General comments

Overall, I'd like to thank the author for their efforts revising the paper. The individual ideas are now much clearer and better explained. In particular, I think the ideas about what causes deviations from linearity and how these can be examined through impulse response are now much easier to grasp.

Some other especially nice parts:

- lines 49 to 53
- Figure 4
- line 402-413

I also think the paper opens up some nice follow up research on negative emissions and applying these ideas further.

Having said that, I still have some concerns, which I outline below.

Major concerns

Key point of the paper

When re-reading the paper, it wasn't clear to me what it's one, key point was. A few things seemed to potentially be the key point, the ones I could see:

- For economics applications, something easier to implement and understand than a simple climate model is needed. Here is a set of analysis that shows that a pulse response/Green's function based approach is a good approximation, some regimes to be aware of where the pulse response might start to break down (related to state- and scenario-dependence) and a non-linear equation for the relationship between temperature and cumulative CO2 that better captures state-dependence (which could also be useful where that state-dependence is important).
- Here we show how to understand non-linearities in the relationship between cumulative CO2 emisssions and temperature from the point of view of Green's functions/impulse response.
- Here we provide a method that allows us to predict a model's level of deviation from linearity between cumulative CO2 emissions and temperature based on knowledge of its pulse response alone

If any of these are the points, I think there are things missing. Respectively:

- The paper doesn't really go into enough detail to actually give climate economists enough information to understand how using a Green's function approach breaks down and under what conditions and how large this breakdown is compared to other uncertainties (for example, simply in just the size of the TCRE), particularly given how thin the section on climate uncertainty is (perhaps there is enough to tell a climate economist how much error is being introduced using the same parameter set the author used in this paper, but there is very little that could tell an economist how big the error is if they assume a TCRE at the top of the IPCC range, for example).

- If showing how to understand non-linearities in the relationship between cumulative CO2 emisssions and temperature from the point of view of Green's functions/impulse response is the point, then build up the paper that way and use the optimisation and comparison to FaIR as validation, rather than starting from optimisation and working the other way (which is much more confusing in my opinion). If this is the case, I would cut most of the climate economics stuff and save that for a future paper (given what we know about deviations, this is what it means for climate economics).
- If you want to provide a method that allows us to predict a model's level of deviation from linearity between cumulative CO2 emissions and temperature based on knowledge of its pulse response alone, then you need to verify this across a much larger part of FaIR's parameter space and arguably with many more models (or you have to caveat your conclusions appropriately). Further, if this is the intended point, I think some analysis to investigate the number of impulse response experiments required to make robust conclusions would be important (given that the number of impulse response type runs we could do with ESMs will be limited).

I think this concern could be thought about another way. In short, after I read this paper, what should I come away thinking? Should I think that we absolutely have to included non-linearities when doing carbon budget calculations? Should I reconsider using a simple pulse response in my climate economics work? Overall, I read the paper and thought, "A lot of this is quite interesting", but I had no idea what I was meant to takeaway from it nor what the key message was.

Fixing this point would also greatly help the paper's communication. It is greatly improved from the first iteration, but the paper is still quite slow. If the paper's point were clearer, then much of the text could be cut (or moved to the supplementary, if the author can't bear to part with it) because it would be clear that it wasn't directly relevant to the main point.

State-dependence of TCRE

Assuming TCRE = -a T + b seems to not be a great choice to me. Specifically, its limits don't make sense. It seems an odd idea to me to suggest that TCRE could to zero (or even negative), particularly because if TCRE is zero then you can neither warm nor cool, no matter how much more CO2 you emit or remove. Can you re-think this functional form or put some limits on the domain of applicability you think it could have.

Minor concerns

Overblown conclusions

There are only a couple of these, but I would strongly encourage the author to be more cautious with their wording when making conclusions. A key example is in the abstract (line 18-20),

“The analysis shows that using the Green’s function approach to diagnose a model’s carbon budget scenario-dependency, along with the method of deriving the non-linear carbon budget equation, both do not depend on the complexity of the chosen climate model.”

