

# Narrowing the Range of Future Climate Projections Using Historical Observations of Atmospheric CO<sub>2</sub>

BEN B. B. BOOTH, GLEN R. HARRIS, AND JAMES M. MURPHY

*Met Office Hadley Centre, Exeter, United Kingdom*

JO I. HOUSE

*Cabot Institute, Department of Geography, University of Bristol, Bristol, and College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom*

CHRIS D. JONES AND DAVID SEXTON

*Met Office Hadley Centre, Exeter, United Kingdom*

STEPHEN SITCH

*College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom*

(Manuscript received 29 February 2016, in final form 8 November 2016)

## ABSTRACT

Uncertainty in the behavior of the carbon cycle is important in driving the range in future projected climate change. Previous comparisons of model responses with historical CO<sub>2</sub> observations have suggested a strong constraint on simulated projections that could narrow the range considered plausible. This study uses a new 57-member perturbed parameter ensemble of variants of an Earth system model for three future scenarios, which 1) explores a wider range of potential climate responses than before and 2) includes the impact of past uncertainty in carbon emissions on simulated trends. These two factors represent a more complete exploration of uncertainty, although they lead to a weaker constraint on the range of future CO<sub>2</sub> concentrations as compared to earlier studies. Nevertheless, CO<sub>2</sub> observations are shown to be effective at narrowing the distribution, excluding 30 of 57 simulations as inconsistent with historical CO<sub>2</sub> changes. The perturbed model variants excluded are mainly at the high end of the future projected CO<sub>2</sub> changes, with only 8 of the 26 variants projecting RCP8.5 2100 concentrations in excess of 1100 ppm retained. Interestingly, a minority of the high-end variants were able to capture historical CO<sub>2</sub> trends, with the large-magnitude response emerging later in the century (owing to high climate sensitivities, strong carbon feedbacks, or both). Comparison with observed CO<sub>2</sub> is effective at narrowing both the range and distribution of projections out to the mid-twenty-first century for all scenarios and to 2100 for a scenario with low emissions.

## 1. Introduction

Most current socioeconomic storylines project future increases in emissions of radiatively active greenhouse gases over the next century (Van Vuuren et al. 2011). Considerable advances have been made in understanding

both the processes that determine the fraction of these emissions that remain in the atmosphere and the physical feedbacks that determine the climate response to these changes. However, current uncertainties in the role of many of these processes lead to a broad spread of projected climate changes for a set of future emissions under a given socioeconomic storyline. At the same time, decision-makers are looking for greater certainty in the magnitude of projected changes. A key factor contributing to future projection spread concerns processes that control the uptake of atmospheric carbon into the land and ocean. This is evident from the projected spread of future global CO<sub>2</sub> changes in multimodel carbon cycle

---

Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JCLI-D-16-0178.s1>.

---

Corresponding author e-mail: Ben B. Booth, [ben.booth@metoffice.gov.uk](mailto:ben.booth@metoffice.gov.uk)

DOI: 10.1175/JCLI-D-16-0178.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

ensembles from the Coupled Climate Carbon Cycle Model Intercomparison Project (C4MIP; Friedlingstein et al. 2006) and phase 5 of the Coupled Model Intercomparison Project (CMIP5; Arora et al. 2013) as well as in general circulation model (GCM) experiments designed to explore carbon cycle parametric uncertainty in a single model (Booth et al. 2012). Uncertainty in carbon cycle processes is one of the dominant sources of spread in current global climate projections (e.g., Bodman et al. 2013; Harris et al. 2013; Tachiiri et al. 2013). This spread carries through directly into regional climate projections. For example, in the most recent climate projections that underpin U.K. adaptation planning (UKCP09; Murphy et al. 2009), we found that the carbon cycle contributed a substantial amount to the total spread for many climate variables and locations. Because of this there is a considerable interest in identifying criteria by which these ranges may be narrowed, by using observations to rule out less plausible simulations.

There are three current approaches to making better use of observations that are being pursued by the Earth system modeling community: benchmarking, process evaluation, and identification of emergent constraints. The first aims to build a set of observed metrics against which Earth system models are routinely evaluated. Currently such evaluation lags behind analogous benchmarking in physical climate model developments, but progress is being made (e.g., Blyth et al. 2011; Luo et al. 2012). The second approach focuses on improvements in model processes (and hence in increased confidence in the model response). For example, Cadule et al. (2010) focus on the ability of simulations to capture observed seasonal and interannual variations in atmospheric CO<sub>2</sub>, variations that are driven by changes in temperature and moisture that would also be expected to act under a changing climate. By developing models that are better able to capture the exchange of carbon between the land/oceans and the atmosphere to large variations in short-term drivers of change (from diurnal up to decadal), greater confidence can be placed on their ability to capture responses to future long-term changes. Indeed, insights such as these can also feed into emergent constraints. The interannual CO<sub>2</sub> responses to temperature have been recently linked to the magnitude of the tropical carbon cycle response to climate-driven changes (Cox et al. 2013). How tight this relationship is has been questioned (Wang et al. 2014), but the constraint would still appear to indicate that larger future tropical carbon-climate feedbacks are less realistic.

Emergent constraints represent a way forward by linking key observable properties of the real system to future response, allowing us to reduce the range of projected future climate changes. Cox et al. (2013)

represent a bottom-up approach that narrows the range of model processes that can be considered plausible. Other approaches take a top-down or integrated view to look at how observables can constrain emergent properties of the climate system. One such observable, historical CO<sub>2</sub> concentration could be considered a top-down constraint since it depends on the emergent response to a large number of processes and interactions in the climate system. Comparisons with observed changes in atmospheric CO<sub>2</sub> have been explored in both simple climate models (SCMs) [energy balance models (EBMs) combined with a simple global carbon cycle representation] and full GCMs or Earth system models. Ricciuti et al. (2008), Tachiiri et al. (2013), and Bodman et al. (2013) have shown the utility of atmospheric CO<sub>2</sub> measurements in narrowing the projected range of future climate changes from SCMs. The strength of simple model frameworks is that, computationally, they can explore a wide range of potential feedbacks. Their usefulness comes from their ability to reproduce the gross behavior of more complex models using simple, often linear, relationships. There is an interest, therefore, in whether CO<sub>2</sub> measurements are also effective in narrowing GCM responses. A number of recent papers using full-complexity general circulation models (Friedlingstein et al. 2014; Hoffman et al. 2014; Murphy et al. 2014) identified relationships between simulated present-day CO<sub>2</sub> biases and projected CO<sub>2</sub> changes at the end of the century. The first of these (Friedlingstein et al. 2014) showed that 7 out of 11 emission-driven CMIP5 GCM simulations tended to project larger future concentrations of CO<sub>2</sub> than the standard reference RCP CO<sub>2</sub> concentration. Friedlingstein et al. (2014) cautioned, however, that this should not be interpreted as the RCP reference concentration being biased low, because these models projecting larger future concentrations tended to be biased high in the present day. Friedlingstein et al. (2014) do not go quite as far as interpreting the historical comparison as a constraint on future CO<sub>2</sub> change, pointing to a potential for compensating ocean and land carbon errors to lead to apparently good agreement in atmospheric CO<sub>2</sub>. In contrast, Hoffman et al. (2014) argue that just such a comparison with historical CO<sub>2</sub> measurements can be interpreted as a constraint on future changes. This is because long time scales inherent in the carbon cycle system mean that present-day biases can be expected to remain persistent throughout the twenty-first century, irrespective of how the simulations balance atmospheric CO<sub>2</sub> with the ocean and land carbon stores. By using the CMIP5 relationship between historical and future CO<sub>2</sub> Hoffman et al. (2014) obtain a narrow observationally constrained estimate of RCP8.5 CO<sub>2</sub> of  $947 \pm 35$  ppm ( $1.96\sigma$ ) in 2100.

In this paper we return to the question of how much information the observed CO<sub>2</sub> changes have in potentially narrowing the range of future GCM CO<sub>2</sub> projections. We extend the earlier analysis (Friedlingstein et al. 2014; Hoffman et al. 2014) by 1) using trends instead of absolute concentrations, 2) using a GCM ensemble that is designed to sample a wider range of physical and carbon cycle feedbacks (Lambert et al. 2013) while still producing credible simulations of present-day climate (Murphy et al. 2014), and 3) including an estimate of the impact of uncertainty in past carbon emissions (from land use and fossil fuels) on historical simulated CO<sub>2</sub> trends. This latter factor either has not been accounted for in previous models (e.g., Lambert et al. 2013) or has been only partially sampled (CMIP5). One motivation for this work is to explore how observed CO<sub>2</sub> records could be used in the future to constrain GCM climate projections, or national climate scenarios derived from them, such as future updates to UKCP09 (Murphy et al. 2009). In addition, insights from this work have wider implications for how we consider the role of carbon cycle processes in future changes. In this paper we show how including a wider range of potential model feedbacks and including estimates of past carbon emission uncertainty can both broaden the distribution of projected future CO<sub>2</sub> that can be considered plausible, compared to previous studies.

