First of all, I would like to thank the reviewer for carefully scrutinizing the manuscript, resulting in a very insightful and helpful review. Below I will reply to the reviewer's comments argument by argument, hereby highlighting the reviewer's elaborations in italic font. In particular, I will refer to the paper that was pointed out as missing literature. Indeed and regrettably, the Nichols et al. (2020) paper was overlooked, but it was a pleasure reading it and comparing it to my work. I believe that these two research pieces complement each other well, which provides motivation to improve my manuscript.

Before going into the reply, I would like to emphasize the key insights that my manuscript provides to the community. The manuscript provides the following novelties:

1. Suggests a simple method for future users to diagnose scenario-dependent deviations from the carbon budget approach for models of any complexity (not restricted to simple climate models).
    - From the pulse experiment, one can judge how scenario-independent TCRE is for fixed cumulative emissions
    - This has been shown in the manuscript in Sect. 3
    - The full model and Green's equation give the deviation in the same order of magnitude
2. It provides a formula for the state-dependent carbon budget approach, an alternative to the one provided by Nicholls et al. (2020).
    - In comparison to assuming the function in advance and then calibrating it to peak warming, the equation derived in the manuscript is motivated by thermodynamic properties —extract TCRE as a function of T, use a first-order Taylor expansion between TCRE and T, and then solve the differential equation.
    - In a sense, I provide an orthogonal approach that complements the one provided by Nicholls et al.
    - Both versions allow for both convex and concave relationship between cumulative emissions and temperature increase
3. I will show the qualitative behaviour of the pulse response under uncertainty in climate parameters (the demonstration is given in this reply)
    - By inspecting the behaviour of the pulse response under higher climate sensitivity, one can show that there is a limit to the carbon budget approach for a combination of high climate sensitivities and higher cumulative emissions.
    - The state-dependent equation under climate uncertainty is ready to be derived, allowing for expanding the utilization of Eq. 6 (manuscript) in the climate-economic assessments under uncertainty

Please find my replies as follows.

## "General comments"

The reviewer captured the points of the paper correctly, with one correction on the sentence: „*They note that neither Green's formalism nor their new equation can capture scenario-dependence*". This is not quite the case. One can see that in the manuscript in Figures 1 and 2, right panels, that for various setups, Green's equation can indeed capture the scenario dependency of the full FaIR to a very high extent, especially in the mitigation scenarios (lower total $F_{tot}$).

Hence, one innovation the paper presents is that once a pulse response of a climate model was generated, Green's equation can predict the order of magnitude of the maximum possible scenario dependence for such a budget. Hereby, Green's method is not restricted to a "simple" climate model. One can, in theory, use the pulse response generated by ESM and feed it into Green's equation, run the min/max optimization (Eq 4), and get the deviation while avoiding running the full-fledged (!) ESM in the optimization scheme (which would likely be impossible due to computational costs). In essence, the manuscript suggests that Green's equation is a practical tool to check how the carbon budget approach works for any model. In a potential new version of the manuscript, this innovation would be made clearer.

## "Missing key literature":

As I mentioned in the introduction, there was an oversight when scanning the literature resulting in claiming novelty in the state-dependent carbon budget equation (Manuscript, Eq 6). The reviewer points out that the equation serving the same purpose was derived in Nicholls et al. (2020), Eq. 4*. (I would like to reflect on Eq. 4* and my understanding of Nicholls et al. (2020) in order to compare Eq. 4* and Eq. 6, so I can simultaneously reflect on the questions raised in the review section). In the following, I reflect on the Nicholls et al. (2020) paper in high detail in order to clarify (what I see as) differences between the approaches.

