Global Biogeochemical Cycles

RESEARCH ARTICLE

10.1029/2021GB007019

Key Points:

- We use a single-model large ensemble to estimate uncertainties from internal climate variability in the global carbon budget
- The land sink accounts for most internal climate uncertainty which may permit 9–18 Pg C yr^{-1} in allowable emissions by 2050 (for 3°C warming)

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

T. F. Loughran,
t.loughran@lmu.de

Citation:

Loughran, T. F., Boysen, L., Bastos, A., Hartung, K., Havermann, F., Li, H., et al. (2021). Past and future climate variability uncertainties in the global carbon budget using the MPI Grand Ensemble. *Global Biogeochemical Cycles*, *35*, e2021GB007019. https://doi.org/10.1029/2021GB007019

Received 29 MAR 2021
Accepted 28 JUL 2021

Past and Future Climate Variability Uncertainties in the Global Carbon Budget Using the MPI Grand Ensemble

T. F. Loughran[1], L. Boysen[2], A. Bastos[1,3], K. Hartung[1,4], F. Havermann[1], H. Li[2], J. E. M. S. Nabel[2], W. A. Obermeier[1], and J. Pongratz[1,2]

[1]Department of Geography, Ludwig Maximilian University, Munich, Germany, [2]Max Planck Institute for Meteorology, Hamburg, Germany, [3]Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, Jena, Germany, [4]Now at Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

Abstract Quantifying the anthropogenic fluxes of CO_2 is important to understand the evolution of carbon sink capacities, on which the required strength of our mitigation efforts directly depends. For the historical period, the global carbon budget (GCB) can be compiled from observations and model simulations as is done annually in the Global Carbon Project's (GCP) carbon budgets. However, the historical budget only considers a single realization of the Earth system and cannot account for internal climate variability. Understanding the distribution of internal climate variability is critical for predicting the future carbon budget terms and uncertainties. We present here a decomposition of the GCB for the historical period and the RCP4.5 scenario using single-model large ensemble simulations from the Max Planck Institute Grand Ensemble (MPI-GE) to capture internal variability. We calculate uncertainty ranges for the natural sinks and anthropogenic emissions that arise from internal climate variability, and by using this distribution, we investigate the likelihood of historical fluxes with respect to plausible climate states. Our results show these likelihoods have substantial fluctuations due to internal variability, which are partially related to El Niño-Southern Oscillation (ENSO). We find that the largest internal variability in the MPI-GE stems from the natural land sink and its increasing carbon stocks over time. The allowable fossil fuel emissions consistent with 3 C warming may be between 9 and 18 Pg C yr^{-1}. The MPI-GE is generally consistent with GCP's global budgets with the notable exception of land-use change emissions in recent decades, highlighting that human action is inconsistent with climate mitigation goals.

1. Introduction

The global carbon budget (GCB) of CO_2 can be decomposed into anthropogenic emissions and natural sinks. Anthropogenic emissions are mostly due to fossil fuel burning and fossil carbonates (E_{FF}), but also from land-use induced land cover change and land management ("land-use change emissions" in the following, E_{LUC}). The emitted CO_2 is then distributed into three natural sinks: it is either assimilated by the land surface via ecosystem productivity (S_{LAND}), absorbed by the ocean via diffusion and photosynthesis of marine organisms (S_{OCEAN}) or accumulated in the atmosphere (atmospheric growth: G_{ATM}) leading to increased atmospheric CO_2 concentrations (Friedlingstein et al., 2020; Le Quéré et al., 2013).

One of the key goals of the Global Carbon Project (GCP) is to evaluate anthropogenic perturbations on the global carbon cycle and to understand the response of the natural carbon sinks to increasing fossil emissions and land-use changes (e.g., Friedlingstein et al., 2020; Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al., 2018; Le Quéré, Andrew, Friedlingstein, Sitch, Pongratz, et al., 2018). These GCBs, conducted almost every year since 2007 (Canadell et al., 2007), provide an important understanding of the efficiency and potential saturation of the natural sinks. This, in turn, is essential knowledge for predicting the future sink capacities and, therefore, the required strength for future climate mitigation targets and of "allowable" emissions under given climate targets. A comprehensive understanding of uncertainties in these budgets is essential for guiding policy and decision-making.

The components of the GCP carbon budgets are associated with large uncertainties, which are based on a combination of observation and model uncertainties. Fossil emissions are based on energy and fuel consumption data whereby the uncertainties lie in the fuel consumption, fuel carbon content, and combustion efficiency (Andres et al., 2012). The E_{LUC} estimate is based on three bookkeeping models, in which estimates of land-use transitions are combined with observation-based carbon densities to track terrestrial emissions and removals according to empirical temporal response curves for each ecosystem (Hansis et al., 2015; Houghton & Nassikas, 2017). The corresponding estimates for E_{LUC} uncertainty have low confidence and are based on expert knowledge, which considers the bookkeeping models and the range of the 17 global dynamical vegetation models (DGVMs; Friedlingstein et al., 2020). The ocean sink estimate is based on the standard deviation of nine global ocean biogeochemical models and their consistency with observed CO_2 partial pressure-based flux estimates. The terrestrial sink in earlier budgets was estimated as a residual from all other terms or based on DGVMs from the 2019 budget onwards. The estimates of both S_{LAND} and S_{OCEAN} are evaluated to have medium confidence (Friedlingstein et al., 2020). When estimating the land sink with DGVMs, the G_{ATM} cannot be matched, leading to a "budget imbalance" term of ~0.4 Pg C yr^{-1}. While atmospheric measurements of CO_2 concentration are relatively more accurate, there are substantial interannual variations (IAV) driven by natural climate variability (Conway et al., 1994; Dlugokencky & Tans, 2018).

From such GCBs, it is possible to quantify the future emissions to stay within a given trajectory of climate change (Millar et al., 2017; Rogelj et al., 2016). However, estimating these "allowable emissions" from historical budgets actually requires considering an additional source of uncertainty: the internal variability of the climate system. The uncertainties in the GCP budgets are related to observational and model uncertainties while uncertainties associated with internal climate variability are not directly addressed.

Much of the IAV in CO_2 concentration and its impacts on the regional (Zhu et al., 2018) and global carbon sinks (Ballantyne et al., 2012; Bastos et al., 2013) is driven by internal variability in the climate system. Internal variability arises from stochastic processes and feedbacks in the coupled ocean-atmosphere system (e.g., El Niño-Southern Oscillation [ENSO]) and is difficult to predict due to high sensitivity to initial conditions and the chaotic evolution of the Earth system (Deser et al., 2012). Traditionally, internal variability in weather and climate forecasts is accounted for by performing ensemble forecasting, that is, running multiple simulations of the same (or several) models started from perturbed initial conditions, in order to estimate the distribution of future climate states (Deser et al., 2012).

The importance of considering the full range of potential climate states due to internal climate variability is particularly pertinent to future estimates of the carbon budget, where the exact climate state (and consequently the strength of the natural sinks) in a given year is unknown. Using only one realization may not robustly capture these future states. Furthermore, we cannot assume that the variance of the natural CO_2 fluxes is stationary under increasing atmospheric CO_2. It is not possible to estimate the range of plausible carbon budget fluxes due to internal climate variability using only one instance of historical observations or observationally forced model simulations. Using ensemble simulations will allow for a more robust calculation of future trends in the mean and variability of the carbon budget terms (e.g., Kay et al., 2015).

Since the historical observation-based carbon budget uncertainty only considers one realization of internal climate variability, the influence of internal climate variability on each budget term is unknown. Therefore, we ask the following research questions:

- How large is the uncertainty from internal climate variability in the global carbon budget (GCB) terms and how does it compare to the variability of the latest GCB (GCB2020) values?
- How likely were the historical carbon fluxes with respect to the distribution of possible fluxes from internal climate variability and what drove those anomalies?
- How will the carbon budget components and their internal variability change in the future (e.g., under RCP4.5)?

In this study, we estimate uncertainties associated with internal climate variability for each component of the carbon budget using a large ensemble of single-model simulations from the Max Planck Institute Grand Ensemble project (MPI-GE; Maher et al., 2019). We compare the results of the estimates for internal climate variability uncertainties to the uncertainties of the recent GCB2020 (Friedlingstein et al., 2020). Furthermore, we discuss the suitability and possible limitations of using a large ensemble of simulations for better understanding variability and uncertainties associated with E_{LUC} and S_{LAND} and how many ensemble members are required to answer these questions.

10.1029/2021GB007019

Table 1
Experiment Simulations

	LUC	No LUC
Historical (1850–2005)	mpige-LUC-hist cbal-LUC-hist	cbal-noLUC-hist
RCP4.5 (2006–2099)	mpige-LUC-rcp4.5 cbal-LUC-rcp4.5	cbal-noLUC-rcp4.5

Note. Each experiment has 100 ensemble members. The MPI-GE simulations have been labeled with the prefix "mpige," while the CBALONE simulations are labeled as "cbal." The scenarios are labeled with the suffix "hist" for the historical scenario and "rcp4.5" for the future scenario. Both scenarios for CBALONE are simulated with land-use change (labeled with LUC) and without land-use change using 1850 land-use throughout the simulation (labeled with noLUC). There are only 97 ensemble members for the CBALONE RCP4.5 scenario because a few MPI-GE output files required by CBALONE contained erroneous data.

Abbreviations: CBALONE, Carbon Balance ALONE; MPI-GE, Max Planck Institute Grand Ensemble.

2. Methods

2.1. Overview of Models and Simulations

The methods used to generate the ensemble of climate realizations as part of the MPI-GE project are fully described in Maher et al. (2019). Therefore, we only give a summary here. The MPI-GE is a single-model large ensemble project that uses the Max Planck Institute Earth System Model (MPI-ESM; for a full description see Giorgetta et al., 2013) version 1.1. The MPI-ESM is composed of an atmospheric component provided by ECHAM 6.3.01p3 (Stevens et al., 2013) run at T63L47 resolution (~1.8° and 47 vertical layers), an ocean component provided by MPIOM 1.6.1p1 (Marsland et al., 2003) run at GR15L40 resolution (~1.5°), the ocean biogeochemistry model HAMOCC5.2 (Ilyina et al., 2013), and the land component JSBACH3 (Goll et al., 2015; Reick et al., 2013). Hundred ensemble members are generated by branched initialization (every ~6–24 years) from a subsample of years from a pre-industrial control (piControl) simulation. The piControl simulation, as well as the subsequent historical and future simulations, follow the protocol of concentration-driven Earth system model runs of the Coupled Model Intercomparison Projects (CMIP), in this case specifically CMIP5 (Taylor et al., 2012).

