Key Points:
- Two climate reconstructions commonly used to force terrestrial biosphere models over the industrial era differ in climate
- Climate forcing produces uncertainty in the carbon cycle, and climate uncertainty exceeds model structural uncertainty in many regions
- Models differ in their sensitivity to climate forcing, especially in boreal North America and Eurasia

Supporting Information:
- Supporting Information S1

Correspondence to:
G. B. Bonan,
bonan@ucar.edu

Model Structure and Climate Data Uncertainty in Historical Simulations of the Terrestrial Carbon Cycle (1850–2014)

Gordon B. Bonan[1] , Danica L. Lombardozzi[1] , William R. Wieder[1] , Keith W. Oleson[1] , David M. Lawrence[1] , Forrest M. Hoffman[2,3] , and Nathan Collier[2]

[1]National Center for Atmospheric Research, Boulder, CO, USA, [2]Computational Earth Sciences Group, Oak Ridge National Laboratory, Oak Ridge, TN, USA, [3]Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, Knoxville, TN, USA

Abstract The divergence among Earth system models in the terrestrial carbon cycle has prompted interest in how to reduce uncertainty. Previous studies have identified model structural uncertainty arising from process parameterizations and parameter values. The current study highlights the importance of climate forcing in generating carbon cycle uncertainty. We use simulations in which three models (Community Land Model version 4 (CLM4), CLM4.5, CLM5) with substantially different carbon cycles are forced with two climate reconstructions (CRUNCEPv7, Global Soil Wetness Project 3 version 1 (GSWP3v1)) to examine the contributions of model structure and climate to uncertainty in the carbon cycle over the period 1850–2014. Climate uncertainty for global annual net biome production exceeds one third of total uncertainty (defined as the sum of climate and model structure uncertainty) in the first half of the twentieth century, but declines after the 1950s. Global annual gross primary productivity, net primary productivity, heterotrophic respiration, and vegetation and soil carbon stocks have substantial climate uncertainty (relative to total uncertainty) throughout the simulation period. Climate forcing contributes more than one half of total uncertainty for these carbon cycle fluxes and stocks throughout boreal North America and Eurasia, some midlatitude regions, and in eastern Amazonia and western equatorial Africa during the decade 2000–2009. Comparison with observationally based data sets of the carbon cycle using model benchmarking methods provides insight into strengths and deficiencies among models and climate forcings, but we caution against overreliance on benchmarking to discriminate among models. The conceptualization of uncertainty arising from this study implies embracing multiple feasible model simulations rather than focusing on which model or simulation is best.

1. Introduction

With the advent of Earth system models that simulate physical, chemical, and biological processes relevant for climate, it is possible to move beyond climate prediction to consider Earth system prediction including terrestrial and marine ecosystems and biogeochemical cycles (Bonan & Doney, 2018). The utility of Earth system models for predicting ecological states and fluxes requires an understanding of the sources of uncertainty in model simulations. For climate prediction, the sources of uncertainty can be ascribed to natural variability, model structure, and socioeconomic scenarios, and uncertainty is assessed based on the spread among an ensemble of simulations using analysis of variance (Hawkins & Sutton, 2009; Yip et al., 2011). In this framework, natural variability is the unforced climate variability internal to the system and is determined for a particular model using a multimember ensemble of simulations in which the model is initialized with different states, thereby generating a spread of climate trajectories. Model structural uncertainty arises from differences among models and is assessed using a multimodel ensemble of simulations. Scenario uncertainty arises from different depictions of forcings (e.g., greenhouse gas concentrations, land use, and land cover change). As shown by Hawkins and Sutton (2009) for global mean surface air temperature, natural variability is particularly important at time scales up to a few decades. Model structural uncertainty can also be large, but scenarios are the largest source of uncertainty at multidecadal time scales. At regional scales, natural variability is a larger source of uncertainty.

The same framework of natural variability, model structural uncertainty, and scenario uncertainty informs carbon cycle projections generated by Earth system models. At decadal time scales, natural variability

imparts large unforced variability in the terrestrial carbon cycle that prevents detection of forced change in land carbon uptake at time scales less than a few decades (Lombardozzi et al., 2014). There is considerable spread among CMIP5 Earth system model projections of terrestrial carbon uptake (Jones et al., 2013), and model structural uncertainty, not scenarios, is the dominant contributor to uncertainty in global land carbon uptake throughout the twenty-first century (Hewitt et al., 2016; Lovenduski & Bonan, 2017).

The large divergence among Earth system models precludes a robust assessment of how the terrestrial carbon cycle might change in the future and has prompted much interest in methods to reduce carbon cycle uncertainty. Structural differences among the terrestrial biosphere models used in Earth system models produce divergent carbon cycle simulations (Huntzinger et al., 2013, 2017), and many studies have focused on structural error resulting from incomplete understanding of how to appropriately represent ecosystem processes and their responses to environmental change. For example, studies have examined uncertainty in terms of particular process parameterizations such as ecosystem nitrogen losses (Meyerholt & Zaehle, 2018; Thomas et al., 2013), biological nitrogen fixation (Wieder et al., 2015), soil organic matter decomposition (Koven et al., 2015; Wieder et al., 2018), photosynthetic temperature acclimation (Lombardozzi et al., 2015), and photosynthetic triose phosphate utilization (Lombardozzi et al., 2018). Structural uncertainties are particularly evident in determining simulated responses to elevated concentrations of atmosphere CO_2 (Huntzinger et al., 2017; Wieder et al., 2019; Zaehle et al., 2014). Other studies have examined parameter uncertainty within a particular model (Booth et al., 2012; Boulton et al., 2017; Li et al., 2018; Poulter et al., 2010).

Earth system models differ in their climate projections, and climate also contributes to carbon cycle spread. Nishina et al. (2015) forced several terrestrial biosphere models with climate projections for the twenty-first century obtained from several different climate models run with multiple scenarios. Using analysis of variance, they found a dominant contribution to carbon cycle uncertainty from model structure, although the particular climate forcing also contributed to uncertainty. Two studies using a single terrestrial biosphere model (LPJ-GUESS) point to the importance of climate forcing. Ahlström et al. (2017) forced LPJ-GUESS with the climate output from 15 CMIP5 models for the period 1850–2100 and found that differences in climate contributed substantially to carbon cycle uncertainty. Wu et al. (2017) forced LPJ-GUESS with six alternative historical climate data sets for the twentieth century and found considerable spread in gross primary productivity. Simulations with JULES using several different forcing data sets also found that the particular climate forcing contributes to uncertainty in gross primary production over the period 2001–2010 (Slevin et al., 2017). Conversely, comparison of three terrestrial biosphere models (Community Land Model version 4 (CLM4), JULES, ORCHIDEE) in both uncoupled simulations forced with the CRUNCEP climate reconstruction and in coupled simulations with their host Earth system model found that gross primary production in the coupled simulations was not that different from the uncoupled simulations over the period 1990–2009 (Anav et al., 2015).

In this study, we further examine the role of climate in generating carbon cycle uncertainty. We utilize a suite of carbon cycle simulations in which three terrestrial biosphere models are forced with two commonly used reconstructions of historical climate over the industrial era (1850–2014) to examine the contributions of model structure and climate data to uncertainty in the carbon cycle. Our results show the importance of climate forcing in generating carbon cycle uncertainty. Comparison of the simulations with observationally based data products highlights challenges in using model benchmarking (Collier et al., 2018) to reduce uncertainty. From these results, we develop insights into the sources of carbon cycle uncertainty, how uncertainty can be reduced, and how model improvements should be judged.

2. Methods

2.1. Model Simulations

The three models used for this study are versions of the CLM that differ markedly in their carbon and nitrogen cycles (Wieder et al., 2019). CLM4 has strong nitrogen downregulation of potential (nonnitrogen limited) gross primary production (GPP), low soil carbon stocks, a weak CO_2 fertilization response, and a weak terrestrial carbon sink. CLM4.5 has improved GPP (but still using the CLM4 concept of nitrogen downregulation), vertically resolved soil carbon with much larger soil carbon stocks compared with CLM4, and a larger terrestrial carbon sink. CLM5 allows for flexible vegetation C:N in response to nitrogen availability so

that photosynthetic capacity directly varies with leaf nitrogen (thereby eliminating the CLM4 and CLM4.5 concept of potential GPP), optimally allocates leaf nitrogen through the canopy, and includes the carbon cost of nitrogen uptake. There is an overall trend from CLM4 to CLM4.5 and CLM5 characterized by a poleward shift in carbon uptake, less sensitivity to nitrogen addition, and a larger CO_2 fertilization response (Wieder et al., 2019).

The simulations reported in the current study are land only, in which the models are forced with two different climate reconstructions. The CRUNCEP forcing has been used in the Global Carbon Project (Le Quéré et al., 2016, 2018), the companion study TRENDY (Sitch et al., 2015), and the MsTMIP model intercomparison (Huntzinger et al., 2013); here we use CRUNCEPv7. The GSWP3v1 forcing is used in land-only simulations for the Land Use Model Intercomparison Project (Lawrence et al., 2016) and the Land Surface, Snow and Soil moisture Model Intercomparison Project (van den Hurk et al., 2016).

The six simulations (three models with two climate forcings) span the period 1850–2014. Land use and land cover change, atmospheric CO_2 concentration, and nitrogen deposition are specified from transient data sets. The GSWP3 and CRUNCEP climate forcings only extend back to 1901; the 20-year period 1901–1920 is cycled for the period 1850–1900. Initial conditions for all simulations were generated by spinning-up the models to steady state for 1850, in which land cover, atmospheric CO_2 concentration, and nitrogen deposition are for the year 1850 and climate forcing is for 1901–1920.

2.2. Analyses

The six simulations (three models with two climate forcings) were analyzed to partition total uncertainty into climate and model contributions using fixed-effect analysis of variance for two groups (climate) with $n = 3$ models within each group. The within-group variation arises from model differences, and the between group variation is from climate. This is equivalent to the methodology introduced by Hawkins and Sutton (2009) and extended by Yip et al. (2011) in a formal analysis of variance decomposition. Total uncertainty is the variance across the six-member ensemble, climate uncertainty is the variance of the GSWP3 and CRUNCEP multimodel means, and model uncertainty is the average of the multimodel variances for GSWP3 and CRUNCEP. We use standard deviation (i.e., the square root of variance) to illustrate the absolute magnitude of uncertainty (so as to have the same units as the mean), but follow the convention of using variance to ascertain the relative contribution of models and climate forcing to total uncertainty.

