Dear Editor,

We are grateful for having been granted the chance to improve our ms based on three referees' comments. The main structural changes are as follows:

- We clarify the scope of our article much more explicitly in the beginning of the introduction to avoid further misunderstandings like those that apparently misled reviewer 1 in our version 1.
- We deliver a physical mechanism to explain the structural difference PH99/AOGCMs prior re-calibration. We devote a whole new chapter to this question.
- We deliver the whole matrix of possible calibration vs validation combinations, spanned by the four RCP options to verify our approach.
- We reformulated our ms more like a warning before a naive usage of PH99 rather than an advertisement.
- We involved a native speaker to proofread our ms.

As the ms is already unusually long, we abstain from involving Ruelle's theory.
Rather we stay with a very basic approach (1 box vs 2 box intercomparison) that we expect would immediately appeal to the intuition of most readers.
We would be delighted to see that this version of the ms, upgraded along the suggestions of the referees, is found suitable for publication in ESD.

Sincerely,
Mohammad M. Khabbazan

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 #1

Received and published: 6 June 2017

Original comments by the referee are highlighted in italic font.

The manuscript by Khabbazan and Held assesses the performance of a very simplified climate module currently in use in some IAMs. In particular, the study is motivated by the need to adjust the existing tools to the capability of this module in the light of assessments of below 2°C scenarios. To that end, it is fitted to different CMIP5 RCP2.6 AOGCMs.

This is the very point of the first version of our ms indeed.

The manuscript contains no fundamental flaws although a re-read is in order and the literature list should be checked.

Done.

One key paper (Foster et al. 2013) is for example missing from the literature list. I presume it's Forster, P. M., T. Andrews, P. Good, J. M. Gregory, L. S. Jackson, and M. Zelinka, 2013: Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. J. Geophys. Res. Atmos., 118, 1139– 1150

The referee is right. This is the very reference that represents the data basis for our work. We sincerely apologize for this flaw. We will make sure that flaws like this one cannot occur in V2.

More fundamentally, however, the scientific advancement presented of this study in my assessment is rather poor and it neglects important recent literature in this context (in fact, the literature list is rather short and at least 3 years old).

In our new introduction we now clearly state our research question. Furthermore we added a semi-analytic section which explains the physical mechanism of the observed PH99-AOGCM discrepancy. We recommend recalibrating PH99 before using it and thereby we also allow for a re-interpretation of existing literature. Finally we demonstrate how PH99 can be re-calibrated such that it would match AOGCM output for *any* RCP. Thereby we have extended the scope of our article. Regarding omitted literature, to our impression the referee mainly refers to the effects outlined in the following paragraph. As outlined below, our ms is *not* about the issue raised by the referee in that paragraph.

On the methodological approach: What's the justification of using the PH99 model (apart from it 'being there')? The authors argue that it's computational efficiency, ...

PH99 is can be interpreted as an energy balance model (for details of the justification see Petschel-Held et al., 1999, and references therein; the authors' list contains Klaus Hasselmann! Furthermore, the model was validated in Kriegler and Bruckner, 2004 – however the validation was done in a different way and also did not have the forcing reconstructions by Forster et al., 2013, at hand). Even today, computational efficiency is key, e.g. when it comes to decision-making under endogenous learning (see e.g. Webster et al., 2012).

...but how exactly are they convinced that their treatment of non-CO2 GHGs is appropriate. For strong mitigation pathways, these 'minor' differences may become very important, last but not least to determine net-zero global GHG forcing etc. I'd think they would need validate their fit using other strong mitigation scenarios with different non-GHG trajectories (if no others are available then from the GeoMIP experiment) rather than RCP4.5. In particular, it appears that non-CO2 gases obscure our assessments of ECS (see e.g. Myre et al. 2016)

As it stands, I'm not convinced that the simplified model is capable of including non- GHG forcing in a sufficient fashion for the question at hand (i.e. staying below 2°C or 1.5°C).

Apparently here we provoked a key misunderstanding what our article is about. It is *not* about how to generate a global total forcing out of regional forcings. Our ms is about how to get from a global total forcing to global mean temperature and whether PH99 does a good enough job in that, i.e. could substitute for AOGCMs or more complex integrated assessment climate modules in that regard. We have made this point much clearer in the introduction by phrasing the concrete research question. We nevertheless include all RCPs for validation including RCP6.0 which contains a strong ozone component. Also here, PH99 proves successful.

Finally, any author within integrated assessment of the coupled climate-economy-problem has to deal with how to construct a meaningful global forcing out of regionally disaggregated AOGCM forcings, as AOGCMs cannot directly be utilized in economic optimizations. This is a difficult discussion indeed, but *any* climate module within integrated assessment would face this issue, not only PH99. This discussion, however, is not the subject of our ms.

On the application: I didn't fully the motivation for step 1. What was the reasoning for the authors to assume that their PH99 model would work with AOGCM diagnostics from Forsters et al. directly? Obviously, the derived feedback response time parameter 1/alpha of 34.5 years in the multi-model mean is quite unphysical. It seems that the PH99 model is not equipped to be used in that context.

This is the very point. We want to highlight that *current practice* in integrated assessment of directly prescribing ECS and other parameters from AOGCMs leads to biases. As an AOGCM also contains time scales faster than 35 years, it is not immediately clear that an average time scale of smaller and larger time scales would be meaningless. We more clearly cite the relevant literature on FUND, MIND, PAGE.

In a next step, the authors find that with two free parameters they are capable of achieving better fits. That's not particularly surprising, but a physical interpretation of these differences is virtually absent? ECS is substantial decreased by almost 1°C. Can this be understood? The authors continue with fitting derived and fitted ECS and TCR, but I would rather like to see a physical interpretation or an extension of the PH99 model that

would correct for this. The authors should also consider their approach in the light of alternative simplified approaches out there i.e. based on a response function approach as in

Ragone et al. (2016).

We did not find it easy to dig out what literature the referee is referring to as no exact reference is supplied. We can only guess that the referee refers to Ruelle's response theory. However we find that Ruelle's response theory is much more general (in allowing for the forcing entering the ODE system augmented by some function of the system's state) than our system at hand. Hence we prefer for the context of this article to stick to rather elementary explanations in terms of 1 mode trying to substitute a superposition of 2 modes – this is what the problem at hand simply is about (the length of our ms is already way above 30 pages). In fact, by having found that the discrepancy we report can be explained by the move from a 2-box to a 1-box model, we have generated an anchor for explaining the reported ECS effect. A key problem of the 1-box model is that it replaces the slow-component response by the averaged faster one and hence would lead to an overreaction for peak-and-decline forcing scenarios such as RCP2.6. In that sense we expect RCP2.6 as particularly difficult to emulate. Kriegler and Bruckner, 2004, could not do this validation, as mitigation scenario forcing reconstructions were not available at that time.

The authors then want to apply an effective correction for their dubious model in the first place. Their results here appear to be prone to outliers. Compare e.g. the low ECS outlier in Fig. 5. When removing it, I guess even a linear fit would deliver decent results and I'm not sure I can deduce any robust trends from these graphs…

V2 hopefully avoids the impression that we plainly 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 spite of the discovered problems. In addition to computational efficiency, for some applications also analytic tractability or conceptual simplicity might be arguments for using PH99.

Furthermore we have done the sensitivity study the referee suggested by repeating some analyses w/o outliers. W/o outliers linear fits would make PH99 parameters predictable by ECS indeed (see ms-Fig. 6). The scenario fit results deliver similar quality as before (see ms-Fig. 5). However we would like to stress that our ms recommends direct correction of ECS along ms-Fig.6 where outliers play a less prominent role. V2 have a short discussion of the outlier issue.

On the application of this. It seems that the model that is used in these IAMs has many flaws.

The question then becomes why it is used at all? And not abandoned for a carbon budget approach that would be even more computational effective and can be determined with more complex models also for these low emissions scenarios (i.e. Rogelj 2016). That becomes in particular relevant since the mitigation challenge ahead is to define pathways that hold warming 'well below 2°C'. "Below 2°C" was interpreted as a 66% chance of non-exceedance (IPCC 2014). What's the added value of using a PH99 model in this context? Would they select an ECS at the 66% quantile and then use this as a basis for the IAM derivations? And if so, why not use carbon budgets directly?

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. In any case, 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).

In summary we are optimistic that a new version V2 could be acceptable for the reviewer as (i) it clarifies its scope: being about the link from total forcing to temperature and (ii) it delivers a physical interpretation of the observed effects. 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 1

Forster, P. M., Andrews, T., Good, P., Gregory, J. M., Jackson, L. S., & Zelinka, M. Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. Journal of Geophysical Research: Atmospheres, 118(3), 1139-1150, 2013.

Kriegler, E. and Bruckner, T.: Sensitivity analysis of emissions corridors for the 21st century, Climatic Change, 66, 345–387, 2004.

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

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.

Webster, Mort, Nidhi Santen, and Panos Parpas. An approximate dynamic programming framework for modeling global climate policy under decision-dependent uncertainty. Computational Management Science 9.3: 339-362, 2012.

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. PH99 is physical in the sense that it is an energy balance model, that larger forcing results in larger GMT rise and that there is linear delay between forcing and response. However the time-scale and forcing-shape specific calibration necessary for PH99 points to a not so physical model.

Also, in V2 we hopefully avoid the impression that we would advertise utilizing PH99 without reservation. 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 hopefully is improved in that regard. A native speaker was involved in correcting the ms.

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 – however with limited accuracy for RCP 8.5.

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-CO_2 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 show in V2 that PH99 emulates *all* RCPs by the identical recalibration. This also contains RCP6.0 with some stronger ozone component. We also show from a semi-analytical treatment that PH99 is a good emulator of a whole 2-dimensional manifold of mitigation forcing scenarios. 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 is much clearer about this scope. The effects by moving from a 2-box to a 1-box model plays 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 is stopped in V2. Instead the underlying physical mechanisms will be explained from which also the observed discrepancy will be derived. No re-calibration is necessary for generic RCP forcing scenarios.

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. In particular we deliver, as we believe, a clear-cut interpretation of the observed discrepancy PH99-AOGCM.

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 now explains why a lower ECS in conjunction with a lower time-scale is adequate to emulate AOGCMs for the time-horizon relevant for climate policy evaluation.

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 CO_2 and non-CO_2 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. 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.*

V2 complies with the reviewer's suggestion – wording changed accordingly.

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 have eliminated 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.

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 have a short discussion on that issue in V2. Results on the PH99 parameter-based fit method (see ms-Fig. 5 and ms-Fig. 6) show that eliminating outliers and then moving to a linear fit (suggested by referee #1) will 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 have been implemented in V2. Please notice that we have also asked a native speaker to proofread our ms.

In summary we are optimistic that a new version V2 could be acceptable for the reviewer as (i) it clarified its scope: being about the link from total forcing to temperature, (ii) 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.

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 #3

Received and published: 27 November 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.

General Comments

This manuscript investigates the performance of a one-box energy balance model (PH99) as an AOGCM emulator for strong mitigation scenarios. The authors find that this simple climate model (SCM) consistently over-predicts future temperatures when the ECS and TCR are transferred directly from AOGCMs. Fitting the PH99 directly to the AOGCM temperature time series eliminates this bias, and reveals that the AOGCMs time series imply a substantially lower ECS and higher TCR than what they had transferred directly. The manuscript briefly discusses the physical interpretation of this discrepancy, and also explore alternative ways of fitting the one-box model that might be more reasonable for extrapolation in parameter space (of the kind performed when these SCMs are used to investigate the optimal dynamic behaviour of a decision maker under uncertainty).

We feel our ms perfectly perceived by the referee.

Before continuing further, I want to briefly highlight two important factors that might reasonably affect how you read this review. First, I have not only read the manuscript, but also the previous reviews and the responses from the authors. My comments primarily address the manuscript itself, but I will also sometimes explicitly agree or disagree with comments that have been made earlier in the process. Second, I have not approached this manuscript as a climate physicist, but rather from the perspective of a researcher who uses the integrated assessment models with SCMs like PH99. My comments will therefore differ in spirit from those of earlier reviewers, and I focus more on issues I believe to be more relevant to those who would use this research.

While we appreciate comments from all relevant disciplines, we are happy that among the referees there is a referee who uses integrated assessment models.

Specific comments

My overall assessment is that this manuscript offers an interesting contribution and should be published.

Thank you!

A previous reviewer expressed concern that the scientific contribution may be inadequate, but I feel that this comment does not adequately consider the policy influence that the one-box model wields (or rather, simple integrated assessment models that use PH99 in one form or another). For instance, two of the three climate-economy models used by the US federal government to calculate the social cost of carbon incorporate one-box energy balance models (notwithstanding recent political developments). The policy analysis in the Stern Review, which was commissioned and used by the UK government to formulate climate policy, was also based on a coupled climate-economy model that incorporated a one-box energy balance model. By their simplicity, these SCMs are also have come to serve as tools for translating new climate science for communities that use climate information but generally lack extensive physics training. Even a relatively small improvement in our understanding and handling of these models would provide a significant contribution.

We feel our ms perfectly perceived by the referee. We added a reference to the Stern review.

I do have some concerns about the manuscript, though. First, I think there are parts of the manuscript that will be difficult to decypher for many of the researchers that actually work with SCMs in the context of simple climate-economy models. Second, I think that authors have tended to focus excess attention on concerns related to interpolation and extrapolation of parameter values, at the expense of a fuller and clearer discussion of the physical interpretation of their primary findings. I discuss each point in turn, and offer a few minor comments at the end. Let me state clearly, though, that I expect these concerns can be fully redressed, so I wouldn't consider them reasons for rejecting the manuscript.

We are thankful to the reviewer for the time and efforts. We are open to any comments that can enhance the manuscript.

1. *The heart of this manuscript, as I see it, is the direct transfer of AOGCM characteristics (section 2.1). The central issue is whether or not it is appropriate to use this physical method for deducing the parameter values in one-box model. The subsequent question about whether alternative methods for fitting the parameter values perform better, is also tied to this baseline method. So section 2.1 is really the foundation for all of the analysis in this paper. Yet two cruicial pieces seem to be missing from it.*

   First, at this point in the manuscript the authors should be offering a childishly clear explanation of how (and which) AOGCM outputs can be used to deduce the values of α and μ, which can then be plugged into equations (2) and (3) to retreive the implicit ECS and TCR, respectively. But I must admit to having some difficulty following their derivations (e.g. not understanding how h is determined in equation (7) where both h and μ appear to be unknowns, and not seeing any expression for α in terms of AOGCM output). A climate physicist will perhaps be so familiar with this material as to be able to perform these calculations with little prompting from the authors, but as a presumptive member of the intended audience, I would appreciate it if the authors exercised greater pedagogical care in this section.

We added a clean recipe how to deduce PH99's parameters from ECS & TCR.

Second, and just as important, is that the authors have not offered any information to suggest that this is how modellers are currently choosing values for the ECS and TCR. In my experience, many users will not themselves try to deduce these parameters from AOGCM outputs, but rather plug in values of ECS and TCR reported in IPCC chapters or specific academic papers without fully understanding how these values are inferred from AOGCM simulations (and sometimes adding a bit of 'calibration' to make sure the results don't look too dissimilar from MAGICC, say). If those reported ECS and TCR values are derived in this way generally, the authors should state this clearly and cite examples. If not, they should consider whether it would be more appropriate to use a different baseline.

We thank the referee for pointing us to another misunderstanding we might have provoked. In fact, the referee's perception of standard practice is actually what we wanted to refer to in our manuscript, as we believe that users of PH99 have not themselves tried to deduce these parameters from AOGCM outputs. However, given that insight, then comparing AOGCM and PH99 output while both models would be characterized by identical ECS and TCR and be driven by identical radiative forcing is then the most direct way we can currently imagine to demonstrate a bias encoded in PH99. We added a paragraph at the end of Section 1 accordingly.

As a suggestion, I think it would be worthwhile to run the model using the actual parameter values assumed in FUND and MIND as a baseline (and maybe PAGE, which the authors seem to have ignored, even though it incorporates a one-box model). If I were to speculate, I would guess that using the default parameter values from these models will give even more discrepant predictions, so in addition to being more relevant to current practice, it might illustrate your point even more clearly.

Thanks for pointing us to PAGE. We added it to V2. Furthermore we explicitly commented on the standard value of ECS=3°C as utilized in FUND and PAGE. The PH99-equivalent should be about 4°C.

2. *SCMs are used for two distinct purposes: (1) as devices for summarizing and*

communicating climate science to other modelling communities, and (2) as computationally efficient AOGCM emulators. The analysis performed in this manuscript has important implications for both uses, but the authors are failing to distinguish clearly between them. This creates unnecessary confusion (seen especially clearly in the exchanges with previous reviewers), and has in my opinion led to an unbalanced treatment.

The authors appear to recognise the role of PH99 as a communication device when they, in their Discussion (section 5), briefly mention the idea that the 'transient climate sensitivity' might be lower than the ECS, as a potential physical explanation for the lowering of the ECS when the parameters are calculated by fitting the one-box model to AOGCM temperatures instead of derived from AOGCM forcings. But what is the chief physical mechanism behind this? And does this mean that PH99 users should interpret $\frac{\mu}{\alpha}\ln(2)$ *as the 'transient climate sensitivity' rather than the ECS? Can we do this without undermining the physical basis for the one-box model? And what about the higher TCR value that you get when fitting the one-box parameters instead of transfering them directly? What is the physical interpretation of this second important change? You also acknowledge that measurement error in AOGCM outputs could lead to biased values of the ECS and TCR in the one-box model, but can you do anything to show that the biases would actually go in the direction of inflating the ECS and deflating the TCR? Perhaps you could just add a random sample of Gaussian deviations to your input data and feed them through the non-linear PH99 mapping to see what the resulting distribution of ECS and TCR would look like?*

We invented a whole new Section to deliver on the 'chief physical mechanism'. In that vein we can explain why ECS must be downsized in PH99 and accordingly also the response time scale.

I realise I have given you a lot of questions to answer, but I really do feel that this part of the discussion has been unduly neglected, and the paper would benefit greatly from extending it. It seems a very interesting fact that, for a given TCR, the ECS value transferred directly from an AOGCM is systematically higher than the value that would yield the best fit to that same AOGCM (and vice versa for TCR). Anything the authors are able to do to help the reader understand the causes of the differences between fitted and transferred parameter values, and *how this might affect the physical interpretation of the one-box model paramters, would be very welcome.*

We are very grateful for the referee sharing our impression that inferred ECS and TCR values are biased as against their fitted counterparts does represent ‘a very interesting fact’. The latter and its practical consequences is the key motivation of having this ms! We feel that V2 does deliver here, indeed.

