Reply to reviewer #1

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

February 5, 2026

We thank the reviewer for the detailed suggestions to improve the submitted paper, which we consider as very helpful. In particular we are very grateful for letting us know that some key elements of our study need further clarification. Since all other remarks require only "technical" changes (some more explanations on the mathematics used; further hints to follow mathematical derivations; improvements in wording) that can easily be dealt with, we concentrate in this reply only on those key elements, namely on a clarification "what linearity in this context actually means, and how it is or not related to system memory/inertia" (the reviewer's third and "most important" point).[1]

Linearity

The question of linearity refers to Eqs. (11) and (23) in the manuscript, which are very similar in form but have fundamentally different interpretations. Eq. (11) is the variant most often found in studies, reading[2]

$$\Delta C_{L+O} = \beta \Delta C_A + \gamma \Delta T. \tag{11}$$

In the literature, this equation is understood as a "linearization"[3], meaning that it is the result of a Taylor expansion of ΔC_{L+O} into ΔC_A and ΔT, retaining only the terms of first order in ΔC_A and ΔT. In this expansion, β and γ are the respective partial derivatives of ΔC_{L+O} evaluated at $\Delta C_A = 0$, $\Delta T = 0$, i.e. at the preindustrial point where the scenario sets in. Being evaluated at this point, β and γ are constants, meaning that their values are independent of ΔC_A and ΔT. The consequence of this constancy is that each of the two right-hand-side terms of (11) contributing to ΔC_{L+O} doubles, triples, quadruples etc. if the value of ΔC_A or ΔT doubles, triples, quadruples etc., i.e. the value of these terms depends linearly on the respective value of ΔC_A or ΔT. This is why in the literature the α-β-γ formalism is understood to be a linear formalism.

Next it is explained why Eq. (23) of the manuscript is not a linear equation. It may be noted that (as we show in the manuscript), Eq. (23) is justified in applications to transient scenario simulations, while Eq. (11) is not. Eq. (23) reads

$$\begin{aligned} \Delta C_{L+O}(\Delta C_A, \Delta T) &= \frac{\Delta C_{L+O}(\Delta C_A, 0)}{\Delta C_A} \Delta C_A + \frac{\Delta C_{L+O}(0, \Delta T)}{\Delta T} \Delta T \\ &= \beta(t) \Delta C_A(t) + \gamma(t) \Delta T(t). \end{aligned} \tag{23}$$

[1] We use the notation introduced in the reviewed manuscript.

[2] Actually, Eq. (11) consists of two equations, the other one, being equally fundamental to the α-β-γ formalism, is $\Delta T = \alpha \Delta C_A$. But we don't consider this second equation here in order to keep focused on the issue of linearity. Nevertheless, all considerations in this reply apply also to that equation.

[3] See e.g. `https://en.wikipedia.org/wiki/Linearization`

To understand these equations it is worth to give some additional explanations (also found in the manuscript): The first line is obtained by starting from carbon conservation under the additional assumption that the response in land and ocean carbon $\Delta C_{L+O}(\Delta C_A, \Delta T)$ is an additive combination of the biogeochemical response $\Delta C_{L+O}(\Delta C_A, \Delta T = 0)$ and the radiative response $\Delta C_{L+O}(\Delta C_A = 0, \Delta T)$ so that $\Delta C_{L+O}(\Delta C_A, \Delta T) = \Delta C_{L+O}(\Delta C_A, 0) + \Delta C_{L+O}(0, \Delta T)$. The first line of (23) then follows by plugging into the first term the identity $1 = \Delta C_A/\Delta C_A$, and similarly into the second term $1 = \Delta T/\Delta T$. Next one defines the ratios appearing in that first line as β and γ, which gives on the way to the second line the intermediate equation

$$\begin{aligned} \Delta C_{L+O}(\Delta C_A, \Delta T) =& \beta(\Delta C_A)\Delta C_A + \gamma(\Delta T)\Delta T \\ \text{where } \beta(\Delta C_A) :=& \frac{\Delta C_{L+O}(\Delta C_A, 0)}{\Delta C_A} \text{ and } \gamma(\Delta T) := \frac{\Delta C_{L+O}(0, \Delta T)}{\Delta T}. \end{aligned} \tag{R1-1}$$