## 2. Model simulations and relationships between past and future CO<sub>2</sub>

A perturbed parameter ensemble (PPE) of the general circulation model based on the carbon cycle configuration of HadCM3 (HadCM3C) was used in this study. HadCM3C (Booth et al. 2012) consists of 57 individual model variants that sample uncertainties in the atmospheric physics, ocean physics, aerosol formation and removal pathways, and land carbon cycle (Lambert et al. 2013). These model variants were picked following a process designed to avoid the risk of using implausible areas of parameter space (Collins et al. 2011; Lambert et al. 2013) and are able to simulate historical climate to a level comparable with CMIP3 and CMIP5 models (Murphy et al. 2014). Each of the 57 model variants was driven by the historical and RCP8.5 emission-driven scenarios. These are referred to as the E-driven historical and RCP8.5 in Taylor et al. (2012) and esmHistorical and esmrcp85 in the CMIP5 data archive description. In addition, these 57 model variants have also been used to simulate a midrange emission scenario (SRES A1B) and an aggressive mitigation scenario (RCP2.6). The implementation of boundary conditions for these is described in more detail in Booth et al. (2013).

The parametric uncertainties explored in each of the PPE components (atmospheric and ocean physics, land carbon and aerosol sulfur cycle) were elicited to explore uncertainties in different processes. Individual parameters were identified that were both uncertain and important for a process response. The two components most relevant to global CO<sub>2</sub> projections are uncertainties in the atmospheric and land carbon components (Lambert et al. 2013). The atmospheric component explored the most parameters (29 parameters, selected from seven physics schemes: boundary layer, convection, land surface, dynamics, large-scale cloud, radiation, and sea ice; Collins et al. 2011; Lambert et al. 2013). The most important ones in terms of influence on projected CO<sub>2</sub> changes are parameters that influence the magnitude of temperature response (mainly in convection and cloud schemes), but there may be other influences via parameters that impact rainfall distributions, land surface, and land–sea contrasts. Six land surface parameters were explored in the land carbon cycle (Booth et al. 2012). These explored uncertainties in leaf nitrogen (and its associated impact on photosynthesis), sensitivity of stomatal opening to CO<sub>2</sub> changes, temperature dependence of photosynthesis (via the carboxylation pathway) and temperature dependence of soil respiration ( $Q_{10}$ ), soil moisture controls on evaporation, and a competition parameter that influenced the simulated vegetation distributions. These parameters affect different processes in the land surface and lead to a broad range of land carbon responses to climate change. These are discussed in detail in Booth et al. (2012), who show that the temperature dependence of photosynthesis represents the most important uncertainty for global CO<sub>2</sub> with secondary influences from the other parameters.

Previous assessment of carbon cycle models included analysis of Earth system model simulations in the E-driven historical and RCP8.5 scenarios (Taylor et al. 2012) of the current CMIP5. For comparison with these studies, data from the emission-driven historical (esm-Historical) and RCP8.5 (esmrcp85) scenarios were downloaded from the Earth System Grid Federation (Taylor et al. 2012). Where two models shared very similar formulation (GFDL-ESM2G and GFDL-ESM2M shared common carbon cycle and atmospheric representation, differing in their ocean coordinate system) then only one model is shown (GFDL-ESM2G in this case). The CMIP5 models used in this analysis were BNU-ESM, CanESM2, CESM1(BGC), GFDL-ESM2G, HadGEM2-ES, INM-CM4, ISPL-CM5A-LR, MIROC-ESM, MPI-ESM-LR, and MRI-ESM1.

These GCM ensembles do show a positive correlation between historical trends in CO<sub>2</sub> and the magnitude of

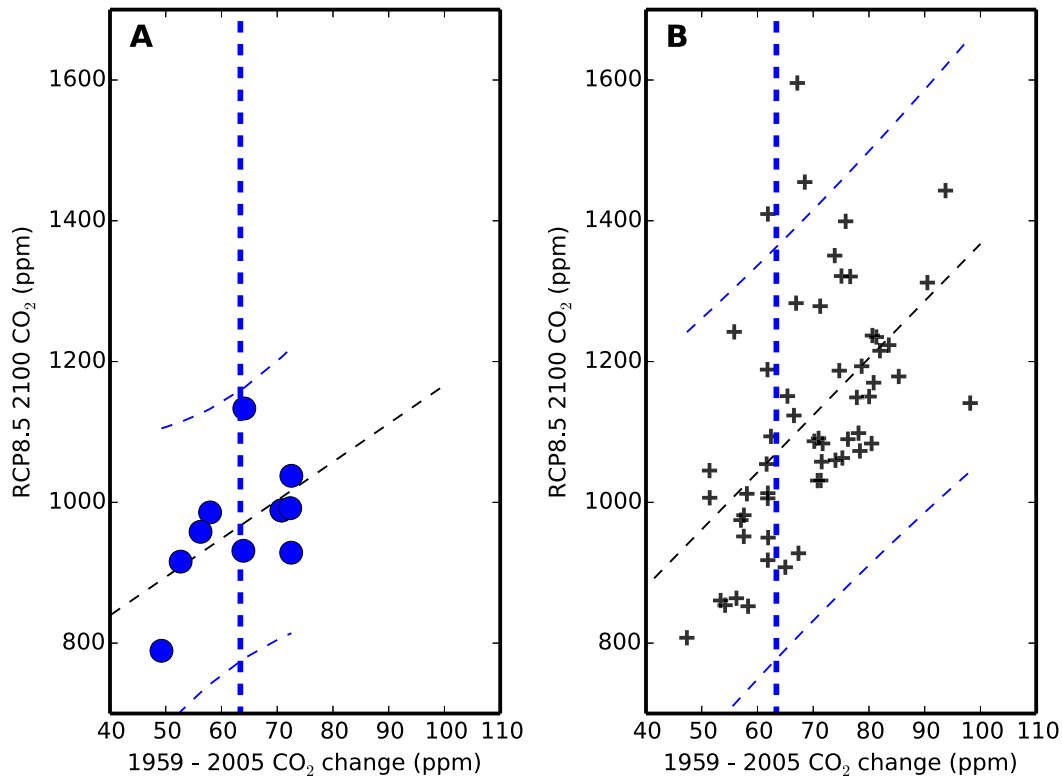


FIG. 1. The relationship between simulated end-of-century CO<sub>2</sub> concentration under RCP8.5 and the simulated historical change in CO<sub>2</sub> during 1959 and 2005. (a) Future concentrations and historical (1959–2005) trends are shown for nine CMIP5 models (BNU-ESM, CanESM2, CESM1(BGC), GFDL-ESM2G, HadGEM2-ES, INM-CM4, ISPL-CM5A-LR, MIROC-ESM, MPI-ESM-LR, and MRI-ESM1) along with a best-fit regression line (black dashed) and confidence interval (thin blue dashed). (b) As in (a), but for the 57-member PPE in this study. Best-fit regression lines (black dashed) and confidence limits (thin blue dashed) are shown for both panels. Historical observations (thick blue dashed) are based on observed concentration data provided as part of the CMIP5 driving historical dataset (Taylor et al. 2012).

projected CO<sub>2</sub> in 2100 under RCP8.5. Figure 1a shows this relationship for the CMIP5 simulations. There are some differences compared to Hoffman et al. (2014), who used the modeled–observed CO<sub>2</sub> mole fraction for a single year (2010), whereas we compare to the contemporary CO<sub>2</sub> trends (1959–2005). Using the trends as a comparison, the simulation with the highest future CMIP5 projection lies closer to the observed estimate than the Hoffman et al. (2014) analysis using 2010 CO<sub>2</sub> concentrations. Qualitatively, however, this comparison of CO<sub>2</sub> trends shows a similar relationship to that in Hoffman et al. (2014), with larger future changes associated with larger historical trends. Also shown (Fig. 1b) is the same comparison for the 57-member PPE of GCM simulations, designed to sample a range of physical and carbon cycle feedbacks (Lambert et al. 2013; Murphy et al. 2014). In this larger ensemble, there is also a positive correlation between models with larger historical trends and 2100 CO<sub>2</sub> concentrations, with a 0.53 Pearson correlation coefficient compared to 0.54 for CMIP5

models. The larger ensemble size of the PPE enables an estimate of the relationship between past and future trends that is statistically significant ( $p$  value =  $2.2 \times 10^{-5}$  vs 0.11 for CMIP5).

The emission-driven PPE ensemble does have a number of evident differences compared to its CMIP5 counterpart. The most obvious is that a number of the PPE model variants project significantly larger future CO<sub>2</sub> concentrations than found for the CMIP5 models. This wider range is a result of the experimental design of the PPE (Lambert et al. 2013), where the GCM configurations were selected to sample a wide range of physical and carbon cycle feedbacks. The CMIP5 ensemble, in contrast, represents a collection of GCMs developed by individual modeling centers to produce “best estimates” of the potential physical and carbon cycle processes. At the same time, the CMIP5 ensemble explores a broader range of model structures (process uncertainty), which might be expected to contribute to broader spread. It is interesting, therefore, that the PPE

produces a wider diversity of model responses (suggesting perhaps that efforts to produce best estimates appear to outweigh any increase in spread due to different model structures). In contrast, models at the higher end of the PPE range result from exploring wider ranges for climate sensitivity and carbon cycle feedbacks and from interactions between them (Lambert et al. 2013). The PPE, for example, samples a number of climate sensitivities that lie above the largest of the CMIP5 models considered here but still in the tails of what is considered plausible in the IPCC distributions (Booth et al. 2013). It should be noted that the IPCC “likely” estimates include climate sensitivity values as low as 1.5 K; however, neither CMIP5 nor the PPE explore climate sensitivities below 2 K. This is a factor that may potentially lead to an undersampling of past and future CO<sub>2</sub> changes in the low end of the model distributions (Fig. 1). The other difference evident in the comparison is that the PPE gives a wider range of possible future CO<sub>2</sub> concentrations that are still relatively consistent with the observed historical trend, based on the subset of ensemble members that are closest to the observed change (e.g., within 10 ppm). For example, PPE simulations can be found that match the observed historical trend yet simulate 2100 concentrations larger than all members of the emission-driven CMIP5 ensemble. The experimental design (Lambert et al. 2013) and limited number of experiments mean that it is not possible to link this wider behavior to individual parameters. However, our results suggest (see supplementary Fig. S1 and accompanying discussion) that it is the land carbon cycle formulation (rather than other uncertainties such as in climate sensitivity) that is the most important factor in exploring higher future climate changes that are not evidently inconsistent with past changes. While these historically plausible but high future global CO<sub>2</sub> simulations are noteworthy, the more general finding is that the majority of PPE high-end 2100 concentrations arise in model variants that tend to overestimate the historical trend, in common with CMIP5.

