Reply on

Interactive comment on **"On the Future Role of the most Parsimonious Climate Module in Integrated Assessment"** ***by*** **Mohammad M. Khabbazan and Hermann Held**

Anonymous Referee #2

Received and published: 16 June 2017

First of all, we would like to thank the referee for an exceptionally thorough and thoughtful review!
Original comments by the referee will be highlighted in italic font below.

1 General Comments

Khabbazan and Held's paper checks the performance of a one box energy balance model (PH99), currently in use in the integrated assessment models FUND and MIND, against output from AOGCMs before suggesting a simple, improved *way to use it in future. Their major conclusion is that, for strong mitigation scenarios, prescribing ECS and TCR to PH99 from Forster et al. (2013) with no further calibration implicitly causes researchers to sample much larger temperature responses than they intend to. They show that a simple fitting exercise rectifies this and validate the fit by checking PH99's performance under one other scenario. This scenario is very similar to the one they used for fitting. They then explore different methods to map AOGCM ECS and TCR onto 'effective' PH99 values which could provide researchers with a simple method of revealing the temperature response they are actually considering.*

We feel our ms perfectly perceived by the referee.

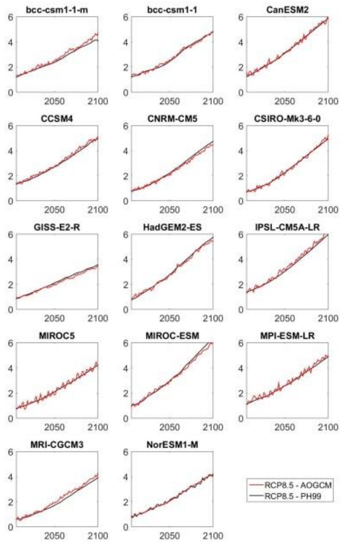
My major concerns focus on whether the analysis shows that PH99 is a valid energy balance model rather than a fitting tool.

That is a difficult question, an almost philosophical one. If we were invited on a new version (termed 'V2' thereafter) we would present a physical mechanism that represents one option to explain the discrepancies produced by PH99. We can show that the discrepancies already occur when replacing a 2-box model like the one implemented in DICE (Nordhaus and Sztorc, 2013) by PH99. So if one had the request PH99 should be able to work with AOGCM parameters, then PH99 would

no longer qualify as an energy balance model. On the other hand, by interpreting the observed effects one could still give it some physical meaning.
Also, V2 will avoid the impression that we advertised utilizing PH99. We simply want to state how to interpret older work based on PH99, and how it could be used if someone wants to use it in the future. In addition to computational efficiency, for some applications also analytic tractability or conceptual simplicity might be arguments for using PH99.

I also think that the writing style could be greatly improved.

V2 will be improved in that regard.



Reply-Figure 1: Intercomparison of global mean temperature from various AOGCMs and PH99 for RCP8.5. PH99 was fitted to RCP2.6.

I think the authors point out some key errors which arise if PH99 is used without care and explore a few ways for modellers to quickly relate their parameters to AOGCM ECS and TCR values. However, given that the authors argue for mapping

AOGCM properties onto 'scenario-class-specific values before using them in PH99', which appears to undermine any physical basis for PH99, I'm left wondering if this paper highlights the limitations of PH99 rather than providing strong arguments for its use.

In fact we expand on the scope of utilizing PH99. We can show that after proper recalibration on AOGCM-RCP2.6, that calibration can be utilized for all RCPs – see Reply-Figure 1 on RCP 8.5 for illustration.

2 Major concerns

1. *The re-callibration of PH99 is only validated for RCP4.5. There is no other testing of the performance outside of RCP2.6 and RCP4.5, two very similar scenarios, nor testing of the effect of different non-CO2 forcing pathways. Thus the authors have shown that a good fit to AOGCM GMT output can be done with two free parameters and that this fit is good for a similar scenario. I wonder if testing over a greater range of other scenarios would strengthen the justification for the use of PH99.*

We will show in V2 that PH99 emulates *all* RCPs by the identical recalibration. This also contains RCP6.0 with some stronger ozone component. Furthermore, this ms is *not* about how to generate a meaningful global forcing from non-CO2 agents, but how to prognose global mean temperature from total global forcing. V2 will be much clearer about this scope. The effects by moving from a 2-box to a 1-box model will play a prominent role in V2. How to get to a total forcing, however, is a discussion that hits *any* climate module utilized in integrated assessment, and is beyond the scope of this ms.