The JSBACH3 component simulates transitions in land cover types with respect to both natural vegetation dynamics and land-use changes. However, we utilize a smaller standalone sub-component of JSBACH3 called Carbon Balance ALONE (CBALONE) to differentiate the emissions due to land-use change from the remaining net land sink (as is done in, e.g., Roeckner et al., 2010). As in all Earth system model simulations that perform historical or scenario simulations, anthropogenic and natural effects occur concurrently, that is, the simulations only provide the net land-atmosphere exchange (i.e., $S_{LAND} + E_{LUC}$). Only instantaneous emissions to the atmosphere can be derived directly from the historical or scenario simulations (as, e.g., in Lawrence et al., 2012). These, however, neglect legacy emissions that result in particular from the slow decay of wood products, harvested material left on site, and the adjustment of soil carbon stocks to the altered land-use over decades to centuries, but also comprise slow carbon uptake in processes like forest regrowth. In order to capture all fluxes from land-use change (instantaneous and legacy), additional simulations are essential that exclude the land-use change forcing, such that by difference to the historical or scenario simulation E_{LUC} can be isolated (Pongratz et al., 2014). Note that effects of altered atmospheric CO_2 concentrations by E_{LUC}, with emissions creating a compensating carbon sink in land and ocean (the "land-use feedback"), are excluded in our concentration-driven feedback (Pongratz et al., 2014). Similarly, since CBALONE is driven by the climate from the coupled simulation, changes in surface climate due to land-use change also act the same way in both simulations. Thus, the difference between the simulations with (MPI-GE) and without land-use change (CBALONE) cancels these effects (apart from secondary-order terms) and excludes resulting feedbacks. This is essential to make our estimates consistent with the methodology used in the GCB2020 for the terrestrial budget terms.

CBALONE includes only the long-term dynamics associated with carbon turnover rates and vegetation biogeography. We force CBALONE with daily data from 100 climate realizations taken from the MPI-GE, both with and without anthropogenic land-use change (LUC and noLUC simulations, respectively) comparable to the approach taken by the GCP (Friedlingstein et al., 2020). The land-use change transition data utilized by MPI-GE and CBALONE are taken from the Land-Use Harmonization 2 project (LUH2; Hurtt et al., 2011). While the carbon fluxes from CBALONE did not exactly match JSBACH3 estimates, they consistently simulate JSBACH3 fluxes to within 5% accuracy (Figure S6). Therefore, the CBALONE simulations with land-use change are required so that E_{LUC} could be calculated independent of the small CBALONE error (in absence of the error, the net land-atmosphere exchange could have been directly provided by the MPI-GE simulations).

The climate realizations used to force CBALONE were taken from existing daily output from the MIP-GE historical and RCP4.5 scenarios (1850–2099; Table 1). We chose the RCP4.5 scenario as a scenario of

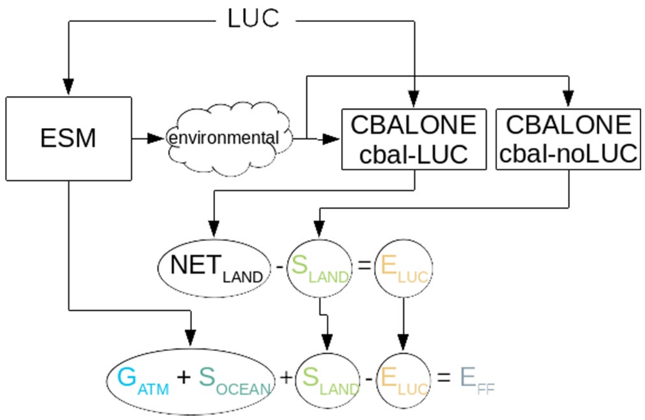

Figure 1. Workflow schematic for simulations and carbon budget decomposition for each ensemble member. Variables from Max Planck Institute Grand Ensemble (MPI-GE) labeled "*environmental*" include leaf area index, net primary productivity (NPP), topsoil temperature, maximum 10 m wind speed, air temperature, and precipitation (see Section 2.2).

medium climate change that estimates the CO_2 emissions under climate policies designed to limit global warming to no more than 3 C over present-day temperatures, allowing us to create uncertainty estimates of fossil emissions that are consistent with this goal. The daily model output variables that are used to force CBALONE include 2 m air temperature, soil temperature, precipitation, net primary productivity (NPP) per plant functional type (PFT), leaf area index (also per PFT), and maximum wind. These variables are marked as "environmental" in Figure 1.

2.2. Carbon Budget Decomposition

The carbon budget is decomposed here into the various source and sink terms as in Friedlingstein et al. (2020), utilizing output from the MPI-GE and the CBALONE simulations. The monthly CBALONE output is aggregated to annual values for comparison to the GCB2020. The cbal-noLUC simulation provides land-atmosphere exchange that would occur without land-use changes, and thus S_{LAND} is calculated as the net biome productivity (NBP) from this simulation. Equation 1 clarifies components of NBP taken from the model, where NPP is NPP, RH is heterotrophic respiration, fFire is carbon flux due to wildfires, fHarvest is carbon flux due to crop and wood harvest, fGrazing is carbon flux due to herbivorous grazing, and fLCC is the instantaneous emissions from land-use induced land cover changes. The fLCC term is zero in the cbal-noLUC simulations.

$$\mathrm{NBP} = S_{LAND} = \mathrm{NPP} + \mathrm{RH} + \mathrm{fFire} + \mathrm{fHarvest} + \mathrm{fGrazing} + \mathrm{fLCC} \quad (1)$$

E_{LUC} is calculated as the difference in NBP between the cbal-LUC and cbal-noLUC simulations (Equation 2; note that fluxes to the natural sinks are negative values and fluxes to the atmosphere are positive consistent with Friedlingstein et al., 2020). Correspondingly, the NBP from the cbal-LUC simulation is equivalent to the net land-atmosphere exchange (NET_{LAND}).

$$E_{LUC} = \mathrm{NBP}\,|_{\text{cbal-LUC}} - \mathrm{NBP}\,|_{\text{cbal-noLUC}} = NET_{LAND} - S_{LAND} \quad (2)$$

G_{ATM} and S_{OCEAN} are taken directly from the MPI-GE output. The implied "compatible" emissions (also E_{FF}) are calculated as the residual of all other terms in the budget (Equation 3; Figure 1), as described in C. Jones et al. (2013) and Roeckner et al. (2010). These are the emissions that would need to occur for CO_2 to be conserved given particular atmospheric concentration, land-use emissions, and natural sink fluxes. This is different from the GCB2020 approach, where all terms were determined independently based on model or observational estimates, which requires a budget imbalance term to be added.

$$E_{FF} + E_{LUC} = G_{ATM} + S_{OCEAN} + S_{LAND} \quad (3)$$

We calculated the full decomposition of the carbon budget for each ensemble member of the historical and RCP4.5 scenarios and compare it to the GCB2020 (Friedlingstein et al., 2020) as the best estimate of the real global carbon cycle. Decadal averages of the MPI-GE ensemble mean and standard deviation are calculated for a direct comparison with the decadal mean and uncertainties presented in the GCB2020. To assess the magnitude of the uncertainties due to internal climate variability compared to the magnitude of the budget terms, we further calculate the signal-to-noise ratio (SNR) of each term as the ensemble mean divided by the ensemble standard deviation.

2.3. Interannual Variability

While internal climate variability may contribute to IAV in carbon fluxes to the natural sinks, there are also variations driven by non-internal climate-related factors, for example, changes in anthropogenic activity ($E_{FF} + E_{LUC}$) and volcanism. An assessment of uncertainties based on temporal standard deviations would be a mixture of internal and non-internal variability, while an ensemble standard deviation at a given time step would reflect variations only due to internal climate variability. In order to assess future uncertainties, it is important that the model can simulate historical IAV appropriately. Here, we assess the ability of individual MPI-GE and CBALONE ensemble members to adequately represent the temporal standard deviation of the historical year-to-year climate variations in each GCB2020 budget term. Therefore, we define a reference IAV as the temporal standard deviation of annual fluxes over the base period 1961–1990 (World Meteorological Organization standard reference period). All models have unique imperfections in their ability to simulate the statistical properties of the carbon fluxes such as mean and standard deviation, which we refer to as model bias. Furthermore, each may have a different trend over the base period which would artificially alter the IAV. To remove the model biases in the ensemble mean of the MPI-GE, we detrend the budget terms of each ensemble member before calculating IAV using an ordinary least-squares regression (OLR) of the ensemble mean over the historical period 1959–2005. We also detrended each model used in the GCB2020 and calculate the IAV over the same period.

2.4. Probability of Exceedance of Past Budget Terms

To evaluate how likely past carbon fluxes were compared to the range of possible climate states due to internal variability, we describe here a measure of the probability of exceedance. Supposing a relatively small amount of CO_2 uptake by the land surface in a particular year, it is quite likely that under more favorable climate conditions for carbon storage this land CO_2 uptake would be exceeded. Therefore, we aim to calculate the probability that the MPI-GE members are greater than the GCB2020 multi-model mean (which we assume to be the closest estimate to historical CO_2 fluxes). Each budget term for the MPI-GE and GCB2020 is OLR detrended in the same way as described above (Section 2.3) except that we use the 1959–2018 period (i.e., the longest available common period for GCB2020 and the MPI-GE simulations). For each year and budget term, we calculate the corresponding cumulative distribution functions ("exceedance") of the MPI-GE ensemble members using a kernel density estimator (Scott, 2015). We then evaluate the GCB2020 terms on the complement of the cumulative distribution functions (1—Pr.) to find their occurrence probability (e.g., see Figure S3). Since we use a cumulative distribution, the complement probability is the "exceedance probability" of the ensemble spread being larger than the historical value. Unusually large historical fluxes will therefore have low probability of exceedance. This is similar to the probability of exceedance calculations from studies on climate extremes (e.g., Suarez-Gutierrez et al., 2020).

Finally, we assess the relationship of the GCB2020 exceedance probabilities for S_{LAND} and S_{OCEAN} fluxes to ENSO, since this is the most prominent mode that drives internal climate variability (Dannenberg et al., 2015; Zhang et al., 2019). We use the annual mean Niño 3.4 index from the NOAA Climate Prediction Center (Climate Prediction Center, 2017) which uses ERSST V5 (Huang et al., 2017) sea surface temperatures averaged over the region 5°N–5°S, 170°–120°W. We then calculate the Pearson's correlation coefficient and the OLR between the exceedance probabilities of the natural sinks and the Niño 3.4 index. We test the significance of this correlation using a two-sided t-test under the null hypothesis that a relationship between the exceedance probabilities of the GCB2020 fluxes and ENSO state can be rejected at the 95% confidence level. Since these methods assume normally distributed data, we beforehand tested the normality of the budget terms and their probabilities using the Shapiro-Wilk test for normality (Shapiro & Wilk, 1965). We found that all budget terms (except for G_{ATM}) are normally distributed in the 1850–2018 period.