### 3. Methodologies for analysis of constraints

#### a. Past anthropogenic emission uncertainties

Past anthropogenic CO<sub>2</sub> emissions have arisen from two major sources, burning of fossil fuels (FF) and carbon released to the atmosphere due to past land-use changes (LUC). Estimates of FF are more closely constrained than estimates of LU change contributions. This is because inventory estimates of past FF usage are available, on a sector-by-sector and country-by-country basis. Le Quéré et al. (2015) estimate the uncertainty in

recent FF emissions at 5% of the total emissions. In absolute magnitude, this represents a small uncertainty for most of the historical period but becomes more significant as FF emissions have ramped up in recent decades and with coal representing a larger fraction of the energy mix (roughly 0.5 GtC yr<sup>-1</sup> in 2010; see Fig. 2a).

LUC emissions are based on changes in land cover and, to a variable extent, land management practices (Houghton 2003). Uncertainties in calculated LUC change arise from difficulties in estimating historical land-use and land-cover change, the carbon density of vegetation and soils, different definitions and methodological approaches, and representation of processes such as fire and nitrogen cycling, among other things (Houghton et al. 2012; Pongratz et al. 2014). During the last decade the uncertainty from historical land-cover and land management change amounted to around 0.5 GtC yr<sup>-1</sup>, comparable to the uncertainty associated with FF (Ciais et al. 2013; Le Quéré et al. 2015). However, during the earlier decades, LUC emissions were a relatively larger fraction of the total emissions, and it is this LUC uncertainty that dominates the total anthropogenic CO<sub>2</sub> emission uncertainty for most of the historical period (Brovkin et al. 2004).

Figure 2b shows estimates of historical LUC emissions, published in Houghton et al. (2012), based on estimates from dynamic global vegetation models (DGVMs)—spatially resolved land surface models (of a complexity akin to that found in the state-of-the-art GCMs) coupled to simplified energy balance and ocean carbon cycle components. These explore two kinds of LUC uncertainty. The first arises from differences in estimates of past land-cover change, sampled by driving a single DGVM, the integrated science assessment model (ISAM; Jain et al. 2013) using three different estimates of land-cover change, based on HYDE (Goldewijk et al. 2011), SAGE (Ramankutty and Foley 1999), and Houghton (Houghton et al. 2012). The second is due to differences in the representation of land carbon cycle processes between different DGVMs. This is sampled by differences in four DGVMs [LPJ-wsl (Poulter et al. 2010); BernCC (Stocker et al. 2011); VISIT (Kato et al. 2013); ISAM (Jain et al. 2013)] driven by a common land-cover dataset, HYDE (Goldewijk et al. 2011). We assume that the land cover and DGVM process uncertainties are independent. We combine these two uncertainties into 12 LUC emission time series assuming that fractional differences between the three different land-cover datasets can be added to fractional differences between the four DGVM emissions (Fig. 2b). The 12 time series explore a range of mean 2001–06 LUC emissions between 0.37 and 1.64 Gt yr<sup>-1</sup>.



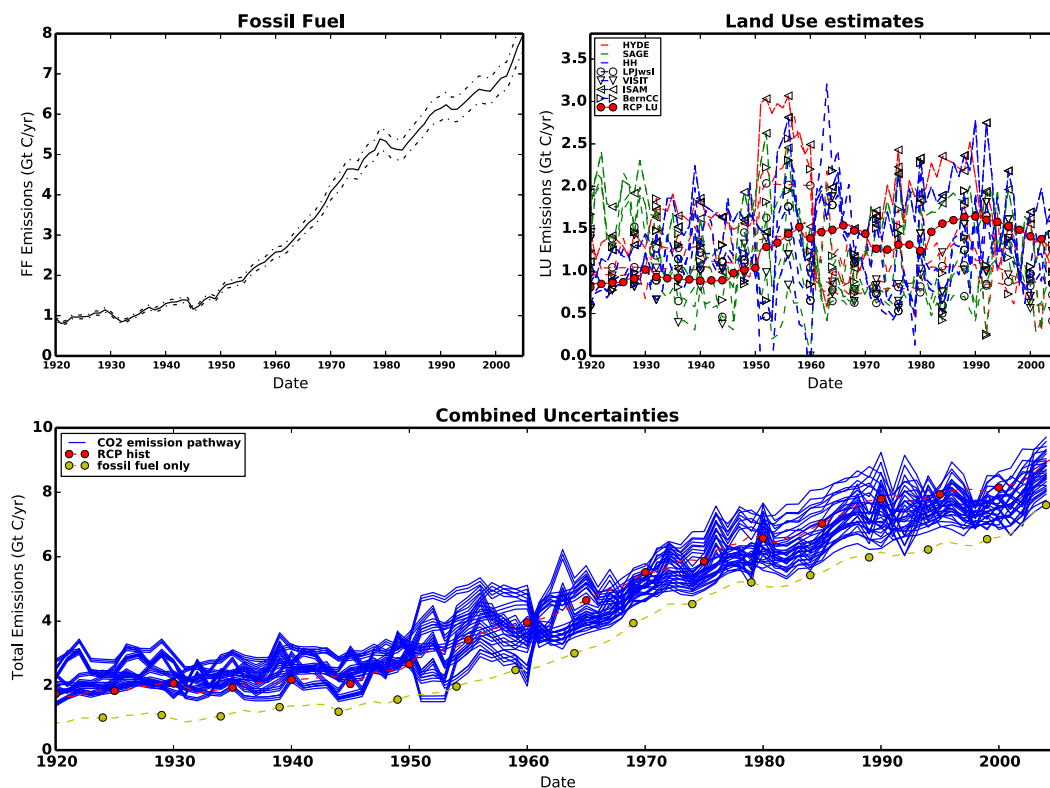


FIG. 2. Sampling uncertainty in  $\text{CO}_2$  emissions. (a) The historical  $\text{CO}_2$  emissions arising from FF with the central estimate (solid) and the estimated two-standard-deviation uncertainty (dashed). (b) The estimates of LU contributions estimated from DGVMs. The time series represent two types of uncertainty, arising from uncertainty in reconstructions of past land-cover change (colors) and differences in land surface models (line style). The three land-cover datasets are based on HYDE (Goldewijk et al. 2011), SAGE (Ramankutty and Foley 1999), and Houghton (Houghton et al. 2012) and used as input for ISAM DGVM. The four DGVMs are LPJwsl, VISIT, BernCC, and ISAM. These two types of uncertainty are combined to produce 12 time series (using the approach described in section 2). Standard RCP LU emissions, based on Houghton et al. (2012)'s bookkeeping estimate, are shown for comparison (red). (c) How the three FF time series [from (a)] are combined with the 12 LUC time series [from (B)] to make 36 total anthropogenic  $\text{CO}_2$  emission time series that were subsequently used in the rerun simulations of historical  $\text{CO}_2$  trends. Also shown, for comparison, are the estimates historical emissions used within CMIP5 (red) and the FF component (central estimate) of the emissions (yellow).

Uncertainties in the components of past emissions represent different things. The FF emission uncertainties can be expressed as a standard deviation, and hence a Gaussian error model could be fitted if desired. On the other hand, current LUC cover estimates arise from a number of factors—some data driven (such as land-cover dataset differences explored here), some dependent on representation of land surface processes in models, some methodological (how these changes are implemented in the models), and others more fundamental, relating to the way we define land-use emissions (Pongratz et al. 2014). Given the very limited sample, it would be hard to justify translating this into a Gaussian uncertainty estimate in a comparable sense. Here we simply combined each LUC emissions estimate with a high, standard, and low FF emissions estimate ( $\mu + 2\sigma$ ,  $\mu$ , and  $\mu - 2\sigma$  respectively; Fig. 2a), and explored

whether the observed  $\text{CO}_2$  trend lies within the simulated spread for each PPE member. In total 36 emission time series are used in this paper, which we obtained by combining the 12 LUC emission time series with these 3 FF emissions time series (Fig. 2c).

#### b. Simple climate model

Ideally each of the individual PPE variants would have been rerun 36 times, sampling each of the emission time series described above. The computational costs of doing so with the full-complexity Earth system model is prohibitive, so instead the historical response to each of these emission time series is approximated using an SCM tuned to replicate each of the HadCM3C PPE variants. The SCM used here is described in detail in the supplementary material to Harris et al. (2013). Here we summarize its main features, describe the calibration of

its input parameters, and specify how radiative forcing is applied.

