Relating Climate Sensitivity Indices to projection uncertainty

Many thanks to the reviewers for their constructive reviews, which I address point-by-point in the following:

Reviewer 1

Reviewer Point P 1.1 — The manuscript in the current version has substantially improved compared to the initial draft. The order of the sections is more suitable for the ESD format and a number of details have been added to clarify the methodology. Particularly, I appreciated that the authors now explicitly discuss the choice of the two timescales for the optimisation in section 1 and, again, in the conclusions. Still, not only Proistosescu and Huybers, 2017, but also other authors (Caldeira et al. 2013; Joos et al. 2013; Tsutsui 2017) have chosen the three timescales for a pulse-response exponential relaxation model, similarly through an educated guess based on heuristic considerations. Therefore, I wonder if the author might explicitly mention the motivation behind his choice, the same way he did in the response to the reviewers. Provided that the author accounts for this general comment, and for the specific suggestions listed in the following, I thereby recommend that the manuscript is accepted for publication.

Reply: Thanks, once again for a careful and very constructive review. Regarding this suggestion, I have included an additional paragraph in the methods sections when introducing the 2 timescale model, defending the logic in the same manner as used in the paper response. I've also included an additional supplemental figure 5 to illustrate the logic.

Reviewer Point P 1.2 — - l. 66 and elsewhere: there are probably some LaTeX problems with the citation format for several references throughout the manuscript

Reply: Noted and fixed throughout..

Reviewer Point P 1.3 — - Eq.2: the equations appear a bit packed here. Consider introducing a broader spacing

Reply: Done

Reviewer Point P 1.4 — - ll. 87-88: as far as my memory goes, the value of 3.7 Wm-2 for the CO2 radiative forcing is computed at doubling compared to pre-industrial baseline (see IPCC AR4, sect. 2.3). Can the authors explain why they refer to it as for CO2 quadrupling?

Reply: Apologies - it should indeed read 7.4Wm^{-2}, thanks for catching that.

Reviewer Point P 1.5 — - l. 174: this is the only explicit reference to Figure S3 in the manuscript. Nevertheless, this figure contains quite a relevant part of the information that this work aims at conveying. The author might consider moving it from the supplementary material, or at least attach it to a more extensive discussion, since it is crucial in order to support the discussion on the role of the historical forcing for the role of the metrics in the CMIP models (ll. 192-198);

Reply: I've put the revised corner plot in the main text as suggested and expanded the discussion of the constrained MCMC posterior distribution

Reviewer Point P 1.6 — - Figure S4: I think this figure might benefit by reducing the thickness of the lines, as those are barely legible in the current version;

Reply: Done.

Reviewer Point P 1.7 — - ll. 226-227: I do not understand the author's choice to generate noise that has the parameters of an hypothetical multi-model mean noise. Particularly, it is not clear to me why the statistics of the noise including model would would be more representative of the metrics relation with the warming than the actual model if a model-dependent noise is applied;

Reply: To clarify, it's not the noise characteristics of the multi-model mean - it's the parameters for variance and lag-1 autocovariance are computed for each member of the ensemble, and our noise model uses the central estimate for those parameters. Using the actual noise from the CMIP models was considered but in practise creates more uncertainties and subjective decisions - how should internal variability be separated from forced response and long term drift? An AR1 acting as a stochastic component for weather noise in the simple model with the same order of magnitude internal variability as we see in GCMs was chosen as a simple sensitivity to a well defined source of noise. I appreciate, of course, that there is certainly an assumption of AR1 noise being representative of climate noise - which I already discuss, but I would argue that the current model is informative as a simple sensitivity test for how additive noise impacts the model fit.

Reviewer Point P 1.8 — - ll- 281-283: the point that I made in my previous revision is not that there is an indefinite number of exponentials that can be combined to describe the response function. Rather, I wanted to stress that other methods, as those based on the Ruelle's response theory (cfr. Ragone et al. 2016, Lucarini et al. 2017, Lembo et al. 2019; I believe that the correct reference here is "Lembo V., Lucarini V., and Ragone F., 2019, Beyond Forcing Scenarios: Predicting Climate Change through Response Operators in a Coupled General Circulation Model, ArXiv, 1912.03996"), allow to extract a (linear) response function without making any assumption on the parameters describing its optimal fit. Consequent to the theory, it can be proven (see Lucarini, V., 2018, Revising and extending the linear response theory for statistical mechanical systems: Evaluating observables as predictors and predictands) that any Green function can be expressed as an infinite sum of exponentials.

Reply: Thanks for this clarification, I've incorporated your central point (that a general linear response function can be described with a decaying exponential basis set) in the new discussion on the motivation for the 2 timescale model in Section 1.1

0.1 Technical comments

Reviewer Point P 1.9 — - l. 28: replace "Response at" with "Response at time of";

Reply: Done

Reviewer Point P 1.10 — - l. 33: replace ”given the” with ”given that the”;

Reply: Done

Reviewer Point P 1.11 — - l. 73: insert blank space before ”(Smith”;

Reply: Done

Reviewer Point P 1.12 — - l. 112: replace ”inefficient requires” with ”inefficient, as it requires”;

Reply: Done

Reviewer Point P 1.13 — - l. 250: replace ”given we” with ”given that we”;

Reply: Done

Reviewer 2

Reviewer Point P 2.1 — The author significantly improved the paper. A detailed and comprehensive description of the model and the optimization method has been included, which substitutes the former, in my view insufficient, appendix. Now, the setup of the model and the optimization procedure defining the parameter distribution (i.e. the model ensemble) becomes much clearer. An attempt is made to evaluate/validate the model by applying it to historical CMIP simulations and projections. The results indicate that, in general, the model performs well. In addition, it is shown that considering the drift can indeed improve (at least a little) the estimate of TCR (and other metrics). Both tests add more confidence to the main conclusions. Furthermore, the author now discusses the potential effect of the assumption of two timescale on the results, and gives some references to the linear response theory, which helps to assess the results. Overall, I believe the manuscript now provides sufficient new and valuable information to warrant publication. I only have very few minor/technical points:

Reply: Many thanks for the constructive review, and positive assessment of the revised study

Reviewer Point P 2.2 — - Abstract: The acronyms ECS and TCR are used without previous definition.

Reply: Thanks - fixed.

Reviewer Point P 2.3 — - L93: $F_{otherAnt}- > F_{other}$

Reply: Fixed

Reviewer Point P 2.4 — - Citations: it seems that sometimes the citation styles ’(authors, year)’ and ’authors (year)’ are mixed-up (e.g. L90, L110)

Reply: fixed throughout

1 Changes in manuscript

I have endeavored to address the minor issues brought up by reviewers in the second round of review:

- Added discussion in Section 1.1 describing the logic for the use of a 2 timescale model, and wider issues of how pulse-response models relate to generic linear response models
- corrected latex citation issues
- reformatted Eq 2
- moved figure S3 to the main text (new Figure 3), and added more extensive discussion of posterior distribution of the MCMC. Included analytical solution for TCR to provide insight into how constrained transient warming would impact parameter constraints.
- fixed minor technical issues

Relating Climate Sensitivity Indices to projection uncertainty

Benjamin Sanderson[1,2]

[1]CERFACS, Toulouse, France
[2]NCAR, Boulder CO, USA

Correspondence: Benjamin Sanderson (sanderson@cerfacs.fr)

Abstract. Can we summarize uncertainties in global response to greenhouse gas forcing with a single number? Here we assess the degree to which traditional metrics are related to future warming indices using an ensemble of simple climate models together with results from CMIP5 and CMIP6. We consider Effective Climate Sensitivity (EffCS), Transient Climate Response at CO_2 quadrupling (T140) and a proposed simple metric of temperature change 140 years after a quadrupling of carbon dioxide (A140). In a perfectly equilibrated model, future temperatures under RCP(Representative Concentration Pathway)8.5 are almost perfectly described by T140, whereas in a mitigation scenario such as RCP2.6, both ~~ECS~~ Equilibrium Climate Sensitivity (ECS) and T140 are found to be poor predictors of 21st century warming, and future temperatures are better correlated with A140. However, we show that T140 and EffCS calculated in full CMIP simulations are subject to errors arising from control model drift and internal variability. Simulating these factors in the simple model leads to greater error in the measured value of T140 than for EffCS. As such, if starting from a non-equilibrated state, measured values of Effective Climate Sensitivity can be better correlated with true ~~TCR~~ Transient Climate Response (TCR) than measured values of TCR itself. We propose that this could be an explanatory factor in the previously noted surprising result that EffCS is a better predictor than TCR of future transient warming under RCP8.5.

Introduction

Summarizing the response of the Earth System to anthropogenic forcers with metrics has long been practised as a way to illustrate uncertainty in Earth system response to greenhouse gases. For example, the concept of the Equilibrium Climate Sensitivity (ECS), the equilibrium global mean temperature increase which would be observed in response to a doubling of atmospheric carbon dioxide concentrations (Hansen et al., 1984) has existed for over 50 years (Charney et al., 1979) and significant amount of literature has been devoted to constraining its value (Knutti et al., 2017).

The Earth system responds to a step-change in forcing on timescales ranging from days to millennia (Knutti and Rugenstein, 2015), so an ‘Effective Climate Sensitivity’ (EffCS hereon) is often used as a proxy for decadal to centennial feedbacks. EffCS is generally calculated in a coupled atmosphere-ocean model from the output of the ‘abrupt4xCO2’ simulation, a standard experiment in which CO2 concentrations are quadrupled instantaneously from pre-industrial levels and the model is allowed to evolve (Gregory et al., 2004).

EffCS is calculated by assuming that a model is associated with a single feedback parameter (i.e. a rate of change of top of atmosphere radiative flux per unit surface temperature increase), allowing the equilibrium temperature response to a step change forcing to be predicted by linear extrapolation (we refer to this approach henceforth as the Constant Feedback (CF) approximation, with EffCS referring to the estimate of ECS made using this approach). Another metric, the Transient Climate Response at the time of CO_2 doubling (TCR) or quadrupling (T140) is calculated from an '1pctCO2' idealized experiment in which CO_2 concentrations are increased by 1 percent each year, starting from a pre-industrial state, resulting in linearly increasing forcing.

Although it was generally assumed that TCR would be a better predictor of transient warming under a high emissions scenario such as RCP8.5 (Riahi et al., 2011), a complication has arisen due to the fact that EffCS seems to be better correlated than TCR with 21st Century warming from present day levels under a business-as-usual scenario (Grose et al., 2018). The reason for this is not yet well understood given that the radiative pathway in RCP8.5 leading up to 2100 is relatively similar to that of the 1 percent annual increase experiment used to measure T140. Furthermore, neither EffCS nor TCR is well correlated with end of century temperatures in a mitigation scenario (Grose et al., 2018) such as RCP2.6 (Van Vuuren et al., 2011), which calls in to question the relevance of such summary metrics in the discussion of mitigation adaptions.

Similarly, a number of studies have shown that the EffCS approximation does not well describe the true equilibrium behaviour of most models (Knutti et al., 2017). When GCM abrupt-4xCO2 simulations are continued for thousands of years, many are found to deviate significantly from the linear trend-line one would fit to a 150 year simulation (Andrews et al., 2015; Knutti et al., 2017; Senior and Mitchell, 2000; Rugenstein et al., 2016).

The conceptual models representing the evolving feedbacks as a function of timescales vary slightly between studies - either modulating the efficacy of deep ocean heat uptake (Geoffroy et al., 2013; Winton et al., 2010; Held et al., 2010) or by representing the climate system as sum of warming patterns which emerge on different adjustment timescales (Armour et al., 2013; Rugenstein et al., 2016), each associated with their own feedback parameter. However, the analytical set of solutions for the temperature response to a step change in forcing is the same in either case - a superposition of decaying exponential modes with different timescales varying between a few years and a few centuries (Proistosescu and Huybers, 2017). It has been shown that the implications of these additional degrees of freedom, and ambiguity over contributions from different timescales of response might imply that EffCS may not be strongly constrained by temperature change over the last century (Proistosescu and Huybers, 2017; Andrews et al., 2018), and that the Long Term Equilibrium (LTE) sensitivity may be greater than that implied by estimates which use the CF framework (Otto et al., 2013; Lewis, 2013).

