On the Future Role of the 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. In the following, we test the validity of a 1-box climate model as an emulator for Atmosphere-Ocean General Circulation Models (AOGCMs). The 1-box climate model is currently employed in the integrated assessment models FUND, MIND and PAGE. For our assessment, we 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 1-box model, global mean temperature (GMT) projections are generically too 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 needs and 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 1-box model offers a good emulator of these AOGCMs, provided their ECS and TCR values are universally mapped onto effective 1-box counterparts and a certain time horizon is not exceeded. We contend that this 1-box model could be used in future integrated assessment within certain limits, in particular when computationally demanding decision-making under climate response uncertainty needs to be modelled. However, then the 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 as being influenced by a larger ECS than claimed.

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

1 Introduction

Climate-economy integrated assessment models (IAMs) are 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 the most sophisticated climate models, Atmosphere-Ocean General Circulation Models (AOGCMs). These climate modules (hereafter: 'simple climate models' (SCMs)) foster computational efficiency and hence allow researchers to project a broader set of scenarios in orders of

magnitude less time. For IAMs based on a decision-analytic framework involving intertemporal welfare optimization, SCMs are in fact indispensable, as these IAMs' numerical solvers would need to access the climate module anywhere from ten thousand to one hundred thousand times before numerical convergence were flagged.

Currently the need to qualify the degree of accuracy with which SCMs mimic AOGCMs or properly represent ensembles of AOGCMs is increasingly being recognized (Calel & Stainforth, 2017; van Vuuren et al., 2011a), as this aspect might have immediate monetary consequences in connection with derived policy scenarios (Calel & Stainforth, 2017). Van Vuuren et al. (2011a) found that IAMs tend 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 AOGCMs' behavior 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., 2011a) represents the most complex SCM currently in use. In some sense one could even call MAGICC an Earth System Model of Intermediate Complexity. It has demonstrated its capacity to emulate all AOGCMs' GMT even more precisely than the standard deviation of interannual GMT variability (Meinshausen et al., 2011a), with a fixed set of parameters, utilized for the whole range of RCPs (representative concentration pathways, van Vuuren et al., 2011b). This represents the current gold standard of AOGCM emulation using SCMs.

In the following, we address the most extreme opposite end of the scale of complexity within the model category of SCMs: the 1-box model as introduced by Petschel-Held et al. (1999) (hereafter: 'PH99'). 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 of this model as assessed in the literature is as follows: by fitting PH99 to GMT time series, it can be used as a diagnostic instrument, as Andrews & Allen (2008) have done. However its main application is as an emulator of AOGCMs. In conjunction with the most parsimonious carbon cycle model (described in Petschel-Held et al. (1999) as well), PH99 has been used to derive 'admissible' greenhouse gas emission scenarios in view of prescribed GMT targets (Bruckner et al., 2003; Kriegler & Bruckner, 2004). Furthermore, the following climate-economic IAMs are currently utilizing PH99: FUND (Anthoff & Tol, 2014), 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., 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 less than 0.2 K for the 21st century (see their Fig.5).

We believe that further validation is both necessary and possible 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 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 additional assumption that non-CO2 greenhouse gas forcing approximately balances aerosol cooling. Here we employ recently diagnosed forcings for 14 CMIP5-AOGCMs by Forster et al. (2013). Third and lastly, we find current practices – directly prescribing equilibrium climate sensitivity (ECS) (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 to calibrate PH99 (see e.g. Anthoff & Tol (2014)) – 'inadequate' in the context of 2° stabilization scenarios. 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 calibrating PH99 by mapping these climate system properties to respective effective values, which are suitable for a centennial time horizon (but likely not beyond it), before using them in PH99.

In doing so, we comply with a need that Calel & Stainforth (2017) recently identified: to achieve the application-specific re-calibration of PH99 as a valid future approach to emulation. In this way, PH99 could complement the use of increasingly complex climate modules, ranging from DICE's 2-box model (Nordhaus, 2013) to the complex upwelling-diffusion climate module used in MAGICC (Meinshausen et al. (2011a)). The potential benefits of doing so are two-fold: firstly, the most parsimonious SCM, PH99, ensures maximum transparency. Secondly, in the context of numerically solving decision-making under climate response uncertainty (Kunreuther et al., 2014), having to simultaneously deal with dozens, hundreds or even thousands of alternate climate 'states of the world' (the economist's term for the uncertain system property) poses a significant challenge for numerical solvers and memory. In 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.

This article is organized as follows. Section 2 introduces the data-based part of our analysis. We call for a 3-step procedure, including: (i) a conventional, though not naïve calibration of PH99 with regard to climate sensitivity and transient climate response (i.e. the GMT change in response to a 1%/yr. increase in the CO2 concentration until doubling compared to the pre-industrial value); (ii) an AOGCM-specific calibration; and (iii) the validation of the former. In 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 independent scenarios. Note that, in Sect. 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. This results in a larger emulation

error than achieved in Section 3, but one that would nevertheless suffice for most applications. In 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 findings for the integrated assessment community, 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

Among others, one of the most extensively used most parsimonious climate emulators is the 1-box global energy balance model, Eq. (1), introduced by Petschel-Held et al. (1999), which projects the atmospheric GMT anomaly compared to its preindustrial level. For a CO2-only forcing scenario, PH99 reads

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

Here T denotes the GMT anomaly, c is the CO2 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 CO2 concentration compared to its pre-industrial value:

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

also in line with Petschel-Held et al. (1999) and Kriegler and Bruckner (2004). In Appendix 1 we briefly derive 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:

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

(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

We first check whether simply calibrating PH99 from AOGCM-specific ECS and TCR data would deliver good emulations for 2°-target-compatible scenarios. After a technical derivation, we summarize this method of mapping AOGCMs' ECS and TCR onto PH99's two parameters.

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 (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:

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

Hence the CO2-carrying summand would become the CO2 forcing and could now be generalized to the total forcing *F*. If we then divide by the still-to-be-determined factor *h*, we obtain:

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

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 CO2-only case of Eq. (4):

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

Q_2 denotes the additional forcing from the doubling of the CO2 concentration compared to its pre-industrial value and is listed for any of the above AOGCMs (see Forster et al., 2013, Table 1). When comparing Eq. (1) and Eq. (6), we obtain:

$$\mu = \frac{Q_2}{h \ln 2} \quad (7)$$

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

The derivation displayed so far can be summarized in terms of the 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 pre-industrial value. To do this, for each AOGCM we calculate the mean 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 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 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. Hence we use the difference between the respective 2071-2100 GMT time averages of PH99 and the AOGCM as an error metric. If the deviations are tolerable, the climate module is validated; if they are intolerable, we must proceed with steps two and three.