I would, compensatingly, recommend shortening the discussion of the second use of PH99, as an AOGCM emulator, which currently takes up the majority of sections 4 and 5. I think it is interesting to consider the advantages and disadvantages of alternative methods for choosing ECS and TCR for emulation purposes, but I often felt lost in this discussion and think it can be done more concisely. The fitting method, perhaps unsurprisingly, does a pretty good job of fitting the AOGCM temperature time series. But the key drawback of the fitting-method is that it’s inappropriate for obtaining probability distributions for the ECS and TCR that can be used to simulate PH99 under uncertainty. These kinds of simulations are now standard practice for economic assessments of climate policy based on coupled climate-economy models, so this is indeed an important issue to wrestle with.

The quadratic and cubic fitting in Figure 5 seems useful mostly as a cautionary example of ‘what not to do.’ The authors already explain that it’s likely to lead to unphysical parameter values, and as a previous reviewer pointed out, the curvature seems largely a consequence of a single AOGCM run with a low ECS. Overall, the authors can probably devote less space on this particular exercise and be even clearer that it is ill-advised. Instead, they should focus on the more physical interpolation/extrapolation methods considered in section 4, and try to offer users more concrete advice about when they might prefer the Lorenz curve method, or the ECS-to-ECS and TCR-to-TCR fit, or the ECS/TCR-to-ECS fit, or when all three are likely to perform poorly.

We are thankful to the referee for this informative comment. We have included the suggestion by referees to omit an obvious outlier and repeat the regression (see ms-Fig. 5 and ms-Fig. 6). W/o outliers linear fits would make PH99 parameters predictable by ECS indeed (see ms-Fig. 6). The scenario fit results deliver similar quality as before (see ms-Fig. 5). However we would like to stress that our ms recommends direct correction of ECS along ms-Fig.6 where outliers play a less prominent role and keep all the rest of the analysis in that section as it is for a better comparison. More specifically regarding the Lorenz curve, as it is vastly used by by MIND users, we see the benefits of keeping it in V2 for interested readers.

I think a slight reorganization of sections 4 and 5 would probably be the most effective way of accomplishing all of this. The new section 4 would take the first two paragraphs of the current section 5 as its starting point, but elaborate along the lines I have discussed above in order to offer a discussion of the physical interpretation of fitted ECS and TCR relative to the directly transferred ones. The new section 5 would merge the current section 4 with the remainder of the current section 5, in order to offer a discussion of the appropriate and inappropriate ways to interpolate and extrapolate ECS and TCR values in PH99, in light of their physical reinterpretation in the new section 4. This separation would also make it much clearer how the choice of parameter values for PH99 depends on whether one is using it as a communications device or as an emulator.

We very carefully considered this advice. In the end also the practitioner must know what could be done in case PH99 should be utilized further. V2 now maps application onto an appropriate fitting method much clearer. Please notice that we have added a section (section 5) as a response to the reviewers' call for clearer physical interpretation.

Technical corrections

1. *p. 1, line 9 (and throughout): The manuscript refers to FUND and MIND as two coupled climate-economy model that employ a one-box model. PAGE does as well (see discussion in Calel & Stainforth, 2017, BAMS, already cited).*

   We included PAGE in V2.

2. *p. 2, line 30: Typically, these models are used to study optimal climate policy, so it would be good if you could cite a few studies where these models are used specifically to study 2 degree stabilisation scenarios.*

For MIND and PAGE, such papers are cited. FUND generically does not result in 2° paths, however, the bias found by us extends to any concave forcing, hence any paper cited on FUND would deliver here.

3. *p. 5, line 12: Typo. "APGCM" should be "AOGCM."*

Corrected for.

4. *p. 5, lines 17-22: While RCP4.5 is certaintly out-of-sample, it's less obvious to me that it serves the purpose of validating the method for 2 degree stabilization scenarios. As a validation exercise, wouldn't it be preferable to fit α and μ using RCP4.5 and then drive the one-box model using the RCP2.6 forcings?*

We had thought that fitting a climate model to a climate state #1 (defined by RCP2.6) and then validating by a climate state #2 (defined by RCP4.5) even further away from the preindustrial state would represent a tougher test than the reversed order of calibration and validation. But we are more than happy to also test the order as suggested by the referee as we expect an even better fit. V2 now includes calibration of PH99 to any RCP and validating it as against any other RCP.

5. *p. 8, line 11: Typo. "againsta" should be "against a."*

Corrected. Please notice that we have also asked a native speaker to proofread our ms.

6. *p. 8, line 15: Typo. "radiative active" should be "radiative activity" or "radiative forcing."*

Corrected.

7. *p. 8, line 19: I think it's inaccurate to say that "studies based on PH99 implicitly worked with ECS values that were larger than announced." I think you've made the point that they might be using a higher ECS than would be appropriate for emulating AOGCMs, but this is quite different. They declare their ECS values, and the question raised in this manuscript is whether they shouldn't be using ECS but rather some 'effective ECS' or 'transient climate sensitivity' instead. Please rephrase this.*

§ erased.

8. *p. 8, lines 20-28: The discussion of log-Normal distributions seems to come out of nowhere. I*

think the reorganization I have suggested above may resolve this, but please make an effort to link this more strongly to the rest of the discussion.

Done.

9. *p. 8, line 29: Typo. "boefore" should be "before."*

Corrected.

10. *p. 10, line 7: Typo. "generally" should be "it generally."*

Corrected.

On the Future Role of the ~~most~~Most Parsimonious Climate Module in Integrated Assessment

Mohammad M. Khabbazan and Hermann Held

Research Unit Sustainability and Global Change, Center for Earth System Research and Sustainability, Universität Hamburg, Grindelberg 5, 20144 Hamburg, Germany

Correspondence to: Mohammad M. Khabbazan (mohammad.khabbazan@uni-hamburg.de)

Abstract. ~~We~~ In the following, we test the validity of a ~~one~~1-box climate model as an emulator for Atmosphere-Ocean General Circulation Models (AOGCMs~~) when the application is confined to the subset of scenarios approximately in line with the 2° target.~~). The ~~one~~1-box climate model is currently ~~in use~~employed in the integrated assessment models FUND, MIND and ~~MIND~~PAGE. For our assessment, we ~~crucially~~primarily rely on 14 recent CMIP5 AOGCM diagnostics of the total radiative forcing for various representative concentration pathways. Our findings are two-fold. Firstly, when directly prescribing AOGCMs' respective equilibrium climate sensitivities (ECSs) and transient climate responses (TCRs) to the ~~one~~1-box model, global mean temperature (GMT) projections are generically too ~~large~~high by 0.5 K at peak temperature~~.~~, although the model was validated in the past. Accordingly, corresponding integrated assessment studies might tend to overestimate mitigation ~~need~~needs and ~~cost.~~costs. We semi-analytically explain this discrepancy as resulting from the information loss produced by replacing a 2-box with a 1-box model. Secondly, the ~~one~~1-box model ~~becomes an excellent~~offers a good emulator of ~~those~~these AOGCMs ~~once~~, provided their ECS and TCR values are universally mapped onto effective ~~one~~1-box ~~intrinsic~~ counterparts~~.~~ and a certain time horizon is not exceeded. We ~~suggest utilizing~~contend that this ~~one~~1-box model could be used in future integrated assessment ~~also in the future~~within certain limits, in particular when computationally demanding decision-making under climate response uncertainty ~~continues~~needs to be modelled. However, then the ~~roles of~~ ECS and TCR values must be transformed beforehand, depending on the time horizon of application and on the question whether an unmitigated increase in radiative forcing or a more concave, kink-linear forcing style, e.g. for mitigation scenarios, is expected. Results that are based on the model and have already been published are still just as informative as intended by their respective authors; however, they should be re-interpreted~~. For the MIND model as used over the past 5 years, even the thereby determined effective~~ as being influenced by a larger ECS ~~values comply with the ranges explicated by IPCC AR5, however now at the high-end.~~than claimed.

Keywords: climate sensitivity, emulator, integrated assessment, mitigation scenarios, reduced climate models

1 Introduction

Climate-economy integrated assessment models (IAMs) represent an indispensable tool when derivingare used to derive welfare-optimal climate policy scenarios (Kunreuther et al., 2014) or constrained welfare-optimal scenarios that would comply with a prescribed policy target (Clarke et al., 2014). Most of them employ relatively simple climate modules emulating highlythe most sophisticated climate models, Atmosphere-Ocean General Circulation Models (AOGCMs). These climate modules (called henceforthhereafter: ‘simple climate models’ (SCMs)) offerfoster computational efficiency and hence allow for projectingresearchers to project a broader set of scenarios in a reducedorders of magnitude less time. For IAMs based on a decision-analytic framework involving intertemporal welfare optimization, SCMs are in fact indispensable, as thosethese IAMs’ numerical solvers would callneed to access the climate module anywhere from ten thousandsthousand to one hundred thousand times before numerical convergence iswere flagged.

The Currently the need to qualify the degree of accuracy ofwith which SCMs in mimickingmimic AOGCMs or properly representingrepresent ensembles of AOGCMs is increasingly being recognized (Calel & Stainforth, 20162017; van Vuuren et al., 2011)2011a), as this aspect might have immediate monetary consequences encoded in connection with derived policy scenarios (Calel & Stainforth, 20162017). Van Vuuren et al. (20112011a) found that IAMs tendedtend to underestimate the effects of greenhouse gas emissions.

Due to the centennial-scale quasi-linear properties of AOGCMs’ global mean temperature (GMT) dynamics, SCMs have proven capable of emulating theAOGCMs’ behavior of AOGCMs regarding GMT change, deviations being a function of spread of forcing, SCM complexity (Meinshausen et al., 2011a) and quality of SCM calibration. The climate component of MAGICC (Meinshausen et al., 20112011a) represents the most complex SCM currently in use. In some sense one could even call MAGICC an Earth System Model of Intermediate Complexity. Its abilityIt has demonstrated its capacity to emulate all AOGCMs’ GMT even more precisely than the standard deviation of interannual GMT variability has been demonstrated (Meinshausen et al., 20112011a), with a fixed set of parameters, utilized for the whole range of RCPs (representative concentration pathways, van Vuuren et al., 2011a2011b). This represents the current gold standard of AOGCM emulation throughusing SCMs.

HereIn the following, we address the most extreme opposite end of the scale of complexity within the model category of SCMs: the one1-box model as introduced by Petschel-Held et al. (1999) (calledhereafter: ‘PH99’ from now on). Its). We pose the question: ‘To what extent is PH99’s temperature equation able to correctly map globally averaged radiative forcing anomalies onto GMT anomalies?’ The current role is of this model as describedassessed in the following. Byliterature is as follows: by fitting PH99 to GMT time series one, it can utilize itbe used as a diagnostic instrument, as described in Andrews & Allen (2008).) have done. However its main application is functioning as an emulator of AOGCMs. In conjunction with the most parsimonious carbon cycle model (also described in Petschel-Held et al. (1999))) as well), PH99 has been utilized for derivingused to derive ‘admissible’ greenhouse gas emission scenarios in view of prescribed GMT targets (Bruckner et al., 19992003; Kriegler & Bruckner, 2004). Furthermore, the following climate-economic IAMs are currently utilizeutilizing

PH99: FUND (Anthoff & Tol, 2014~~) and~~), MIND (Edenhofer et al., 2005~~).~~) and PAGE (Hope, 2006) – the last of which was used in the 'Stern Review' to the UK government (Stern, 2007). While MIND has since been succeeded by the IAM REMIND (Luderer et al., 2011) when it comes to spatial resolution or representing the energy sector by dozens of technologies, it currently serves as a state-of-the-art IAM for decision-making under uncertainty (Held et al., 2009; Lorenz et al., 2012; Neubersch et al., 2014; Roth et al., 2015) or joint mitigation-solar radiation management analyses (Roshan et al., ~~2016;~~acc.; Stankoweit et al., 2015).

Kriegler and Bruckner (2004) validated PH99 in conjunction with a simple carbon cycle model. When diagnosing the effect of the IS92a emissions scenario (Kattenberg et al., 1996) on GMT, they demonstrated deviations of ~~below~~less than 0.2 K for the 21st century (see their Fig.5).

~~None the less, we put forward~~We believe that further validation is both necessary and possible ~~on~~at a higher level of consistency. Firstly, the respective GMT time series as checked in Kriegler and Bruckner (2004) is convexly increasing. However in the context of scenario generation ~~compatible~~in keeping with the 2° target (UNFCCC, 2016~~)~~)), validation along GMT stabilization or even peaking scenarios is crucial, displaying a qualitatively different shape from IS92a. Secondly, in Kattenberg et al. (1996) the forcing was reconstructed by the ~~auxiliary~~additional assumption ~~of~~that non-CO_2 greenhouse gas forcing approximately ~~balancing~~balances aerosol cooling. Here we ~~utilize~~employ recently diagnosed forcings for 14 ~~CMIP('Coupled Model Intercomparison Project')5~~CMIP5-AOGCMs by Forster et al. (2013). ~~Finally~~Third and lastly, we find current ~~practice,~~practices – directly prescribing equilibrium climate sensitivity (ECS) ~~and~~ (see Hope (2006) or Anthoff & Tol (2014); prescribing the value of 3°, which was generally considered to be the 'best estimate' for years; all the MIND-based work on decision-making under ECS uncertainty (see citations above)); and using a second, time-scale~~-~~-relevant property ~~for calibrating PH99, inadequate~~to calibrate PH99 (see e.g. Anthoff & Tol (2014)) – 'inadequate' in the context of 2° stabilization scenarios. ~~Instead~~In this regard, 'inadequate' implies that PH99 cannot emulate an AOGCM with similar ECS and TCR to a sufficient degree of accuracy. Needless to say, we are not claiming that the previously published IAM-based work mentioned above is 'worthless'; rather, we argue that the parameters and probability density distributions need to be re-interpreted, essentially because larger (but still meaningful) ECS values have inadvertently been utilized. Hence we propose ~~to calibrate~~calibrating PH99 ~~in~~by mapping ~~those~~these climate system properties ~~on~~to respective effective ~~scenario-class-specific values~~ values, which are suitable for a centennial time horizon (but likely not beyond it), before using them in PH99.

~~Hereby~~In doing so, we comply with a ~~quest recently formulated by~~ need that Calel & Stainforth (~~2016):~~2017) recently identified: to ~~establish~~achieve the application-specific re-calibration of PH99 as a valid future approach to emulation. ~~Thereby~~In this way, PH99 could complement ~~utilizing ever more~~the use of increasingly complex climate modules, ranging from ~~the two~~DICE's 2-box model ~~of DICE~~ (Nordhaus, 2013) to the ~~rather~~ complex upwelling-diffusion climate module ~~of~~used in MAGICC (Meinshausen et al. (~~2013~~2011a)). The ~~benefit of PH99 would be~~potential benefits of doing so are two-fold: firstly, the most parsimonious SCM, PH99, ensures maximum transparency. Secondly, in ~~view~~the context of numerically solving decision-making under climate response uncertainty (Kunreuther et al., 2014), having to simultaneously deal with dozens~~ to ~~, hundreds or even thousands of ~~alternative~~alternate climate ‘states of the world’ ~~in parallel~~(the economist’s term for the uncertain system property) poses a significant challenge for numerical solvers ~~with regard to the available~~and memory. In ~~that~~this regard, PH99 appears particularly attractive. At the same time, our article represents a warning: if PH99 is to be used in the future, it should be done in a re-scaled manner, adjusted to the time horizon under investigation.

~~The remainder of this~~ This article is organized as follows. Section 2 introduces the data-based part of our~~ method of~~ analysis~~,~~ ~~comprising of~~. We call for a 3-step procedure~~.~~, including: (i) ~~A traditional, if~~ a conventional, though not naïve~~,~~ calibration of PH99 ~~by~~ with regard to climate sensitivity and transient climate response (i.e. the GMT change in response to a 1%/yr. increase ~~of~~in the CO_2 concentration until doubling ~~as against~~compared to the pre-industrial value~~),~~); (ii) ~~a~~an AOGCM-specific calibration~~,~~; and (iii) the validation of the former. In ~~Sect.~~Section 3 we first demonstrate that (i) would lead to emulation errors of up to 0.5 K for scenarios approximately compatible with the 2° target. We then show that this emulation error can generically be reduced to 0.1 K when choosing AOGCM-specific calibrations of PH99. This calibration is subsequently validated by ~~an~~ independent ~~scenario. In~~ scenarios. Note that, in Sect. ~~4 we show how~~ 3, we focus on only RCP2.6 scenario for calibration and use RCP4.5 and RCP8.5 for validation and leave further analyses, which show that PH99 can be generally calibrated to and validated by a variety of scenarios, for the sake of brevity, to Appendix 2. In Section 4 we present a scheme of how to calibrate PH99 for a given ECS, thereby avoiding AOGCM-specific calibrations. ~~While this~~This results in a larger emulation error than achieved in ~~Sect. 3 (up to 0.2 K instead of up to 0.1 K), it~~Section 3, but one that would ~~still~~nevertheless suffice for most applications. In ~~Sect. 5~~Section 5 we explain the observed discrepancy between PH99 and AOGCMs as reported for step one of Section 2 by pursuing a semi-analytical, physically-based approach. In Section 6 we discuss the implications of our ~~numerical~~ findings for the integrated assessment community~~. In Sect. 6 we conclude and outline~~, while Section 7 presents our conclusions and outlines further research needs.

Before we proceed, a brief note on the role of AOGCM data in our article might be in order. We compare PH99 to AOGCM data because we utilize AOGCMs here as the entities closest to ‘reality’ available on the ‘model market.’ We do not, however, claim that IAM modelers were using them or should be using them. AOGCM data is used to demonstrate how ECS and TCR data can skew the calibration of PH99, and how it should be corrected. The same correction should in principle be used for ECS data inferred from *any* source, e.g. abstract distributions such as those presented in Bindoff et al., 2013. Mirroring PH99 in AOGCM data, however, is currently the most direct way to infer the quality of a (not) re-calibrated PH99.

2 Method

~~The~~ Among others, one of the most extensively used most parsimonious climate ~~emulator~~emulators is the ~~one~~1-box global energy balance model, Eq. (1), introduced by Petschel-Held et al. (1999), ~~that~~which projects the atmospheric GMT anomaly ~~with respect~~compared to its preindustrial level. For a CO_2-only forcing scenario, PH99 reads

$$\dot{T} = \mu \ln(c) - \alpha T \qquad (1)$$

~~Hereby~~Here T denotes the GMT anomaly, c is the CO_2 concentration in units of its pre-industrial level, and α and μ are constant tuning parameters.

