~~An emergent constraint~~ Emergent constraints on ~~Transient Climate Response~~ 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. ~~The transient climate~~ 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~~(Arora et al., 2019).~~. 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' ~~range of TCR was given as~~ ranges (16–84% confidence) of TCR (1.0~~to 2.5 K, with a central estimate which was~~ –2.5 K) and ECS (1.5–4.5 K) were broadly consistent with the ensemble ~~mean of the~~ of CMIP5 Earth System Models (ESMs) available at the time~~(1.8 ± 0.4 K). Many~~. However, many of the latest CMIP6 ESMs have larger climate sensitivities, with ~~6 of 23~~ 5 of 34 models having TCR values above 2.5 K, and an ensemble mean TCR of ~~2.1~~ 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 ~~range of TCR~~ 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 ~~further~~ constrain the likely range of TCR to ~~1.5-2.2~~1.3–2.1 K, with a central estimate of ~~1.82~~1.68 K. We reach this conclusion through an emergent constraint approach which relates the value of TCR linearly to the global warming from ~~1970 onwards. We confirm~~ 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~~(Jiménez-de-la Cuesta and Mauritsen, 2019). Our emergent constraint on TCR benefits from both the large range of TCR values across~~. 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 ~~, and also from the extension of the historical simulations into a period when the uncertain changes in aerosol forcing have had a far less significant impact on the trend in~~ 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 ~~perturbations of~~ changes in the Earth's energy budget, ~~normally termed~~ 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 ~~parameters~~ quantities are used to characterise the climate sensitivity of Earth System Models (ESMs). Equilibrium ~~Climate Sensitivity~~ climate sensitivity (ECS) is an estimate of the eventual steady-state global warming at double CO_2. Transient ~~Climate Response~~ 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 ~~typically around half of~~ less than ECS values because of 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 ~~significant~~ decades of advances in climate science, ~~both~~ 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 ~~AR4~~ 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~~or variations~~ , 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; Po-Chedley et al., 2018; Rypdal et al., 2018; Cox et al., 2018b)~~ (Brown et al., 2018; Bretherton and Caldwell, 2019; Cox et al., 2018b; Gregory et al., 2020) , 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 ~~(Jiménez-de-la Cuesta and Mauritsen, 2019)~~ (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 ± stdev) of ECS (~~3.2~~ 3.3 ± 0.7 K) and TCR (1.8 ± ~~0.4~~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 ~~almost half~~ more than one third of the new CMIP6 models now have ECS values over 4.5 K (Forster et al., 2019), and ~~more than a quarter~~ 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 ~~1970 onwards (Jiménez-de-la Cuesta and Mauritsen, 2019)~~1975 onwards (Jiménez-de-la Cuesta and Mauritsen, 2019; Tokarska et al., 2020) , but using the CMIP6 models and observational data up to ~~2018.~~ 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 ~~describing~~ 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 ~~our~~ the emergent constraint on sound physical reasoning~~, as outlined below~~.

~~By definition~~From physical principles, we expect values of TCR to be very well-correlated with simulated global warming across a model ensemble~~, if all models are driven by a~~ . 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. ~~These are the idealised conditions used to estimate TCR, so only internal variability would deny a perfect correlation between TCR and the modelled rate of global warming in this case. However, historical global warming has not~~ Historical global warming has been driven by a qualitative similar forcing, albeit somewhat less rapid. Instead of 1.0%~~per year~~ , CO_2~~increase, but instead by a smaller near-exponential rate of CO_2 increase (~~, 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~~(especially methane and nitrous oxide)~~, and partially offset by the cooling effects of anthropogenic aerosols.

The radiative effects of the ~~known increases~~ 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)~~ (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 ~~TCR~~ climate sensitivity and strong aerosol cooling, and a model with low ~~TCR~~ climate sensitivity and weak aerosol cooling.

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 ~~decade after 1970~~ decades after 1975 when total aerosol load from global SO_2 and NH_3 emissions were similar to values over the last decade (Stevens et al., 2017). For this reason, we follow Jiménez-de-la Cuesta and Mauritsen (2019) in focusing on global warming since ~~1970.~~ 1975. However, in addition we explore a range of start and finish dates to assess 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.

