Response to reviewer 1.

We thank the reviewer for the comments, which have lead to us further clarifying the methods and conclusions of our study. Our responses are below in bold.

Major issues:

1) Section 2.1 is used to formally argue for using 1975 as a starting year rather than 1970 as used in most other studies. However, based on the description I was unable to understand what is being plotted, but if I am right, I don't think it is relevant. I think what is plotted in panel (b) is the standard deviation of forcing with respect to pre-industrial, given that it rises with time. However, for the purpose of the emergent constraint here, this is completely irrelevant, rather what is interesting as a measure of noise is the standard deviation in the forcing change *after* the given year. Likewise, the signal should be the forcing change afterwards.

Reply: although we choose a default start year of 1975, in Figure 4c we have also assessed the robustness of our constraint to different start years (including 1970 as used by JM19). In the revised manuscript we point more clearly to this sensitivity study. We have also clarified the caption of Figure 1 as requested.

2) I remain completely confused as to why the authors apply both a running mean and block averaging. In their reply they state that no data outside the blocks is used. If this is true, then years within the block will have been given different weight in order to have results depend on the running mean length. Also, how is it possible to apply a 20-year running mean (Figure 5b) within a block of only 11 years?`

Reply: we do not separately "apply both a running mean and block averaging" – these are two different descriptions for the same procedure. We have made this clearer in the revised manuscript, by adding: "*(or equivalently, the difference in smoothed T between 2014 and 1980)".*

3) In linear statistical model used to estimate TCR an offset is assumed (lines 145-147). As for the case of ECS, this offset should be around zero since a world with zero TCR should experience zero warming. Nevertheless, the offset is very large 0.8 or 0.6 K. What is the meaning of this, and why not assume up front that the offset is close to zero?

Reply: the nonzero offset (regression dilution) is a consequence of minimizing the error in the dependent variable. Using ORD, we don't find an offset significantly different from zero, so we don't attach any physical significance to it. When trying to make a prediction, as we are in this study, it is better to allow a non-zero offset (which implies a 2-parameter rather than a 1-parameter fit to the emergent relationship), to allow for a possible model misspecification (Hahn, G. J. (1977). J. Qual. Techn. 9(2)). We make these issues clearer in our revised manuscript.

4) I found the discussion of the different regression methods somewhat one sided. The authors argue against Orthogonal distance regression, but why would one assume zero error in the post-1975 warming estimates from models? Intuitively I would think that the relative error there should actually be larger than the error in estimating TCR.

Reply: we do not assume zero error in the post-1975 warming. Similar to OLS, we minimize the error in the dependent variable. Our methodology for taking account of error in the post-1975 warming estimates is discussed in the Appendix. We now point to this more clearly in both our methods and results sections.

5) Perhaps not a very major point, but the paper uses many different uncertainty ranges, 16-83, 16-84, 5-95, or even 90 percent confidence with unspecified range. IPCC's 'likely' statement means greater than 66 percent probability, not the 16-84 percentile range. Sometimes these are taken to be the same, but they are not. Practically all studies with which results are compared provide 5-95 percent confidence intervals, and I would recommend that this range is primarily used. I understand that the authors might also want to compare with IPCC, but then they should also state that the 'likely' range is not the same as that provided, but a lower-bound. IPCC can and will add various types of uncertainty not accounted for here and obtain broader likely ranges than found in individual studies.

Reply: the 16-83% was a typo, which is now corrected. We thank the reviewer for pointing this out. We have also replaced the 90 with 5-95%, and clarified other instances that this is the primary use.

6) The authors should explain that it is not obvious how to handle the cases of negative ECS with large dT. By removing these cases the error distribution is inevitably skewed and therefore biased towards higher ECS. This is arguably equally unreasonable as keeping the cases in. A reasonable alternative is to remove an equal number of cases from the high end of the error distribution to alleviate the bias.

Reply: the negative fitted ECS values are an artifact of the assumptions of no deep ocean temperature rise. We have expanded the discussion of negative ECS to explain this point.

7) I was confused over the discussion of the different model fits (lines 241-259), which doesn't seem to lead to any new conclusions. Perhaps I missed something, though, the authors might want to clarify what they learned from the exercise.

Reply: we have rewritten the results section to highlight the motivation for the different model fits. We have also expanded the explanation in the methods section by adding (to line 164): '*which, if too steep, could lead to an upper estimate of ECS biased high*'.

Minor or technical issues:

31, This sentence is confusing, to first order the reason the ratio is decreasing is because of deep ocean heat uptake. Pattern effects make this effect stronger.

Reply: we have included '*deep ocean heat uptake*'.

36, Perhaps specify that AR5 used models and historical warming, rather than 'multiple lines of evidence'.

Reply: for TCR they also used detection and attribution studies, and reconstructions of RF (see p.1112 AR5 WGI). For conciseness, we prefer the current wording.

40, The Gregory et al. paper was published 2019.

Reply: corrected, thank you.

76, Perhaps reference to Kiehl (2007) would be appropriate.

Response: done.

80, Which 'last decade'?

Response: changed to '*recent*', as the source is from 2017, so doesn't capture 2017-2019.

105, I guess you also need abrupt4xCO2

Response: yes, added.

140, I would suggest to use broadly used nomenclature 'ocean heat uptake efficacy (representing a pattern effect) and \gamma the deep ocean heat uptake parameter'

Response: done.

143, what is 'JM19'?

Response: now defined.

155, perhaps refer to Hansen et al. 1985 for unrealised warming?

Response: done.

179-180, These specific models can not be distinguished in the plot.

Response: distinguishing individual models in the plot would make it overly complicated, in our view.

181, which 'dip'?

Response: we have added '*temperature*' to be more specific (refers to previous sentence).

185-186, how was the confidence interval calculated?

Response: **this is explained in the Appendix, which is now more explicitly pointed-to from the main manuscript.**

210-211, It seems to me that the uncertainty increases (surprisingly) little with increasing start year (Fig. 5c). Instead it is the median that increases. Also, there is no reason the uncertainty range should revert to the CMIP6 range, it can be much broader if things are done correctly: Imagine an ensemble with very little spread in TCR, then the uncertainty in the relationship should be very large and the posterior spread of TCR too.

Reply: as a consequence of using linear regression, the uncertainty range does revert to the range. In the case of low spread in TCR, the slope parameter will get close to zero, and the uncertainty around the slope will encompass the entire ensemble spread. No matter the observation, the result will revert to the ensemble statistics. Only (statistical) model uncertainty can lead to larger spread.

Figure 5c, which 'Model range'?

Response: we have included an explanation in the caption: '*For comparison, the 5-95% model range and IPCC range are shown, both assuming normal distributions'.*

Figure 6b, greens are hard to distinguish, they look the same in print. Also yellow is a poor choice. There are many other colours to choose from.

Response: we changed the colours; yellow to light brown, light green to dark red. We also changed the markers for the alternative models to crosses, to further emphasize that the dark green dots correspond to the dark green dots on the left-hand panel.

279, delete 'tests'

Response: done.

290, I believe this should be 1.9-3.4 K

Response: corrected.

Response to reviewer 2.

We thank the reviewer for their review, who prevented a silly mistake from being printed and whose suggestions mathematics more transparent. Answers in bold.

Page 7 must be rewritten as follows (eq.3 is obviously wrong! It must include s, not s prime and be in the form of JM19's Eq.5).

Response: we have corrected eq.3

Lines 143-147:

…. We follow the approximations in Williamson et al. (2018) and JM19 in assuming no change in deep ocean temperature, and assuming the upper ocean to be in equilibrium. These assumptions are reasonable for timescales larger than a decade, but smaller than a century (JM19). As a result (see JM19) we have:

$\Delta T0$=0, $\Delta T(\lambda+\varepsilon\gamma)=F$ and $TSR(\lambda+\varepsilon\gamma)=F2x$. Consequently,

$TSR=F2x/(\lambda+\varepsilon\gamma)=s\Delta T$. (3)

Here s is a forcing parameter, defined as $s=F2x/F$. The choice of …..

Response: we have included the extra equations in a slightly different wording.

Lines 155-156:

. …. (Rugenstein et al., 2019). Using (2) and assuming both upper and deep ocean to be in equilibrium we have $\Delta T0=\Delta T$ and $ECS=F2x/\lambda$. Expressing λ using (3) one can relate TCR and ECS via: …..

Line 159:

Where $s'=F/F2x$ and….

Response: we have now included the definition of ECS, the reference to eq (2) and reiterated the assumptions.

