Supplementary information

Relationship between model formulation and CO2 changes

The PPE model variants show a wider range of future CO2 responses consistent with historical trends, than earlier CMIP5 based analysis suggested. Ideally these simulations could be used to identify which model parameters, and hence processes, were responsible for models consistent with historical trends but exploring larger future changes. Unfortunately, the ensemble was not designed to address this question. Instead it was setup to sample interactions between uncertainties in four climate model components explored in earlier modelling work: atmospheric and land surface physical processes (Collins, 2011), land carbon cycle processes (Booth, 2012), ocean physics (Brierley, 2010) and aerosols. Specifically, each of the original 17 atmospheric; land carbon; aerosol and ocean physical configurations were coupled together so that resulting ensemble of model variants (used in this analysis) sampled a range of interactions between each component. This design is detailed in Lambert et al, 2013.

![Figure S1: Role of land carbon cycle configuration in past-future relationships in CO2 responses. This figure reproduces the same data shown in the main manuscript (Figure 1b) for the PPE model variants. Here, colours are used to indicate individual land carbon cycle configurations. Points with the same colours share the same configuration.](http://dx.doi.org/10.1175/JCLI-D-16-0178.s1)
ensemble, coupled to different configurations from the other 3 components, leading to 64 model variants. The 57 used in this study were reduced from the original 64 by removing model variants that failed to simulate gross properties of the climate (see Lambert, 2013). The climate sensitivity (largely driven by differences in the atmospheric physics configurations) and the carbon cycle response where the two factors that most influence the global CO2 response (Booth, 2013). The presence of multiple version of each atmospheric physics and land carbon cycle configurations provide a clue as to the processes leading to model variants with large future changes in CO2 whilst reproducing recent trends. Figure S1 shows that 4 model variants (red) in or near this large future but historical plausible region share a common land carbon configuration. Carbon cycle configurations tend to provide a reasonable indication of where a particular model variant will sit in this response space. For example, the 4 of the 5 lowest future responses share the same land carbon cycle configurations (yellow, Fig. S1). Similarly common land carbon configurations can be found in the high future response, large historical trend region (e.g. the light and dark blue, salmon and pink land carbon configurations in Fig. S1). The inference is that it is uncertainty in the formulation of the land carbon cycle response that is most important to understand if we are going to rule out either low or high projections of future CO2 change.

However, the picture may not be as clear as these clusters at first suggest. Some land carbon configurations (e.g. grey and orange configurations in Fig. S1) appear in quite different parts of this historical/future CO2 response space. In these cases, this suggests that it is the interaction of these common land carbon configurations with other components that lead to this spread. Clearly more work is needed to tease apart the relative importance of the various processes. The authors are looking at the potential for new experiments that would enable the individual effects to be estimated. This is envisaged to be an important direction for further work.

Compensating errors and CO2 as a model constraint

The main manuscript revisits earlier work that used observed-simulated comparisons to down weight less plausible historical simulations and hence down weight future projected CO2 changes that are associated with these less plausible simulations. In doing so, this approach assume that models capture or not, the observed changes for the correct reasons.

Friedlingstein, 2013 highlighted the potential risks in taking such an approach, pointing to models that achieve a good simulation of observed CO2 due to compensation of errors in the processes that control CO2 changes. They showed that the INM model, which produces global CO2 values close to observed, does so by combining an erroneously large ocean carbon uptake coupled to a land carbon uptake that is considered too small. Hoffman, 2014 also noted this potential for compensation between the processes that determine atmospheric CO2 changes, but argued that on century timescales was too short for biases in these component fluxes to lead to errors in the projected CO2 changes. The evidence, that underpinned this statement, was that the CMIP5 models showed consistent ranking through time. So the models that showed the larger (smallest) CO2 changes in present day producing the largest (smallest) projected CO2 values at the end of this century.

The responses of the PPE model variants presented in the main manuscript suggest, however, that potential biases may play a more important role during this coming century. The PPE simulations shows much less linearity in the relationship between past and future changes (Figure 1, main manuscript). If the strength of the future CO2 changes is more dependent on either the land carbon or ocean carbon cycle component, then whether or not present day biases exist in either the land or
ocean carbon uptake becomes a more important question.

![Figure S2: The relative uptake of historical anthropogenic carbon emissions into the Atmosphere, ocean and land components. This Ternary plot illustrates the fraction of 1980 to 1999 anthropogenic emissions that reside in the atmosphere, ocean and land (where 100% in each of these corresponds to the top, bottom left and bottom right corners, respectively). The observed estimates including uncertainties in the anthropogenic emissions and observational errors (based on Sabine, 2004) are shown by the contoured coloured region. The corresponding values for each of the PPE model variants (crosses) and for HadGEM2ES (larger green star) are also shown. Note, no estimate of the internal variability is accounted for in this figure. Doing so may reconcile a number of the simulations, that currently are shown outside the observed contours, with the observed values.](http://dx.doi.org/10.1175/JCLI-D-16-0178.s1)

Sabine, 2004 provides estimates of the fraction of 1980-1999 anthropogenic emissions that resides in each of the 3 components (atmosphere, ocean and land stores). Figure S2 illustrates how the 57 PPE model variants compare to this estimate. The models span a range of carbon fractions (just as they explored a range of historical CO2 trends, see Figure 1 main manuscript). What is encouraging is that the response space of the different model variants spans the observed estimates. So most models that do well on the atmospheric fraction (which we'd expect to be closely related to good simulation of observed atmospheric CO2 changes) tend to fall within the space consistent with
the observed ocean and land carbon uptake estimates. Whilst the scatter of the PPE model variants covers the region consistent with good agreement with the observed estimates of all three components (atmospheric, oceanic and land fractions), it also includes models that match the atmospheric fraction correctly but do so due to either larger ocean/smaller land uptake or smaller ocean/larger land uptake than the observed estimates.

Presenting the results in this way points to a future direction for this work, where multiple component fluxes are constrained. The probabilistic framework that this work would take is outlined in the Section 5 of the main paper.

Additional References:
