, and less than 3% failing to overlap uncertainty estimates. This consistency means we can have reasonable trust in the credibility of future projections based on CMIP6 multi-model mean and we can use the ensemble of models to explore the future carbon balance of each region under different future scenarios and on different time horizons. We find that by mid-century, projected changes are relatively insensitive to scenario but that the choice of scenario has set the world on very different paths beyond this time horizon, with a very strong dependence on scenario by 2100. At mid-century, most regions are a carbon source even under aggressive mitigation pathways and by 2100 all regions are a carbon source under high emissions scenarios.

We demonstrated a very simplistic filtering approach to show the impact of imposing agreement between the CMIP6 ensemble and regional assessments. This analysis provides insight into model-by-model systematic biases and guides development, and regional insight into the importance of different processes. We find that more than half the models perform well at more than half the regions but no model performs well in all regions or for all output variables. Models with greater complexity, especially inclusion of the nitrogen cycle, appear to perform better than lower-complexity models but the sample size does not allow robust attribution of skill to process inclusion. Hence we recommend that analyses use as many CMIP6 models as possible and do not rely on single models unless directly evaluated for a specific study.

For many reasons this filtering approach is not robust enough to be applied as an actual constraint and more research is required into potential techniques to do this. Possible constraints on future projections could include: adding observationally constrained estimates (e.g., Chadburn et al., 2017) as a correction term in permafrost regions to account for CMIP models absence of process-completeness; adjusting regional projections, especially in tropical forests, to account for national-scale land-use policies; or applying emergent constraints which have been shown to be regional in their applicability (e.g., interannual variability can be used to constrain tropical carbon cycle sensitivity to warming (Cox et al., 2013) and seasonal cycle response mid-latitude sensitivity to $CO_2$ (Wenzel et al., 2014)) or applied to sensitivity of processes (such as GPP response to doubling CO2). Canadell et al. (2021) show a synthesis of powerful emergent constraints in their Figure 5.28 which could be used alongside regional evaluations from RECCAP2.

We argue though that our regional analysis of these global models brings insight into their behavior and offers the potential to reduce the spread across model projections at global scale and we encourage the land modeling community to more regularly consider detailed regional assessments of global models.

## Conflict of Interest

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

## Data Availability Statement

All CMIP6 model output datasets analyzed during this study are available online at https://esgf-node.llnl.gov/search/cmip6/ and code required to reproduce figures is available at https://github.com/ChrisJones-MOHC/RECCAP2Future_2023 (ChrisJones-MOHC, 2023) and Zenodo at https://doi.org/10.5281/zenodo.8420250.

## References

Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., et al. (2019). ESD reviews: Model dependence in multi-model climate ensembles: Weighting, sub-selection and out-of-sample testing. *Earth System Dynamics*, *10*(1), 91–105. https://doi.org/10.5194/ESD-10-91-2019

Ahlström, A., Raupach, M. R., Schurgers, G., Smith, B., Arneth, A., Jung, M., et al. (2015). The dominant role of semi-arid ecosystems in the trend and variability of the land $CO_2$ sink. *Science*, *348*(6237), 895–899. https://doi.org/10.1126/SCIENCE.AAA1668

Anav, A., Friedlingstein, P., Kidston, M., Bopp, L., Ciais, P., Cox, P., et al. (2013). Evaluating the land and ocean components of the global carbon cycle in the CMIP5 Earth system models. *Journal of Climate*, *26*(18), 6801–6843. https://doi.org/10.1175/JCLI-D-12-00417.1

Argles, A. P. K., Moore, J. R., & Cox, P. M. (2022). Dynamic Global Vegetation Models: Searching for the balance between demographic process representation and computational tractability. *PLOS Climate*, *1*(9), e0000068. https://doi.org/10.1371/JOURNAL.PCLM.0000068

Armstrong McKay, D. I., Staal, A., Abrams, J. F., Winkelmann, R., Sakschewski, B., Loriani, S., et al. (2022). Exceeding 1.5°C global warming could trigger multiple climate tipping points. *Science*, *377*(6611), eabn7950. https://doi.org/10.1126/SCIENCE.ABN7950

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

Assis, T. O., de Aguiar, A. P. D., von Randow, C., de Paula Gomes, D. M., Kury, J. N., Ometto, J. P. H. B., & Nobre, C. A. (2020). $CO_2$ emissions from forest degradation in Brazilian Amazon. *Environmental Research Letters*, *15*(10), 104035. https://doi.org/10.1088/1748-9326/AB9CFC

