Reply to reviewer #2, Chris Jones

Concerning the review on Reick and Torres Mendonça "On the foundation of the α-β-γ approach to climate-carbon feedbacks" submitted to ESD.

March 6, 2026

We very much thank the reviewer Chris Jones for this great comment on our study – great because it adds to our rather mathematical view a pragmatic and strategic perspective on the subject that we – intentionally – didn't address in the paper, but which is definitely needed to run an enterprise like C[4]MIP. In fact, we are tempted to consider the reviewer's comments as a start for a discussion about the usefulness of the α-β-γ approach that – as a side effect – we hoped to provoke with our paper. Likely such discussions have been taking place already for some time in the core group designing the C[4]MIP experimental protocol – the problems with linearity are known at least since the Gregory et al. 2009 paper co-authored by the reviewer –, but, as we can say from ourselves, participating in C[4]MIP for long only from the periphery, we were not aware of them. And, looking into the literature, we think that we are not an exception. Accordingly, we hope that our study will help us all to better understand what we are doing when applying the formalism, what we can expect from it, and what options are available to overcome its limitations.

To answer the reviewer's comments, we reproduce them below (in italic), breaking them down into separate points. But we omit the reviewer's introductory remarks, in particular his summary of our study, by which we feel well understood.

1. *My interpretation of your study is that maybe the previous analyses listed here as "wrongly" stating linear feedback approach, are not necessarily wrong, but more the case they have rather lazily used the phrase "linear".*

It may indeed be that in some of those analyses it is simply laziness, but for quite a number of them this is definitely not the case, as can be seen from the wording: In the original paper by Friedlingstein et al. (2003) we read "that the linear perturbation assumption . . . holds . . . i.e. that the changes are small enough to ignore higher-order terms". And for other examples cited in Appendix A this is also obvious, namely when the formalism is introduced as an "expansion" (Boer and Arora, 2013), by a "Taylor series" (Boer and Arora, 2013; Schwinger et al., 2014), or when the sensitivities are explicitly written as derivatives (Williams et al., 2019).

2. *It is valuable to bring a level of robustness to the definition, but the use of the 1% simulations still appears valid. Your figure 2 panels c,d,e (alpha, beta and gamma) actually show a relatively straight line for each. For example, plotting delta-C against delta-T ("gamma") we see a very close to straight line relationship, and gamma is the gradient of this line. It is thus (for this scenario) approximately constant in time and its definition as DC/DT is a good measure of sensitivity of the carbon cycle to climate change. This*

straight line is likely what many studies refer to as "linear" even though you show this is not the same as having a linear expansion of a feedback. In this sense, the current use of the 1% runs to measure a system sensitivity (the same applies to alpha and beta – also approx straight line responses) – at least up to 2xCO2 (it may break down at higher levels) remains valid. I think this is what you mean by saying that alpha-beta-gamma can still be valid as "difference quotients" (i.e. the gradient of these straight lines) rather than linear expansion coefficient.

The reviewer suggests here a pragmatic interpretation of the sensitivities and thus of the α-β-γ formalism as a whole by considering the responses to be *approximately* linear. We don't think that this is a viable way to escape the fact that the sensitivities are mathematically zero. And, interestingly, such an interpretation has so far not been presented in the literature.

First, while indeed in 1% simulations *some* of the responses defining the associated sensitivities look for *some* models like a straight line at small perturbations, this is not true in general. As we discuss in our paper, this is typically true for α because the climate response to a CO_2 rise is rather fast. For γ the response looks for several models (like MPI-ESM shown in Fig. 2 of our manuscript) quite linear, but there are also models, like CMCC-ESM2, for which this is definitely not the case (see figure R2-1 below). And for β a linear approximation is for almost all CMIP6 models whose behaviour we checked a rather bad approximation, which is plausible as here a memory of several decades determines the system behaviour. And against the claim of the reviewer this non-linear behaviour is also seen for MPI-ESM, as we show in Fig. 2 of the manuscript: The inset to the bottom left plot, showing how β behaves in time, clearly demonstrates that during the first 70 years, where CO_2 approximately doubles, β changes from zero to about 0.9, which indicates a pronounced nonlinearity. A similar behaviour is seen in Fig. R2-1 below for CMCC-ESM2.