Emergent constraints on TCR and ECS from simulated historical warming in CMIP5 and CMIP6 models

Femke J.M.M. Nijsse[1], Peter M. Cox[1], and Mark S. Williamson[1, 2]

[1]University of Exeter, College of Engineering, Mathematics and Physical Sciences, EX4 4QE, Exeter, UK
[2]University of Exeter, Global Systems Institute, EX4 4QE, Exeter, UK

Correspondence: Femke J.M.M. Nijsse (f.j.m.m.nijsse@exeter.ac.uk)

Abstract. Climate sensitivity to CO_2 remains the key uncertainty in projections of future climate change. Transient climate response (TCR) is the metric of temperature sensitivity that is most relevant to warming in the next few decades, and contributes the biggest uncertainty to estimates of the carbon budgets consistent with the Paris targets. Equilibrium climate sensitivity (ECS) is vital for understanding longer-term climate change and stabilization targets. In the IPCC 5th Assessment Report (AR5), the stated 'likely' ranges (16–84% confidence) of TCR (1.0–2.5 K) and ECS (1.5–4.5 K) were broadly consistent with the ensemble of CMIP5 Earth System Models (ESMs) available at the time. However, many of the latest CMIP6 ESMs have larger climate sensitivities, with 5 of 34 models having TCR values above 2.5 K, and an ensemble mean TCR of 2.0 $\pm$ 0.4 K. Even starker, 12 of 34 models have an ECS value above 4.5 K. On the face of it, these latest ESM results suggest that the IPCC likely ranges may need revising upwards, which would cast further doubt on the feasibility of the Paris targets.

Here we show that rather than increasing the uncertainty in climate sensitivity, the CMIP6 models help to constrain the likely range of TCR to 1.3–2.1 K, with a central estimate of 1.68 K. We reach this conclusion through an emergent constraint approach which relates the value of TCR linearly to the global warming from 1975 onwards. This is a period when the signal-to-noise ratio of the net radiative forcing increases strongly, so that uncertainties in aerosol forcing become progressively less problematic. We find a consistent emergent constraint on TCR when we apply the same method to CMIP5 models. Our constraints on TCR are in good agreement with other recent studies which analysed CMIP ensembles. The relationship between ECS and the post-1975 warming trend is less direct and also non-linear. However, we are able to derive a likely range of ECS of 1.9–3.4 K from the CMIP6 models by assuming an underlying emergent relationship based on a 2-box energy balance model. Despite some methodological differences, this is consistent with a previously-published ECS constraint derived from warming trends in CMIP5 models to 2005. Our results seem to be part of a growing consensus amongst studies that have applied the emergent constraint approach to different model ensembles and to different aspects of the record of global warming.

1 Introduction

The key uncertainty in projections of future climate change continues to be the sensitivity of global mean temperature to changes in the Earth's energy budget, called 'radiative forcing'. This sensitivity is usually characterised in terms of the global

mean temperature that would occur if the atmospheric carbon dioxide concentration was doubled, for which the radiative forcing is reasonably well-known.

Two related quantities are used to characterise the climate sensitivity of Earth System Models (ESMs). Equilibrium climate sensitivity (ECS) is an estimate of the eventual steady-state global warming at double CO_2. Transient climate response (TCR) is the mean global warming predicted to occur around the time of doubling CO_2 in ESM runs for which atmospheric CO_2 concentration is prescribed to increase at 1% per year. Across an ensemble of ESMs, TCR values are less than ECS values because of deep ocean heat uptake, which leads to a lag in the response of global temperature to the increasing CO_2 concentration (Hansen et al., 1985). The ratio of TCR over ECS tends to decrease with increasing ECS, and depends on spatial pattern effects (Armour, 2017).

Despite decades of advances in climate science, the Earth's ECS and TCR remain uncertain. The 'likely' range of ECS (66% confidence limit) has been quoted as 1.5 K to 4.5 K in all of the five Assessment Reports (ARs) of the Intergovernmental Panel on Climate Change (IPCC) starting in 1990, aside from the fourth AR which moved the likely lower range temporarily to 2 K. Similarly the likely range of TCR is given as 1 K to 2.5 K in the IPCC AR5, based on multiple lines of evidence.

There have been numerous attempts to constrain ECS using the record of historical warming or palaeoclimate data (Knutti et al., 2017), and more recently using emergent constraints which relate observed climate trends, variations or other variables to ECS using an ensemble of models (Caldwell et al., 2018; Cox et al., 2018a). However, debate still rages about the likely range of ECS ~~(Brown et al., 2018; Bretherton and Caldwell, 2019; Cox et al., 2018b; ?)~~(Brown et al., 2018; Bretherton and Caldwell, 2019; Cox et , in part because observed global warming is a rather indirect measure of global warming at equilibrium. On the other hand, TCR is more closely related to the rate of warming, and therefore ought to be more amenable to constraint by the record of global warming (Bengtsson and Schwartz, 2013; Gregory and Forster, 2008; Jiménez-de-la Cuesta and Mauritsen, 2019; Tokarska et al., 2020). Nevertheless, the accepted likely range of TCR has also resisted change (Knutti et al., 2017), for reasons we will discuss in this paper. At the time of the AR5, the CMIP5 ESMs produced central estimates (mean $\pm$ stdev) of ECS (3.3 ± 0.7 K) and TCR (1.8 ± 0.3 K), that were broadly consistent with these IPCC likely ranges. However, there has been a general drift upwards towards higher climate sensitivities in the new CMIP6 ESMs, such that more than one third of the new CMIP6 models now have ECS values over 4.5 K (Forster et al., 2019), and five have TCR values over 2.5 K (Table 1). If the real climate system is similarly sensitive, the Paris climate targets will be much harder to achieve (Tanaka and O'Neill, 2018).

Therefore some key science- and policy-relevant questions arise:

(a) *Are such high climate sensitivities consistent with the observational record?*

(b) *If so, do the CMIP6 models demand an upward revision to the IPCC likely ranges for climate sensitivity?*

We address these questions in this paper by evaluating the historical simulations of global warming from the CMIP6 models. In particular, we explore an emergent constraint on TCR based on global warming from 1975 onwards (Jiménez-de-la Cuesta and Mauritsen, 2019; Tokarska et al., 2020), but using the CMIP6 models and observational data up to 2019.

Emergent constraints are increasingly used to assess future change by exploiting statistical relationships in multimodel ensembles between an observable and a variable describing future climate (Cox et al., 2018a; Hall et al., 2019). In the work

presented here, we use the latest CMIP6 multimodel ensemble to define an emergent relationship between historical warming (expressed in terms of GMST, the observable) and TCR (the variable related to future climate). In line with published recommendations (Hall et al., 2019; Klein and Hall, 2015), we check the robustness of the resulting emergent constraint against the CMIP5 ensemble, using exactly the same methodology as for CMIP6. We also follow the suggestion of Hall et al. (2019) in striving to base the emergent constraint on sound physical reasoning.

From physical principles, we expect values of TCR to be very well-correlated with simulated global warming across a model ensemble. By definition, TCR is a measure of warming from a simulation that is driven by an exponential 1.0% per year increase in CO_2. Historical global warming has been driven by a qualitative similar forcing, albeit somewhat less rapid. Instead of 1.0%, CO_2, the atmospheric CO2 concentration has increased at about 0.5% per year since 2000 (Dlugokenchy and Tans, 2019)), augmented by additional positive radiative forcing from other well-mixed greenhouse gases, and partially offset by the cooling effects of anthropogenic aerosols.

The radiative effects of the rise in greenhouse gas concentrations are relatively well-known (Myhre et al., 2013), and are broadly similar in different ESMs. By contrast, the radiative forcing due to changes in anthropogenic aerosols, especially indirect effects via changes in cloud brightness and lifetime, are poorly constrained (Myhre et al., 2013; Bellouin et al., 2019).

These uncertainties in aerosol forcing have hindered attempts to constrain TCR or ECS from the rate of warming, especially during the pre-1980 period when the burning of sulphurous coal led to increases in CO_2 and increases in sulphate aerosols, that went up almost together (Andreae et al., 2005). As a result it has been difficult to distinguish, based purely on the observational record of global warming, between a model with high climate sensitivity and strong aerosol cooling, and a model with low climate sensitivity and weak aerosol cooling (Kiehl, 2007).

In order to minimise the effects of uncertainties in aerosol forcing, we need periods in which aerosol radiative forcing changes relatively little compared to the change in radiative forcing due to CO_2 and other well-mixed greenhouse gases. Fortunately, this applies to the decades after 1975 when total aerosol load from global SO_2 and NH_3 emissions were similar to ~~recent values~~ values over the last decade (Stevens et al., 2017). For this reason, we ~~follow Jiménez-de-la Cuesta and Mauritsen (2019) in focusing~~ focus on global warming since 1975. However, ~~in addition we explore a range of start and finish dates to assess~~ we also test the robustness of our ~~TCR constraint, and to test the hypothesis that the relationship between TCR and warming rate is emerging strongly now because of the declining importance of changes in aerosol forcing~~conclusions to different start dates (see Figure 5(c)), including the start year of 1970 as used by Jiménez-de-la Cuesta and Mauritsen (2019) (hereafter JM19).