2.2 Step Two

For each AOGCM, α and μ are tuned such that the difference between PH99 and the AOGCM GMT anomaly for the RCP2.6 scenario in the period 2006-2100 is minimized 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 2071-2100 means of GMT between PH99 and the respective 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. (2011a) subtracted (low-pass filtered) control runs. We avoid 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. (2011a) low-pass filtered both with a cut-off frequency of 1/20 yrs. In the interest of parsimonious analysis we avoided low-pass filtering here, as the 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

Lastly, we validate the PH99 model versions generated in step two. For this purpose, independent temperature and forcing paths 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 and RCP8.5 scenarios. Similar to steps one and two, the deviations in 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

shows the calculated α and μ together with the feedback response time 1/α 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 a minimum and maximum of 2.11 K and 4.67 K, respectively. The mean value of the time scales is roughly 35 years.

Figure 1 represents the projected PH99 temperature evolution for the scenario RCP2.6 of each GCM in 2006-2100, using the data from Table 1 and the RCP2.6s' forcings. PH99 clearly overestimates the temperature anomaly for all GCMs, especially over the last 30 years. The absolute values of the deviations of mean temperature over the last 30 years (hereafter: MTD) from the AOGCM data are shown in Figure 2. The MTD ranges from 0.22 K for MRI-CGCM3 to approximately 0.79 K for HadGEM2-ES. On average, the deviations are ca. 0.45 K. This is clearly a large error, both in units of annual GMT standard deviation as well as the climate policy dimension. A proclaimed goal of the 2015 Paris Agreement (UNFCCC, 2016) consists 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 for the whole period 2006-2100 are minimized in a least squares manner as represented in Figure 3 and Figure 4. From the thereby adjusted α and μ we derive the ECS and TCR, which are presented in Table 2. MTDs for the various AOGCMs are shown in Figure 2.

The results 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 ECS decreases by 0.9 K (from 3.35 K to 2.46 K), the average 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 roughly 35 years to less than 12 years.

For validation we move on to step three. In this regard, we utilize the RCP4.5 temperature and forcing data as provided by Forster et al. (2013). In Figure 3 and Figure 4 the respective GMT trajectories for any AOGCM are contrasted with the PH99-generated ones, whereby α and μ are fixed to their values as determined in step two. The MTDs are shown in Figure 2. The results confirm that the climate module is so well trained in the second step that it can suitably mimic the actual temperatures for RCP4.5 and RCP8.5. As shown, the average MTD is approximately 0.05 K for RCP4.5 and about 0.14 for RCP8.5. For RCP4.5, the deviations for three of the GCMs, namely CCSM4, CNRM-CM5 and NorESM1-M, are even better than those diagnosed for RCP2.6 in step two. See Appendix 2 for further analyses.

4 A mapping of 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 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 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/μ and ECS, Lorenz et al. (2012) suggest the following approximation:

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

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

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

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

Figure 6 relates α and μ (from Table 2) to the ECS (from Table 1), using linear, quadratic and cubic polynomial approximations. For the case of linear approximation, we put the model GISS_E2_R out as an outlier. Figure 5 indicates that on average 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 0.3 K for IPSL-CM5A-LR, which is roughly a third of the maximum in the quadratic approximation that is revealed for CSIRO-Mk3-6-0.

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

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

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

and

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

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

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

$$TCR_{\mathrm{PH99}} \approx m\, ECS_{\mathrm{AOGCM}} + n \tag{11}$$

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

The respective MTDs are shown in 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) clearly offers the better overall approximation.

Using the ETE has four major advantages over all other options dealt with here, especially for the IAM community. Firstly, its approximation is better than all options but the cubic fit. Even though the cubic fit may yield a better approximation, in our analysis it is only better by 0.03 K 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 the cubic fit because one can easily use a broader range of climate sensitivities, for example, from 1 K to 9 K, which may not be accurately determined by the cubic fit. Thirdly, the ETE not only yields a better approximation; prior knowledge regarding the TCR is 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.

$$\begin{aligned} C\ \frac{\mathrm{d}T_{2\mathrm{B}}}{\mathrm{d}t} &= F - \lambda T_{2\mathrm{B}} - \gamma(T_{2\mathrm{B}} - T_0) \\ C_0\ \frac{\mathrm{d}T_0}{\mathrm{d}t} &= \qquad\qquad \gamma(T_{2\mathrm{B}} - T_0) \end{aligned} \tag{12}$$

$T_{2\mathrm{B}}$ 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}= (ln2)/γ≈70yrs 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)

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

Then the defining condition for θ, after some manipulation, can be written as $h(\theta / t_{\mathrm{TCR}})$ = (1/2) $h(\tau / t_{\mathrm{TCR}})$. 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 $h(\theta/t_{\mathrm{TCR}})$ = (1/2) $h(\tau/t_{\mathrm{TCR}})$.

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:

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

hereby $\lambda_1=\lambda\, ECS_{2\mathrm{B}}/ECS_{1\mathrm{B}}$ 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

The previous section offers a key mechanism to explain why, for given ECS and TCR, GMT scenarios generated by PH99 are biased towards higher temperatures. One should not forget about potential additional mechanisms. 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 undertaking (see e.g. Meinshausen et al., 2011a, for a discussion). A bias to the high end here would also result in 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.

If 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 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 in fact been utilized in PH99 for the MIND model since Lorenz et al. (2012), one could ask for the back-transformed ECS values according to our Figure 7. The respective values are 1.2 K, 3.6 K and 9.0 K. From Figure 13, which reflects IPCC AR5's synopsis of current knowledge regarding 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 erroneous, but simply need to be re-interpreted as rather high-end representatives within the collection of ranges as seen in IPCC AR5.

7 Summary and Conclusion

We utilize recent data on total radiative forcing (Forster et al., 2013) from 14 state-of-the-art CMIP5 Atmosphere Ocean General Circulation Models (AOGCMs) in order to test the validity of the 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 PAGE.

We find that when prescribing the equilibrium climate sensitivity (ECS) and transient climate response (TCR) of these AOGCMs to the emulator PH99, global mean temperature (GMT) is generically projected 0.5 K higher. In contrast, by directly fitting PH99 to the RCP2.6 time series and validating 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 numerically demonstrate that PH99 can be used to excellently emulate AOGCMs within integrated assessment of the 2° target, provided its ECS and TCR are re-interpreted as effective, scenario-class-specific values and mapped from original ECS and TCR values. We suggest a first version of such a mapping.

Furthermore we explain the observed discrepances 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 PAGE, may need to be re-interpreted in the sense that higher ECS values had effectively been used. Even when having dealt with distributions of ECS as for the MIND model, ECS values re-interpreted in this manner are still within the range outlined by IPCC AR5 (see Figure 13).

For future work, we propose the following steps: (i) 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. (ii) Future discussions with the AOGCM community should illuminate to what extent the further explanations we suggested might also apply, thereby potentially reducing the need to correct for PH99. (iii) An AOGCM-independent, yet scenario-class-specific 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 one based on observational data and mediated by more complex SCMs.

In summary, PH99 could continue to be 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 of TCR in PH99

We recapitulate Eq. (1) as

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

TCR is defined as the temperature change in response to a 1%/yr. increase in CO2 concentration, starting from preindustrial conditions. Hence the concentration, expressed in units of the pre-industrial concentration, reads

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

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 is to vanish, its solution reads

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

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 find the same result using an expression provided in Andrews & Allen, 2008, when we plug in our expression for t_2 into 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. performed the statistical analysis. H.H. provided the analytic analysis. M.M.K. suggested and developed the alternative scheme. Both participated in the writing of the article.