3. Results

3.1. Temporal Evolution of Budget Components and Internal Climate Variability Uncertainties

The historical period and RCP4.5 scenario have globally increasing CO_2 fluxes from the atmosphere to the land and ocean sinks until about 2040 before decreasing thereafter (see Figure 2) due to assumed RCP4.5 mitigation measures. The decrease in land and ocean sink is because G_{ATM} in RCP4.5 decelerates after 2040 resulting in an atmospheric concentration of ~525 ppm CO_2 by 2100 (Thomson et al., 2011). The compatible fossil emissions in the MPI-GE (E_{FF} in Figure 2) share similar temporal evolution of the natural sinks. On the other hand, E_{LUC} is driven by the LUH2 land-use data set and is independent of fossil emissions, which increases until about 1990 before becoming a weak net sink from around 2020 onward under the RCP4.5 scenario (Figures 2 and S1b). Within the period 1970–2010, the ensemble means of the G_{ATM} and E_{FF} terms

10.1029/2021GB007019

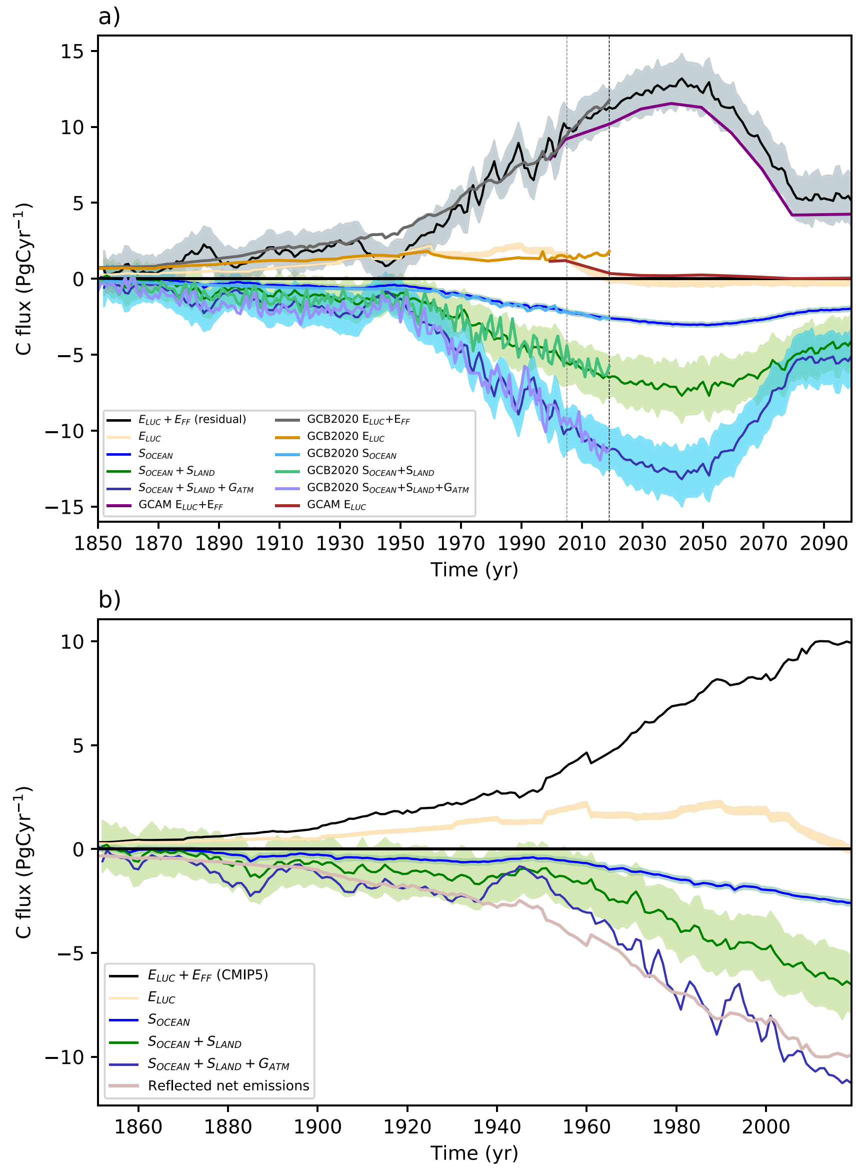

Figure 2. Stacked decomposition of the CO_2 budget terms from the Max Planck Institute Grand Ensemble (MPI-GE) for the historical (1850–2005) and RCP4.5 (2006–2099) scenarios (a) (unstacked plots of the individual terms can be found in Figure S1). Thick lines mark the ensemble mean and shading marks the range of the ensemble ±1 standard deviation. Overlaid are the global carbon budget (GCB)2020 budget terms for comparison. Vertical lines mark the end of the historical period (2006) and the end of the latest Global Carbon Project (GCP) budget (2019). An alternative budget using the Coupled Model Intercomparison Project (CMIP)5 E_{FF} taken from Andres et al. (2012) is also provided (b). The pink line shows the reflected net emissions, the difference with the net natural sinks would give the simulated B_{IM} term in Figure S1f.

show annual to decadal-scale variations, which are a known feature of the CO_2 concentration forcing used in the historical period (caused by the introduction of additional CO_2 observation stations in the 1960s, see Figure 10 of Meinshausen et al., 2017) and are not internally driven variations in the MPI-ESM. The S_{LAND} and S_{OCEAN} do not immediately respond to such rapid changes in G_{ATM} since they are dominated by the climate state and its variability. It then follows that these variations are evident in the residual E_{FF} term.

The budget terms in Figure 2 are stacked for S_{LAND} and G_{ATM}, and hence the shown standard deviation of the ensemble members for these terms aggregates according to a normal sum distribution (i.e., $\sigma[S_{OCEAN} + S_{LAND}] = \sqrt{[\sigma^2\{S_{OCEAN}\} + \sigma^2\{S_{LAND}\}]}$). The atmospheric concentration is prescribed to be the

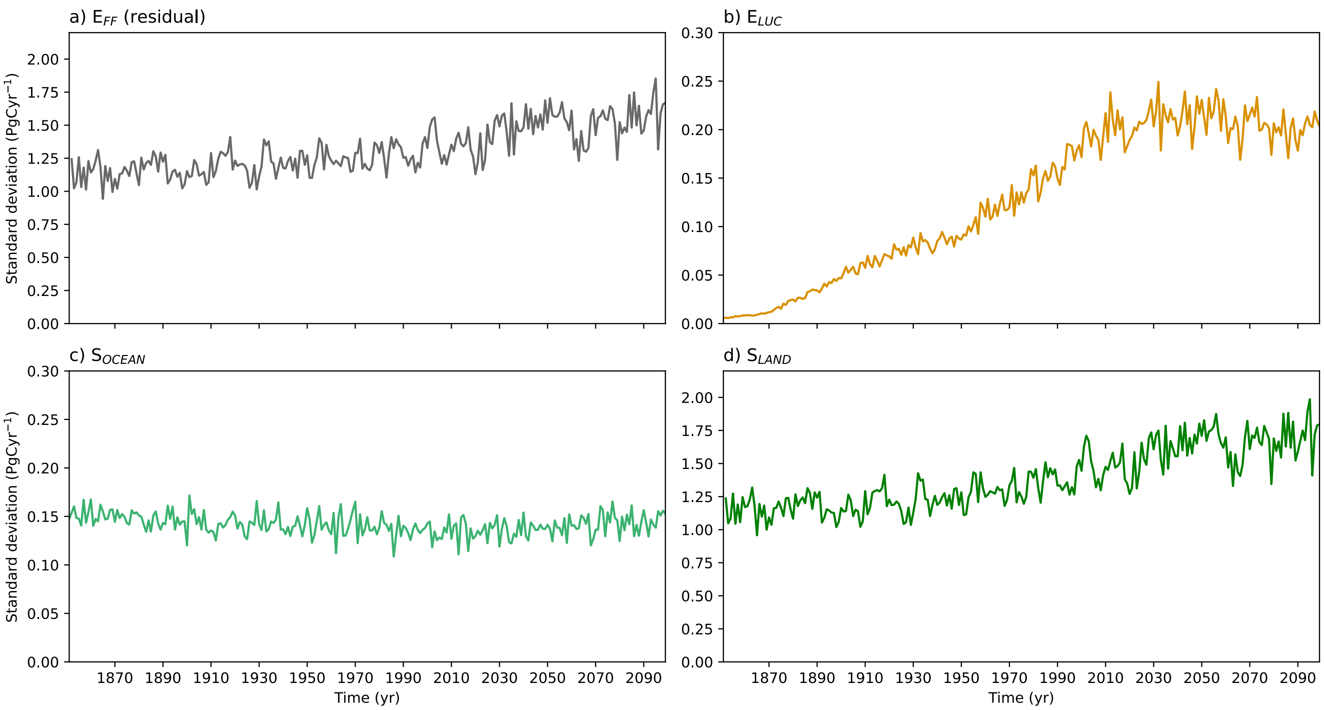

Figure 3. Yearly ensemble standard deviation for each carbon budget term. The emissions are on the top ((a) residual E_{FF} and (b) E_{LUC}) and the natural sink terms are on the bottom ((c) S_{OCEAN} and (d) S_{LAND}).

same for all ensemble members, and so G_{ATM} has no ensemble standard deviation. The standard deviation of residual E_{FF} is inherited directly from the net natural sinks and E_{LUC} because it is calculated as a residual in the budget. S_{OCEAN} has a stable standard deviation of ∼0.15 Pg C yr^{-1} (Figure 3c), which does not have a trend. S_{LAND} has the largest standard deviation throughout the historical period and the RCP4.5 scenario (see Figure 3d), therefore, the standard deviation of the net of natural sinks in Figure 2 (and consequently residual E_{FF}) mostly originates from S_{LAND}. Standard deviation increases with time for residual E_{FF} and S_{LAND} (Figures 3a and 3d) from ∼1 Pg C yr^{-1} in 1850 to ∼1.5 Pg C yr^{-1} in 2100. E_{LUC} standard deviation gradually increases from almost 0 to ∼0.2 Pg C yr^{-1} by 2010 and later (Figure 3b).

The importance of internal climate-driven variations (Figure 3) relative to the ensemble mean state can be better understood by analyzing the SNRs (Figure 4). Values greater than one indicate that the mean state dominates the signal, whereas values less than one indicate that the internal climate variability uncertainty is the dominant factor in the carbon fluxes. For residual E_{FF} and S_{LAND} (Figures 4a and 4d), internal variations are more relevant up until 1970. After that, the mean carbon fluxes (i.e., the forced signal) are much larger than the variations due to internal climate variability, for example, ∼2.5–3 times greater for S_{LAND}. S_{OCEAN} generally follows the same pattern (Figure 4c); the internal climate variability remains several times smaller than the mean carbon flux to the ocean from about 1890 onward. On the other hand, the standard deviation in E_{LUC} is as large as the mean from 2010 onward (Figure 4b), however, this is likely a consequence of the simulation setup: land-use changes begin in 1850 but the full range of variation from the legacy emissions of land-use change does not manifest until several decades later. This means the E_{LUC} SNR is effectively only valid under the future scenario when the mean E_{LUC} is small.