This state of understanding leads to a number of emerging critical questions which we discuss in this paper - can we explain the non-intuitive result that EffCS is a better predictor then T140 of end-of-century temperatures under RCP8.5? Which summary metrics of global sensitivity to greenhouse gas forcing are most useful for effective policy decisions? Finally, do the implicit structural assumptions underpinning the applicability of these metrics to the real world cause us to mis-categorize and potentially underestimate future warming risk?

1 A simple model example

We begin by considering an idealized ensemble of climate model simulations. We use a two timescale thermal response model, conceptually representing the deep ocean (with a response timescale of a century or more) and shallow ocean response timescales (with a response timescale of 10 to 50 years). Such a model, although simple, is capable of resolving evolving feedback amplitudes and can emulate the climatological responses of complex Earth System Models on two timescales. Such a model makes a structural assumption that the Earth can be modelled as a discrete sum of linear decaying exponential responses to forcing, but this model has been found to well describe GCM evolution on a century timescale (Proistosescu and Huybers, 2017; Geoffroy et al., 2013) and is sufficiently complex to illustrate the limitations of defining system sensitivity through TCR or EffCS.

To efficiently describe the response of the system to a generic forcing, this study employs a linear Green's function which describes the forcing by convolution with an impulse response Ruelle (1998) (in this case, the step change in CO2 forcing). This approach can be used to approximate and simplify global climate dynamics Ragone et al. (2016); Lucarini et al. (2017), and its computational efficiency allows Markov-Chain Monte Carlo parameter estimation for the physical parameters. Furthermore (and critically for this study), the pulse-response formulation can be used to self-consistently relate different metrics of climate sensitivity on a range of timescales ~~Lucarini et al. (2017)~~(Lucarini et al., 2017).

1.1 Model Formulation

The two-timescale impulse response model follows the thermal feedback-timescale implementation from the FAIR simple climate model (Smith et al., 2018; Millar et al., 2017), which follows Hasselmann et al. (1993):

$$\frac{dT_n}{dt} = \frac{q_n F - T_n}{d_n}; T = \sum_n T_n; n = 1,2, \tag{1}$$

where T_n is global mean temperature and for each timescale n. T_n is the component of warming associated with that timescale, q_n is the feedback parameter and d_n is the response timescale. ~~We consider the~~

Note the use of $n = 2$ timescales is a structural choice, used here both for relevance to parameterisation choices in simple models in common use (Smith et al., 2018; Geoffroy et al., 2013; Goodwin et al., 2018; Meinshausen et al., 2011) and because the parameters of two timescale model can be readily interpreted and unambiguously fitted to complex model output. $n = 1$ timescales provides a significantly poorer fit to temperature evolution in abrupt4xCO2 CMIP simulations (see Supplemental material, figure S5). Notably, some authors have considered 3 timescale models (Caldeira and Myhrvold, 2013; Joos et al., 2013; Tsutsui, 2 or general linear response functions (Ragone et al., 2016; Lucarini et al., 2017; Lembo et al., 2019) which allow (effectively) for an unlimited number of exponential response modes Lucarini (2018). While we observe a small further improvement in fit is apparent for some models with $n = 3$ modes, not all models appear to express 3 response timescales, which causes unstable fitting behaviour in those cases and a difficulty in comparing and interpretation of the values of fitted parameters across CMIP.

Nevertheless further understanding the feedback timescale dynamics of different CMIP models is an important topic for further research.

Total heat flux into the system R is divided into shallow and deep ocean ~~to be functions~~ fluxes, defined as a function of the same ~~timescale~~two timescales:

$$R_n = r_n(F - T_n/q_n); R = \sum_n R_n; \sum_n r_n = 1; n = 1,2$$

2

$$R = \sum_{n=1}^{2} R_n; \tag{2}$$

$$\text{s.t.} \sum_{n=1}^{2} r_n = 1; \tag{3}$$

$$R_n = r_n(F - T_n/q_n);$$

where r_n is an efficacy factor for heat absorbed by the deep ($n = 1$) or shallow ($n = 2$) ocean, which sum to unity given the boundary condition that $R(0) = F(0) = F_{4xCO2}$ at $t = 0$ (allowing just one degree of freedom r_1 - the fraction of heat which is allocated to deep ocean storage).

The particular solutions for temperature and radiation response to a step change in forcing F_{4xCO2} at time $t = 0$ can be expressed as a sum of exponential decay functions:

$$T_p(t) = F_{4xCO2} \sum_{n=1}^{2} q_n(1 - exp(-t/d_n)) \tag{4}$$

$$R_p(t) = F_{4xCO2} \sum_{n=1}^{2} r_n(exp(-t/d_n)), \tag{5}$$

where $T_P(t)$ is the annual global mean temperature and $R_p(t)$ is the net top-of atmosphere radiative imbalance at time t, and F_{4xCO2} is the instantaneous global mean radiative forcing associated with a quadrupling of CO$_2$, taken here to be ~~3.7~~7.4Wm^{-2} (Myhre et al., 2013).

We define a historical forcing timeseries as a function of CO$_2$ concentrations $C(t)$ and a non-CO2 forcing timeseries $F_{nonCO2}(t)$ (both taken from ~~(Meinshausen et al., 2011)~~Meinshausen et al. (2011)):

$$F(t) = \frac{F_{4xCO2}}{ln(4)} ln\left(\frac{C(t)}{C_0}\right) + f_r F_{aer} + F_{other}, \tag{6}$$

where f_r is a free parameter to allow scaling of aerosol forcing (conceptually allowing for forcing uncertainty in the historical timeseries), and ~~$F_{otherAnt}$~~ F_{other} is all other anthropogenic and and natural forcers (summed from ~~(Meinshausen et al., 2011)~~

Meinshausen et al. (2011)). The thermal response is calculated by expressing the numerical time derivative of the forcing timeseries $F(t)$ where the change in forcing in a given time-step in a given year $\Delta F(t')$ is $[F(t') - F(t'-1)]$. The forcing timeseries can thus be expressed a series of step functions, and T_p from equation 4 can be used to calculate the integrated thermal response.

$$T(t) = \sum_{t'=0}^{t} \Delta F(t') \sum_{n=1}^{2} q_n \left(1 - exp\left(\frac{-(t-t')}{d_n}\right)\right), \tag{7}$$

Heat fluxes into the deep ($D(t)$) and shallow ($H(t)$) ocean components are represented by numerical integration of the slow (n=1) and fast (n=2) pulse response components of $R_p(t)$ in Equation 5:

$$D(t) = r_1 \sum_{t'=0}^{t} \Delta F(t') exp\left(\frac{-(t-t')}{d_1}\right), \tag{8}$$

$$H(t) = (1 - r_1) \sum_{t'=0}^{t} \Delta F(t') exp\left(\frac{-(t-t')}{d_2}\right), \tag{9}$$

1.1.1 Model Optimization

The model input time-series for calibration are observed CO_2 concentrations, along with radiative estimates from Meinshausen et al. (2011) of non-CO2 forcing agents. We optimize the thermal model parameters for 2 timescales and the non-CO2 forcing factor (see Table 1).

A Markov-Chain Monte-Carlo (MCMC) optimization procedure produces an ensemble of parameter configurations such that the density of the simulations in parameter space reflects the likelihood as reflected in a cost function (as represented by a number of pre-defined likelihood metrics). MCMC algorithms employ a random walk in parameter space which ultimately seeks to produce a representative sample of the distribution.

The classical approach to this random walk is the Metropolis Hastings algorithm ~~MacKay and Mac Kay (2003)~~(MacKay and Mac Kay, 2 , which iteratively moves a set of 'walkers' or sample points throughout the parameter space. This approach, however is computationally inefficient, as it requires the specification of the transition distribution with a large number of degrees of freedom. Here, we follow the ~~(Goodman and Weare, 2010)~~ Goodman and Weare (2010) MCMC implementation which updates a walker position using a vector defined stochastically from the remaining ensemble of walkers. This approach has fewer degrees of freedom and is a well-tested approach for multidimensional optimization problems ~~Foreman-Mackey et al. (2013)~~ (Foreman-Mackey et al., 2013). We use flat initial parameter distributions as shown in Table 1, 200 walkers and 50,000 iterations for each optimization.

Cost functions are computed for global mean temperature, shallow and deep ocean content:

$$E_T = \sum_t \left(\frac{(T(t) - T_{obs}(t))}{\sqrt{2}\sigma_T} \right)^2 \tag{10}$$

$$E_H = \sum_t \left(\frac{(H(t) - H_{obs}(t))}{\sqrt{2}\sigma_H} \right)^2, \tag{11}$$

$$E_D = \sum_t \left(\frac{(D(t) - D_{obs}(t))}{\sqrt{2}\sigma_D} \right)^2, \tag{12}$$

where T_{obs} are HadCRUT 4.6 ensemble median global mean temperature anomalies ~~Morice et al. (2012)~~ (Morice et al., 2012) relative to a 1850-1900 baseline and σ_T is defined as the standard deviation of HadCRUT 1850-1900 values. Shallow and Deep Ocean heat fluxes are taken as the 0-300m and 300m+ heat content derivatives respectively in ~~(Zanna et al., 2019)~~ Zanna et al. (2019), with σ_H and σ_D taken as 1850-1900 standard deviations from the same dataset.

Flat priors are used for all parameters, with an additional prior on true equilibrium climate sensitivity using the likely value and upper bound on Equilibrium Climate Sensitivity from Goodman and Weare (2010) to specify the median and 90th percentile of a gamma distribution for equilibrium sensitivity (i.e. warming as t$\rightarrow \infty$).

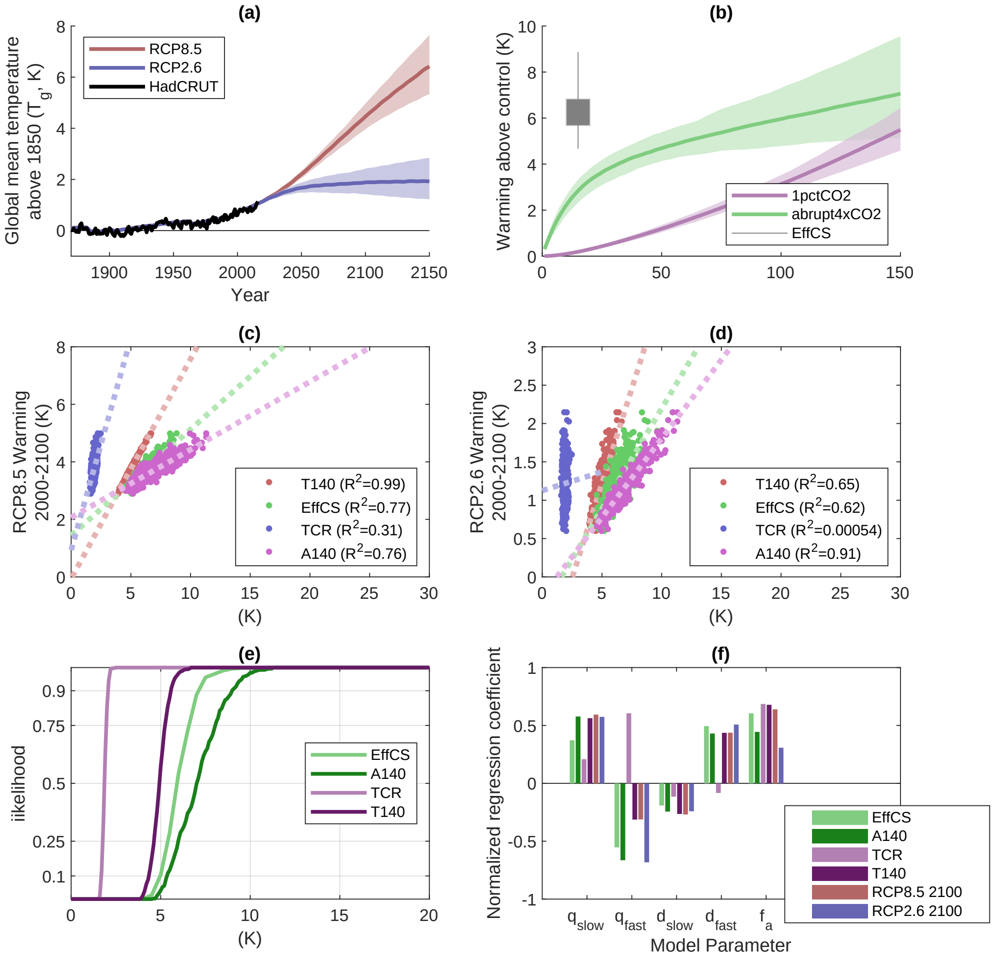
We demonstrate that this technique is able to capture the broad uncertainty associated with future projections of CMIP models by using pre-2020 temperatures in RCP8.5 to calibrate the simple model outlined above (Figure S3). In most cases, the future projection for each scenario falls within the distribution arising from the MCMC ensemble fit, with some specific exceptions - FIO-ESM, FGoals-G2, CCSM4 (which share some common heritage) and the GISS models. As such, the observationally fitted MCMC ensemble explores broadly comparable uncertainty to that seen in the bulk of the CMIP ensemble, with the caveat that the ensemble tends to under-sample cases where there is little or no long term warming response to emissions.