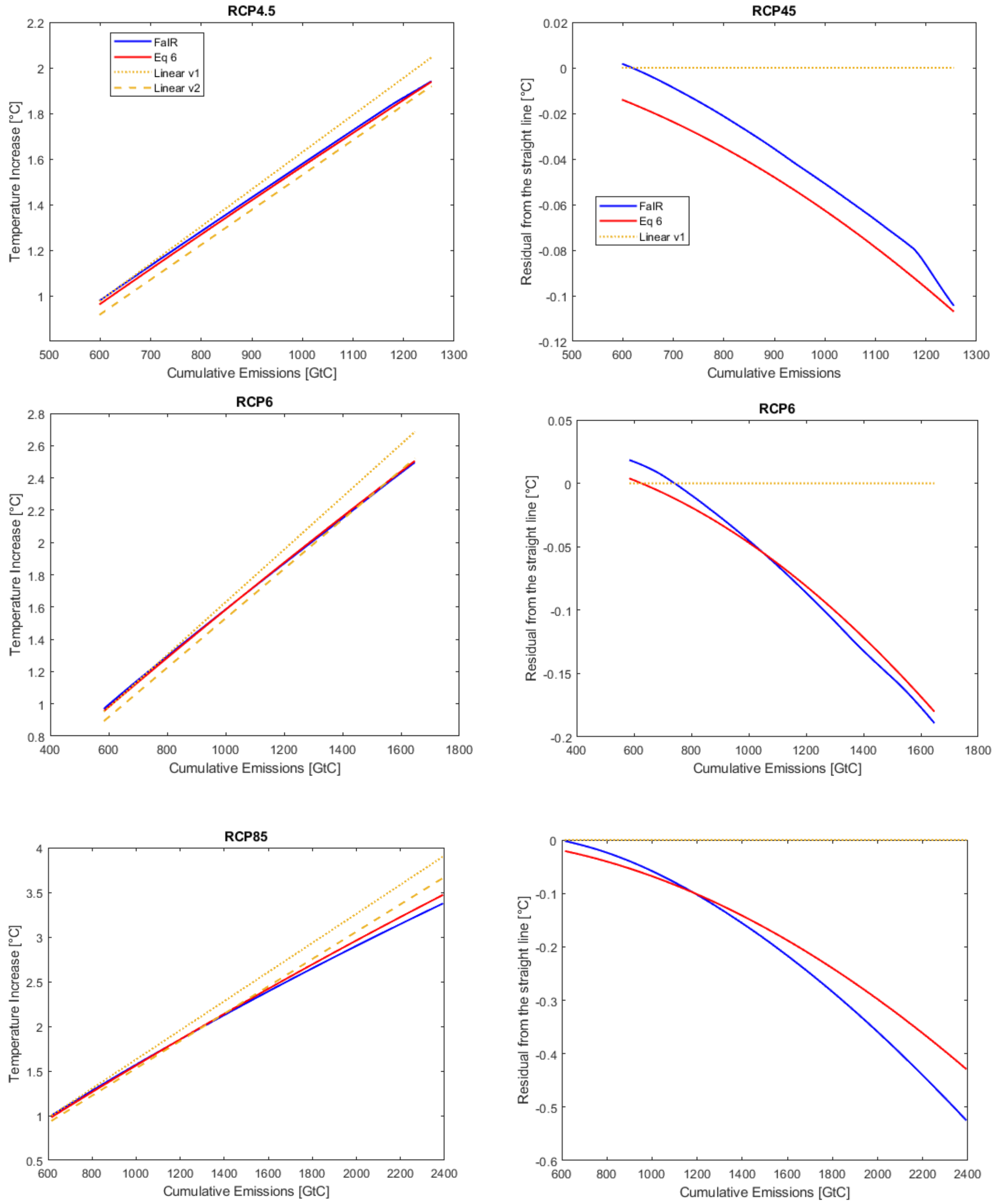
Firstly, while reading Nicholls et al. (2020), it seems that the dots (Figs 1, 3 & 4) are referred to as both the peak warming and the warming at the time of the emission cessation. In my manuscript, I argue that those two are closely related. Additionally, looking at Fig 3, one can detect a large spread between the dots: the same cumulative emissions values give largely different temperature value (e.g. ~0.4 °C difference for ~1250 GtCO2). Unfortunately, I did not grasp whether the difference comes from different parameters, difference in non-CO2 forcers, or the warming definition (mentioned above). In any case, this leads to Eq. 4* and Eq. 6 not being comparable to a full extent, as Eq. 6 (manuscript) relates to the transient warming (not a peak warming) and also excludes non-CO2 forcers (as the reviewer already pointed out and I refer to it later in the reply).