3.2. Comparison to GCB2020

3.2.1. Comparison of Means

We compare here the GCB2020 mean of each budget term to the ensemble mean of the MPI-GE for each decade, before comparing the variances in the following sections. First, the residual E_{FF} mean increases faster in the MPI-GE than observed in the GCB2020 (Figure 5a). Initially, MPI-GE residual E_{FF} in the 1960s

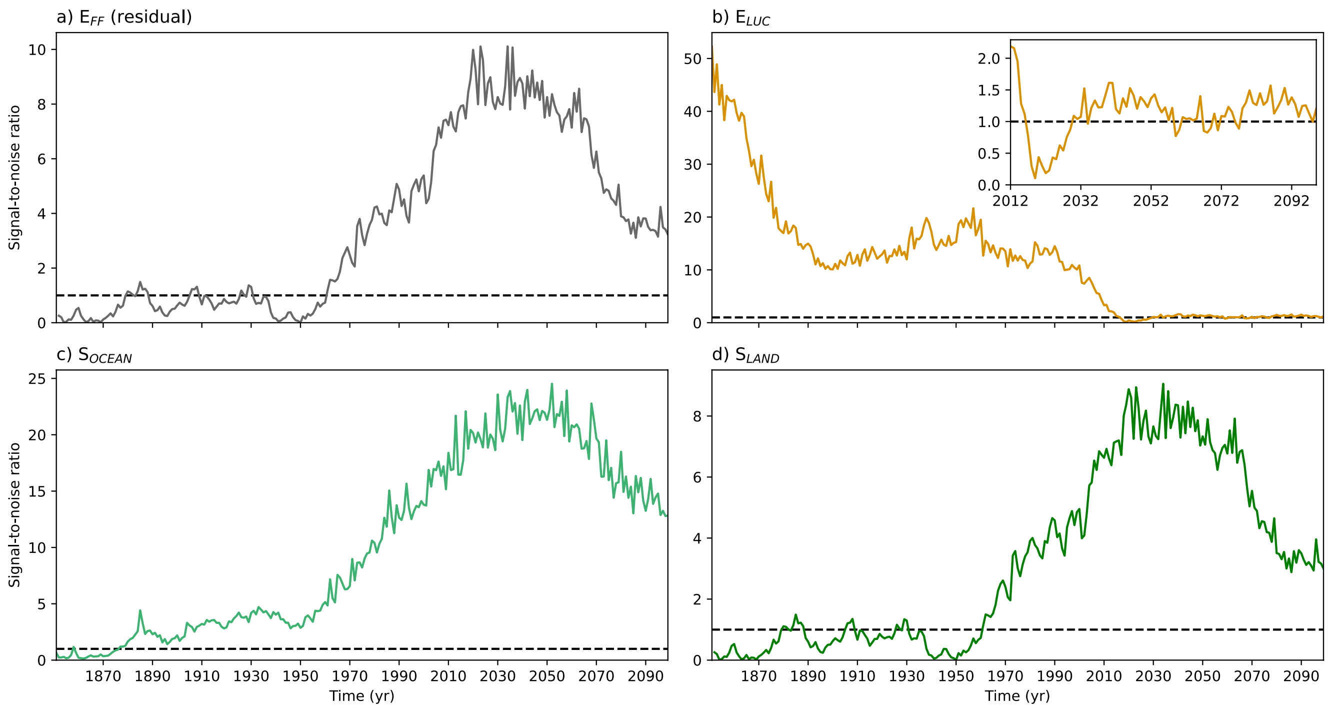

Figure 4. Yearly signal-to-noise ratio (SNR) for each budget term in the Max Planck Institute Grand Ensemble (MPI-GE). Dashed lines delineate ratio 1, where the standard deviation of the respective flux equals the mean flux. E_{LUC} has an inset plot with the post-2010 period zoomed in, when variations from legacy land-use fluxes have fully established.

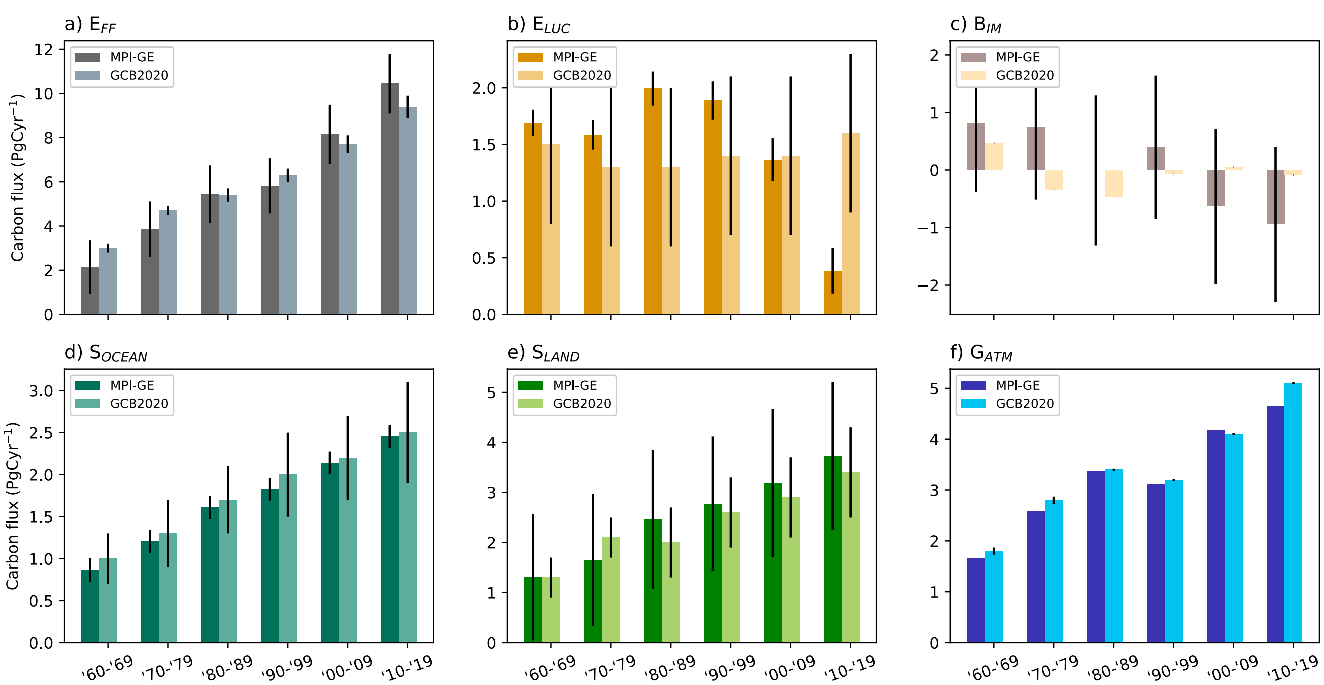

Figure 5. Decadal average of carbon flux budget terms (bars), and the uncertainty expressed as ±1 standard deviation from the mean (error whiskers). The Max Planck Institute Grand Ensemble (MPI-GE) uncertainties are ensemble standard deviations and the global carbon budget (GCB)2020 uncertainties are multi-model standard deviations. The dark bars are the MPI-GE and the lighter bars are the GCB2020 values taken from Friedlingstein et al. (2020). The top row (a) and (b) are the emissions and the simulated budget imbalance term (c) as shown in Figure 2b, and the bottom row (d, e, and f) are the sink terms.

is less than the GCB2020 estimate by 0.8 Pg C yr^{-1} while it is greater than it by 1.3 Pg C yr^{-1} in the 2010–2018 decade. However, the range of GCB2020 means is well within the range of values simulated by the MPI-GE. Second, there are large differences in the mean E_{LUC} fluxes between MPI-GE and GCB2020 (Figure 5b). MPI-GE E_{LUC} is larger compared to GCB2020 in decades prior to 2000, however, these values are also within the large uncertainty ranges of the GCB2020. In recent decades, the MPI-GE estimates lower E_{LUC} than the GCB2020. Third, S_{LAND} tends to be slightly higher in the MPI-GE for almost all decades (Figure 5e). Fourth, S_{OCEAN} mean fluxes in MPI-GE and GCB2020 are very similar (Figure 5d). Lastly, G_{ATM} in MPI-GE has similar decadal variations as GCB2020, both displaying a dip in the 1990s, and there is no consistent bias (Figure 5f).

3.2.2. Un-Bias-Corrected Comparison of Uncertainties

The uncertainty ranges in Figure 5 are based on ensemble standard deviations for MPI-GE (and therefore reflect internal climate variability uncertainties) and multi-model standard deviation for GCB2020. These ranges can tell us two things: how realistic the MPI-GE range of fluxes is compared to observations, and how large uncertainties associated with internal climate variability are compared to other sources of uncertainty (e.g., from observational measurements or the differing process representations in the different GCB2020 models). Therefore, we will determine here whether the GCB2020 mean state lies outside the MPI-GE uncertainty ranges for each budget term.

Residual E_{FF}, B_{IM} (based on the budget in Figure 2b) and S_{LAND} (Figures 5a, 5c and 5e) have larger standard deviations in the MPI-GE compared to GCB2020, that is, internal variability is a larger source of error than observational and model uncertainty (more detail follows in Section 3.2.3). The GCB2020 mean for these budget terms falls within the uncertainty range due to internal climate variability, demonstrating the capability of MPI-GE to capture the observed carbon flux.

On the other hand, E_{LUC} and S_{OCEAN} have a narrower range of internal climate variability uncertainty in the MPI-GE compared to the modeled uncertainty in the GCB2020 (Figures 5b and 5d). While the GCB2020 mean is within the MPI-GE uncertainty for S_{OCEAN} for most decades (indicating consistency between the two), E_{LUC} GCB2020 means are outside the corresponding MPI-GE ranges for nearly all decades. However, the uncertainty ranges of MPI-GE and GCB2020 overlap for both S_{OCEAN} and E_{LUC}, that is, certain ensemble members match certain GCB2020 models. Only, the E_{LUC} 2009–2018 mean and standard deviation of the GCB2020 is outside the standard deviation range of uncertainty due to internal climate variability, indicating clear inconsistency (see discussion Section 4.1).

There is no uncertainty range for G_{ATM} from MPI-GE (Figure 5f) since all ensemble members are prescribed with the same atmospheric CO_2 concentration. The error whiskers in the G_{ATM} GCB2020 are derived from various observational uncertainties, which are very small compared to the terms that are simulated by dynamical models (S_{LAND}, S_{OCEAN}, and E_{LUC}). Because the MPI-GE CO_2 concentration starting 2006 is derived from the Global Change Assessment Model (GCAM; Thomson et al., 2011), the difference in G_{ATM} between MPI-GE and the GCB2020 for the last two decades may in part be due to the differences in carbon cycle processes that are represented in MPI-ESM and GCAM.