Long name	Symbol	Min	Max
Thermal equilibration of deep ocean Sensitivity (KWm^{-2})	q_1	0	10*
Thermal adjustment of upper ocean Sensitivity (KWm^{-2})	q_2	0	10
Thermal equilibration of deep ocean timescale $(years)$	d_1	100	4000
Thermal adjustment of upper ocean timescale $(years)$	d_2	10	100
Fraction of forcing in deep ocean response	r_1	0.	1
Non-CO2 Forcing ratio	f_r	.7	1.3

Table 1. A table showing model parameter values and minimum and maximum values allowed in model optimization.

The physical parameters of this simple model are constrained by historical carbon dioxide concentrations together with observed global mean temperatures from 1870 to present day (together with aggregate forcing estimates representing other anthropogenic emissions (Meinshausen et al., 2011), which are not the focus of this study). The posterior parameter distribution for the model can then be used to project the corresponding range of response in probabilistic projections of the future scenarios or in idealized experiments which simulate a range of self-consistent values for various climate sensitivity metrics.

Figure 1. An observationally constrained ensemble of simple models. (a) shows the global mean temperature both historically and under the RCP2.6 and RCP8.5 scenarios. Black lines show the HadCRUT data used in calibration, whereas shaded regions show the 10-90% range of scenario projections in the posterior simple model ensemble distribution. (b) shows the corresponding time-series posterior distributions for the abrupt4xCO2 and 1pctCO2 simulated experiments, with grey errorbars showing range of EffCS for CO_2 quadrupling (boxes and whiskers show 25-75th and 1-99th percentiles respectively) . (c/d) show relationships between different sensitivity indicators and 2000-2100 temperature changes under RCP8.5/RCP2.6 respectively (e) shows the posterior cumulative probability density functions for the 4 sensitivity variables considered and (f) shows the parameter regression coefficients relating the 5 normalized model input parameters to the 4 normalized sensitivity metrics.



1.1.2 Idealized Simulations

Effective Climate Sensitivity is measured by implementing a step-change abrupt CO_2 quadrupling, and following ~~(Gregory et al., 2004)~~ Gregory et al. (2004) to assess the linear extrapolation of warming at the point of net top of atmosphere energetic balance. A140 is calculated as the average of year 131-150 of the abrupt4xCO2 simulation. TCR and T140 are calculated as the average of years 61-80 and 131-150 respectively of the 1pctCO2 simulation (during which the CO2 concentrations are doubled and quadrupled, respectively), where CO_2 concentrations are increased annually by 1pct resulting in a linear increase in climate forcing. RCP scenario temperature trajectories are calculated for each parameter set using concentration and forcing timeseries from ~~(Meinshausen et al., 2011)~~ Meinshausen et al. (2011) from 1850 until 2300.

Resulting EffCS values (to a doubling of CO2) range from 2.4 to 4.6K (5th and 95th percentiles), and values of TCR from 1.6 to 2.2K (Figure 1(b,e)). This results in a range of 21st century warming under two scenarios considered, RCP2.6(RCP8.5) 2100 warming ranges from 1.4 to 2.4 K (3.8 to 5.1K) respectively (5th and 95th percentiles, see Figure 1(a)).

We then consider in the context of this observationally constrained ensemble of simple models, what idealized metrics of system response are most informative for describing 21st century warming. We consider four metrics: the EffCS, TCR/T140 (transient warming under an annual compounded 1 percent increase in CO_2 concentrations at time of CO_2 doubling/quadrupling, corresponding to years 70 and 140 of the simulation). We also introduce A140 as a possible metric for consideration, defined as the global mean warming above pre-emission levels in the abrupt4xCO2 simulation calculated 140 years after time of CO_2 quadrupling (here and throughout estimated as the mean from years 131-150). Figure 2 illustrates how ensemble spread would be impacted for a set of different scenarios if each of these metrics were constrained to lie within a narrow range (nominally the 45-55th percentile range of values present in the entire observationally constrained ensemble).

In the high emissions, RCP8.5 scenario (Riahi et al., 2011), 2000-2100 warming is nearly perfectly described ($R^2 = 0.99$) by T140, the transient climate response after 140 years in a 1 percent CO2 simulation (Figure 1(c) and Figure 2(k)). The corresponding response after only 70 years, TCR, is a much poorer predictor at $R^2 = 0.31$).

These results are physically intuitive. The climate forcing and rate of change of forcing in RCP8.5 at the end of the 21st century are of similar magnitude to those in year 140 of the 1 percent CO2 simulation, and so it is unsurprising that T140 is an efficient predictor for RCP8.5. TCR is a poor predictor in the simple model ensemble largely because TCR itself is already highly constrained by historical warming (Figure 1(e)), and thus the ensemble is effectively conditioned on a value of TCR and it has little additional explanatory value in explaining the ensemble variance in the RCP projections (Figure 2(f,g)).

EffCS and A140 are also well correlated with the RCP8.5 warming ($R^2 = 0.77$ and 0.76 respectively), but less so than T140. For the mitigation scenario RCP2.6, the most effective predictor of 2000-2100 warming is A140 ($R^2 = 0.91$). Both EffCS and T140 are weakly correlated ($R^2 = 0.62$ and 0.65 respectively), and TCR shows no significant correlation.

To help understand these relationships, we can perform a regression analysis of the metrics as a function of model ensemble parameters (Figure 1(f)) which suggests A140 and RCP2.6 warming from 2000 to 2100, are controlled by the *difference* between the slow and fast components of sensitivity. We can understand this in the context of the way the model is constrained by historical temperatures.

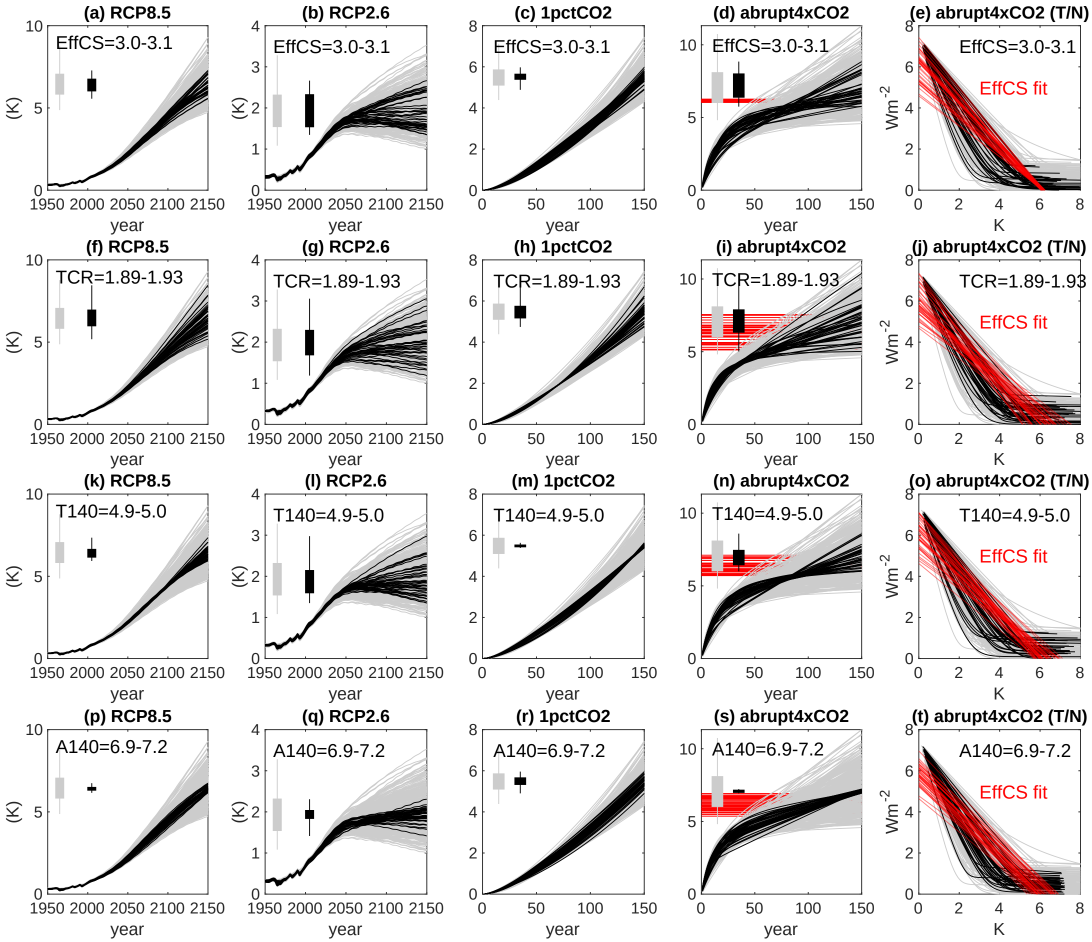
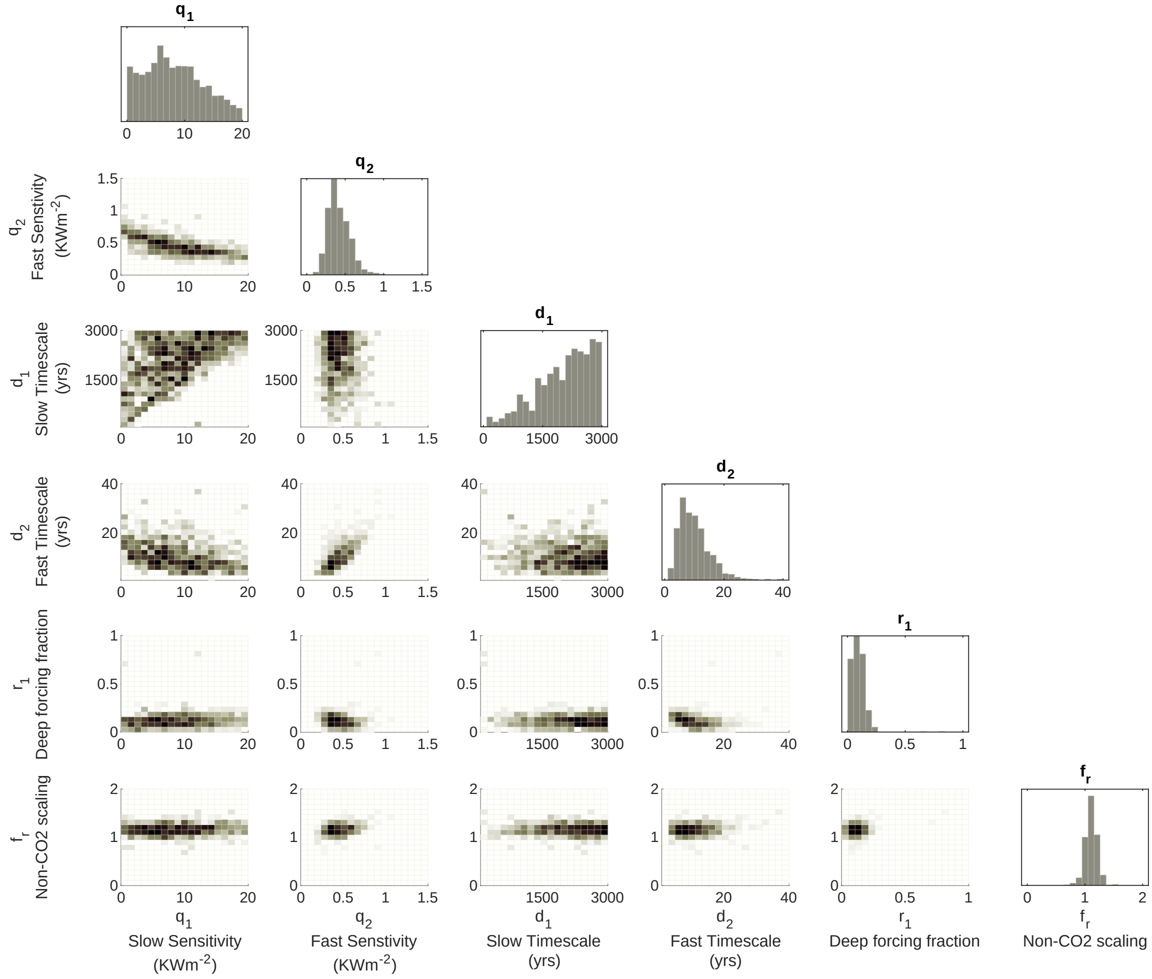