Competing Interests

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to thank J. Marotzke for 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 making the authors aware of Geoffroy et al. (2013), which discusses negligible AOGCM drift. The authors are also grateful to E. Roshan for her help with the visualizations and providing quantiles of Wigley's & Raper's (2001) distribution on ECS. We thank M. Fentem for proofreading the second version of our manuscript from a native speaker's perspective. M.M.K. 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, 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.

Calel, R. and Stainforth, D. A.: On the Physics of three Integrated Assessment Models, Bull. Amer. Meteor. Soc., 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, 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.

Roshan, E., Khabbazan, M. M., Held, H., Cost-Risk Trade-off of Mitigation and Solar Geoengineering – Considering Regional Disparities under Probabilistic Climate Sensitivity, 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.

Stern, N.: The Stern Review – The Economics of Climate Change, Cambridge, 2007.

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., 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, 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: 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: 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

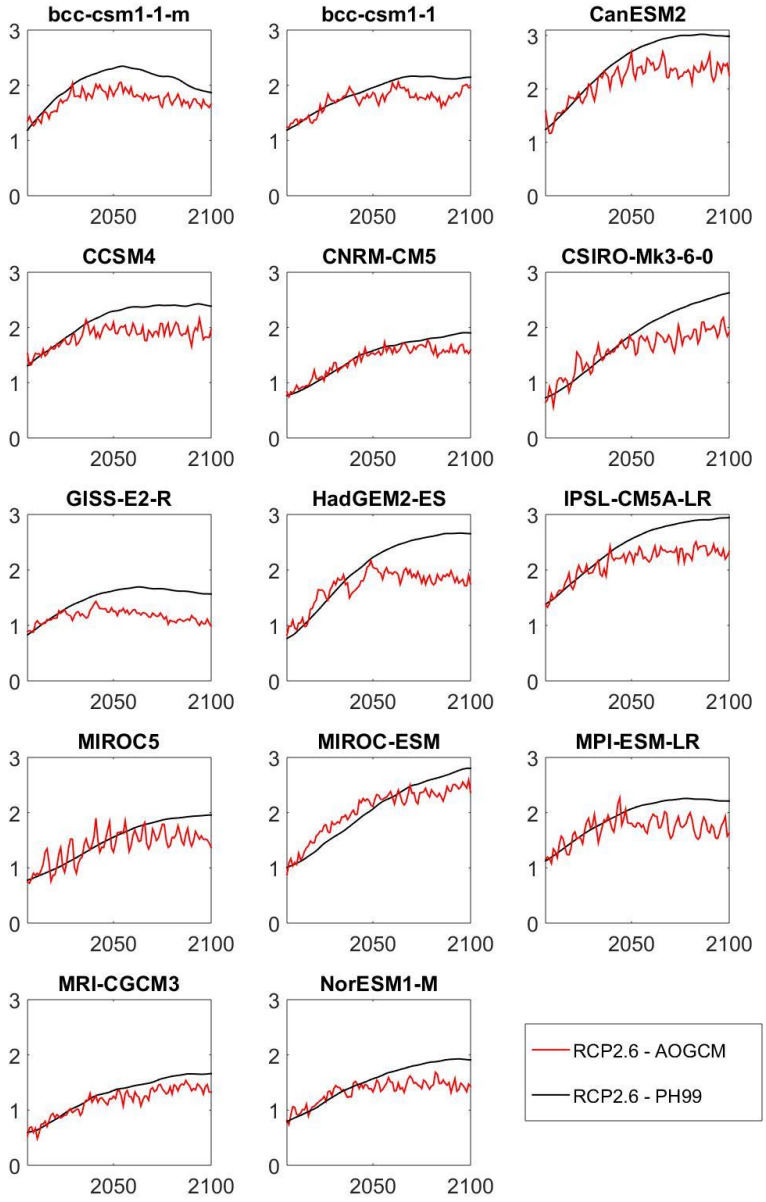

Figure 1: Comparison of temperature paths projected by PH99 (black curve), calibrated by an AOGCM's ECS and TCR, to the corresponding AOGCM's temperature paths (red curve). Deviations on the order of 0.5 K for 2100 are observed.

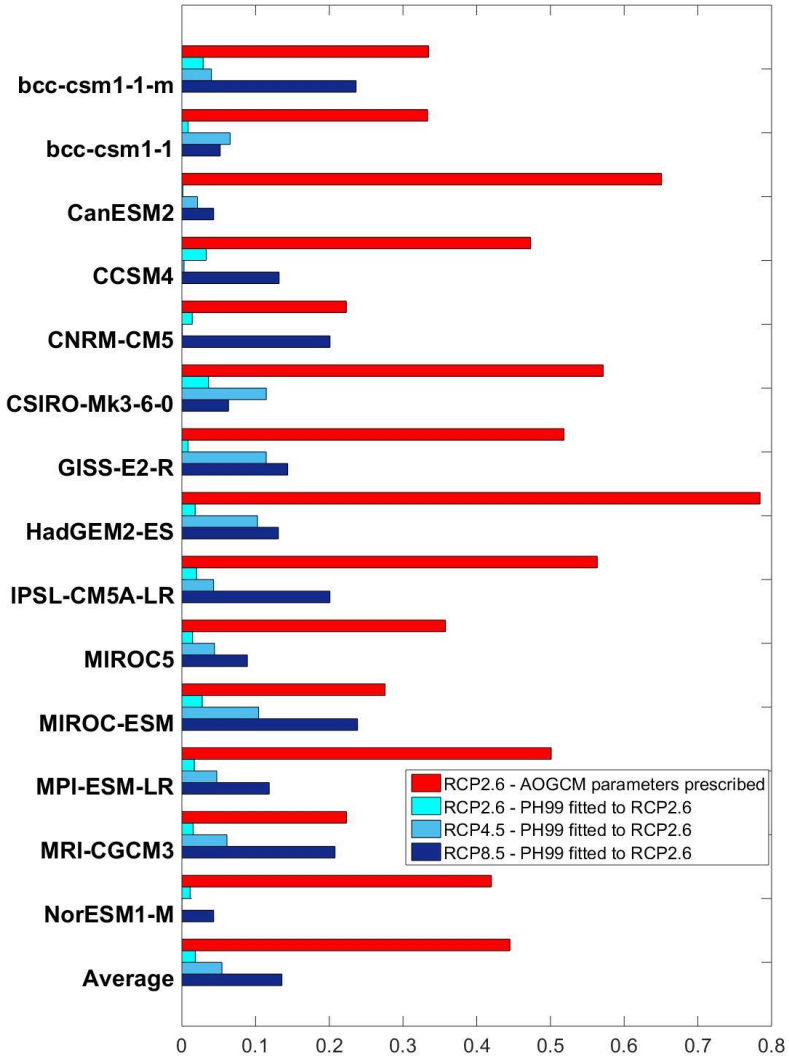

Figure 2: Modulus of deviations of GMT mean values of PH99 over the period 2071-2100 from corresponding AOGCM means. The red bars show the deviations for RCP2.6 when α and μ are from Table 1 and not fitted. The cyan bars show the deviations in RCP2.6 when α and μ are fitted to the AOGCM's RCP2.6 data. The light blue bars show the deviations for RCP4.5 when α and μ are kept 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).