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) 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 ~~constraint~~ 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.

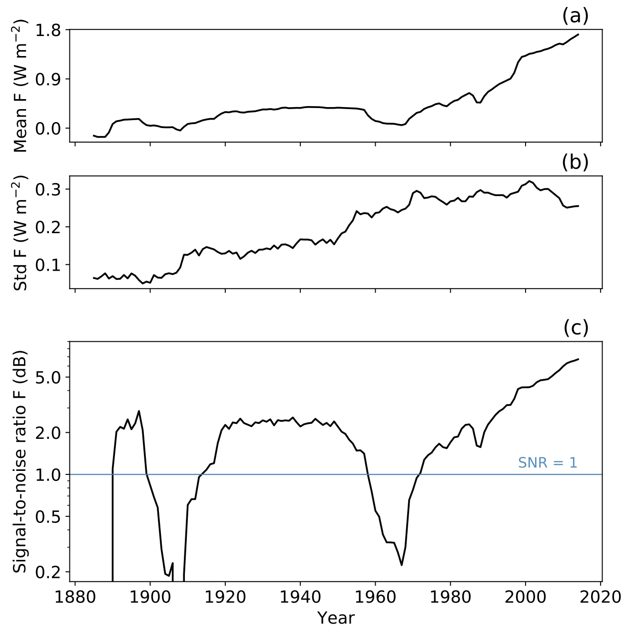


Figure 1. Effective radiative forcing over the historical period, calculated from 22 CMIP6 models: (a) ensemble mean; (b) ensemble standard deviation; (c) signal-to-noise ratio.

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

$$F = \Delta N + \lambda \Delta T \quad (1)$$

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.

2.2 Selection of CMIP6 model runs

Table 1. List of CMIP6 models used in this study and their ERF 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 ~~the climate feedback parameter~~ λ are computed using the Gregory method (Gregory, 2004). ~~Starred models had full model output available and are~~ Models above the ~~ones included~~ horizontal line were used in the ~~emergent constraint~~extended simulations to 2019. Models below the line did not have SSP simulations available at time of writing. ~~Values~~ Consistently derived values for CMIP5 are ~~from (Flato et al., 2013), and model selection the same as~~ displayed in ~~(Nijsse et al., 2019).~~the Supplementary Table 1.