To establish an emergent constraint on ECS, we investigate the appropriate functional form between observed warming and climate sensitivity. Due to the slow response of the ocean, this is not expected to be linear, and using a set of assumptions, ~~Jiménez-de-la Cuesta and Mauritsen (2019)~~ JM19 proposed an analytical form based on a two-layer box model. By computing the model parameters directly per model, we investigate the appropriateness of this analytical function, and use it to derive an emergent constraint.

The remainder of this paper is organised as follows: in Section 2 we describe our methodological choices; Section 3 contains the emergent constraints on TCR and ECS and Section 4 contains the discussion and conclusions. More technical details concerning the regression methods are given in the Appendix.

Table 1. List of CMIP6 models used in this study and their ~~ERF~~ effective radiative forcing at CO_2 doubling $F_{2\times}$, the climate feedback parameter λ, equilibrium climate sensitivity (ECS) and transient climate response (TCR). Mean values are reported for models with multiple realisations. The values of $F_{2\times}$, ECS and λ are computed using the Gregory method (Gregory, 2004). Models above the horizontal line were used in the extended simulations to 2019. Models below the line did not have SSP simulations available at time of writing. Consistently derived values for CMIP5 are displayed in the Supplementary Table 1.

Centre	**Model**	$F_{2\times}$	λ	**ECS**	**TCR**	**n**	Δ **T**	**std(Δ T)**
BCC	BCC-CSM2-MR	3.01	0.98	3.07	1.59	1	0.64	
CAMS	CAMS-CSM1-0	3.95	1.71	2.31	1.72	1	0.44	
CAS	FGOALS-f3-L	3.95	1.31	3.03	2.01	1	0.70	
CCCma	CanESM5	3.63	0.64	5.66	2.66	50	1.27	0.10
CNRM-CERFACS	CNRM-CM6-1	3.54	0.72	4.94	2.08	10	0.73	0.11
CNRM-CERFACS	CNRM-ESM2-1	3.09	0.66	4.66	1.92	5	0.65	0.15
CSIRO-ARCCSS	ACCESS-CM2	3.21	0.67	4.81	2.00	1	0.77	
CSIRO	ACCESS-ESM1-5	2.71	0.68	3.97	1.91	3	0.84	0.10
EC-Earth-Consortium	EC-Earth3-Veg	3.32	0.77	4.34	2.57	2	0.97	0.23
EC-Earth-Consortium	EC-Earth3	3.30	0.78	4.22	2.38	10	0.72	0.16
INM	INM-CM4-8	2.61	1.42	1.84	1.32	1	0.61	
INM	INM-CM5-0	2.88	1.49	1.93	1.40	1	0.55	
IPSL	IPSL-CM6A-LR	3.32	0.72	4.63	2.32	6	0.85	0.10
MIROC	MIROC-ES2L				1.55	1	0.62	
MIROC	MIROC6	3.76	1.47	2.56	1.52	3	0.50	0.04
MOHC	HadGEM3-GC31-LL	3.38	0.60	5.62	2.45	4	1.07	0.19
MOHC	UKESM1-0-LL	3.56	0.66	5.41	2.72	5	1.13	0.13
MPI-M	MPI-ESM1-2-HR	3.58	1.20	2.99	1.64	2	0.65	0.07
MRI	MRI-ESM2-0	3.36	1.07	3.14	1.56	5	0.73	0.06
NCAR	CESM2-WACCM	3.08	0.63	4.90	1.92	3	0.97	0.15
NCAR	CESM2	3.13	0.59	5.30	2.04	3	0.82	0.01
NCC	NorESM2-LM	3.06	1.13	2.69	1.46	3	0.63	0.18
NOAA-GFDL	GFDL-CM4	2.91	0.71	4.09	1.97	1	0.86	
NOAA-GFDL	GFDL-ESM4	3.51	1.31	2.68	1.53	2	0.79	0.15
NUIST	NESM3	3.73	0.78	4.76	2.73	2	0.93	0.17
UA	MCM-UA-1-0				1.94	1	0.81	
Mean		3.69	0.95	3.90	1.96	4.9	0.78	0.12
Standard deviation		0.40	0.34	1.18	0.42	9.4	0.19	0.06
AS-RCEC	TaiESM1				2.34			
BCC	BCC-ESM1	3.03	0.89	3.39	1.74			
E3SM-Project	E3SM-1-0	3.23	0.60	5.38	2.99			
NASA-GISS	GISS-E2-1-G	3.89	1.43	2.71	1.68			
NASA-GISS	GISS-E2-1-H	3.55	1.14	3.12	1.89			
MOHC	HadGEM3-GC31-MM	3.36	0.61	5.52	2.37			
MPI-M	MPI-ESM1-2-HR	3.58	1.20	2.99	1.64			
SNU	SAM0-UNICON	3.83	1.02	3.76	2.25			
Mean CMIP6		3.70	0.95	**3.91**	**2.01**			
Standard deviation CMIP6		0.39	0.33	1.17	0.42			
Mean CMIP5		3.58	1.06	**3.31**	**1.79**			
Standard deviation CMIP5		0.22	0.29	0.76	0.34			

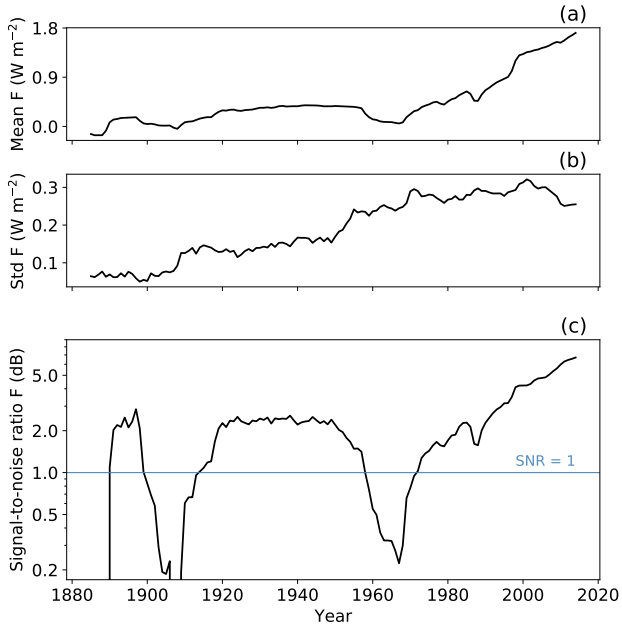


Figure 1. Effective radiative forcing over the historical period, calculated from 22 CMIP6 models as $F = \Delta N + \lambda \Delta T$: (a) ensemble mean; (b) ensemble standard deviation; (c) signal-to-noise ratio. Model means are calculated first, and then the ensemble mean is calculated.

2 Methodology

2.1 Choice of period over which to calculate warming trends

To constrain climate sensitivity using observed warming, we seek a period for which the forcing is relatively similar across models. In order to identify such a period we compute the effective radiative forcing F (ERF) for each model run using

$$F = \Delta N + \lambda \Delta T \tag{1}$$

following Forster et al. (2013). Here ΔN is the difference in net top of the atmosphere radiative flux and ΔT is the difference in near-surface temperature, both computed as global annual-mean anomalies relative to the initial state. We calculate the signal-to-noise ratio of F at each time as the model mean F divided by the standard deviation of F across the model ensemble.

Figure 1 shows how the signal to noise ratio of the ERF varies from 1880 to 2010. It is notable that the signal-to-noise ratio increases rapidly from around 1975, as relatively well-known greenhouse gas forcing continues to increase but the uncertain aerosol forcing begins to saturate. We have therefore focused our analysis on the post-1975 warming~~.~~, but we also performed a sensitivity analysis by varying the start year between 1960 and 2005.

2.2 Selection of CMIP6 model runs

We use all currently available CMIP6 models that have control (piControl), historical, a Shared Socioeconomic Pathway simulation (SSP1-2.6, SSP2-4.5, SSP3-7.0 or SSP5-8.5) and one percent CO_2 increase per year (1pctCO2) experiments. We extend the historical simulations from 2014 to 2019 using the Shared Socioeconomic Pathways (SSPs) scenario runs. Additional warming over this 5 year period varies very little across the SSPs, so we use SSP2-4.5 as this has the largest number of participating models at the time of writing.

2.3 Calculation of model sensitivity

From the 1pctCO2 experiment TCR is determined as the average temperature difference from the corresponding piControl run between 60 to 80 years after the start of the simulation (IPCC, 2013a). ECS is computed using the Gregory method (Gregory, 2004) on the first 150 year of the abrupt-4xCO2 simulations. The values of ECS and TCR that we derived are given in table 1.