One must note that such a linear approximation would only make sense if it could be applied to all three sensitivities, because already when it fails for one of them, the α-β-γ formalism is not anymore consistently applicable: Even if linearity holds only in an effective way, to arrive by linearity arguments at the gain formula – and we consider this formula to be the main result of the formalism – linearity must hold for each of the α, β and γ responses. So in our view, such a pragmatic approach fails.

And even if it didn't fail, we don't see any added value of such a linear fit of model behaviour. One could in principle use it to scale up and down the response between different forcing sizes – but only for that particular scenario simulated with that particular model. But why should we do so? We have anyway the data of the whole simulations. It would also not help to compare the strengths of the feedbacks among different models (one of the reasons to compute the sensitivities): The typical practice is to calculate the sensitivities for a particular level of CO_2 rise (usually $4\times CO_2$) and compare their values at that level. For this practice a linear fit is not needed: As we have shown in our paper, the α-β-γ formalism works consistently even when the sensitivities, defined as time-dependent difference quotients, don't behave linearly. One must accept that

the feedbacks are not only scenario but also time dependent (even for 1% simulations, see bottom right panel in Fig. 2 of our manuscript and Fig. R2-1 in this reply), and therefore the practice to compare their strength among different models *for a particular scenario* and *at a particular time horizon* is perfectly fine. Insofar, our paper provides the theoretical foundation for this practice. If one wants, one can go a step further, and compare the time development of α, β, and γ through the whole simulation period, so that one sees also how the *development* of the feedback strength differs among models. Because linearity – in whatever sense – is not needed, one can in this way even compare the strength of feedbacks in other than 1% scenarios, like overshoot or negative emission scenarios, that gain increasing importance in policy considerations.

Overall, by claiming that the sensitivities can still be valid as "difference quotients" we do not mean – as the reviewer suggests – that this is because they can be considered as gradients of ("relatively") straight lines in response plots (which is in particular not the case for the β response, see above). We claim this validity because, as we have demonstrated in our manuscript, the α-β-γ formalism provides a consistent description of climate-carbon feedbacks for transient simulations without any linearity assumption. And in this description, the difference quotients defining α, β, and γ should be interpreted as a kind of *mean sensitivity* (as opposed to an instantaneous sensitivity defined via a derivative), that can be obtained for any type of response, may it be linear or nonlinear. The consequence is that for transient simulations all feedback metrics are scenario and – typically – also time dependent. And this must indeed be expected for systems with memory. How useful such metrics are is a question to be discussed.

3. *My main comment is that it would be useful to give more depth to the physical interpretation of the feedbacks and processes. We should remember the intended use of this analysis and therefore what we learn about the Earth System from these experiments. So overall, while I like the mathematical rigour and analysis you bring, I miss a little the improved utility of your suggestions. While we can of course change our analysis, what will be the benefit? We can (and likely will) continue to use alpha-beta-gamma in the 1% runs to both see how models have evolved since earlier generations, and also to explore policy relevant metrics like TCRE. If we adopt instead you recommendations what will be able to do that is new or improved (other than a moral high ground of being more mathematically correct!).*