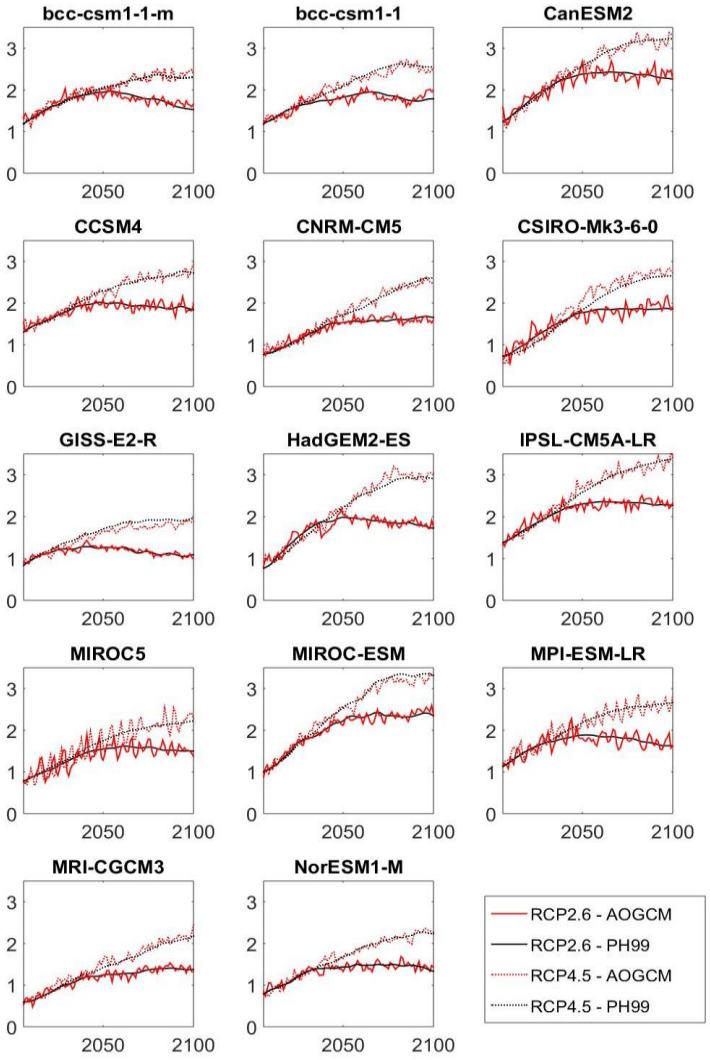

Figure 3: Comparison of temperature evolutions projected by the climate module PH99 (solid and dotted black curves) to the actual AOGCM's temperature (solid and dotted red curves). α and μ have been tuned to fit the PH99 temperature path (solid black curve) to the respective AOGCM's RCP2.6 temperature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

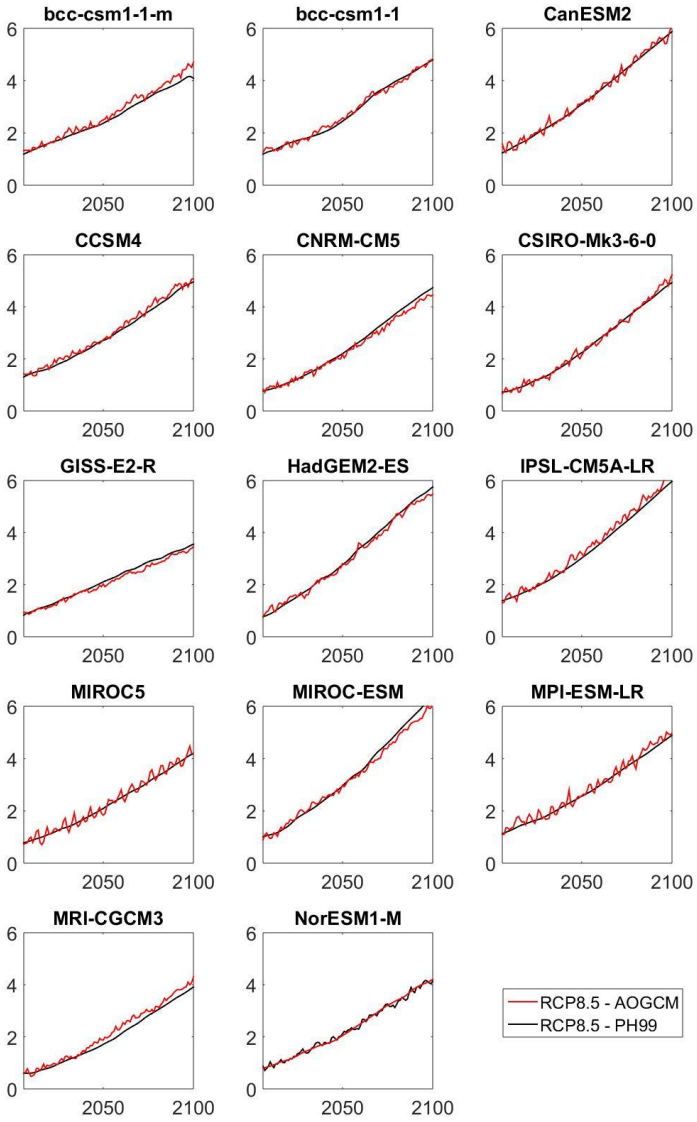

Figure 4: Comparison of temperature evolutions projected by the climate module PH99 (black solid curves) in RCP8.5 scenario to the actual AOGCM's temperature (red solid curves) in RCP8.5 scenario. α and μ are taken from the second step, where PH99 is calibrated to RCP2.6 scenario.