2.4 Calculation of warming trend

Historical warming (our observable) is found from the historical and SSP simulations using the global annual mean surface air temperature (GMSAT) smoothed with a equally weighted running mean. Some of these models have multiple runs starting from different initial conditions, forcing time series or parameter settings. We use all available runs.

We use smoothed GMSAT to calculate warming. This is to limit the random effect of internal variability on the forced change we wish to constrain. We choose a centred 11-year running mean to remove shorter interannual and mid-term variability from sources such as ENSO, as well as reducing the effect of longer period modes of natural variability. We have tested the robustness of the constraint on TCR to the length of the running mean. It remains relatively invariant past a length of 8 years, suggesting most of the internal variability in GMSAT resides in shorter periods.

Warming ΔT is calculated as the difference in ~~smoothed~~ GMSAT between two periods, typically the years 1975–1985 and 2009–2019 (or equivalently, the difference in smoothed temperature between 2014 and 1980). We have chosen the end year to be 2019 to maximise the chances of discrimination between high and low sensitivity models. As the forcing from CO_2 increases with time, the warming in more sensitive models is more likely to diverge from less sensitive ones as we extend the period over which we calculate the trend. Extending to 2019 also allows us to include the most recent observational data and to eliminate possible effects from the warming slowdown between 2000–2012. This slowdown has been attributed to a combination of internal variability and decreased forcing, amongst other things (Medhaug et al., 2017). We assess the impact of the slow-down by comparing emergent constraints derived from time-series truncated to have different end years.

2.5 Theoretical basis

2.5.1 Transient Climate Response (TCR)

Once choices of length of running mean and start and end years for calculation of ΔT are fixed (our observable), we can fit an emergent relationship between the observable and our values of TCR via linear regression. Linear regression is performed using a hierarchical Bayesian model which can take into account all the different simulations per model: models with more simulations have a better-constrained post-1975 warming. This results in a set of 127 simulations from 26 different models. The regression method is further described in Appendix A. The choice of linear regression is justified by considering a two-layer energy balance model (Winton et al., 2010; Geoffroy et al., 2013a):

$$C\frac{dT}{dt} = F - \lambda T - \epsilon\gamma(T - T_0)$$
$$C_0\frac{dT_0}{dt} = \gamma(T - T_0). \tag{2}$$

Here $\cancel{\Delta T}$ T is the top layer temperature anomaly, $\cancel{\Delta T_0}$ T_0 the deep ocean temperature anomaly, λ ~~the climate sensitivity~~ is the climate feedback parameter, ϵ ~~the ocean pattern efficacy~~ is the ocean heat uptake efficacy (reflecting a pattern effect), and γ is the ocean heat uptake parameter (Winton et al., 2010). The parameters C and C_0 are the heat capacity of the upper ocean and deep ocean, respectively. We will refer to this model as EBM-ϵ, or EBM-1 if ϵ is set to 1. We follow the approximations in Williamson et al. (2018) and JM19 in assuming no change in deep ocean temperature ($T_0 = 0$), and assuming the upper ocean to be in equilibrium ($dT/dt = 0$). These assumptions are reasonable for timescales larger than a decade, but smaller than a century (see JM19), and lead to the following relationship:

$$\text{TCR} = s'\Delta T + \eta \tag{3}$$

Here $\cancel{s'}$ s is a forcing parameter, defined as ~~$F/F_{2\times}$ and η an offset .~~ $F_{2\times}/F$, and ΔT is the difference in temperature between two periods. For fitting, we include an offset η, so that TCR $= s\Delta T + \eta$, allowing for a possible model mis-specification and regression dilution (Hahn, 1977). A hierarchical linear regression was adopted which includes both uncertainty in ΔT and TCR (see Appendix). The choice of 1975 for the starting period minimises the uncertainty in our estimate of TCR. ~~As a function of start period~~However, uncertainty is relatively flat ~~and minimal between periods centred around~~ for starting periods between 1975 ~~to~~ and 1990. We ~~further investigate~~ also investigated the sensitivity of ~~this~~ our TCR constraint to the final year, the length of the running mean, the model selection, and the method of regression (see Figure 5).

2.5.2 Equilibrium Climate Sensitivity (ECS)

Similarly to the constraint on TCR, we use the warming between 1975–1985 and 2009–2019 to find an emergent constraint on ECS. The relationship between climate sensitivity and observed warming or TCR is not expected to be linear, as a smaller

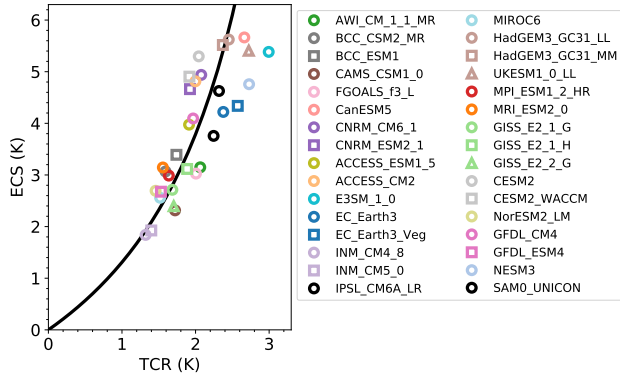


Figure 2. Scatter plot of TCR values plotted against ECS values for all CMIP6 models with both available at the time of submission. Models from the same modelling group are plotted with the same colour. Plot markers differentiate models from the same modelling centre. Black line uses the average ocean heat uptake parameters as fitted from the Geoffroy et al. (2013b) procedure, listed in Supplementary Table 3 and related ECS and TCR via: $\mathrm{ECS} = \mathrm{TCR}/(1 - e'\mathrm{TCR})$, with ~~e'=0.24~~$e' = 0.24$, the model mean.

fraction of equilibrium warming is typically realised in models with high climate sensitive within the first decades of warming ~~(Rugenstein et al., 2019). Using the same assumption as in the above section~~(Hansen et al., 1985; Rugenstein et al., 2019) . Using Eq. 2, $\mathrm{ECS} = F_{2\times}/\lambda$, and again assuming the upper ocean to be in equilibrium and the deep ocean temperature to not change, TCR and ECS are related via:

$$\mathrm{ECS} = \mathrm{TCR}/(1 - e'\mathrm{TCR}). \tag{4}$$

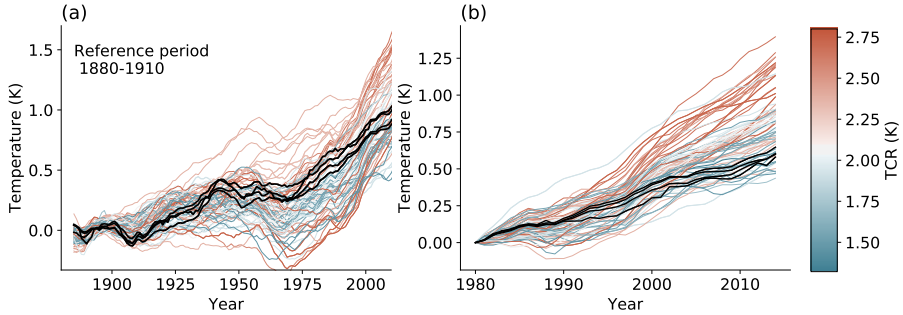
So that the relationship between ECS and ΔT ends up as:

$$\mathrm{ECS} = \frac{\Delta T}{s' - e'\Delta T}. \tag{5}$$

~~Again,~~ The forcing parameter is denoted by s'~~is a forcing parameter~~ , defined as $\Delta F/F_{2\times}$ and e' is the ocean heat uptake parameter defined as $\epsilon\gamma/F_{2\times}$, The function has an asymptote at $s' - e'\Delta T = 0$, and turns negative for larger ΔT values. As negative ECS values are unphysical, we modify the equation by keeping ECS at infinity for ~~$\Delta T > \frac{s'}{e'}$.~~ $\Delta T > s'/e'$. The appearance of negative ECS for high ΔT is an artefact of the no deep ocean temperature rise assumption: it corresponds to an equilibrium between the heating effect of $F - \lambda\Delta T$, balanced by $-\epsilon\gamma\Delta T$. In reality, this last term cancels completely with $\epsilon\gamma\Delta T_0$ at equilibrium.

To test the validity of these assumptions, we perform two checks. Firstly, by explicitly simulating the two box model, we investigate to what extent the analytical functional form deviates from the true functional form. We are especially interested in the upper region of this functional form, which, if too steep, could lead to an upper estimate of ECS biased high.

Figure 3. Global mean surface temperature of the 26 CMIP6 models named in Table 1. To avoid visual over-representation, a maximum of ten realisations per model are plotted. An 11-year running mean was used. (a) Temperature anomaly ΔT using 1880-1910 as reference period. (b) Temperature anomaly ΔT relative to the 1975–1985 mean.



Secondly, we fit the ocean heat uptake and forcing parameters for all CMIP6 models, following the two algorithms described in Geoffroy et al. (2013a, b), with slight modifications to ensure solutions exist for all models described in the Supplementary Information.