In our opinion, the mathematical rigour we bring here has more benefits than leading to "a moral high ground of being more mathematically correct": it provides e.g. an understanding why the practice of applying the α-β-γ approach to transient simulations yields only scenario-dependent metrics of climate-carbon feedbacks. It also backs our practice in applying it to transient simulations by providing a modified interpretation. At the same time it allows a clearer view on what the framework may be good for and for what it cannot be used: It allows for quantifying how the strength of feedbacks develops within arbitrary scenarios, and how this strength differs between realizations of the same scenario by different models. It cannot be used to understand the reasons for differences in the contributions of feedbacks between different scenarios because memory
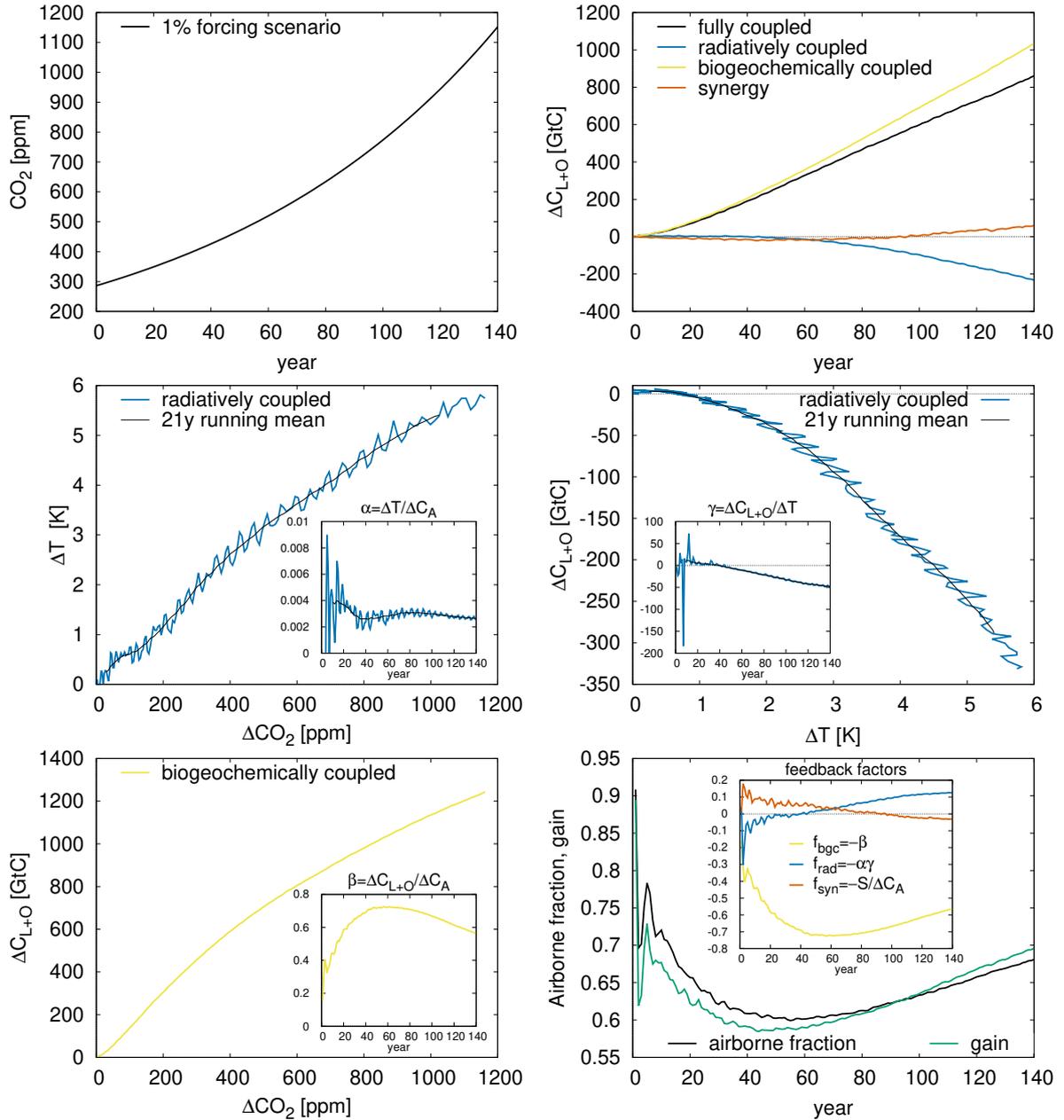
Figure R2-1: Another example for a time dependent feedback analysis of CMIP6 C^4 MIP 1%-simulatons, here for the scenario simulations performed with CMCC-ESM2. In contrast to MPI-ESM (see Fig. 2 in our manuscript), this model shows very clear nonlinear behaviour for the dependence of land and ocean carbon on temperature (middle right plot) so that even in an approximate sense $\gamma(t) = \Delta C_{L+O}/\Delta T$ is not constant. – We added here, as suggested by the reviewer, also a running mean to the plots showing the results from the radiatively coupled simulations. And the reviewer is right: We use here, consistently with the definitions in the mansuscript, carbon and not CO_2 units for the sensitivities, i.e. α [K/GtC], β [GtC/GtC], and γ [GtC/K].