A two-box EBM is used to predict changes in the globally averaged land and ocean surface air temperatures in response to global average radiative forcing of the climate. Ocean heat uptake is represented by a vertical diffusion–advection equation in which the thermal diffusivity is determined separately for each of the perturbed ocean parameter configurations (Collins et al. 2011) that contribute to the PPE design. The land and ocean climate feedback parameters  $\lambda_L$  and  $\lambda_O$  used to configure the EBM are assumed to depend only on the atmospheric configuration and not vary when the carbon cycle, sulfur cycle (aerosols), and ocean parameters are perturbed. They are diagnosed, therefore, from the different atmospheric GCM configurations, using simulations forced by a 1% yr<sup>-1</sup> increase in CO<sub>2</sub> concentration. This is achieved by estimating  $\lambda_{O,L}$  as the regression coefficient between the radiative forcing and top-of-atmosphere radiative response  $\Delta Q - N_{O,L}$  and the change in surface temperature  $\Delta T_{O,L}$  over both ocean and land (Murphy 1995; Forster and Taylor 2006), where  $\Delta Q_{CO_2} = 5.4 \ln\{[\text{CO}_2(t)]/[\text{CO}_2(0)]\} \text{ W m}^{-2}$  (Myhre et al. 1998) and  $N_{O,L}$  refers to the average ocean and land net downward top-of-the-atmosphere radiative fluxes.

A carbon cycle component is implemented in the SCM, which allows the atmospheric concentrations to be calculated for prescribed scenarios of CO<sub>2</sub> emissions. Uptake and release of CO<sub>2</sub> from the vegetation and soil land reservoirs is estimated in a similar manner to that described in Jones et al. (2006), while uptake of CO<sub>2</sub> by the oceans is represented using the impulse-response method of Joos et al. (1996). The SCM contains parameters that control net primary production (NPP), litter production, soil respiration, and ocean CO<sub>2</sub> uptake. Optimal values required for the SCM to reproduce each of the 57 GCM PPE variants are determined by varying these SCM parameters and identifying SCM parameter sets that minimize the error in reproducing each of these four CO<sub>2</sub> fluxes (NPP, litter, soil respiration, and ocean CO<sub>2</sub> uptake) in the corresponding GCM (Harris et al. 2013). One difference from Harris et al. (2013) is that this study calibrates the carbon cycle parameters of the SCM using CO<sub>2</sub> fluxes diagnosed directly from the 57 PPE simulations themselves [Harris et al. (2013) fitted instead to a separate but related experiment described in Booth et al. (2012)]. Another calibration difference is that mean square errors for four scenarios (RCP2.6, RCP8.5, A1B, and historical) are jointly used to determine the carbon parameters for each SCM configuration. The non-CO<sub>2</sub> forcing (which includes other greenhouse gases, aerosol, and solar and volcanic forcing) is diagnosed directly from the global

surface temperature response  $\Delta T$  and net TOA radiative imbalance  $N$  in the PPE simulations, subtracting off forcing implied by the diagnosed CO<sub>2</sub> concentrations; that is,  $\Delta Q_{\text{nonCO}_2} = N + \lambda \Delta T - \Delta Q_{\text{CO}_2}$  (where  $\Delta Q_{\text{CO}_2}$  is calculated as described above). The evolution of the coupled CO<sub>2</sub> and temperature responses for each member of the PPE to each of the new 36 historical CO<sub>2</sub> emission time series is predicted by the SCM using these estimates for the non-CO<sub>2</sub> forcing, with climate feedbacks, ocean thermal diffusivity, and carbon cycle configuration obtained for the corresponding members in the component ensembles (Lambert et al. 2013).

The simple model is used to explore what historical CO<sub>2</sub> changes each of the 57 PPE variants would have simulated if they had instead been driven by the 36 historical emission time series (they are not used to explore future emission uncertainty in this analysis). The SCM provides a close but not perfect reproduction of the GCM carbon cycle response. For each SCM–GCM/PPE pair there is a small bias in the SCM's reproduction of the GCM's CO<sub>2</sub> using the standard CMIP5 historical emission time series. We assume that this is systematic for each pairing and add/subtract this bias from each of the 36 historical CO<sub>2</sub> simulated by the SCM for this study (i.e., the adjustment is time dependent and is assumed independent of historical emission uncertainty). The simple model is therefore able to explore what CO<sub>2</sub> changes each model variant (from the PPE) would have produced had it been instead driven by each of the alternative historical emission time series presented here.

### c. Choice of historical time period

In this study we choose to compare the simulated and observed trends for the period 1959 to 2004. The following criteria were used to select this period.

- The comparison would be based on trends, not absolute CO<sub>2</sub> in a specified year (or years), since predicted GCM CO<sub>2</sub> values for a given year can vary owing to differences in their initial preindustrial state, as well as their response to carbon and climate change. Here, we want to focus on metrics that highlight the latter and not the former.
- The length of the trend period should be maximized, while still falling within the Keeling record (1959–present). The CO<sub>2</sub> reference concentration compared to both the PPE and emission-driven CMIP5 GCMs was based on CMIP5 historical datasets (Taylor et al. 2012), representing a merged record based on both Mauna Loa direct measurements and ice-core records to obtain a globally representative time series.
- The selected period should maximize the information available from datasets on past emissions (currently

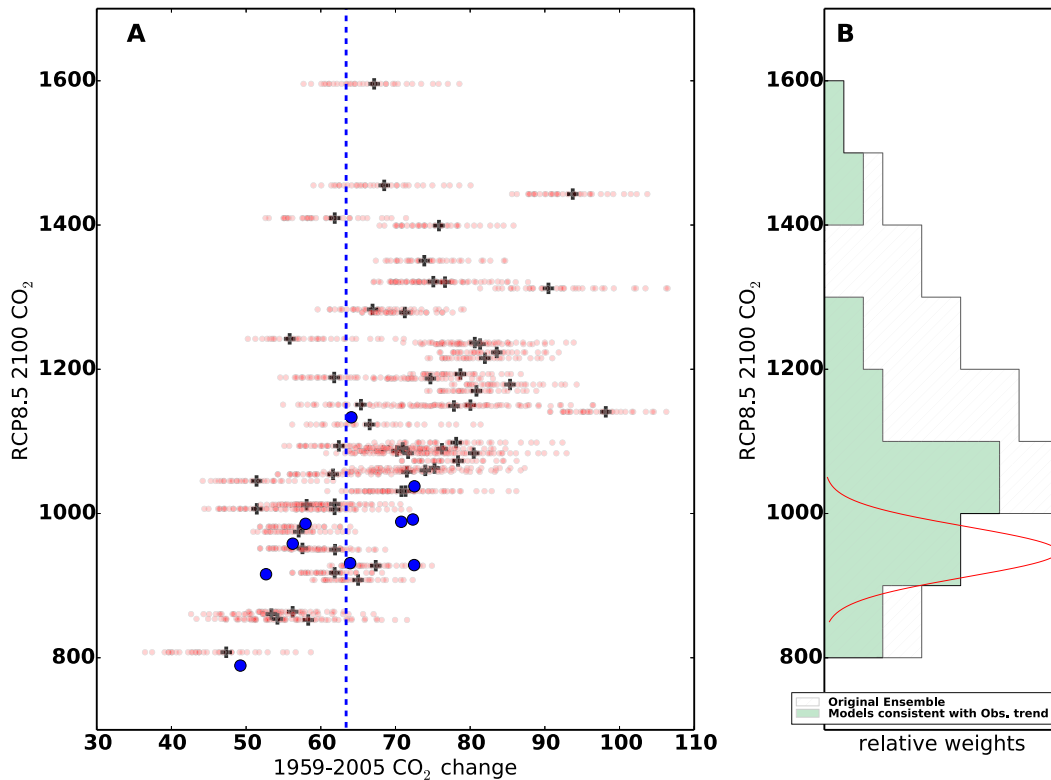


FIG. 3. The relationship between simulated historical  $\text{CO}_2$  trends and the magnitude of projected  $\text{CO}_2$  concentration in 2100 for the RCP8.5 scenario. (a) Black crosses show the GCM  $\text{CO}_2$  concentration in 2100 plotted against the projected historical  $\text{CO}_2$  trend over the period 1959–2005 for each of the 57 PPE configurations, while red dots indicate the historical trends that would have been simulated if it had been driven by each of the 36 historical emission time series. The spread of historical simulations (red dots) provides information on whether or not the observed trend (blue horizontal line) would have fallen within the distribution of historical trends for a particular model configuration had these simulations sampled anthropogenic  $\text{CO}_2$  emission uncertainty. CMIP5 models (blue dots) are included for comparison. (b) The histogram of RCP8.5  $\text{CO}_2$  concentration by the end of the century for the PPE (hatched), alongside the histogram of projected changes for only those model variants that envelop the observed historical  $\text{CO}_2$  trend (green). The resulting histogram is more tightly constrained, thinning out many, but not all, of the higher-end projected changes. For comparison, the red Gaussian distribution illustrates the constrained  $\text{CO}_2$  distribution developed by Hoffman et al. (2014) from a slightly larger set of CMIP5 models.

all emission datasets, described in section 2a, cover the period 1920–2005).