Centre	Model	F~~$_{2xCO_2}$~~$_{2\times}$ ~~(W m^{-2})~~	λ	E
~~*~~heightBCC	BCC-CSM2-MR	3.01	0.98	3.
~~BCC~~ CAMS	~~BCC-ESM1~~ CAMS-CSM1-0	~~3.03~~ 3.95	~~0.93~~ 1.71	~~3.~~
~~*CAMS~~ CAS	~~CAMS-CSM1-0~~ FGOALS-f3-L	~~3.39~~ 3.95	~~1.76~~ 1.31	~~2.~~
~~*~~CCCma	CanESM5	3.63	0.64	5.
~~*CNRM-CERFACS~~ CNRM-CERFACS	CNRM-CM6-1	3.54	0.72	4.
~~*CNRM-CERFACS~~ CNRM-CERFACS	CNRM-ESM2-1	3.09	0.66	4.
~~E3SM-Project~~ CSIRO-ARCCSS	~~E3SM-1.0~~ ACCESS-CM2	~~3.23~~ 3.21	~~0.60~~ 0.67	~~5.~~
~~*EC-Earth-Consortium~~ CSIRO	ACCESS-ESM1-5	2.71	0.68	3.
EC-Earth-Consortium	EC-Earth3-Veg	3.32	0.77	4.
~~*~~EC-Earth-Consortium	EC-Earth3	3.30	0.78	4.
INM	INM-CM4-8	2.61	1.42	1.
INM	INM-CM5-0	2.88	1.49	1.
IPSL	IPSL-CM6A-LR	3.32	0.72	4.
~~*~~MIROC	~~MIROC6~~ MIROC-ES2L	~~3.75~~	~~1.47~~	~~2.~~
~~*~~MIROC	~~MIROC-ES2L~~ MIROC6	~~-~~3.76	~~-~~1.47	~~-~~2
MOHC	HadGEM3-GC31-LL	~~3.39~~3.38	~~0.62~~ 0.60	~~5.~~
~~*~~MOHC	UKESM1-0-LL	3.56	0.66	5.
~~*~~MPI-M	MPI-ESM1-2-HR	3.58	1.20	2.
MRI	MRI-ESM2-0	3.36	1.07	3.
NCAR	~~2.70~~ CESM2-WACCM	~~1.72 NASA-GISS~~ 3.08	~~GISS-E2-1-H~~ 0.63	~~3.~~
~~*~~NCAR	CESM2	~~3.12~~ 3.13	~~0.63~~ 0.59	~~5.~~
NCC	NorESM2-LM	~~-~~3.06	~~-~~1.13	~~-~~2
~~NOAA-GFDL~~ NOAA-GFDL	GFDL-CM4	~~2.90~~ 2.91	0.71	4.
NOAA-GFDL	GFDL-ESM4	3.51	1.31	2.
NUIST	NESM3	~~-~~3.73	~~-~~0.78	~~-~~4
~~SNU~~ UA	~~SAM0-UNICON~~ MCM-UA-1-0	~~3.85~~	~~1.06~~	~~3.~~
~~UA~~ height	**Mean**	~~MCM-UA-1-0~~ 3.69	~~-~~0.95	~~-~~3
Standard deviation		0.40	0.34	1.
AS-RCEC	~~**CMIP6 mean**~~ TaiESM1	~~3.40 (3.42*)~~	~~0.90 (*0.89)~~	~~4.~~
BCC	~~**CMIP6 standard deviation**~~ BCC-ESM1	~~0.32 (0.35*)~~ 3.03	0.89	3.
E3SM-Project	E3SM-1-0	3.23	0.60	5.
NASA-GISS	GISS-E2-1-G	3.89	1.43	2.
NASA-GISS	GISS-E2-1-H	3.55	1.14	3.
MOHC	HadGEM3-GC31-MM	3.36	0.61	5.
MPI-M	MPI-ESM1-2-HR	3.58	1.20	2.
SNU	SAM0-UNICON	3.83	1.02	3.
Mean CMIP6		3.70	0.95	**3.**
Standard deviation CMIP6		0.39	0.33 ~~(*0.35)~~	~~1.~~
Mean CMIP5		~~**CMIP5 mean**~~ 3.58	~~3.44~~ 1.06	~~1.~~
Standard deviation CMIP5		~~**CMIP5 standard deviation**~~ 0.22	~~0.46~~ 0.29	~~0.~~

~~Method~~We ~~use the CMIP6 multimodel ensembleto find an emergent relationship between historical warmingand TCR. We~~ use all currently available CMIP6 models that have control (piControl), historical, a Shared Socioeconomic Pathway ~~3-7.0 (ssp370~~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 ~~. Values of TCR~~ (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 ~~ssp370~~ SSP simulations using the global annual mean surface air temperature (GMSAT) smoothed with a ~~centred~~ 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. ~~This results in a set of 95 simulations from 13 different models.~~

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 ~~our~~ the constraint on TCR to the length of the running mean. It remains relatively invariant past a length of ~~5 years~~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 ~~1970-1980 and 2008-2018~~1975–1985 and 2009–2019. We have chosen the end year to be ~~2018~~ 2019 to maximise the chances of discrimination between high and low sensitivity models~~i. e. as~~. As the forcing from CO_2 increases with time, the warming in more sensitive models is more likely to diverge from less sensitive ones ~~resulting in stronger statistical relationships between TCR and ΔT. Although we use 2018 as~~ as we extend the period over which we calculate the trend. Extending to 2019 also allows us to include the ~~end year of annual GMSAT, we report the central year of the smoothed timeseries in the following figures i. e. the central year of annual GMSAT smoothed with a centred running mean of 11 years would be shown as 2013.~~