The simulations were assessed using the ILAMB benchmarking package (Collier et al., 2018). Scores generated by ILAMB assess the carbon cycle in terms of vegetation biomass, burned area, gross primary production, leaf area index, net biome production, net ecosystem production, ecosystem respiration, and soil carbon, and higher scores indicate greater fidelity to the benchmarking data sets. Collier et al. (2018) provide details on the scoring methodology and data sets. Here we used a subset of the ILAMB benchmarks. Global net biome production (NBP) for 1959–2014 is from the Global Carbon Project (Le Quéré et al., 2016) and is taken as the balance between fossil fuel emissions (FF), atmospheric growth rate (ATM), and ocean uptake (OCE) so that NBP = FF – ATM – OCE. As in Le Quéré et al. (2016), the estimated uncertainty (Pg C/year) for each year is $(0.05 \times FF)^2 + (0.1)^2 + (0.5)^2$, in which FF uncertainty is proportional to the emissions and ATM (0.1 Pg C/year) and OCE (0.5 Pg C/year) uncertainties are constant. Hoffman et al. (2014) provide estimates for the historical period 1850–2010. GPP, ecosystem respiration (ER), and net ecosystem production (NEP = GPP – ER) are from the GBAF data set (Jung et al., 2010). Vegetation carbon (VEGC) is evaluated for the tropics (Saatchi et al., 2011) and also using two global data sets (GEOCARBON, GlobalCarbon) as described by Collier et al. (2018). Soil carbon (SOMC) is evaluated using the Harmonized World Soil Database as processed by Todd-Brown et al. (2013) and also for permafrost regions using the Northern Circumpolar Soil Carbon Database (Hugelius et al., 2013). The simulations were compared with CMIP5 models to determine if the spread among our simulations is comparable to that of the CMIP5 multimodel ensemble (Anav et al., 2013).

3. Results

3.1. Comparison With CMIP5 Models

The six CLM simulations capture the spread in the carbon cycle compared with CMIP5 models. The CMIP5 models have a large spread in NBP over the period 1959–2005 and span the estimated historical uncertainty;
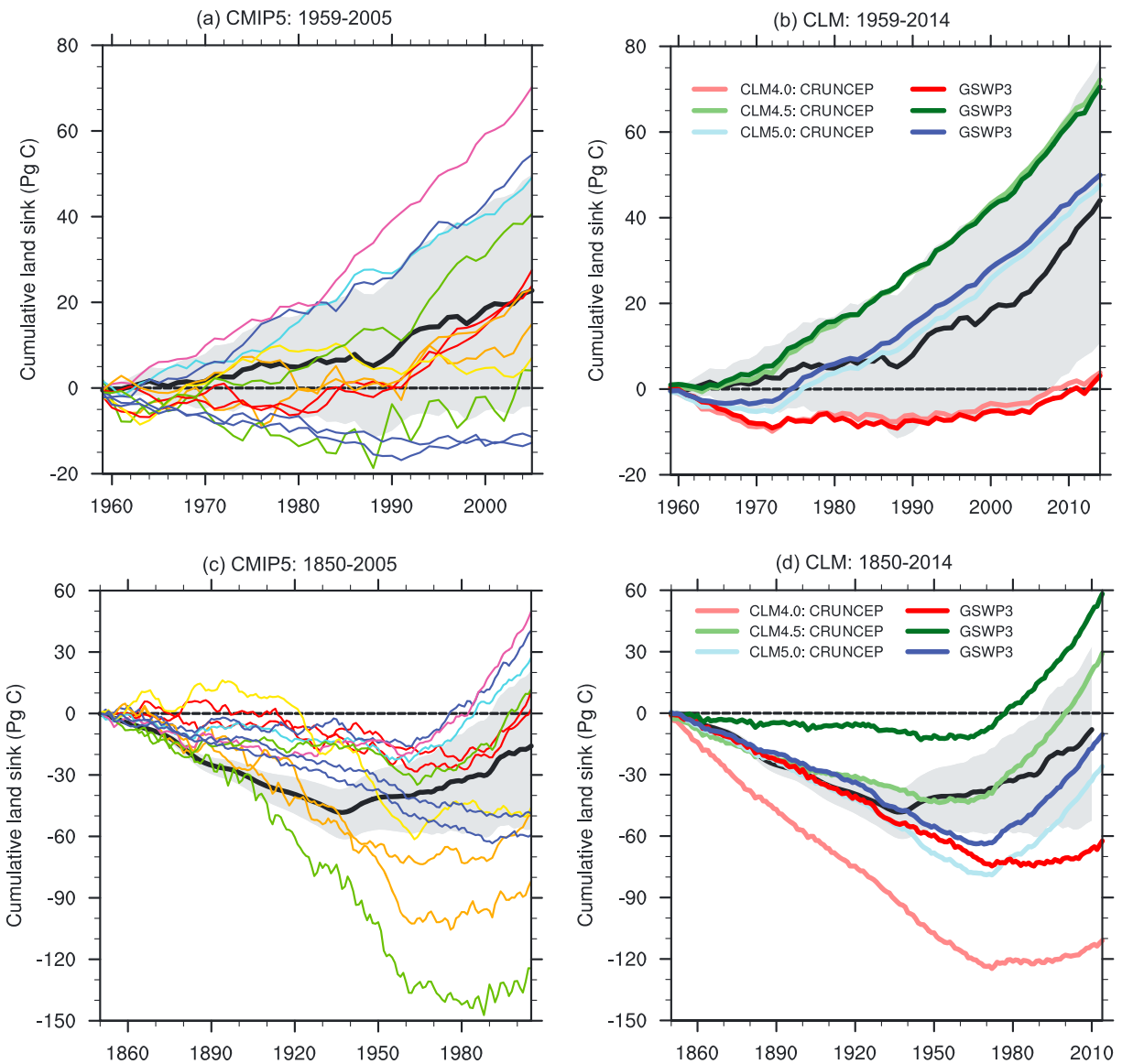
Figure 1. Cumulative land carbon sink (NBP) for CMIP5 models (1850–2005) and CLM (1850–2014). CMIP5 data are for the 12 models used in Lovenduski and Bonan (2017). CLM simulations are for the three model versions (CLM4, CLM4.5, CLM5) with CRUNCEP and GSWP3 forcing. Shown are comparisons with the Global Carbon Project (Le Quéré et al., 2016) for (a) CMIP5 (1959–2005) and (b) CLM (1959–2014). The thick black line shows the historical estimate, and the shading denotes the estimated uncertainty from the Global Carbon Project. CMIP5 models are color coded based on fidelity to the Global Carbon Project, as described by Lovenduski and Bonan (2017). Also shown are comparisons with Hoffman et al. (2014) for (c) CMIP5 and (d) CLM over the full simulation period. The historical carbon budget is for the period 1850–2010. Uncertainty estimates are from Hoffman et al. (2014).

the CLM simulations have a similar spread (Figures 1a and 1b). Few of the CMIP5 models fall within the observationally based uncertainty estimates for the period 1850–2005; the CLM simulations have a similarly large spread (Figures 1c and 1d). GPP has a comparable range as the CMIP5 models, excluding six models with anomalously large GPP (Figure S1a in the supporting information). VEGC and SOMC have a narrower range compared with the CMIP5 models, but some of those models have anomalously large pools (Figure S1b in the supporting information).

3.2. Model Benchmarking

The ILAMB benchmarking shows a progressive improvement in the carbon cycle from CLM4 to CLM4.5 and CLM5, but also points to the opportunities and challenges in discriminating among models
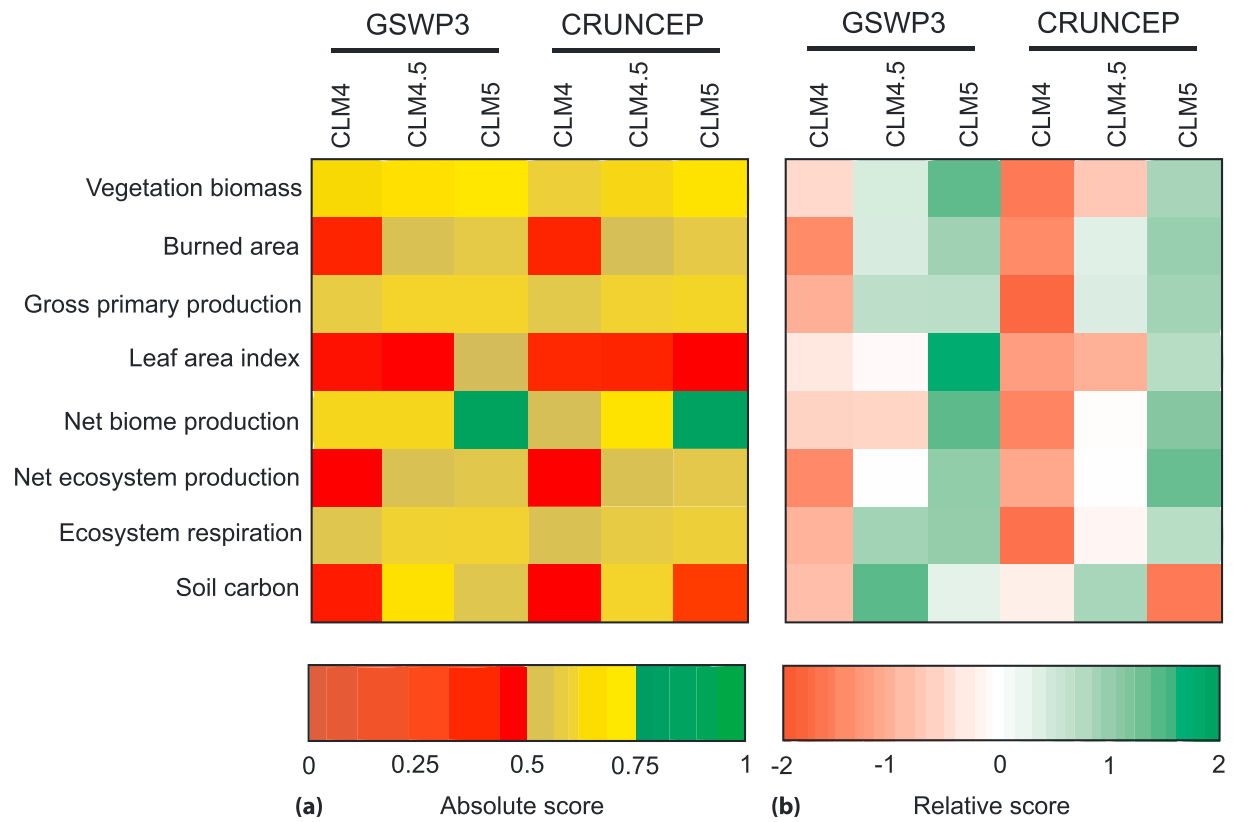
Figure 2. ILAMB carbon cycle summary scores for CLM4, CLM4.5, and CLM5 forced with GSWP3 and CRUNCEP. Shown are (a) the absolute score and (b) the relative score for various pools and fluxes.