Note that here β and γ are not *constants* as above, but introduced as *functions* of their respective arguments ΔC_A and ΔT. Applying this equation to scenario simulations, ΔC_A and ΔT are time dependent. To make this time dependence explicit, one had e.g. to write $\Delta C_{L+O}(\Delta C_A(t), \Delta T(t))$ instead of $\Delta C_{L+O}(\Delta C_A, \Delta T)$, which would be rather clumsy. Therefore it is shorter to write instead $\Delta C_{L+O}(t)$. Proceeding similarly with the other terms in Eq. (R1-1) one obtains

$$\Delta C_{L+O}(t) = \beta(t)\Delta C_A(t) + \gamma(t)\Delta T(t), \tag{R1-2}$$

Except for having made explicit here also the time dependence at the left hand side, this is our Eq. (23) of the manuscript. And in this form it looks very similar to Eq. (11). But in contrast to Eq. (11) it is not obtained by a linearization, and in fact this equation is not linear. This is seen best by its intermediate form (R1-1): When doubling, tripling, quadrupling etc. ΔC_A, the first right hand side term is not doubling, tripling, quadrupling as well because also the factor $\beta(\Delta C_A)$ changes when ΔC_A changes. Similarly the second term doesn't behave linearly when changing ΔT. So the reason why Eq. (23) is in contrast to Eq. (11) not a linear equation is that now β and γ aren't constants. And this is also why Eq. (23) can't be understood as a linearization. Nevertheless, as shown in the manuscript, the whole α-β-γ formalism can be derived by this equation, so that a linearization is not needed.

The role of system memory

Now, how does memory enter the picture and how is it related to linearity/non-linearity of the response? In the following we first make a bit more explicit what is explained on this issue in the main text of the manuscript. Then we have a look at this issue from the more general viewpoint underlying the derivation of the vanishing of the sensitivities presented in Appendix B of the manuscript.

In the main text the relation between memory and linearity is discussed by means of the simple, single time scale system $dX/dt = bF(t) - x/\tau$ (see Eq. (13)). This is a system with memory, where τ is the relaxation time. That the system has memory is also visible from its solution $X(t) = \int_0^t ds \exp((t-s)/\tau)F(s)$, which is mathematically a convolution where not only the forcing at time t contributes to the response, but also values of the forcing at earlier times $s < t$. Taking this system it is shown that when expanding X into a Taylor series of the forcing F at $F = 0$, the linear term vanishes, i.e. the sensitivity $\kappa := dX/dF|_{F=0}$ is zero and the first non-zero terms are non-linear in F. What is missing in the main text is a discussion of the converse case, namely what happens in a system without memory. Using the notation of the single time scale system, this is the case when X is a *function* of the forcing F, i.e. the response X depends only on the *value* of the forcing, and not, as in the convolution above, on the whole history of $F(t)$. Hence, taking $X(F)$ and performing a Taylor expansion into F gives to linear order $X = \kappa F$, where $\kappa := dX/dF|_{F=0}$ is the constant sensitivity, analogue to β and γ in Eq. (11), and there is no reason for κ to be zero. Looking at a transient scenario, F is time dependent and the expansion then reads $X(t) = \kappa F(t)$, i.e. the response $X(t)$ happens instantly at the same time t of the forcing $F(t)$. Accordingly, the non-vanishing of the sensivity and thus the validity of the linearization is equivalent to the assumption that the response follows the forcing instantly, or, mathematically, that the response is a *function* of the forcing and not a *functional* as in the presence of memory.