From Eq. (1) we can readily read the ECS, the equilibrium temperature anomaly in response to a doubling of the CO_2 concentration ~~as against~~compared to its pre-industrial value:

$$ECS = \frac{\mu}{\alpha} \ln(2) \qquad (2)$$

also in line with Petschel-Held et al. (1999) and Kriegler and Bruckner (2004). In Appendix 1 we briefly derive ~~TCR~~the TCR (GMT from a stylized experiment after the CO2 concentration has been exponentially increased with the rate γ (of 1%/yr.) until the concentration has doubled for this model:

$$TCR = \frac{\mu\gamma}{\alpha^2}\left(-1 + 2^{-\frac{\alpha}{\gamma}} + \frac{\alpha}{\gamma}\ln(2)\right) = \frac{\gamma ECS}{\alpha \ln(2)}\left(-1 + 2^{-\frac{\alpha}{\gamma}} + \frac{\alpha}{\gamma}\ln(2)\right) \qquad (3)$$

~~Here, γ denotes a constant rate of increase of the CO_2 concentration. In view of the current definition of TCR, γ amounts to 1%/yr.~~

~~In the following validation of PH99 is performed in three steps.~~

(The right-hand side of the equation has been obtained by utilizing Eq. (2).) Now we propose a 3-step validation approach to clarify PH99's range of applicability.

2.1 Step ~~one: direct transfer of AOGCM characteristics~~One

We first check whether simply calibrating PH99 from AOGCM-specific ECS and TCR data would deliver good emulations for 2°-target ~~compatible scenarios. A technical difficulty arises from the fact that such scenarios are not available for CO_2-only forcing, but for various forcings. We address this difficulty by a chain of arguments along the equations as given below. As starting point we note that PH99 maps total radiative forcing onto temperature. We utilize scenarios generated from 14 AOGCMs (see Table 1) having participated in~~ -compatible scenarios. After a technical derivation, we summarize this method of mapping AOGCMs' ECS and TCR onto PH99's two parameters. ~~CMIP5, because the total forcings of these scenarios are reconstructed in Forster et al. (2013). ECS and TCR for those 14 models are taken from Forster et al. (2013), Table 1. Then model-specific α and μ are derived from Eq. (2) and Eq. (3).~~

A technical difficulty arises due to the fact that such scenarios are not available for CO2-only forcing, but solely for a plethora of simultaneous forcings that would add up to a total forcing. We utilize scenarios generated by 14 AOGCMs (see) from CMIP5, because the total forcings of these scenarios are reconstructed in Forster et al. (2013). The ECS and TCR for these 14 models are taken from Forster et al. (2013),. Then model-specific α and μ are derived from Eq. (2) and Eq. (3). ~~In order to validate PH99, we need to drive Eq. (1) by the total forcing and compare the so derived emulator's GMT with the respective AOGCM's GMT. When inspecting Eq. (1) it becomes apparent that it still needs to be generalized from a CO_2-only to total forcing. We proceed accordingly in retrieving its physical interpretation. When multiplying~~

In order to validate (or invalidate) PH99, we would need a more general version of it in which temperature in Eq. (1) is driven by the total forcing (rather than the CO2 concentration). This would allow us to compare like with like: the total forcing would be mapped onto temperature, just as it is done in Forster et al. (2013). In this regard, we recall the derivation from an energy balance approach, as summarized in Kriegler and Bruckner (2004). If we multiply it by a constant effective oceanic heat capacity *h*, we obtain an equation that governs the heat flux:

$$h\frac{dT}{dt} = -\alpha h T(t) + \cancel{f}F(t) \qquad (4)$$

Hence the CO_2-carrying summand ~~(*f*(*t*))~~ would become the CO_2~~-~~ forcing and ~~can~~could now be generalized to the total forcing ~~*f*. When dividing~~*F*. If we then divide by the still~~-~~ to~~-~~ be~~-~~ determined factor *h*, we obtain:

$$\frac{dT}{dt} = -\alpha T(t) + \cancel{\frac{f(t)}{h}}\frac{F(t)}{h} \qquad (5)$$

~~In order to derive *h*~~Hence if *h* were known, the forcing and GMT taken from Forster et al. (2013) could be used to test PH99. In order to relate *h* to the original parameters of PH99, α and μ, we re-consider the limiting CO_2-only case of Eq. (4):

$$h\frac{dT}{dt} = -\alpha h T(t) + Q_2\frac{\ln c(t)}{\ln 2} \qquad (6)$$

Q_2 denotes the additional forcing from the doubling of the CO_2 concentration ~~as against~~compared to its pre-industrial value and is listed for ~~all~~any of the above AOGCMs (see Forster et al., 2013, Table 1). When comparing Eq. (1) and Eq. (6), we obtain:

$$\mu = \frac{Q_2}{h \ln 2} \; . \qquad (7)$$

~~Equation (7) is~~ in~~-~~ line with Kriegler & Bruckner (2004). Equation (7) allows ~~for determining~~us to determine h, and in turn ~~for~~to use the time-integrating Eq. (5). ~~Thereby from~~Here, the AOGCMs' total climate forcing for the scenario RCP2.6~~,~~ and the temperature paths for the period 2006~~-~~ to 2100 are projected.

~~To derive the initial levels (2006)~~The derivation displayed so far can be summarized in terms of the ~~temperature~~ following recipe to generate PH99's parameters on the basis of AOGCMs' ECS and TCR:

1. Identify PH99's ECS and TCR with an AOGCM's ECS and TCR.
2. Numerically invert Eq. (3), right-hand side expression, to find α (no analytic expression possible).
3. Invert Eq. (2) to find μ.
4. In case a forcing *F*(*t*) beyond CO2 is to be employed, invert Eq. (7) to find *h*, then utilize Eq. (5).

Finally, to avoid differences occurring over the historical period (pre-2006 for the RCPs), we need to initialise PH99 with each AOGCM's 2006 temperature anomaly with respect to the ~~preindustrial~~pre-industrial value. To do this, for each AOGCM we calculate the mean ~~temperatures~~temperature over the period 1881-1910 and set this as the pre-industrial value. We then calculate the mean temperature over the period 1991-2020~~, respectively, as the preindustrial level of temperature and~~ and use this as an indicator for the 2006 temperature level. The difference between these two values is fixed as the initial temperature anomaly for PH99.

Each temperature trajectory should be compared to the temperature data from the ~~related~~corresponding AOGCM. As for GMT-target-constrained economic optimizations (Clarke et al., 2014; Edenhofer et al., 2005~~)~~), the maximum GMT (rather than the whole time series) is of special importance~~, as an error metric,~~. Hence we use the difference between the respective 2071-2100 GMT time averages of PH99 and the AOGCM ~~are subtracted.~~as an error metric. If the deviations are tolerable, the climate module is validated~~. We~~; if they are intolerable, we must proceed with steps two and three~~ if the deviations are found intolerable~~.

2.2 Step ~~two: fitting PH99 to AOGCM scenarios~~Two

For each AOGCM, α and μ are tuned such that the ~~deviations from the annual temperature data of~~ difference between PH99 and the AOGCM GMT anomaly for the RCP2.6 scenario ~~for~~in the period 2006-2100 ~~are~~is minimized ~~in~~using a least squares approach. For further diagnostics we then determine the new ‘effective’ ECS and TCR from Eq. (2) and Eq. (3). As in step one, the deviations in ~~30-year~~2071-2100 means of GMT between PH99 and the respective ~~APGCM~~AOGCM are determined as an accuracy check.

In order to eliminate the effect of AOGCM drift, prior to a similar fitting exercise for MAGICC, Meinshausen et al. (~~2011~~2011a) subtracted (low-pass filtered) control runs. We avoid ~~accounting~~the need to account for AOGCM drift, as our analysis is based on CMIP5 AOGCMs (Forster et al., 2013~~)~~), for which the problem of model drift has essentially been eliminated (Geoffroy et al., 2013, Fig. 2). Furthermore, before comparing SCM and AOGCM time series, Meinshausen et al. (~~2011~~2011a) low-pass filtered both with a cut-off frequency of 1/20 yrs. ~~For~~In the ~~sake~~interest of parsimonious analysis we avoided low-pass filtering here, as the ~~one~~1-box-only climate model PH99 essentially acts as a low-pass filter on the high~~-~~-frequency components of forcing. Therefore we decided to avoid introducing another degree of freedom in terms of a cut-off frequency into our analysis.

2.3 Step ~~three: validation~~Three

~~Finally~~Lastly, we validate the PH99 model versions generated in step two. For this purpose, independent temperature and forcing paths ~~are needed to~~must be run as a nontrivial test to check whether the trained climate module can accurately project other temperature data trajectories. To do so, the values for α and μ determined in step two are implemented in PH99, the latter then being driven by the total climate forcing of the RCP4.5 ~~scenario.~~and RCP8.5 scenarios. Similar to steps one and two, the deviations in ~~final 30-year~~2071-2100 means of GMT between PH99 and the respective APGCM are determined as an accuracy check.

One might be interested in seeing if the calibrated module is capable of mimicking other scenarios such as RCP6.0 or what if PH99 was calibrated to RCP4.5 or others. Stating that, in general, the procedure outlined above brings about similar results, for the sake of brevity of the main text, we present the respective results in Appendix 2.

3 Results

Table 1 shows the calculated α and μ together with the feedback response time $1/\alpha$ calculated in step one. For all of the indicators we also compute the mean values and standard deviations of the samples. The mean value of the ECS for GCM data is 3.35 K, with thea minimum and maximum of 2.11 K and 4.67 K, respectively. The mean value of timescalesthe time scales is 34.5roughly 35 years.

Figure 1Figure 1 represents the projected PH99 temperature evolution for the scenario RCP2.6 of each GCM in 2006-2100, using the data in Table 1from Table 1 and the RCP2.6s' forcings from RCP2.6. PH99 obviouslyclearly overestimates the temperature anomaly for all GCMs during, especially over the last 30 years. The absolute values of the deviations of the mean temperature over the last 30 years, (hereafter: MTD,) from the AOGCM data are shown in Fig. 2.Figure 2. The MTD ranges from 0.22 K for MRI-CGCM3 to aboutapproximately 0.79 K for HadGEM2-ES. On average, the deviations are aboutca. 0.45 K. This is clearly a large error, both in units of annual GMT standard deviation as well as the climate policy dimension. The signatories' A proclaimed goal of the 2015 Paris 2015 agreementAgreement (UNFCCC, 2016) stated goal isconsists in '…holding the increase in the global average temperature to well below 2°C above pre-industrial levels and pursuing efforts to limit the temperature increase to 1.5°C above pre-industrial levels…'. Hence a difference in 0.5 K does matter. Accordingly, we must proceed with step two.

In step two, for any of the GCMs, we tune α and μ such that the GMT deviations from the actual temperature of GCMs for the whole period 2006-2100 are minimized in a least squaresquares manner as represented in Fig. 3.Figure 3 and Figure 4. From the thereby adjusted α and μ we derive the ECS and TCR, which are presented in Table 2.Table 2. MTDs for the various AOGCMs are shown in Fig. 2.Figure 2.

From theThe results we find the following:tell us three main things. Firstly, the average of the absolute values of deviations is significantly reduced when α and μ are tuned. Indeed, the MTD average drops to below 0.02 K. Secondly, while the average of ECSECS decreases by 0.9 K (from 3.35 K to 2.46 K,), the average of TCR increases by 0.14 K (from 1.90 K to 2.04 K.). Thirdly, the mean value of feedback response times decreases significantly, from aboutroughly 35 years to belowless than 12 years.

For validation we move on to step three. WeIn this regard, we utilize the RCP4.5 temperature and forcing data as provided by Forster et al. (2013). In Fig. 3In Figure 3 and Figure 4 the respective GMT trajectories for any AOGCM are contrasted with the PH99-generated ones, whereby $\alpha\alpha$ and μ *are fixed to their* valuevalues as determined in step two. The MTDs are shown in Fig. 2.Figure 2. The results confirm that the climate module is so well trained in the second step that it can appropriatelysuitably mimic the actual temperatures estimated by the AOGCMs for RCP4.5, hence also out of the sample it was fitted to. and RCP8.5. As shown, the average value MTD is approximately 0.05 K for RCP4.5 and about 0.05 K. The14 for RCP8.5. For RCP4.5, the deviations for three of the GCMs, namely CCSM4, CNRM-CM5, and NorESM1-M, are even better than the ones asthose diagnosed for RCP2.6 in step two. See Appendix 2 for further analyses.

4 A mapping of ~~*α*, *μ*, and~~ *ECS* onto their PH99-specific counterparts *α* and *μ*

Finally, we attempt to abstract from fitting PH99 to individual AOGCMs and provide an approximate way to calibrate PH99 within the cloud of AOGCMs ~~by~~ simply by knowing the ECS. Then PH99 could be utilized for any ECS in analyses where the ECS is uncertain. However, before diving into our suggestions, it is worthwhile ~~that we~~to first take a look at one of the existing options and utilize the curve suggested by Lorenz et al. (2012~~)~~), which correlates α and μ to the ECS. Using a sample from Frame et al. (2005) and assuming a strict relationship between $1/\mu$ and ECS, Lorenz et al. (2012) suggest the following approximation:

$$\frac{1}{\mu} \approx \frac{1}{\bar{\mu}} - 10\exp(-0.5\,ECS) \qquad (8)$$

where $\bar{\mu}$ is the mean value of μ in the sample (see Fig.7 in Lorenz et al., ~~2011~~2012, all quantities measured in the units utilized in Kriegler & Bruckner, 2004). Knowing μ, Eq. (2) is used to determine α. ~~Equation~~In turn, Eq. (2) and Eq. (8) have been repeatedly used in ~~the~~ studies employing MIND and concerning uncertainties ~~on~~and ECS~~.~~ (Neubersch et al., 2014; Roshan et al., acc., Roth et al., 2015).

We employ Eq. (2) and Eq. (8) for all ~~*ECS*~~ECSs from ~~Table 1~~Table 1 and show the MTDs for the RCP2.6 scenario in ~~Fig. 4. Obviously~~Figure 5. Clearly, on average, employing Lorenz's curve does not result in a better situation than step one. However~~ hereby one~~, this might not ~~have compared~~necessarily be a case of comparing like with like. ~~The~~At the time of Frame et al. (2005), the two-dimensional uncertainty information~~ as of Frame et al. (2005)~~ was obtained by reconstructing the 20th century's warming signal from fingerprinting by means of a single AOGCM and then using ~~these~~this observational data as a constraint. It is well~~-~~ known that observational constraints may lead to different distributions than ensembles of AOGCMs ~~would~~ do (Andrews & Allen, 2008). Nevertheless we include this piece of information here ~~to highlight that~~for the ~~two methods differ and a conscious decision is needed which one to use~~sake of completeness.

Given the inferred estimations ~~shown~~ in ~~Table 2,~~Table 2, one ~~can~~ could attempt to directly relate α and μ to the *ECS*. ~~For this~~To do so, we generate polynomial fits (of orders 2 and 3) of α and μ against all ~~AOGCM's~~AOGCMs' ECSs. The attempt to predict a two-dimensional manifold from ECS alone implicitly exploits the fact that ~~AOGCM's~~AOGCMs' TCRs can ~~well~~ be predicted ~~from~~well using ECSs (see e.g. Meinshausen et al., 2009) in a statistical sense. Therefore, another option would be deriving α and μ analytically (~~as~~like in the first step) when the inferred *ECS* and *TCR* are correlated to the *ECS* and *TCR* of AOGCMs.

~~Figure 5~~Figure 6 relates α and μ (from ~~Table 2)~~Table 2) to the ECS (from ~~Table 1),~~Table 1), using linear, quadratic and cubic polynomial approximations. ~~Figure 4~~For the case of linear approximation, we put the model GISS_E2_R out as an outlier. Figure 5 indicates that on average ~~both~~all approximations mimic the actual temperature paths better than a non-fitted one. The cubic estimation projects significantly smaller deviations compared to the quadratic approximation~~.~~ and slightly smaller deviations compared to the linear approximation. The maximum MTD in the cubic approximation is ~~about~~ 0.3 K for IPSL-CM5A-LR, which is ~~about~~roughly a third of the maximum in the quadratic approximation that is revealed for CSIRO-Mk3-6-0.

~~In the following we describe~~We ask for alternative ways to ~~extrapolate~~map ECS and TCR from the 14 utilized AOGCMs ~~on any *ECS*~~onto PH99-intrinsic properties, going beyond the scheme displayed in ~~Fig. 4.~~Figure 6. As one option, shown in ~~Fig. 6,~~Figure 7, we linearly regress the *ECS* and *TCR* values inferred from step two against their original AOGCM counterparts respectively and obtain

$$ECS_{\mathrm{PH99}} \approx a\, ECS_{\mathrm{AOGCM}} + b \qquad (9)$$

with a= 0.5846, b= 0.5095 K, and R-square=0.8158, as long as $ECS_{\mathrm{PH99}} < ECS_{\mathrm{AOGCM}}$.

and

$$TCR_{\mathrm{PH99}} \approx c\, TCR_{\mathrm{AOGCM}} + d \qquad (10)$$

with c= 0.9763, d= 0.1829 K, and R-square=0.667.

~~Another~~The other option ~~is~~consists in using Eq. (9) along with a linearly regressed TCR_{PH99} over ECS_{AOGCM}, that is

$$TCR_{\mathrm{PH99}} \approx m\, ECS_{\mathrm{AOGCM}} + n \qquad (11)$$

with m= 0.4582, n= 0.5044 K, and R-square=0.7876.

The respective MTDs are shown in ~~Fig. 4.~~Figure 5. Although both approximations mimic the actual temperature paths better than a non-fitted one, regressing both the inferred effective ECS and TCR solely against AOGCMs' ECS (hereafter~~,~~: ETE~~, shown in Blue in Figure 4) obviously is the~~) clearly offers the better overall ~~better~~ approximation.

~~The use of~~Using the ETE has ~~a three-fold advantage~~four major advantages over all other options ~~described above~~dealt with here, especially for the IAM community. Firstly, its approximation is better than all options but the ~~cubit~~cubic fit. Even though the cubic fit may ~~give~~yield a better approximation, in our analysis it is only better by 0.03 K~~. Still the ETE~~ at the expense of a non-intuitive shape that might result in even worse deviations for out of sample data. Secondly the ETE still has an advantage over ~~it~~ the cubic fit because~~, secondly,~~ one can easily ~~make~~ use ~~of~~a broader range of climate sensitivities, for example, ~~ranging~~ from 1 K to 9 K, which may not be accurately determined by the cubic fit. ~~Lastly, by use of~~ Thirdly, the ETE~~,~~ not only ~~it projects~~yields a better approximation~~, but also~~; prior ~~probabilistic~~ knowledge ~~on~~ regarding the TCR is ~~not a necessary input – which is convenient if not yet available~~no longer a decisive factor. Please note that prior knowledge regarding the TCR can make approximations better. However, as we tested, for example, in the case of linearly regressing both the inferred effective ECS and TCR against both AOGCMs' ECS and TCR, the R-squares for Eq. (9) and Eq. (11) only improve by 6% and 7% respectively, and the MTD is no better than the cubic fit. Finally, in the case of ETE, we do not need to re-evaluate our sample and possibly drop any model as an outlier. For the sake of brevity, we do not go beyond linear approximation here.