(Figure 2). CLM4 has the lowest benchmarking scores, CLM4.5 has higher scores, and CLM5 has the highest scores. SOMC is a prominent exception and has a lower score for CLM5 compared with CLM4.5. GSWP3 results in scores for CLM4.5 and CLM5 that are generally comparable to or larger than CRUNCEP. However, comparison of the absolute scores for CLM4.5 and CLM5 show that many of the differences in scores are small (less than ±0.05 for VEGC, burned area, GPP, NEP, ER, and overall summary; see Table S1 in the supporting information). Larger improvements are seen from CLM4.5 to CLM5 in leaf area index and NBP but with degradation in SOMC. Summaries of particular strengths and weakness of model versions and forcing data for NBP and component fluxes are presented below.

Over the period 1959–2014, CLM5 closely matches the Global Carbon Project estimate for global NBP and is improved compared with CLM4 and CLM4.5 (Figure 1b). The difference between GSWP3 and CRUNCEP is small compared with the differences among models, but all six simulations are within (or close to) the observationally based uncertainty. Over the full historical period 1850–2014, four of six simulations are within the estimated uncertainty prior to midtwentieth century (Figure 1d). Differences among the four simulations only emerge later in the twentieth century. CLM4.5 (CRUNCEP) and CLM5 (GSWP3) both are within the uncertainty estimate for the late twentieth century. CRUNCEP reduces carbon uptake compared with GSWP3, more so for CLM4 and CLM4.5 than for CLM5.

The strength and geographic location of the terrestrial carbon sink differs among models and climate forcings (Figure 3). For CLM4, the southern equatorial tropics is a small carbon sink during the latter half of the twentieth century when forced with GSWP3, but is a source of carbon with CRUNCEP. With both forcings, CLM4.5 has a strong tropical sink throughout the twentieth century (weaker with CRUNCEP) and a small northern sink (stronger with CRUNCEP). CLM5 has a smaller tropical sink compared with CLM4.5 (stronger for CRUNCEP than for GSWP3) and has a northern high-latitude carbon sink similar to CLM4.5 (stronger for GSWP3 than for CRUNCEP).
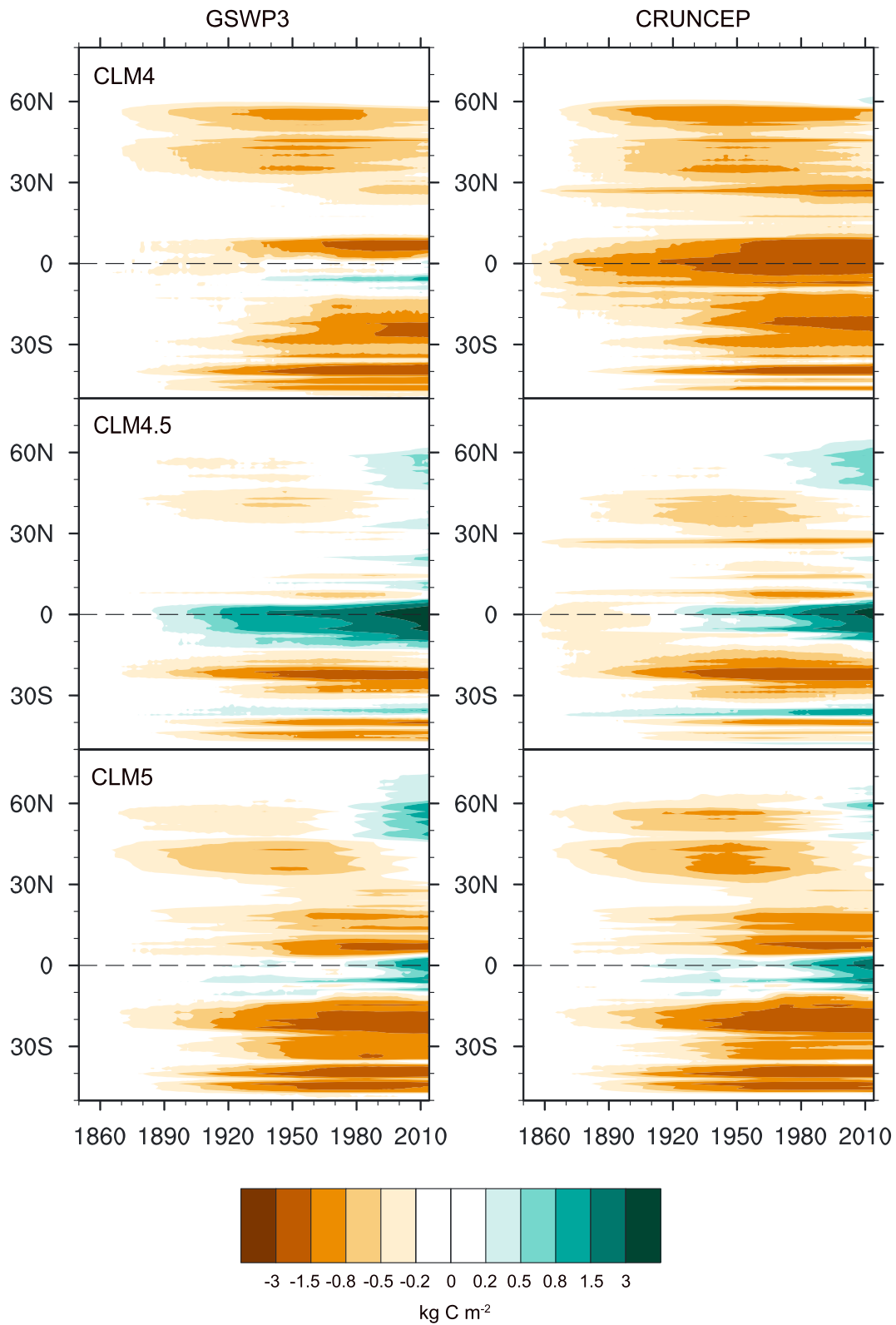
Figure 3. Zonal mean plots of the cumulative land carbon sink over the period 1850–2014 for (top row) CLM4, (middle row) CLM4.5, and (bottom row) CLM5 forced with (left column) GSWP3 and (right column) CRUNCEP. Positive values show a carbon sink, and negative values show a carbon source.
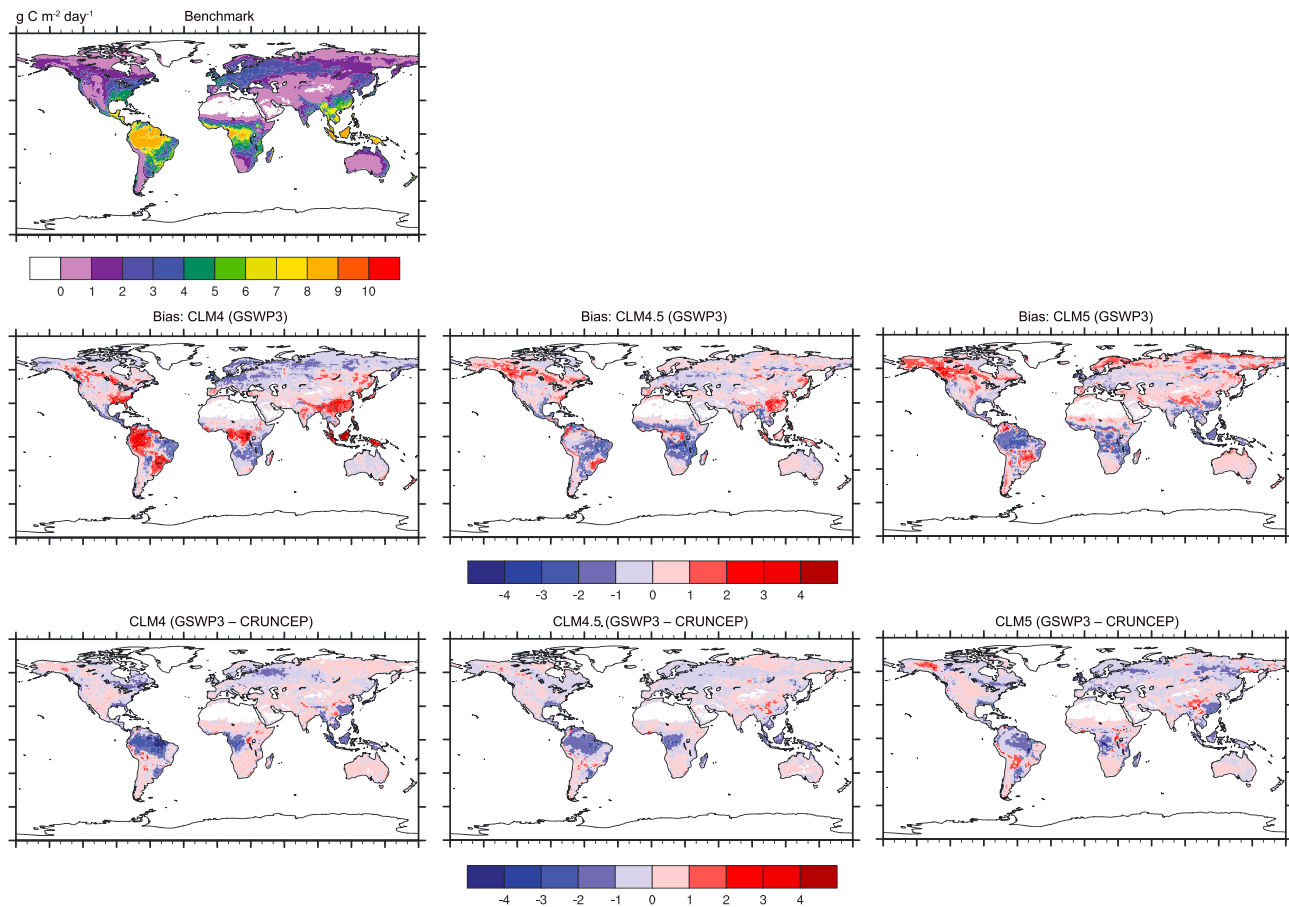
Figure 4. Annual gross primary production (GPP) for the period 1982–2008. Shown are the (top row) ILAMB benchmark data set; (middle row) model biases for CLM4, CLM4.5, and CLM5 with the GSWP3 forcing; and (bottom row) the difference with CRUNCEP for each model. The benchmark data set is from Collier et al. (2018), as described in section 2.2. Bias is calculated as the difference between the model and the benchmark. Positive values (red) show where the model exceeds the benchmark. Climate forcing differences for each model are expressed as GSWP3–CRUNCEP. Positive values (red) show where GSWP3 values are greater than CRUNCEP.