Figure 2. An illustration of how constraining different types of global sensitivity metric impact the idealized spread of global mean temperature evolution under different scenarios. Each row illustrates one constraint, Effective Climate Sensitivity to CO_2 doubling (EffCS), TCR (70 year, CO_2 doubling), T140 (140 year, CO_2 quadrupling) and A140. Lines in grey show the entire posterior distribution of models from Figure 1, while lines in black show the 45-55th percentiles of the distribution of the respective quantity. The first four columns show global mean temperature time-series of a scenario or idealized experiment - RCP8.5, RCP2.6, 1 percent ramping CO2, abrupt CO2 quadrupling (the 5th column shows energetic imbalance as a function of surface temperature in the abrupt4xCO2 experiment). Histograms show the resulting distribution of temperature in 2150 (RCP8.5/2.6) or year 140 (1pctCO2, abrupt4xCO2) for the complete distribution (grey) and 45-55th percentile range (black). Red lines show the distribution of values of effective climate sensitivity (4th column) and the trend lines used to compute it (5th column).

Figure 3. A 'corner-plot' showing the posterior parameter distribution attained by MCMC calibration of the simple climate model. Diagonal plots show posterior histograms for each of the parameter values optimized in the calibration, where the x-axis range reflects the bounding values of the initial flat prior distribution. Off-diagonal plots show pairwise distributions of parameters in the posterior distribution.



There is a ~~weak~~ trade-off between fast and slow components of climate sensitivity in the posterior parameter distribution of the ensemble (see ~~supplementary figure~~ Figure 3), which broadly determines the fraction of equilibrium warming associated with current forcing levels that has already been experienced. ~~If~~ There is also a correlation between fast sensitivity and fast timescale. These relationships should be broadly expected if we consider that the observed transient warming of the model has been constrained by the model. If we consider the analytical expression for TCR (warming after 70 years of 1 percent annual increase in CO2) in a 2 timescale model (from 7 following Smith et al. (2018)):

$$TCR = F_{2xCO2}\left[\left(q_1\left(1-d_1/70\left(1-e^{-70/d_1}\right)\right)\right)+q_2\left(1-d_2/70\left(1+e^{-70/d_2}\right)\right)\right], \tag{13}$$

where F_{2xCO2} is the forcing from a doubling of atmospheric CO_2, q_1, d_1 are the fast sensitivity and timescale and q_2, d_2 are the slow sensitivity and timescale. In the limit that $d_1 << 70$ and $d_2 >> 70$, we obtain the following:

$$TCR \approx F_{2xCO2}\left(\frac{q_1}{1+\frac{d_1}{70}}+\frac{q_2}{2}\left(\frac{70}{d_2}\right)^2\right), \tag{14}$$

This expression explains the primary features apparent in the MCMC posterior distribution, if we consider that the observations broadly fix the value of TCR: a inverse relationship is expected between q_1 and q_2, and we observe this in Figure 3. The fast component (left hand term in equation 14) is constrained by the historical warming timeseries to be non-zero (see Figure 3) - and there is a tight proportionality in constrained values of q_1 and d_1. Only the lower bound of the slow timescale d_2 is constrained for a given value of q_2, i.e. the transient warming alone provides no information on the upper bound of the slow response timescale.

Thus if a greater fraction of today's observed warming is explained with the faster component of model response, ~~there is~~ we would expect less unrealized warming in a mitigation scenario later in the century. This causes large uncertainties in RCP2.6 evolution in the constrained ensemble, even if in the case that we had confidence in the values of EffCS, TCR or T140 ~~are known~~ (Figure 2b,g,l).

The constrained distribution for fast-timescale sensitivity is near-Gaussian, and non-zero in all ensemble members, whereas slow-timescale sensitivity is more weakly constrained by the observations ranging from near-zero to large (20K/Wm^{-2}) long term equilibrium responses. The slow feedback component strongly controls A140 and RCP2.6 warming (Figure 1(d,f), Figure 2q).

RCP8.5 warming and T140, however are associated with a near-linear increase in forcing throughout the simulation which results in a near-linear temperature increase. The relative fraction of warming associated with fast- and slow-timescale feedbacks remains constant over time, and thus warming to date (effectively fixing TCR, subject to aerosol forcing uncertainty) better constrains relative error in future response in a non-mitigation scenario (Figure 2f).

2 Considering the multi-model ensemble

But how do the findings in the simple model framework reconcile with findings in the CMIP5 and CMIP6 multi-model ensembles? Firstly, it is plausible that there is some commonality in the lack of skill of TCR (the transient response after 70 years) in our simple model ensemble and in the CMIP ensembles. In our simple model case, the ensemble members were explicitly calibrated to reproduce the 20th and early 21st century warming - which is a very strong constraint on the value of TCR in this idealized setup.

Earth System Model calibration is conducted in a much larger parameter space by groups with a wide range of objectives which complicate interpretation (Mauritsen et al., 2012; Sanderson and Knutti, 2012), but simulations are generally only published using models which are able to adequately describe the 20th century and thus might be subject to a similar effective constraint on TCR which renders the metric ineffective for describing variance in the future evolution of the model. But there remains a direct contradiction for T140, where the simple model suggests T140 should be a better predictor than EffCS for non-mitigation warming in the 21st century whereas the opposite was found in the CMIP correlations (see Supplementary material, Figure S2 and ~~(Grose et al., 2018)~~Grose et al. (2018)).

To understand this, we need to consider how the properties of the simple model ensemble differ from the CMIP archive. Although the thermal response of the simple model is broadly able to represent the climatological response of CMIP models to step forcing and transient forcing in CO_2 over a century timescale ~~((Geoffroy et al., 2013; Proistosescu and Huybers, 2017))~~(Geoffroy et al., 2013; Proistosescu and Huybers, 2017), it contains no internal climate variability and all experiments in Section 1 are conducted from an idealized, perfectly spun up state.

Both of these assumptions are not true for CMIP5 or CMIP6. Measurement of EffCS and TCR are complicated by internal variability (Knutti and Rugenstein, 2015), and many models still exhibit some temperature drift in the control simulation from which the ‘1pctCO2’ simulations and ‘abrupt4xCO2’ simulations are branched (Figure 4). This creates uncertainty from two sources - firstly, it is not always apparent at what point during the control simulations the 1pctCO2 simulation has been branched, thus there is uncertainty in how the anomaly should be measured. Secondly, there is the potential for an unknown contribution of control drift to be erroneously included in the temperature evolution of the 1pctCO2 and abrupt4xCO2 simulations.

To assess the contribution of control drift bias in sensitivity metrics, we implement idealized representations of non-equilibration into our simple model from Section 1. We then create an idealized distribution of drift similar to that seen in the CMIP ensembles in the simple model ensemble by initializing the model 500 years before the experiment begins, defining an effective ‘baseline’ period from which anomalies are measured to be the average temperature between years 400 and 500. Climate internal variability is represented by a 2nd order autoregressive model, which is fitted to each CMIP model in turn. The ensemble-mean autoregressive parameters are used to create artificial ‘noisy’ simulations by linearly adding noise generated from the autoregressive model to the output of the simple model.

We consider the range of control drifts observed in the CMIP5 and CMIP6 ensembles (illustrated in Figure 4(L)) which range from -.3 to +.6K /century in the CMIP5 and CMIP6 models considered in this study. An idealized distribution of drift in the simple model ensemble is created by initializing the model 500 years before the abrupt4xCO2 or 1pctCO2 simulation with a non-zero, constant forcing drawn from a flat distribution ranging from -1 to $+1Wm^{-2}$, which results in a distribution of control drift of -.4K to +.4K per century (i.e. broadly comparable to the CMIP case). For each simulation we consider a baseline for temperature to be defined by the average global mean temperature in years 400-500.

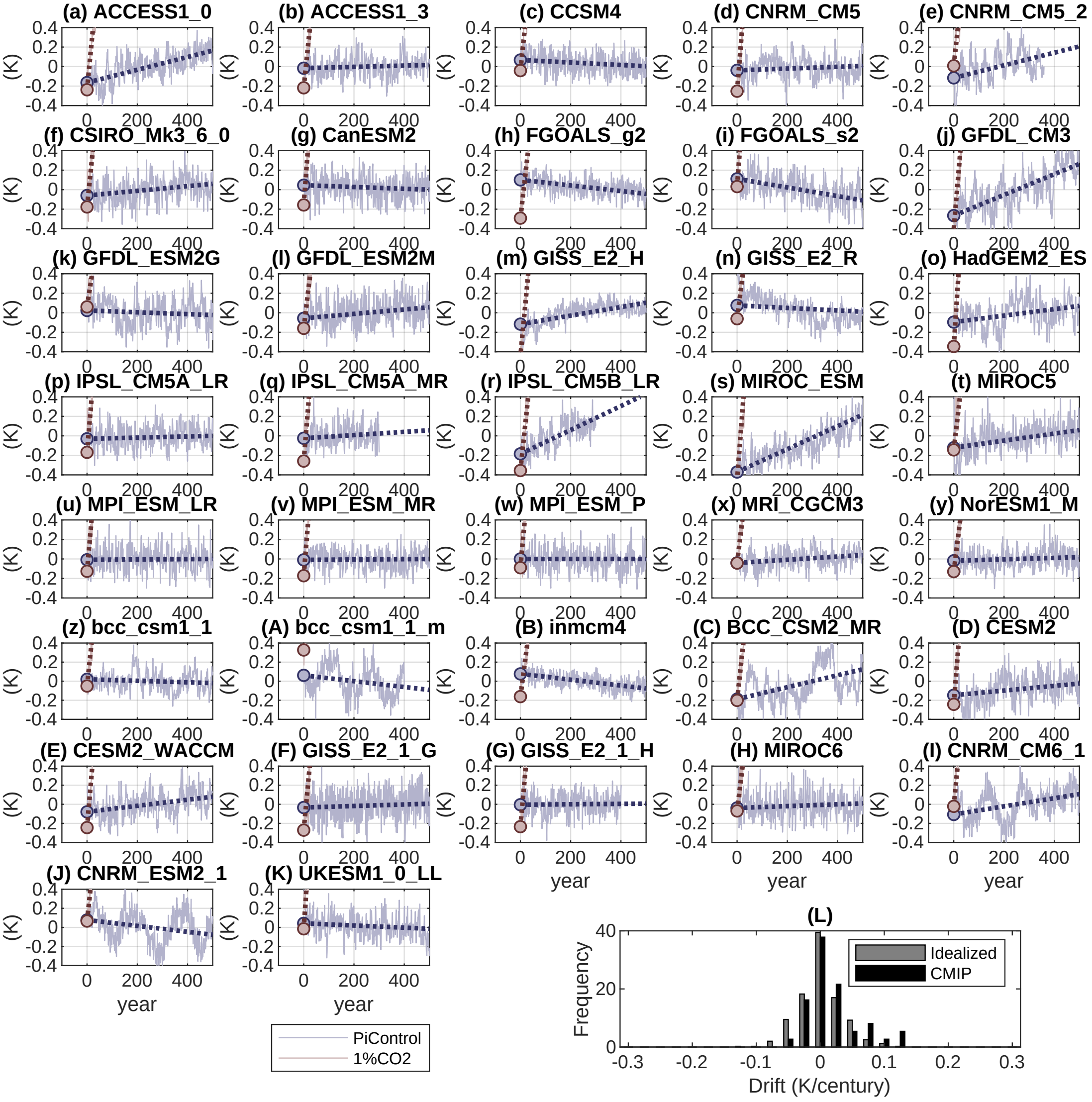


Figure 4. (a-K) Control simulation global mean temperatures from a selection of models in the CMIP5 and CMIP6 ensembles. Control simulations (blue) and initial years of 1pctCO2 simulations (pink) are plotted. Dotted lines show linear fit to the available timeseries. Blue and pink circles show the intersection of the linear temperature fit at the start of the simulation. (L) histogram showing the distribution of control model trend in CMIP (black) and in idealized ensemble of non-equilibrated simple models considered in Figure 5 (grey).