In this work, the last two criteria (on availability of past LUC uncertainty information and length of the Mauna Loa record) had the largest impact on our choice of period; 1959 to 2005 represents the period where all the LUC datasets provide data while also satisfying the other criteria. The effects of this choice on our results are discussed below in section 5.

#### 4. Influence of land-use uncertainty on $\text{CO}_2$ constraints

The SCM provides useful insight into the range of historical atmospheric  $\text{CO}_2$  trends that we would have expected if the GCMs had been rerun with these

different historical land-use emission time series. Figure 3a updates the constraint evidence shown in Fig. 1b with this additional information. For each of the 57 GCM variants, the 36 historical simulated concentration trends are shown (red dots) relative to the original GCM (black cross). Uncertainty in the past carbon emissions leads to a spread in the historical  $\text{CO}_2$  concentration trends (red dots in Fig. 3a). As a consequence including emission uncertainty enables the question of consistency between  $\text{CO}_2$  observations and simulations to be evaluated based on whether the range of historical trends overlap with the observations.

Similarly to the previous papers on  $\text{CO}_2$  constraints (Friedlingstein et al. 2014; Hoffman et al. 2014), we find that ruling out models that can be considered inconsistent with historical data leads to a shift to a distribution



of lower projected change in 2100 (Fig. 3b). Although the impact of applying our constraint is less evident on the total range of CO<sub>2</sub>, especially for RCP8.5 at the end of the century, this is mainly due to just a few model variants with large future increases in CO<sub>2</sub> but historically consistent responses. There is, however, a noticeable and robust impact on the distribution of responses since many of the 30 simulations inconsistent with historical trends are from the top end of the distribution, while models in the lower and central parts of the distribution remain included. So, for RCP8.5, 18 of the 26 models projecting changes above 1100 ppm (in 2100) are excluded. This leads to a reduction of the interquartile range of RCP8.5 from 1012–1223 ppm down to 951–1137 ppm in 2100, for A1B from 716–865 ppm down to 692–792 ppm, and for RCP2.6 from 423–472 ppm down to 417–445 ppm.

We can extend this analysis to look at other future periods and other emission scenarios (Fig. 4). The full range of CO<sub>2</sub> projections, as well as the factors that lead to future spread, has previously been documented in Booth et al. (2013) for three scenarios: RCP8.5, SRES A1B, and RCP2.6. Figure 4 illustrates the impact on the CO<sub>2</sub> projections when the 30 GCMs inconsistent with the historical trends are removed. This demonstrates that the historical CO<sub>2</sub> trend is much more effective at narrowing the range of projected CO<sub>2</sub> changes earlier in the century, and for lower emission scenarios throughout, than it is for RCP8.5 in 2100. High-end projected changes can be ruled out for all scenarios during the first half of this century. The 90% confidence ranges for 2050 CO<sub>2</sub> concentration are reduced from 475–750, 475–700, and 375–525 ppm down to 500–675, 475–625, and 375–500 ppm for RCP8.5, A1B, and RCP2.6, respectively. The reasons why the historical record is more effective at narrowing the absolute range of future changes in this earlier part of the century is related to time scales. The processes that lead to changes in the effectiveness of either the land or ocean carbon cycles have not had a chance to evolve much over the 45 years following the historical constraint period, in contrast to RCP8.5 projections in the latter part of the coming century. It is less obvious why the high-end 2100 CO<sub>2</sub> projected changes are more effectively ruled out in low emission scenarios, compared to RCP8.5. None of the simulations in the upper third of the RCP2.6 range for our ensemble for the full twenty-first-century period can be considered consistent with historical trends. For A1B, a central business-as-usual scenario, only a single simulation projects changes in the upper third of the distribution that is consistent with the historical trend. We discuss the factors that lead to different levels of constraint across scenarios in section 6.

## 5. Robustness of historical comparison

The analysis in this paper used a 1959-to-2005 time period (see section 3c for discussion). Repeating this analysis using alternative time periods does lead to changes in the details of which specific models can be excluded, which is discussed in this section. However, the wider qualitative conclusions of this paper do appear to be robust to the choice of time period. Figure 5a shows the historical CO<sub>2</sub> changes (with respect to 2005) for both observations and three model variants chosen to illustrate the time dependence. Because of the uncertainty in past carbon emissions explored here, each of the variants produces a range of historical trends (colored plumes in Fig. 5a). For many of the PPE members, the range of simulated trends remains inconsistent with the observed trend (such as model 3 in Fig. 5a) regardless of the time period chosen. At the same time, very few model variants (anywhere in the distribution of future responses) remain consistent with the observations for the whole time series. Observed and model CO<sub>2</sub> show different evolution through the historical time period and the observations can often fall within the distribution of potential simulated trends for one period and lie just outside for another (models 1 and 2 in Fig. 5a are two examples of this behavior). While longer-term CO<sub>2</sub> responses remain broadly similar, many simulations produce decadal time-scale variability, such as responses to large volcanic eruptions (Agung and to a lesser extent El Chichón and Pinatubo), not found in the observed record. The consequence of this is that the details of which models can be excluded or not change depending on the time period used (Fig. 5b). Visually the impact of the choice of period is most evident in the tails of the distribution but occurs for model variants projecting changes across the full range. For PPE members close to the observed trends, the choice of time period will influence whether it is excluded or not. However, the impact of the time period on the wider distribution appears to be considerably less sensitive. For example, when excluding simulations with RCP8.5 2100 CO<sub>2</sub> concentrations in excess of 1100 ppm, changing the period to either 1920–2005 or 1980–2005 leads to 18 or 20 of the 26 models being excluded, respectively (compared to 18 models for the 1959–2005 period).

In this paper we used trends over a particular period to exclude or retain model variants to illustrate the potential information in the comparison with observed CO<sub>2</sub>. Also, we simply included or excluded model variants, rather than attempting to assign a weight based on the relative probability that each model variant is consistent with observations. In principle, these limitations could be addressed by using a multivariate Bayesian

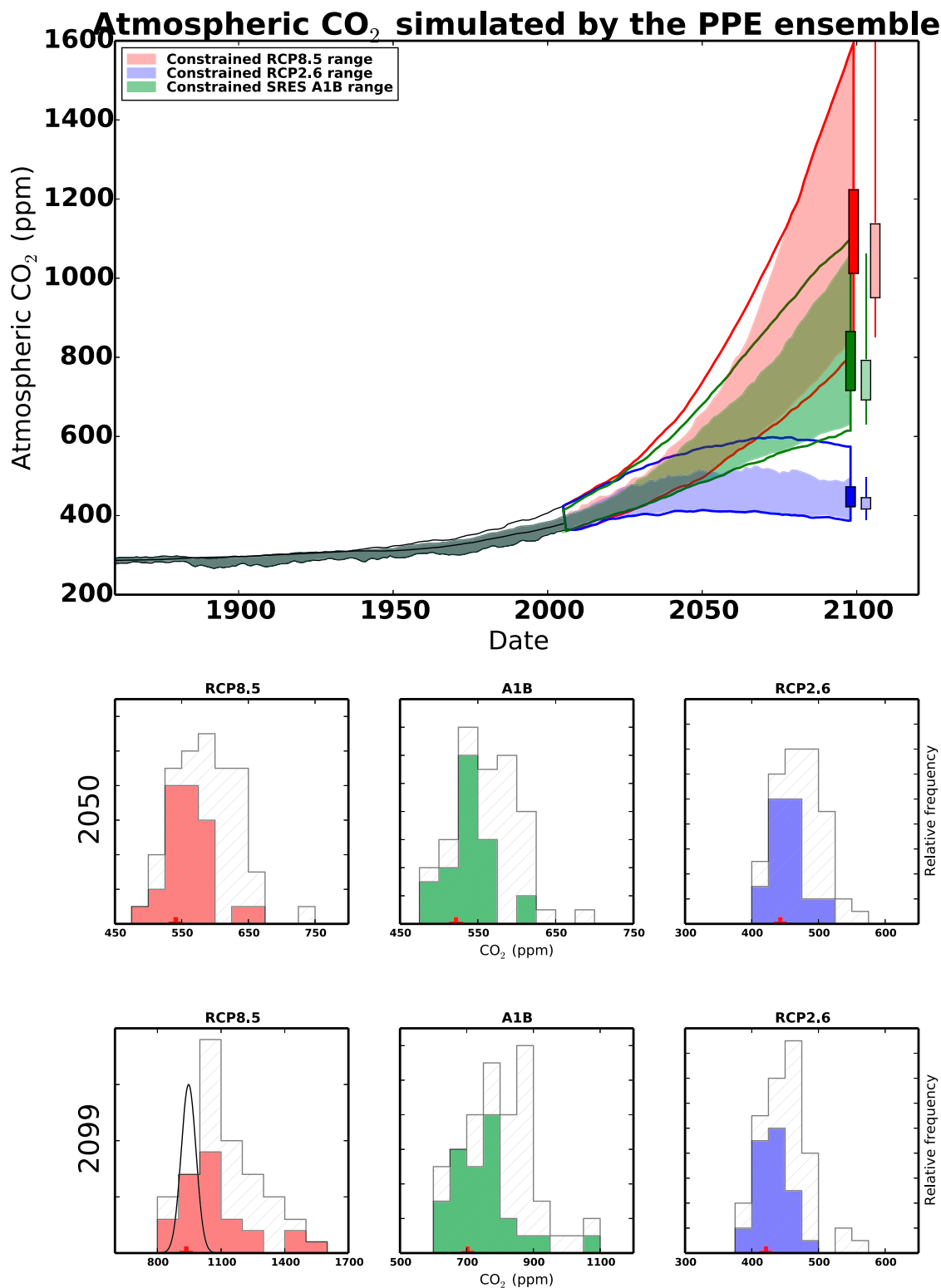


FIG. 4. Projected future changes in CO<sub>2</sub> concentration with and without the constraint from historical trends. (top) The plume of simulated CO<sub>2</sub> changes from three scenarios: RCP8.5 (red), SRES A1B (green), and RCP2.6 (blue). The full ranges (open plumes, bounded by thick color lines) and constrained ranges (colored plumes) are shown. Vertical bars show the range of CO<sub>2</sub> in 2100 for the unconstrained (strong colors) and constrained (light colored) ranges, with different thickness corresponding to the full range (thin line) or 25%–75% interquartile range (thick). The six histograms beneath show the relative frequency of the constrained (colored) and unconstrained (hatched) distributions for (middle) 2050 or (bottom) 2100 for the three scenarios (columns). The RCP reference concentrations are also indicated (red crosses), with the constrained estimate from Hoffman et al. (2014) also shown for 2100 RCP8.5 for comparison (bell curve).