There is absolutely no way you can conclude that using FaIR and the one-box model. They are far too simple. Computational constraints mean we can probably never make a statement like this, because we’ll never be able to do the experiments with an ESM. Something like the following would be an appropriate softening,

“The analysis finds that using the Green’s function approach to diagnose a model’s carbon budget scenario-dependency, along with derivation of a non-linear carbon budget equation, doesn’t depend on the complexity of the two simple models used here, leaving investigation with other and more complex models to future work.”

Exploration of scenario space

The author says that their use of optimisation means they explore a greater amount of the scenario space than other papers. That’s probably technically true, but I think it is a bit of a stretch to say that this is a really novel aspect. Nicholls et al. (2020) used all the SR1.5 CO2 pathways, which cover an already wide range of different rates of mitigation considered plausible.

The optimisation goes further than this in terms of pathway exploration, but whether this is a sensible extension or not is questionable I think, particularly looking at the triangle shape of the pathways in Figure 4c and 4d, which I think prove that the author’s constraints don’t actually prevent unrealistic pathways.

I think it should be noted that the author also applies (arguably arbitrary) contraints on emissions in their optimisation. I think this undermines lines 95-96, “Through the optimization scheme, the full portfolio of emission pathways is tested.” Given that a full portfolio can’t truly ever be claimed, I would suggest removing lines like these, rather prefering statements like, “a wide range” or “range that can be explored freely by the optimiser”.

I found the discussion in 4.1 far more convincing than the optimisation in terms of explaining how you can go from impulse response to scenario deviation. Having read this section, you could cut the optimisation completely and just construct, by hand, pathways that lead to this maximial scenario dependence based on the insights presented in Section 4.1.

In the author’s reply, they said, “I think it is important to keep [the optimisation], since it is the key point that differs how the scenario-dependent effects are examined in the manuscript, compared to the previous literature”. As I’ve said above, I don’t think this element is that novel or key in terms of making this manuscript stand out (and as I've said further above, it wasn't obvious to me that this was the key difference/point of this paper).

Freely evolving case

I still don't understand the freely evolving case. After t* there is no constraint, so couldn't your program just dump out 1000 GtC in a single year? That would cause quite different responses no? It would be helpful if you could explain how emissions after t* are decided in the case that there are no constraints on this period (which is what I understand the freely evolving case to be)? Or is the point if this freely evolving case that you remove the constraint that E(t*) = 0 but leave all other constraints the same? If yes, I think it would help to clarify this in the manuscript.

T_left

As far as I can tell, T_left is only required for comparison to FaIR, and cancles out in the calculation of T_d (because it is the same in both T_max and T_min). If that is correct, why is T_left introduced in 3.1.4 and not in the section where you compare the pulse response based method and FaIR (where it would seem to belong better as essentially a correction factor to deal with different start dates rather than something fundamental to your overall methods)?

Dependence on optimisation year

There is now a supplementary figure showing the deviation as a function of the optimisation year. However, it is only done for a single F no? Don't you need to vary both F and t* to make any conclusion of robustness in this relationship?

Technical corrections

- In response to reviewer 1's point 3, the author said, "However, the inspection of the pulse responses was exactly the reason why I used GAMS and I do not see a way to do it in FaIR." I will just note that it is possible to run FaIR concentration-driven in Python, you just have to do a bit of digging to find the configuration. If you raise an issue on their GitHub, I'm sure they will reply fairly quickly.
- line 61-63: "Regardless of ZEC, the linear segmented framework concept itself has been challenged by Nicholls et al. (2020), who claim that its assumption of a linear relationship between peak warming and cumulative emissions leads to unrealistically low budgets." I think you may have misunderstood the paper. When I look at it, the paper doesn't come to this conclusion at all, particularly given how wide the uncertainties are where they're presented in the main text. The paper has a look at the implications of including non-linearities and basically concludes, they're

pretty small in the context of other uncertainties so ignoring them in the segmented framework is not a terrible approximation.