GSWP3 reduces annual GPP compared with CRUNCEP in many regions of the world (Figure 4). CLM4 is biased high in the tropics, much of the Northern Hemisphere midlatitudes, and in boreal North America; GSWP3 reduces these biases compared with CRUNCEP. CLM5 has a large negative bias in the tropics and a positive bias throughout much of boreal North America and Eurasia. By reducing GPP, GSWP3 increases the tropical negative bias and reduces the boreal positive bias compared with CRUNCEP. CLM4.5 is similar to CLM5, but with smaller boreal GPP. A prominent exception occurs in Alaska and NW Canada, where GPP increases with GSWP3 (especially for CLM5). ILAMB benchmarking shows that CLM4.5 and CLM5 are improved compared with CLM4, especially in terms of bias and spatial distribution scores (Table S2 in the supporting information). Differences between CLM4.5 and CLM5 and between GSWP3 and CRUNCEP are less evident.

Differences among models and climate forcings in annual net primary production (NPP) and heterotrophic respiration (HR) are similar to GPP. Annual NPP in the tropics and Northern Hemisphere midlatitudes decreases from CLM4 to CLM5, but northern high-latitude NPP increases (Figure S2 in the supporting information). GSWP3 decreases NPP compared with CRUNCEP in many regions. HR follows similar patterns as NPP, and GSWP3 decreases HR compared with CRUNCEP in many regions (Figure S3 in the supporting information). However, GSWP3 increases NPP and HR in Alaska and NW Canada, most prominently for CLM5. The sensitivity of NPP and HR to climate forcing is robust across models in the tropics, but is larger in northern high latitudes for CLM5.
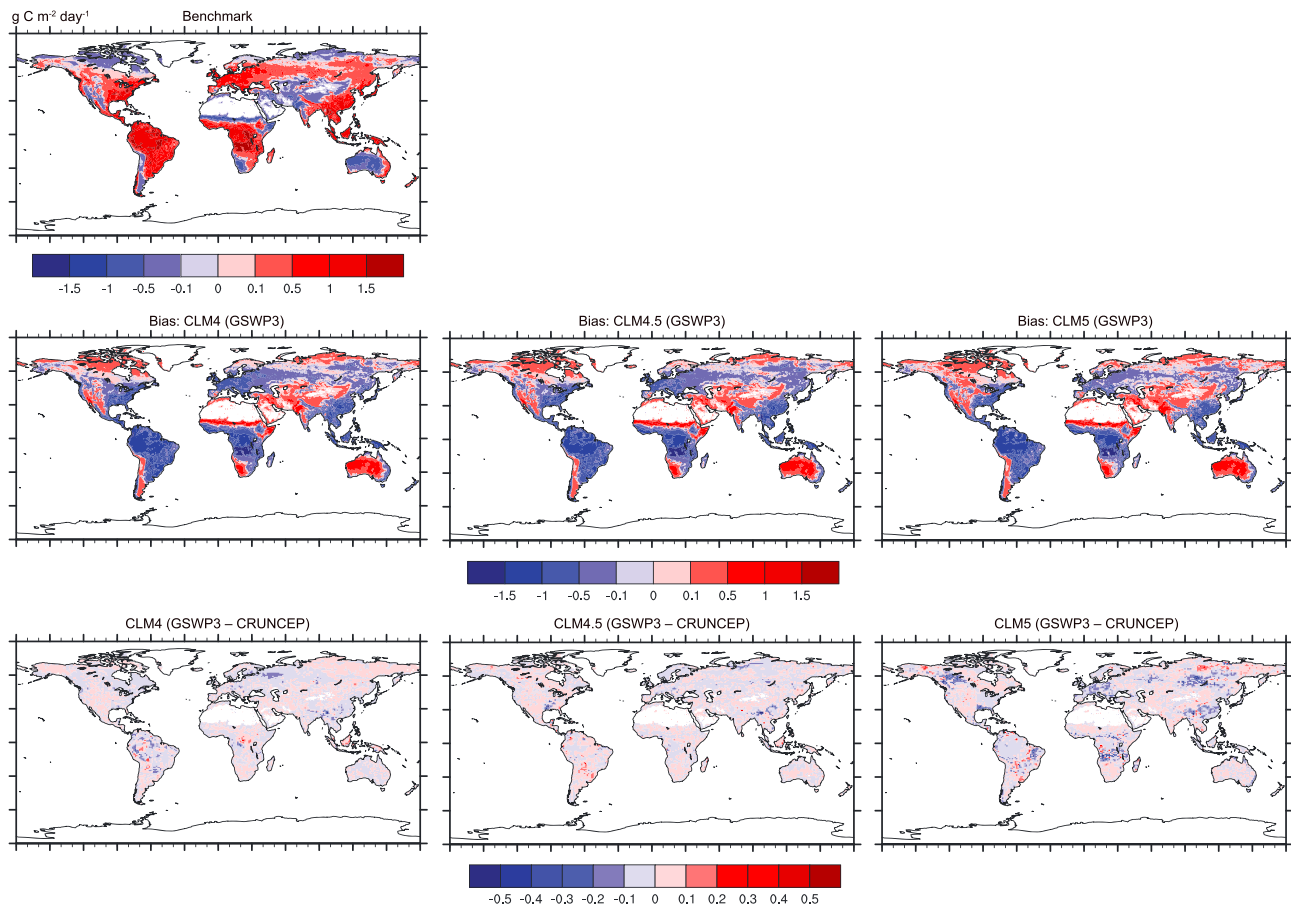
Figure 5. Same as in Figure 4 but for annual net ecosystem production (NEP).

All three models are biased low in NEP throughout much of the world (Figure 5). With GSWP3, CLM5 simulates larger Northern Hemisphere middle- and high-latitude NEP compared with CLM4 and CLM4.5, seen in a reduced negative bias in Europe and SE United States. CLM5 NEP is generally smaller for GSWP3 than for CRUNCEP. CLM4 and CLM4.5 are less sensitive to climate forcing. CLM5 performs better than CLM4 and CLM4.5 in all ILAMB metrics (Table S3 in the supporting information). The primary difference in benchmarking scores is among models. There is less difference due to climate forcing, but GSWP3 produces a poorer seasonal cycle across all models.

There is a shift across models from CLM4 to CLM5 in terms of less tropical VEGC and more boreal VEGC (Figure S4 in the supporting information). The reduction in tropical biomass reduces a positive bias in the models (not shown). CLM5 has a large positive biomass bias in boreal North America and much of Eurasia (not shown); CLM4.5 has a smaller positive bias. Similar to GPP and NPP, GSWP3 decreases VEGC compared with CRUNCEP (except in Alaska and NW Canada), more so for CLM5. CLM5 has higher ILAMB scores than CLM4 and CLM4.5 when compared with the tropics and GlobalCarbon data sets, and GSWP3 improves scores (Table S4 in the supporting information). Model and climate differences are smaller when compared with GEOCARBON.

CLM5 simulates more SOMC in northern high latitudes compared with CLM4 and CLM4.5 when forced with GSWP3 (Figure S5 in the supporting information). GSWP3 decreases SOMC compared with CRUNCEP, most prominently for CLM5 in the arctic and less so for CLM4 and CLM4.5. CLM4.5 has the highest ILAMB scores in permafrost regions, and GSWP3 produces a slightly better simulation (NCSCD; Table S5 in the supporting information). For CLM5, however, GSWP3 is much better than CRUNCEP. CLM4 significantly underestimates SOMC in the arctic compared with NCSCD, resulting in a low overall
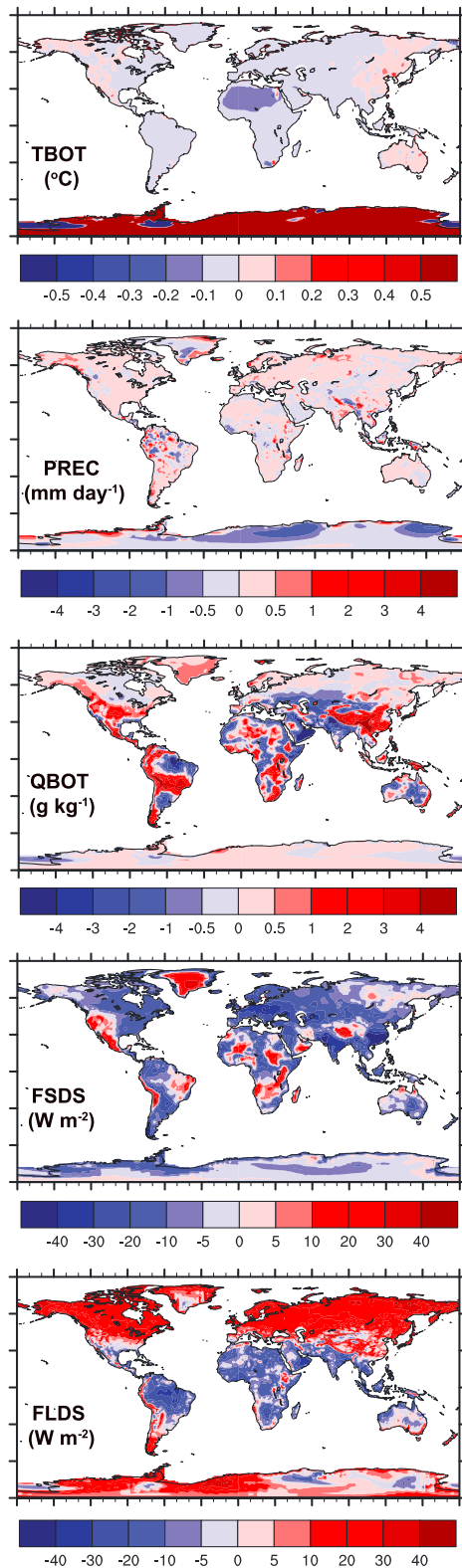
Figure 6. Climate differences between GSWP3 and CRUNCEP for the period 2000–2009. Shown are annual mean surface air temperature (TBOT), precipitation (PREC), surface specific humidity (QBOT), total incoming solar radiation (FSDS), and incoming longwave radiation (FLDS). Positive values (red) show where GSWP3 values are greater than CRUNCEP.

ILAMB score. Similar results are seen globally, and GSWP3 improves scores for CLM4.5 and CLM5 (HWSD; Table S5 in the supporting information).

3.3. Climate Differences

GSWP3 and CRUNCEP have noticeable differences in climate, as shown in Figure 6 for the period 2000–2009. Temperature differs by less than 0.1 °C in most regions of the world. Precipitation differences are prominent in tropical South America, SE Asia, and Alaska and NW Canada. Specific humidity, downwelling solar radiation, and downwelling longwave radiation show more widespread differences between the two data sets. GSWP3 has smaller specific humidity in eastern Amazonia compared with CRUNCEP. Differences are particularly evident during boreal summer, when GSWP3 has smaller specific humidity across much of the Northern Hemisphere middle and high latitudes except for central North America and SE Asia (not shown). Solar radiation is less in GSWP3 over much of the Northern Hemisphere but is larger in eastern Amazonia. Longwave radiation is larger over much of the Northern Hemisphere, but is smaller in the tropics and subtropics. For CLM5, the forcing differences result in soils that are several degrees warmer across the arctic for GSWP3 compared with CRUNCEP (Figure S6 in the supporting information).