Bastos, A., Janssens, I. A., Gouveia, C. M., Trigo, R. M., Ciais, P., Chevallier, F., et al. (2016). European land $CO_2$ sink influenced by NAO and East-Atlantic Pattern coupling. *Nature communications*, *7*(1), 10315. https://doi.org/10.1038/ncomms10315

Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., Watson-Parris, D., et al. (2020). Bounding global aerosol radiative forcing of climate change. *Reviews of Geophysics*, *58*(1), e2019RG000660. https://doi.org/10.1029/2019RG000660

Bloom, A. A., Exbrayat, J.-F., van der Velde, I. R., Feng, L., & Williams, M. (2016). The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proceedings of the National Academy of Sciences*, *113*(5), 1285–1290. https://doi.org/10.1073/pnas.1515160113

Booth, B. B. B., Harris, G. R., Murphy, J. M., House, J. I., Jones, C. D., Sexton, D., & Sitch, S. (2017). Narrowing the range of future climate projections using historical observations of atmospheric $CO_2$. *Journal of Climate*, *30*(8), 3039–3053. https://doi.org/10.1175/JCLI-D-16-0178.1

Braswell, B. H., Schimel, D. S., Linder, E., & Moore, B. (1997). The response of global terrestrial ecosystems to interannual temperature variability. *Science*, *278*(5339), 870–872. https://doi.org/10.1126/SCIENCE.278.5339.870

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., & Knutti, R. (2020). Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth System Dynamics*, *11*(4), 995–1012. https://doi.org/10.5194/ESD-11-995-2020

Burke, E., Chadburn, S., & Huntingford, C. (2022). Thawing permafrost as a nitrogen fertiliser: Implications for climate feedbacks. *Nitrogen*, *3*(2), 353–375. https://doi.org/10.3390/NITROGEN3020023

Burke, E. J., Jones, C. D., & Koven, C. D. (2013). Estimating the permafrost-carbon climate response in the CMIP5 climate models using a simplified approach. *Journal of Climate*, *26*(14), 4897–4909. https://doi.org/10.1175/JCLI-D-12-00550.1

Burton, C., Betts, R., Cardoso, M., Feldpausch, T. R., Harper, A., Jones, C. D., et al. (2019). Representation of fire, land-use change and vegetation dynamics in the Joint UK Land Environment Simulator vn4.9 (JULES). *Geoscientific Model Development*, *12*(1), 179–193. https://doi.org/10.5194/gmd-12-179-2019

Canadell, J. G., Monteiro, P. M. S., Costa, M. H., Cotrim da Cunha, L., Cox, P. M., Eliseev, A. V., et al. (2021). Global carbon and other biogeochemical cycles and feedbacks supplementary material. In *Climate Change 2021: The Physical Science Basis. Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. https://doi.org/10.1017/9781009157896.007

Carvalhais, N., Forkel, M., Khomik, M., Bellarby, J., Jung, M., Migliavacca, M., et al. (2014). Global covariation of carbon turnover times with climate in terrestrial ecosystems. *Nature*, *514*(7521), 213–217. https://doi.org/10.1038/nature13731

Cavallaro, N., Shrestha, G., Birdsey, R., Mayes, M. A., Najjar, R. G., Reed, S. C., et al. (Eds.). (2018). *U.S. Global Change Research Program* (p. 878). https://doi.org/10.7930/SOCCR2.2018

Chadburn, S. E., Burke, E. J., Cox, P. M., Friedlingstein, P., Hugelius, G., & Westermann, S. (2017). An observation-based constraint on permafrost loss as a function of global warming. *Nature Climate Change*, *7*(5), 340–344. https://doi.org/10.1038/NCLIMATE3262

ChrisJones-MOHC. (2023). ChrisJones-MOHC/RECCAP2Future_2023: RECCAP2 future (1.0). *Zenodo*. https://doi.org/10.5281/zenodo.8420250

Chylek, P., Folland, C., Klett, J. D., Wang, M., Hengartner, N., Lesins, G., & Dubey, M. K. (2022). Annual mean Arctic amplification 1970–2020: Observed and simulated by CMIP6 climate models. *Geophysical Research Letters*, *49*(13), e2022GL099371. https://doi.org/10.1029/2022GL099371