Predictor	T140	EffCS	A140	RCP8.5 2000-2100	RCP2.6 2000-2100
T140 (true)	1.00	0.78	0.77	0.99	0.65
EffCS (true)	0.78	1.00	0.70	0.77	0.62
A140 (true)	0.77	0.70	1.00	0.76	0.91
T140 (drift)	0.74	0.58	0.59	0.73	0.50
EffCS (drift)	0.73	0.94	0.67	0.73	0.59
A140 (drift)	0.74	0.67	0.95	0.73	0.86
T140 (noise)	0.99	0.77	0.76	0.98	0.65
EffCS (noise)	0.78	1.00	0.69	0.77	0.61
A140 (noise)	0.78	0.70	1.00	0.77	0.91
T140 (drift+noise)	0.70	0.55	0.55	0.69	0.47
EffCS (drift+noise)	0.72	0.93	0.65	0.71	0.58
A140 (drift+noise)	0.73	0.66	0.94	0.72	0.85

Table 2. A table showing R^2 regression statistics relating a set of predictors to a set of unbiased model properties. Predictors are Transient Climate Sensitivity at quadrupling of CO2 (T140), Effective Climate Sensitivity (EffCS) and warming 140 years after a quadrupling of CO2 (A140), additional rows show these values measured experiments conducted with unequilibrated base climates (drift), additive autoregressive noise (noise) and a combination of both factors (drift+noise). 'True' output model properties (T140, EffCS, A140, RCP8.5 and RCP2.6 warming from 2000 to 2100) are derived from the equilibrated model without noise.

To represent the first order effect of climate noise, we fit a 2nd order autoregressive model to the detrended global mean temperature timeseries in each available model in the CMIP5/6 ensemble. Taking CMIP mean parameters for the variance and autoregressive parameters, we generate noise for each realization of the simple model (though we note, in practise that the noise characteristics vary by CMIP model).

The results are illustrated in Figure 5(a), where the simple model ensemble is initialized in a non-equilibrium state with additive Gaussian noise. With these additional sources of error, both EffCS and A140 are not strongly impacted when measured in the noisy/unequilibrated model variants (Figure 5(b,c)), but the T140 measurement is strongly degraded (Figure 5(d)). Indeed, in this ensemble the biased measurements of EffCS or A140 are slightly better correlated with true T140 than the biased measurement of T140 itself. This provides a possible explanation for why T140 may be a poor predictor of RCP8.5 warming in CMIP.

In our simple framework, the reasons for the more accurate measurement of EffCS are primarily associated with the lack of equilibration. Simply adding noise from the autoregressive model has little effect on the accuracy of EffCS, T140 or A140 (where both T140 and A140 are estimated using the average of years 131 to 150 in the simulation, see Table 2).

Both A140 and EffCS are less sensitive to non-equilibrated initial states than T140. The former experiences the same variance due to the uncertain climate drift, but the absolute value of A140 tends to be larger than T140, thus there is less relative error in its estimation. The effect on the drift on EffCS is muted because the near-linear climate drift primarily biases the estimation of

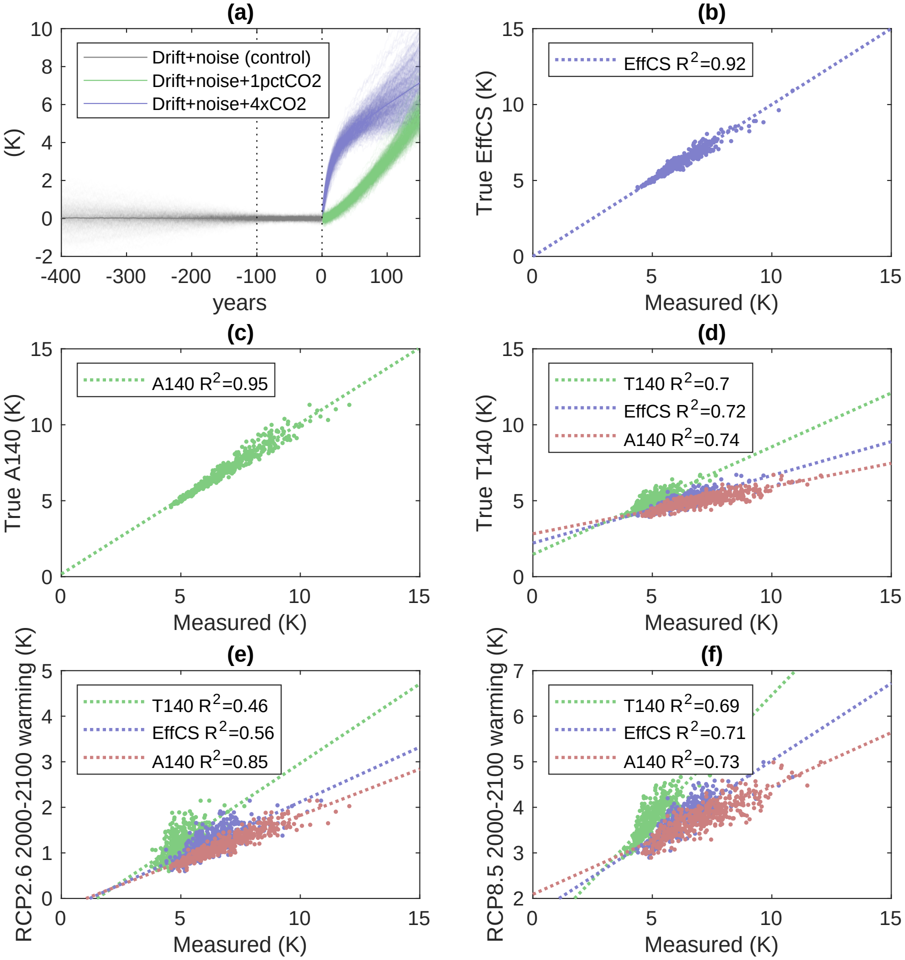
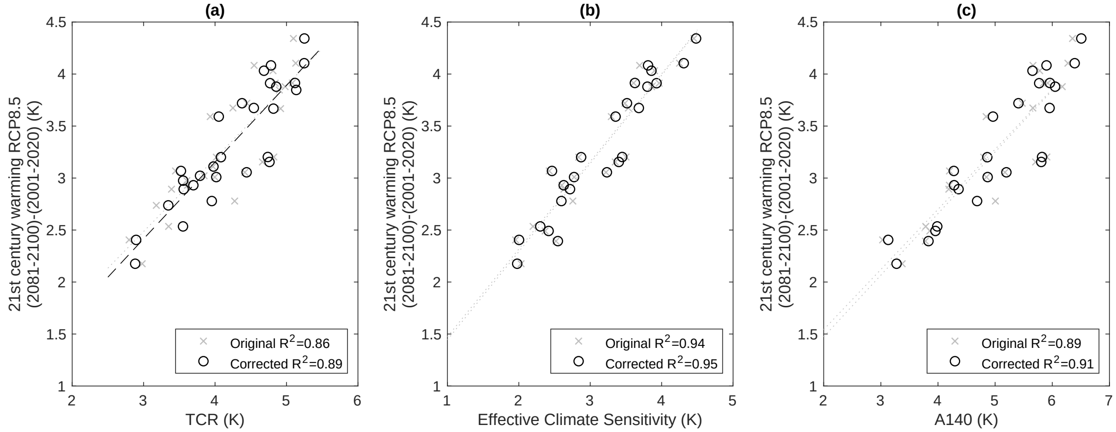


Figure 5. An idealized ensemble of simple models, where model parameters are identical to those considered in Figure 1(b), but models are initialized in a non-equilibrium state such that the baseline period is subject to some control drift, and model output is also subject to interannual variability of a similar magnitude to models in the CMIP archive. (a) shows global mean temperature evolution for the control period (gray), abrupt4xCO2 simulation (blue) and 1pctCO2 simulation (green). (b,c) show the true value of (EffCS,A140) as calculated in the noise-free, equilibrated simulations, plotted as a function of the measured value of (EffCS,A140) in a noisy, non-equilibrated simulations. (d,f,g) shows the true value of (T140,RCP2.6,RCP8.5 2000-2100 warming) plotted as a function of the measured values of T140, EffCS and A140 respectively.

Figure 6. Plots showing the correlation between TCR (a), EffCS (b) and A140 (c) with 21st century warming, here represented by the difference between 2001-2020 and 2081-2100 global mean temperatures in the 1st ensemble member for each model in the CMIP5 archive for the RCP8.5 scenario. Each plot shows the ’original’ calculation, where the baseline temperatures (and TOA fluxes for EffCS) are taken as the PIControl mean. In the ’corrected’ calculation, a correction term for the baseline temperature and control drift is applied. Correlation coefficients are shown for the original and corrected cases.



slow rather than fast feedbacks (see Supplementary Figure S1). Because EffCS is primarily a measure of fast-mode feedback strength (see Figure 1(f)), its value is less impacted if experiments are started from a non-equilibrium state.

There is some evidence that the lack of equilibration has an outsized effect on the estimation of TCR in the CMIP models. In Figure 6, we attempt to unbias the estimate of TCR in two ways. Firstly, we estimate the baseline temperature by regressing the temperatures in the first 20 years of the 1 percent CO2 ramp experiment as a function of time (see Supplementary Figure S4). Anomalies in temperature (and TOA fluxes for ECS) are measured relative to the corrected baselines derived from the 1pctCO2 simulation, and estimated linear pre-industrial trends are subtracted from the 1pctCO2 and abrupt4xCO2 timeseries. This pre-processing of the temperature timeseries improves the correlation between TCR and 21st century warming under RCP8.5 from 0.86 to 0.89. It also improves the correlation between EffCS and 21st century warming slightly from 0.94 to 0.95 (and A140 from 0.89 to 0.91).

These ‘corrected’ values (listed in Table 3) are estimates only, given that we would expect the regression estimate based on a short 20 year period to be itself subject to internal variability noise, and we are assuming that the abrupt4xCO2 simulation and 1pctCO2 simulation have the same baselines. However, the improvement in correlation with future warming seen over the case with the pre-industrial average baseline supports the hypothesis that control drift adds uncertainty to the estimation of all quantities (and particularly TCR). However, it is not a complete explanation - and even after this adjustment, EffCS remains better correlated to RCP8.5 transient warming than TCR in the multi-model ensemble.