~~Other reasons for choosing 2018 include being able to use the~~ most recent observational data and to eliminate possible effects from the warming slowdown between ~~2000-2012~~2000–2012. This slowdown has been attributed to a combination of internal variability and decreased forcing, amongst other things (Medhaug et al., 2017). We ~~have investigated whether this reduced forcing makes a difference to our emergent constraint by extending the historical CMIP6 simulations from 2014 to 2018 using the SSP scenario simulations. The different SSP scenarios have very similar greenhouse gas emissions for the 2015-2019 period, and the choice of SSP does not significantly alter our findings. We use SSP3-7.0 as this SSP has the largest number of model simulations at the time of writing. Choosing either 2014 or 2018 does not significantly affect our results.~~ assess the impact of the slow-down by comparing emergent constraints derived from time-series truncated to have different end years.

~~We have chosen the starting period to be 1970-1980 when aerosols forcing was similar to today's values. This choice also minimised the uncertainty in our estimate of TCR. As a function of start period, uncertainty is relatively flat and minimal between periods with central years of 1975 and 1985.~~

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-aerosol warming.~~ 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):

$$C\frac{d\Delta T}{dt} = F - \lambda\Delta T - \epsilon\gamma(\Delta T - \Delta T_0) \qquad C_0\frac{d\Delta T_0}{dt} = \gamma(\Delta T - \Delta T_0) \tag{2}$$

Here ΔT is the top layer temperature anomaly, ΔT_0 the deep ocean temperature anomaly, λ the climate sensitivity parameter, ϵ the ocean pattern efficacy and γ 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,

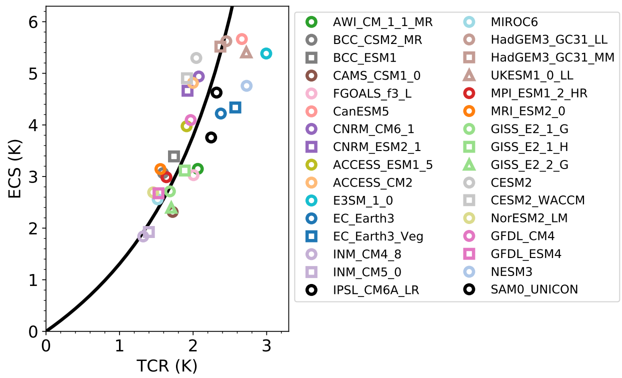


Figure 2. Scatter plot of TCR values plotted against ECS values for all CMIP6 models with both available at the time of submission. Models from the same modelling group are plotted with the same colour. Plot markers differentiate models from the same modelling centre. Black line uses the average ocean 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, the model mean.

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, 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), and lead to the following relationship:

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

Here s' is a forcing parameter, defined as $F/F_{2\times}$ and η an offset. The choice of starting period minimises the uncertainty in our estimate of TCR. As a function of start period, uncertainty is relatively flat and minimal between periods centred around 1975 to 1990. We further investigate the sensitivity of this 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 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, 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, 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 $\Delta T > \frac{s'}{e'}$.

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.

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.

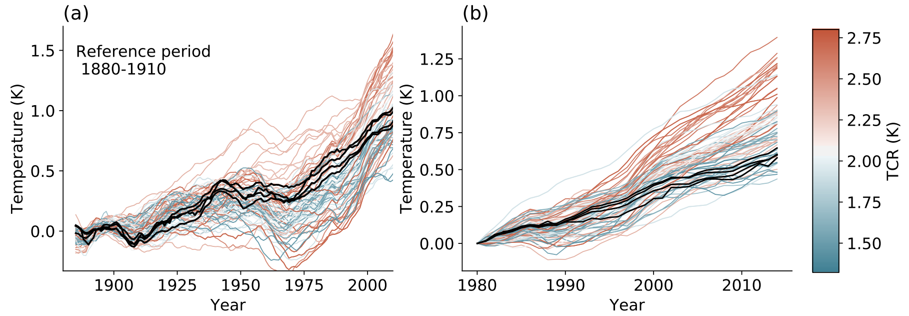
Using these fitting parameters, we investigate the physical basis of Eq 5 with the EBM-ϵ and EBM-1 models. If this function derived from the two-box model is a faithful representation, $\frac{\Delta T}{s' - e'\Delta T}$ should be better related to ECS with individual model parameters than with the bulk 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 ~~ΔT~~ temperature anomaly over the period 1880 to ~~2018 simulated by 13~~ 2019 simulated by 26 different CMIP6 models running a total of ~~95~~ 127 simulations smoothed with a 11-year running mean~~relative to the mean of the reference period .~~. 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 ~~various observational datasets of~~ observational global warming datasets over the same period (Morice et al., 2012; Rohde et al., 2013; Lenssen et al., 2019; Zhang et a . Models with higher TCR either show large ~~ΔT~~ warming at the end of the period, or portray a strong aerosol cooling over the 20th century, particularly visible as a dip around ~~1970~~ 1960-1970 (notably CNRM-ESM1, UKESM1-0-LL and EC-Earth-Veg). ~~The major uncertainty for historical radiative forcing comes from aerosols (Bellouin et al., 2019) .~~ Figure 3b shows the same information for the end of the historical period although the reference period is now chosen to be ~~1970-1980, close to~~ 1975-1985, after the dip. The positive correlation intuitively expected between TCR and