<u>Comparison between Eq. 6 (Manuscript) and Eq. 4* (Nicholls et al. (2020)):</u>

Coming to the derivation of Eq. 4*, the authors assumed the logarithmic relationship between cumulative emissions and temperature as a starting point. Here, I question whether the logarithmic relationship between atmospheric carbon and radiative forcing is a valid ground to assume the logarithmic relationship between (cumulative) emissions and temperature; my suspicion comes from a convex relationship between cumulative emissions and atmospheric concentration that is a necessary link between emissions and temperature. Hence, the assumption is that a concave (logarithmic) relationship outweighs a convex one.

Coming back to my manuscript. Contrary to Nicholls et al. (2020), I do not assume a relationship in advance but derive it from approximating the pulse response with a TCRE that is dependent on the cumulative emissions and temperature increase.. Moreover, the method that I suggest (Eq. 5), does not require a pulse response at all, but one can map TCRE directly from the run (as temperature over cumulative emissions) and then proceed with deriving the equation. After mapping the TCRE that varies with temperature, I derive the equation using a physical quantity (temperature). Additionally, Eq. 6 can also be convex if parameter a changes sign. This has not been empirically observed in my experiment, but Eq. 6 does not theoretically restrict convexity. Furthermore, Eq 6. in the manuscript is strictly concave because only a single parameter combination (connected to the temperature saturation and carbon cycle weakening terms) was used. As I will show later, the pulse representation provides a hint about where (or when) the concavity in question breaks down. The discussion will show how the (manuscript) work continues with different parameter combinations, hence then including uncertainty.

To complete the comparison, I generated graphs inspired by Fig. 4b/d (Nicholls et al.) using the method and the runs from my manuscript. The following figures' left panels are to be compared with Fig. 4b, while the right panels are comparable with 4d. Note that the graphs are for demonstration purposes only.




At first glance, the right panels suggest that FaIR (manuscript) deviates much from linearity much more than MAGICC. This is not the case because I used GtC as an emissions metric and count from the preindustrial era, while Nichols et al. use GtCO2 and count from 2010. Therefore, the last figure (right panel RCP85) covers by far the highest range of cumulative emissions.
Furthermore, comparing the right panels with the Fig 4b in Nicholls et al. shows that Eq. 6 arguably emulates FaIR better than Eq. 4* MAGICC. However, this is no actual proof that my approach is more precise due to the much smaller emission pathway sample size.

Concluding the comparison („Missing key literature“) section with the reviewer remark:„*Missing this paper (unfortunately) undermines the author's claims of complete novelty*.“. Regardless of a comprehensive scrutinization of the two approaches, the reviewer is, without a doubt, right about my statement of novelty. I would gladly revise the statements and incorporate new findings from Nicholls et al. 2020 in the revision.

## „Writing“:

"*The paper reads like it hasn't been proof-read for basic errors (typos, missing words etc.) or overall structure and fluency*". I truly regret that errors on this level occurred in the discussion paper. Measures have been taken that in case a revised manuscript is allowed for, that version will see a thorough proofreading by a native speaker. Thank you for pointing this out.

On the Heaviside function and its purpose in the text:
I fully see the reviewer's point about the triviality of the finding that the Heaviside function by itself gives the perfect budget, as it is (extremely) intuitive. The idea to put it in the text was not so much to show that perfect budget can be represented by the Heaviside function, but rather that using Heaviside in a Green's formulism is an intermediate, first-hand proof that Green's formalism is one way to describe the carbon budget equation in a <u>general</u> way (in which, I later show, it is possible to show scenario-dependent effects as mentioned in the "General comments" reply). Finally, indeed because of the pulse being close to Heaviside function (later after the relaxation), we have a very weak (and fainting) scenario-dependency of the cabon budget (Manuscript,Figs 1,2,3).
However, I see the logic behind putting this part in the appendix or the supplement and am happy to do so in case it is overall better.

"*In general, I would reconsider whether the more trivial observations would instead be better in a supplementary or simply removed (for example, the discussion of different ways one can reduce emissions and still have the same cumulative emissions in section 3.1.1)*."
I am not sure if we mean the same thing, but the idea behind the end of that section was to show why for different run setups, we have different feasibility limits on k, which then effectively also affects the results, the closer we are to the feasibility limit.