Model	EffCS (org)	EffCS (corr)	A140 (org)	A140 (corr)	T140 (org)	T140 (corr)	RCP8.5 2000-2100	RCP2.6 2000-2100
ACCESS1_0	3.48	3.53	5.48	5.60	4.45	4.57	3.72	-
ACCESS1_3	3.30	3.38	4.84	5.02	3.93	4.11	3.59	-
BNU_ESM	3.86	3.80	6.17	6.05	4.98	4.86	3.88	0.63
CCSM4	2.84	2.87	4.80	4.86	4.02	4.08	3.20	0.44
CESM1_CAM5_1_FV2	3.31	2.89	5.29	4.44	-	-	-	-
CNRM_CM5	3.22	3.28	5.17	5.30	4.42	4.54	3.06	0.68
CNRM_CM5_2	3.37	3.37	5.11	5.12	4.29	4.29	-	-
CSIRO_Mk3_6_0	3.53	3.63	5.66	5.86	4.25	4.45	3.67	1.09
CanESM2	3.61	3.59	5.92	5.89	5.08	5.05	3.91	0.92
FGOALS_s2	3.85	3.78	5.90	5.76	4.76	4.62	-	-
GFDL_CM3	3.69	3.87	5.66	6.02	4.55	4.90	4.08	1.25
GFDL_ESM2G	2.37	2.34	3.86	3.80	-	-	2.49	-0.08
GFDL_ESM2M	2.52	2.60	3.78	3.93	-	-	2.39	0.32
GISS_E2_H	2.20	2.42	3.79	4.23	3.35	3.79	2.53	0.36
GISS_E2_R	2.03	2.01	3.37	3.34	2.98	2.94	2.18	0.09
HadGEM2_ES	4.25	4.34	6.27	6.45	5.13	5.30	4.10	0.87
IPSL_CM5A_LR	3.90	3.92	5.78	5.78	4.81	4.81	4.03	0.80
IPSL_CM5A_MR	3.96	4.01	5.84	5.93	4.84	4.93	3.91	0.59
IPSL_CM5B_LR	2.43	2.54	4.20	4.43	3.45	3.67	3.07	-
MIROC_ESM	4.45	4.51	6.35	6.56	5.10	5.30	4.34	1.26
MIROC5	2.60	2.62	4.20	4.27	3.61	3.68	2.93	0.62
MPI_ESM_LR	3.50	3.45	5.91	5.82	4.82	4.74	3.20	0.43
MPI_ESM_MR	3.35	3.42	5.71	5.84	4.66	4.80	3.15	0.36
MPI_ESM_P	3.34	3.31	5.71	5.64	4.57	4.49	-	-
NorESM1_M	2.63	2.68	4.19	4.29	3.39	3.49	2.89	0.55
bcc_csm1_1	2.77	2.77	4.85	4.87	4.00	4.02	3.01	0.52
bcc_csm1_1_m	2.76	2.68	5.00	4.84	4.27	4.11	2.78	0.30
inmcm4	1.96	2.00	3.03	3.13	2.80	2.89	2.41	-
BCC_CSM2_MR	2.87	2.91	4.75	4.89	3.88	4.02	-	-
CESM2	4.70	4.80	7.20	7.40	5.01	5.20	-	-
CESM2_WACCM	4.32	4.43	6.51	6.74	4.68	4.91	-	-
GISS_E2_1_G	2.61	2.66	4.18	4.27	1.95	2.04	-	-
GISS_E2_1_H	2.99	3.09	4.94	5.13	4.11	4.31	-	-
MIROC6	2.40	2.40	3.96	3.98	3.47	3.49	-	-
CNRM_CM6_1	4.69	4.67	6.75	6.71	5.49	5.46	-	-
CNRM_ESM2_1	4.35	4.30	6.16	6.07	5.12	5.02	-	-
UKESM1_0_LL	5.26	5.14	7.66	7.41	6.36	6.11	-	-
E3SM_1_0	5.26	4.68	-	-	-	-	-	-

Table 3. A table showing various sensitivity metrics estimated from the CMIP5 and CMIP6 ensembles (in K), using both pre-industrial average baseline temperatures (org) and baseline temperatures estimated from a regression fit to the first 20 years of the 1cptCO2 simulation (corr), where the linear fit is used to estimate temperatures and radiative fluxes at t=0. Warming is shown (where available) for corresponding RCP2.6 and RCP8.5 simulations, where the difference between 2001-2020 and 2081-2100 in the first ensemble member for the corresponding model is used to assess 21st century warming.

3 Conclusions

The question of which metric of climate sensitivity is most useful for summarizing uncertainty in future projections is conditional on a number of factors. Any single metric of sensitivity, even if known perfectly, cannot constrain Earth System response on all timescales and scenarios. We have shown here that one can produce a number of model variants which can exhibit the same value of EffCS or TCR, but with a range of responses, especially in a mitigation scenario such as RCP2.6.

In an idealized environment where models can be brought to a complete equilibrium control state, and ensemble sizes for '1pctCO2' simulations are large enough to avoid the effects of internal variability, the T140 metric would be the best idealized warming measure for century-scale warming under a high emissions scenario. However, the presence of even moderate control drift can act as a significant source of error in the measurement of T140, and so here we find that EffCS is likely to be a more accurate practical sensitivity metric in Earth System Model applications where full equilibration is difficult to achieve.

EffCS itself has limitations, it is relatively insensitive to slow timescale feedbacks, which means that it poorly correlated with century-scale warming under RCP2.6 (where a large fraction of warming occurs due to slow feedback response to historical emissions), and for warming on multi-century timescales under a high emissions scenario (where concentrations stabilize post-2100). We find that a simple, but useful alternative is to simply use the mean warming from years 131-150 of the abrupt-4xCO2 simulation - which is comparably skilled to EffCS in predicting RCP8.5 warming in 2100, but more sensitive to century timescale feedbacks than EffCS - so therefore it is better correlated with RCP2.6 end of century warming.

It is notable that the most common metrics of sensitivity, EffCS, T140 and TCR, provide very little guidance on peak warming expected under climate mitigation. The focus on these metrics has also given rise to the issue that slow feedbacks in Earth System Models are not well constrained by the set of experiments currently conducted by default in CMIP. The standard 150 year simulation used to calculate Effective Climate Sensitivity does not constrain true Equilibrium Climate Sensitivity, and only a limited set of CMIP-class models have run models for long enough to be informative about equilibrium response (Rugenstein et al., 2019).

It should be noted that these conclusions are derived from the consideration of a relatively simple two-timescale pulse response model. In this model, we can show that certain sensitivity metrics are insufficient to constrain future projections, and that non-equilibration may confound measurement. However, the constrained distributions for the metrics are subject to the structural assumptions of the model. The real world may have more than two response timescales ~~Aengenheyster et al. (2018)~~(Aengenheyster et al., 2018), or may be better described as a continuous sum ~~Ragone et al. (2016); Lembo et al. (2019)~~(Ragone et al., 2016
. Further work should identify how such complexity impacts uncertainty in relevant climate metrics.

The diversity of simulated global mean dynamical response to greenhouse gas forcing over the coming centuries can be represented in simple models with a relatively small number of parameters (Smith et al., 2018; Meinshausen et al., 2011), but we cannot reduce uncertainty in climate projections on all timescales to a single degree of freedom. Summary metrics of climate response have value if the context of those metrics (and their range of applicability in relation to projection uncertainty) is well understood, but their limitations should be kept in mind.

Data availability. CMIP5 and CMIP6 data are available through a distributed data archive developed and operated by the Earth System Grid Federation (ESGF).

Code and data availability. Code for this study is available on Github at https://github.com/benmsanderson/matlab_pulse

Author contributions. The author performed all analysis and writing for this project

Competing interests. The author declares no competing interests

Acknowledgements. This work is funded by the French National Research Agency, project number ANR-17-MPGA-0016. Benjamin Sanderson is an affiliate scientist with the National Center for Atmospheric Research, sponsored by the National Science Foundation.

References

Aengenheyster, M., Feng, Q. Y., Van Der Ploeg, F., and Dijkstra, H. A.: The point of no return for climate action, Earth System Dynamics, 9, 2018.

Andrews, T., Gregory, J. M., and Webb, M. J.: The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models, Journal of Climate, 28, 1630–1648, 2015.

Andrews, T., Gregory, J. M., Paynter, D., Silvers, L. G., Zhou, C., Mauritsen, T., Webb, M. J., Armour, K. C., Forster, P. M., and Titchner, H.: Accounting for changing temperature patterns increases historical estimates of climate sensitivity, Geophysical Research Letters, 45, 8490–8499, 2018.

Armour, K. C., Bitz, C. M., and Roe, G. H.: Time-varying climate sensitivity from regional feedbacks, Journal of Climate, 26, 4518–4534, 2013.

Caldeira, K. and Myhrvold, N.: Projections of the pace of warming following an abrupt increase in atmospheric carbon dioxide concentration, Environmental Research Letters, 8, 034 039, 2013.

Charney, J., Arakawa, A., Baker, D., Bolin, B., Dickinson, R., Goody, R., Leith, C., Stommel, H., and Wunsch, C.: Carbon Dioxide and Climate: A Scientific Assessment: Report of an Ad Hoc Study Group on Carbon Dioxide and Climate, Woods Hole, Massachusetts, July 23-27, 1979 to the Climate Research Board, Assembly of Mathematical and Physical Sciences, National Research Council, National Academies, 1979.

Foreman-Mackey, D., Hogg, D. W., Lang, D., and Goodman, J.: emcee: the MCMC hammer, Publications of the Astronomical Society of the Pacific, 125, 306, 2013.

Geoffroy, O., Saint-Martin, D., Bellon, G., Voldoire, A., Olivié, D., and Tytéca, S.: Transient climate response in a two-layer energy-balance model. Part II: Representation of the efficacy of deep-ocean heat uptake and validation for CMIP5 AOGCMs, Journal of Climate, 26, 1859–1876, 2013.

Goodman, J. and Weare, J.: Ensemble samplers with affine invariance, Communications in applied mathematics and computational science, 5, 65–80, 2010.

Goodwin, P., Katavouta, A., Roussenov, V. M., Foster, G. L., Rohling, E. J., and Williams, R. G.: Pathways to 1.5 C and 2 C warming based on observational and geological constraints, Nature Geoscience, 11, 102–107, 2018.

Gregory, J., Ingram, W., Palmer, M., Jones, G., Stott, P., Thorpe, R., Lowe, J., Johns, T., and Williams, K.: A new method for diagnosing radiative forcing and climate sensitivity, Geophysical Research Letters, 31, 2004.

Grose, M. R., Gregory, J., Colman, R., and Andrews, T.: What Climate Sensitivity Index Is Most Useful for Projections?, Geophysical Research Letters, 45, 1559–1566, 2018.

Hansen, J., Lacis, A., Rind, D., Russell, G., Stone, P., Fung, I., Ruedy, R., and Lerner, J.: Climate sensitivity: Analysis of feedback mechanisms, Climate processes and climate sensitivity, 29, 130–163, 1984.

Hasselmann, K., Sausen, R., Maier-Reimer, E., and Voss, R.: On the cold start problem in transient simulations with coupled atmosphere-ocean models, Climate Dynamics, 9, 53–61, 1993.

Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., and Vallis, G. K.: Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing, Journal of Climate, 23, 2418–2427, 2010.

Joos, F., Roth, R., Fuglestvedt, J., Peters, G., Enting, I., Bloh, W. v., Brovkin, V., Burke, E., Eby, M., Edwards, N., et al.: Carbon dioxide and climate impulse response functions for the computation of greenhouse gas metrics: a multi-model analysis, Atmospheric Chemistry and Physics, 13, 2793–2825, 2013.

Knutti, R. and Rugenstein, M. A.: Feedbacks, climate sensitivity and the limits of linear models, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 373, 20150 146, 2015.

Knutti, R., Rugenstein, M. A., and Hegerl, G. C.: Beyond equilibrium climate sensitivity, Nature Geoscience, 10, 727, 2017.

Lembo, V., Lunkeit, F., and Lucarini, V.: TheDiaTo (v1. 0)–a new diagnostic tool for water, energy and entropy budgets in climate models, Geoscientific Model Development, 12, 3805–3834, 2019.