3.4. Uncertainty Analysis

Total uncertainty in simulated global NBP is large relative to the mean and varies considerably from year to year (Figure 7). Climate uncertainty is largest prior to the 1950s (contributing up to 80% or more of total uncertainty in some years), but individual years thereafter can also have large climate uncertainty (20% or more of total uncertainty in the 1980s–2010s). Model uncertainty (as an absolute amount) can also vary considerably between years. NBP has largest total uncertainty in the tropics (where the mean is largest), but uncertainty is also comparably large in SE Asia and in boreal North America and Eurasia (Figure 8). Climate uncertainty as a fraction of total uncertainty can be locally large, and NW Canada, portions of the Russian arctic, and eastern Amazonia stand out regionally.

GSWP3 decreases simulated global GPP, NPP, HR, NEP, VEGC, and SOMC compared with CRUNCEP, and climate forcing contributes considerably to total uncertainty in these components of the carbon cycle (Figure 9). The total uncertainty for GPP is 10–12 Pg C/year and declines slightly over time with less model uncertainty (Figure 9a). Climate uncertainty is constant (about 6 Pg C/year) and increases to about 40–50% of total uncertainty in the 1990s and 2000s. Total uncertainty for NPP and HR is about 3–4 Pg C/year and arises mostly from climate forcing (Figures 9b and 9c). Total uncertainty for NEP is about 1–2 Pg C/year during the latter half of the twentieth century and is large relative to the mean (Figure 9d). Climate uncertainty is constant and is generally a small contribution to total uncertainty after the 1950s, but prior to that time individual years can have a large contribution of climate forcing to total uncertainty (similar to NBP). In addition to decreasing VEGC compared with CRUNCEP (Figure 9e), GSWP3 reduces the spread among models (note the similarity of CLM4 and CLM5 with GSWP3 but not with CRUNCEP). Total uncertainty is 60–80 Pg C and arises mostly from climate forcing. Total uncertainty for SOMC is about 460 Pg C and arises
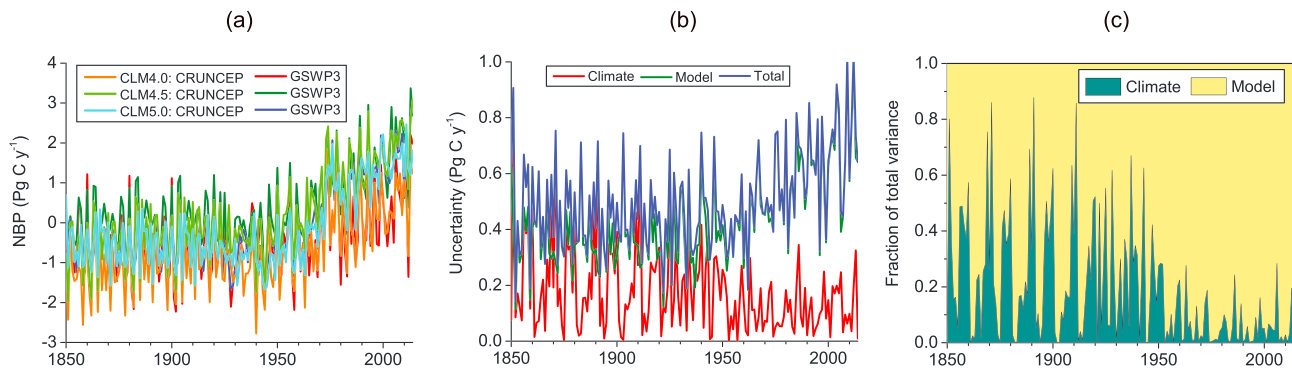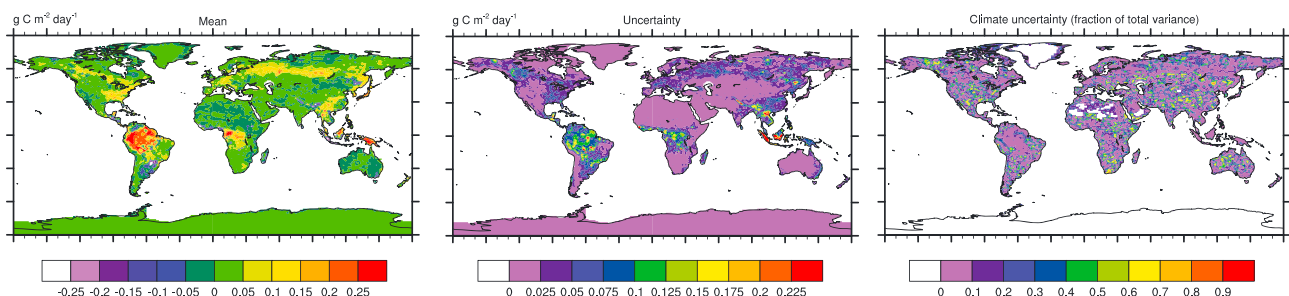
Figure 7. Uncertainty analysis for global net biome production (NBP) for the period 1850–2014. Shown are (a) global annual NBP for the six simulations; (b) climate, model, and total uncertainty; and (c) the partitioning of total uncertainty between model and climate forcing (as a fraction of total variance). Uncertainty is quantified using analysis of variance as described in section 2.2.

mostly from models (Figure 9f). Climate uncertainty is constant and is about 20% of total uncertainty.

There is a distinct geographic pattern to carbon cycle uncertainty in the CLM simulations. The largest GPP total uncertainty occurs in the tropics (where the mean is largest) and also SE United States and SE Asia (Figure 10a). However, Alaska, Scandinavia, and portions of the Siberian arctic also have large uncertainty. The contribution of climate to total uncertainty can be regionally large, most prominently in boreal North America, boreal Eurasia, eastern Amazonia, and western equatorial Africa where climate uncertainty exceeds 60% of total uncertainty. Uncertainty in NPP (Figure 10b) and HR (Figure 10c) have a similar geographic pattern as GPP, but with a larger contribution of climate uncertainty in SE United States and SE Asia. Total uncertainty in NEP is large in central North America, Europe, and SE Asia (Figure 11a). The contribution of climate to total uncertainty is similar to that for NBP. Total uncertainty in VEGC is large in forest ecosystems in the tropics, SE Asia, and boreal North America and Eurasia, where the mean is large (Figure 11b). Climate uncertainty generally follows a similar spatial pattern as NPP and is regionally large, most prominently in parts of boreal North America, boreal Eurasia, SE United States, Amazonia, and western equatorial Africa where climate uncertainty exceeds 60% of total uncertainty. SOMC total uncertainty is largest in northern high latitudes where the mean is largest (Figure 11c). Climate uncertainty generally follows a similar spatial pattern as NPP, but is more widespread in the tropics and northern latitudes, and exceeds 60% of total uncertainty in the tropics, SE United States, SE Asia, and parts of boreal North America and Eurasia.

4. Discussion

The quest to reduce uncertainty in terrestrial carbon cycle simulations commonly focuses on biogeochemical process parameterizations and parameter estimation. Some studies have begun to examine the importance of climate forcing in generating carbon cycle uncertainty (Ahlström et al., 2017; Anav et al., 2015; Nishina et al., 2015; Slevin et al., 2017; Wu et al., 2017), and the results of the present study provide evidence that