Ciais, P., Bastos, A., Chevallier, F., Lauerwald, R., Poulter, B., Canadell, P., 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). *Geoscientific Model Development*, *15*(3), 1289–1316. https://doi.org/10.5194/GMD-15-1289-2022

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., et al. (2013). Carbon and other biogeochemical cycles. In Intergovernmental Panel on Climate Change (Ed.), *Climate Change 2013—The Physical Science Basis* (pp. 465–570). Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.015

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., et al. (2013). Long-term climate change: Projections, commitments and irreversibility. In *Climate Change 2013—The Physical Science Basis* (pp. 1076–1136). https://doi.org/10.1017/cbo9781107415324.025

Cox, P. M., Pearson, D., Booth, B. B., Friedlingstein, P., Huntingford, C., Jones, C. D., & Luke, C. M. (2013). Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature*, *494*(7437), 341–344. https://doi.org/10.1038/nature11882

Davies-Barnard, T., Meyerholt, J., Zaehle, S., Friedlingstein, P., Brovkin, V., Fan, Y., et al. (2020). Nitrogen cycling in CMIP6 land surface models: Progress and limitations. *Biogeosciences*, *17*(20), 5129–5148. https://doi.org/10.5194/bg-17-5129-2020

Davies-Barnard, T., Valdes, P. J., Singarayer, J. S., Wiltshire, A. J., & Jones, C. D. (2015). Quantifying the relative importance of land cover change from climate and land use in the representative concentration pathways. *Global Biogeochemical Cycles*, *29*(6), 842–853. https://doi.org/10.1002/2014GB004949

Erb, K. H., Luyssaert, S., Meyfroidt, P., Pongratz, J., Don, A., Kloster, S., et al. (2017). Land management: Data availability and process understanding for global change studies. *Global Change Biology*, *23*(2), 512–533. https://doi.org/10.1111/gcb.13443

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

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., et al. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, *9*(2), 102–110. https://doi.org/10.1038/s41558-018-0355-y

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Bakker, D. C. E., Hauck, J., et al. (2022). Global carbon budget 2021. *Earth System Science Data*, *14*(4), 1917–2005. https://doi.org/10.5194/ESSD-14-1917-2022

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

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., et al. (2022). Global carbon budget 2022. *Earth System Science Data*, *14*(11), 4811–4900. https://doi.org/10.5194/ESSD-14-4811-2022

Gasser, T., Crepin, L., Quilcaille, Y., Houghton, R. A., Ciais, P., & Obersteiner, M. (2020). Historical $CO_2$ emissions from land use and land cover change and their uncertainty. *Biogeosciences*, *17*(15), 4075–4101. https://doi.org/10.5194/BG-17-4075-2020

Gedney, N., Huntingford, C., Comyn-Platt, E., & Wiltshire, A. (2019). Significant feedbacks of wetland methane release on climate change and the causes of their uncertainty. *Environmental Research Letters*, *14*(8), 084027. https://doi.org/10.1088/1748-9326/ab2726

Gloor, M., Gatti, L., Brienen, R., Feldpausch, T. R., Phillips, O. L., Miller, J., et al. (2012). The carbon balance of South America: A review of the status, decadal trends and main determinants. *Biogeosciences*, *9*(12), 5407–5430. https://doi.org/10.5194/BG-9-5407-2012

Goll, D. S., Brovkin, V., Parida, B. R., Reick, C. H., Kattge, J., Reich, P. B., et al. (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

Graven, H. D., Keeling, R. F., Piper, S. C., Patra, P. K., Stephens, B. B., Wofsy, S. C., et al. (2013). Enhanced seasonal exchange of $CO_2$ by northern ecosystems since 1960. *Science*, *341*(6150), 1085–1089. https://doi.org/10.1126/SCIENCE.1239207

Harris, N. L., Gibbs, D. A., Baccini, A., Birdsey, R. A., de Bruin, S., Farina, M., et al. (2021). Global maps of twenty-first century forest carbon fluxes. *Nature Climate Change*, *113*(11), 234–240. https://doi.org/10.1038/s41558-020-00976-6