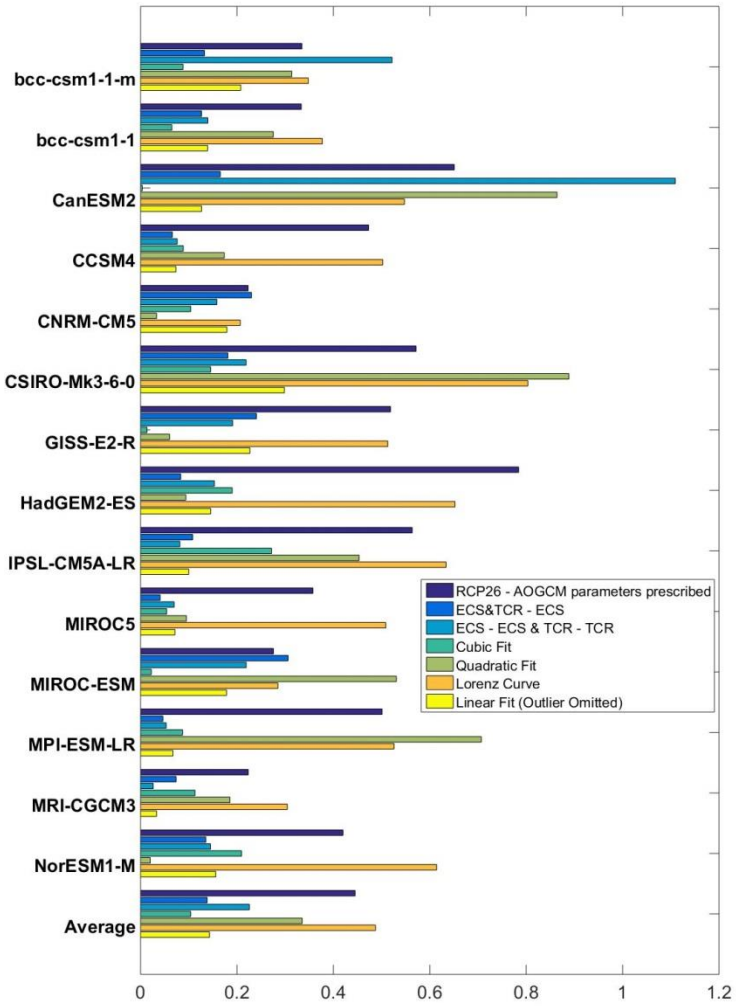

Figure 5: Modulus of mean temperature deviations over the period 2071-2100 (MTD) for PH99 from AOGCMs when α, μ, ECS, and TCR from Table 2 are related to ECS and TCR in Table 1. Using linear (yellow bars), quadratic (light green bars), and cubic functions (dark green bars), α and μ are related to ECS 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 (light blue bars). The 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 orange bars indicate MTD using Lorenz's curve.

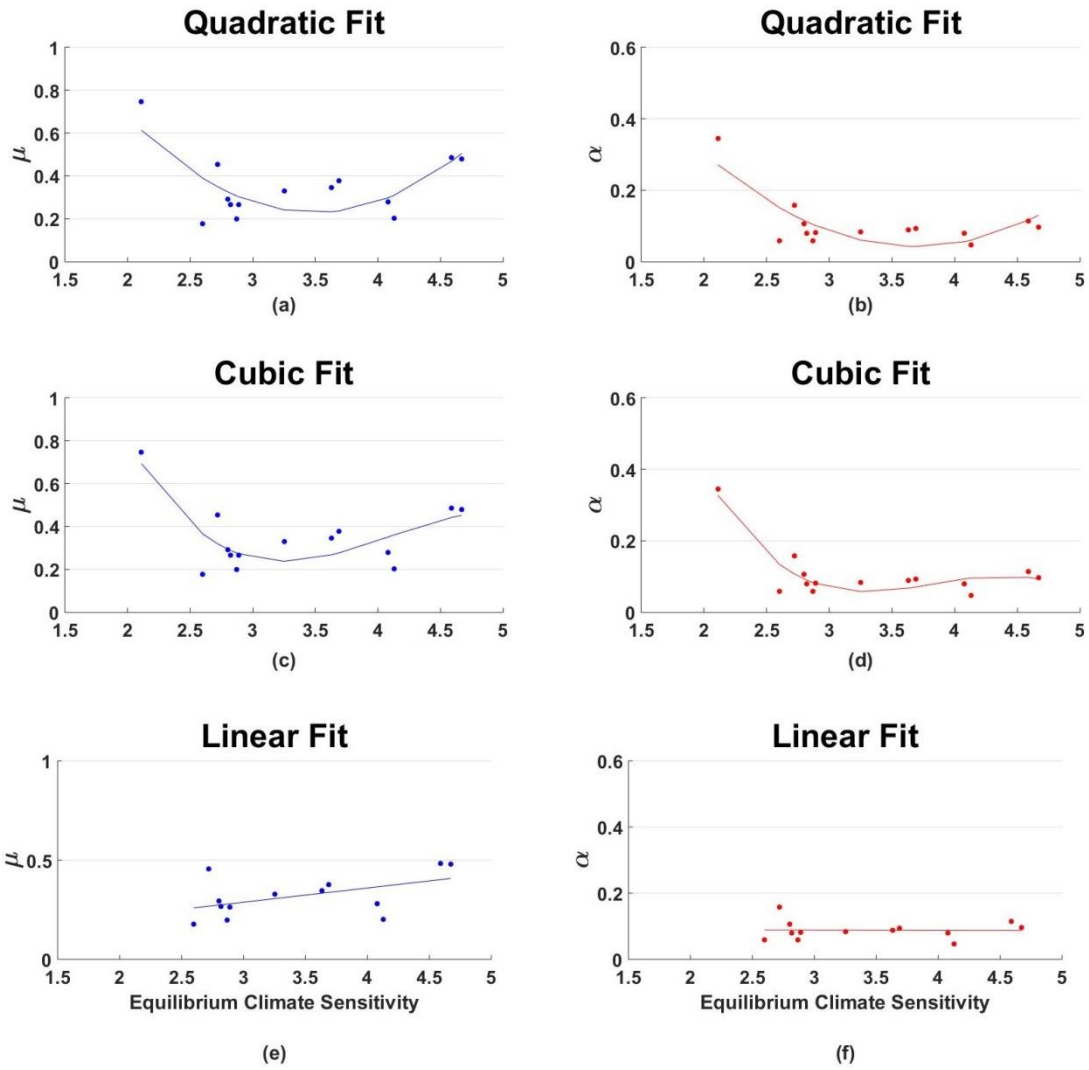

Figure 6: Quadratic (up), cubic (middle), and linear (down) relationships of μ (left) and α (right) in Table 2 to ECS in Table 1. Notice that in the linear case the model GISS_E2_R (the upper left sample), as an outlier, is out.