5 An analytic interpretation of the AOGCM-PH99 intercomparison

In the following, we explain why PH99 systematically overestimates maximum GMT for peaking scenarios when fitted for exponentially growing scenarios. We trace back all of the effects reported so far to the information loss incurred by replacing a 2-box SCM (as utilized in DICE (Nordhaus, 2013)) with a 1-box SCM like PH99. We then also investigate the quality of alternative fitting schemes based on our semi-analytic analysis, which complements our previously mentioned AOGCM-based validation.

Following Geoffroy et al. (2013) we introduce a 2-box SCR as a more universal emulator of AOGCMs' mapping from radiative forcing onto temperature.

$$C\,\frac{dT_{2B}}{dt} = F - \lambda T_{2B} - \gamma(T_{2B} - T_0) \qquad (12)$$

$$C_0\,\frac{dT_0}{dt} = \gamma(T_{2B} - T_0)$$

T_{2B} denotes the 2-box analogue of the 1-box temperature T in Eq. (1). The upper and the lower equation represent the upper and the lower ocean, respectively.

In order to contrast PH99 with this 2-box model, we search for analytic approximations of generic shapes of the forcing F(t) and examine the long-term perspective on various RCPs as depicted in Meinshausen et al. (2011b) – an excerpt is included in Figure 8 for the reader's convenience. Particularly in view of the peaking, mitigation-oriented lowest forcing scenario, we approximate forcing paths in three phases: zero forcing, linear increase, and linear decrease, under a continuity assumption.

$$F(t) = \begin{cases} 0 & \text{for } t < 0 \\ k_1 t & \text{for } 0 \le t \le t_1 \\ k_2(t - t_1) + k_1 t_1 & \text{for } t > t_1 \end{cases}$$

We approximately identify $t1$ with the year 2035 and t=0 with 100 years earlier, i.e. we assume a ramp-up time $t1$ for the forcing of roughly 100 years. Furthermore, $k2$<0 and $|k2/k1| =: \varepsilon \ll 1$. From Figure 8 we find a generic value of ε=0.2. For $0 \le t \le t_1$ we draw on Geoffroy et al. (2013)

$$T_{2B}(0 \le t \le t_1) = \frac{k_1}{\lambda}\left(t - \tau_f a_f\left(1 - e^{-\frac{t}{\tau_f}}\right) - \tau_s a_s\left(1 - e^{-\frac{t}{\tau_s}}\right)\right)$$

This represents two linear modes of amplitudes a_f and a_s, delayed by the characteristic time scales of a fast and a slow mode, τ_f and τ_s, respectively, and continuously matched to the initial condition '0' by an exponential. In Geoffroy et al. (2013) the 2-box model is fitted to 16 AOGCMs. After having reviewed their results for our order-of-magnitude estimates of PH99's accuracy, we can make the following two simplifying assumptions: (i) both amplitudes a_f and a_s approximately equal 1/2 (see their Fig. 3a – amplitudes range from 0.35 to 0.65), (ii) $\tau_f \approx 0$ (values range from 1 yr. to 5.5 yrs., see their Table 4; for centennial effects, this mode would nearly match the equilibrium response). Furthermore we can see that τ_s ranges from 100 yrs. to 300 yrs. for 15 out of 16 AOGCMs. Hence the 2-box model is characterized by a marked time-scale separation between the two linear modes. With the aid of these two approximations, the last equation can be simplified to

$$T_{2B}(0 \le t \le t_1) \approx \frac{k_1}{\lambda}\left(t - \frac{\tau}{2}\left(1 - e^{-\frac{t}{\tau}}\right)\right) \text{ with } \tau := \tau_s.$$

We then extend the analytic range of that formula, given the two approximations above, for $t > t_1$:

$$T_{2B}(t > t_1) \approx \frac{k_1}{\lambda}\left(-\varepsilon t + (1+\varepsilon)t_1 + \frac{\tau}{2}\left(\varepsilon + e^{-\frac{t}{\tau}} - (1+\varepsilon)e^{-\frac{(t-t_1)}{\tau}}\right)\right)$$

The analogous expression for the 1-box model is

$$T(0 \le t \le t_1) = \frac{k_1}{\lambda_{1B}}\left(t - \theta\left(1 - e^{-\frac{t}{\theta}}\right)\right), \theta := \frac{1}{\alpha}, \lambda_{1B} := \frac{Q_2}{ECS_{1B}},$$

the λ-analogue for the 1-box model, λ_{1B}, being inferred from Eqs. (4) and (7), and

$$T(t > t_1) = \frac{k_1}{\lambda_{1B}}\left(-\varepsilon(t - \theta) + (1 + \varepsilon)t_1 + \theta\left(e^{-\frac{t}{\theta}} - (1 + \varepsilon)e^{-\frac{(t - t_1)}{\theta}}\right)\right).$$

5.1 Explaining the PH99-AOGCM discrepancy for equal ECS and TCR values

We are now prepared to mimic Step One in Section 2: we calibrate the 1-box model such that it is characterized by the same ECS and TCR as the 2-box model. As $\lambda = Q_2/ECS_{2B}$, equal ECS values for both models deliver $\lambda_1 = \lambda$.

We introduce t_{TCR} as the moment in time when T needs to be evaluated in order to determine the TCR. In Appendix 1 we recapitulate that $t_{TCR} = (\ln 2)/\gamma \approx 70$yrs for a rate of 1%/yr. In order to determine θ from the TCR, it proves useful to define the auxiliary function (see Figure 9)

$$h(x) := \left(1 - e^{-\frac{1}{x}}\right)x, \quad \text{resulting in} \quad \lim_{x\to 0} h(x) = 0, \lim_{x\to\infty} h(x) = 1, \quad h(x) \approx x \text{ for } x \ll 1.$$

Then the defining condition for θ, after some manipulation, can be written as $h(\theta / t_{TCR}) = (1/2)\, h(\tau / t_{TCR})$. From this, we can already get a first impression of the scale of θ, prior to numerical inversion: as τ is generically markedly larger than t_{TCR}, the right-hand side of the defining equation above approximates ½. Further, if we boldly assume a slight time-scale separation between θ and t_{TCR}, the former being smaller than the latter, then the linear approximation of h would apply and $\theta \approx t_{TCR}/2 \approx 35$ yrs. For a centered value of τ=250 yrs, this approximation is confirmed in a direct numerical treatment of $h(\theta/t_{TCR}) = (1/2)\, h(\tau/t_{TCR})$.

Hence from the twin time-scale separation of 'the 1-box model mode,' 'defining time scale for TCR,' and the 'slow mode of the 2-box model' we have explained why TCR-oriented fitting exercises of the 1-box model would generically result in time scales of roughly 30 to 40 years (see e.g. Anthoff & Tol, 2014; Kriegler & Bruckner, 2004). The factor ½ between the 1-box model's time scale and the TCR-defining time scale goes back to Geoffroy et al.'s (2013) observation that the fast and the slow mode both enter the superposition result with approximately equal weights of ½. The slow mode is then too slow to be of much relevance for TCR – a phenomenon not revealed by the 1-box model.

We are now equipped to compare the two models' temperature projections and apply the 3-phase forcing as defined above for ε=0.2. $a1/\lambda$ is chosen such that peak temperatures enter the 2° regime for illustrative purposes. We exploit the coincidence that t_{TCR} just happens to approximately correspond to our starting year 2006 for PH99 (because 2035-100+70=2005). Hence the formulas for the 1-box model do not need to be adapted for an explicit initial condition for this purpose. Figure 10 shows that by construction, both temperature responses match at $t_{TCR} \approx 70$ yrs., although the 1-box model's maximum exceeds the exact maximum by ½°. This phenomenon can be explained as follows. As the 1-box model responds with a finite time scale, its derivative must be continuous in response to a continuous forcing. Hence the leading term is quadratic when the forcing starts. In contrast, the 2-box model contains a virtually degenerate time scale (the fast one); hence its leading term is linear. If the two curves are to nevertheless match at t_{TCR}, the 1-box model's derivative at t_{TCR} must transcend the 2-box model's derivative. This, together with the right-bending kink in the 2-box model's response at *t1*, leads to a larger maximum in the 1-box model. In summary, on time-scales much smaller than the slow mode, the slow mode, compared to the fast mode, cannot develop yet; hence the fast mode will dominate the slow mode. As such, fitting a 1-modal model in a convex regime is likely to yield poor predictions of a temperature maximum for mitigation-based forcings. This explains the discrepancies found in our PH99-AOGCM comparison when directly transferring AOGCMs' ECS and TCR onto PH99. Figure 10 further suggests that if PH99 were used to predict correct maxima and emulate AOGCMs in this time regime, it would need to be used with a markedly smaller time scale. However, a simple reduction in time scale would lead to a new inter-model discrepancy before the kink; hence the overall amplitude of PH99's response would need to be reduced as well. The latter scales with the ECS; hence the ECS must be reduced by a certain factor towards a new 'effective ECS,' which could also be called a 'transient climate sensitivity.'

5.2 Testing the validity of a recalibrated PH99 for a 2-box model

We now systematically test the previously suggested adjustment formulas Eqs. (9) to (11) for a range of *t1* and ε values, emulating alternative forcing scenarios, given the ECS and τ for the 2-box model. We test for the centerred ECS values of 3 K and 4 K and a slow mode's time scale, which generically ranges from 100 yrs. to 300 yrs (see Geoffroy et al., 2013). Given the fact that here, the GMT of PH99 is to be mapped to the GMT of the 2-box model, and the TCRs have been transformed, we need an expanded version of the original GMT formula for PH99 for which an initial value at $t_0 < t_1$ (t_0 representing the initialization year 2006) is explicitly foreseen:

$$T_{\text{init}}(t > t_1) = \frac{k_1}{\lambda_{1\text{B}}}\left(-\varepsilon(t-\theta)+(1+\varepsilon)t_1 + \theta\left(\mathrm{e}^{-\frac{t}{\theta}} - (1+\varepsilon)\mathrm{e}^{-\frac{(t-t_1)}{\theta}}\right) + (T_{2\text{B}}(t_0) - T(t_0))\mathrm{e}^{-\frac{(t-t_0)}{\theta}}\right),$$

hereby $\lambda_1 = \lambda\, ECS_{2\text{B}}/ECS_{1\text{B}}$ and θ numerically determined from the request $T(t_{TCR}) = TCR1B$. We find numerically that θ is on the order of 10 years, and the ECS needs to be reduced by 1/4 to 1/3. Figure 11 shows the relative deviations of the GMT maxima of the 1-box and the 2-box model for the extrapolation scheme ETE (Eqs. (9) and (11)). In a certain regime, the extrapolation delivers sufficiently accurate results, however, not everywhere. When utilizing the mapping scheme represented by Eqs. (9) and (10), the results look similar. The overall impression is that the mapping removes the bias, however would deliver not an as universal correction as found for the direct intercomparison between PH99 and AOGCMs. We cannot exclude the possibility that AOGCMs are easier to emulate as they contain many more time scales than the 2-box model and their effects might in part cancel.

While we observe a qualitative gain, Figure 11 reveals there is still room for improvement. Accordingly, we further transform the ECS to request perfect matching for *t1*=100 yrs, ε=0.2; the results can be seen in Figure 12. The fit is much further improved such that a major fraction of (t1, ε) values would lead to a relative error of <5%, and another large fraction to a relative error of <10%. As the standard deviation of annual GMT is between 0.1°C and 0.2°C and a typical application might be a cost-effectiveness analysis of the 2°C target, such errors might still seem tolerable. However we observe structural problems for very small values of ε, the latter implying very late assumption of a maximum. Hence here, the slow mode becomes more relevant, then no longer consistent with the original calibration. The calibration is valid for a time horizon on the order of t_1 to 2 t_1.

6 Discussion

~~Why are~~ The previous section offers a key mechanism to explain why, for given ECS and TCR, GMT scenarios generated ~~from~~by PH99 are biased towards higher temperatures~~? In the literature it has been discussed, for decades, that in systems more complex than just one box, transient warming dynamics are governed by a 'transient climate sensitivity' markedly lower than ECS (for an overview see, e.g. Meinshausen et al., 2011, in conjunction with their Fig.1). Andrews & Allen (2008) warned againsta straightforward usage of PH99 in that regard. However these diagnostics mainly referred to convexly increasing GMT paths while RCP2.6 leads to quasi-stabilizing GMT paths.~~

~~Hence~~ . One should not forget about potential additional mechanisms~~ for explanation should not be ruled out ex ante. This means, that, firstly~~. Firstly, the statistical errors in determining AOGCMs' ECS, TCR~~,~~ and Q_2 may lead, mediated through the nonlinear mapping on PH99's parameters, to an overall bias in PH99's GMT. Furthermore, diagnosing the total radiative forcing active in an AOGCM is a complex ~~issue~~undertaking (see e.g. ~~again~~ Meinshausen et al., ~~2011~~2011a, for a discussion). A bias to the high~~-~~ end here would also result in ~~too~~inaccurately large GMT responses by PH99. However, in the context of this article, we would like to focus on our main finding and contend that the information loss when moving from a 2-box to a 1-box model is the key source of the observed discrepancy – last but not least, we find Figure 10 compelling in this regard. ~~However, assuming that all AOGCM-based inputs into the current analysis were unbiased, our findings would imply that past integrated assessment studies based on PH99 implicitly worked with ECS values that were larger than announced, as shown in Fig. 6. Since Lorenz et al. (2012), the probability density function by Wigley & Raper (2001), a log-normal function, has been utilized within the MIND model. In the remainder of this paragraph we argue why we would also prefer log-normal distributions for ECS in the future. The rationale behind using a log-normal function is our personal conviction that constraining ECS by paleo data in addition to instrumental records results in thin-tailed distributions. This conviction rests on an approach successfully applied at an Earth system model of intermediate complexity (Schneider von Deimlinig et al., 2006). In contrast, relying solely on the instrumental record must result in fat-tailed distributions (Roe & Baker, 2007). However constraining ECS by paleo data still does not seem to be on the same level of quality than doing so by the modern instrumental record. A log-normal distribution would then describe the borderline case of a fat and a thin tail, hence expresses mainly relying on the observational record, but also allowing for Bayesian learning from paleo data.~~

Suppose we accepted above suggestion to map ECS values onto effective, scenario class-adjusted values boefore usage in PH99. Would a re-scaled version of that lognormalIf the reader will join us in exploring this line of reasoning, it raises a key question: Can PH99 be seen as a ‘physical model’ and if so, what are the implications for users? It is readily apparent that a 1-box model cannot mimic a 2-box model, characterized by a marked time-scale separation for all forcings at all times. However it is equally clear that the simplest temperature equation is in fact the one that treats the ocean as a single box. It would still explain warming with forcing in a quasi-linear manner, though with some delay. If we are willing to accept that the calibration of PH99 is time-horizon-specific, i.e., that a distinction has to be made between a ramp-up phase and a peak-and-decline phase, then PH99 still holds some semi-physical meaning. If, however, the need to augment the range of validity of a single calibration is seen as the very definition of an ‘unphysical’ model, then we would have to recognize that PH99 is more an efficient emulator than a physical model. In this context we would like to recall that virtually every model has a limited range of validity – and as such, PH99 is no different from most other models.

For future applications we can conclude that PH99 must be applied and interpreted with greater care than in the past, if it is not to be replaced by an at least 2-box model as suggested by Geoffroy et al. (2013) and implemented in DICE (Nordhaus (2013)). However, when investigating the 1-box / 2-box-models’ differences, our research also suggests that within the class of peak-and-decline (or at least stabilizing) scenarios PH99 provides excellent emulation for various scenarios such as RCP2.6, 4.5, 6.0 and, to a lesser extent, 8.5. The latter becomes and remains convex longer, limiting the accuracy of the RCP2.6-based calibration.

What are the ramifications of our findings for previous publications based on PH99? Those authors who claimed to have worked with ECS=3°C would have effectively worked with ECS≈4°C. Much of the work done in our group, based on MIND in conjunction with PH99 and the log-normal distribution for ECS by Wigley & Raper (2001), has essentially been based on a distribution still comply with our recent knowledge on ECS?shifted to larger ECS values. The 5%, 50%,% and 95% % quantiles of the log-normal distribution by Wigley & Raper (2001) are 1.2 K, 2.6 K, and 5.8 K, respectively. When interpreting these values as PH99 values, as they have been in fact been utilized in PH99 for the MIND model since Lorenz et al. (2012) for the MIND model,), one could ask for the back-transformed ECS values according to our Fig. 6.Figure 7. The respective values are 1.2 K, 3.6 K, and 9.0 K. From Fig. 7, from the Figure 13, which reflects IPCC AR5’s synopsis of current knowledge onregarding ECS (Bindoff et al., 2013)), we can see that these are still in- line with the range spanned by instrumental studies. Hence the results obtained by PH99 in conjunction with the distribution by Wigley & Raper (2001) are not errornouserroneous, but simply need to be re-interpreted as rather high-end representatives within the collection of ranges as describedseen in IPCC AR5.