2. *The initial testing of the performance of PH99 against AOGCMs reveals a key, hidden, bias of this model if used without validation in strong mitigation scenarios. This is a good bit of analysis. As a result of this analysis, the authors advocate mapping AOGCM climate system properties onto 'scenario-class-specific values before using them in PH99'. Whilst this seems to be necessary for acceptable performance of PH99, it also appears to undermine any physical basis for PH99. If you have to re-callibrate PH99 every time you want to use it in a different scenario class then its parameters lose all physical meaning and instead simply become fitting parameters. Thus the authors appear to advocate shifting PH99 from an energy balance model to a function that can be fitted to AOGCM data and then used for a limited range of scenarios?*

We now know that we undersold PH99 in that regard. This underselling will be stopped in V2. Instead the underlying physical mechanisms will be explained from which also the observed discrepancy will be derived.

3. *I don't think I am wrong in saying that this model is ultimately meant to be used by those who are looking for simple emulators of global mean temperature response and hence may not be climate modellers themselves. If this target audience can't pick up this paper and get some sense of what is going on then they will struggle to use any of the fits provided. A paper on 'the most parsimonious climate module' should have a style which reflects its title. Given that parsimonious is synonymous with 'simple' in this context, it makes sense for the communication to be as plain, clear and simple as possible too. With this goal in mind, I suggest numerous technical corrections and ask for multiple clarifications.*

We fully agree and are extremely open to the detailed changes the reviewer detailed below.

4. *The exploration of different possible parameterisations of the relationship between AOGCM ECS/TCR and effective ECS/TCR is, in my opinion, worthwhile. My impression is that they recognise that a parameterisation would be nice but don't have strong enough evidence to recommend any of the ones they have tried and so the results here are underwhelming.*

V2 would strive at discussing those fitting procedures with the new physical interpretation at hand.

3 Specific Comments

1. *As an exercise, the fitting that is done is scientifically sound re methods, assumptions, results, and reproducibility as far as I can tell. I can also see that it would be useful for modellers who wish to use a simple emulator but don't wish to do the calibration themselves.*

Thank you!

2. *I think this paper shows that PH99 is closer to a fitting tool rather than a physical model. Hence I wonder, if modellers are after computational simplicity and a fitting tool, why wouldn't they use a simple carbon budget target or emissions pathway to constrain their model. There is already research on how emissions pathways and targets map to temperature targets so this could be used to back out emissions constraints from a given temperature target for a given scenario class. This approach seems far simpler than introducing an energy balance module which requires atmospheric concentration and radiative forcing input, has little physical basis and hasn't been validated over a wide range of CO2 and non-CO2 scenarios so might not produce realistic temperature*

projections anyway.

We are grateful for this exciting suggestion. It is a fascinating question indeed when a dynamic climate module could simply be replaced by the carbon budget approach to deliver similar – or even better quality – in emulation of an AOGCM. This discussion is beyond the scope of this ms, however we would highlight this option in V2. Here we simply would like to stress that there are applications in climate economics where timing matters, such as cost benefit analyses (see e.g. Nordhaus and Sztorc, 2013) or cost risk analyses (see e.g. Neubersch et al., 2014).

3. *The introduction calls IAMs an 'indispensable tool'. I acknowledge that this comment is made in the context of 'driving welfare-optimal climate policy scenarios' so it is accurate. However given that there are many who disagree with using economic analyses for determining 'welfare-optimal scenarios' because of the need to monetise many things which arguably can't be monetised (e.g. the environment), using this term seems to open the paper up to unwanted distractions. I think this could be avoided with a simple re-wording; calling IAMs a 'tool which are used to derive welfare-optimal scenarios' rather than an 'indispensable tool used to derive welfare-optimal scenarios'. This change would avoid opening up an economic debate (in the reader's mind) which is completely outside the scope of this paper.*

For V2 we will comply with the reviewer's suggestion.