3.2.3. Bias-Corrected Comparison of Uncertainties

To more directly evaluate the magnitude of the historical uncertainties associated with internal climate variability compared to the GCB2020, Figure 6 shows the standard deviations where the biases in the means have been removed (centered). The models used in the GCB2020 estimates are forced by only one realization of the climate state—the actual historical climate evolution. Therefore, the plausible carbon fluxes under different climate states cannot be inferred using only the GCB2020, and while the models used in the GCB2020 do contain internal climate variability, the multi-model standard deviations only account for model uncertainty, but not that from natural variability. If we assume that there is no or negligible uncertainty due to internal climate variability associated with the multi-model GCB2020 standard deviation and that the standard deviation of the MPI-GE is entirely due to internal climate variability, then we can find the proportion of the total uncertainty attributable to internal climate variability (i.e., the sum of GCB2020 and MPI-GE uncertainties; red lines in Figure 6). The importance of internal climate variability decreases with time for S_{LAND} and residual E_{FF} and the MPI-GE land sink uncertainty increases faster than the multi-model uncertainty in the GCB2020. For the 2009–2018 decade, the contribution of internal climate variability to

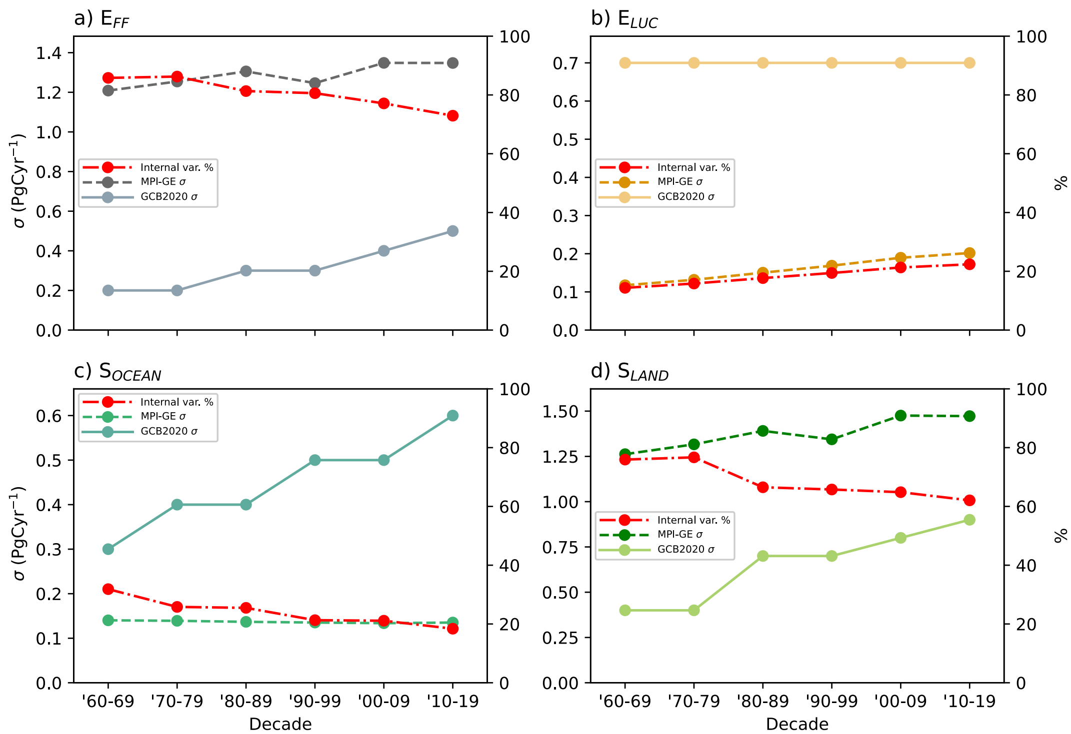

Figure 6. Centered standard deviation of carbon flux from the multi-model global carbon budget (GCB)2020 (solid lines) and ensemble standard deviation from the Max Planck Institute Grand Ensemble (MPI-GE) (dashed lines). The relative contribution of internal climate variability uncertainty is marked in red (dot-dashed lines corresponding to the right-hand axis). The color-coding is the same as that used in Figures 2–5.

total uncertainty is 70% for the residual E_{FF} and 60% for S_{LAND}. A constant multi-model uncertainty was assumed for E_{LUC} in the GCB2020 and therefore the MPI-GE E_{LUC} uncertainty increases gradually relative to it. By the 2009–2018 decade, the uncertainty due to internal climate variability would account for 22% of the total E_{LUC} uncertainty. Lastly, approximately 20% of total uncertainty is from internal climate variability uncertainty for S_{OCEAN}.

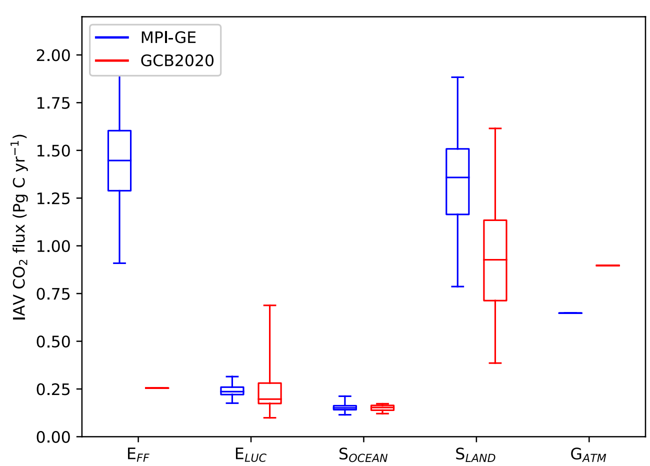

Figure 7. Box and whisker plots of interannual variability (IAV) calculated as the standard deviation over the base period 1961–1990 for the Max Planck Institute Grand Ensemble (MPI-GE; blue) and the global carbon budget (GCB)2020 (red). The ranges shown here are derived from the ensemble members for MPI-GE, and from multiple model simulations for the GCB2020. The boxes mark the median and inter-quartile range, and the whiskers mark the full range of values.

3.2.4. Interannual Variability

The ability of individual ensemble members to capture the IAV (in the base period 1961–1990) for each term compared to the GCB2020 IAVs is shown in Figure 7. The ranges of the IAVs generally have good overlap for the E_{LUC} and S_{OCEAN} budget terms. This means that individual MPI-GE members can simulate a plausible range of IAV values that are not significantly different from the published values from the GCB2020. S_{LAND}, however, shows some IAV bias in the MPI-ESM compared to other models in the GCB2020. IAV in MPI-GE S_{LAND} tends to be on average 0.4 Pg C yr^{-1} larger than other models. A higher IAV may contribute to the large ensemble spread in the MPI-GE for S_{LAND} (compare to Figure 5). There are large differences between MPI-GE and GCB2020 for E_{FF}, and G_{ATM} (Figure 7). Evaluation of G_{ATM} is difficult because there is no associated uncertainty range; the GCB2020 only has one potential realization of past emissions and observed CO_2 concentration, and the MPI-GE atmospheric CO_2 concentrations are prescribed. The observationally-based GCB2020 uncertainties are only 0.02 Pg C yr^{-1} for G_{ATM} and at most 0.5 Pg C yr^{-1} for residual E_{FF} and if we use these values as a range on top of the GCB2020 IAV, MPI-GE is still outside these ranges.

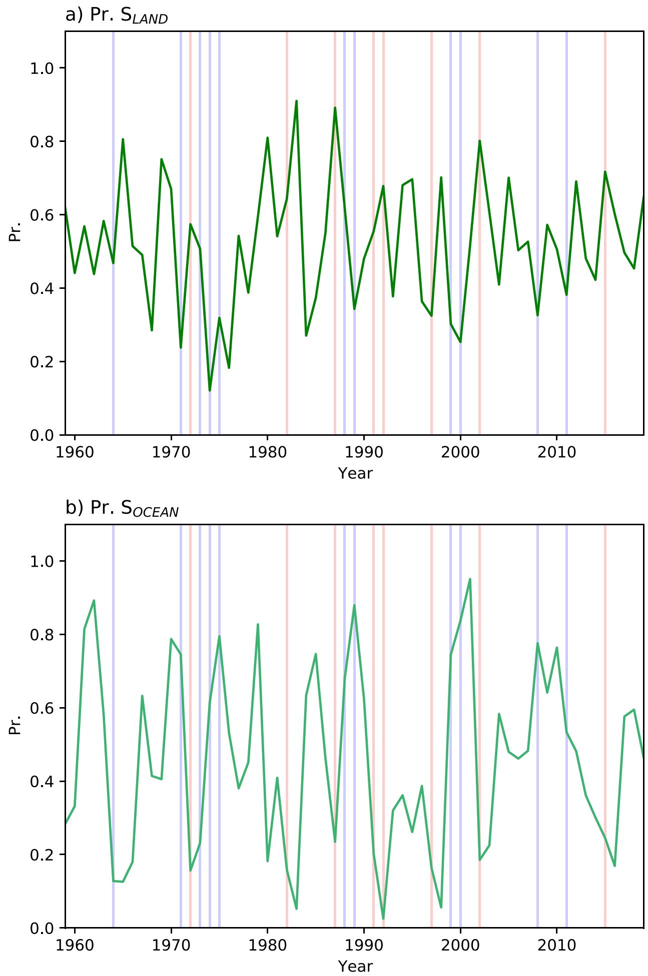

Figure 8. Probability of exceedance that the Max Planck Institute Grand Ensemble (MPI-GE) carbon fluxes are greater than the historical global carbon budget (GCB)2020 mean. Lower values indicate years where the carbon flux to the respective sink was *unusually* high compared to the MPI-GE *distribution (vice versa for large values)*. The vertical lines mark El Niño (red) and La Niña (blue) years where Niño 3.4 index is greater than 1 standard deviation from the mean.

3.3. The Relationship of Historical Probabilities to ENSO

To investigate a potential source of the IAV and uncertainty from internal climate variability, we examine here the exceedance probabilities and the relationship to ENSO. Figure 8 shows the probability of the magnitude of the past carbon fluxes in GCB2020 with respect to the distribution of the MPI-GE. Higher values indicate years where the carbon flux for the respective sink was unusually small compared to the MPI-GE distribution and thus were more likely to be exceeded under more favorable climate conditions. S_{LAND} and S_{OCEAN} have large annual variations in exceedance probability. For example, since 1960 there were three years where the historical S_{LAND} was so high, related to La Niña, that it had a chance of less than 20% to be exceeded and 5 years with S_{LAND} so low that it had a chance of more than 80% to be exceeded (Figure 8a). This highlights the importance of using a large ensemble to capture the high variability in S_{LAND} (see Section 4.5). The cause of these year-to-year variations may come from a variety of internal climate variability modes. To investigate potential drivers, Figure 9 shows that there are significant correlations between the Niño 3.4 index and S_{OCEAN} or S_{LAND} exceedance probability of −0.61 and 0.56 respectively (see also Text S1 and Figure S2).