Using these fitting parameters, we investigate the physical basis of Eq 5 with the EBM-ϵ and EBM-1 models. If this function derived from the two-box model is a faithful representation, $\frac{\Delta T}{s'-e'\Delta T}$ $\Delta T/(s'-e'\Delta T)$ should be better related to ECS with individual model parameters than with the bulk fitted parameter. Figure 2 plots model TCR versus ECS, related via Eq. 4, using the ensemble mean of the fitted ocean parameters.

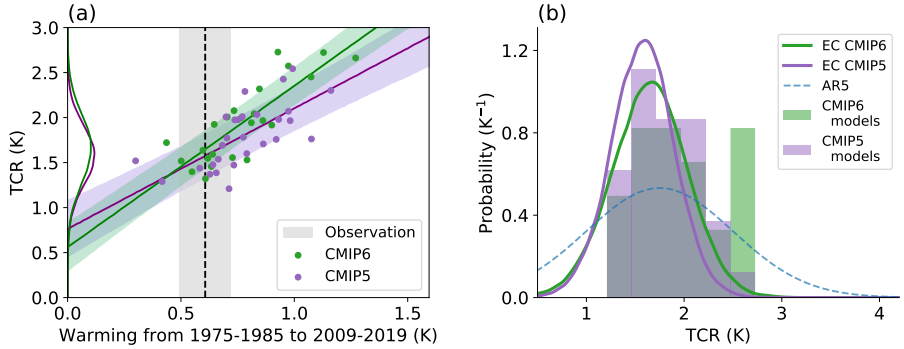
3 Results

3.1 Transient Climate Response (TCR)

Figure 3a shows the temperature anomaly over the period 1880 to 2019 simulated by 26 different CMIP6 models running a total of 127 simulations smoothed with a 11-year running mean. The reference period in this case is 1880-1910. Model runs have been colour coded by their TCR value, with darker red indicating models with higher TCR, and darker blue indicating lower TCR. Black lines are observational global warming datasets over the same period (Morice et al., 2012; Rohde et al., 2013; Lenssen et al., 2019; Zhang et al., 2019). Models with higher TCR either show large warming at the end of the period, or portray a strong aerosol cooling over the 20th century, particularly visible as a dip around 1960-1970 (notably CNRM-ESM1, UKESM1-0-LL and EC-Earth-Veg). Figure 3b shows the same information for the end of the historical period although the reference period is now chosen to be 1975-1985, after the temperature dip. The positive correlation intuitively expected between TCR and temperature increase ΔT is much clearer for this time interval.

The ΔT for each model simulation in Fig. 3b is used for the emergent constraint on TCR in Fig. 4a. Observational warming (black vertical dashed line) is the mean of HadCRUT4 (Morice et al., 2012), Berkeley Earth (Rohde et al., 2013), GISSTEMP4 (Lenssen et al., 2019) and NOAA v5 (Zhang et al., 2019). The 90% observational confidence interval (grey shaded vertical

Figure 4. **(a)** Emergent constraint on TCR against historical warming ΔT. ΔT is calculated from the difference between 1975-1985 and 2009-2019 of a timeseries of GMSAT. Linear regression is performed with all CMIP5 and CMIP6 simulations. Shaded areas indicate a 90% prediction interval (see Appendix A). The vertical dotted line is the mean value of the observations and y-axis shows the probability distribution of both generations of ensembles. **(b)** Comparison of probability distributions for the transient climate response using post-1975 warming using CMIP5 and CMIP6 simulations. The probability distribution in the fifth IPCC assessment is not fully specified, so the figure shows a normal distribution with the same likely range as IPCC.



area) is a combination of the observational uncertainty and the internal variability. To avoid double-counting observational uncertainty, the 90% regression confidence interval details the uncertainty of the best estimate of ΔT versus TCR (see Appendix for details). The models from the previous CMIP5 generation generally fall within the prediction interval of the CMIP6 emergent constraint: the emergent constraint is robust across generations (Klein and Hall, 2015). The best estimate (1.68 K) from this emergent constraint is higher than the best estimate using the larger set of models that have historical simulations up to 2014, but no future scenarios (median: 1.54 K, 5–95% range: 0.76–2.30 K). This can mostly be explained by the fact that 2004-2014 overlaps with the slow-down in surface temperature increase over the 2000-2012 period, but the wider range of models also impacts the regression.

Figure 4b shows the probability density functions (pdf) of TCR derived from the emergent constraint for both CMIP6 and the earlier CMIP5 model ensembles. For comparison, the raw model range in each CMIP is plotted as a histogram, as well as the reported IPCC AR5 likely range (assuming a normal distribution). Both CMIP5 and CMIP6 pdfs are very similar (central estimates differ by 0.1 K) even though CMIP6 contains many more high TCR models. As a continuation of the historical CMIP5 simulation, RCP8.5 is chosen. The tighter constraint in CMIP5 is mostly a consequence of differences in internal variability, which is 42% larger in CMIP6 than in CMIP5, in line with the findings of Parsons et al. (2020).

3.1.1 Period selection

Estimates of TCR depend on the final year chosen for the emergent constraint. Uncertainty in the estimate of TCR reduces as time increases and the central estimate converges as shown in Fig. 5a. Later end years are favoured as the signal-to-noise ratio of the net radiative forcing increases monotonically after 1975 (see Figure 1). In the 21st century, the climate impact of volcanoes has been dominated by smaller eruptions (Stocker et al., 2019). The scenarioMIP simulations used for 2015-2019 include a

small background forcing from volcanoes (O'Neill et al., 2016). We estimate errors from a potential mismatch between model and real forcing to be relatively small.

To mitigate the effect of internal variability, we use a running mean of GMSAT. Figure 5b shows the likely range of TCR as a function of the length of the running mean. Since we use all available simulations including multiple realisations of the same model in the emergent constraint, the effect of internal variability is already reduced and the length of the running mean on the estimate of TCR is small - the central estimate and the likely range remain relatively invariant past a window length of 8 years.

Figure 5c shows the effect of the start year on the emergent constraint. Uncertainty in the estimated value of TCR is ~~minimal and~~ relatively flat between start years of 1975 and 1990. Uncertainty ~~from start years of~~ for start years from 1990 onwards increases until the estimate and the uncertainty revert towards the raw CMIP6 ensemble statistics (no predictive power) at later years.

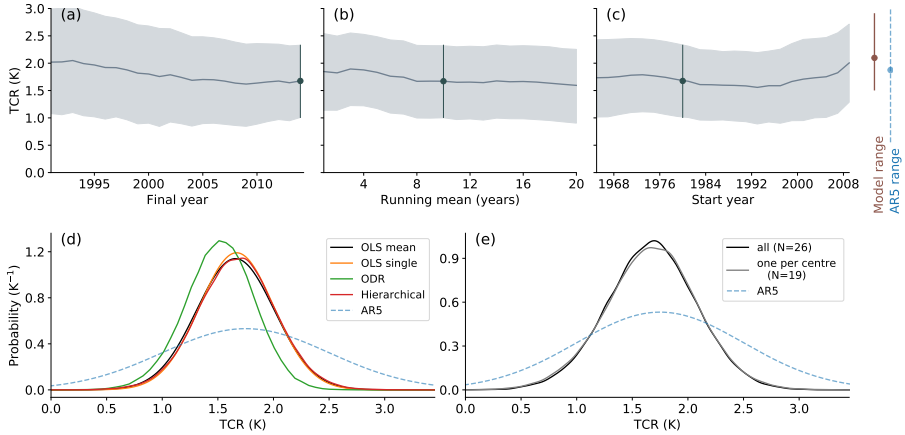
3.1.2 Regression method

When only one realisation per model is used for ordinary least square regression, regression dilution takes place in which the slope is underestimated (Cox et al., 2018b). This has the potential to lead to a slight overestimation of TCR (Fig. 5d), as the observed warming is on the lower end of the model range. ~~Jiménez-de-la Cuesta and Mauritsen (2019)~~ JM19 used the average warming for models with multiple simulations. As not all models provide a sufficient amount of simulations, they state that this leads to a minor inflation of the estimation of uncertainty. Although we use a hierarchical Bayesian model as the default (details in Appendix A) we have investigated three other regression methods used in the emergent constraint literature: ordinary least squares (OLS) with only one realisation per model, OLS on the mean warming per model and orthogonal distance regression (Fig. 5d). While the first three give very similar results, orthogonal distance regression gives a somewhat lower estimate of TCR. Orthogonal distance regression assumes that there are both errors in the predictor and in the predictand, which leads to a steeper slope. As our observation lies under the average, a steeper slope results in a smaller mean TCR value. Orthogonal distance regression is known to sometimes overcompensate for errors in the independent variable, for instance in the case the statistical model is not perfectly known; if the model deviates from being a perfectly straight line (Carroll and Ruppert, 1996).