The rest of the section on writing consists of very well points that should be incorporated in the revision in any case.

## "Key insights get lost“:

Thank you for this valuable comment. I will expand Section 4 and probably focus much more on it in the new revision with fresh results that include uncertainty in parametrization (revealed in the "Lack of exploration of climate uncertainty" section of the review.

## "Journal fit“:

Again, a very valuable comment. I was thinking about whether to keep the focus on the economics part and possibly change the journal, but I changed my perspective after introducing the uncertainty. I think discussing pulse representation with climate uncertainty fits the ESD and its audience.

## “Clarity of methods presentation“:

“*I would suggest removing discussion of the optimisation in the abstract*”.
The reason for putting this in the abstract is to emphasize that this is a fully novel approach to calculating the carbon budget deviations when compared to previous literature. I found it important to emphasize that in this work, we do not test a portfolio of emissions but generate a maximally possible deviation – and hence the optimization. Maybe I could simplify it, but I still consider that it has its place in the abstract.

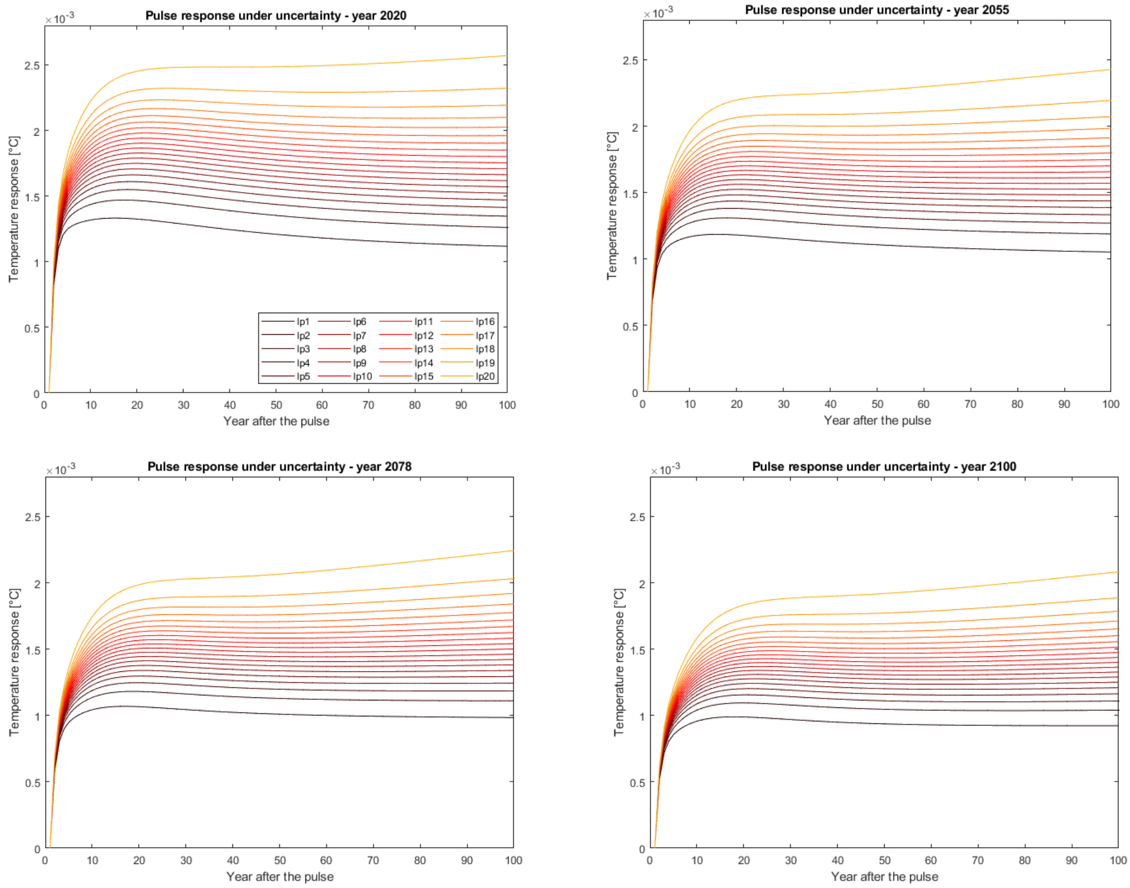
"*On this point, the open budget isn’t clearly presented*.”
The condition E(t>t*=0) is met only for the net-zero budget (I think this version goes in the same line as the budget considered in Nicholls et al) and not for the open budget (which would then go in line with the transient response interpretation). "*Which would seem odd, as then the cumulative emissions that the system sees are now changing from the specified limit up to t**.” They are in fact, not changing, as they are set to be a fixed value ($F_{tot}$) at the time t*. The only thing that this condition does is that it allows (or constrains) emitting after t*, which is reflected only in the fact that because of the slope restriction, the net-zero budget requires the emissions to decrease prior to t* to reach zero at t*, while the open budget does not since then they can take any value at t*.

