Response to review by C. D. Jones

We thank the reviewer for their considered and constructive comments. Please find below our responses to the reviewer's comments. We have uploaded our proposed revised manuscript that addresses this and the other reviewer's comments in a separate Author Comment on the article's Discussion page.

This is a nicely designed study, and well presented manuscript which attempts to develop and document a simple ("stylized") model of the global climate-carbon cycle system, but in a way which enables analytical analysis of its behaviour and feedback mechanisms. The intended aim is to facilitate improved understanding of the system dynamics and more readily quantify which processes contribute to certain feedbacks and long-term responses.

Overall, I very much like the approach and the intention – such modelling studies can develop improved insight and can strip back confounding and complicating issues of model complexity to reveal more fundamental underlying behaviour. I have a few comments below and a few queries about the extent to which the intentions have been realised – I think these can be readily addressed with revised text and some more context/explanation.

We thank the reviewer for his support.

Overarching comments:
1. There are numerous simple/stylized carbon cycle models in the literature. You cite Raupach (2013) which I like very much. There is also one I developed and have published with several times, including as recently as last year (Jones et al., Tellus, 2003; Jones et al., Tellus, 2006; Jones et al., ERL, 2016). Then there is the MAGICC model often used in IAMs, the Joos IRF, and also the Oscar model which some of the current authors know very well. So I wonder if a bit more explanation is needed for why a new model is required – could you start from one of the existing ones and achieve the same thing? I guess your main driver is the ability to analytically derive the feedback functions – but it's not clear to me the same is not possible from these previous models (I haven't tried it with the Jones et al simple model, but may do!) – Mike Raupach derived eigenmodes, so I would imagine feedback metrics could follow also, but again I haven't tried.

A key motivation for the proposed model is that it contains mechanistic representations (albeit highly aggregated and stylised) of key climate-carbon processes. In contrast, few, if any, of the models cited above explicitly include a solubility or biological ocean pump. In comparison to many of the cited models, we also substantially simplify the representation of the terrestrial carbon cycle, in order to simplify the analysis. We suspect that a similar analytical feedback analysis could in theory be applied to most of the simple models that the reviewer cites, but the partitioned terrestrial carbon stocks in most of these models would complicate the analysis, and parametric fits to the ocean carbon cycle would make the results less meaningful. We will make clearer the added value of the model in the revised version (see list beginning bottom page 3).

2. On a similar line – I was looking forward to seeing the analytical derivation of feedback factors, but then realised this was not as "tractable" as your title suggests, and you have to make a lot of assumptions in your sections 4.2 and 4.3. This seems a shame – if this is the case, have you not lost your unique and attractive feature? The resulting expressions are still useful, but not as analytical as you suggest. I also wondered, on seeing the expressions in table 3, if you had compared these to the expressions of Ric Williams et al – who have done a similar based assessment of terms controlling ocean heat/carbon uptake and TCRE. (See e.g., Richard G Williams et al 2016 Environ. Res. Lett. 11 015003)

We agree with the reviewer that while our feedback results are analytical (in the sense of closed-form mathematical expressions) they are not exact. It would be an interesting challenge to develop a model for which exact results could be achieved, but we suspect this would be at the cost of a mechanistic representation. In the revised version of the manuscript, we will clarify our use of the term of 'analytical'.

Williams et al. (2016) split apart three key factors influencing TCRE: (1) influence of CO2 emissions on radiative forcing from CO2; (2) influence of radiative forcing from CO2 on total radiative forcing; and (3) influence of total radiative forcing on temperature. They then use time series output from ESMs to drive each of these factors and calculate TCRE over time. For example, land and ocean uptake (which influence factor 1) are based directly on ESM output. In contrast, we formulate mechanistic models for land and ocean uptake. Our treatment of factor 3 is similarly highly stylised and we do not explicitly model ocean heat uptake. Regarding factor 2, we only consider CO2 forcing.

In one subsection, Williams et al. analytically calculate an equilibrium TCRE. They use a similar formulation for ocean chemistry based on the Revelle (buffer) factor as ours. However as theirs is an equilibrium calculation, they unlike us do not consider time scales introduced by mechanisms such as the solubility and biological pumps. They also neglect land carbon uptake on this long time scale.

3. You discuss (page 18, line 16) that you might want to develop this model further to include other mechanisms and forcings. Modellers have attempted a two-box approach for the ocean – by splitting upper and deep ocean before. But I would suggest maybe a two-box approach for the land also – not splitting veg vs soil, but maybe by tropics and extra-tropics as these can have very different responses (even by sign) to changes in climate. IPCC AR5 fig 6.22 shows the latitudinal distribution of "gamma" – there is a clear change of sign towards high latitudes, and so a tropics/high-latitude 2-box approach might be a nice (and novel) development

To compartmentalise land carbon by tropics and extra-tropics is an interesting suggestion which could be followed in future studies. We will raise this idea in the revised version of the manuscript, see section 6. In doing so, one would define carbon pools not by residence times but by climate sensitivities. This is really interesting but requires some more thinking.