3.1.3 Model selection

Model selection can prevent double counting of very similar models (Sanderson et al., 2015; Cox et al., 2018a). As models from the same centre can have very dissimilar climate sensitivities (Chen et al., 2014; Jiménez-de-la Cuesta and Mauritsen, 2019) and sensitivity can change drastically with only small adjustments to parameters (Zhao et al., 2016), we initially use all available models in the CMIP5 and CMIP6 ensemble. Figure 5e shows that this choice does not significantly change the best estimate of the transient response, and that using one model per modelling centre only very slightly increases the variance, even as models from one modelling centre are relative similar (Fig. 2).

Figure 5. Robustness of the result to various parameter choices and the choice of regression method. Unless stated differently, start year is 1975, all years up to 2018 are used, and the length of the running mean is 11 years. For comparison, the 5-95% model range range and IPCC range are shown, both assuming normal distributions. **(a)** 5-95% TCR range as a function of the final year (blue line central estimate). **(b)**: 5-95% TCR range as a function of length of running mean. **(c)** 5-95% TCR range as a function of start year. **(d)** Pdf of TCR from different regression methods: the hierarchical Bayesian model is compared to three other linear regression methods used in the emergent constraint literature: ordinary least squares (OLS) with only one realisation per model and OLS on the mean warming per model and orthogonal distance regression (ODR). **(e)** Resulting pdfs on TCR from stricter model selection (one model per modelling centre) compared to regression using all models and the IPCC AR5 range.



Study	Ensemble	Period	Median	5-95% range	16-84% range
Jiménez-de-la Cuesta and Mauritsen (2019)	CMIP5	1970 - 2005	~~1.67~~ 1.7 K	~~1.17~~ 1.2 – ~~2.16~~ 2.2 K	
Nijsse et al. (2020)	CMIP5	1970 - 2005	~~1.72~~ 1.7 K	~~1.11~~ 1.1 – ~~2.34~~ 2.3 K	~~1.37~~ 1.4 – ~~2.08 k~~ 2.1 K
Tokarska et al. (2020)	CMIP6	1981 - 2017	~~1.60~~ 1.6 K		1.2 – 2.0 K
Nijsse et al. (2020)	CMIP6	~~19750~~ 1975 - 2019	~~1.68~~ 1.7 K	~~1.001~~ 1.0 – ~~2.325~~ 2.3 K	~~1.293~~ 1.3 – ~~2.051~~ 2.1 K

Table 2. Emergent constraint on TCR depending on choices of ensemble and period. Results from ~~Jiménez-de-la Cuesta and Mauritsen (2019)~~ JM19 and Tokarska et al. (2020) are also shown for comparison.

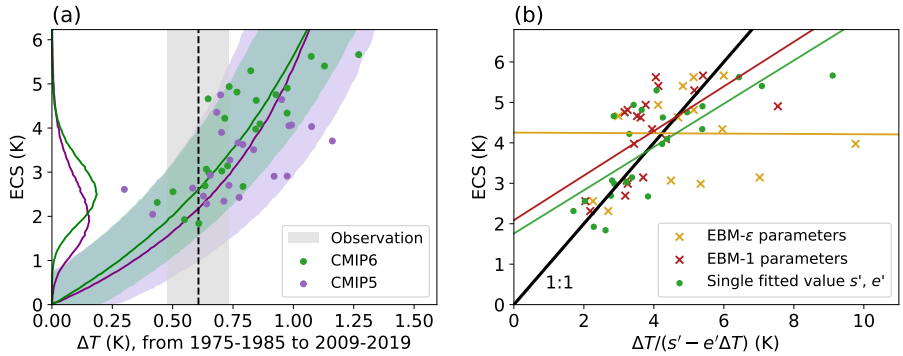


Figure 6. a) Emergent constraint for ECS, using the functional form of Eq. 5. The shaded area includes the 5-95% confidence interval. b) Comparison of emergent constraint fitted parameters, with using model values for s' and e'. The coloured lines are OLS fits for the three cases, and the black line indicates the 1:1-line. Three values for the EBM-ϵ model are not shown as their $\Delta T/(s' - e'\Delta T)$ are between 75 and 90 K.

Ensemble	Median	5-95% range
CMIP5 1970-2005	~~2.33~~ 2.3 K	~~0.70~~ 0.7 – ~~8.41~~ 8.4 K
CMIP5 hist + RCP85	~~2.19~~ 2.2 K	~~0.96~~ 1.0 – ~~4.10~~ 4.1 K
CMIP6 1970-2005	~~2.53~~ 2.5 K	~~0.97~~ 1.0 – ~~8.62~~ 8.6 K
CMIP6 historical	~~1.89~~ 1.9 K	~~0.98~~ 1.0 – ~~3.26~~ 3.3 K
CMIP6 hist + SSP2-4.5	~~2.62~~ 2.6 K	~~1.53~~ 1.5 – ~~3.99~~ 4.0 K

Table 3. Emergent constraint on ECS depending on ~~choices~~ choice of ensemble and period.

3.2 Equilibrium Climate Sensitivity (ECS)

Figure 6a shows the emergent constraint on ECS. For CMIP5, the ~~90~~5–95% confidence interval lies between 0.96–4.09 K. The constraint is stronger for CMIP6, with the ~~90~~5–95% confidence interval spanning 1.52–4.03 K. Further results are shown in Table 3.

The results are highly dependent on the time interval chosen. For shorter intervals, the theoretical functional form shows an increased steepness for higher values of ΔT, making it more difficult to ~~constraint~~constrain. For instance, taking the time period in line with ~~Jiménez-de-la Cuesta and Mauritsen (2019)~~JM19, i.e. 1970–1989 versus 1994–2005, we obtain a 5–95% interval of 0.70~~– 8.41~~ –8.41 K for CMIP5, significantly wider than ~~what was found in the former paper~~found in JM19, which reported a 5–95% confidence interval of 1.72-–4.12 K. The major differences lie in the definition of the theoretical function, where we have cut off the unphysical branch, and a correction of a coding error.

In Fig. 6b the dark green dots represent expected ECS from observed warming (using Eq. 5) and true ECS, using the fitted parameters from Fig. 6a. The light green dots denote the same, but now every model uses its own ocean parameters, $F_{2\times}$ and model forcing computed using Eq. 1. The yellow data shows the expected ECS computed from the EBM-1 model. Full parameter fits for both models are found in Supplementary Tables 2 and 3.

The EBM-ϵ model performs poorly for large values of the ocean heat uptake ~~pattern~~ efficacy parameter ϵ. Models with ϵ around 1.8 in particular show an expected ECS far above a realistic range, with one expected ECS reaching a value of 89 K. Eq. 5 is nonlinear and small errors in parameter estimation quickly lead to large errors in ECS. For the EBM-ϵ model in particular, high internal variability may skew the parameter estimate upwards.

The EBM-1 fit leads to an improved estimation of ECS compared to the Eq. 5 fit in 53% of the cases, whereas the EBM-ϵ model leads to an improvement in 34% of cases. This pattern in similar in the case only historical models are used, with 66% and 42% improved respectively.

3.2.1 Functional form

Explicitly simulating the two-layer model ~~indicates~~ shows that the steepness of the graph is overestimated: assuming no deep ocean temperature rise ($T_0 = 0$) dampens the temperature response of the upper ocean. Geoffroy et al. (2013a) ~~included a~~ derived an analytical solution to the two-box model of Eq. 2 under the weaker assumption of a linearly increasing forcing, which ~~again showed a larger temperature rise. However, by using an~~ also showed a less steep increase of ECS with ΔT for high values of ΔT. This leads the question whether the upper range of ECS is overestimated. In Supplementary Fig. 1, we show this is not the case: by using a decreased ocean heat uptake parameter e' and forcing, the two analytical solutions do overlap~~(Supplementary Fig. 1), which indicates~~ , which demonstrates that using the approximated Eq. 5 in the regression should not lead to biased results in the emergent constraint, but simply that the fitted parameters will be slightly different from the model parameters. This also explains why the regression using model parameters in Fig. 6b is not significantly better than using the overall fitted parameters of Fig. 6a.

4 Discussion and Conclusion

The emergent constraint found on TCR in this paper is very similar to the one found in ~~Jiménez-de-la Cuesta and Mauritsen (2019)~~ JM19 and Tokarska et al. (2020). The most important determinant of the constraint is the periods taken. We have slightly expanded on the amount of models compared to a Tokarska et al. (2020), taking a different period, and we compared further regression choices.

Our best estimate for TCR from the CMIP6 models is 1.68 K, which remains close to the centre of the likely range (1–2.5K) given in the IPCC AR5 (IPCC, 2013b). The emergent constraint on TCR from the CMIP6 models is however strong enough to indicate a much tighter likely range of TCR (16-84%, 1.29–2.05 K).