Figure 3. Global mean surface temperature of ~~13 individual~~ 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 ~~the mean GMSAT between~~ 1880-1910 as reference period. (b) Temperature anomaly ΔT relative to the ~~1970-1980~~ 1975–1985 mean. ~~To avoid overrepresenting a single model, a maximum of ten realisations per model are plotted.~~



temperature increase ΔT is ~~more clearly seen~~much clearer for this time interval.

The ΔT for each model simulation in Fig. 3b is used ~~in our~~ 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 confidence interval (grey shaded vertical area) is a combination of the observational uncertainty and the internal variability. 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 (~~1.82~~ 1.68 K) from this emergent constraint is ~~very similar to~~ higher than the best estimate using the larger set of models that have historical simulations up to 2014, but no future scenarios (~~1.77~~median: 1.54 K, ~~1.31 - 2.22 K, likely range). The set of CMIP6 models used in our emergent constraint are listed in table 1 along with their TCR and ECS values, the latter being determined from Gregory plots (Gregory, 2004) on the first 150 year of~~ 5–95% range: 0.76–2.30 K). This can mostly be explained by the fact that 2004-2014 overlaps with the ~~abrupt-4xCO2 simulations~~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 ~~range. The IPCC pdf is not specified, here we take it as~~ likely range (assuming a normal distribution). Both CMIP5 and CMIP6 pdfs are very similar (central estimates differ by ~~0.05~~0.1 K) even though ~~raw model means in~~ CMIP6 ~~and CMIP5 differ by 0.23 K~~contains many more high TCR models. As a continuation of the historical CMIP5 simulation, ~~RCP4~~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).

3.2 ~~Robustness to parameter choices~~

~~We have assessed how robust our estimate of TCR is to the various choices of parameters:~~

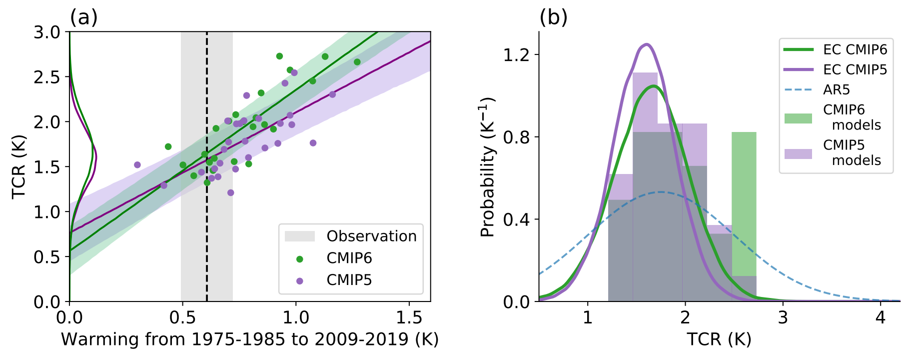
3.1.1 ~~End year~~Period selection

Estimates of TCR depend on the final year chosen for the emergent constraint. ~~However, uncertainty~~ Uncertainty in the estimate of TCR reduces as time increases and the central estimate converges as shown in ~~Figure~~ Fig. 5a. Later end years are ~~intuitively preferred (i.e. 2018) as the increased CO_2 forcing with time leads to more separation in models warming predictions over their internal variability. This increases correlation of warming with TCR and reduces uncertainty in the best estimate. We believe this might also be due to the reducing relative effect of aerosol forcing compared to forcing from CO_2 at later years.~~ 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.