4

is not explicitly represented. In particular this last point calls for considering alternative methods to analyze feedbacks in the coupled climate-carbon system. While all this is not completely new – here and there one finds something on these points in the literature – this knowledge has so far not been systematically brought together – this is what our mathematical analysis also provides.

Concerning the suggestion of the reviewer "to give more depth to the physical interpretation of the feedbacks and processes", we don't see how we could follow it: Our investigation is about a particular *method* to analyze feedbacks in transient simulations, not about the feedbacks themselves – addressing their physical nature is, besides some introductory remarks, out of the scope of our study.

For this reply we have assumed that the reviewer's comment refers to our analysis of the standard α-β-γ formalism. But it may be that the comment was meant to address the alternative approaches we suggest in the outlook. For our opinions on the usefulness of these approaches please see our replies to points 4, 5 and 6 below.

4. *With this in mind, it is not fully clear to me what you are saying we should (or shouldn't) be doing next. The manuscript does not say that alpha-beta-gamma should not be applied to CMIP7 simulations, and it is very likely that it will be. So maybe the recommendation here is that we should be more careful in presentation of this and not refer to it as a linearised feedback analysis, but more a measure of system sensitivity? It remains important to separate the sensitivity of the climate-carbon cycle system to CO2 and climate and we know that these metrics combine to inform us of TCRE (the Transient Climate Response to Emissions) which is a key determinant of the remaining carbon budget. So can you precise exactly how you recommend doing this?*

We didn't write the paper with the intention of giving recommendations for the future of C^{4}MIP. Instead, our hope is that with our study at hand it gets clearer what one is actually doing when applying the α-β-γ formalism. We think that the paper contributes to better understand past results (e.g that the studies by Zickfeld et al. (2011) and Schwinger et al. (2014) actually investigated the synergy, not the nonlinearity of feedbacks) and to see clearer what the limits of this approach are (e.g. it turns out not to be restricted to exponential scenarios, but to be applicable even to overshoot scenarios) and what the alternatives to this approach could be.

Concerning TCRE, even with our re-interpretation of the α-β-γ formalism, this – as other metrics – can be analyzed as before. E.g. the methodology introduced in (Jones and Friedlingstein, 2020) to split the model spread seen in TCRE into contributions from the spread in the different sensitivities is not affected by our study. Nevertheless, by leaving an interpretation of the formalism on the basis of linearity behind, one must realize that the sensitivities, and all metrics based on them, are, even within a single scenario, not constants, but time-dependent quantities – and this is also true for TCRE. This is clearly seen in Fig. R2-2. The consequence is that an additional uncertainty arises from the change of the metrics in time. In our opinion it would thus make sense – as

suggested in our paper – to analyze such time dependencies. Moreover, as is also seen in that figure, for some of the models the synergy is another source of uncertainty, and this uncertainty cannot be attributed to CO_2 or climate alone – this may also be worth investigating in view of the immense importance of TCRE in climate policy.

But maybe the reviewer aims with his request for recommendations at methodologies beyond the standard α-β-γ formalism. As already mentioned above, we didn't write the paper with the intention of giving recommendations. But indeed we think that it sharpens the view on our current practice of analyzing feedbacks in the climate-carbon system, and may thereby be helpful in discussions on how to proceed. Our personal view is that to understand the *dynamics* of the feedbacks in the climate-carbon system, that causes e.g. the non-constancy of TCRE seen in Fig. R2-2, one needs to investigate its memory structure. And for this the appropriate instrument is the generalized α-β-γ formalism. By this instrument one can e.g. understand from an investigation of the relevant time scales in the system how it comes to an *approximate* constancy of TCRE even across radically different scenarios, as recently observed in (Winkler et al., PLOS, 2024) – we have a paper on this in preparation. This doesn't mean that the standard α-β-γ formalism should not be used any more, but it seems that its potential to gain further insight into climate-carbon feedbacks is likely exploited. Nevertheless, as a descriptive instrument its application may still be useful. But these are our personal views that are beyond the scope of our submitted study.