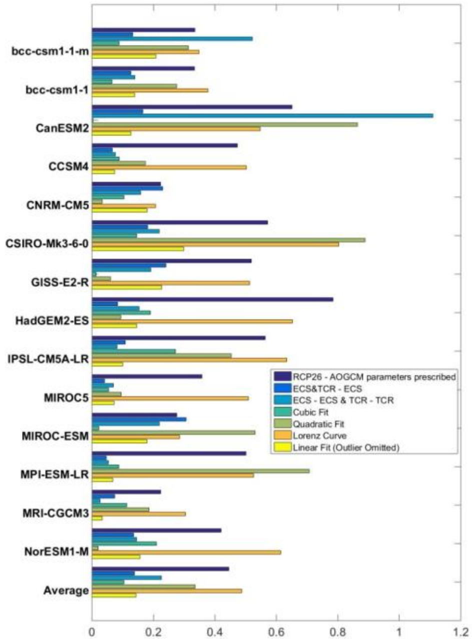
4. *page 8, line 22: 'personal conviction'. I don't think personal convictions have any place in scientific papers. Either the evidence is there to support using log-normal distributions or it's not. I also don't understand what the sentence beginning with 'This conviction rests' means. Does it mean 'Schneider von Deimlinig et al. show that constraining ECS by paleo data results in thin-tailed distributions'? If yes, then there is no need for a 'personal conviction', circling back to my first point.*

We will eliminate this paleo discussion from V2. For V2 it is sufficient to say that lognormal distributions for ECS are used in the literature and that under an affine transformation (as suggested in ms-Fig.6) they would be mapped onto lognormal distributions the quantiles of which are still compatible with what is reported in IPCC AR5 WGI on ECS.
Yes, the point was that Schneider von Deimling et al., 2006, would allow for a thin tail on ECS. The term 'conviction' might not have been the accurate term for what we tried to express. Our point was that a single paper does not yet trigger a paradigm shift. Major fractions, if not the majority of climate researchers – in contrast to us – doubt that ECS can be constrained by paleo data. This fundamental question is still open and hence also subject to personal believe and intuition that might guide the choice of the always in-part

subjective distributions of ECS (as any of them rests on Bayesian learning that needs a subjective prior as an input).

5. *I really appreciated the discussion of low pass filtering and think this was well done.*

Thank you.



Reply-Figure 2: The quality of fit does not change by avoiding outliers and using a linear fit instead.

6. *In section 4 (page 7, lines 32-34), the authors state that 'regressing both inferred effective ECS and TCR solely against AOGCMs' ECS obviously is the overall better approximation'. Whilst this is borne out by taking a pure average of all the results, there are clearly two strong outliers which are having a major effect on the performance of the ECS-ECS & TCR-TCR mapping. I wonder what is causing such large outliers (they seem hugely anomalous) and if removing them would be justified. If they are removed, how much does this change the conclusions.*

We will have an extended discussion on that issue in V2. First results on the PH99 parameter-based fit method (see ms-Fig. 5) show that eliminating outliers and then moving to a linear fit would not significantly change the quality of emulation (see Reply-Fig. 2).

4 Technical Corrections

The innumerous suggestions made by the referee appear very meaningful and appropriate to us and will be implemented in V2.

In summary we are optimistic that a new version V2 could be acceptable for the reviewer if (i) it clarified its scope: being about the link from total forcing to temperature, (ii) if it delivered a physical interpretation of the observed effects, (iii) reflected all of the technical corrections. We are grateful for the referee’s comments as they will have triggered an – in our view – considerably upgraded version of our ms.

References Review 2

Neubersch, D., Held, H., & Otto, A. (2014). Operationalizing climate targets under learning: An application of cost-risk analysis. Climatic change, 126(3-4), 305-318.

Nordhaus, W., Sztorc, P., DICE 2013R: Introduction and User’s Manual, dicemodel.net , 2013.

Petschel-Held, G., Schellnhuber, H.-J., Bruckner, T., Toth, F. L., and Hasselmann, K.: The tolerable windows approach: Theoretical and methodological foundations, Climatic Change, 41, 303–331, 1999.

Schneider von Deimling, T., Held, H., Ganopolski, A., and Rahmstorf, S.: Climate sensitivity estimated from ensemble simulations of glacial climate, Clim Dyn, 27, 149–163, doi:10.1007/s00382-006-0126-8, 2006.