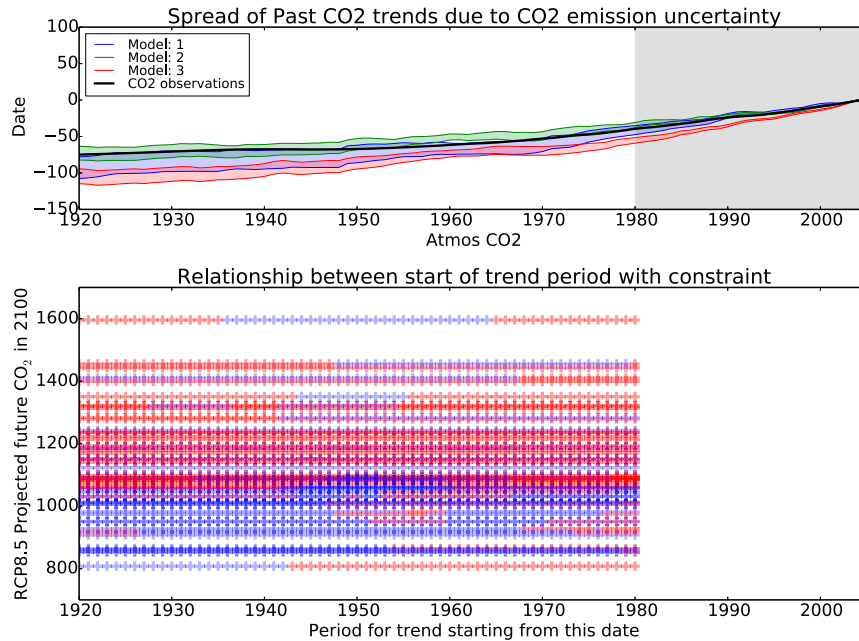


FIG. 5. The relationship between the time period of emission trends (anomalized with respect to 2005) and inclusion or exclusion of particular model variants from the PPE. (top) The range of historic CO<sub>2</sub> changes consistent with past emission uncertainty for three selected model variants (colored plumes) compared with historical observations (black line). The simulated and observed time series are all anomalized with respect to 2005 to keep consistency with the common end point used in the lower panel. (bottom) The relationship between simulated 2100 CO<sub>2</sub> concentrations and whether this simulation is excluded (red) or retained (blue) is based on the choice of the start date of the model-observed historical trend comparison. All trends extend to 2005, and the start dates in (bottom) correspond to the nongray time period in (top). So colored crosses in 1920 indicate whether the 1920–2005 comparison with observed trends excluded or retained this particular future change, and so on for the range of start dates.

approach to provide probabilistic projections accounting for multiple constraints (e.g., by considering trends over several historical periods or including additional variables such as mean observed climate and observed changes in ocean heat content and surface temperature). Sexton et al. (2012) and Harris et al. (2013) describe a suitable statistical framework for application to PPE experiments, in which probabilistic projections are obtained by integrating over the climate model parameter space and weighting projections associated with specific model variants according to relative likelihood, obtained by calculating the multivariate distance of each variant from a set of historical observables. In such a framework, those model variants where the observations lie close to but sometimes outside the model range (such as models 1 and 2 in Fig. 5a) could still be expected to receive some weight, whereas variants that show changes far from the observed (such as model 3 in Fig. 5a) would be substantially down weighted. This approach was not adopted here, since a formal Bayesian calculation requires estimation of additional sources of uncertainty affecting relative likelihood, including

observational uncertainty (likely small in this case), uncertainty in emulated estimates of results for parameter combinations for which no climate model simulation exists (e.g., Sexton et al. 2012), and uncertainty in the structural component of model error, representing systematic differences between the modeled and real-world climate that cannot be resolved by varying uncertain model parameters (Sexton et al. 2012). Without such a wider assessment there is a risk that use of a single constraining variable, such as the CO<sub>2</sub> trend examined here, and a simplified representation of model-observation misfit, could lead to overly constrained predictions due to compensating errors. The sensitivity of model exclusion to choice of trend period highlights the potential value of a more likelihood-based approach in the future. However, the simple approach here of excluding models based on CO<sub>2</sub> trends is useful in illustrating the value of this particular observable quantity and motivates future consideration of a more comprehensive approach, such as that outlined above.

As just discussed, when using a single simulated-observed comparison metric it neglects the potential

for compensating errors that may lead to plausible simulated historical CO<sub>2</sub> trends for the wrong reasons and hence may give a misleading picture of the more plausible future changes. This issue has been highlighted previously by [Friedlingstein et al. \(2014\)](#) where they showed that a number of the CMIP5 models matched the atmospheric CO<sub>2</sub> but did so owing to larger-than-observed uptake in the ocean coupled with very little uptake on the land. Figure S2 (supplementary information) illustrates the fraction of emitted carbon partitioned in the atmosphere, ocean, and land fractions by the 57 model variants during 1980–99. Encouragingly, the spread of simulated fractions encompasses the observed estimates, suggesting there is no systematic error in the ensemble reproducing the observed carbon partitioning during this period. Of the 57 PPE members, 17 simulate atmospheric fractions within observational uncertainties (Fig. S2). Among these, 6 members, a minority of the 17, capture the observed atmospheric fraction at the expense of compensating error, owing to either a larger-than-observed ocean uptake balanced by smaller-than-observed land uptake, or vice versa. Except for one model variant, however, these differences are small compared to the observational errors and may fall within internal variability, an additional source of model–observation misfit that is not accounted for in Fig. S2. Nevertheless, these results suggest that applying a constraint based solely on reproducing atmospheric CO<sub>2</sub> trends may be a necessary, but perhaps not sufficient, criterion for reducing uncertainty in future CO<sub>2</sub> changes. It is an aspect of future work to refine the constraint by incorporating additional observational metrics such as ocean carbon uptake and global/regional temperature changes. Doing so would tackle potential compensation of errors in the relative carbon uptake in the land and ocean and the magnitude of global climate change and climate–carbon cycle feedbacks, respectively.

## 6. Discussion and conclusions

In this study we have found that requiring simulated trends in atmospheric CO<sub>2</sub> concentration to be consistent with historical CO<sub>2</sub> trends can be effective in constraining the distribution of projected carbon cycle changes, ruling out 30 of the 57 simulations in a perturbed parameter ensemble (PPE) constructed from HadCM3C, many of which lie on the high end of possible future responses. The impact of this constraint is also apparent in the time scales that potential mitigation can be expected to become evident. With the raw ensembles we need to wait until 2068 before we reach a point where a high-end RCP2.6 run no longer overlaps

with the low-end business-as-usual RCP8.5 simulation [or 2037 if the interquartile range (IQR) from both ensembles is used]. By excluding implausible historical simulations, the date that the RCP2.6 concentrations no longer overlap with the RCP8.5 scenario reduces to 2053 (or down to 2029 if the IQR is used).

Our results can be compared with previous studies that have attempted to constrain future simulated CO<sub>2</sub> concentration based on historical CO<sub>2</sub> simulation–observation comparisons ([Friedlingstein et al. 2014](#); [Hoffman et al. 2014](#)). The smaller number of simulations available in the emission-driven CMIP5 experiment led these previous studies to focus on the range of CO<sub>2</sub> responses rather than the distribution. It is in the range that the apparent differences are most evident, particularly for end-of-century CO<sub>2</sub> responses to high future emission scenarios (RCP8.5). Previous work by [Hoffman et al. \(2014\)](#) suggested that historical CO<sub>2</sub> observations could represent an exceptionally tight constraint on RCP future concentration ( $947 \pm 35$  ppm in 2100). Here, individual PPE simulations consistent with historical trends can be found that lie at either end of the simulated PPE range (from 854 to 1455 ppm), despite the general tendency to rule out many of the models with high-end changes. Two main factors contribute to the weaker constraint on the absolute range (than on the distribution generally). First, we account for uncertainty in historical fossil fuel and land-use carbon emissions. Second, the PPE allows consideration of a much larger sample of physical and carbon feedbacks. Here we discuss both of these factors in further detail.