The study suggests other options

5. *Firstly that the analysis could be applied to a steady state response. It's not fully clear exactly what such an experiment would look like. For the climate system, a 2xCO2 or 4xCO2 provides this framing – a top-of-atmosphere imbalance of energy is imposed and the system comes to a new steady state as loss of energy increases to balance the imposed forcing. Is it the recommendation that a step-CO2 change also provides an analysis of a new steady state carbon cycle? The land and ocean sinks will eventually saturate and we can define the response in the same way as the thermal system? As with the climate system, we know this will take many centuries, so we may want an equivalent of a "Gregory plot" to help estimate an effective equilibrium response? It may be beyond the current study to fully define any new analysis, but maybe a few words of what you expect it to look like would be useful – now is timely to test new experiments with some fast ESMs before a full C4MIP set of CMIP7 experiments is decided.*

As explained in the manuscript, by calculating the strengths of *equilibrium* climate-carbon feedbacks these values would be quantitatively comparable with the well known atmospheric feedbacks (Planck feedback, lapse rate feedback, . . .), a comparison that is to date missing. Such equilibrium values characterize the Earth system as such, independent of any scenario. But indeed, the computational costs may be high to run the necessary step response experiments to equilibrium, so resorting to a Gregory type analysis may therefore indeed be a good idea (and this seems feasible: see (Torres-Mendonça et al., 2021, `doi.org/10.5194/npg-28-533-2021`, Figure 9c)). But we think that such a comparison could be a valuable scientific aim so that its worth to explore how it could be achieved.
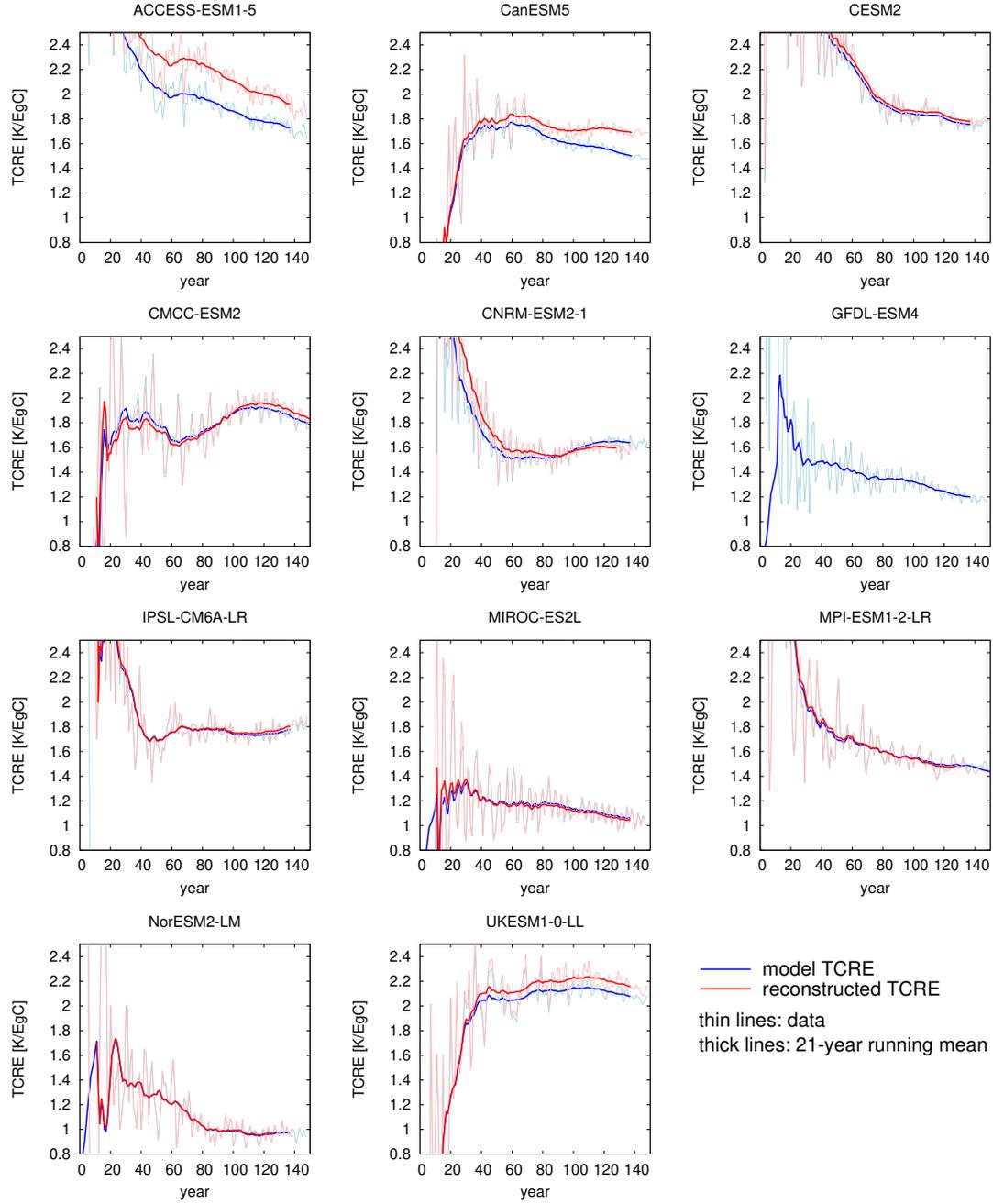
6