In fact Fig. 6 might even exaggerate the need for correcting ECS and TCR values. As we purposely avoided usage of a low-pass filter before fitting PH99 to AOGCMs output, improved goodness of fit was obtained for slightly emulating natural variability (see the small wiggles of PH99-scenarios as displayed in Fig. 3). This in turn might have led to too small inferred feedback response times which usually correlate with too low ECS values. Future work will include a sensitivity study regarding the effects of cut-off frequency of low-pass filtering. However this possibility cannot eliminate the apparent need

~~to re-interpret ECS and TCR of PH99, as it is obvious from our observed emulator-AOGCM difference of 0.5 K in GMT (see Fig. 1).~~

~~Furthermore, if such a transformation were found necessary by future work, how would one generate it? Our paper solely rests on the properties of 14 CMIP5 AOGCMs that may not sample our current joint distribution of ECS/TCR well. In fact Andrews & Allen (2008) pointed to the possibility that ensembles of GCMs misrepresented an actual joint distribution as inferred from the instrumental record. Hence we propose deriving two separate transfomation schemes: one for a specified set of AOGCMs as our current work does and a further one for probabilistic information generated from relating a state-of-the-art SCM or Earth system model of intermediate complexity to the instrumental record (examples for the latter: see the probabilitic studies as cited in Fig. 7).~~

~~In any case such transformation rules must be able to tackle the whole range of potential ECS and TCR values in order to enable us to transform density functions. Quite the contrary, our rather pragmatic polynomial fit as of Fig. 5 should not be extrapolated beyond the range spanned by the 14 AOGCMs utilized in this study. Particularly a 3rd order polynomial would generate unphysical effects. Furthermore such a transformation should treat ECS and TCR as independent entities, while we treated ECS as the only predictor in Fig. 4 and Fig. 5, just to demonstrate the chance of finding a model-independent transformation.~~

~~Finally when adjusting PH99 according to such a transformation or even directly fitting to time series of an AOGCM, PH99 can emulate an AOGCM to a sufficient level of accuracy. Hereby we utilized RCP2.6 scenarios for fitting, and RCP4.5 scenarios for validation. As economic optimizers would, in any practical sense, employ only variations of RCP2.6 scenarios when finding a constrained welfare optimum compatible with the 2° target, those deviations from the RCP2.6 scenario would generically be an order of magnitude smaller than the difference as against the RCP4.5 scenario used for validation. If emulation quality for the RCP4.5 scenario were sufficient, the more so we expect sufficiency for scenarios approximantely in line with the 2° target!~~

~~6 Summary and Conclusion~~

7 Summary and Conclusion

We utilize recent data on total radiative forcing (Forster et al., 2013) ~~of~~from 14 ~~CMIP5~~ state-of-the-art CMIP5 Atmosphere Ocean General Circulation Models (AOGCMs) in order to test the validity of the ~~one~~1-box climate module by Petschel-Held (1999, ‘PH99’) for scenarios approximately compatible with the 2° target. PH99 is currently utilized within the integrated assessment models FUND, MIND and ~~MIND~~PAGE.

We find that when prescribing the equilibrium climate sensitivity (ECS) and transient climate response (TCR) of these AOGCMs to the emulator PH99, ~~generally leads to an overestimation of~~ global mean temperature (GMT) ~~by~~ is generically projected 0.5 K~~. Quite the contrary, when~~ higher. In contrast, by directly fitting PH99 to the RCP2.6 time~~-~~ series and validating ~~by~~with the RCP4.5 and RCP6.0 series, we find that PH99 can emulate AOGCMs to a degree of accuracy better than 0.1 K. Even for RCP8.5 the error is on the same order of magnitude, although somewhat larger.

We ~~propose several explanations for the intolerable discrepancy of emulators and AOGCM when directly prescribing ECS and TCR. The first candidate to be checked in future investigations would be RCP2.6 scenarios being characterized by transient climate sensitivities markedly smaller than ECS, as already known for purely convex temperature scenarios.~~ ~~However we find~~numerically demonstrate that PH99 can be used ~~for emulating~~to excellently emulate AOGCMs within ~~an accuracy~~integrated assessment of ~~0.1 K to 0.~~the 2~~ K if~~° target, provided its ECS and TCR are re-interpreted as effective, ~~2°C-~~ scenario-class~~–~~-specific values and mapped from original ECS and TCR values. We suggest a first version of such a mapping.

~~Older~~Furthermore we explain the observed discrepancies and the need to reduce PH99's ECS compared to the AOGCM's ECS as being due to the information loss produced by approximating a 2-box-based energy balance model with a 1-box-based model. The slow mode of the 2-box model is so slow that in a climate-policy-relevant context it can unfold only up to a relatively small extent; hence for practical purposes the 2-box model's ECS cannot fully develop. Accordingly, adjusting the ECS to lower values also proves to be compatible with reducing PH99's response time. When comparing PH99 and AOGCMs, the match is even better – a phenomenon the explanation of which is beyond the scope of this article.

Hence older work based on PH99, executed within FUND, MIND and ~~MIND~~PAGE, may need to be re-interpreted in the sense that higher ECS values ~~of ECS~~ had effectively been ~~operated. For~~used. Even when having dealt with distributions of ECS as for the MIND model, ~~even~~ECS values re-interpreted ~~values of ECS~~in this manner are still within the range outlined by IPCC AR5 (see ~~Fig. 7).~~Figure 13).

For future work, we propose the following steps: (i) ~~It has to be checked to what extent the transformations on ECS and TCR we found are robust under low-pass filtering scenarios before fitting. (ii)~~ By comparison with more sophisticated, multi-box climate modules it should be tested again whether the effect of a transient climate sensitivity (and TCR) alone could explain our observed PH99-AOGCM discrepancy. (~~iii~~ii) Future discussions with the AOGCM community should illuminate to what extent the further explanations we suggested might also apply, thereby potentially reducing the ~~correction~~ need to correct for PH99. (~~iv~~iii) An AOGCM-independent, yet scenario-class-specific ~~2~~two-dimensional mapping from ECS/TCR onto ECS/TCR and designed for PH99~~,~~ should be derived in conjunction with two-dimensional distributions inferred from observations as done in Frame et al. (2005). The IAM community could then be offered both options for emulation~~,~~: the one presented here, trained by AOGCMs, and ~~the~~ one based on observational data and mediated ~~via~~by more complex SCMs.

~~Thus in both cases, the use of~~In summary, PH99 could continue to be ~~continued~~used as the most parsimoneous emulator of AOGCMs, and is especially efficient for decision-making under climate response uncertainty. However its calibration proves to be much more involved than previously assumed. Future users should carefully consider whether they actually want to use PH99, or whether they prefer a less parsimonious solution.

Appendix 1: An Analytic Expression ~~for~~of TCR ~~within~~in PH99

We recapitulate Eq. (1) as

$$\dot{T} = \mu \ln(c) - \alpha T \qquad \text{(A1)}$$

TCR is defined as the temperature change in response to a 1%/yr. increase in CO_2 concentration, starting from preindustrial conditions. Hence the concentration, expressed in units of the pre-industrial concentration, reads

$$c = \exp(\gamma t) \qquad \text{(A2)}$$

~~hereby~~with γ denoting the above rate of change. As Eq. (A1) represents a linear ordinary differential equation with constant coefficients, and the initial temperature anomaly ~~vanishes~~is to vanish, its solution reads

$$T = \mu\gamma \exp(-\alpha t) \int t \exp(\alpha t) \mathrm{d}t = \frac{\exp(-\alpha t)\, \mu\gamma\, (1 + \exp(\alpha t).(-1+\alpha t))}{\alpha^2} \qquad \text{(A3)}$$

Temperature should be evaluated at t_2 when the concentration is doubled. t_2 is determined by $c(t_2)=2 \Rightarrow t_2=\ln 2/\gamma$. From this and Eq. (A3) we conclude Eq. (3). (In fact we ~~retrieve~~find the same result ~~from~~using an expression ~~given~~provided in Andrews & Allen, 2008, when we plug in our expression for t_2 into ~~their expression that~~theirs, which is phrased in terms of ECS.)

Appendix 2: Further Analysis on Calibration and Validation

As further validation of the trained PH99 calibrated to RCP2.6, Figure 14 shows the respective GMT trajectories of AOGCMs for RCP6.0 scenario contrasted with its respective PH99-generated ones whereby α and μ are fixed to their value as determined in step two. MTDs are shown in the 3rd columns of Table 3. The missing models are due to either lack of temperature trajectories for AOGCM or lack of total forcing. Notice that 1st, 2nd, and 4th columns are exactly the numbers related to the Figure 2. The results confirm that the climate module is so well trained in the second step that it can appropriately mimic the actual temperatures for RCP6.0. As shown, the average value of MTD is about 0.06 K for RCP6.0. Columns 5th thereafter in Table 3 show MTDs in the situations when PH99 is calibrated to the other RCP scenarios and is validated as against the others. Over all, the results show that PH99 would be well trained by being calibrated to any RCP scenario.

Authors' Contributions

M.M.K. ~~did~~performed the statistical analysis~~, wrote the codes, and prepared the visualizations.~~. H.H. ~~suggested and developed step one of the~~ provided the analytic analysis~~,~~. M.M.K. suggested and developed the alternative scheme. ~~H.H. wrote the framing sections of the manuscript, M.M.K. wrote the descriptive and analysing sections.~~ Both participated in the writing of the article.

Competing Interests

The authors declare that they have no ~~conflict~~conflicts of interest.

Acknowledgments

The authors would like to thank J. Marotzke for ~~pointing them~~drawing their attention to the Forster et al. (2013) article, discussing these results on total forcing and providing the relevant data. In addition, the authors would like to thank C. Li for supporting the data handling process and ~~pointing~~making the authors ~~to~~aware of Geoffroy et al. (2013~~ describing~~), which discusses negligible AOGCM drift. The authors are also grateful to E. Roshan for her help ~~on visualization~~with the visualizations and providing ~~quantilies~~quantiles of Wigley's & Raper's (2001) distribution on ECS~~ and to B. Blanz~~. We thank M. Fentem for ~~carefully reading~~proofreading the second version of our manuscript~~.~~ from a native speaker's perspective. M.M.K. ~~has been~~was supported by the Cluster of Excellence 'Integrated Climate System Analysis and Prediction' (CliSAP, DFG-EXC177).

References

Andrews, D. G. and Allen, M. R.: Diagnosis of climate models in terms of transient climate response and feedback response time, Atmospheric Science Letters, 9, 7–12, 2008.

Anthoff, D. and Tol, R. S. J.: The Climate Framework for Uncertainty, Negotiation and Distribution (FUND): Technical description, Version 3.6~~,~~ (retrieved November, ~~2013~~2016), 2014.

Bindoff, N. L., Stott, P. A., AchutaRao, K. M., Allen, M. R., Gillett, N., Gutzler, D., Hansingo, K., Hegerl, G., Hu, Y., Jain, S., and others: Detection and attribution of climate change: From global to regional, 2013.

Bruckner, T., Petschel-Held, G., Leimbach, M., and Toth, F. L.: Methodological aspects of the tolerable windows approach, Climatic Change, 56, 73–89, 2003.

~~Bruckner, T., Petschel-Held, G., Leimbach, M., and Toth, F. L.: Methodological aspects of the tolerable windows approach, Climatic Change, 56, 73–89, 2003.~~

Calel, R. and Stainforth, D. A.: On the Physics of three Integrated Assessment Models, Bull. Amer. Meteor. Soc., ~~doi:10.1175/BAMS-D-16-0034.1, 2016.~~98(6), 1199–1216 (2017).

Clarke, L., Jiang, K., Akimoto, K., Babiker, M., Blanford, G., Fisher-Vanden, K., Hourcade, J.-C., Krey, V., Kriegler, E., Löschel, A., and others: Assessing transformation pathways, 2014.

Edenhofer, O., Bauer, N., & Kriegler, E. (2005). The impact of technological change on climate protection and welfare: Insights from the model MIND. Ecological Economics, 54(2-3), 277-292.

Forster, P. M., Andrews, T., Good, P., Gregory, J. M., Jackson, L. S., & Zelinka, M. (2013). Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. Journal of Geophysical Research: Atmospheres, 118(3), 1139-1150.

Frame, D. J., Booth, B. B. B., Kettleborough, J. A., Stainforth, D. A., Gregory, J. M., Collins, M., & Allen, M. R. (2005). Constraining climate forecasts: The role of prior assumptions. Geophysical Research Letters, 32(9).

Geoffroy, O., Saint-Martin, D., Olivié, D. J. L., Voldoire, A., Bellon, G., and Tytéca, S.: Transient Climate Response in a Two-Layer Energy-Balance Model. Part I: Analytical Solution and Parameter Calibration Using CMIP5 AOGCM Experiments, J. Climate, 26, 1841–1857, doi:10.1175/JCLI-D-12-00195.1, 2013.

Held, H., Kriegler, E., Lessmann, K., & Edenhofer, O. (2009). Efficient climate policies under technology and climate uncertainty. Energy Economics, 31, S50-S61.

Hope, C.: The Marginal Impact of CO2 from PAGE2002: An Integrated Assessment Model Incorporating the IPCC's Five Reasons for Concern, The Integrated Assessment Journal, 6 (1), 19-56, 2006.

Kattenberg, A., Giorgi, F., Grassl, H., Meehl, G. A., Mitchell, J. F., Stouffer, R. J., Tokioka, T., Weaver, A. J., and Wigley, T. M.: Climate models—projections of future climate, Climate Change 1995: The Science of Climate Change. Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change, 285–357, 1996.

Kriegler, E. and Bruckner, T.: Sensitivity analysis of emissions corridors for the 21st century, Climatic Change, 66, 345–387, 2004.

Kunreuther, H., Gupta, S., Bosetti, V., Cooke, R., Dutt, V., Ha-Duong, M., Held, H., Llanes-Regueiro, J., Patt, A., Shittu, E., and others: Integrated risk and uncertainty assessment of climate change response policies, 2014.

Lorenz, A., Schmidt, M. G. W., Kriegler, E., and Held, H.: Anticipating Climate Threshold Damages, Environ Model Assess, 17, 163–175, doi:10.1007/s10666-011-9282-2, 2012.

Luderer, G., Leimbach, M., Bauer, N., and Kriegler, E.: Description of the ReMIND-R model, Potsdam Institute for Climate Impact Research. Retrieved from: https://www.pik-potsdam.de/research/sustainable-solutions/models/remind/REMIND_Description.pdf, 2011.

Meinshausen, M., Raper, S. C. B., and Wigley, T. M. L.: Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – Part 1: Model description and calibration, Atmos. Chem. Phys., 11, 1417–1456, doi:10.5194/acp-11-1417-2011, ~~2011~~2011a.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J. F., ..., and Thomson, A. G. J. M. V.: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. Climatic change, 109(1-2), 213, 2011b.

Meinshausen, M., Meinshausen, N., Hare, W., Raper, S. C. B., Frieler, K., Knutti, R., Frame, D. J., and Allen, M. R.: Greenhouse-gas emission targets for limiting global warming to 2 C, Nature, 458, 1158–1162, 2009.

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. D.: The climate casino: Risk, uncertainty, and economics for a warming world, Yale University Press, 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.

~~Roe, G. H. and Baker, M. B.: Why is climate sensitivity so unpredictable?, Science (New York, N.Y.), 318, 629–632, 2007.~~

Roshan, E., ~~Mohammadi~~ Khabbazan, M. M., Held, H~~.: A Scheme For Jointly Trading Off Costs And Risks Of Solar Radiation Management And~~., Cost-Risk Trade-off of Mitigation ~~Under Long-Tailed~~and Solar Geoengineering – Considering Regional Disparities under Probabilistic Climate Sensitivity~~ Probability Density Distributions, EAERE Zurich, 22-25 June 2016~~, Environmental and Resource Economics, manuscript EARE-D-17-00532R1 accepted for publication.

Roth, R., Neubersch, D., Held, H.: Evaluating Delayed Climate Policy by Cost-Risk Analysis, EAERE Helsinki, 24-27 June 2015.

~~Schneider von Deimling, T~~Stern, N.: The Stern Review – The Economics of Climate Change, Cambridge, 2007.

~~., 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.~~

Stankoweit, M., Schmidt, H., Roshan, E., Pieper, P., and Held, H.: Integrated mitigation and solar radiation management scenarios under combined climate guardrails, in: EGU General Assembly Conference Abstracts, 7152, 2015.

van Vuuren, D. P., ~~Edmonds, J. A., Kainuma, M., Riahi, K., and Weyant, J.: A special issue on the RCPs, Climatic Change, 109, 1–4, doi:10.1007/s10584-011-0157-y, 2011.~~

~~van Vuuren, D. P.,~~ Lowe, J., Stehfest, E., Gohar, L., Hof, A. F., Hope, C., Warren, R., Meinshausen, M., and Plattner, G.-K.: How well do integrated assessment models simulate climate change?, Climatic Change, 104, 255–285, doi:10.1007/s10584-009-9764-2, ~~2011~~2011a.

van Vuuren, D. P., Edmonds, J. A., Kainuma, M., Riahi, K., and Weyant, J.: A special issue on the RCPs, Climatic Change, 109, 1–4, doi:10.1007/s10584-011-0157-y, 2011b.

UNFCCC: United Nations Framework Convention on Climate Change. Adoption of the Paris Agreement. Conference of the Parties on its twenty-first session, 21932, 2016.

Wigley, T. M. and Raper, S. C.: Interpretation of high projections for global-mean warming, Science (New York, N.Y.), 293, 451–454, doi:10.1126/science.1061604, 2001.

Table ~~1:~~1: PH99 parameters (α and μ) and feedback response times ($1/\alpha$) utilizing data (ECS and TCR) from AOGCMs.

	PH99 Parameters		Climate Sensitivities		Feedback Response Times
	α [1/yrs]	μ [K/yrs]	ECS [K]	TCR [K]	$1/\alpha$ [yrs]
bcc_csm1_1_m	0.052	0.217	2.87	2.10	19.1
bcc_csm1_1	0.033	0.132	2.82	1.70	30.8
CanESM2	0.038	0.204	3.69	2.40	26.1
CCSM4	0.035	0.145	2.89	1.80	28.7
CNRM_CM5	0.038	0.177	3.25	2.10	26.5
CSIRO_Mk3_6_0	0.019	0.111	4.08	1.80	53.2
GISS_E2_R	0.048	0.147	2.11	1.50	20.8
HadGEM2_ES	0.027	0.177	4.59	2.50	37.4
IPSL_CM5A_LR	0.022	0.130	4.13	2.00	45.9
MIROC5	0.027	0.107	2.72	1.50	36.6
MIROC_ESM	0.021	0.140	4.67	2.20	48.0
MPI_ESM_LR	0.027	0.143	3.63	2.00	36.7
MRI_CGCM3	0.034	0.127	2.60	1.60	29.5
NorESM1_M	0.023	0.093	2.80	1.40	43.5
Multimodel Mean	**0.032**	**0.146**	**3.35**	**1.90**	**34.5**
Standard Deviation	0.010	0.036	0.792	0.342	10.350

Table 2:2: PH99 parameters (α and μ), climate sensitivities (ECS and TCR), and feedback response times ($1/\alpha$) after fitting PH99 GMT time series to AOGCM RCP2.6 GMT time series.