4. Discussion

In summary, S_{LAND} has the largest uncertainty, which emphasizes the dominant role of internal climate variability on the land sink (Figure 3d). This uncertainty gradually increases over time to approximately ± 1.5 Pg C yr^{-1}. While the global S_{LAND} flux and CO_2 concentration increases until the middle of the 21st century (Figure 2), afterward its SNR of the mean flux nevertheless decreases (Figure 4b). The internal climate variability uncertainty in E_{LUC} is relatively smaller at approximately ± 0.2 Pg C yr^{-1} (Figure 3b). However, the trend in E_{LUC} variability is likely due to a combination of sensitivity to initial conditions and the time delay associated with legacy land-use change emissions. The S_{OCEAN} variations from internal climate variability are similarly small as those in E_{LUC} but show almost no trend (Figure 3c). The S_{LAND} internal climate variability accounts for about 70% of the total uncertainty that results from both internal variability and uncertainties from models and observations (Figure 6d), much more than for E_{LUC} (approximately 22%) and S_{OCEAN} (approximately 19%). The standard deviations of the MPI-GE compare well with the uncertainty ranges of the GCB2020 for most budget terms: with respect to the ensemble standard deviation against multi-model standard deviations (usually at least an overlap, Figure 5), and with respect to individual ensemble IAV against individual model IAV in the GCB2020 (Figure 7). Finally, we show that the effect of internal climate variability on the historically observed exceedance probabilities of carbon fluxes to the land and ocean have significant but moderate correlations to ENSO (Figure 9).

4.1. Differences Between MPI-GE and GCB2020

One of the most striking differences between the MPI-GE and the GCB2020 estimates is in E_{LUC}, where the forced ensemble mean signal from land-use change in the RCP4.5 scenario differs from the observed LUH2 data in the last historical decade. The MPI-GE E_{LUC} transitions to a net sink at around 2020, while the forcing used in GCB2020 estimates sustained E_{LUC} until this period (Bastos et al., 2020; Friedlingstein et al., 2020). Given that the variance of E_{LUC} ensemble members is quite small compared to the forced mean response, the disparity between the RCP4.5 land-use change and the GCB2020 becomes evident. The RCP4.5 scenario is characterized by a high CO_2 price that encourages investment into agricultural intensification rather than

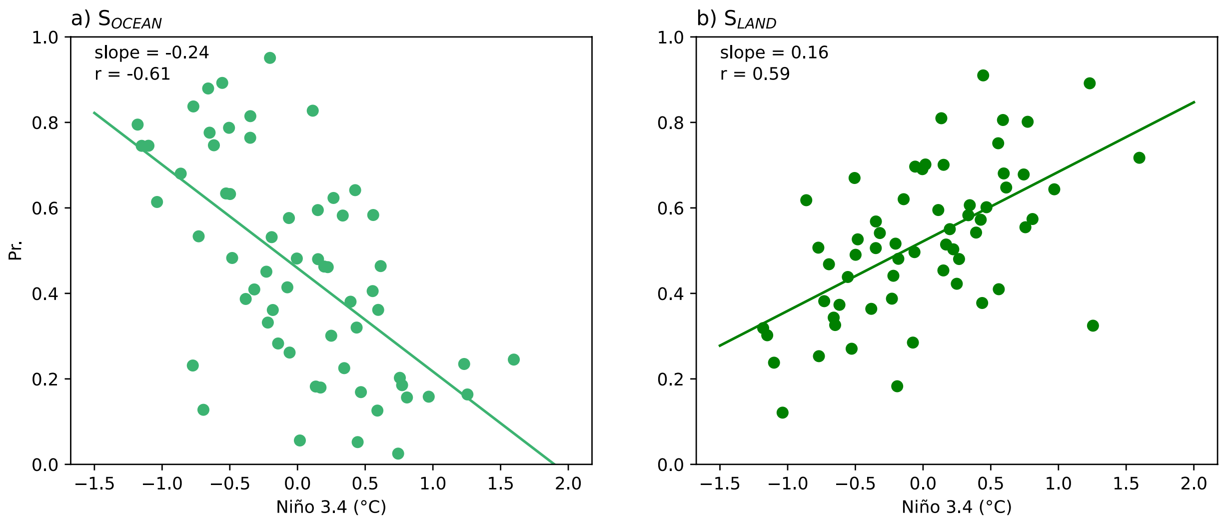

Figure 9. Regression and correlation analysis between Niño 3.4 index and the probability of exceedance for carbon fluxes (a) S_{OCEAN} and (b) S_{LAND}. The units of the slope are in °C^{-1}.

expansion. Consequently, re-/afforestation would occur following widespread abandonment of agricultural lands and substantial deforestation reduction since 2007 (Thomson et al., 2011). Despite the process of forest regrowth (such as that in North America and Europe; Doelman et al., 2020) being slow, the MPI-GE reduction in E_{LUC} associated with stopping deforestation globally (in particular the Amazon and other tropical regions) is quick and modeling studies simulate substantial carbon uptake by re-/afforestation and reduced deforestation. For example, Sonntag et al. (2016) estimate an uptake of about 200 Pg C over the 21st century with RCP4.5 land-use change in an RCP8.5 climate compared to unmitigated deforestation. However, the trajectory of RCP4.5 land-use change has not been followed until now, and so the land-use-related mitigation potential remains untapped. This explains the large divergence of our results from the GCB2020 estimates for the last 15 years.

There are also considerable differences in the "compatible" residual E_{FF} in the MPI-GE compared to the GCB2020 values. If we assume the GCB2020 estimate to be the closest estimate to the mean in reality, then the MPI-GE first underestimates the E_{FF} then overestimates it. The discrepancy may arise due to the closure of the carbon balance and the consequent effect that S_{LAND} has on the compatible emissions. On the other hand, the GCB2020 has an imbalance term that includes carbon fluxes that remain unaccounted for. This term would include errors introduced by the calculation of budget terms independently (e.g., model bias errors in E_{LUC} and S_{LAND}, e.g., Dai & Fung, 1993), errors from incomplete coverage of observations, and minor terms that are not included in the budget decomposition. For these reasons, we would not expect the MPI-GE to accurately reproduce E_{FF}.

Lastly, another approach to evaluating the MPI-GE against the GCB2020 is to verify that there are no trends in the budget imbalance relative to the GCB2020. If the compatible residual E_{FF} in the MPI-GE budget is replaced with the CMIP5 E_{FF} values (Figure 2b), a budget imbalance term (B_{IM}) can be calculated that is the residual carbon flux that is not accounted for under each ensemble member's climate state. This simulated B_{IM} term (Figure S1f) derived from the MPI-GE is largely consistent with the B_{IM} from the GCB2020 and shows no significant long-term trends over the analysis period. Both MPI-GE and GCB2020 show as a positive B_{IM} around the 1950s and again more briefly in the 1990s (suggesting either an overestimate in the emissions or underestimate in the sinks). While Friedlingstein et al. (2020) could not directly attribute a cause to the B_{IM}, they suggest that its variations originate mostly from S_{LAND} and S_{OCEAN}. Specifically, they suggest that it could originate from internal variability which models cannot capture with a single realization. However, the multiple realizations in the MPI-GE B_{IM} range also show positive values in the 1950s, which suggests that it is more likely from common deficiencies in model physics, resolution, or forcing data. In particular, the land-use forcing could explain the 1950s B_{IM}, as the LUH2 forcing creates large emissions in the 1950s (e.g., Hansis et al., 2015) not captured by datasets based on other land-use forcing such as FAO (Houghton & Nassikas, 2017).

4.2. Allowable Emissions Under RCP4.5

The standard deviations in the MPI-GE (Figure 2) are derived either directly from the ensembles or are inferred from other budget terms, and therefore they should be interpreted with care. The standard deviation of residual E_{FF} is mostly derived from S_{LAND} due to its calculation as a residual. In this case, the ranges here are merely a range of emissions that are compatible with the likely range of climate states and the corresponding strengths of the ocean and land sinks. Therefore, the residual E_{FF} uncertainty estimates from MPI-GE should not be interpreted as variations in fossil fuel emissions due to internal climate variability-related global demand.

The net sinks and the corresponding compatible residual E_{FF} range are still useful when deciding what the allowable future emissions may be. They indicate the allowable emissions (accounting for internal climate variability) if appropriate policies are implemented to successfully mitigate climate change in a manner that is consistent with the RCP4.5 scenario. Therefore, the maximum and minimum ensemble ranges of 9–18 Pg C yr^{-1} in residual E_{FF} at 2050 denote allowable emissions under this scenario (2019 was 9.95 Pg C yr^{-1} as per the GCB2020). In Figure 2, the ±1 standard deviation range of the ensemble is shown instead. In the comparison, it is clear that extreme outliers occur mainly at the maximum end. These maximum values may occur before fossil emissions have to drop steeply in the MPI-GE and level off at around 5 Pg C yr^{-1} if the 3 C target is to be met by 2100. This evolution matches well the fossil emissions estimates from GCAM (Thomson et al., 2011) but allows some higher peak emissions than the Integrated Assessment Model assumed, suggesting smaller assumed sinks and slightly larger E_{LUC} in the simplified carbon cycle of this assessment model (see Figure 2 to compare to E_{FF} and E_{LUC} from GCAM).

As highlighted by Mankin et al. (2020), decision makers need to be provided the full range of possible outcomes in order to make appropriate decisions. For example, policy decisions based only on the most likely outcome may lead to a blowout of greenhouse gas inventory targets, particularly if S_{LAND} performs poorly within a given 5-year accounting period of the Paris Agreement's Global Stocktake (UNFCCC, 2015, 2017). On the other hand, caution should be taken when considering the efficacy of past decision-making because internal variability uncertainties can potentially obfuscate emission reduction efforts such as re-/afforestation.

4.3. Trends in Uncertainty

The increase in standard deviation in the ensemble members for S_{LAND} may be due to an increase in the variability in the climate state as is expected under a warming climate. For example, Maher et al. (2019) find an increase in the global mean precipitation variability in the MPI-GE 1% CO_2 scenario. The trend in S_{LAND} internal variability can also potentially arise from the increase in the magnitude of fossil emissions, which is initially forced in the MPI-GE as the prescribed atmospheric CO_2 concentration. Larger emissions would result in higher atmospheric CO_2 concentrations and increased potential carbon uptake by vegetation via so-called CO_2 fertilization (Walker et al., 2021). This combined with the effect of unfavorable climatic conditions (i.e., heat and drought stress) on the carbon uptake by plants acting on an increased carbon stock, results in a larger variance depending on the climate conditions. The increasing internal variability makes it more likely that S_{LAND} becomes near-neutral by the end of the century compared to the start of the historical period (Figure S1d). This contrasts somewhat with S_{OCEAN}, which has a relatively lower variance and does not have a trend in the historical or future periods under the RCP4.5 scenario (a similar standard deviation is found by Li & Ilyina, 2018). However, under higher emissions scenarios S_{OCEAN} has been shown to also have increasing trends in CO_2 flux standard deviation (see Figure 1 of Maher et al., 2019).