We find a consistent emergent constraint from the CMIP5 models against observed global warming from 1975 to 2019 (16-84%, 1.27–1.88 K). Furthermore, both of these likely ranges overlap strongly with the emergent constraint on TCR derived by Jiménez-de-la Cuesta and Mauritsen (2019) using a similar method, but only considering global warming from 1970 to 2005 (5-95%, 1.17–2.16 K). In terms of the classification proposed by Hall et al. (2019), we therefore now have a *confirmed* emergent constraint on TCR, with consistency across generations and a sound theoretical framework.

Equilibrium climate sensitivity is likely between 1.9 and 3.4 K (~~16-83~~16-84% percentile). This finding strengthens previous evidence that ECS very unlikely above 4.5 K (Cox et al., 2018a; Jiménez-de-la Cuesta and Mauritsen, 2019; Goodwin et al., 2018). For instance, Goodwin et al. (2018) used history matching, a simple emulator, and observations of surface temperature, ocean heat uptake, and carbon fluxes to estimate climate sensitivity and concluded upon a 5-95% range of 2.0 to 4.3 K. Renoult et al. (2020) ~~using~~ used a combined emergent constraint of the last glacial maximum and mid-Pliocene Warm Period to constrain ECS to 1.1–3.9 K, with the same best estimate of 2.6 K.

Does the presence of many models with ECS over 4.5 K ~~tests~~ mean that the CMIP5 generation was better or more useful for understanding climate sensitivity than CMIP6? From the point of view of emergent constraints the answer is clearly no, as model spread helps capture the shape of the emergent relationship.

In the future, we hope that this TCR constraint will become the basis for constraints also on TCRE (transient climate response to emissions), but this will require the inclusion of additional constraints on land and ocean carbon uptake.

However, we are now in a position to answer the questions that we posed in Section 1:

(a) *Are such high climate sensitivities consistent with the observational record?*

No, models with high ECS (>4.5K) and high TCR (>2.5 K) do not appear to be consistent with observed global warming since 1975 (Figure 3b).

(b) *If so, do the CMIP6 models demand an upward revision to the IPCC likely ranges for climate sensitivity?*

No, instead emergent constraints on TCR (Fig. 4) and ECS (Fig 6) suggest narrower likely ranges for TCR (1.3–2.1 K) and ECS (1.9~~–4.0~~–3.4 K).

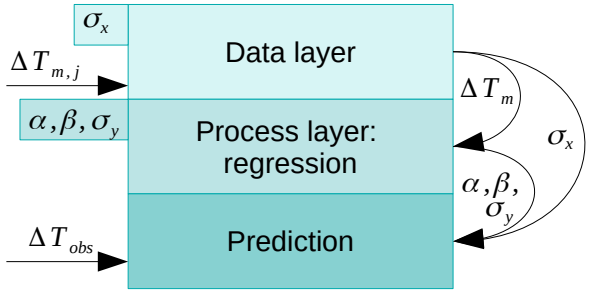
Code availability. The code to analyse the data and produce the figures is available upon request to the corresponding author.

Data availability. CMIP5 and CMIP6 data can be accessed through ESGF nodes.

Appendix A: Hierarchical linear regression

To systematically include the information from all model realisations, we use a hierarchical Bayesian model (Sansom, 2014). This model includes two layers: the normal linear regression (process layer) and a layer that computes the expected warming per model from all its initial value realisations (data layer). To include the initial value ensemble, we assume that each model m has a "true" or "best" value for warming over the last decades denoted by ΔT_T. We further assume that every realisation j of a model gives a value of ΔT that is drawn from a normal distribution with mean ΔT_T and a standard deviation σ_x that is the same across all models. Our hierarchical model consists of two steps: for each model the best estimate of historical warming

Figure A1. Schematic of the hierarchical Bayesian model employed. The data layer models a best estimate of historical warming for each model. With this estimate, a regression is performed between historical warming and TCR in the process layer. Using information from both layers and observed warming, a probability density function is estimated for TCR is the final step.



is computed and with this value a simple linear regression is performed:

$$\Delta T_{m,j} | \Delta T_m, \sigma_x \quad \sim \text{normal}(\Delta T_m, \sigma_x)$$
$$\text{TCR}_m | \alpha, \beta, \sigma_y \quad \sim \text{normal}(\alpha + \beta \Delta T_m, \sigma_y)$$

The probability density function for TCR is then sampled from the observed warming between 1975–1985 and 2009–2019 ΔT_{obs} using the emergent constraint. The observational uncertainty σ_{obs} is taken as the sample standard deviation of the four observational datasets.

$$\text{TCR}_{pred} = \text{normal}\left(\alpha + \beta\,\text{normal}\left(\Delta T_{obs}, \sqrt{\sigma_x^2 + \sigma_{obs}^2}\right), \sigma_y\right)$$

The second layer corresponds with normal linear regression, while the first layer makes an estimate of the true ΔT_m. Note that especially for models with only few initial value member, the "best" ΔT_m does not necessarily correspond with the mean value of these ensemble members, but will instead lie closer to the regression line.

As no warming is expected if climate sensitivity were zero, we expect the regression to pass through the intercept and chose a prior for the intercept α of normal(0, 1). Weakly informative priors are chosen for the slope β, the uncertainty of the regression σ_y and the internal variability σ_x:

$$\alpha \quad \sim \text{normal}(0, 1);$$
$$\beta \quad \sim \text{normal}(2, 10);$$
$$\sigma_y \quad \sim \text{half-normal}(0.5, 10);$$
$$\sigma_x \quad \sim \text{half-normal}(0.2, 0.5);$$

Author contributions. All authors contributed towards the design of the study. MSW led on the data collection, FJMMN led on the data analysis with contributions from MSW and PMC. All authors contributed equally to the manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgements. This work was supported by the European Research Council ECCLES project, grant agreement number 742472 (F.J.M.M.N., P.M.C. and M.S.W.); the EU Horizon 2020 Research Programme CRESCENDO project, grant agreement number 641816 (P.M.C. and M.S.W.). We also acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups (listed in Table 1) for producing and making available their model output. We further thank Diego Jiménez-de-la-Cuesta for kindly sharing their code.

References

Andreae, M. O., Jones, C. D., and Cox, P. M.: Strong present-day aerosol cooling implies a hot future, https://doi.org/10.1038/nature03671, 2005.

Armour, K. C.: Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks, Nature Climate Change, 7, 331–335, https://doi.org/10.1038/nclimate3278, 2017.

Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., Watson-Parris, D., Boucher, O., Carslaw, K., Christensen, M., Daniau, A., Dufresne, J., Feingold, G., Fiedler, S., Forster, P., Gettelman, A., Haywood, J., Malavelle, F., Lohmann, U., Mauritsen, T., McCoy, D., Myhre, G., Mülmenstädt, J., Neubauer, D., Possner, A., Rugenstein, M., Sato, Y., Schulz, M., Schwartz, S., Sourdeval, O., Storelvmo, T., Toll, V., Winker, D., and Stevens, B.: Bounding aerosol radiative forcing of climate change, Reviews of Geophysics, https://doi.org/10.1029/2019RG000660, 2019.

Bengtsson, L. and Schwartz, S. E.: Determination of a lower bound on Earth's climate sensitivity, Tellus, Series B: Chemical and Physical Meteorology, 65, 1–16, https://doi.org/10.3402/tellusb.v65i0.21533, 2013.

Bretherton, C. and Caldwell, P.: Combining Emergent Constraints for Climate Sensitivity, https://doi.org/10.1002/ESSOAR.10501381.1, 2019.

Brown, P. T., Stolpe, M. B., and Caldeira, K.: Assumptions for emergent constraints, Nature, 563, E1–E3, https://doi.org/10.1038/s41586-018-0638-5, 2018.

Caldwell, P. M., Zelinka, M. D., and Klein, S. A.: Evaluating Emergent Constraints on Equilibrium Climate Sensitivity, Journal of Climate, https://doi.org/10.1175/JCLI-D-17-0631.1, 2018.

Carroll, R. J. and Ruppert, D.: The Use and Misuse of Orthogonal Regression in Linear Errors-in-Variables Models, https://doi.org/10.1080/00031305.1996.10473533, 1996.

Chen, X. L., Zhou, T. J., and Guo, Z.: Climate sensitivities of two versions of FGOALS model to idealized radiative forcing, Science China Earth Sciences, 57, 1363–1373, https://doi.org/10.1007/s11430-013-4692-4, 2014.

Cox, P. M., Huntingford, C., and Williamson, M. S.: Emergent constraint on equilibrium climate sensitivity from global temperature variability, Nature, 553, 319–322, https://doi.org/10.1038/nature25450, 2018a.