	PH99 Parameters		**Climate Sensitivities**		**Feedback Response Times**
	α [1/yrs]	μ [K/yrs]	ECS [K]	TCR [K]	$1/\alpha$ [yrs]
bcc_csm1_1_m	0.058	0.199	2.37	1.79	17.20
bcc_csm1_1	0.080	0.267	2.32	1.90	12.51
CanESM2	0.093	0.377	2.81	2.37	10.74
CCSM4	0.082	0.264	2.24	1.85	12.21
CNRM_CM5	0.084	0.329	2.73	2.26	11.97
CSIRO_Mk3_6_0	0.079	0.280	2.45	2.00	12.61
GISS_E2_R	0.345	0.746	1.50	1.44	2.90
HadGEM2_ES	0.114	0.485	2.94	2.57	8.75
IPSL_CM5A_LR	0.046	0.201	3.01	2.11	21.58
MIROC5	0.158	0.455	1.99	1.81	6.32
MIROC_ESM	0.096	0.478	3.45	2.93	10.41
MPI_ESM_LR	0.088	0.344	2.70	2.26	11.33
MRI_CGCM3	0.059	0.178	2.09	1.58	16.93
NorESM1_M	0.105	0.292	1.92	1.66	9.49
Multimodel Mean	**0.106**	**0.350**	**2.46**	**2.04**	**11.78**

Standard Deviation	0.074	0.152	0.512	0.409	4.639

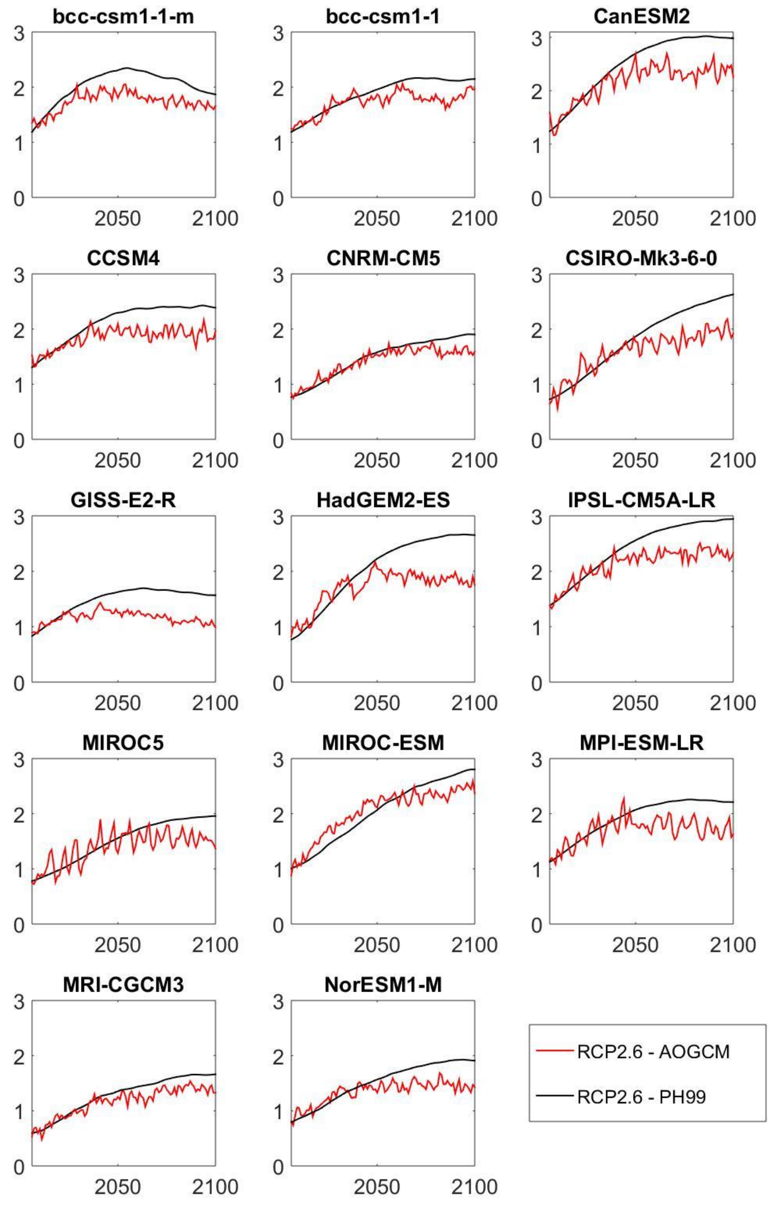
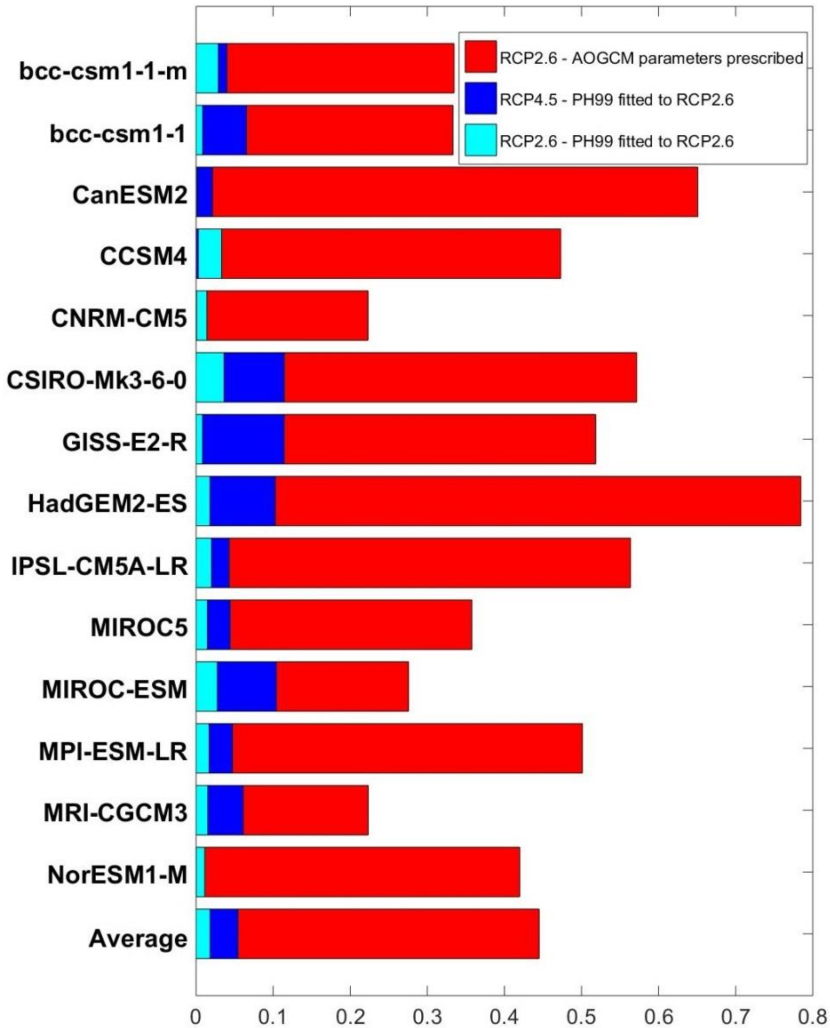


Figure 1: ~~The comparison~~Comparison of temperature paths projected by PH99 (black curve), calibrated by an AOGCM's ECS and TCR, to the corresponding AOGCM's temperature paths (red ~~curves~~curve). Deviations on the order of 0.5 K for 2100 are observed.


RCP2.6 - AOGCM parameters prescribed
RCP4.5 - PH99 fitted to RCP2.6
RCP2.6 - PH99 fitted to RCP2.6
bcc-csm1-1-m
bcc-csm1-1
CanESM2
CCSM4
CNRM-CM5
CSIRO-Mk3-6-0
GISS-E2-R
HadGEM2-ES
IPSL-CM5A-LR
MIROC5
MIROC-ESM
MPI-ESM-LR
MRI-CGCM3
NorESM1-M
Average
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

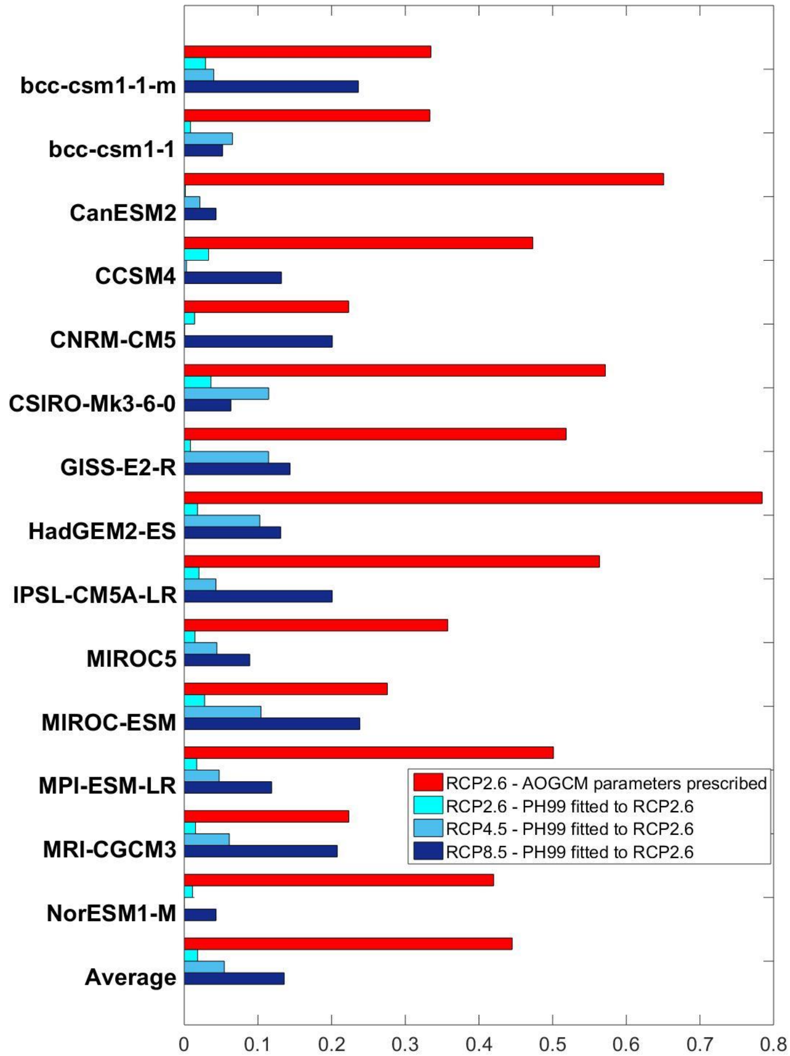


Figure 2: Modulus of deviations of ~~PH99 last 30 years~~ GMT mean values of PH99 over the period 2071-2100 from corresponding AOGCM means. The red bars show the deviations for RCP2.6 when α and μ are from Table 1 and not fitted. The cyan bars show the deviations in RCP2.6 when α and μ are fitted to ~~AOGCMs~~the AOGCM's RCP2.6 data. The light blue bars show the deviations for RCP4.5 when α and μ are kept ~~as before~~at their RCP2.6-fitted values (validation). The dark blue bars show the deviations for RCP8.5 when α and μ are kept at their RCP2.6 fitted values (validation).

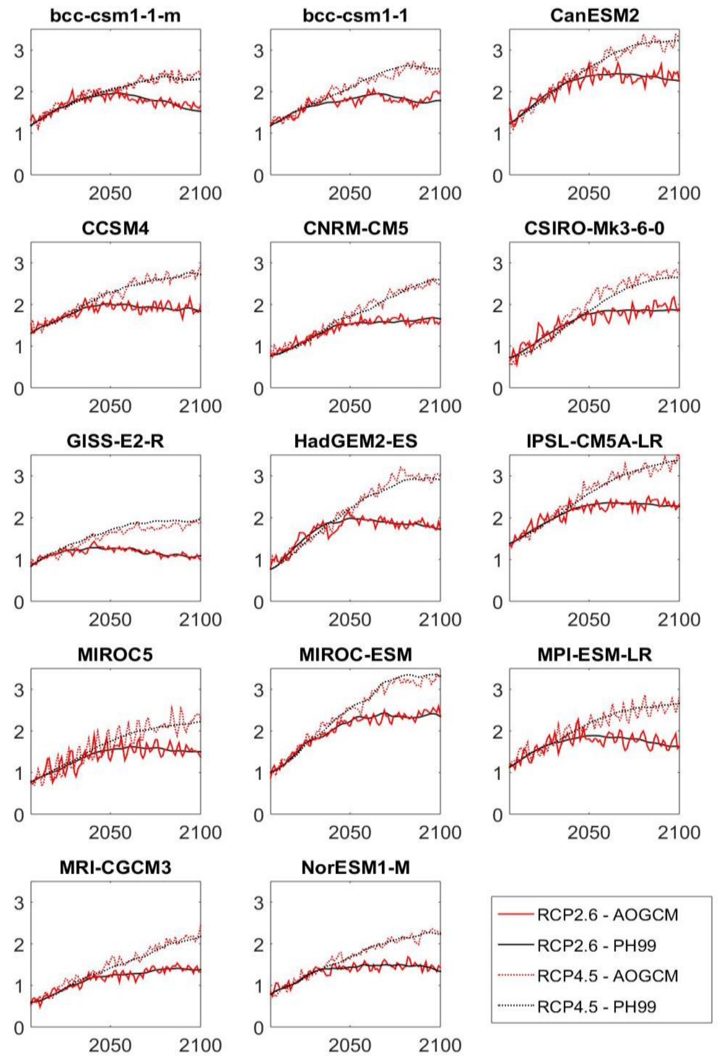
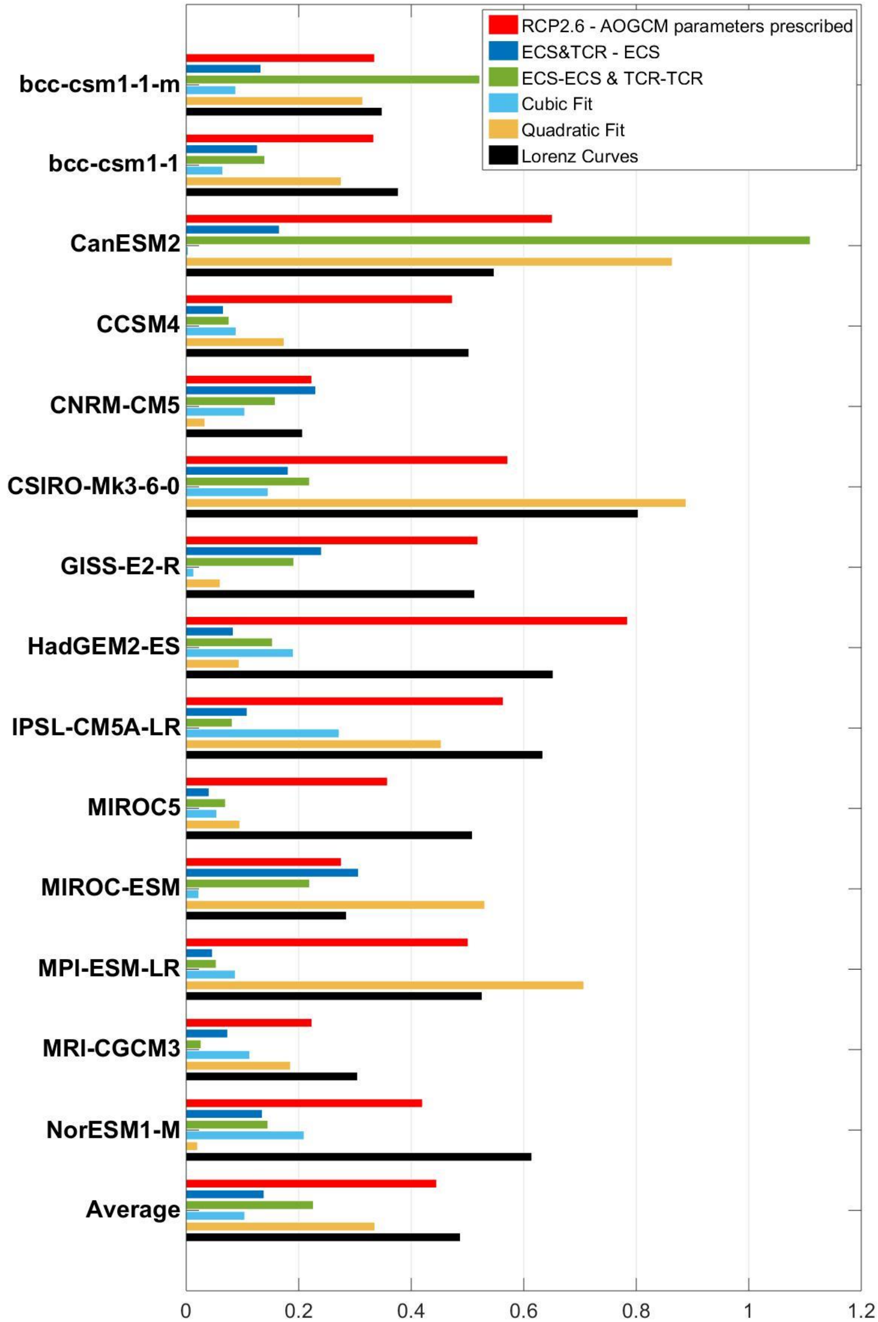


Figure 3: ~~The comparison~~Comparison of temperature evolutions projected by the climate module PH99 (~~black ~~solid and dotted black curves) to the actual AOGCM's temperature (~~red ~~solid and dotted red curves). ~~α~~α and μ ~~are~~have been tuned to fit the PH99 temperature path (solid black ~~solid ~~curve) to the respective ~~AOGCM~~AOGCM's RCP2.6 temperature ~~paths (red ~~path (solid red curve). Using the fitted ~~α~~α and μ, and taking the forcing reconstructed for RCP4.5 into account, PH99 also reproduces the projected RCP4.5 (~~black ~~dotted ~~curves). Red ~~black curve). The dotted ~~curves show~~red curve shows the actual RCP4.5 temperatures.~~ Hence the validation is successful.~~


RCP2.6 - AOGCM parameters prescribed
ECS&TCR - ECS
ECS-ECS & TCR-TCR
Cubic Fit
Quadratic Fit
Lorenz Curves
bcc-csm1-1-m
bcc-csm1-1
CanESM2
CCSM4
CNRM-CM5
CSIRO-Mk3-6-0
GISS-E2-R
HadGEM2-ES
IPSL-CM5A-LR
MIROC5
MIROC-ESM
MPI-ESM-LR
MRI-CGCM3
NorESM1-M
Average
0
0.2
0.4
0.6
0.8
1
1.2