## „Lack of exploration of climate uncertainty“:

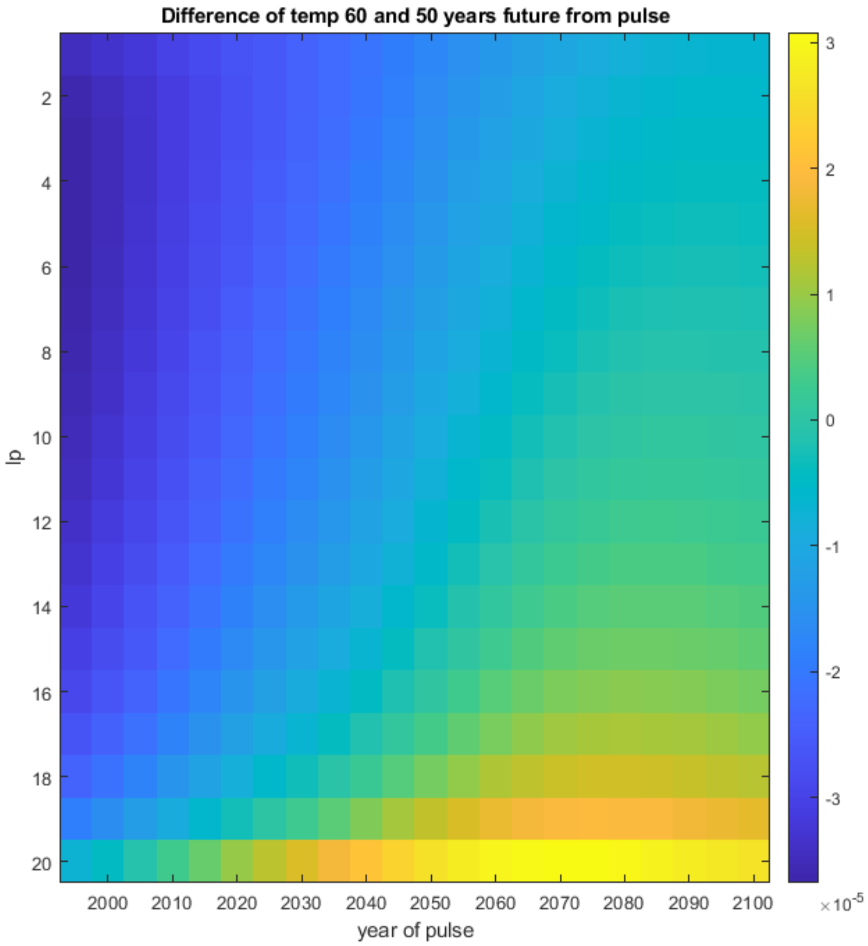
In this section, I can finally show what I was referring to in the sections above; I will show preliminary graphs that depict how I expanded this work to include the uncertainty in the climate response.

To represent uncertainty, we use 20 different parametrizations (lp1, lp2, … lp20) that cover 20 different equiprobable climate sensitivities in accordance with the IPCC AR6 range fitted to log-normal distribution (Forster et al., 2021). Lp1 represents the lowest climate sensitivity (CS), and lp20 the highest.

In the original manuscript, Figure 4 showed the pulse response of one parameter setup under different climatic conditions (different years in the RCP6 run). Conversely, the next figures show the pulse responses of different parameter setups (lp’s) in one year of the RCP6 run.