Figure 8. Uncertainty analysis for net biome production (NBP) for the period 2000–2009. Shown are the (left column) multimodel mean, (middle column) total uncertainty, and (right column) climate uncertainty expressed as a fraction of total uncertainty.
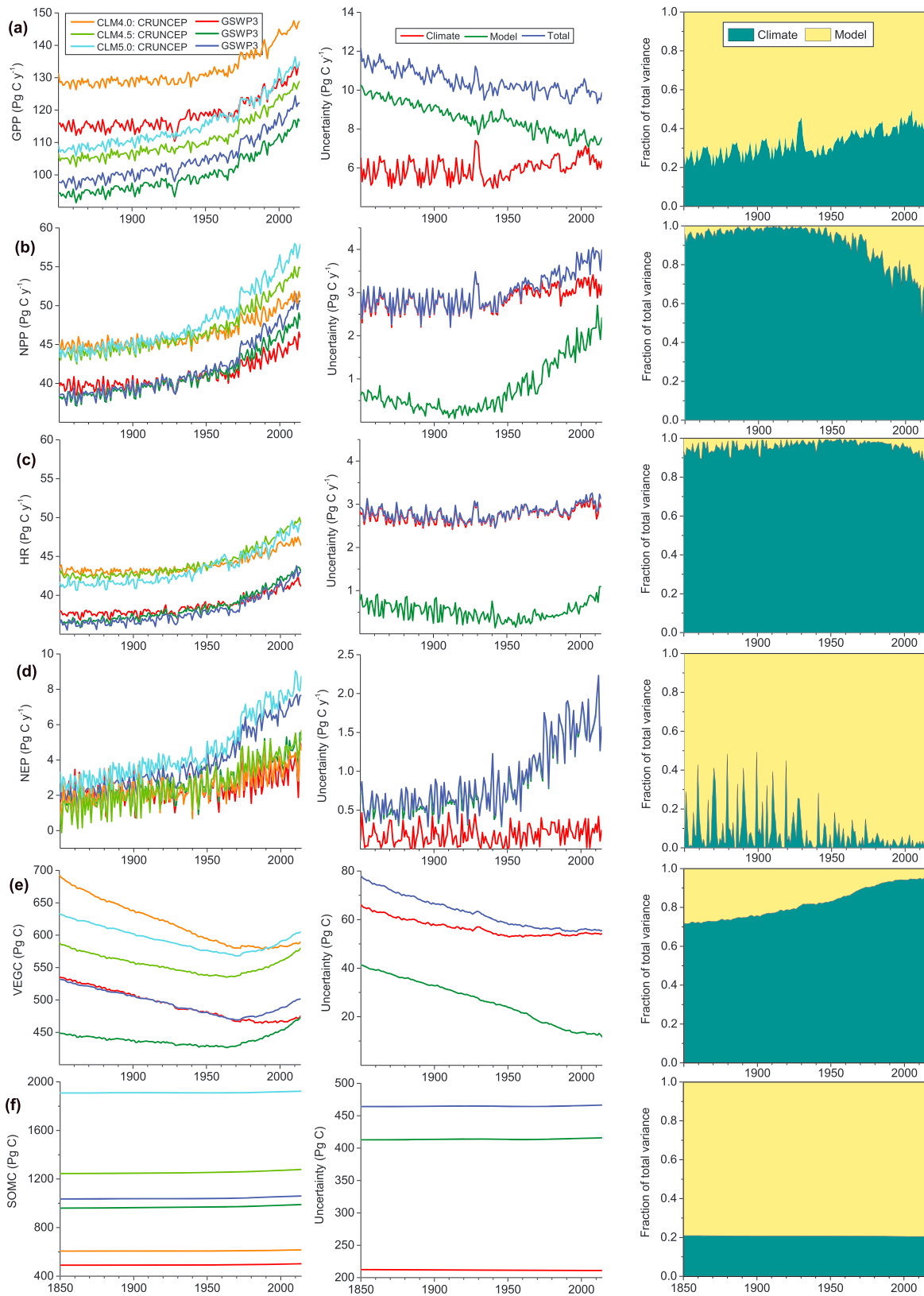
Figure 9. Carbon cycle uncertainty for the period 1850–2014 for (a) gross primary production (GPP), (b) net primary production (NPP), (c) heterotrophic respiration (HR), (d) net ecosystem production (NEP), (e) vegetation carbon (VEGC), and (f) soil carbon to a depth of 1 m (SOMC). Shown are (left column) the six simulations; (middle column) climate, model, and total uncertainty; and (right column) the partitioning of total uncertainty between model and climate forcing.
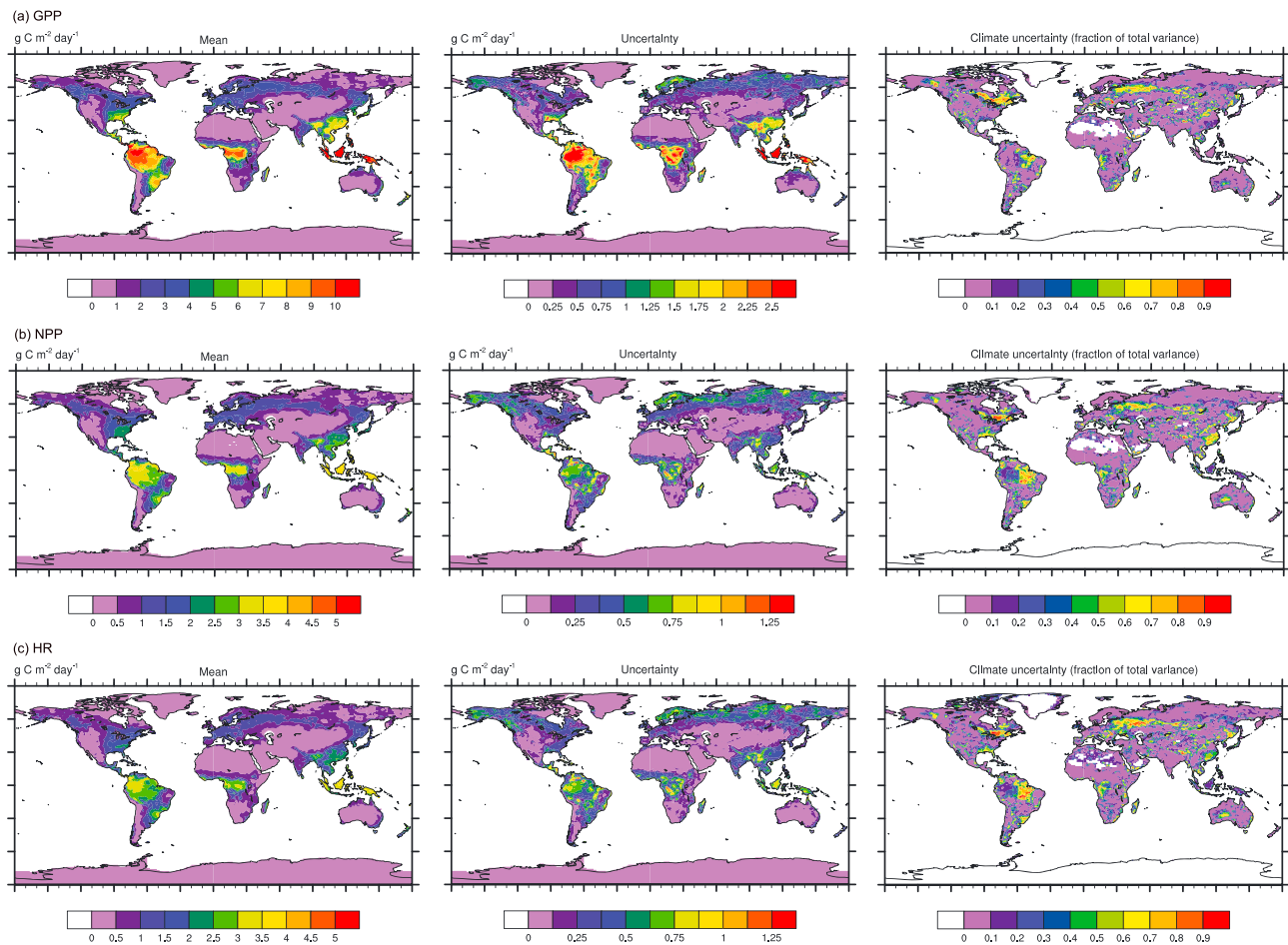
Figure 10. Carbon cycle uncertainty for the period 2000–2009. Shown are the (left column) multimodel mean, (middle column) total uncertainty, and (right column) climate uncertainty expressed as a fraction of the total uncertainty for (a) gross primary production (GPP), (b) net primary production (NPP), and (c) heterotrophic respiration (HR).

uncertainty in climate forcing can exceed uncertainty arising from model structure, even when terrestrial biosphere models are forced with best-estimate climate reconstructions over the industrial era. Our six-member ensemble of three versions of CLM (CLM4, CLM4.5, CLM5) forced with two climate reconstructions (GSPW3, CRUNCEP) captures the spread of CMIP5 Earth system model simulations (Figures 1 and S1 in the supporting information), suggesting that our results are generalizable across a wider range of models. Our results emphasize that although model uncertainty is large, climate uncertainty also makes an important contribution to the total carbon cycle uncertainty. Within this ensemble of models, the ILAMB benchmarking tool provides valuable insights into particular strengths and deficiencies in model versions and climate forcings (Figure 2), but we also caution against an overreliance on these scores in evaluating models.

4.1. Climate Forcing Versus Model Uncertainty

Our simulations reveal significant model uncertainty, but also highlight that climate forcing is a large source of uncertainty in the carbon cycle even when constrained by best climate reconstructions over the industrial era. Climate uncertainty is large for global NBP prior to the 1950s, but can also be large for particular years in the 1980s–2010s (Figure 7). Climate uncertainty as a fraction of total uncertainty can be locally large, especially in NW Canada, portions of the Russian arctic, and eastern Amazonia (Figure 8). Other terms in the global carbon cycle (GPP, NPP, HR, VEGC, SOMC) also have substantial climate uncertainty (Figure 9). Climate forcing contributes more than 60% of total uncertainty for these terms throughout boreal North America and Eurasia, in eastern Amazonia and western equatorial Africa, and also in SE United States

AGU
100
ADVANCING EARTH
AND SPACE SCIENCE

Global Biogeochemical Cycles

10.1029/2019GB006175
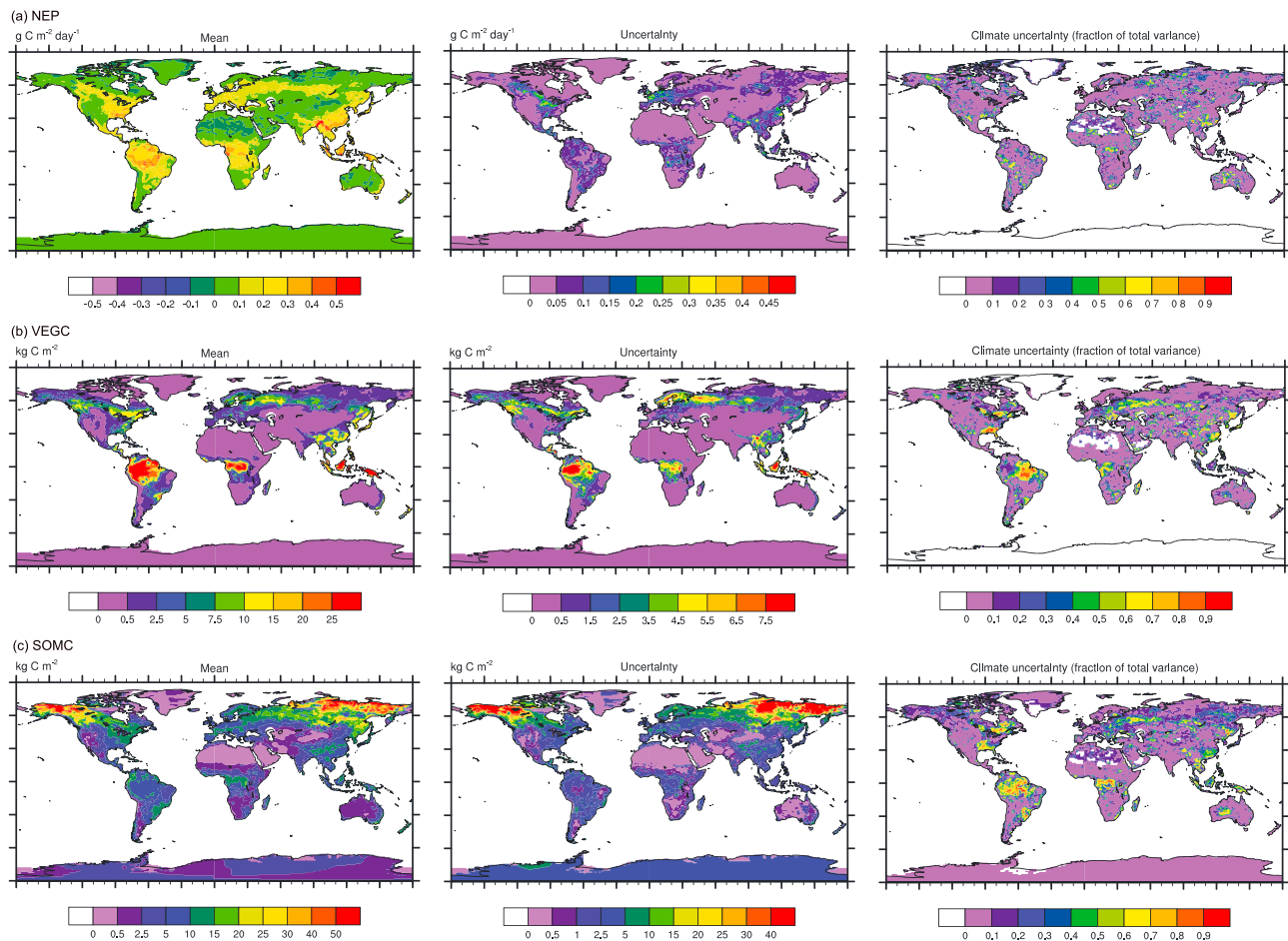
Figure 11. Same as in Figure 10 but for (a) net ecosystem production (NEP), (b) vegetation carbon (VEGC), and (c) soil organic matter carbon to a depth of 1 m (SOMC).

and SE Asia (Figures 10 and 11). Other climate reconstructions are also available, but not all cover the same time period. Wu et al. (2017), for example, used six climate reconstructions in their analysis of GPP with LPJ-GUESS. CLM5 simulations using the WFDEI/WATCH reconstruction (Weedon et al., 2011) produce a different carbon cycle compared with GSWP3 and CRUNCEP (see the data availability for data access). Slevin et al. (2017) used several reconstructions, including variants of WFEDI/WATCH. These additional climate reconstructions would likely alter our estimated climate uncertainty.