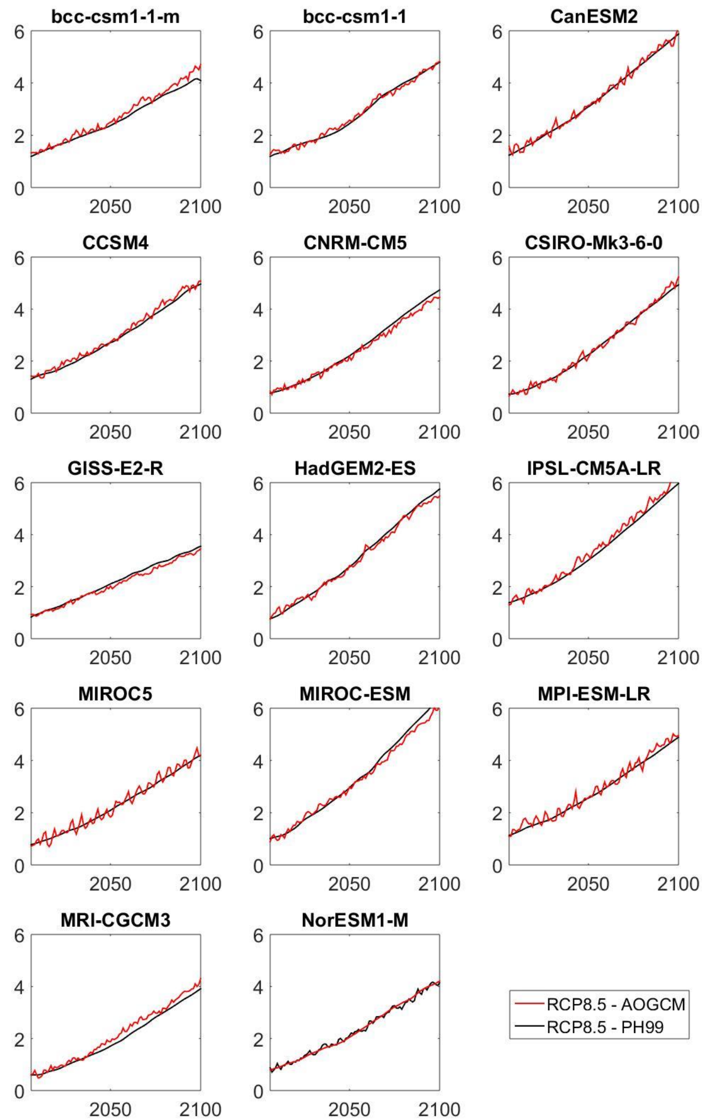


Figure 4: **Comparison of temperature evolutions projected by the climate module PH99 (black solid curves) in RCP8.5 scenario to the actual AOGCM's temperature (red solid curves) in RCP8.5 scenario. α and μ are taken from the second step, where PH99 is calibrated to RCP2.6 scenario.**

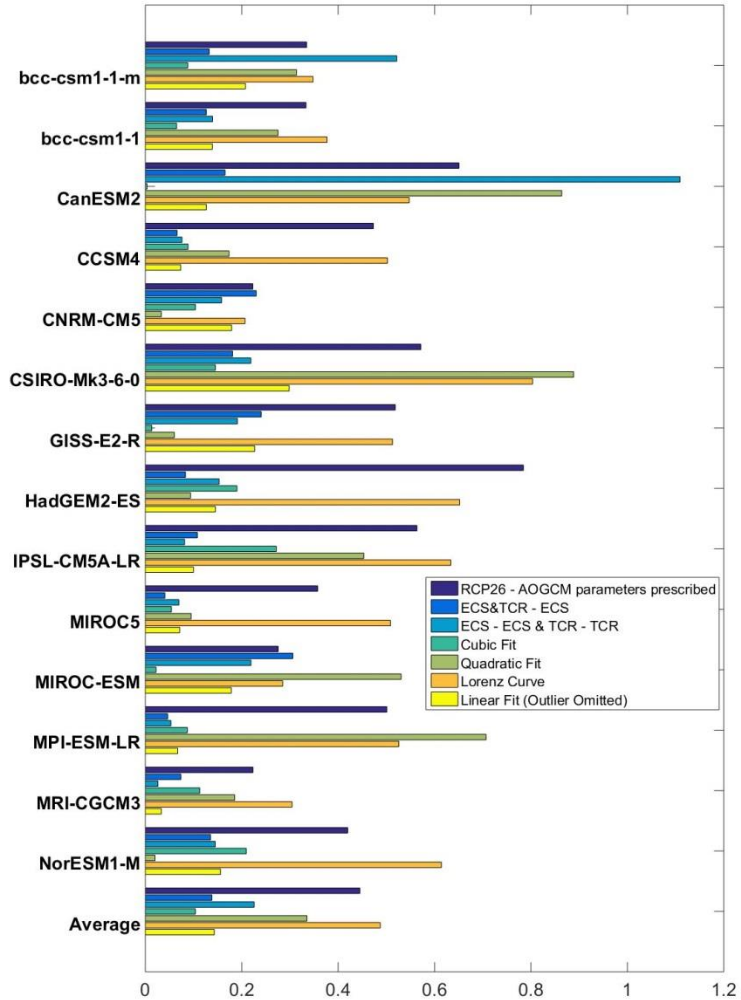
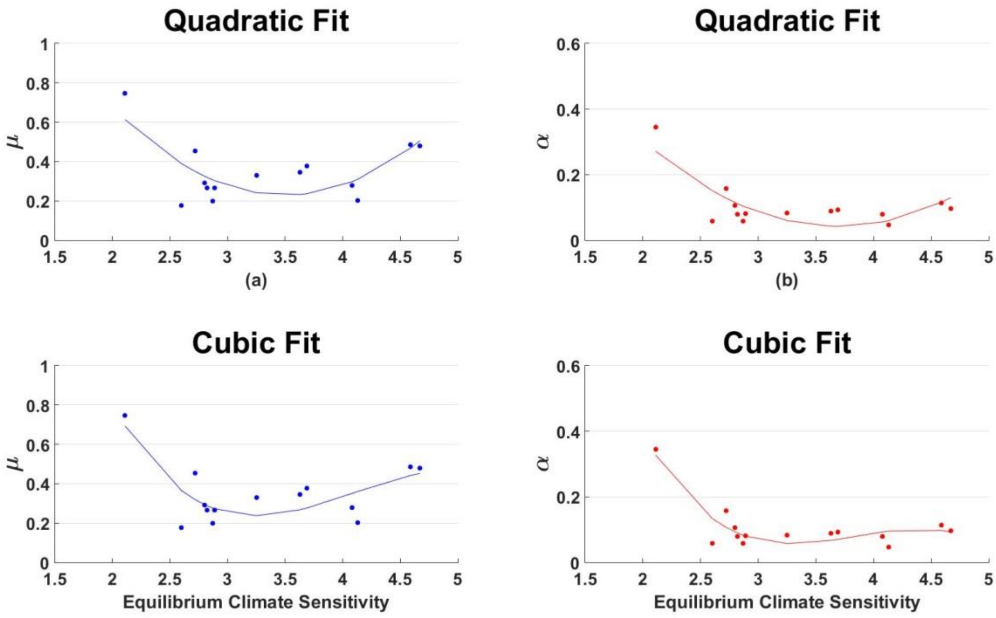


Figure 5: Modulus of mean temperature deviations over the ~~last 30 years~~period 2071-2100 (MTD) for PH99 ~~vs.~~from AOGCMs when α, μ, ECS, and TCR from Table 2 are related to ECS and TCR in Table 1. Using linear (yellow bars), quadratic (~~orange~~light green bars~~)~~), and cubic functions (~~cyan~~dark green bars), α and μ are related to ECS when outlier is put out for the linear case. Using linear fits, ECS and TCR are related to ECS (blue bars). Using linear fits, ECS and TCR are related to ECS and TCR respectively (~~green~~light blue bars). The ~~red~~dark blue bars show the deviations for RCP2.6 when α and μ are from Table 1 and not fitted (the same as Fig.2). The ~~black~~orange bars indicate MTD using Lorenz's curve.


Quadratic Fit
Quadratic Fit
Cubic Fit
Cubic Fit
μ
α
Equilibrium Climate Sensitivity
(a)
(b)

(c)

(d)

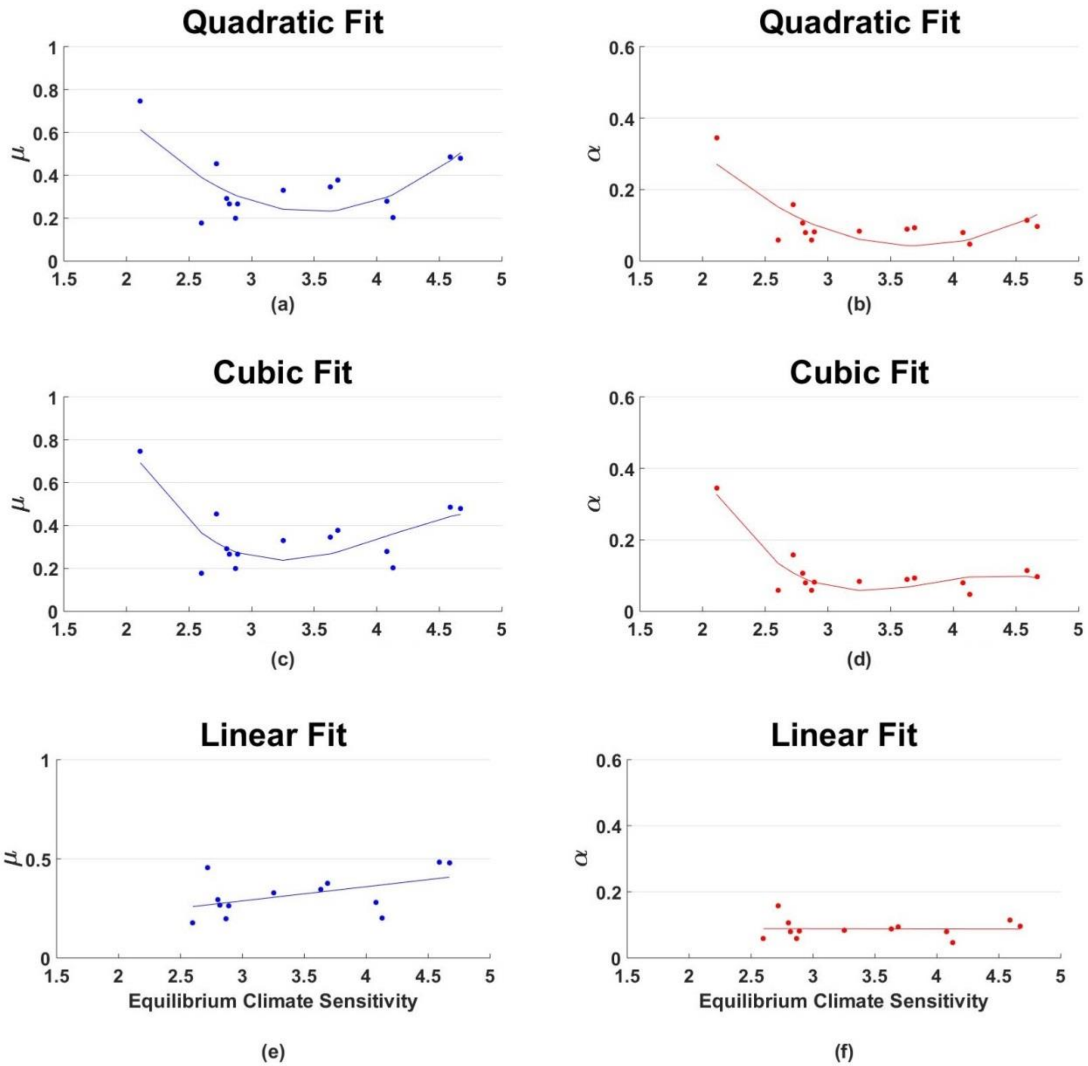


Figure 6: Quadratic (up~~)~~), cubic (middle), and ~~cubic~~linear (down) relationships of μ (left) and α (right) in Table 2 to ECS in Table 1. Notice that in the linear case the model GISS_E2_R (the upper left sample), as an outlier, is out.

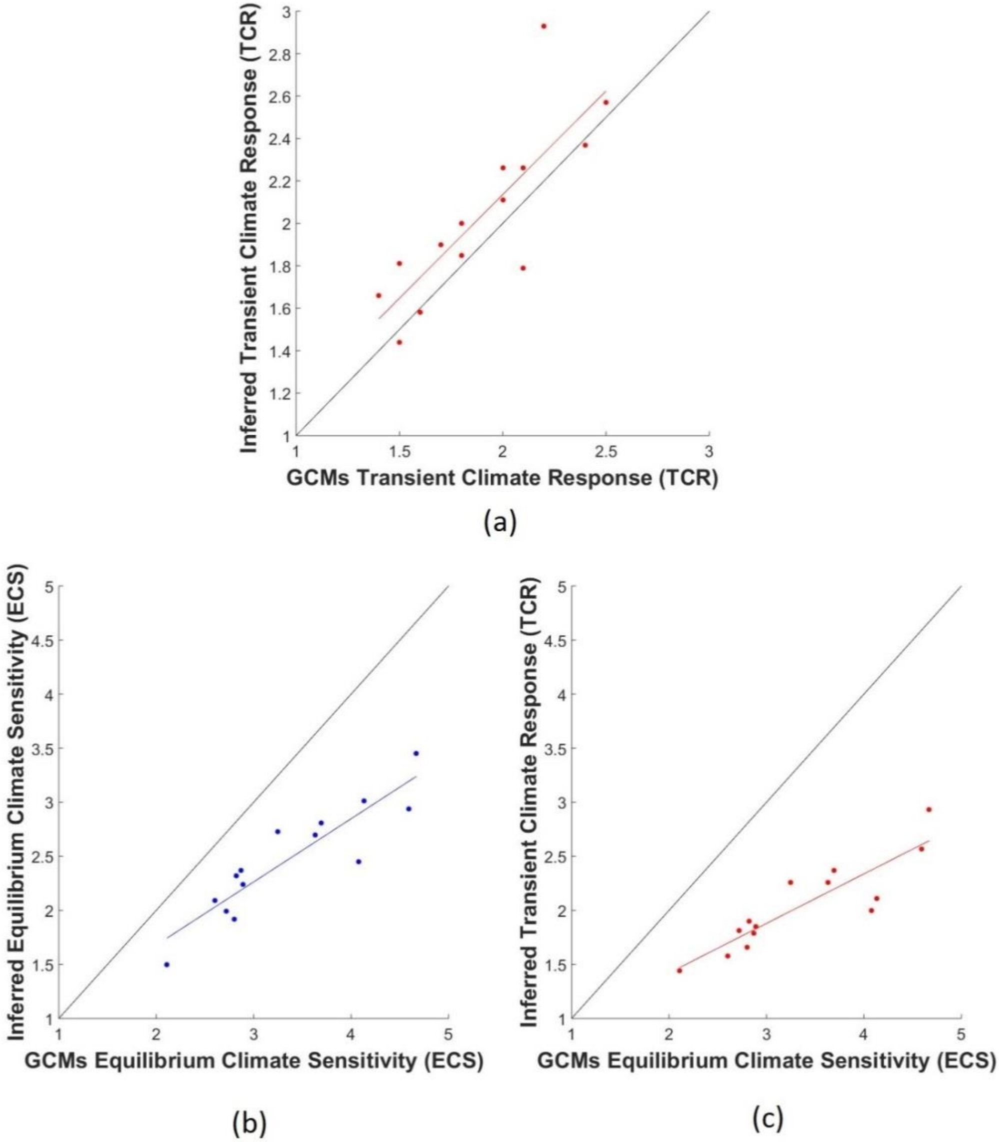
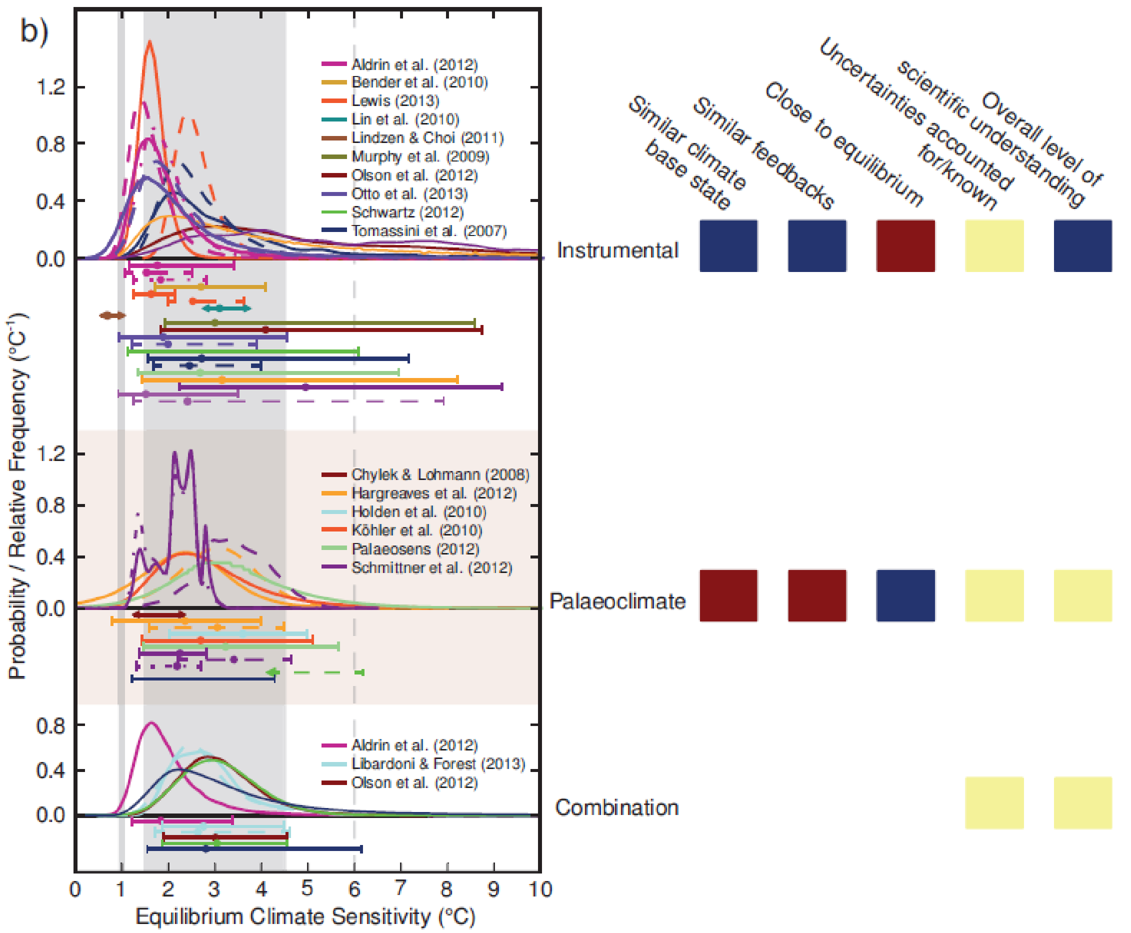