Historical uncertainty in anthropogenic carbon emissions has not been considered in previous studies looking at the potential information in the observed CO<sub>2</sub> changes ([Friedlingstein et al. 2014](#); [Hoffman et al. 2014](#); [Murphy et al. 2014](#)). Including an estimate of emission uncertainty broadens the range of models that could be consistent with observed changes. We find that although the risk of both low- and high-end CO<sub>2</sub> response is reduced by our constraint, we are unable to exclude all models that predict extreme responses. Sensitivity tests, where the LUC emission uncertainty was reduced by two-thirds (not shown), do find a greater number of potential projected future changes excluded (41 out of 57), emphasizing the value in reducing historical LUC emission uncertainty for future climate projections. Doing so, however, will be a significant challenge in climate research. While there is increasing empirical data available to distinguish and characterize the current sources of uncertainty ([Goldewijk and Verburg 2013](#)) from, for example, satellite inventories of land cover and biomass, issues of definitional differences and model differences will still need to be overcome ([Houghton et al.](#)

2012; Pongratz et al. 2014). Regarding the high or low tails of the distribution, it is still possible to identify models with plausible historical trends that span the range of 2100 RCP8.5 CO<sub>2</sub> concentrations, even when emission uncertainty is reduced, but the impact on the distribution will be stronger than the impact on the range.

The other factor that leads to some of the high-end 2100 CO<sub>2</sub> projections being retained for RCP8.5, albeit a minority, is the use of model variants from a perturbed parameter ensemble. These simulations were explicitly designed to explore a wide range of potential physical and carbon feedbacks (Lambert et al. 2013). The multimodel ensemble (CMIP5), in contrast, is a collection of models designed to produce plausible estimates of these feedbacks, albeit with a structural diversity of modeling assumptions. There are two potential aspects that may point to why the PPE produces a wider range of future behavior. The finding that the scenario with the weakest constraint on high-end 2100 CO<sub>2</sub> changes is the one with the strongest warming (RCP8.5) does point to climate–carbon cycle feedbacks as one of the primary processes capable of driving large future CO<sub>2</sub> responses in models with only modest trends historically. Some of the key processes explored in the PPE (Booth et al. 2012) do have important temperature controls on land carbon cycle processes. We lack the availability of idealized experiments for the PPE, required to explicitly diagnose the strength of these feedbacks; however, it is quite possible that the PPE includes model variants with stronger climate–carbon cycle feedbacks than models included in CMIP5. The other aspect that may contribute is the range of temperature responses in the PPE. The PPE includes a number of ensemble members with larger climate sensitivities than found in CMIP5 emission-driven runs (Booth et al. 2013). This, in itself, would lead to stronger warming and hence drive larger fluxes of carbon from the land to the atmosphere via climate–carbon cycle feedbacks, irrespective of whether the PPE explores larger climate–carbon cycle feedbacks than CMIP5. Both these factors highlight that any potential constraint on future CO<sub>2</sub> change, particularly for high-end scenarios, needs to account for the potential for both these factors playing a role in the real world. This suggests that until we are able to reject the range of these underlying feedbacks on other grounds, caution will need to be taken in putting too much weight on observationally constrained future CO<sub>2</sub> ranges based on simulations sampling smaller ranges to these feedbacks (such as CMIP5; Hoffman et al. 2014).

Hoffman et al. (2014) argued that the strong constraint on the future CO<sub>2</sub> range from historical observations implied that models could achieve more

confident projections by tuning models to reproduce past observed CO<sub>2</sub> values. These new results suggest caution in such an interpretation on a number of grounds. First, overly tight constraints on future CO<sub>2</sub> can be obtained when uncertainty in historical emissions is neglected. Future simulated–observed comparisons (e.g., with the CMIP6 generation of models) may benefit from explicitly sampling anthropogenic emission uncertainty. It is hoped that this study would raise awareness of the dependence on past carbon emission estimates, especially the LU component. Second, until the simulations used in these studies sample the full range of climate sensitivities and carbon feedbacks, doing so with a more limited sample is always likely to lead to an overly constrained estimate that does not reflect the wider uncertainties. Third and in a similar vein, any inference of such a constraint depends on the underlying models representing the full range of relevant processes. Currently important processes such as the nitrogen cycle (neglected by most CMIP5 and the PPE simulations) and dynamic vegetation (neglected in many CMIP5 simulations) are not represented in all underlying simulations. This is a caveat that needs to be borne in mind. Last, as highlighted by Friedlingstein et al. (2014) and discussed in the previous section, the potential for compensation of errors (e.g., land vs ocean carbon uptake; magnitude of climate change vs climate–carbon cycle feedbacks) means that any comparison against CO<sub>2</sub> trends alone represents a necessary but not sufficient constraint on the range of model responses.

This paper also has a few implications for CO<sub>2</sub> estimates that will be prescribed for concentration-driven experiments in the forthcoming model intercomparison project, CMIP6. For CMIP5 the reference concentration (prescribed to the standard, non-emission-driven experiments) was homogenized to be consistent with the BernCC carbon cycle and climate sensitivity used in the earlier SRES scenarios (Meinshausen et al. 2011). These results suggest that the standard reference concentration profiles, prescribed for concentration-driven CMIP5 experiments (both RCP8.5 and others), may be biased low. Although the standard RCP8.5 reference concentration falls within the range of emission-driven simulations with consistent historical trends, 21 of these 27 PPE simulations produce larger CO<sub>2</sub> concentrations in 2100. Indeed when using the IQR of emission-driven simulations consistent with the historical trends (951–1137 ppm) the reference concentration (936 ppm) does not even fall within this range. This is not new. Friedlingstein et al. (2014) noted a low bias to the reference concentration compared to CMIP5 emission-driven simulations. However, the presence of historical CO<sub>2</sub> biases in many of the CMIP5 high-end simulations meant that they did not



draw any strong conclusions on this. This paper shows that simulations that would be considered high end in the CMIP5 context can be consistent with observed trends, a point that reinforces the idea that the reference CMIP5 scenario concentration profiles are likely to have a low bias. This has implications for the estimates of allowable emissions taken from CMIP5 concentration-driven simulations, which consequentially would be on the lower end of expectations based on the presented results. Decisions for the reference scenario for CMIP6 will need to balance the advantages for choosing a reference scenario consistent with previous assessments on one hand, with the benefits for choosing a more central estimate of the carbon cycle feedbacks on the other.

The main result from this work is the impact on the distribution of future CO<sub>2</sub> projections consistent with the historical trends, rather than on the outliers. Using the IQR as a robust metric of uncertainty, applying our constraint based on CO<sub>2</sub> observations shifts the RCP8.5 2100 range down by roughly 80 ppm (from 1012–1223 ppm down to 951–1137 ppm). For lower emission scenarios, and for earlier times in the century, using CO<sub>2</sub> observations also narrows this IQR, as well as shifting it lower. For example, the IQR for RCP2.6 at 2100 contracts from 50 ppm (423–472 ppm) down to 28 ppm (417–445 ppm) and midcentury RCP8.5 IQR contracts from 63 ppm (551–614 ppm) down to 44 ppm (535–579 ppm). In summary, we show both that historical CO<sub>2</sub> observations are effective at narrowing uncertainty in future projections but also that the observationally constrained ranges (whether based on the full spread of outcomes or the IQR) remain larger than earlier emission-driven CMIP5 estimates suggested (Hoffman et al. 2014). We attribute this to a wider sampling of potential physical and carbon cycle feedbacks in our simulations and inclusion of estimates of past carbon emission uncertainty. Accounting for the broader range of both historical carbon emissions and future carbon cycle responses will be important if we are to use historical CO<sub>2</sub> observations to inform future climate projections.

*Acknowledgments.* This work was supported by the Joint U.K. DECC/DEFRA Met Office Hadley Centre Climate Programme (GA01101). Chris Jones's contribution was supported by the CRESCENDO project under the European Union's Horizon 2020 research and innovation programme, Grant Agreement 641816. Jo House was supported by a Leverhulme Early Career Fellowship and EU FP7 Project LUC4C (603542). Stephen Sitch was supported by the EU FP7 through Project LUC4C (GA603542). We acknowledge the contribution of the CMIP5 modeling groups through the World Climate Research Programme's Working Group on Coupled

Modelling (2011, with subsequent updates): phase 5 of the Coupled Model Intercomparison Project, Earth System Grid Federation (<https://pcmdi9.llnl.gov/projects/cmip5/>; accessed January 2014).