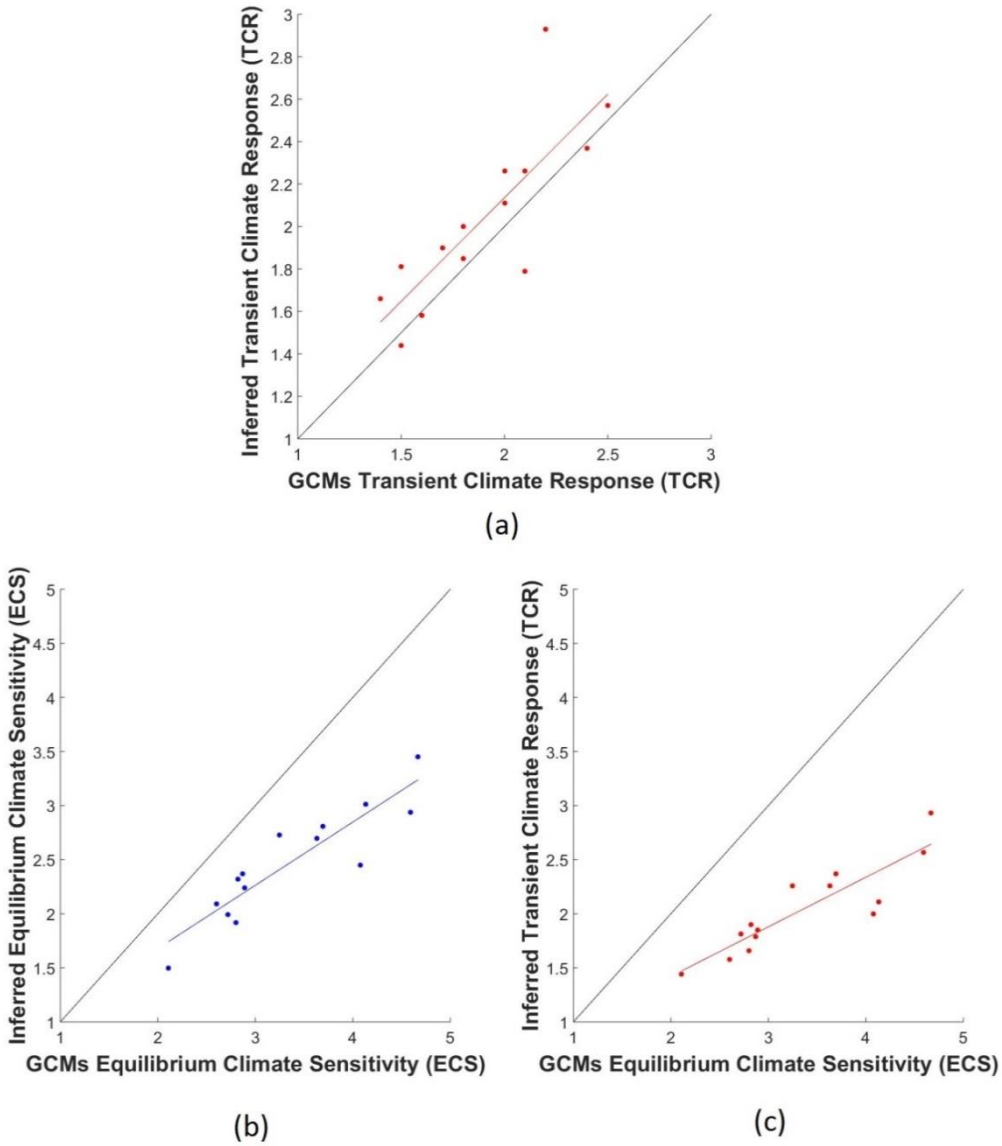

Figure 7: Inferred effective TCR vs. AOGCMs' TCR (a), inferred effective ECS vs. AOGCMs' ECS (b), and inferred effective TCR vs. AOGCMs' ECS (c). While the TCRs differ by less than 0.2 K, the ECSs differ by up to 2 K. This opens the door for a discussion as to whether PH99 should be calibrated using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

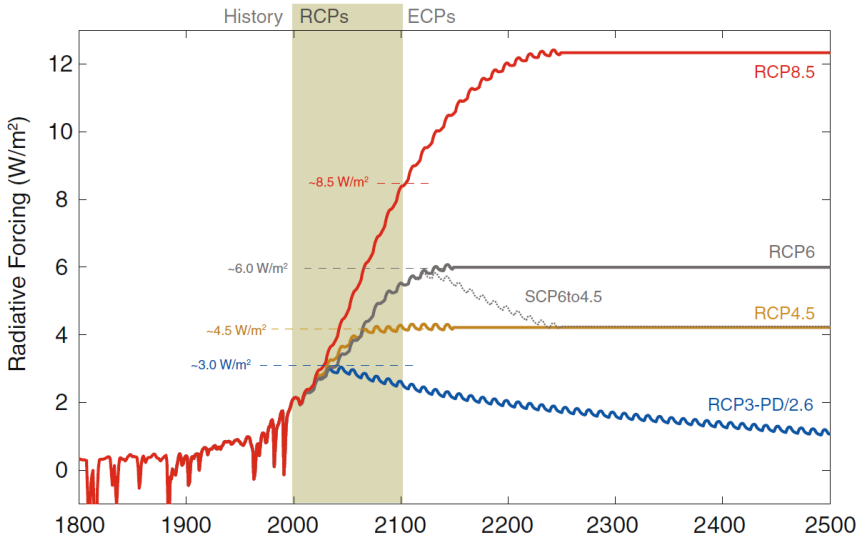

Figure 8: Total radiative forcing (anthropogenic plus natural) for RCPs – supporting the original names of the four pathways, as there is a close match between peaking, stabilization and 2100 levels for RCP2.6 (also called RCP3-PD), RCP4.5 & RCP6, and RCP8.5, respectively (taken from Meinshausen et al.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

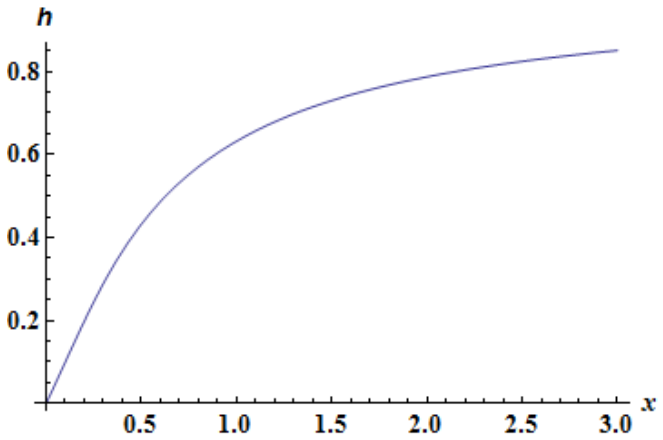

Figure 9: The auxiliary function *h*(*x*), which links the slow time scale of the 2-box model and the time scale of the 1-box model.

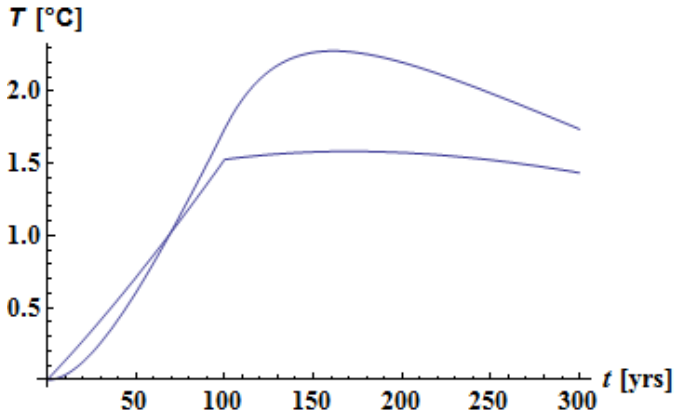

Figure 10: 1-box vs. 2-box model in response to kink-linear forcing as a stylized interpretation of mitigation-oriented forcing paths and for equal levels of ECS and TCR in both models. Kink-linear curve: 2-box model, smooth curve: 1-box model. The temperature development of the 1-box model overshoots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