Figure 7: Inferred effective TCR vs. AOGCMs' TCR (a), inferred effective ECS vs. AOGCMs' ECS (b), and inferred effective TCR vs. AOGCMs' ECS (c). While the TCRs differ by less than 0.2 K, the ECSs differ by up to 2 K. This opens the door for a discussion as to whether PH99 should be calibrated ~~by~~using scenario~~-~~-class-adjusted effectively lower ECS values~~ of ECS~~.



b)
Aldrin et al. (2012)
Bender et al. (2010)
Lewis (2013)
Lin et al. (2010)
Lindzen & Choi (2011)
Murphy et al. (2009)
Olson et al. (2012)
Otto et al. (2013)
Schwartz (2012)
Tomassini et al. (2007)
Chylek & Lohmann (2008)
Hargreaves et al. (2012)
Holden et al. (2010)
Köhler et al. (2010)
Palaeosens (2012)
Schmittner et al. (2012)
Aldrin et al. (2012)
Libardoni & Forest (2013)
Olson et al. (2012)
Probability / Relative Frequency (°C⁻¹)
Equilibrium Climate Sensitivity (°C)
Similar climate base state
Similar feedbacks
Close to equilibrium
Uncertainties accounted for/known
Overall level of scientific understanding
Instrumental
Palaeoclimate
Combination

Figure 7:

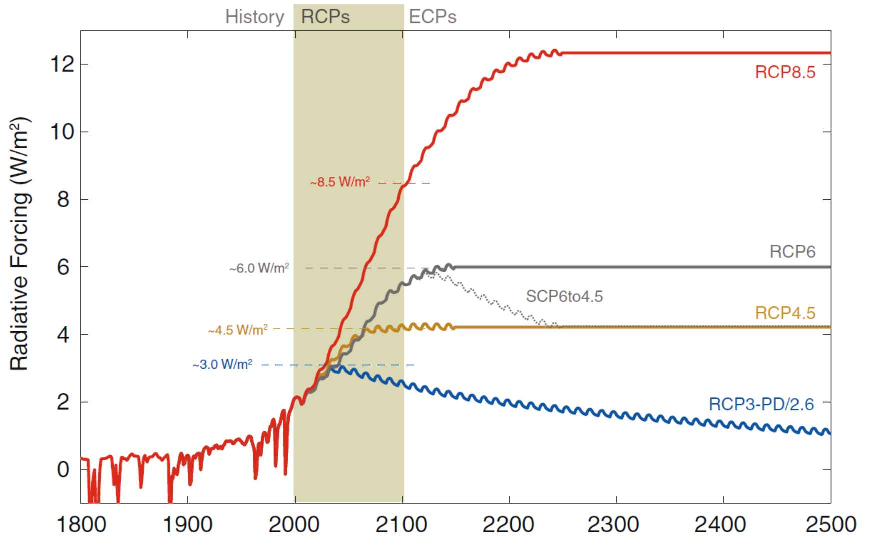


Figure 8: Total radiative forcing (anthropogenic plus natural) for RCPs – supporting the original names of the four pathways, as there is a close match between peaking, stabilization and 2100 levels for RCP2.6 (also called RCP3-PD), RCP4.5 & RCP6, and RCP8.5, respectively (taken from Meinshausen et al. (2011b)).

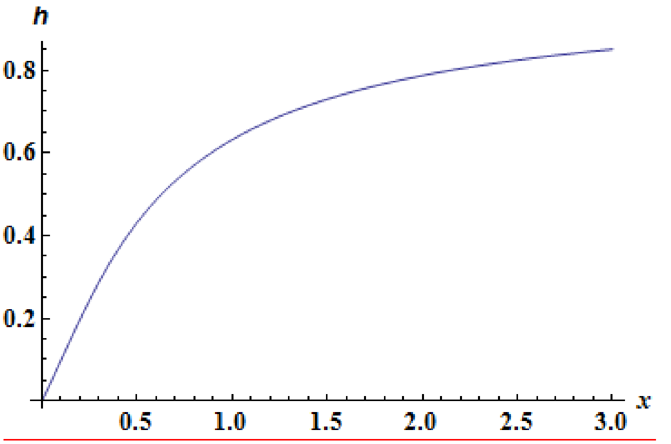


Figure 9: The auxiliary function *h*(*x*), which links the slow time scale of the 2-box model and the time scale of the 1-box model.

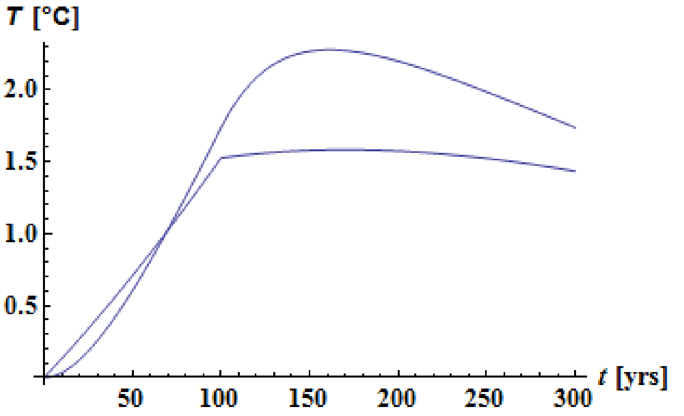


Figure 10: 1-box vs. 2-box model in response to kink-linear forcing as a stylized interpretation of mitigation-oriented forcing paths and for equal levels of ECS and TCR in both models. Kink-linear curve: 2-box model, smooth curve: 1-box model. The temperature development of the 1-box model overshoots the maximum of the 2-box model by roughly 50%.

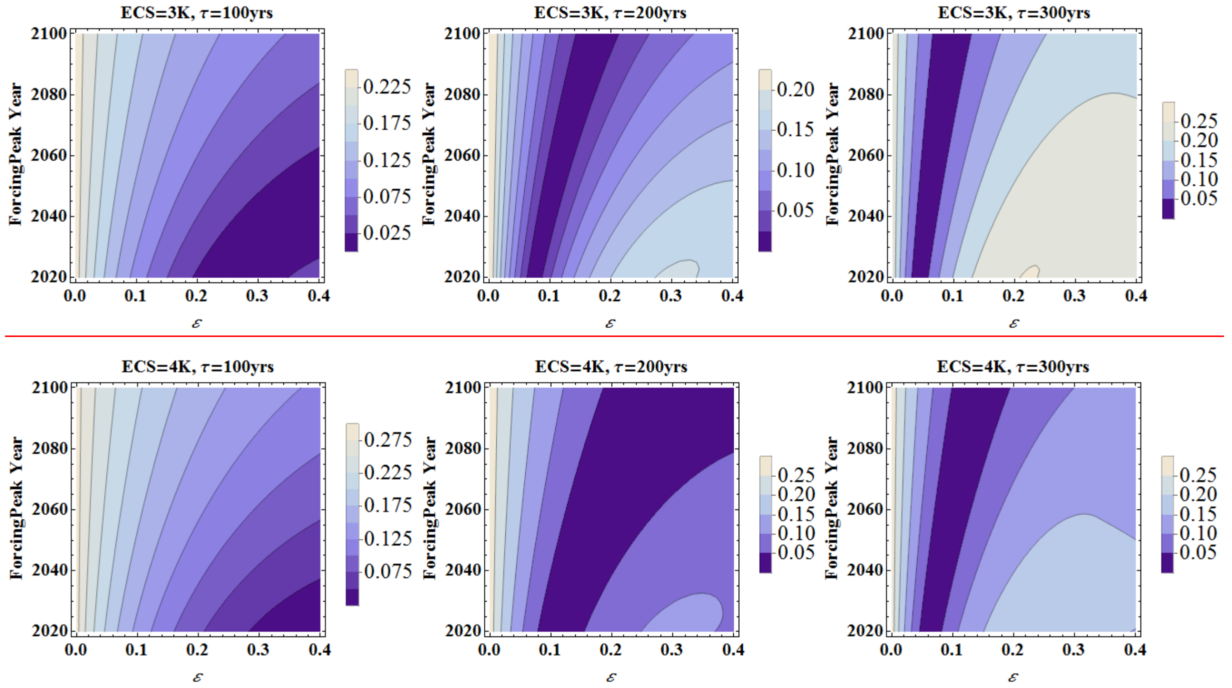


Figure 11: : Comparing GMT maxima of the 2-box model and the 1-box model, the latter being adjusted to the former by prescribing the linearly transformed ECS and TCR according to the scheme ETE. Abscissa: ε, ordinate: changed peaking year t_1, however transformed to years, for the 2-box ECS of 3 K and 4 K, and τ=100, 200, 300 yrs, respectively. The relative error (max. GMT difference normalized by the max. GMT of the 2-box model) is markedly smaller than for the case of prior adjustment.

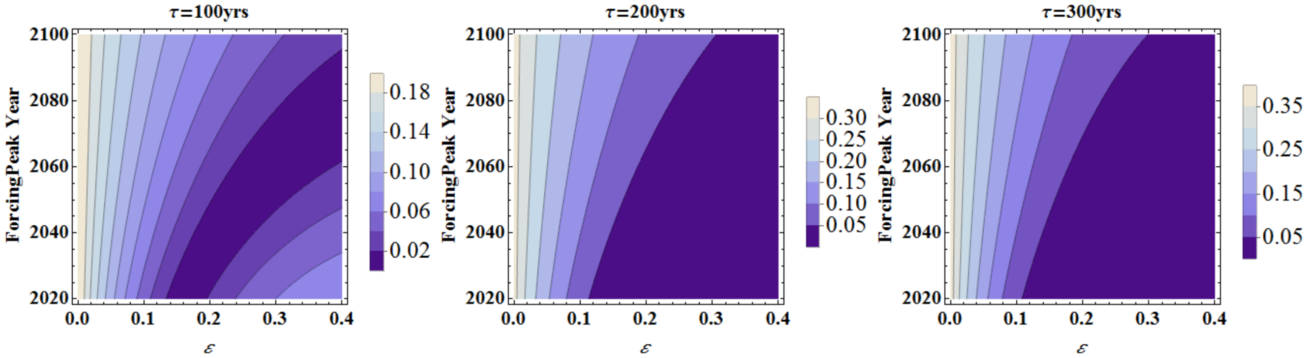


Figure 12: Similar to the previous figure (relative max. GMT error with abscissa: ε, ordinate: t_1 [yrs],), however for a further adjusted ECS of the 1-box model, such that perfect matching is achieved for t_1=100 yrs, ε=0.2, and a 1-box time scale of 12 yrs. For most of the parameter settings, the relative error is below 10%.

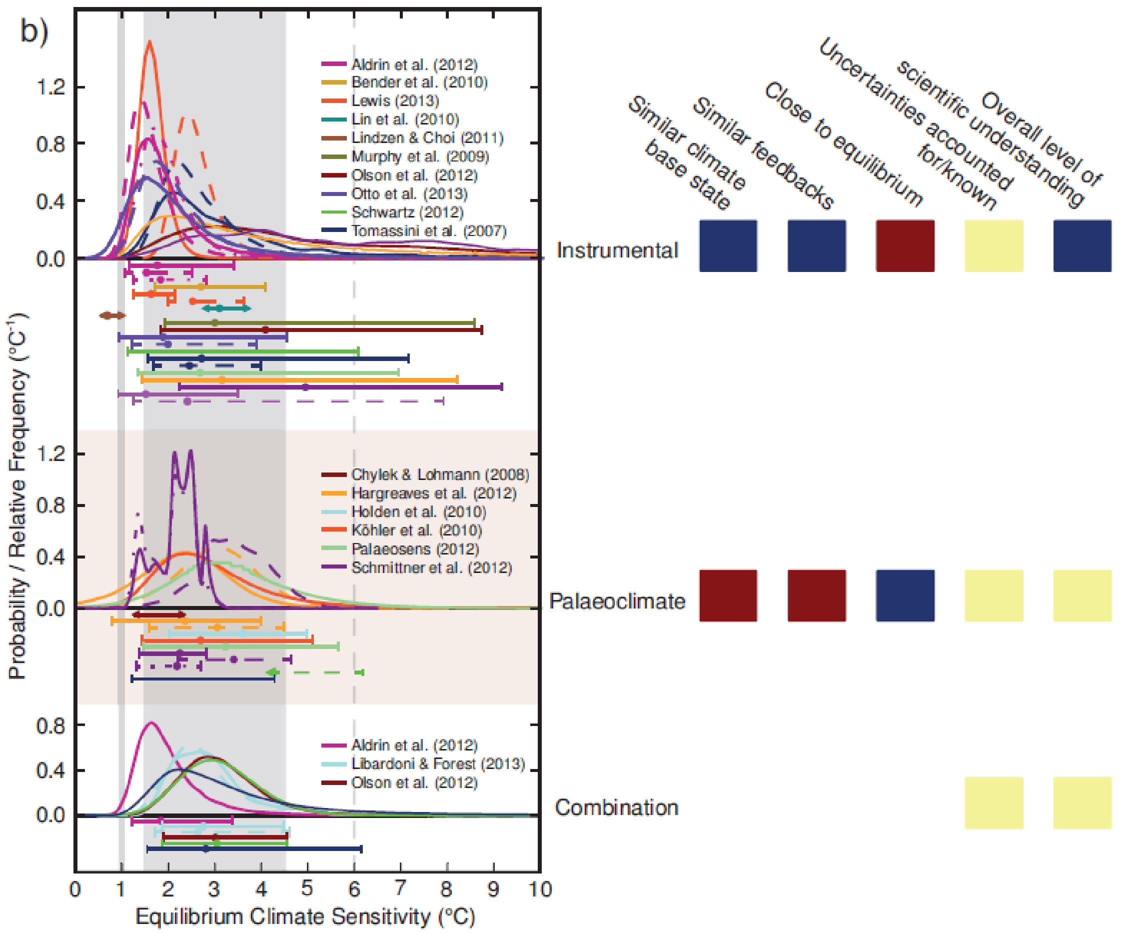


Figure 13: Probability density distributions of ECS according to IPCC AR5 WG-I (Bindoff et al., 2013, Fig. 10.20).

Figure 14: The comparison of temperature evolutions projected by the climate module PH99 (black solid curves) in RCP6.0 scenario to the actual AOGCM's temperature (red solid curves) in RCP6.0 scenario. α and μ are taken from the second step, where PH99 is calibrated to RCP2.6 scenario

Table 3: Modulus of mean temperature deviations over the period 2071-2100 (MTD) for PH99 from corresponding AOGCM. In the first 4 columns, PH99 is calibrated to RCP 2.6. In the second 4 columns, PH99 is calibrated to RCP 4.5.

	Calibrated to RCP 2.6				**Calibrated to RCP 4.5**		
	MTD RCP2.6	MTD RCP4.5	MTD RCP6.0	MTD RCP8.5	MTD RCP2.6	**MTD RCP4.5**	MTD RCP6.0
bcc_csm1_1_m	**0.029**	0.040		0.236	0.018	**0.007**	
bcc_csm1_1	**0.009**	0.066		0.052	0.064	**0.021**	
CanESM2	**0.001**	0.021		0.043	0.039	**0.003**	
CCSM4	**0.033**	0.003	0.069	0.132	0.024	**0.005**	0.064
CNRM_CM5	**0.014**	0.001		0.201	0.005	**0.012**	
CSIRO_Mk3_6_0	**0.036**	0.115	0.040	0.063	0.017	**0.015**	0.168
GISS_E2_R	**0.008**	0.114	0.094	0.144	0.064	**0.003**	0.027
HadGEM2_ES	**0.018**	0.103	0.036	0.131	0.057	**0.020**	0.097
IPSL_CM5A_LR	**0.020**	0.043	0.050	0.201	0.121	**0.013**	0.017
MIROC5	**0.015**	0.044	0.029	0.089	0.032	**0.009**	0.034
MIROC_ESM	**0.028**	0.104	0.079	0.238	0.140	**0.012**	0.044
MPI_ESM_LR	**0.017**	0.047		0.119	0.108	**0.015**	
MRI_CGCM3	**0.015**	0.061	0.083	0.208	0.001	**0.007**	0.004
NorESM1_M	**0.011**	0.000	0.026	0.043	0.024	**0.000**	0.006
Multimodel Mean	**0.018**	0.054	0.056	0.136	0.035	**0.005**	0.039
Standard Deviation	**0.010**	0.041	0.025	0.071	0.044	**0.006**	0.053

Table 3 (continued): Modulus of mean temperature deviations over the period 2071-2100 (MTD) for PH99 from corresponding AOGCM. In the third 4 columns, PH99 is calibrated to RCP 6.0. In the fourth 4 columns, PH99 is calibrated to RCP 8.5.

	Calibrated to RCP 6.0				**Calibrated to RCP 8.5**		
	MTD RCP2.6	MTD RCP4.5	**MTD RCP6.0**	MTD RCP8.5	MTD RCP2.6	MTD RCP4.5	MTD RCP6.0
bcc_csm1_1_m					0.287	0.257	
bcc_csm1_1					0.091	0.008	
CanESM2					0.008	0.025	
CCSM4	0.038	0.067	**0.018**	0.086	0.059	0.004	0.010

CNRM_CM5					0.117	0.151	
CSIRO_Mk3_6_0	0.161	0.199	**0.019**	0.062	0.119	0.019	0.034
GISS_E2_R	0.041	0.037	**0.019**	0.046	0.045	0.023	0.011
HadGEM2_ES	0.146	0.233	**0.021**	0.063	0.146	0.252	0.073
IPSL_CM5A_LR	0.016	0.077	**0.001**	0.095	0.052	0.078	0.030
MIROC5	0.067	0.079	**0.011**	0.032	0.025	0.006	0.019
MIROC_ESM	0.187	0.070	**0.005**	0.198	0.309	0.235	0.140
MPI_ESM_LR					0.011	0.082	
MRI_CGCM3	0.092	0.068	**0.003**	0.042	0.008	0.014	0.055
NorESM1_M	0.068	0.021	**0.016**	0.136	0.070	0.055	0.054
Multimodel Mean	0.091	0.095	**0.007**	0.084	0.096	0.086	0.029
Standard Deviation	0.060	0.072	**0.008**	0.053	0.096	0.096	0.041