- line 73-74: "Nicholls et al. (2020) have derived the non-linear carbon budget equation by positing a logarithmic relationship between cumulative emissions and temperature increase". This might be a slight mischaracterisation, it is a logarithm but it can actually go both ways so it isn't always logarithmic saturation (which is how this text reads).
- line 91-94: "At its core, this paper endeavors to define and assess both the scenario- and state-dependent deviations (non-linearities) of the carbon budget approach. It demonstrates that a temperature response to an emission pulse, i.e., the pulse response representation, offers a very convenient tool for doing so." I would blend this sentence into one because the deviations aren't novel to this paper, but the pulse response representation is (noting also the major concern about the key point of this paper being unclear as it is currently written).
- lines 102-104: "allows us to calculate the maximum possible scenario dependency of the ESM models" should be "allows us to approximate the maximum possible scenario dependency of the ESM models" (as the author says, you can never do this with an ESM because it costs too much so you're left with approximations at best)
- lines 113-114: "For climate economists, it reveals the consequences of using models with incorrect pulse representation, in terms of their inability to adhere to the carbon budget approach". If this is the point of the paper, make it more obvious (see also major concern).
- lines 148-150: "Specifically, the sets used in this paper are tuned to the MIROC-ES2L, BCC-CSM2-MR, MPI-ESM1-5, CNRM-ESM2-1, and ACCESS- 150 ESM1-5 models." You should cite the model description papers here (those model developers put huge effort into making the models available and citations like this are key for recognising that effort).
- line 182: "The difference is that, in Eq. (6)" Should this refer to Eq. (2)?
- line 185: "To make use of Eq. (2), one must opt for a shape matching Green's function fg." What does this mean (I think the words 'opt' and 'a shape matching' are the things that make this most unclear)? "One must find an appropriate Green's function"?
- Figure 1: where is "TCREv2" on this figure
- lines 192-193: "In the year of pulse response generation tp, the emission pathway necessary to keep the level of atmospheric concentration Ca(tp) constant is generated" Why do you need this? Keeping Ca constant leads to ongoing warming. Don't you want emissions that keep temperatures constant (or just add the pulse to the RCP emissions and don't worry about the fact that temperature is changing)?

- line 214: “would show deviations” -> “would not show deviations”?
- line 279: “again, independent of ZEC, as explained above”. You get to this point later in the manuscript, where as you say it’s only independent of ZEC if the ZEC is also path-independent, which I don’t think it is so ‘approximately independent of ZEC’ throughout is probably better/necessary here (possibly also with a forward-reference to your later discussion)
- line 352: “a deviation of 0.15C is produced.” What is this in percentage terms?
- line 357: “In the supplement material, various combinations of the same cumulative emissions and different t*’s show that the deviation not being a function of the optimization year is a robust result.” Tell the reader which supplementary figure. Looking at the figure, it is only done for a single F? Don’t you need to vary both F and t* to make this conclusion of robustness?
- line 393: “Fig. 1, left graph shows the FaIR-generated Green’s function (blue).” You can delete sentences like this, they add nothing on top of what comes after. The context can simply go at the end of the point you’re making. In this case, at the end of the next sentence.
- Figure 9d: something wrong? Red line in ACCESS disappears
- line 558: “Additionally, it would be interesting to see to which extent FaIR tuned to a CMIP6 model reproduces the behavior of its corresponding ESM under the same setup”, Doesn’t Leach do this as part of their analysis i.e. don’t you already have the answer? Or do you mean something else?
- lines 617-619: “In the context of adhering to the temperature target, the declining temperature following emission cessation leads to non-intuitive policy recommendations, namely, to perpetually (albeit at a decreasing rate) continue emitting in order to adhere to the target.” I’m not sure I agree with this? Doesn’t this just mean that you’ve assumed the target is a stabilisation target whereas there is no precedent for that in e.g. the Paris Agreement (which isn’t a stabilisation target, but rather a ‘do not exceed’ target, and arguably a temperature decline target given Article 4’s wording). I would re-word this.