The trend in E_{LUC} may arise for several reasons. First, the legacy effects of land-use change (mostly from wood harvest) take time to manifest. The anthropogenic pools in which CBALONE stores deforested biomass decay to the atmosphere at time scales of 1–100 years. The variance of the ensemble members therefore not only depends on the climate variability of the current year but also on that of preceding years. Consequently, it would take at least 100 years for the full variance due to land-use change to manifest. Similarly,

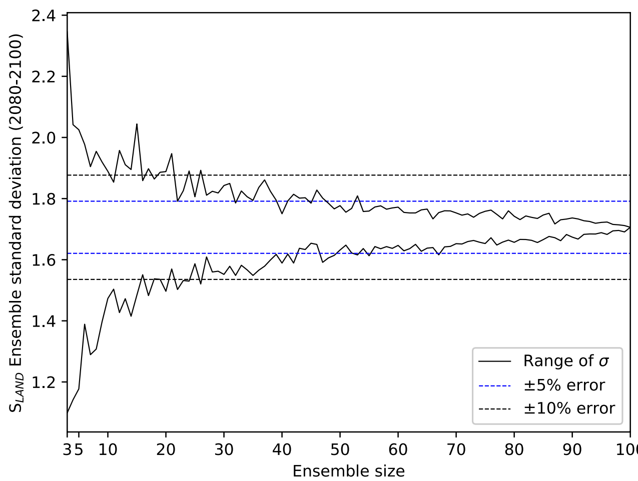

Figure 10. Range of ensemble standard deviation (2080–2100) as a function of sample size from 30 subsamples for S_{LAND}.

the carbon pool of woody, slowly decomposing litter left on site after clearing or harvesting will build up over time as land-use transitions occur. Thus, more litter is available to react to the climate-dependent microbial decomposition. Note that while the study of Yue et al. (2020) included this effect in their assessment of the contribution of land-use to the interannual variability of the land carbon pools, their high IAV of E_{LUC} (30%–45% of net land exchange IAV, compared to 15% in this study) also originates from attributing part of S_{LAND} (the part on managed land) to E_{LUC}. Internal variability alone, our study shows, is about 0.25 Pg C yr^{-1} standard deviation for E_{LUC} in recent decades (Figure 3) or 20% of the total uncertainty (model plus internal; Figure 6). IAV of E_{LUC} in the MPI-GE is only slightly larger than in the GCB2020 (Figure 7), indicating that the main driver is not internal climate variability, but land-use forcing.

While the data analyzed in this study is annual and much of the analysis concerns IAV, we conducted simulations for several centuries, and therefore, the longer time scale variations must also be considered. There are centennial-scale internal variations in the land carbon content in JSBACH3 and CBALONE (see Figure 2 in Schneck et al., 2013) which could influence trends and variability of S_{LAND} and E_{LUC} for simulations that run for several hundred years. These variations have a periodicity of ~250 years and consist of a change in the total land carbon content of ~8 Pg C. This corresponds to an average land carbon flux of 0.03 Pg C yr^{-1} or roughly 2% of the MPI-GE S_{LAND} standard deviation. Schneck et al. (2013) suggest that these long-duration variations in land carbon content are linked to variations in anthropogenic land cover changes, and the modulation of soil respiration by long-term changes in temperature from volcanism and solar forcing. Since the duration of the MPI-GE and CBALONE simulation in this study is 250 years, it is possible that these long-term variations may affect the estimates of internal climate variability uncertainty in S_{LAND}.

4.4. ENSO as a Potential Source of Variability

ENSO is positively correlated with S_{LAND} exceedance probabilities and negatively correlated with S_{OCEAN} exceedance probabilities, which is consistent with how ENSO affects CO_2 fluxes to the land surface and ocean. During La Niña, cool and moist mean global conditions tend to encourage vegetative productivity on land and increase land carbon storage, while El Niño drought conditions put widespread stress on ecosystems and suppress productivity (Gonsamo et al., 2016; C. D. Jones et al., 2001). Meanwhile, over the ocean, stronger pacific equatorial up-welling during La Niña brings dissolved inorganic carbon-rich subsurface water to the surface, thereby favoring CO_2 outgassing and reducing net CO_2 uptake (Feely et al., 1999; Jones et al., 2001). The cooler sea surface temperatures during La Niña events can increase the dissolution of CO_2 and can reduce CO_2 outgassing, but this is a smaller term relative to the up-welling-induced CO_2 outgassing. This could explain the diverging response of S_{OCEAN} to ENSO compared to that of S_{LAND}. The moderate correlation suggests that while ENSO may have a considerable impact on IAV in CO_2 fluxes, it is very likely that other climate modes and internal dynamics are also important. No significant correlations with other climate modes could be found at the global scale, however, the impacts of climate modes on regional budgets may be considerable.

4.5. Importance of Ensemble Size

Lastly, it is important to discuss the effect of ensemble size on the results and whether or not using 100 members is enough or more than necessary. A framework to assess this is demonstrated in Milinski et al. (2020). In accordance with this framework, our goal is to quantify variability using the metric of ensemble standard deviation, to within 5% accuracy of the full 100-member variance. We estimate standard deviation using 30 iterations of subsample sizes from 3 to 100 members without replacement. Figure 10 suggests that at least 40 ensemble members are required to capture the standard deviation of S_{LAND} to within ±5% accuracy. Since S_{LAND} has the largest standard deviation of all budget terms, the accuracy of a sub-sample of the carbon budget decomposition would depend on this term. The other budget terms (Figure S5) do not display variations as large as S_{LAND}, and therefore, 40 members are sufficient for those terms. Whether this result is representative of other models that simulate internal variability through ensemble simulations depends on the budget terms. In the absence of extensive multi-model large ensemble projects that provide the full suite of budget terms, including the split into S_{LAND} and E_{LUC}, we evaluated this based on the IAV in the models participating in the GCB2020 simulations that are forced with observed climate (Figure 7). A key assumption is that MPI-GE is capable of accurately representing IAV, and the fact that MPI-GE slightly overestimates S_{LAND} IAV by 0.4 Pg C yr^{-1} compared to other models in the GCB2020 suggests that the minimum 40 ensemble members required here may be a conservative estimate.

5. Conclusion

In this study, we use a large ensemble of single-model simulations from the MPI-GE and a subcomponent of JSBACH3 (called CBALONE) to decompose the global anthropogenic carbon budget into fossil and land-use change emissions, atmospheric growth, and natural land and ocean sinks. Through its 100 ensemble members, the MPI-GE captures the uncertainties associated with internal climate variability, which we compare to the 2020 GCB's uncertainty and interannual variability, and calculate exceedance probabilities of the past carbon fluxes with respect to a full range of climate variability states. We estimate about 40 ensemble members are required to capture internal variability in S_{LAND}, and thus, all budget components. Contrary to S_{LAND}, to reduce uncertainty in S_{OCEAN} and E_{LUC} estimates, we must prioritize reducing observational error and model spread rather than capturing internal variability. Despite its high internal variability, S_{LAND} (or S_{OCEAN}) is likely not the reason behind the high budget imbalance found in previous studies for the 1950s, which suggests common model deficiencies or biases in the land-use forcing.

We also present a novel estimate of the uncertainty in land-use change emissions associated with internal climate variability at approximately ±0.2 Pg C yr^{-1}, which we estimate would account for about 20% of the total (internal and multi-model) land-use change emissions uncertainties. Land-use change emissions thus contribute little to interannual variability of the annual carbon budget and are driven rather by land-use forcing than by climate variability.

We investigate future changes in the GCB under RCP4.5 and demonstrate upper and lower bounds on the allowable future CO_2 emissions depending on climate variations. The RCP4.5 scenario exemplifies a future where climate policies are implemented to limit warming to less than 3 C over present-day conditions. Our study largely confirms that the allowable emissions under the assumptions of the socioeconomic model GCAM are compatible with RCP4.5, though slightly higher emissions of up to 13 Pg C yr^{-1} on average would be allowed in the MPI-ESM. The minimum of the full ensemble range is 9 Pg C yr^{-1} and would be the lower risk limit to ensure we stay below 3°C warming for all possible climate states, while the maximum of 18 Pg C yr^{-1} would be the higher risk limit for the climate states leading to stronger land CO_2 uptake. Our results suggest that internal variability of the natural land sink increases over the 21st century, which puts the steady persistence of carbon removal by land ecosystems at risk. We also show that even when accounting for random variations in climate and natural sinks, the emissions in recent decades for land-use change—characterized by continuing global deforestation—are dangerously inconsistent with the RCP4.5 goals and further erode our ability to successfully mitigate future warming.

Data Availability Statement

The authors also thank the German Climate Computing Centre (DKRZ), for providing computational resources. CMIP5 emissions data are available from RCP Database http://www.iiasa.ac.at/web-apps/tnt/RcpDb and the GCP Global Carbon Budgets data are available from https://www.globalcarbonproject.org/carbonbudget/archive.htm.

Acknowledgments
The authors would like to acknowledge the generous help and comments from the Max Planck Institute Grand Ensemble working group. H. Li was supported by the European Union's Horizon 2020 research and innovation program under grant agreement no. 821003 (4C). Open access funding enabled and organized by Projekt DEAL.

References

Andres, R. J., Boden, T. A., Bréon, F.-M., Ciais, P., Davis, S., Erickson, D., et al. (2012). A synthesis of carbon dioxide emissions from fossil-fuel combustion. *Biogeosciences*, *9*, 1845–1871. https://doi.org/10.5194/bg-9-1845-2012

Ballantyne, A. P., Alden, C. B., Miller, J. B., Tans, P. P., & White, W. C. (2012). Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years. *Nature*, *488*, 70–72. https://doi.org/10.1038/nature11299

Bastos, A., O'Sullivan, M., Ciais, P., Makowski, D., Sitch, S., Friedlingstein, P., et al. (2020). Sources of uncertainty in regional and global terrestrial CO_2 exchange estimates. *Global Biogeochemical Cycles*, *34*. https://doi.org/10.1029/2019GB006393

Bastos, A., Running, S. W., Gouveia, C., & Trigo, R. M. (2013). The global NPP dependence on ENSO: La Niña and the extraordinary year of 2011: Global NPP and ENSO: The 2011 La Niña. *Journal of Geophysical Research: Biogeosciences*, *118*, 1247–1255. https://doi.org/10.1002/jgrg.20100

Canadell, J. G., Le Quere, C., Raupach, M. R., Field, C. B., Buitenhuis, E. T., Ciais, P., et al. (2007). Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proceedings of the National Academy of Sciences*, *104*, 18866–18870. https://doi.org/10.1073/pnas.0702737104

Climate Prediction Center. (2017). *Monthly atmospheric SST indices*. Retrieved from https://www.cpc.ncep.noaa.gov/data/indices/

Conway, T. J., Tans, P. P., Waterman, L. S., Thoning, K. W., Kitzis, D. R., Masarie, K. A., & Zhang, N. (1994). Evidence for interannual variability of the carbon cycle from the National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory Global Air Sampling Network. *Journal Geophysical Research*, *99*, 22831–22855. https://doi.org/10.1029/94JD01951

Dai, A., & Fung, I. Y. (1993). Can climate variability contribute to the "Missing" CO_2 sink? *Global Biogeochemical Cycles*, *7*, 599–609. https://doi.org/10.1029/93GB01165

Dannenberg, M. P., Song, C., Hwang, T., & Wise, E. K. (2015). Empirical evidence of El Niño-Southern Oscillation influence on land surface phenology and productivity in the western United States. *Remote Sensing of Environment*, *159*, 167–180. https://doi.org/10.1016/j.rse.2014.11.026

Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012). Uncertainty in climate change projections: The role of internal variability. *Climate Dynamics*, *38*, 527–546. https://doi.org/10.1007/s00382-010-0977-x

Dlugokencky, E., & Tans, P. (2018). *Trends in atmospheric carbon dioxide*. National Oceanic & Atmospheric Administration, Earth System Research Laboratory (NOAA/ESRL).