Figure R2-2: Development of TCRE over time in C^{4}MIP 1%-simulations performed with a suite of Earth system models as part of CMIP6. The plot shows TCRE obtained in two ways: The blue lines, denoted as 'model TCRE', were calculated from its definition as the quotient of temperature change and cumulated emissions in the FULL simulation – the result is shown as thin curves together with a 21-year running mean. The red lines, denoted as 'reconstructed TCRE', were obtained as in (Jones and Friedlingstein, 2020) from the α, β, and γ sensitivities, namely by $TCRE = \alpha/(1 + \beta + \alpha\gamma)$ – once more the result of these calculations is shown as thin lines together with a 21-year running mean. Here we have used γ as calculated from the RAD simulation (and not from FULL - BGC), so that in the TCRE formula the synergy is excluded (see the discussion in our manuscript and point 12. below). Therefore the difference between the two ways of calculations quantifies the contribution of the synergy to TCRE, i.e. that part of the feedback that cannot be attributed to the biogeochemical or radiative feedback alone. If we had used here γ calculated from FULL - BGC, the two curves would be identical because γ calculated in this way includes the synergy (compare Eqs. (30) and (31) in the manuscript). – By using the terms 'model TCRE' and 'reconstructed TCRE' we follow (Jones and Friedlingstein, 2020).

We will add a few more words on this in the revised manuscript.

6. *Secondly the study propose a more generalised alpha-beta-gamma analysis based on Mendonca et al. This is well noted and C4MIP will explore how to incorporate this into plans for CMIP7.*

We are happy to read this! As already noted in our reply to point 4, personally we think that by this approach one can make indeed a big step forward: In our answer to comment 2. we discussed how to consistently analyze feedbacks for a particular scenario. But, ideally, one would like to understand not only how the system is responding in a particular scenario, but rather what *inherent characteristics* make it respond in that particular way. This is what the generalised α-β-γ analysis has to offer, and, to a limited extent, also a Hansen-type analysis of equilibrium feedbacks. As we express in the last paragraph of the submitted paper, in our view the whole point of constructing a forcing-feedback framework for complex systems with feedback loops is to help disentangle what is internal (feedback processes) from what is external (forcing) to the system, so one can better understand the system's behaviour under any scenario. This is not achieved by the standard α-β-γ approach, even under our re-interpretation, because in this approach the quantified feedbacks depend on the scenario itself. But this can be achieved by means of the generalized α-β-γ framework: Hereby feedbacks are quantified as inherent system properties, independently of the scenario. This leads to a model of the global climate-carbon system by which one can i) make predictions about its behaviour for different scenarios; and ii) study the system itself, separately from the effect of the scenario. This would mean a paradigm shift: one would not anymore focus on scenarios, but on the system alone. Of course one could still study particular scenarios, but by such an approach it would get much clearer how the system and the forcing are individually contributing to the response. And this approach has the very practical consequence of obtaining a model for the global climate-carbon system in terms of the underlying feedbacks, by which one can predict its response and investigate its inherent properties.