Lewis, N.: An objective Bayesian improved approach for applying optimal fingerprint techniques to estimate climate sensitivity, Journal of Climate, 26, 7414–7429, 2013.

Lucarini, V.: Revising and extending the linear response theory for statistical mechanical systems: evaluating observables as predictors and predictands, Journal of Statistical Physics, 173, 1698–1721, 2018.

Lucarini, V., Ragone, F., and Lunkeit, F.: Predicting climate change using response theory: Global averages and spatial patterns, Journal of Statistical Physics, 166, 1036–1064, 2017.

MacKay, D. J. and Mac Kay, D. J.: Information theory, inference and learning algorithms, Cambridge university press, 2003.

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungclaus, J., Klocke, D., Matei, D., et al.: Tuning the climate of a global model, Journal of advances in modeling Earth systems, 4, 2012.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M., Lamarque, J.-F., Matsumoto, K., Montzka, S., Raper, S., Riahi, K., et al.: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, Climatic change, 109, 213, 2011.

Millar, R. J., Nicholls, Z. R., Friedlingstein, P., and Allen, M. R.: A modified impulse-response representation of the global near-surface air temperature and atmospheric concentration response to carbon dioxide emissions, Atmospheric Chemistry and Physics, 17, 7213–7228, 2017.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, Journal of Geophysical Research: Atmospheres, 117, 2012.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhang, H.: Anthropogenic and natural radiative forcing, pp. 659–740, Cambridge University Press, Cambridge, UK, https://doi.org/10.1017/CBO9781107415324.018, 2013.

Otto, A., Otto, F. E., Boucher, O., Church, J., Hegerl, G., Forster, P. M., Gillett, N. P., Gregory, J., Johnson, G. C., Knutti, R., et al.: Energy budget constraints on climate response, Nature Geoscience, 6, 415, 2013.

Proistosescu, C. and Huybers, P. J.: Slow climate mode reconciles historical and model-based estimates of climate sensitivity, Science Advances, 3, e1602 821, 2017.

Ragone, F., Lucarini, V., and Lunkeit, F.: A new framework for climate sensitivity and prediction: a modelling perspective, Climate Dynamics, 46, 1459–1471, 2016.

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., and Rafaj, P.: RCP 8.5—A scenario of comparatively high greenhouse gas emissions, Climatic Change, 109, 33, 2011.

Ruelle, D.: General linear response formula in statistical mechanics, and the fluctuation-dissipation theorem far from equilibrium, Physics Letters A, 245, 220–224, 1998.

Rugenstein, M., Bloch-Johnson, J., Gregory, J., Andrews, T., Mauritsen, T., Li, C., Frölicher, T., Paynter, D., Danabasoglu, G., Yang, S., Dufresne, J.-L., Cao, L., SchmidtG.A., Abe-Ouchi, A., Geoffroy, O., and Knutti, R.: Equilibrium climate sensitivity estimated by equilibrating climate models, Geophysical Research Letters, in press, 2019.

Rugenstein, M. A., Caldeira, K., and Knutti, R.: Dependence of global radiative feedbacks on evolving patterns of surface heat fluxes, Geophysical Research Letters, 43, 9877–9885, 2016.

Sanderson, B. M. and Knutti, R.: On the interpretation of constrained climate model ensembles, Geophysical Research Letters, 39, 2012.

Senior, C. A. and Mitchell, J. F.: The time-dependence of climate sensitivity, Geophysical Research Letters, 27, 2685–2688, 2000.

Smith, C. J., Forster, P. M., Allen, M., Leach, N., Millar, R. J., Passerello, G. A., and Regayre, L. A.: FAIR v1. 3: A simple emissions-based impulse response and carbon cycle model, Geoscientific Model Development, 11, 2273–2297, 2018.

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M. M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M.: Climate Change 2013 The Physical Science Basis, IPCC, 2013.

Tsutsui, J.: Quantification of temperature response to CO 2 forcing in atmosphere–ocean general circulation models, Climatic Change, 140, 287–305, 2017.

Van Vuuren, D. P., Stehfest, E., den Elzen, M. G., Kram, T., van Vliet, J., Deetman, S., Isaac, M., Goldewijk, K. K., Hof, A., Beltran, A. M., et al.: RCP2. 6: exploring the possibility to keep global mean temperature increase below 2 C, Climatic Change, 109, 95, 2011.

Winton, M., Takahashi, K., and Held, I. M.: Importance of ocean heat uptake efficacy to transient climate change, Journal of Climate, 23, 2333–2344, 2010.

Zanna, L., Khatiwala, S., Gregory, J. M., Ison, J., and Heimbach, P.: Global reconstruction of historical ocean heat storage and transport, Proceedings of the National Academy of Sciences, 116, 1126–1131, 2019.

Appendix : Supplementary Material

Figure S1. Plots illustrating how different types of sensitivity metric are influenced by climatological drift. Each line describes the evolution of the model (with default parameters), where the control simulation is initialized 500 years in advance of the sensitivity experiment with a non-zero forcing ranging from -1 to 1 Wm^{-2}. (a) shows the global mean temperature time evolution of the abrupt4xCO2 simulations (blue) and the 1pctCO2 simulation (green), with box-whisker plots showing the range of biased values which are measured due to climate drift for A140, T140 and EffCS. (b) shows the trend lines used to compute the EffCS estimates from the simple model. Blue lines show an example model configuration response to an abrupt 4xCO2 perturbation in for the equilibrated case (dashed), and end members (+/- $1\mathrm{Wm}^{-2}$ imbalance). The red shaded area shows the range of fitted trend lines consistent with (a).

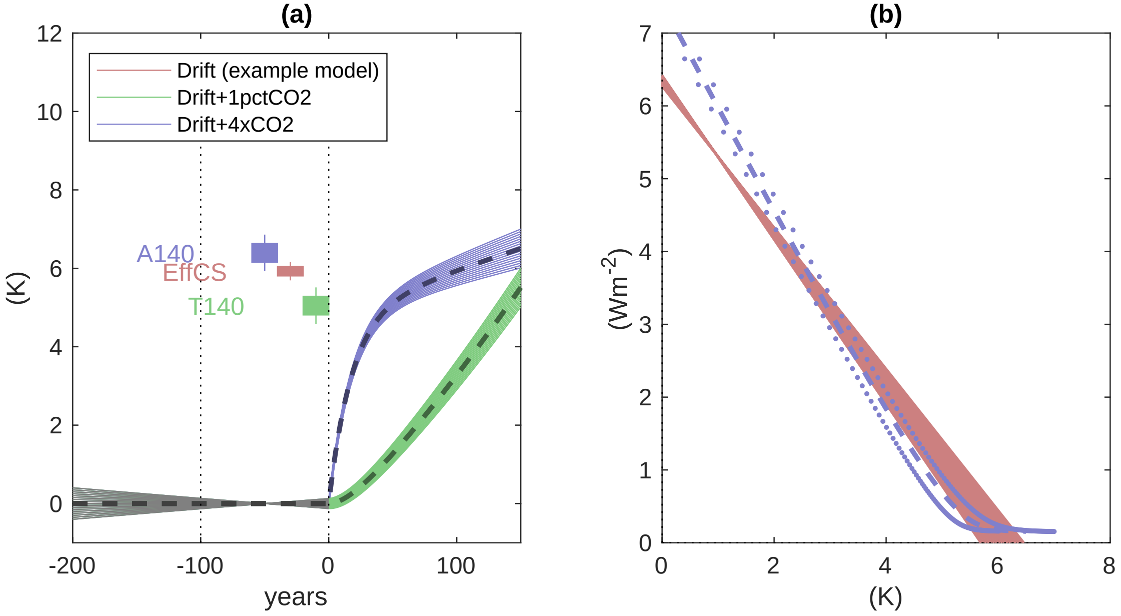
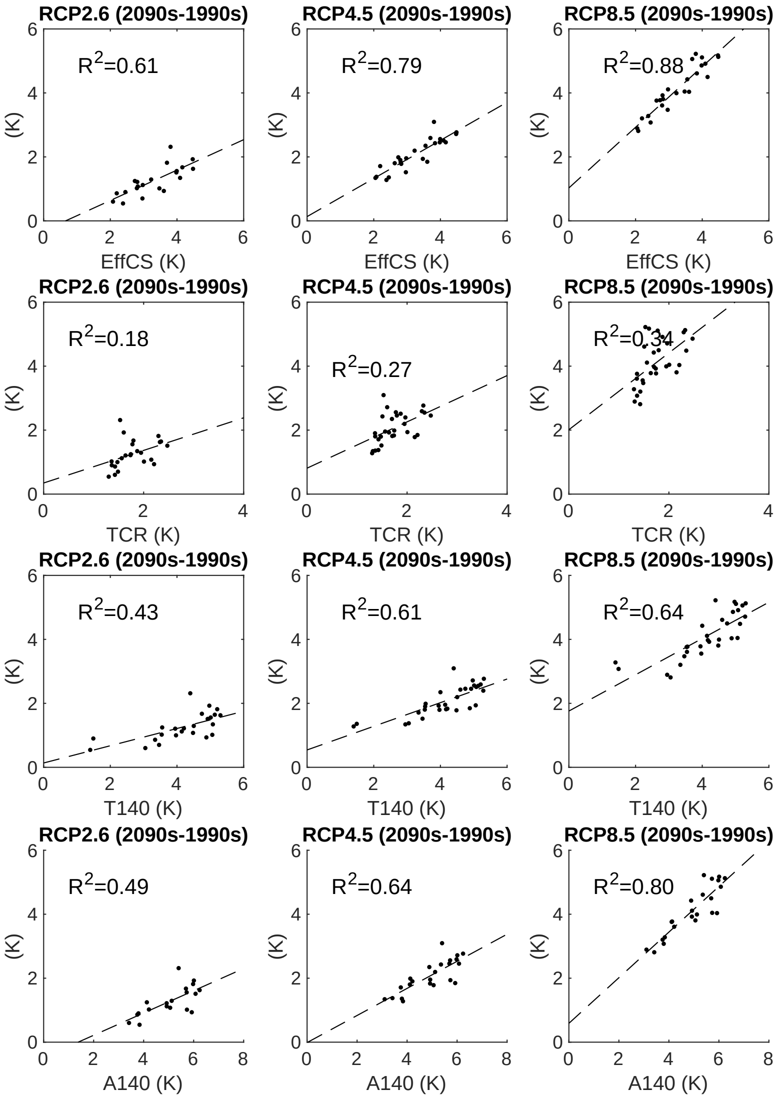


Figure S2. Scatterplots of 21st century warming (difference between 20 year means in 2081-2100 and 1981-2000) and a range of sensitivity metrics for CMIP5. TCR, T140 and EffCS are reported values from ~~(Stocker et al., 2013)~~Stocker et al. (2013), A140 is calculated as the year 131-150 average global mean temperature above the control level (taken as the last 100 years of the relevant control simulation). Columns represent different RCPs, rows represent different sensitivity metrics considered in the text. Each point represents a single model from the archive. Only results from the 1st initial condition ensemble member are considered for each model (thus the plots are subject to initial condition variability).



~~A ‘corner-plot’ showing the posterior parameter distribution attained by MCMC calibration of the simple climate model. Diagonal plots show posterior histograms for parameter values optimized in the calibration, while the horizontal range indicates the bounding values of the initial flat prior distribution. Off-diagonal plots show pairwise distributions of parameters in the posterior distribution.~~

Figure S3. A demonstration of the simple model fitting strategy applied to historical simulations for a range of models in the CMIP5 archive. A pulse-response model is fitted treating each model’s global mean temperature output in turn as truth for the period 1870-2019 (black line). 10th-90th percentiles of fitted temperature response for historical (grey area) and future projections are shown for RCP8.5 (pink area) and RCP2.6 (blue area) concentration pathways. Dotted lines show the median temperature in the ensemble projection, while solid colored lines show the evolution of the actual GCM for the corresponding scenario.