A few effects can be detected. Firstly, one assumes that using the same method as provided in the paper, one can derive Eq. 6 dependent on CS too, with mapping TCRE (lambda) with both corresponding temperature and CS (not only temperature as in the manuscript).
However, the current hypothesis is that this method will not work for higher CS values (especially not lp20 for example). To understand it, remember that the state-dependent TCRE is an approximation of a pulse response in time when the pulse already reaches relaxation (near constant phase). As we can see in the figures above, the combination of higher climate sensitivities and different climate conditions (later years in the RCP6 run), the pulses change the behavior: instead of reaching a relaxation, they shift the trend towards increasing response in time. Hence, the method for deriving Eq. 6 is not compatible anymore.
This effect connects to the discussion about the convexity of the equation in the first review section. The reviewer states that the concavity in the carbon budget equation diminishes towards linearity (or even convexity), depending on the parametrization. In the pulse response representation (figure above), this can actually be detected in the mentioned shifting response trend. In essence, both higher climate sensitivity values (higher lps) and "higher" climatic conditions (later years of the run) mutually influence the system to be more linear (flattening the pulse response) until some threshold is reached, after which the response starts increasing later in time (most visible for lp20 where this effect is present in every year).
Following up on these findings and keeping in mind that the relaxation of the pulse is necessary for the carbon budget to be meaningful, we add one more finding. The new analysis with uncertainty provides insights into the values of climate sensitivity for which FaIR allows for a carbon budget approach at all. The following figure maps all of the pulse runs for all of the years and all of the CS values, calculating the proximity to the carbon budget (pulse relaxation) for every combination.




We can see how the carbon budget adherence changes with year of the pulse and climatic condition. This is still a work in progress, but in my view, it is moving towards a final product for the ESD.

## „Validation of model and code“:

I am happy to upload the code.

## „Claims about FaIR being 'best' “:

I must admit that this was just a poor choice of wording. The authors suggest using FaIR carbon cycle in combination with the 2-box temperature cycle (effectively FaIR), but never state that it is the best. I think I would follow the reviewer’s advice and simply remove the assessment since, as reviewer suggests, it is not relevant for the paper.

Concerning introducing negative emissions, I thank the reviewer for the hint about FaIR’s ability to capture them. Depending on the scope of the paper, I will see whether and how to include them in the analysis.

## „Lower cumulative emissions level“:

Introducing the lower cumulative emissions level could be done but the analysis would be limited due to lesser feasibility (the feasibility limit is discussed in the contested redundant text at the end of 3.1.1.). Changing the way the emission rate is restricted (non-linear restriction) could maybe help in this case.

## „Leftover temperature“:

Regarding plotting, I think it is a good idea -- it makes perfect sense to include T_left in the supplementary.
“*Could you just run the Green’s model over the historical period too*?” Not really, since I generate a pulse response used in Green’s equation in present day conditions. The pulse response function also changes if we go back in history (like ti changes when we project it into the futue, Fig 4, albeit other way around). This means that we would also get some absolute temperature difference between full model and Green’s equation.

## „Claims about time evolution of T_d“:

Regarding “*However, I am not convinced that this is a general thing rather than just being a model feature of FaIR, which appears to always reaches the same equilibrium for the same*”. In a sense this was exactly my point when extrapolating the conclusion to other simple climate models. The Green’s representation (using a pulse response function) shows whether a climate model can or cannot adhere to the carbon budget approach (and declining of the T_d). Hence, I claim that other models that do not reach relaxation in the pulse response will not adhere to the carbon budget.

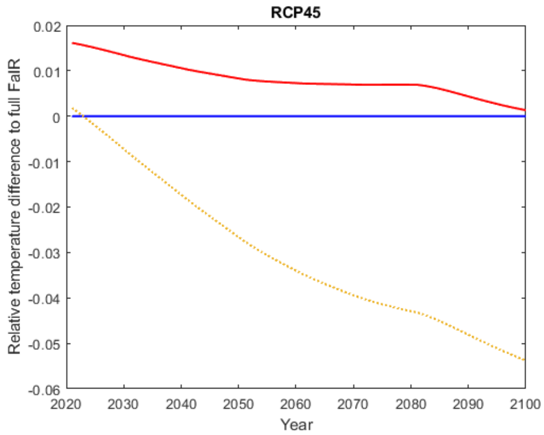
Considering the second paragraph, I think I already reflected on most of the points of the reviewer by now. Regarding the non-CO2 forcers, I purposely left them out because I constrained my research only to carbon budget effects which, to my understanding, are not applicable to other forcers.