Doelman, J. C., Stehfest, E., Vuuren, D. P., Tabeau, A., Hof, A. F., Braakhekke, M. C., et al. (2020). Afforestation for climate change mitigation: Potentials, risks and trade-offs. *Global Change Biology*, *26*, 1576–1591. https://doi.org/10.1111/gcb.14887

Feely, R. A., Wanninkhof, R., Takahashi, T., & Tans, P. (1999). Influence of El Niño on the equatorial Pacific contribution to atmospheric CO_2 accumulation. *Nature*, *398*, 597–601. https://doi.org/10.1038/19273

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., et al. (2020). Global carbon budget 2019. *Earth System Science Data*, *12*, 3269–3340. https://doi.org/10.5194/essd-12-3269-2020

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., et al. (2013). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, *5*, 572–597. https://doi.org/10.1002/jame.20038

Goll, D. S., Brovkin, V., Liski, J., Raddatz, T., Thum, T., & Todd-Brown, K. E. O. (2015). Strong dependence of CO_2 emissions from anthropogenic land cover change on initial land cover and soil carbon parametrization. *Global Biogeochemical Cycles*, *29*, 1511–1523. https://doi.org/10.1002/2014GB004988

Gonsamo, A., Chen, J. M., & Lombardozzi, D. (2016). Global vegetation productivity response to climatic oscillations during the satellite era. *Global Change Biology*, *22*, 3414–3426. https://doi.org/10.1111/gcb.13258

Hansis, E., Davis, S. J., & Pongratz, J. (2015). Relevance of methodological choices for accounting of land use change carbon fluxes. *Global Biogeochemical Cycles*, *29*, 1230–1246. https://doi.org/10.1002/2014GB004997

Houghton, R. A., & Nassikas, A. A. (2017). Global and regional fluxes of carbon from land use and land cover change 1850-2015: Carbon emissions from land use. *Global Biogeochemical Cycles*, *31*, 456–472. https://doi.org/10.1002/2016GB005546

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et al. (2017). Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *Journal of Climate*, *30*, 8179–8205. https://doi.org/10.1175/JCLI-D-16-0836.1

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., et al. (2011). Harmonization of land-use scenarios for the period 1500-2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climate Change*, *109*, 117–161. https://doi.org/10.1007/s10584-011-0153-2

Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H., & Núñez-Riboni, I. (2013). Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations: The Model Hamocc within Mpi-Esm in Cmip5. *Journal of Advances in Modeling Earth Systems*, *5*, 287–315. https://doi.org/10.1029/2012MS000178

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*, 4398–4413. https://doi.org/10.1175/jcli-d-12-00554.1

Jones, C. D., Collins, M., Cox, P. M., & Spall, S. A. (2001). The carbon cycle response to ENSO: A coupled climate-carbon cycle model study. *Journal of Climate*, *14*, 17. https://doi.org/10.1175/1520-0442(2001)014<4113:TCCRTE>2.0.CO;2

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*, 1333–1349. https://doi.org/10.1175/BAMS-D-13-00255.1

Lawrence, P. J., Feddema, J. J., Bonan, G. B., Meehl, G. A., O'Neill, B. C., Oleson, K. W., et al. (2012). Simulating the biogeochemical and biogeophysical impacts of transient land cover change and wood harvest in the Community Climate System Model (CCSM4) from 1850 to 2100. *Journal of Climate*, *25*, 3071–3095. https://doi.org/10.1175/JCLI-D-11-00256.1

Le Quéré, C., Andres, R. J., Boden, T., Conway, T., Houghton, R. A., House, J. I., et al. (2013). The global carbon budget 1959–2011. *Earth System Science Data*, *5*, 165–185. https://doi.org/10.5194/essd-5-165-2013

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., et al. (2018). Global carbon budget 2018. *Earth System Science Data*, *10*, 2141–2194. https://doi.org/10.5194/essd-10-2141-2018

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*, 405–448. https://doi.org/10.5194/essd-10-405-2018

Li, H., & Ilyina, T. (2018). Current and future decadal trends in the oceanic carbon uptake are dominated by internal variability. *Geophysical Research Letters*, *45*, 916–925. https://doi.org/10.1002/2017GL075370

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., et al. (2019). The Max Planck Institute Grand Ensemble: Enabling the exploration of climate system variability. *Journal of Advances in Modeling Earth Systems*, *11*, 2050–2069. https://doi.org/10.1029/2019MS001639

Mankin, J. S., Lehner, F., Coats, S., & McKinnon, K. A. (2020). The value of initial condition large ensembles to robust adaptation decision-making. *Earth's Future*, *8*. https://doi.org/10.1029/2020EF001610
Marsland, S. J., Haak, H., Jungclaus, J. H., Latif, M., & Röske, F. (2003). The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates. *Ocean Modelling*, *5*, 91–127. https://doi.org/10.1016/S1463-5003(02)00015-X
Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., et al. (2017). Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geoscientific Model Development*, *10*, 2057–2116. https://doi.org/10.5194/gmd-10-2057-2017
Milinski, S., Maher, N., & Olonscheck, D. (2020). How large does a large ensemble need to be? *Earth System Dynamics*, *11*, 885–901. https://doi.org/10.5194/esd-11-885-2020
Millar, R. J., Fuglestvedt, J. S., Friedlingstein, P., Rogelj, J., Grubb, M. J., Matthews, H. D., et al. (2017). Emission budgets and pathways consistent with limiting warming to 1.5°C. *Nature Geoscience*, *10*, 741–747. https://doi.org/10.1038/ngeo3031
Pongratz, J., Reick, C. H., Houghton, R. A., & House, J. I. (2014). Terminology as a key uncertainty in net land use and land cover change carbon flux estimates. *Earth System Dynamics*, *5*, 177–195. https://doi.org/10.5194/esd-5-177-2014
Reick, C. H., Raddatz, T., Brovkin, V., & Gayler, V. (2013). Representation of natural and anthropogenic land cover change in MPI-ESM. *Journal of Advances in Modeling Earth Systems*, *5*, 459–482. https://doi.org/10.1002/jame.20022
Roeckner, E., Giorgetta, M. A., Crueger, T., Esch, M., & Pongratz, J. (2010). Historical and future anthropogenic emission pathways derived from coupled climate-carbon cycle simulations. *Climatic Change*, *105*, 91–108. https://doi.org/10.1007/s10584-010-9886-6
Rogelj, J., Schaeffer, M., Friedlingstein, P., Gillett, N. P., van Vuuren, D. P., Riahi, K., et al. (2016). Differences between carbon budget estimates unravelled. *Nature Climate Change*, *6*, 245–252. https://doi.org/10.1038/nclimate2868
Schneck, R., Reick, C. H., & Raddatz, T. (2013). Land contribution to natural CO_2 variability on time scales of centuries: Land contribution to CO_2 variability. *Journal of Advances in Modeling Earth Systems*, *5*, 354–365. https://doi.org/10.1002/jame.20029
Scott, D. W. (2015). *Multivariate density estimation: Theory, practice, and visualization*. John Wiley & Sons.
Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*, 591–611. https://doi.org/10.2307/2333709
Sonntag, S., Pongratz, J., Reick, C. H., & Schmidt, H. (2016). Reforestation in a high-CO_2 world—Higher mitigation potential than expected, lower adaptation potential than hoped for. *Geophysical Research Letters*, *43*, 6546–6553. https://doi.org/10.1002/2016GL068824
Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., et al. (2013). Atmospheric component of the MPI-M Earth System Model: ECHAM6: ECHAM6. *Journal of Advances in Modeling Earth Systems*, *5*, 146–172. https://doi.org/10.1002/jame.20015
Suarez-Gutierrez, L., Müller, W. A., Li, C., & Marotzke, J. (2020). Hotspots of extreme heat under global warming. *Climate Dynamics*, *55*, 429–447. https://doi.org/10.1007/s00382-020-05263-w
Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*, 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1
Thomson, A. M., Calvin, K. V., Smith, S. J., Kyle, G. P., Volke, A., Patel, P., et al. (2011). RCP4.5: A pathway for stabilization of radiative forcing by 2100. *Climate Change*, *109*, 77–94. https://doi.org/10.1007/s10584-011-0151-4
UNFCCC. (2015). *Paris Agreement*. Article 14.
UNFCCC. (2017). *Report of the Subsidiary Body for Scientific and Technological Advice on its forty-fifth session, held in Marrakech from 7 to 15 November 2016*.
Walker, A. P., De Kauwe, M. G., Bastos, A., Belmecheri, S., Georgiou, K., Keeling, R., et al. (2021). Integrating the evidence for a terrestrial carbon sink caused by increasing atmospheric CO_2. *New Phytologist*, *229*, 2413–2445.
Yue, C., Ciais, P., Houghton, R. A., & Nassikas, A. A. (2020). Contribution of land use to the interannual variability of the land carbon cycle. *Nature Communications*, *11*, 3170. https://doi.org/10.1038/s41467-020-16953-8
Zhang, Y., Dannenberg, M. P., Hwang, T., & Song, C. (2019). El Niño-Southern Oscillation-induced variability of terrestrial gross primary production during the satellite era. *Journal of Geophysical Research: Biogeosciences*, *124*, 2419–2431. https://doi.org/10.1029/2019JG005117
Zhu, Z., Piao, S., Yan, T., Ciais, P., Bastos, A., Zhang, X., & Wang, Z. (2018). The accelerating land carbon sink of the 2000s may not be driven predominantly by the Warming Hiatus. *Geophysical Research Letters*, *45*, 1402–1409. https://doi.org/10.1002/2017GL075808

References From the Supporting Information

Bastos, A., Friedlingstein, P., Sitch, S., Chen, C., Mialon, A., Wigneron, J.-P., et al. (2018). Impact of the 2015/2016 El Niño on the terrestrial carbon cycle constrained by bottom-up and top-down approaches. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*, 20170304. https://doi.org/10.1098/rstb.2017.0304
Poulter, B., Frank, D., Ciais, P., Myneni, R. B., Andela, N., Bi, J., et al. (2014). Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle. *Nature*, *509*, 600–603. https://doi.org/10.1038/nature13376