Haverd, V., Smith, B., Nieradzik, L., Briggs, P. R., Woodgate, W., Trudinger, C. M., et al. (2018). A new version of the CABLE land surface model (Subversion revision r4601) incorporating land use and land cover change, woody vegetation demography, and a novel optimisation-based approach to plant coordination of photosynthesis. *Geoscientific Model Development*, *11*(7), 2995–3026. https://doi.org/10.5194/GMD-11-2995-2018

Hengl, T., De Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*, *12*(2), e0169748. https://doi.org/10.1371/JOURNAL.PONE.0169748

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

Hidy, D., Barcza, Z., Hollós, R., Dobor, L., Ács, T., Zacháry, D., et al. (2022). Soil-related developments of the Biome-BGCMuSo v6.2 terrestrial ecosystem model. *Geoscientific Model Development*, *15*(5), 2157–2181. https://doi.org/10.5194/GMD-15-2157-2022

Hidy, D., Barcza, Z., MarjanoviÀ, H., Sever, M. Z. O., Dobor, L., Gelybó, G., et al. (2016). Terrestrial ecosystem process model Biome-BGCMuSo v4.0: Summary of improvements and new modeling possibilities. *Geoscientific Model Development*, *9*(12), 4405–4437. https://doi.org/10.5194/GMD-9-4405-2016

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

Hungate, B. A., Dukes, J. S., Shaw, M. R., Luo, Y., & Field, C. B. (2003). Nitrogen and climate change. *Science*, *302*(5650), 1512–1513. https://doi.org/10.1126/SCIENCE.1091390

Huntzinger, D. N., Schwalm, C., Michalak, A. M., Schaefer, K., King, A. W., Wei, Y., et al. (2013). The North American carbon program multi-scale synthesis and terrestrial model intercomparison project—Part 1: Overview and experimental design. *Geoscientific Model Development*, *6*, 2121–2133. https://doi.org/10.5194/GMD-6-2121-2013

Intergovernmental Panel on Climate Change (IPCC). (2014). Summary for policymakers. In *Climate Change 2013—The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1–30). Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.004

IPCC. (2018). Special Report on Global Warming of 1.5°C. *Chapter 2: Mitigation pathways compatible with 1.5°C in the context of sustainable development*. Retrieved from https://www.ipcc.ch/site/assets/uploads/sites/2/2019/02/SR15_Chapter2_Low_Res.pdf

IPCC. (2021). Summary for policymakers. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, et al. (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 3–32). Cambridge University Press. https://doi.org/10.1017/9781009157896.001

IPCC. (2022). Summary for policymakers. [P.R. Shukla, J. Skea, A. Reisinger, R. Slade, R. Fradera, M. Pathak, et al. (Eds.)]. In P. R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, et al. (Eds.), *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. https://doi.org/10.1017/9781009157926.001

Jones, C., Robertson, E., Arora, V., Friedlingstein, P., Shevliakova, E., Bopp, L., et al. (2013). Twenty-first-century compatible $CO_2$ emissions and airborne fraction simulated by CMIP5 Earth system models under four representative concentration pathways. *Journal of Climate*, *26*(13), 4398–4413. https://doi.org/10.1175/JCLI-D-12-00554.1

Jones, C. D., & Cox, P. M. (2001). Constraints on the temperature sensitivity of global soil respiration from the observed interannual variability in atmospheric $CO_2$. *Atmospheric Science Letters*, *2*(1–4), 166–172. https://doi.org/10.1006/asle.2001.0041

Jones, C. D., & Friedlingstein, P. (2020). Quantifying process-level uncertainty contributions to TCRE and carbon budgets for meeting Paris Agreement climate targets. *Environmental Research Letters*, *15*(7), 074019. https://doi.org/10.1088/1748-9326/ab858a

Kirschbaum, M. U. F., & Watt, M. S. (2011). Use of a process-based model to describe spatial variation in Pinus radiata productivity in New Zealand. *Forest Ecology and Management*, *262*(6), 1008–1019. https://doi.org/10.1016/J.FORECO.2011.05.036

Kloster, S., Mahowald, N. M., Randerson, J. T., & Lawrence, P. L. (2011). The impacts of climate, land use, and demography on fires during the 21st century simulated by CLM-CN. *Biogeosciences Discussions*, *8*, 9709–9746.

Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., Eyring, V., et al. (2017). A climate model projection weighting scheme accounting for performance and interdependence. *Geophysical Research Letters*, *44*(4), 1909–1918. https://doi.org/10.1002/2016GL072012

Kondo, M., Patra, P. K., Sitch, S., Friedlingstein, P., Poulter, B., Chevallier, F., et al. (2020). State of the science in reconciling top-down and bottom-up approaches for terrestrial $CO_2$ budget. *Global Change Biology*, *26*(3), 1068–1084. https://doi.org/10.1111/GCB.14917

Kondo, M., Sitch, S., Ciais, P., Achard, F., Kato, E., Pongratz, J., et al. (2022). Are land-use change emissions in Southeast Asia decreasing or increasing? *Global Biogeochemical Cycles*, *36*(1), e2020GB006909. https://doi.org/10.1029/2020GB006909

Koven, C. D., Riley, W. J., & Stern, A. (2013). Analysis of permafrost thermal dynamics and response to climate change in the CMIP5 Earth System Models. *Journal of Climate*, *26*(6), 1877–1900. https://doi.org/10.1175/JCLI-D-12-00228.1

Koven, C. D., Riley, W. J., Subin, Z. M., Tang, J. Y., Torn, M. S., Collins, W. D., et al. (2013). The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4. *Biogeosciences*, *10*(11), 7109–7131. https://doi.org/10.5194/BG-10-7109-2013

Kurz, W. A., Dymond, C. C., Stinson, G., Rampley, G. J., Neilson, E. T., Carroll, A. L., et al. (2008). Mountain pine beetle and forest carbon feedback to climate change. *Nature*, *452*(7190), 987–990. https://doi.org/10.1038/nature06777

Lawrence, D. M., Hurtt, G. C., Arneth, A., Brovkin, V., Calvin, K. V., Jones, A. D., et al. (2016). The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: Rationale and experimental design. *Geoscientific Model Development*, *9*, 2973–2998. https://doi.org/10.5194/gmd-9-2973-2016

Lee, J.-Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J. P., et al. (2021). Future global climate: Scenario-based projections and near-term information supplementary material. In *Climate Change 2021: The Physical Science Basis. Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. https://doi.org/10.1017/9781009157896.006

Liang, S., Cheng, J., Jia, K., Jiang, B., Liu, Q., Xiao, Z., et al. (2021). The global land surface satellite (GLASS) product suite. *Bulletin of the American Meteorological Society*, *102*(2), E323–E337. https://doi.org/10.1175/BAMS-D-18-0341.1

Liang, Y., Gillett, N. P., & Monahan, A. H. (2020). Climate model projections of 21st century global warming constrained using the observed warming trend. *Geophysical Research Letters*, *47*(12), e2019GL086757. https://doi.org/10.1029/2019GL086757

Lovenduski, N. S., & Bonan, G. B. (2017). Reducing uncertainty in projections of terrestrial carbon uptake. *Environmental Research Letters*, *12*(4), 044020. https://doi.org/10.1088/1748-9326/aa66b8

MacDougall, A. H., Zickfeld, K., Knutti, R., & Matthews, H. D. (2015). Sensitivity of carbon budgets to permafrost carbon feedbacks and non-$CO_2$ forcings. *Environ. Res. Lett.*, *10*(12), 125003. https://doi.org/10.1088/1748-9326/10/12/125003

Merrifield, A. L., Brunner, L., Lorenz, R., Medhaug, I., & Knutti, R. (2020). An investigation of weighting schemes suitable for incorporating large ensembles into multi-model ensembles. *Earth System Dynamics*, *11*(3), 807–834. https://doi.org/10.5194/ESD-11-807-2020

Murray-Tortarolo, G., Poulter, B., Vargas, R., Hayes, D., Michalak, A. M., Williams, C., et al. (2022). A process-model perspective on recent changes in the carbon cycle of North America. *Journal of Geophysical Research: Biogeosciences*, *127*(9), e2022JG006904. https://doi.org/10.1029/2022JG006904

Notz, D. (2015). How well must climate models agree with observations? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2052), 20140164. https://doi.org/10.1098/RSTA.2014.0164

Parry, I. M., Ritchie, P. D. L., & Cox, P. M. (2022). Evidence of localised Amazon rainforest dieback in CMIP6 models. *Earth System Dynamics*, *13*(4), 1667–1675. https://doi.org/10.5194/ESD-13-1667-2022

Piao, S. L., Ito, A., Li, S. G., Huang, Y., Ciais, P., Wang, X. H., et al. (2012). The carbon budget of terrestrial ecosystems in East Asia over the last two decades. *Biogeosciences*, *9*, 3571–3586. https://doi.org/10.5194/BG-9-3571-2012

Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *Soil*, *7*(1), 217–240. https://doi.org/10.5194/SOIL-7-217-2021

Pongratz, J., Dolman, H., Don, A., Erb, K. H., Fuchs, R., Herold, M., et al. (2018). Models meet data: Challenges and opportunities in implementing land management in Earth system models. *Global Change Biology*, *24*(4), 1470–1487. https://doi.org/10.1111/gcb.13988

Poulter, B., Bastos, A., Canadell, J., Ciais, P., Gruber, N., Hauck, J., et al. (2022). Inventorying Earth's land and ocean greenhouse gases. *Eos, Transactions American Geophysical Union*, *103*. https://doi.org/10.1029/2022EO179084

Pugh, T. A. M., Jones, C. D., Huntingford, C., Burton, C., Arneth, A., Brovkin, V., et al. (2018). A large committed long-term sink of carbon due to vegetation dynamics. *Earth's Future*, *6*(10), 1413–1432. https://doi.org/10.1029/2018EF000935

Qiu, C., Zhu, D., Ciais, P., Guenet, B., & Peng, S. (2020). The role of northern peatlands in the global carbon cycle for the 21st century. *Global Ecology and Biogeography*, *29*(5), 956–973. https://doi.org/10.1111/GEB.13081

Ribes, A., Qasmi, S., & Gillett, N. P. (2021). Making climate projections conditional on historical observations. *Science Advances*, *7*(4), 671–693. https://doi.org/10.1126/SCIADV.ABC0671/SUPPL_FILE/ABC0671_SM.PDF

Rogelj, J., Forster, P. M., Kriegler, E., Smith, C. J., & Séférian, R. (2019). Estimating and tracking the remaining carbon budget for stringent climate targets. *Nature*, *571*(7765), 335–342. https://doi.org/10.1038/s41586-019-1368-z

Rowland, L., da Costa, A. C. L., Oliveira, A. A. R., Oliveira, R. S., Bittencourt, P. L., Costa, P. B., et al. (2018). Drought stress and tree size determine stem $CO_2$ efflux in a tropical forest. *New Phytologist*, *218*(4), 1393–1405. https://doi.org/10.1111/NPH.15024

Salazar, L. F., & Nobre, C. A. (2010). Climate change and thresholds of biome shifts in Amazonia. *Geophysical Research Letters*, *37*(17), 17706. https://doi.org/10.1029/2010GL043538

Sanderson, B. M., Wehner, M., & Knutti, R. (2017). Skill and independence weighting for multi-model assessments. *Geoscientific Model Development*, *10*(6), 2379–2395. https://doi.org/10.5194/GMD-10-2379-2017

Schepaschenko, D. G., Mukhortova, L. V., Shvidenko, A. Z., & Vedrova, E. F. (2013). The pool of organic carbon in the soils of Russia. *Eurasian Soil Science*, *46*(2), 107–116. https://doi.org/10.1134/S1064229313020129/METRICS

Schuur, E. A. G., Abbott, B. W., Commane, R., Ernakovich, J., Euskirchen, E., Hugelius, G., et al. (2022). Permafrost and climate change: Carbon cycle feedbacks from the warming Arctic. *Annual Review of Environment and Resources*, *47*(1), 343–371. https://doi.org/10.1146/ANNUREV-ENVIRON-012220-011847

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., et al. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, *58*(4), e2019RG000678. https://doi.org/10.1029/2019RG000678

Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström, A., et al. (2015). Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences*, *12*(3), 653–679. https://doi.org/10.5194/bg-12-653-2015

Spawn, S. A., Sullivan, C. C., Lark, T. J., & Gibbs, H. K. (2020). Harmonized global maps of above and belowground biomass carbon density in the year 2010. *Scientific Data*, *71*(7), 112. https://doi.org/10.1038/s41597-020-0444-4

Stroeve, J., & Notz, D. (2015). Insights on past and future sea-ice evolution from combining observations and models. *Global and Planetary Change*, *135*, 119–132. https://doi.org/10.1016/J.GLOPLACHA.2015.10.011