In addition, to our understanding the results in IPCC AR5 fig 6.22 are based on ESM runs that do not contain any representation of permafrost carbon, hence this strong difference between arctic and extra-arctic beta values seems to more reflect the vegetation response to climate change. It would be interesting to see such sensitivity study using a fully coupled ESM run including permafrost carbon. For now, we therefore refrain from including these sensitivity gradients in our model.

4. In a few places, you discuss non-linearities – this is good, and important to bring out. It's not necessarily true that land dominates the non-linearity, but this does show up in your rapid-forcing transient runs. If you ran longer, or with slower changing forcing, the ocean would have more chance to exhibit these too – see, e.g. Schwinger et al (J. Climate, 2014). Hence both land and ocean can have pronounced non-linearities and for this reason, C4MIP took the decision to move back towards the Friedlingstein definition of feedbacks as the difference between COUPLED and BGC runs (see Jones et al., GMD, 2016 – documentation of C4MIP). Hence this differs from the Arora et al definition of using the RAD forced run.

We agree with the reviewer: since we chose to investigate effects on a 100-year policy-relevant time scale, many ocean effects are rendered insignificant. We will discuss in the revised version of the manuscript that the dominance of the land in our nonlinearity results is likely due to the time scale simulated (see section 5.3).

We thank the reviewer for drawing attention to the need for clarity in feedback definitions. Most of the the previous studies to which we compare our numerical feedback results use the Arora definition of feedbacks (Arora 2013 and Zickfeld 2011), the exception is Friedlingstein 2006. We have used the Arora definition to be consistent with the majority of cited previous studies, and will clarify which definition we use in the revised manuscript. We are prepared to also calculate the climate-carbon feedbacks under the Friedlingstein definition if the reviewer wishes, however we feel this would further complicate an already large table.

5. My final request would be to ask if you can more directly or relevantly bring this back to complex models – how does this approach help us develop/evaluate/constrain them further? For example you claim in the discussion that the carbon-climate feedbacks are "less sensitive" than the carbon-concentration ones. And that this is due to "the shape of K(Ca,DT)". So how does that help with my ESM? What controls the shape of this in ESMs? And can we measure and constrain it from obs? If so, then your analysis brings a way in which we might narrow the spread in model projections, or at least evaluate a very relevant aspect of the models. If not, then all it does is leave us with a better feel of why the models continue to diverge – if you have any ideas how to make this jump that would be great to see.

We thank the reviewer for this relevant comment. Of course, the divergence amongst ESMs could well be due to diverging parameterisations, as well as different functional forms. As the reviewer suggests, an interesting area for future work would be to study what effects different forms for key functions such as K have on feedback strengths. Other steps to aid development of ESMs could include analysing the effective shape of functional forms such as *K* in ESMs or how to constrain these functional forms from data. These are beyond the scope of the present work but in the revised manuscript we will point to these possible future directions in section 6.

Minor comments:
1. Page 4. Your NPP function of CO2 claims to include the effects of climate change – but surely these also depend on the climate sensitivity. For models with high/low climate sensitivity, there is a different trade-off of the effects of CO2 and climate. So I don't follow how the impact on NPP can be made without reference back to the temperature as well as the CO2

We agree with the reviewer that an accurate treatment of NPP would separately parameterise the effects of CO2, temperature, rainfall, nutrient availability, and so on. We fold all these effects into a CO2 dependence through Keeling's formula. The references in section 2.1 that we cite for Keeling's formula support this simplification. A key assumption to support this folding, as the reviewer implies, is that we fix climate sensitivity to a constant value -- in the revised manuscript we will state this assumption in section 2.1.

2. Page 7. I couldn't quite see if you had a link between ocean heat and ocean carbon uptake – I don't think so. Should these be related? There might be a false extra degree of freedom in your model – I would expect for example rapidly mixed oceans to have high rates of both heat and carbon uptake – and vice versa for poorly mixing oceans. But if you have independent mechanisms of carbon uptake and transient response to climate do you miss this link?

The reviewer is correct that in our model ocean heat uptake (as represented by climate sensitivity) and ocean carbon uptake are parameterised independently. We also agree with the reviewer that higher ocean mixing rates ought to speed up both carbon and heat uptake. We have chosen to focus our model development and analysis on the carbon cycle; future work could involve incorporating mechanisms related to ocean heat uptake such as ocean circulation, and then specifying common drivers on ocean heat and carbon uptake could be worthwhile. We discuss in Section 6 that energy balance is a potential route for further model development.

3. Table 1: please be careful to stress "climate sensitivity" as "transient climate response" – you do say so, but using the wrong name makes it look like a very low value (1.8K)

We thank for the reviewer for the cautionary note. We will stress that the climate sensitivity in our model is transient climate response in the revised version (see section 3).

4. Page 14, line 23 – just to check here you mean "1% increase up to double CO2" and not a step-change to double CO2.

We thank the reviewer for prompting us to clarify this matter. In fact this simulation has a 1% increase up to *quadrupling* CO2. We will clarify this matter in the revised manuscript (see section 4.3).