The GSWP3 and CRUNCEP data sets differ in forcing variables that affect the carbon cycle, especially specific humidity, solar radiation, and longwave radiation (Figure 6). A main difference between climate forcings across all three models, seen both globally (Figure 9) and regionally (Figures 4, 5, S2, and S3), is a less vigorous carbon cycle (reduced GPP, NPP, HR, and NEP) for GSWP3 compared with CRUNCEP. This also manifests in terms of reduced VEGC and SOMC (Figures 9, S4, and S5). An important exception is Alaska and NW Canada, where GPP, NPP, HR, NEP, and VEGC (but not SOMC) are larger for GSWP3. However, NBP shows the opposite response, and CRUNCEP reduces global NBP compared with GSWP3 (Figure 1d). There is also a large difference between models and climate forcings in terms of the strength of tropical versus boreal carbon sinks (Figure 3). One noteworthy point is that CLM5 is more sensitive to climate forcing in boreal North America and Eurasia than is CLM4.5 (see GPP, NPP, HR, and VEGC), and SOMC is particularly sensitive to climate forcing in permafrost regions (seen also in the large CLM5 soil temperature differences between GSWP3 and CRUNCEP; Figure S6 in the supporting information), suggesting that models likely vary in the sensitivity of the carbon cycle to climate. This could relate to different mean states rather than parameterization differences among models, although the exact cause is unclear.

4.2. Model Evaluation

The different carbon cycle simulations among models and climate forcings prompt the question of whether observational data sets can be used to discriminate among the simulations. The generational development of CLM shows increasing fidelity to many observationally based data sets, and ILAMB benchmarking quantifies the improvements in model performance across CLM4, CLM4.5, and CLM5 and between climate forcings (Figure 2 and Table S1 in the supporting information). In general, CLM5 has higher ILAMB scores compared with CLM4 and CLM4.5, and GSWP3 scores for CLM4.5 and CLM5 are comparable to or higher than CRUNCEP. SOMC is an exception and has a lower benchmarking score for CLM5 than for CLM4.5. While the tendency among modelers will be to use ILAMB summary scores (Figure 2b) to demonstrate model improvement or to rank models in their performance, Collier et al. (2018) caution against interpreting too much into the absolute ILAMB scores. Indeed, our comparisons among models and climate forcings suggest that overreliance on the overall summary scoring may provide misguided confidence in model performance at the exclusion of a lower scoring model. Careful analysis of the component metrics that determine the overall scoring is necessary, as is an assessment of spatial biases. Observationally based data products provide a necessary constraint on models, but inference about which version of CLM and which climate forcing is best depends on region and the particular variable of interest. Moreover, the ILAMB analysis assigns relative weights to each data set to qualitatively account for uncertainty in the data sets (Collier et al., 2018), but better methods to quantify data set uncertainty and utilize the uncertainty in the benchmarking scores are required and are a priority for ILAMB.

Rather than ranking models or weighting models based on benchmarking scores, an alternative use of ILAMB benchmarking is to ascertain which models provide equally plausible results (and conversely which models are likely implausible), particularly in light of observational uncertainty. In this context, CLM4 is a different class of model than CLM4.5 and CLM5, with consistently lowest benchmarking scores for various aspects of the carbon cycle (Figure 2 and Table S1 in the supporting information). Perturbation experiments that examine the response of the three models to elevated CO_2 and nitrogen addition also point to deficiencies in CLM4 compared with CLM4.5 and CLM5 (Wieder et al., 2019). However, improvements between CLM4.5 and CLM5 are more nuanced. ILAMB, for example, assigns CLM5 much higher scores for NBP compared with CLM4.5 (GSWP3, 0.87 versus 0.63; CRUNCEP, 0.84 versus 0.69; Table S1 in the supporting information), but analysis of the complete time series over the historical period (1850–2014) shows that both CLM4.5 (CRUNCEP) and CLM5 (GSWP3) are within the estimated uncertainty (Figure 1d).

The benchmarking analyses of GPP and VEGC also highlight the need for caution when evaluating models. The overall ILAMB scores for GPP have small differences between CLM4.5 and CLM5 and between GSWP3 and CRUNCEP for these two models (less than 0.01; Table S1 in the supporting information), but this masks important regional differences that are evident in ILAMB only upon careful examination of the spatial diagnostics it generates. CLM5 exacerbates a low GPP bias in the tropics and a high GPP bias in boreal regions compared with CLM4.5 (Figure 4). GSPW3 reduces GPP, enhancing the low bias in the tropics but improving the high boreal bias (a prominent exception is Alaska and NW Canada). Despite these GPP biases, CLM5 has improved NEP scores in the ILAMB benchmarking (Figure 2 and Table S1 in the supporting information). Biases in GPP propagate to affect vegetation biomass. CLM5 simulates more VEGC than CLM4.5 in boreal forests but less in the tropics (Figure S4 in the supporting information). ILAMB benchmarking shows improved VEGC for CLM5 compared with CLM4.5 (Figure 2 and Table S1 in the supporting information), but this varies regionally and relates to differences in GPP between models. CLM5 reduces a positive tropical biomass bias compared with CLM4.5, and GSWP3 reduces the tropical bias compared with CRUNCEP, so that the ILAMB score substantially improves in the tropics (Table S4 in the supporting information), but this occurs in part because CLM5 has low GPP in the tropics. Conversely, higher GPP with CLM5 accentuates a positive biomass bias in boreal North America and Eurasia and reduces a negative bias in Siberia. Comparison of ILAMB scores for the GEOCARBON and GlobalCarbon data sets gives an indication of the uncertainty in the observational data sets and leads to different inferences about model performance; differences among models are less evident for GEOCARBON (Table S4 in the supporting information).

Soil carbon is a key metric of the global carbon cycle, but one in which CLM5 is prominently deficient in terms of carbon stocks compared with CLM4.5 (Figure 2 and Table S1 in the supporting information). The primary difference between models is in northern high latitudes, where CLM5 simulates more SOMC

compared with CLM4.5 (Figure S5 in the supporting information). In this region, CLM5 is strikingly more sensitive to climate forcing than is CLM4.5, with substantially less SOMC for GSWP3 compared with CRUNCEP. This relates to soils that are several degrees warmer with GSWP3 compared with CRUNCEP (Figure S6 in the supporting information). As a consequence, CLM5 forced with CRUNCEP has large positive biases in SOMC throughout a broad region encompassing Alaska and western Canada and also in Siberia; GSWP3 reduces these biases. This is seen in the ILAMB scoring for CLM5, which is markedly better with GSWP3 compared with CRUNCEP (see NCSCD in Table S5 in the supporting information). In support of this, CLM5 forced with CRUNCEP likely produces soils that are too cold. Moreover, the ILAMB benchmarking focuses on SOMC stocks. A better metric is apparent soil carbon turnover times.

The notion of how to define whether a model is good likely needs to expand beyond benchmarking analysis. Prentice et al. (2015), for example, argued that models must balance reliability (the model must produce correct output), robustness (the results are not overly sensitive to specific assumptions or parameters), and realism (the necessary processes are represented in sufficient detail). Greater process complexity may increase realism but reduce reliability and robustness. Bradford et al. (2016) made just such a point with regard to soil carbon. CLM5, with greater process richness, can be said to be more realistic than CLM4.5, but that authenticity to process detail may come with greater uncertainty. For example, CLM4.5 uses prescribed values of the photosynthetic parameters V_{cmax} and J_{max} obtained from trait databases. This captures prevailing macroscale environmental constraints, but not seasonal or long-term changes in these parameters. CLM5 simulates prognostic V_{cmax} and J_{max} based on dynamic leaf nitrogen concentration and canopy optimization of nitrogen and thereby allows them to vary seasonally and annually. Rather than being a prescribed constraint on the model, V_{cmax} and J_{max} are an emergent test of the model. Yet other factors not included in the model such as phosphorus availability may be needed to simulate realistic photo-synthetic parameters (Norby et al., 2017).

4.3. Carbon Cycle Prediction

The desire to reduce carbon cycle uncertainty has focused attention on the large spread among models and the need for observational data sets with which to constrain models. However, as Lovenduski and Bonan (2017) showed, it is difficult to reduce uncertainty in CMIP5 carbon cycle projections, even when the models are weighted based on fidelity to best estimates of NBP over the historical era. The present study further shows the difficulty in reducing carbon cycle uncertainty, specifically, in this case, uncertainty arising from differences in reconstructed climate over the industrial era. Many of the differences in climate forcing are relatively small (a few grams per kilogram for specific humidity or a few tens of watt per square meter for solar and longwave radiation; Figure 6), and it remains to be seen how similar the forcings must be to reduce uncertainty. Simulations with LPJ-GUESS suggest the importance of solar radiation and precipitation for carbon cycle uncertainty in the tropics (Wu et al., 2017). The carbon cycle simulated by JULES is also sensi-tive to reconstructed air temperature, precipitation, and solar radiation (Slevin et al., 2017). Differences in solar radiation, in particular, affect simulated gross primary production by changing the balance between Rubisco-limited and light-limited CO_2 assimilation. Climate uncertainty is likely to be even larger in Earth system models that simulate divergent climates. Indeed, simulations with LPJ-GUESS over the period 1850–2100 suggest that climate differences among Earth system models contribute substantially to carbon cycle uncertainty (Ahlström et al., 2017).