The relation between memory and the vanishing of the sensitivities, i.e. of the linearity/non-linearity of the response is from a more general point of view proved in Appendix B of the manuscript. There it is explained that generally for not too large forcing F the response X of a system to that forcing may be written as

$$X(t) = \int_0^t K(s)F(t-s). \tag{R1-3}$$

Here, the function K is a characteristic of the considered system that indicates whether the system has "memory" or not. For the simple single time scale system discussed in the manuscript, K is given as $K(t) = \exp(-t/\tau)$ (see above).[4] Let's now apply formula (R1-3) to our situation. For the sake of simplicity we consider the biogeochemical carbon response where $X = \Delta C^{bgc}_{L+O}$, which depends only on the forcing $F = \Delta C_A$[5]. In this case, the response reads

$$\Delta C^{bgc}_{L+O}(t) = \int_0^t K(s)\Delta C_A(t-s). \tag{R1-4}$$

If the system had no memory, the function K would collapse to a Dirac delta function, i.e. $K = a\delta(t)$, where $a \neq 0$ is a constant. Hence, the integral would simplify to

$$\Delta C^{bgc}_{L+O}(t) = a\Delta C_A(t). \tag{R1-5}$$

Setting $a = \beta$, we would then recover the first term of Eq. (11) (with the only difference that in Eq. (11) the time dependence of ΔC_A is implicit). Seeing Eq. (R1-5) as a linearization of ΔC^{bgc}_{L+O} into ΔC_A, β would then be the linear coefficient, which in this memoryless

[4]Note that by a change in the integration variable $t - s \to s$ one can show that $\int_0^t K(s)F(t-s) = \int_0^t K(t-s)F(s)$.

[5]The following argument applies also to the radiatively- and fully-coupled responses.

case would be typically different from zero and thus the traditional interpretation of the framework would be justified.

Nevertheless, we know that in reality the coupled climate-carbon cycle system does have memory. In this more general case, K is not anymore a delta function, but some other function that may start at some non-zero value and then slowly decay back to zero (because forcings at times deeper in the past contribute less to the recent response). What we show in our manuscript (both for the single timescale system and for a general system in Appendix B) is that when writing the response in terms of the forcing as in Eq. (R1-4) that explicitly accounts for the memory, and then expanding this response in the forcing by a Taylor series, the linear coefficient of the response vanishes, leaving only the nonlinear terms. Hence, the memory of the system leads to a Taylor expansion with only nonlinear terms, which cannot be understood as a linearization.

Final remarks

With these remarks we hope to have presented more clearly the problem to base the α-β-γ formalism on the linearization performed to obtain Eq. (11). For completeness it shall be remarked that while these considerations are mathematically true in general, in practice the assumption of a linear response may under certain conditions be a very good approximation of system behaviour even in the presence of memory: this is the case if the time scale at which the perturbation changes is much slower than the reaction time of the system.[6] For the climate system this seems to be the case for the response of temperature in e.g. 1%-simulations: In the α-β-γ formalism this response is characterized by the α sensitivity. Therefore the associated linear equation $\Delta T = \alpha \Delta C_A$ is in practice applicable.[7] But for β this approximation is definitely invalid, and to a lesser extent also for γ, as was recognized already shortly after Friedlingstein et al. developed their α-β-γ formalism (see references in the manuscript).

What is explained here can surely not be included in full length in the revised manuscipt. But we will try to include at least part of it to improve the clarity of the explanations without deteriorating the readability of the text.

Christian Reick and Guilherme Torres Mendonça

[6] Mathematically this can as well be demonstrated by the example of a system with single timescale: Considering its solution $\Delta X(t) = b \int_0^t ds \, \exp((t-s)/\tau) F(s)$ (see Eq. (15)) for the case that the memory time τ is so short that the forcing $F(s)$ under the integral essentially doesn't change while the exponential integral kernel drops to zero, the effect of the integral is essentially to evaluate $F(s)$ at time t. Hence one can approximate that solution by writing $\Delta X(t) \approx bF(t) \int_0^\infty ds \, \exp((t-s)/\tau) = \tau b F(t)$, where τ and b are constants so that this is a linear relation between response and forcing. – We intend to add this consideration to the manuscript.

[7] This is found as second entry to Eq. (11) in the manuscript.