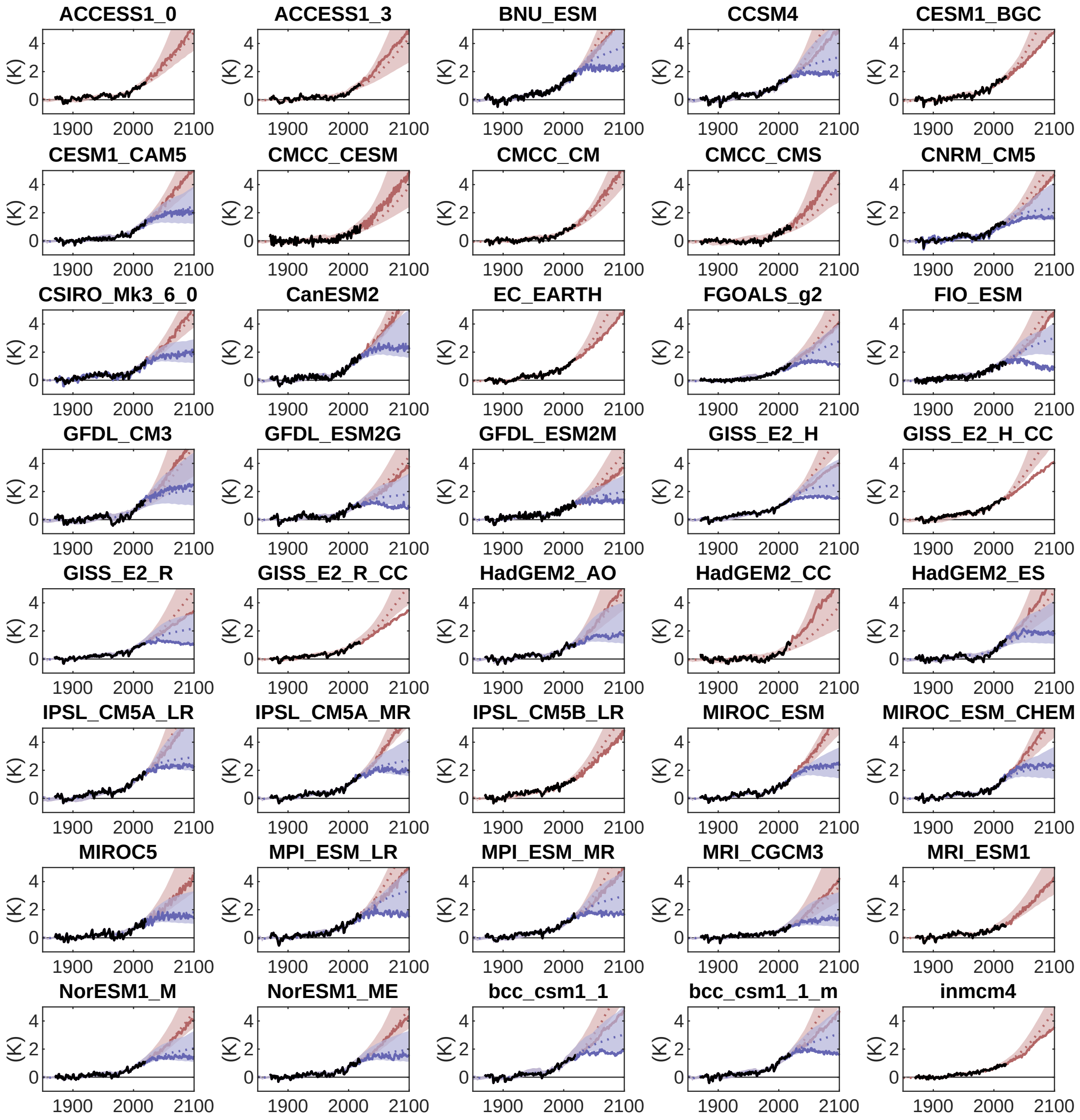


Figure S4. Figure illustrating the 'correction' employed for TCR and ECS in Figure 6. Corrected baseline temperatures are estimated by regression of the first 20 years of the control simulation, and branch-point from the control simulation is identifying by finding the year in which a linear fit to the control model evolution intersects the corrected baseline temperature. Branching in cases where there is no intersection are illustrated by the year in which the trendline is closest to the corrected baseline (either the first or last year).

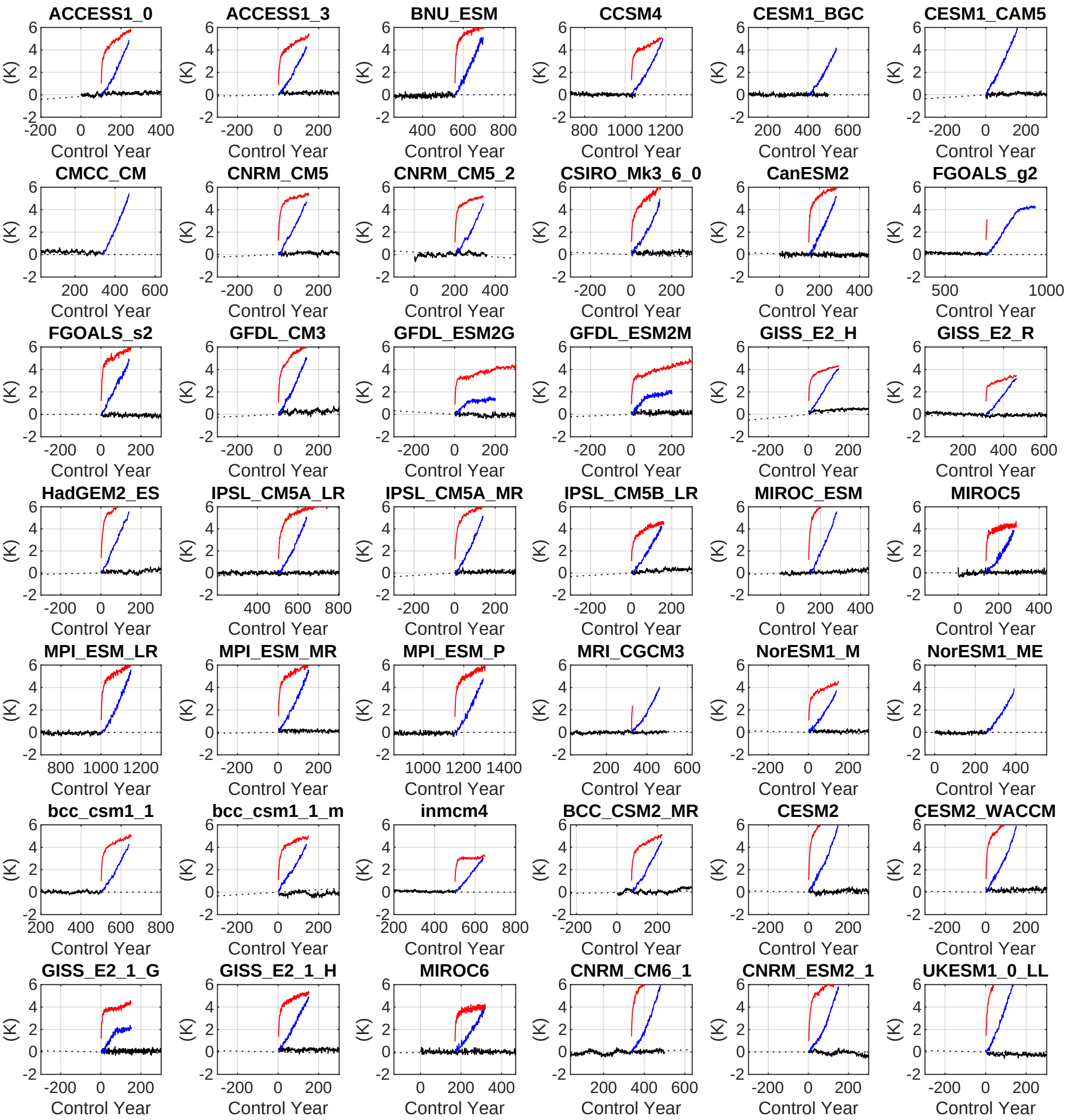
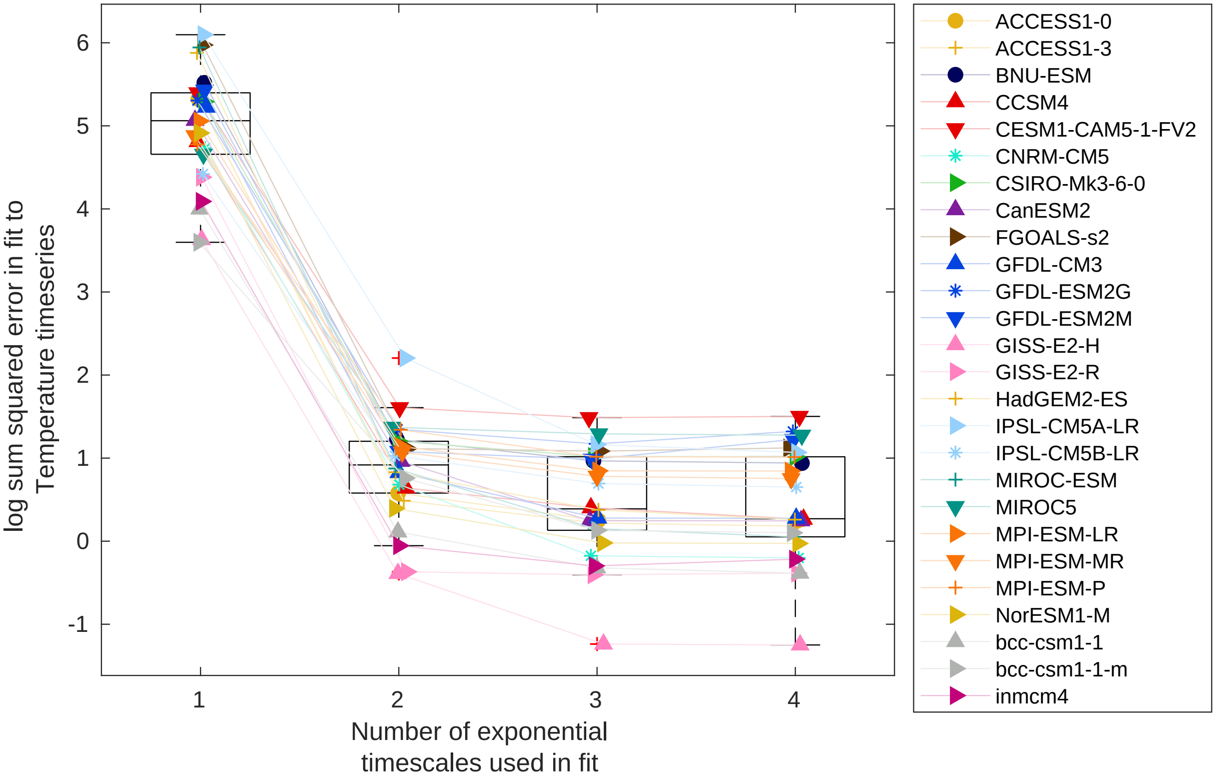


Figure S5. A figure illustrating the log of the sum squared error in best-fitting of the first 140 years of the abrupt4xCO2 time series in global mean temperature for each available model in the CMIP5 ensemble, as a function of the number of exponential modes allowed in the pulse response model. Each point shows the error in fitting one model in the CMIP5 ensemble, with colored lines tracking the error in the fitting as a function of number of allowed modes.


log sum squared error in fit to
Temperature timeseries
Number of exponential
timescales used in fit
ACCESS1-0
ACCESS1-3
BNU-ESM
CCSM4
CESM1-CAM5-1-FV2
CNRM-CM5
CSIRO-Mk3-6-0
CanESM2
FGOALS-s2
GFDL-CM3
GFDL-ESM2G
GFDL-ESM2M
GISS-E2-H
GISS-E2-R
HadGEM2-ES
IPSL-CM5A-LR
IPSL-CM5B-LR
MIROC-ESM
MIROC5
MPI-ESM-LR
MPI-ESM-MR
MPI-ESM-P
NorESM1-M
bcc-csm1-1
bcc-csm1-1-m
inmcm4