The quest to reduce carbon cycle uncertainty is not simple; we do not yet have a complete understanding of whence uncertainty arises, so how, then, can we reduce the uncertainty? Moreover, as Lovenduski and Bonan (2017) suggested, there may be a limit below which uncertainty cannot be reduced. Model structure and parameter estimation contribute to uncertainty. The present study shows that climate forcing also con-tributes to such a limit. If this premise of irreducible uncertainty is correct, there is not a single exact solution that models should strive to attain (thereby by definition reducing uncertainty) but rather a family of plau-sible outcomes. The ILAMB benchmarking of the six simulations reported herein further supports such a conclusion, and shows that some simulations may be easily categorized as deficient, but others may not be so easily distinguished. This conceptualization of uncertainty implies embracing multiple feasible model simulations rather than determining which single model is best or focusing on a single outcome. Beven (2006) addressed this point in the context of hydrologic models with his equifinality thesis, but the same con-cept applies to terrestrial biosphere models and the carbon cycle. Yet as Beven (2006) cautioned, it is easy to

"interpret the resulting ambiguity of predictions as a failure (or at least an undermining) of the science." The carbon cycle community would benefit from a comprehensive discussion of uncertainty, what makes a model good, and the goals of carbon cycle prediction.

Data and Model Availability

Model data are archived and publicly available at the UCAR/NCAR Climate Data Gateway, https://doi.org/10.5065/d6154fwh. CLM5.0 is publicly available through the Community Terrestrial System Model (CTSM) git repository (https://github.com/ESCOMP/ctsm).

Author Contribution

G.B.B. wrote the manuscript with contributions from all authors. D.L.L. and K.W.O. performed the simulations in collaboration with D.M.L. D.L.L. and W.R.W. assisted with data analyses. K.W.O., F.M.H., and N.C. assisted with ILAMB benchmarking analyses.

References

Ahlström, A., Schurgers, G., & Smith, B. (2017). The large influence of climate model bias on terrestrial carbon cycle simulations. *Environmental Research Letters*, *12*(1), 014004. https://doi.org/10.1088/1748-9326/12/1/014004

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., et al. (2015). Spatiotemporal patterns of terrestrial gross primary production: A review. *Reviews of Geophysics*, *53*, 785–818. https://doi.org/10.1002/2015RG000483

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

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1-2), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

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

Booth, B. B. B., Jones, C. D., Collins, M., Totterdell, I. J., Cox, P. M., Sitch, S., et al. (2012). High sensitivity of future global warming to land carbon cycle processes. *Environmental Research Letters*, *7*(2), 024002. https://doi.org/10.1088/1748-9326/7/2/024002

Boulton, C. A., Booth, B. B. B., & Good, P. (2017). Exploring uncertainty of Amazon dieback in a perturbed parameter Earth system ensemble. *Global Change Biology*, *23*(12), 5032–5044. https://doi.org/10.1111/gcb.13733

Bradford, M. A., Wieder, W. R., Bonan, G. B., Fierer, N., Raymond, P. A., & Crowther, T. W. (2016). Managing uncertainty in soil carbon feedbacks to climate change. *Nature Climate Change*, *6*(8), 751–758. https://doi.org/10.1038/10.1038/NCLIMATE3071

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

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1108. https://doi.org/10.1175/2009BAMS2607.1

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

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*, 141–162. https://doi.org/10.1002/2013JG002381

Hugelius, G., Bockheim, J. G., Camill, P., Elberling, B., Grosse, G., Harden, J. W., et al. (2013). A new data set for estimating organic carbon storage to 3 m depth in soils of the northern circumpolar permafrost region. *Earth System Science Data*, *5*(2), 393–402. https://doi.org/10.5194/essd-5-393-2013

Huntzinger, D. N., Michalak, A. M., Schwalm, C., Ciais, P., King, A. W., Fang, Y., et al. (2017). Uncertainty in the response of terrestrial carbon sink to environmental drivers undermines carbon-climate feedback predictions. *Scientific Reports*, *7*(1), 4765. https://doi.org/10.1038/s41598-017-03818-2

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*(6), 2121–2133. https://doi.org/10.5194/gmd-6-2121-2013

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

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., et al. (2010). Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, *467*(7318), 951–954. https://doi.org/10.1038/nature09396

Koven, C. D., Lawrence, D. M., & Riley, W. J. (2015). Permafrost carbon-climate feedback is sensitive to deep soil carbon decomposability but not deep soil nitrogen dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 3752–3757. https://doi.org/10.1073/pnas.1415123112

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*(9), 2973–2998. https://doi.org/10.5194/gmd-9-2973-2016

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

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A. C., et al. (2018). Global carbon budget 2017. *Earth System Science Data*, *10*(1), 405–448. https://doi.org/10.5194/essd-10-405-2018

Li, J., Duan, Q., Wang, Y. P., Gong, W., Gan, Y., & Wang, C. (2018). Parameter optimization for carbon and water fluxes in two global land surface models based on surrogate modelling. *International Journal of Climatology*, *38* (Suppl.1, e1016–e1031. https://doi.org/10.1002/joc.5428

Lombardozzi, D., Bonan, G. B., & Nychka, D. W. (2014). The emerging anthropogenic signal in land–atmosphere carbon-cycle coupling. *Nature Climate Change*, *4*(9), 796–800. https://doi.org/10.1038/NCLIMATE2323

Lombardozzi, D. L., Bonan, G. B., Smith, N. G., Dukes, J. S., & Fisher, R. A. (2015). Temperature acclimation of photosynthesis and respiration: A key uncertainty in the carbon cycle-climate feedback. *Geophysical Research Letters*, *42*, 8624–8631. https://doi.org/10.1002/2015GL065934

Lombardozzi, D. L., Smith, N. G., Cheng, S. J., Dukes, J. S., Sharkey, T. D., Rogers, A., et al. (2018). Triose phosphate limitation in photosynthesis models reduces leaf photosynthesis and global terrestrial carbon storage. *Environmental Research Letters*, *13*(7), 074025. https://doi.org/10.1088/1748-9326/aacf68

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

Meyerholt, J., & Zaehle, S. (2018). Controls of terrestrial ecosystem nitrogen loss on simulated productivity responses to elevated CO_2. *Biogeosciences*, *15*(18), 5677–5698. https://doi.org/10.5194/bg-15-5677-2018

Nishina, K., Ito, A., Falloon, P., Friend, A. D., Beerling, D. J., Ciais, P., et al. (2015). Decomposing uncertainties in the future terrestrial carbon budget associated with emission scenarios, climate projections, and ecosystem simulations using the ISI-MIP results. *Earth System Dynamics*, *6*(2), 435–445. https://doi.org/10.5194/esd-6-435-2015

Norby, R. J., Gu, L., Haworth, I. C., Jensen, A. M., Turner, B. L., Walker, A. P., et al. (2017). Informing models through empirical relationships between foliar phosphorus, nitrogen and photosynthesis across diverse woody species in tropical forests of Panama. *New Phytologist*, *215*(4), 1425–1437. https://doi.org/10.1111/nph.14319

Poulter, B., Hattermann, F., Hawkins, E., Zaehle, S., Sitch, S., Restrepo-Coupe, N., et al. (2010). Robust dynamics of Amazon dieback to climate change with perturbed ecosystem model parameters. *Global Change Biology*, *16*, 2476–2495. https://doi.org/10.1111/j.1365-2486.2009.02157.x

Prentice, I. C., Liang, X., Medlyn, B. E., & Wang, Y.-P. (2015). Reliable, robust and realistic: The three Rs of next-generation land-surface modelling. *Atmospheric Chemistry and Physics*, *15*(10), 5987–6005. https://doi.org/10.5194/acp-15-5987-2015

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., et al. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(24), 9899–9904. https://doi.org/10.1073/pnas.1019576108

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

Slevin, D., Tett, S. F. B., Exbrayat, J.-F., Bloom, A. A., & Williams, M. (2017). Global evaluation of gross primary productivity in the JULES land surface model v3.4.1. *Geoscientific Model Development*, *10*(7), 2651–2670. https://doi.org/10.5194/gmd-10-2651-2017

Thomas, R. Q., Bonan, G. B., & Goodale, C. L. (2013). Insights into mechanisms governing forest carbon response to nitrogen deposition: A model-data comparison using observed responses to nitrogen addition. *Biogeosciences*, *10*(6), 3869–3887. https://doi.org/10.5194/bg-10-3869-2013

Todd-Brown, K. E. O., Randerson, J. T., Post, W. M., Hoffman, F. M., Tarnocai, C., Schuur, E. A. G., & Allison, S. D. (2013). Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences*, *10*(3), 1717–1736. https://doi.org/10.5194/bg-10-1717-2013

van den Hurk, B., Kim, H., Krinner, G., Seneviratne, S. I., Derksen, C., Oki, T., et al. (2016). LS3MIP (v1.0) contribution to CMIP6: The Land Surface, Snow and Soil moisture Model Intercomparison Project—Aims, setup and expected outcome. *Geoscientific Model Development*, *9*(8), 2809–2832. https://doi.org/10.5194/gmd-9-2809-2016

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., et al. (2011). Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *Journal of Hydrometeorology*, *12*(5), 823–848. https://doi.org/10.1175/2011JHM1369.1

Wieder, W. R., Hartman, M. D., Sulman, B. N., Wang, Y.-P., Koven, C. D., & Bonan, G. B. (2018). Carbon cycle confidence and uncertainty: Exploring variation among soil biogeochemical models. *Global Change Biology*, *24*(4), 1563–1579. https://doi.org/10.1111/gcb.13979

Wieder, W. R., Cleveland, C. C., Lawrence, D. M., & Bonan, G. B. (2015). Effects of model structural uncertainty on carbon cycle projections: Biological nitrogen fixation as a case study. *Environmental Research Letters*, *10*(4), 044016. https://doi.org/10.1088/1748-9326/10/4/044016

Wieder, W. R., Lawrence, D. M., Fisher, R. A., Bonan, G. B., Cheng, S. J., Goodale, C. L., et al. (2019). Beyond static benchmarking: Using experimental manipulations to evaluate land model assumptions. *Global Biogeochemical Cycles*. https://doi.org/10.1029/2018GB006141

Wu, Z., Ahlström, A., Smith, B., Ardö, J., Eklundh, L., Fensholt, R., & Lehsten, V. (2017). Climate data induced uncertainty in model-based estimations of terrestrial primary productivity. *Environmental Research Letters*, *12*(6), 064013. https://doi.org/10.1088/1748-9326/aa6fd8

Yip, S., Ferro, C. A. T., Stephenson, D. B., & Hawkins, E. (2011). A simple, coherent framework for partitioning uncertainty in climate predictions. *Journal of Climate*, *24*(17), 4634–4643. https://doi.org/10.1175/2011JCLI4085.1

Zaehle, S., Medlyn, B. E., De Kauwe, M. G., Walker, A. P., Dietze, M. C., Hickler, T., et al. (2014). Evaluation of 11 terrestrial carbon–nitrogen cycle models against observations from two temperate Free-Air CO_2 Enrichment studies. *New Phytologist*, *202*(3), 803–822. https://doi.org/10.1111/nph.12697