Cox, P. M., Williamson, M. S., Nijsse, F. J. M. M., and Huntingford, C.: Cox et al. reply, Nature, 563, E10–E15, https://doi.org/10.1038/s41586-018-0641-x, 2018b.

Dlugokenchy, E. and Tans, P.: Trends in Atmospheric Carbon Dioxide, https://www.esrl.noaa.gov/gmd/ccgg/trends/gl_gr.html, 2019.

Forster, P. M., Andrews, T., Good, P., Gregory, J. M., Jackson, L. S., and Zelinka, M.: Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models, Journal of Geophysical Research Atmospheres, 118, 1139–1150, https://doi.org/10.1002/jgrd.50174, 2013.

Forster, P. M., Maycock, A. C., McKenna, C. M., and Smith, C. J.: Latest climate models confirm need for urgent mitigation, Nature Climate Change, pp. 1–4, https://doi.org/10.1038/s41558-019-0660-0, 2019.

Geoffroy, O., Saint-martin, D., Olivié, D. J., 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, Journal of Climate, 26, 1841–1857, https://doi.org/10.1175/JCLI-D-12-00195.1, 2013a.

Geoffroy, O., Saint-Martin, D., Bellon, G., Voldoire, A., Olivié, D. J., 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, https://doi.org/10.1175/JCLI-D-12-00196.1, 2013b.

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, https://doi.org/10.1038/s41561-017-0054-8, https://www.nature.com/articles/s41561-017-0054-8.pdf, 2018.

Gregory, J. M.: A new method for diagnosing radiative forcing and climate sensitivity, Geophysical Research Letters, 31, L03 205, https://doi.org/10.1029/2003GL018747, 2004.

Gregory, J. M. and Forster, P. M.: Transient climate response estimated from radiative forcing and observed temperature change, Journal of Geophysical Research Atmospheres, 113, 1–15, https://doi.org/10.1029/2008JD010405, 2008.

Gregory, J. M., Andrews, T., Ceppi, P., Mauritsen, T., and Webb, M. J.: How accurately can the climate sensitivity to CO_2 be estimated from historical climate change?, Climate Dynamics, 54, 129–157, https://doi.org/10.1007/s00382-019-04991-y, 2019.

Hahn, G. J.: Fitting Regression Models With No Intercept Term., Journal of Quality Technology, 9, 56–61, https://doi.org/10.1080/00224065.1977.11980770, 1977.

Hall, A., Cox, P., Huntingford, C., and Klein, S.: Progressing emergent constraints on future climate change, Nature Climate Change, 9, 269–278, https://doi.org/10.1038/s41558-019-0436-6, 2019.

Hansen, J., Russell, G., Lacis, A., Fung, I., Rind, D., and Stone, P.: Climate response times: Dependence on climate sensitivity and ocean mixing, Science, 229, 857–859, https://doi.org/10.1126/science.229.4716.857, 1985.

IPCC: Annex III: Glossary, book section AIII, p. 1447–1466, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/CBO9781107415324.031, 2013a.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/CBO9781107415324, 2013b.

Jiménez-de-la Cuesta, D. and Mauritsen, T.: Emergent constraints on Earth's transient and equilibrium response to doubled CO_2 from post-1970s global warming, Nature Geoscience, 12, 902–905, https://doi.org/10.1038/s41561-019-0463-y, 2019.

Kiehl, J. T.: Twentieth century climate model response and climate sensitivity, Geophysical Research Letters, 34, https://doi.org/10.1029/2007GL031383, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007GL031383, 2007.

Klein, S. A. and Hall, A.: Emergent Constraints for Cloud Feedbacks, Current Climate Change Reports, 1, 276–287, https://doi.org/10.1007/s40641-015-0027-1, 2015.

Knutti, R., Rugenstein, M. A., and Hegerl, G. C.: Beyond equilibrium climate sensitivity, https://doi.org/10.1038/NGEO3017, 2017.

Lenssen, N., Schmidt, G., Hansen, J., Menne, M., Persin, A., Ruedy, R., and Zyss, D.: Improvements in the GISTEMP uncertainty model, J. Geophys. Res. Atmos., 124, 6307–6326, https://doi.org/10.1029/2018JD029522, 2019.

Medhaug, I., Stolpe, M. B., Fischer, E. M., and Knutti, R.: Reconciling controversies about the 'global warming hiatus', Nature, 545, 41–47, https://doi.org/10.1038/nature22315, 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, https://doi.org/10.1029/2011JD017187, 2012.

Myhre, G., Shindell, D., Bréon, F.-M., W. Collins, J. F., Huang, J., Koch, D., J.-F. Lamarque, D. L., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhang, H.: Anthropogenic and Natural Radiative Forcing, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Stocker, T., D. Qin, G.-K. P., Tignor, M., Allen, S., J. Boschung, A. N., Xia, Y., Bex, V., and (eds.), P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J. F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, Geoscientific Model Development, 9, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016, 2016.

Parsons, L. A., Brennan, M. K., and Proistosescu, C.: Magnitudes and spatial patterns of interdecadal temperature variability in CMIP6, Geophysical Research Letters, 47, 1–11, https://doi.org/10.1029/2019GL086588, 2020.

Renoult, M., Annan, J. D., Hargreaves, J. C., Sagoo, N., Flynn, C., Kapsch, M.-L., Mikolajewicz, U., Ohgaito, R., and Mauritsen, T.: A Bayesian framework for emergent constraints: case studies of climate sensitivity with PMIP, Climate of the Past Discussions, pp. 1–29, https://doi.org/10.5194/cp-2019-162, 2020.

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., and Mosher, S.: Berkeley Earth Temperature Averaging Process, Geoinformatics & Geostatistics: An Overview, 01, https://doi.org/10.4172/2327-4581.1000103, 2013.

Rugenstein, M., Bloch-Johnson, J., Gregory, J., Mauritsen, T., Li, C., Frölicher, T., Danabasoglu, G., Yang, S., Dufresne, J.-L., Schmidt, G. A., Abe-Ouchi, A., and Geoffroy, O.: Equilibrium climate sensitivity estimated by equilibrating climate models 2, Submitted, https://doi.org/10.1029/2019GL083898, 2019.

Sanderson, B. M., Knutti, R., and Caldwell, P.: A representative democracy to reduce interdependency in a multimodel ensemble, Journal of Climate, https://doi.org/10.1175/JCLI-D-14-00362.1, 2015.

Sansom, P. G.: Statistical methods for quantifying uncertainty in climate projections from ensembles of climate models, Ph.D. thesis, University of Exeter, https://ore.exeter.ac.uk/repository/handle/10871/15292, 2014.

Stevens, B., Fiedler, S., Kinne, S., Peters, K., Rast, S., Müsse, J., Smith, S. J., and Mauritsen, T.: MACv2-SP: A parameterization of anthropogenic aerosol optical properties and an associated Twomey effect for use in CMIP6, Geoscientific Model Development, 10, 433–452, https://doi.org/10.5194/gmd-10-433-2017, 2017.

Stocker, M., Ladstädter, F., Wilhelmsen, H., and Steiner, A. K.: Quantifying Stratospheric Temperature Signals and Climate Imprints From Post-2000 Volcanic Eruptions, Geophysical Research Letters, https://doi.org/10.1029/2019GL084396, 2019.

Tanaka, K. and O'Neill, B. C.: The Paris Agreement zero-emissions goal is not always consistent with the 1.5 °C and 2 °C temperature targets, Nature Climate Change, 8, 319–324, https://doi.org/10.1038/s41558-018-0097-x, 2018.

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., and Knutti, R.: Past warming trend constrains future warming in CMIP6 models, Science Advances, 6, 2020.

Williamson, M. S., Cox, P. M., and Nijsse, F. J. M. M.: Theoretical foundations of emergent constraints: relationships between climate sensitivity and global temperature variability in conceptual models, Dynamics and Statistics of the Climate System, 3, https://doi.org/10.1093/climsys/dzy006, 2018.

Winton, M., Takahashi, K., and Held, I. M.: Importance of Ocean Heat Uptake Efficacy to Transient Climate Change, Journal of Climate, 23, 2333–2344, https://doi.org/10.1175/2009JCLI3139.1, 2010.

Zhang, H.-M., Huang, B., Lawrimore, J., Menne, M., and Smith, T. M.: Global Surface Temperature Dataset (NOAAGlobalTemp), Version 5.0, https://doi.org/10.7289/V5FN144H, 2019.

Zhao, M., Golaz, J. C., Held, I. M., Ramaswamy, V., Lin, S. J., Ming, Y., Ginoux, P., Wyman, B., Donner, L. J., Paynter, D., and Guo, H.: Uncertainty in model climate sensitivity traced to representations of cumulus precipitation microphysics, Journal of Climate, 29, 543–560, https://doi.org/10.1175/JCLI-D-15-0191.1, 2016.