3.1.2 ~~Length of running mean~~

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 ~~our~~ the emergent constraint, the effect of internal variability is already reduced and the length of the

Figure 4. **(a)** Emergent constraint on TCR against historical warming ΔT. ΔT is calculated from the difference between ~~1970-1980~~ 1975-1985 and ~~2008-2018~~ 2009-2019 of a timeseries of GMSAT~~smoothed with a centred 11 year running mean, denoted by their central year on axis label~~. Linear regression is performed with all CMIP5 and CMIP6 simulations. Shaded areas indicate a ~~66~~90% ~~confidence~~ prediction interval (see Appendix A). The vertical dotted line is the mean value of the observations and y-axis shows the ~~horizontal dotted line is the implied central estimate~~ probability distribution of ~~the TCR~~both generations of ensembles. **(b)** Comparison of probability distributions for the transient climate response using ~~post-1970~~ post-1975 warming using CMIP5 and CMIP6 simulations. The probability distribution in the fifth IPCC assessment is not fully specified, ~~and~~ so the figure shows a normal distribution with the same likely range as IPCC.



running mean on the estimate of TCR is small - the central estimate and the likely range remain relatively invariant past a ~~length of 5~~ window length of 8 years.

3.1.2 ~~Start year~~

~~Figure~~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 ~~1985.~~ 1990. Uncertainty from start years of ~~1985 onwards increases although slowly~~ 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 ~~leads to~~ has the potential to lead to a slight overestimation of TCR (~~Figure~~Fig. 5d), as the observed warming is on the lower end of ~~model warming~~the model range. Jiménez-de-la Cuesta and Mauritsen (2019) 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 ~~two~~ three other 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 (Fig. 5d~~. All~~). 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).

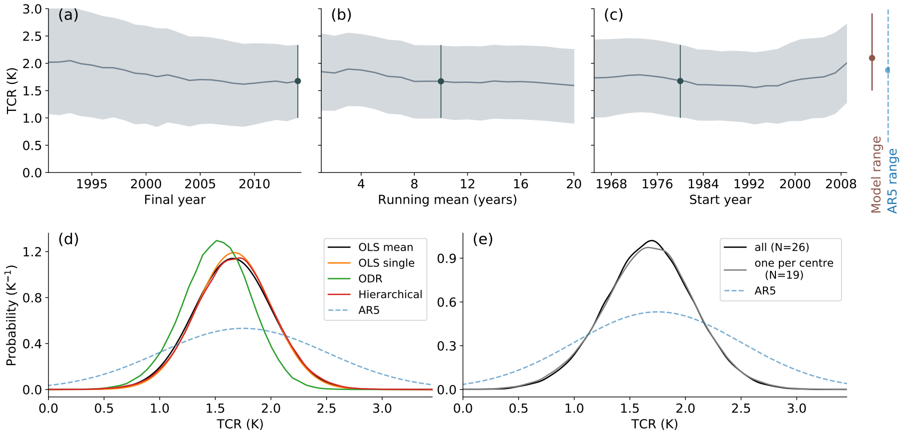
3.1.3 Model selection

~~It has been noted that model~~ 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, ~~but using all models gives a stronger constraint.~~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).

4 ~~Discussion and Conclusion~~

~~An immediate question that may come to mind after constraining TCR, is whether the same information can be used to constrain ECS. There is an approximately linear relationship between ECS and TCR across the~~

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. **(a)** ~~Likely~~ 5-95% TCR range as a function of the ~~end~~ final year (~~dotted~~ blue line central estimate). **(b)**: ~~Likely~~ 5-95% TCR range as a function of length of running mean. **(c)** ~~Likely~~ 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 ~~the~~ regression using all models and the IPCC AR5 range~~using the hierarchical Bayesian model~~.