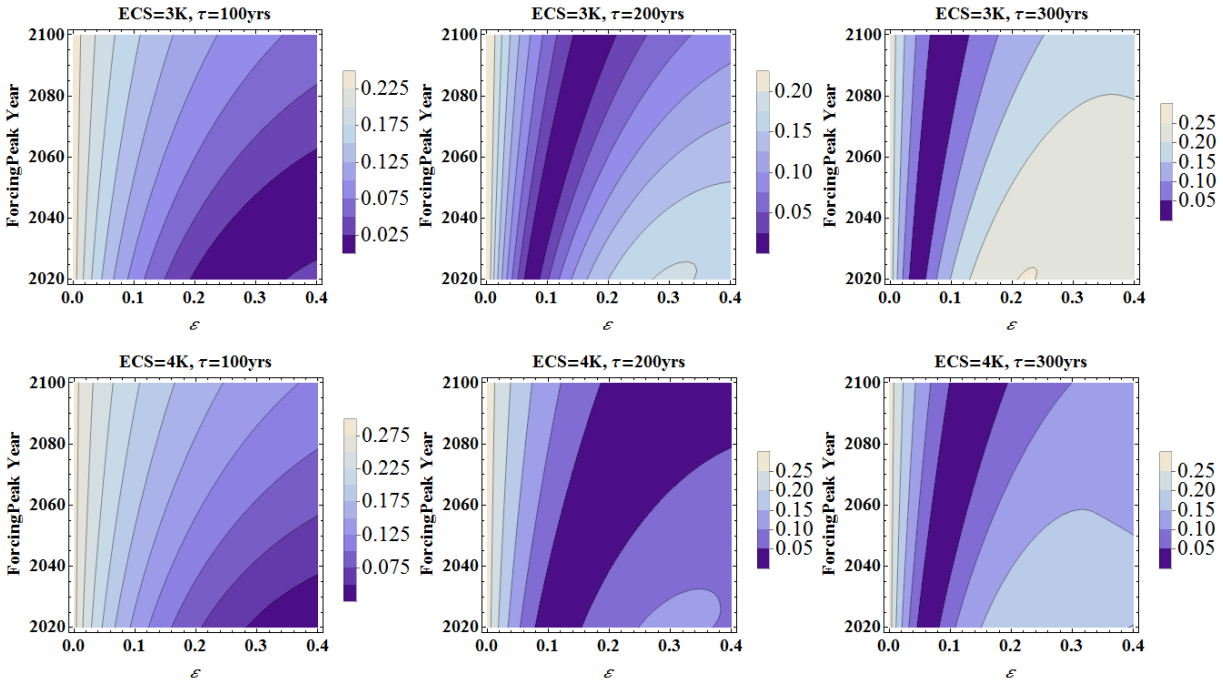

Figure 11: : Comparing GMT maxima of the 2-box model and the 1-box model, the latter being adjusted to the former by prescribing the linearly transformed ECS and TCR according to the scheme ETE. Abscissa: ε, ordinate: changed peaking year t_1 , however transformed to years, for the 2-box ECS of 3 K and 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.

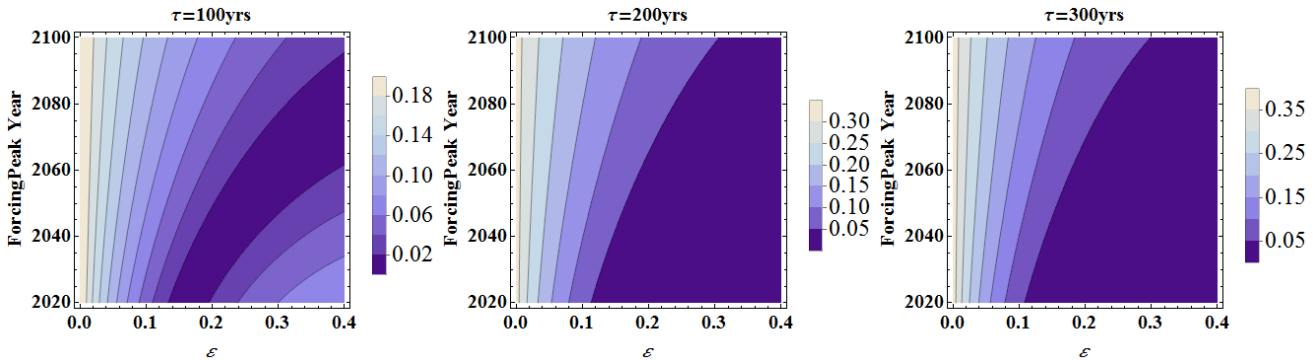

Figure 12: Similar to the previous figure (relative max. GMT error with abscissa: ε, ordinate: t_1 [yrs],), however for a further adjusted ECS of the 1-box model, such that perfect matching is achieved for t_1=100 yrs, ε=0.2, and a 1-box time scale of 12 yrs. For most of the parameter settings, the relative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