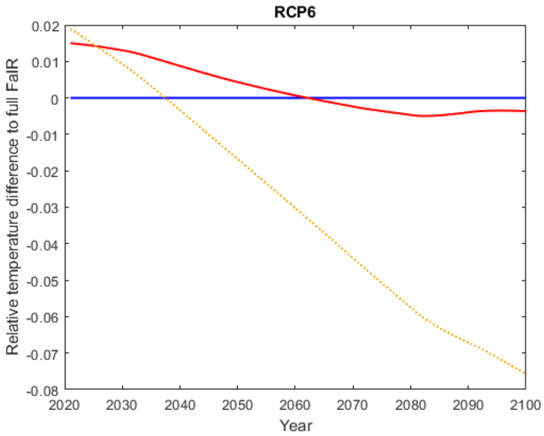
## „Why can't you just use a simple climate model“:

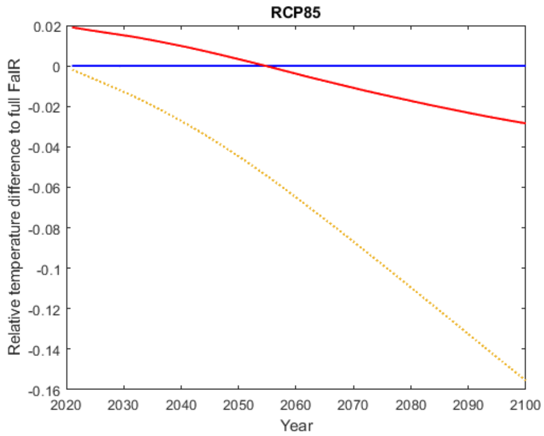
Firstly, the TCRE concept is of interest in its own right. In addition, within climate economics, the implementation of a model of FaIR’s complexity represents a hurdle not every research group is willing to invest in. This holds more so for MAGICC. Finally, the TCRE concept represents an analytic bridge between two target-based decision-analytic concepts: cost effectiveness analysis (CEA) and cost risk analysis (CRA). CEA is a dominant paradigm in IPCC ARs 5-6, WGIII. However, it conceptually cannot deal with decision-making under anticipated future learning, as already pointed out by Blau (1974). For that reason, it might be accused of tackling an ill-posed problem for what reason Schmidt et al. (2011) modified it into CRA as one option for a ‘repaired’ version of CEA. Held (2019) raised the question of under what circumstances CEA-based scenarios can be retroactively justified as good approximations of CRA-based solutions, hence are well-posed under uncertainty. In Held (2019), a set of sufficient conditions is presented, which includes that temperature can be predicted by cumulative emissions (not necessarily linearly). Hence, for a justification of CEA in integrated assessment, the level of accuracy of TCRE for prediction is of high relevance.

## “Figures”:

As requested, here are the plots that correspond to the three runs in Fig. 6, depicting a **<u>relative</u>** deviation from full FaIR (% of the absolute temperature) for each model choice. Thank you for this suggestion, it is indeed quite helpful to get a clearer idea of what is going on.




## References


Blau RA (1974) Stochastic programming and decision analysis: an apparent dilemma. Manage Sci 21(3):271–276

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J. L., Frame, D., ... & Zhang, H. (2021). The Earth’s energy budget, climate feedbacks, and climate sensitivity.

Held, H. (2019). Cost risk analysis: Dynamically consistent decision-making under climate targets. *Environmental and Resource Economics*, *72*(1), 247-261.

Nicholls, Z. R. J., Gieseke, R., Lewis, J., Nauels, A., & Meinshausen, M. (2020). Implications of non-linearities between cumulative CO2 emissions and CO2-induced warming for assessing the remaining carbon budget. *Environmental Research Letters*, *15*(7), 074017.

Schmidt, M. G., Lorenz, A., Held, H., & Kriegler, E. (2011). Climate targets under uncertainty: challenges and remedies: A letter. *Climatic change*, *104*, 783-791.