Earth Syst. Dynam. Discuss., 5, C433–C438, 2014
www.earth-syst-dynam-discuss.net/5/C433/2014/

Earth System Dynamics
Discussions
Open Access

Interactive comment on "Differences in carbon cycle and temperature projections from emission- and concentration-driven earth system model simulations" *by* P. Shao et al.

C.D. Jones (Referee)

chris.d.jones@metoffice.gov.uk

Received and published: 18 September 2014

Review of P.Shao et al, "Differences in carbon cycle and temperature projections from emission- and concentration-driven earth system model simulations"

This manuscript sets out to look at the differences between emissions-driven and concentration-driven simulations of CMIP5 earth system models. Now that models routinely include an interactive carbon cycle component they can either be run in a mode driven by boundary conditions of prescribed atmospheric CO2 pathways ("concentration driven") or more fully coupled by taking CO2 emissions as an input and the models themselves simulate the evolution of atmospehric [CO2] ("emissions driven").

For CMIP5 some models performed the same scenario (RCP8.5) under both protocols so it is reasonable to ask what are the differences and implications of these different experimental designs.

The paper assembles results from these different simulations and shows where the different runs are different or not. But I find some of the logic and order of the analysis does not make sense and there are some gross errors in some of the results due to mistakes with units of carbon fluxes and stores. I have highlighted below some areas where the paper needs to address these issues before it can be accepted for publication. I have also listed a handful of comments which I hope are useful in more minor ways.

Chris Jones

Major comments.

1. On page 997 you try to explain the differences in [CO2] between the C-driven and E-driven runs by looking at the differences in land and ocean carbon uptake. This is the wrong way round. It is the differences in [CO2] which drive the differences in the fluxes, not the other way round. The reason why E-driven runs have different [CO2] from C-driven is because the ESM fluxes differ from the IAM which created the scenario. In the IAM the simple model MAGICC is used to map from emissions to concentration. If the ESM uptake differs from MAGICC then the E-driven run will have different [CO2] from the C-driven run.

This difference between E-driven and C-driven runs then CAUSES (is not caused by) the different land/ocean fluxes in the ESM between the two simulations. You cannot use the land/ocean differences to explain the [CO2] differences.

2. In figure 1 you have mixed up your units of carbon leading to very wrong and misleading numbers in panel 1d. Panel 1a presents the differences in atmospheric CO2 - it is the same as in Friedlingstein et al. (2014) figure 2a, but subtracting the

C-driven [CO2] from the E-driven results. Here you choose to convert the units to Pg(CO2) rather than PgC. In panels 1b and 1c you use units of PgC for land/ocean equivalents. You therefore cannot simply add these together. When you combine them to get panel 1d you therefore have mis-matched units and hence VERY odd results. Did you not wonder why these numebrs are so large and different from other studies? e.g. Jones et al (2013) compare the timeseries and cumulative totals of emissions between C-driven runs and the IAMs and do not see numbers this big (Jones et al figure 5b shows most ESMs are within 100 PgC of the IAM and all are within about 500 PgC). If you convert your panel 1a to PgC before calculating the diagnosed emissions you should get consistent numbers.

i.e. the 2100 values in your figure 1d should give the same values as the red bars/black dots in Jones et al figure 5b.

3. I found the discussion on the climate feedback metrics a bit over simplistic. It is well known (in fact for over a century) that radiative forcing scales with ln(CO2) and not CO2 itself. This is precisely why transient climate sensitivity ("TCR") in climate models is measured using an exponential (1%/yr) rise in CO2 - so that the forcing and hence T response is linear in time. So your finding is of course not surprising that alpha-prime is more constant in time than alpha. If the only point of alpha was a diagnosis of the climate response to CO2 then this would indeed make more sense. But from the C4MIP feedback framework alpha is required to be a linearisation so that the subsequent feedback framework fits. The whole of section 4 of Friedlingstein et al 2006 relies on this linearisation, and alpha is then a vital component of the gain factor, g, in their eqn 7. You cannot simply redefine one term in this framework. If you want to derive an improved framework it would need to be consistent right through all the metrics. Gregory et al (2009, J. Clim., 22, 5232-5250) is a good start at looking at the implications of linearising the responses about different points.