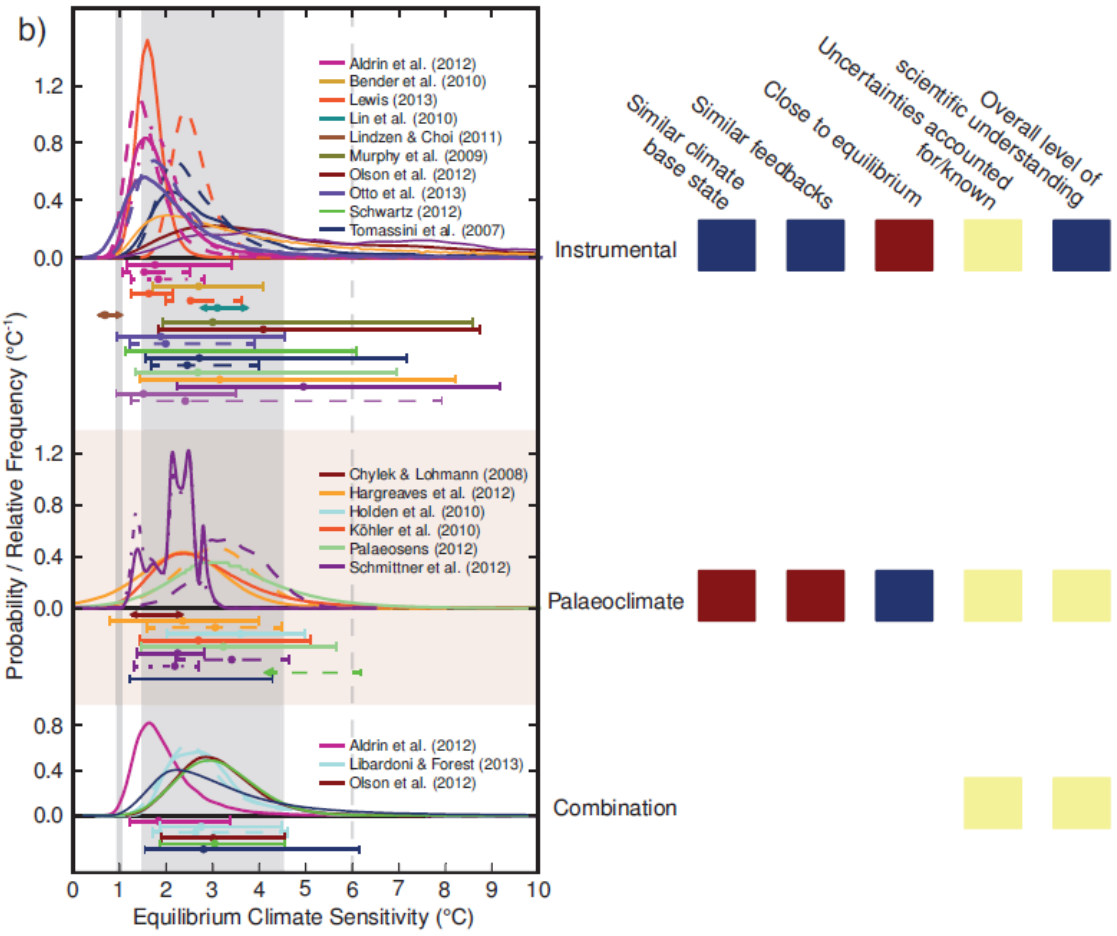

Figure 13: Probability density distributions of ECS according to IPCC AR5 WG-I (Bindoff et al., 2013, Fig. 10.20).

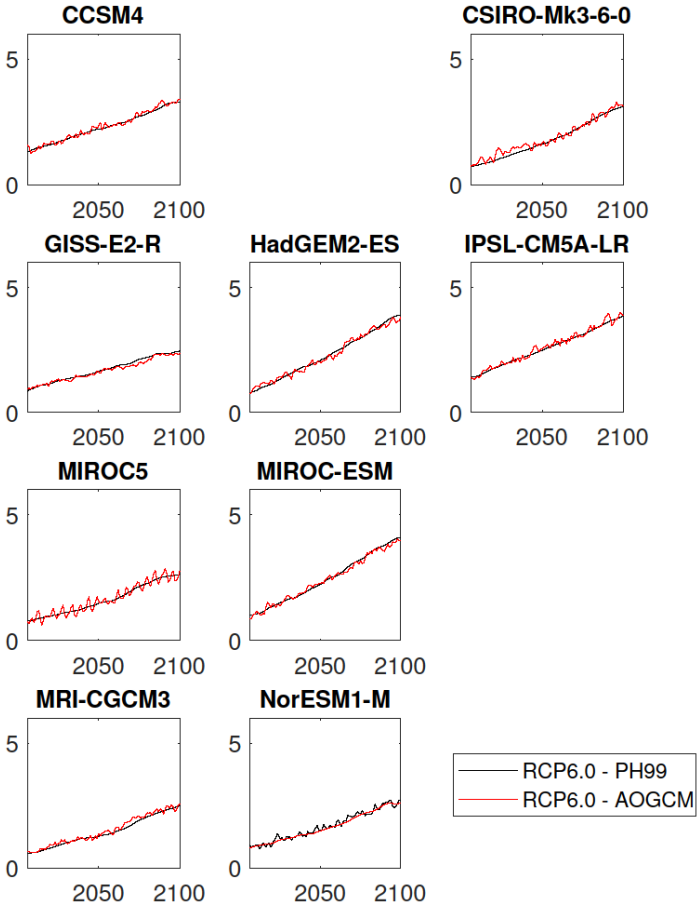

Figure 14: The comparison of temperature evolutions projected by the climate module PH99 (black solid curves) in RCP6.0 scenario to the actual AOGCM's temperature (red solid curves) in RCP6.0 scenario. α and μ are taken from the second step, where PH99 is calibrated to RCP2.6 scenario

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	MTD RCP8.5
bcc_csm1_1_m	**0.029**	0.040		0.236	0.018	**0.007**		0.154
bcc_csm1_1	**0.009**	0.066		0.052	0.064	**0.021**		0.059
CanESM2	**0.001**	0.021		0.043	0.039	**0.003**		0.018
CCSM4	**0.033**	0.003	0.069	0.132	0.024	**0.005**	0.064	0.128
CNRM_CM5	**0.014**	0.001		0.201	0.005	**0.012**		0.273
CSIRO_Mk3_6_0	**0.036**	0.115	0.040	0.063	0.017	**0.015**	0.168	0.278
GISS_E2_R	**0.008**	0.114	0.094	0.144	0.064	**0.003**	0.027	0.015
HadGEM2_ES	**0.018**	0.103	0.036	0.131	0.057	**0.020**	0.097	0.211
IPSL_CM5A_LR	**0.020**	0.043	0.050	0.201	0.121	**0.013**	0.017	0.033
MIROC5	**0.015**	0.044	0.029	0.089	0.032	**0.009**	0.034	0.106
MIROC_ESM	**0.028**	0.104	0.079	0.238	0.140	**0.012**	0.044	0.241
MPI_ESM_LR	**0.017**	0.047		0.119	0.108	**0.015**		0.060
MRI_CGCM3	**0.015**	0.061	0.083	0.208	0.001	**0.007**	0.004	0.061
NorESM1_M	**0.011**	0.000	0.026	0.043	0.024	**0.000**	0.006	0.082
Multimodel Mean	**0.018**	0.054	0.056	0.136	0.035	**0.005**	0.039	0.078
Standard Deviation	**0.010**	0.041	0.025	0.071	0.044	**0.006**	0.053	0.093

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	**MTD RCP8.5**
bcc_csm1_1_m					0.287	0.257		**0.027**
bcc_csm1_1					0.091	0.008		**0.039**
CanESM2					0.008	0.025		**0.010**
CCSM4	0.038	0.067	**0.018**	0.086	0.059	0.004	0.010	**0.004**
CNRM_CM5					0.117	0.151		**0.005**
CSIRO_Mk3_6_0	0.161	0.199	**0.019**	0.062	0.119	0.019	0.034	**0.015**
GISS_E2_R	0.041	0.037	**0.019**	0.046	0.045	0.023	0.011	**0.001**
HadGEM2_ES	0.146	0.233	**0.021**	0.063	0.146	0.252	0.073	**0.017**
IPSL_CM5A_LR	0.016	0.077	**0.001**	0.095	0.052	0.078	0.030	**0.002**
MIROC5	0.067	0.079	**0.011**	0.032	0.025	0.006	0.019	**0.019**
MIROC_ESM	0.187	0.070	**0.005**	0.198	0.309	0.235	0.140	**0.007**
MPI_ESM_LR					0.011	0.082		**0.012**
MRI_CGCM3	0.092	0.068	**0.003**	0.042	0.008	0.014	0.055	**0.027**
NorESM1_M	0.068	0.021	**0.016**	0.136	0.070	0.055	0.054	**0.013**
Multimodel Mean	0.091	0.095	**0.007**	0.084	0.096	0.086	0.029	**0.014**
Standard Deviation	0.060	0.072	**0.008**	0.053	0.096	0.096	0.041	**0.011**