Moreover, in agreement with other C[4]MIP members, we think that a necessary next step in the analysis of climate-carbon feedbacks is to consider also non-exponential scenarios, like the policy-relevant overshoot or negative emission scenarios. And to understand the role of feedbacks for such scenarios one must include the memory structure into the analysis, which naturally calls for employing the generalized α-β-γ framework.

Specific comments

7. *Line 27 – you say "particularly the ocean sink" is expected to weaken. I'm not sure why just the ocean? Results show land sinks also weaken. In fact under stabilisation or overshoot scenarios the ocean sink may persist longer than the land in many models.*

Thank you for pointing this out. We will remove this comment.

8. *Line 155 – end of the background section – I like this overview. A nice summary. You could add a little here maybe that the use of the BGC as the reference in Friedlingstein 2003 led to maybe an over-focus of attention on the gamma term as _the_ carbon cycle*

feedback. The parallels with climate sensitivity led to this, but in that case the background stabilising feedback (Planck feedback of a warming black body) was well known. One key outcome of Gregory et al 2009 was the emphasis on the beta term – for the carbon cycle the stabilising effect (response of sinks to CO2) is actually both larger AND MORE UNCERTAIN than the climate response. Hence the reference to pre-industrial allows a more balanced view of both parts of the response.

Thank you for pointing out the role of the Gregory et al. paper for this important shift in view – this was not clear to us. We will add a comment in this direction in the revised manuscript.

9. *Section 4 ("Critique") – you say that there is a strong assumption that global T is a measure of "climate changes" (including precip etc). Yes, this is vital to understand. It is a (relatively) well justified assumption because we know climate patterns scale well with global T (certainly in a 1% run). But it is actually a very poor assumption when we compare the COU and BGC runs, because the delta-T in the BGC run has very different characteristics from that in the radiatively forced run. Literature on this is lacking, but it is one reason why C4MIP does not attempt a more thorough removal of the small warming in the BGC runs. Because the associated gamma is very different and cannot be transferred between simulations.*

Agreed. We will see how to account for this comment in the revised manuscript.

10. *Figure 2c – can you clarify units here? The x-axis is in "ppm", but the inset I think uses units of GtC for the CA. Could these be made the same? Otherwise by eye the inset does not appear to agree with the main plot*

Yes, in the inset we use carbon instead of CO_2 units in order to be consistent with the main text. And we prefer to stick with carbon units, because otherwise we had to include everywhere the respctive conversion factor between ppm and GtC, which makes the presentation clumsy. But we will add a comment to the figure caption.

11. *Also on figure 2 – given climate variability, I wonder if a 20-year mean could be applied (maybe in addition to annual data)? That at least then reflects more what we would measure in our quoting of alpha/beta/gamma*

We will do so.

12. *On the discussion of the synergy term. Sections 5/6, eqn (29) etc. As above, the maths here is robust, but we should also remember the science and physical understanding of what we are analysing and _why_. So in the case of the synergy term, yes it could be measured as a separate term, or included in the BGC term. But as you say, by design in C4MIP we include it in the RAD term, by defining gamma from COU-BGC instead of the RAD simulation. The reason for this is that there is a physical difference (not just a numerical artefact) in what these represent. The RAD run (climate change with no elevated CO2) measures how much carbon is lost due to warming. The COU-BGC pairing measure how much enhanced uptake is hindered by warming. This is quite a different thing and is actually what we want to know – "to what extent will climate*

change affect the carbon sinks due to elevated CO2". Hence, our definition of gamma from COU-BGC is by design what we want to measure.

Thank you for pointing this out. We understand that for this particular question it makes sense to include the synergy into the RAD term. But e.g. for the question of how appropriate the additivity assumption underlying the calculation of the feedback strength is, a separate calculation of the synergy would be necessary – and the synergy may get rather large as shown in (Zickfeld et al., 2011, Fig. 5) when simulations are prolonged until year 2300. But we will account for the reviewer's remark when revising our manuscript.

Christian Reick and Guilherme Torres Mendonça