## REFERENCES

- Arora, V. K., and Coauthors, 2013: Carbon-concentration and carbon-climate feedbacks in CMIP5 Earth system models. *J. Climate*, **26**, 5289–5314, doi:10.1175/JCLI-D-12-00494.1.
- Blyth, E., D. B. Clark, R. Ellis, C. Huntingford, S. Los, M. Pryor, M. Best, and S. Sitch, 2011: A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale. *Geosci. Model Dev.*, **4**, 255–269, doi:10.5194/gmd-4-255-2011.
- Bodman, R. W., P. J. Rayner, and D. J. Karoly, 2013: Uncertainty in temperature projections reduced using carbon cycle and climate observations. *Nat. Climate Change*, **3**, 725–729, doi:10.1038/nclimate1903.
- Booth, B. B. B., and Coauthors, 2012: High sensitivity of future global warming to land carbon cycle processes. *Environ. Res. Lett.*, **7**, 024002, doi:10.1088/1748-9326/7/2/024002.
- , D. Bernie, D. McNeall, E. Hawkins, J. Caesar, C. Boulton, P. Friedlingstein, and D. M. H. Sexton, 2013: Scenario and modelling uncertainty in global mean temperature change derived from emission-driven global climate models. *Earth Syst. Dyn.*, **4**, 95–108, doi:10.5194/esd-4-95-2013.
- Brovkin, V., S. Sitch, W. Von Bloh, M. Claussen, E. Bauer, and W. Cramer, 2004: Role of land cover changes for atmospheric CO<sub>2</sub> increase and climate change during the last 150 years. *Global Change Biol.*, **10**, 1253–1266, doi:10.1111/j.1365-2486.2004.00812.x.
- Cadule, P., P. Friedlingstein, L. Bopp, S. Sitch, C. D. Jones, P. Ciais, S. L. Piao, and P. Peylin, 2010: Benchmarking coupled climate carbon models against long term atmospheric CO<sub>2</sub> measurements. *Global Biogeochem. Cycles*, **24**, GB2016, doi:10.1029/2009GB003556.
- Ciais, P., and Coauthors, 2013: Carbon and other biogeochemical cycles. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 465–570.
- Collins, M., B. B. Booth, B. Bhaskaran, G. R. Harris, J. M. Murphy, D. M. Sexton, and M. J. Webb, 2011: Climate model errors, feedbacks and forcings: A comparison of perturbed physics and multi-model ensembles. *Climate Dyn.*, **36**, 1737–1766, doi:10.1007/s00382-010-0808-0.
- Cox, P. M., D. Pearson, B. B. Booth, P. Friedlingstein, C. Huntingford, C. D. Jones, and C. M. Luke, 2013: Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature*, **494**, 341–344, doi:10.1038/nature11882.
- Forster, P. M. F., and K. E. Taylor, 2006: Climate forcings and climate sensitivities diagnosed from coupled climate model integrations. *J. Climate*, **19**, 6181–6194, doi:10.1175/JCLI3974.1.
- Friedlingstein, P., and Coauthors, 2006: Climate-carbon cycle feedback analysis: Results from the C4MIP model intercomparison. *J. Climate*, **19**, 3337–3353, doi:10.1175/JCLI3800.1.
- , M. Meinshausen, V. K. Arora, C. D. Jones, A. Anav, S. K. Liddicoat, and R. Knutti, 2014: Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *J. Climate*, **27**, 511–526, doi:10.1175/JCLI-D-12-00579.1.
- Goldewijk, K. K., and P. H. Verburg, 2013: Uncertainties in global-scale reconstructions of historical land use: An illustration using the HYDE data set. *Landscape Ecol.*, **28**, 861–877, doi:10.1007/s10980-013-9877-x.

- , A. Beusen, G. Van Drecht, and M. De Vos, 2011: The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Global Ecol. Biogeogr.*, **20**, 73–86, doi:10.1111/j.1466-8238.2010.00587.x.
- Harris, G. R., D. M. Sexton, B. B. Booth, M. Collins, and J. M. Murphy, 2013: Probabilistic projections of transient climate change. *Climate Dyn.*, **40**, 2937–2972, doi:10.1007/s00382-012-1647-y.
- Hoffman, F. M., and Coauthors, 2014: Causes and implications of persistent atmospheric carbon dioxide biases in Earth system models. *J. Geophys. Res. Biogeosci.*, **119**, 141–162, doi:10.1002/2013JG002381.
- Houghton, R. A., 2003: Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850–2000. *Tellus*, **55B**, 378–390, doi:10.1034/j.1600-0889.2003.01450.x.
- , J. I. House, J. Pongratz, G. R. Van der Werf, R. S. DeFries, M. C. Hansen, C. L. Quéré, and N. Ramankutty, 2012: Carbon emissions from land use and land-cover change. *Biogeosciences*, **9**, 5125–5142, doi:10.5194/bg-9-5125-2012.
- Jain, A. K., P. Meiyappan, Y. Song, and J. I. House, 2013: CO<sub>2</sub> emissions from land-use change affected more by nitrogen cycle, than by the choice of land-cover data. *Global Change Biol.*, **19**, 2893–2906, doi:10.1111/gcb.12207.
- Jones, C. D., P. M. Cox, and C. Huntingford, 2006: Climate-carbon cycle feedbacks under stabilization: Uncertainty and observational constraints. *Tellus*, **58B**, 603–613, doi:10.1111/j.1600-0889.2006.00215.x.
- Joos, F., M. Bruno, R. Fink, U. Siegenthaler, T. F. Stocker, C. Le Quéré, and J. L. Sarmiento, 1996: An efficient and accurate representation of complex oceanic and biospheric models of anthropogenic carbon uptake. *Tellus*, **48B**, 397–417, doi:10.1034/j.1600-0889.1996.t01-2-00006.x.
- Kato, E., T. Kinoshita, A. Ito, M. Kawamiya, and Y. Yamagata, 2013: Evaluation of spatially explicit emission scenario of land-use change and biomass burning using a process-based biogeochemical model. *J. Land Use Sci.*, **8**, 104–122, doi:10.1080/1747423X.2011.628705.
- Lambert, F. H., G. R. Harris, M. Collins, J. M. Murphy, D. M. Sexton, and B. B. Booth, 2013: Interactions between perturbations to different Earth system components simulated by a fully-coupled climate model. *Climate Dyn.*, **41**, 3055–3072, doi:10.1007/s00382-012-1618-3.
- Le Quéré, C., and Coauthors, 2015: Global carbon budget 2015. *Earth Syst. Sci. Data*, **7**, 349–396, doi:10.5194/essd-7-349-2015.
- Luo, Y. Q., and Coauthors, 2012: A framework for benchmarking land models. *Biogeosciences*, **9**, 3857–3874, doi:10.5194/bg-9-3857-2012.
- Meinshausen, M., and Coauthors, 2011: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, **109**, 213–241, doi:10.1007/s10584-011-0156-z.
- Murphy, J. M., 1995: Transient response of the Hadley Centre coupled ocean–atmosphere model to increasing carbon dioxide. Part III: Analysis of global-mean response using simple models. *J. Climate*, **8**, 496–514, doi:10.1175/1520-0442(1995)008<0496:TROTHC>2.0.CO;2.
- , and Coauthors, 2009: UK climate projections science report: Climate change projections. Met Office Hadley Centre Rep., 192 pp.
- , B. B. Booth, C. A. Boulton, R. T. Clark, G. R. Harris, J. A. Lowe, and D. M. Sexton, 2014: Transient climate changes in a perturbed parameter ensemble of emissions-driven Earth system model simulations. *Climate Dyn.*, **43**, 2855–2885, doi:10.1007/s00382-014-2097-5.
- Myhre, G., E. J. Highwood, K. P. Shine, and F. Stordal, 1998: New estimates of radiative forcing due to well mixed greenhouse gases. *Geophys. Res. Lett.*, **25**, 2715–2718, doi:10.1029/98GL01908.
- Pongratz, J., C. H. Reick, R. A. Houghton, and J. I. House, 2014: Terminology as a key uncertainty in net land use and land cover change carbon flux estimates. *Earth Syst. Dyn.*, **5**, 177–195, doi:10.5194/esd-5-177-2014.
- Poulter, B., and Coauthors, 2010: Net biome production of the Amazon basin in the 21st century. *Global Change Biol.*, **16**, 2062–2075, doi:10.1111/j.1365-2486.2009.02064.x.
- Ramankutty, N., and J. A. Foley, 1999: Estimating historical changes in global land cover: Croplands from 1700 to 1992. *Global Biogeochem. Cycles*, **13**, 997–1027, doi:10.1029/1999GB900046.
- Ricciuto, D. M., K. J. Davis, and K. Keller, 2008: A Bayesian calibration of a simple carbon cycle model: The role of observations in estimating and reducing uncertainty. *Global Biogeochem. Cycles*, **22**, GB2030, doi:10.1029/2006GB002908.
- Sexton, D. M., J. M. Murphy, M. Collins, and M. J. Webb, 2012: Multivariate probabilistic projections using imperfect climate models part I: Outline of methodology. *Climate Dyn.*, **38**, 2513–2542, doi:10.1007/s00382-011-1208-9.
- Stocker, B. D., K. Strassmann, and F. Joos, 2011: Sensitivity of Holocene atmospheric CO<sub>2</sub> and the modern carbon budget to early human land use: Analyses with a process-based model. *Biogeosciences*, **8**, 69–88, doi:10.5194/bg-8-69-2011.
- Tachiiri, K., J. C. Hargreaves, J. D. Annan, C. Huntingford, and M. Kawamiya, 2013: Allowable carbon emissions for medium-to-high mitigation scenarios. *Tellus*, **65B**, 20586, doi:10.3402/tellusb.v65i0.20586.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.
- Van Vuuren, D. P., and Coauthors, 2011: The representative concentration pathways: An overview. *Climatic Change*, **109**, 5–31, doi:10.1007/s10584-011-0148-z.
- Wang, X., and Coauthors, 2014: A two-fold increase of carbon cycle sensitivity to tropical temperature variations. *Nature*, **506**, 212–215, doi:10.1038/nature12915.