Todd-Brown, K. E. O., Randerson, J. T., Hopkins, F., Arora, V., Hajima, T., Jones, C., et al. (2014). Changes in soil organic carbon storage predicted by Earth system models during the 21st century. *Biogeosciences*, *11*(8), 2341–2356. https://doi.org/10.5194/bg-11-2341-2014

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., & Knutti, R. (2020). Past warming trend constrains future warming in CMIP6 models. *Science Advances*, *6*(12), 9549–9567. https://doi.org/10.1126/SCIADV.AAZ9549

Van Der Werf, G. R., Randerson, J. T., Giglio, L., Van Leeuwen, T. T., Chen, Y., Rogers, B. M., et al. (2017). Global fire emissions estimates during 1997–2016. *Earth System Science Data*, *9*(2), 697–720. https://doi.org/10.5194/ESSD-9-697-2017

Van Gestel, N., Shi, Z., Van Groenigen, K. J., Osenberg, C. W., Andresen, L. C., Dukes, J. S., et al. (2018). Predicting soil carbon loss with warming. *Nature*, *554*(7693), E4–E5. https://doi.org/10.1038/nature25745

Walker, A. P., Zaehle, S., Medlyn, B. E., De Kauwe, M. G., Asao, S., Hickler, T., et al. (2015). Predicting long-term carbon sequestration in response to $CO_2$ enrichment: How and why do current ecosystem models differ? *Global Biogeochem. Cycles*, *29*(4), 476–495. https://doi.org/10.1002/2014GB004995

Wang, J., Feng, L., Palmer, P. I., Liu, Y., Fang, S., Bösch, H., et al. (2020). Large Chinese land carbon sink estimated from atmospheric carbon dioxide data. *Nature*, *586*(7831), 720–723. https://doi.org/10.1038/s41586-020-2849-9

Wang, J. A., Baccini, A., Farina, M., Randerson, J. T., & Friedl, M. A. (2021). Disturbance suppresses the aboveground carbon sink in North American boreal forests. *Nature Climate Change*, *115*(11), 435–441. https://doi.org/10.1038/s41558-021-01027-4

Wang, Y., Wang, X., Wang, K., Chevallier, F., Zhu, D., Lian, J., et al. (2022). The size of the land carbon sink in China. *Nature*, *603*(7901), E7–E9. https://doi.org/10.1038/s41586-021-04255-y

Weigel, A. P., Knutti, R., Liniger, M. A., & Appenzeller, C. (2010). Risks of model weighting in multimodel climate projections. *Journal of Climate*, *23*(15), 4175–4191. https://doi.org/10.1175/2010JCLI3594.1

Wenzel, S., Cox, P. M., Eyring, V., & Friedlingstein, P. (2014). Emergent constraints on climate-carbon cycle feedbacks in the CMIP5 Earth system models. *Journal of Geophysical Research: Biogeosciences*, *119*(5), 794–807. https://doi.org/10.1002/2013JG002591

Wenzel, S., Cox, P. M., Eyring, V., & Friedlingstein, P. (2016). Projected land photosynthesis constrained by changes in the seasonal cycle of atmospheric $CO_2$. *Nature*, *538*(7626), 499–501. https://doi.org/10.1038/nature19772

Wiltshire, A. J., Burke, E. J., Chadburn, S. E., Jones, C. D., Cox, P. M., Davies-Barnard, T., et al. (2021). JULES-CN: A coupled terrestrial carbon–nitrogen scheme (JULES vn5.1). *Geoscientific Model Development*, *14*(4), 2161–2186. https://doi.org/10.5194/gmd-14-2161-2021

Wiltshire, A. J., von Randow, C., Rosan, T. M., Tejada, G., & Castro, A. A. (2022). Understanding the role of land-use emissions in achieving the Brazilian Nationally Determined Contribution to mitigate climate change. *Climate Resilience and Sustainability*, *1*(1), e31. https://doi.org/10.1002/CLI2.31

Winkler, K., Fuchs, R., Rounsevell, M., & Herold, M. (2021). Global land use changes are four times greater than previously estimated. *Nature Communications*, *12*(1), 2501. https://doi.org/10.1038/s41467-021-22702-2

Zaehle, S., Jones, C. D., Houlton, B., Lamarque, J. F., & Robertson, E. (2015). Nitrogen availability reduces CMIP5 projections of twenty-first-century land carbon uptake. *Journal of Climate*, *28*(6), 2494–2511. https://doi.org/10.1175/JCLI-D-13-00776.1