I also did not understand why you suggest using log(CO2) in the gamma term. You do not discuss this at all in the text, so strange to come out of the blue in the conclusion

section. gamma is defined as the change in land/ocean carbon storage per degree of temperature change. It does not involve [CO2] at all.

Minor Comments.

1. There are many biases and errors in the carbon cycle simulations of CMIP5 ESMs. Do you have a reason why you single out the seasonal cycle of CO2 (in abstract and conclusion) as being particularly urgent to address? Anav et al (2013., J. Clim., 26, 6801–6843) look across many variables for example and see big errors in terrestrial carbon stores. It's not clear the seasonal cycle of the fluxes is a more urgent area to fix than this for example.

As you say, the CO2 seasonal cycle is driven mainly by land fluxes. There are very large uptake (GPP) and release (respiration) fluxes which have large seasonality. The cycle of the net flux is a fine balance between these. So I agree the seasonal cycle of CO2 is at least one good metric of model performance, but it is very hard to know the origins (and hence implications) of any discrepancies. Too large/small GPP might have very different implications from too large/small respiration or simply a mis-match in their phasing. Models may have reasonable components and a very poor net cycle, or may get a good net cycle which hides large but compensating errors in the processes. So I would recommend focus on a much wider set of metrics is required.

2. Your description of the experimental design (E-driven vs C-driven, p.993) is good. You might also site Box 6.4 of Ciais et al (IPCC AR5 WG1 Ch.6) which shows this diagramatically.

3. p. 997, line 1. When you refer to differences of the E-driven run from the C-driven don't use the word "bias". This implies the E-driven run is wrong and the C-driven is correct. We don't know which is right/wrong/better/worse so just say they are different from each other.

4. p.999 / figure 3a. Why does BNU have a significant T difference at the start of the

runs?

5. p.1000. You can define an alpha for runs driven by many climate forcings, but is it meaningful? The temperature in these runs is only partially due to CO2. Idealised simulations with CO2 as the only forcing are better to calculate the sensitivity.

6. p.1002, lines 4-10. The seasonal cycle of temperature is driven by insolation/earth's orbit. I wouldn't expect the seasonal cycle of CO2 to affect it. If it were important then you should look at N. hemi and S. hemi CO2 separately and not a global mean.

7. figure 1 - Does INM really have 1200 PgC different land carbon between these two simulations? I haven't looked at the data, but this seems huge. Is it a real difference or some diagnostic issue? You mention they don't impose LUC, but this is the same for both runs isn't it?

8. table 1. can you define how you calcualte the correlation with HadCRUT - is it based on annual temperatures? or smoothed to give decadal/longer trends? If the former then is this meaningful given we don't expect the models to agree in their phase of internal variability, if the latter then is this meaningful? the MAGNITUDE of response is not captured so simply getting a good correlation is only part ofht erequirement.

9. going through your conclusions:

- [CO2] differs by 2100 by -19 to +207 ppm. OK.

- The diagnosed emissions therefore differ by »1000 PgC. No - this is an error due to mixed up units of carbon/CO2.

- E-driven results have a wider temperature spread than C-driven. OK. But this is not new in itself - see Friedlingstein 2014 Figure 2b. This was also noted in the AR4 projections chapter.

- carbon cycle feedback framework should use ln(CO2) for its temperature sensitivity. No - because then the whole linearisation breaks down. While I acknowledge the

linearisation is not perfect it has proven useful. If you only want to look at the climate sensitivity of the model to CO2 then I agree the C4MIP definition of alpha is not optimal. But within the feedback framework it is still OK.

- use ln[CO2] in the gamma term instead of [CO2]. No - [CO2] is not in gamma in the first place?

- seasonal cycle of CO2 should be used as a benchmark. I agree, but only within a basket of other metrics too. We need to understand why models differ and ensure that they get the right answer for the right reasons.

- ESMs and IAMs should be more consistent. They are already fairly consistent. Jones et al show this (figure 5). They agree well for low scenairos, and less well for higher scenarios. But the agreement is much better than you show in your figure 1d due to your units error.

Interactive comment on Earth Syst. Dynam. Discuss., 5, 991, 2014.