Study	Ensemble	Period	Median	5-95% range	16-84% range
Jiménez-de-la Cuesta and Mauritsen (2019)	CMIP5	1970 - 2005	1.67 K	1.17 – 2.16 K	
Nijsse et al. (2020)	CMIP5	1970 - 2005	1.72 K	1.11 – 2.34 K	1.37 – 2.08 k
Tokarska et al. (2020)	CMIP6	1981 - 2017	1.60 K		1.2 – 2.0 K
Nijsse et al. (2020)	CMIP6	1970 - 2019	1.68 K	1.1 – 2.5 K	1.3 – 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) and Tokarska et al. (2020) are also shown for comparison.

3.1 Equilibrium Climate Sensitivity (ECS)

Figure 6a shows the emergent constraint on ECS. For CMIP5, the 90% confidence interval lies between 0.96–4.09 K. The constraint is stronger for CMIP6~~ensemble, but this shows significant scatter due to variation in ocean heat uptake across the ensemble (see Figure 2). As a result, although our central estimate of TCR corresponds to ECS values slightly over 3K, some models with ECS > 4.5 K also fall within the likely bounds of our TCR constraint. We find no good evidence to support a particular non-linear relationship between ECS and TCR (Rugenstein et al., 2019), and therefore little evidence of a direct emergent constraint on ECS from recent warming alone (Jiménez-de-la Cuesta and Mauritsen, 2019). In the future, we hope that our TCR constraint will become the basis for constraints also on ECS and TCRE (Transient Climate Response to Emissions)~~ , ~~but this will require the inclusion of additional constraints on ocean heat uptake, and land andocean carbon uptake , respectively.~~ with the 90% 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. For instance, taking the time period in line with Jiménez-de-la Cuesta and Mauritsen (2019), i.e. 1970–1989 versus 1994–2005, we obtain a 5–95% interval of 0.70 - 8.41 K for CMIP5, significantly wider than what was found in the former paper, 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.

~~However, we are now in a position to answer the questions that we posed in Section 1, at least with regard to TCR: Are such high climate sensitivities consistent with the observational record?~~

~~No, models with high TCR (>2.5 K) are not consistent with observed global warming since 1970, as demonstrated in Figure 3b. If so, do the~~

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

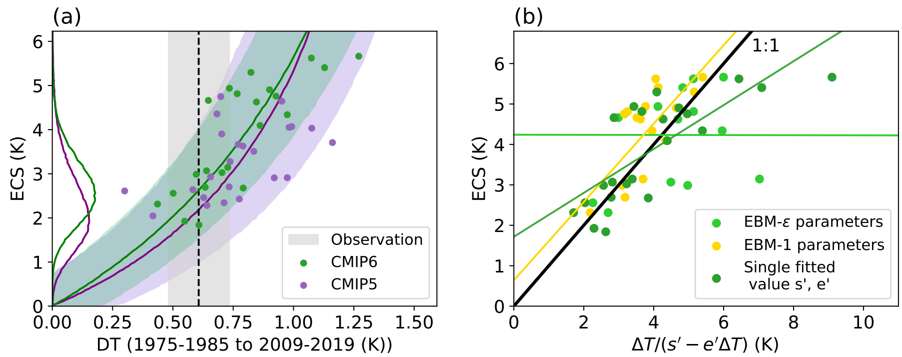


Figure 6. a) Emergent constraint for ECS, using the functional form of Eq. 5. The shaded area includes the 5-95% confidence interval. b) Comparison of emergent constraint fitted parameters, with using model values for s' and e'. The coloured lines are OLS fits for the three cases, and the black line 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
CMIP5 1970-2005
CMIP5 hist + RCP85
CMIP6 ~~models demand an upward revision to the IPCC likely ranges fo[illegible] TCR (Figure 4) does~~
CMIP6 ~~ensemble.~~ historical
CMIP6 hist + SSP2-4.5

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

denote the same, but now every model uses its own ocean parameters, $F_{2\times}$ and model forcing computed using Eq. ??. 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.

~~Instead, the CMIP6 models help to constrain the likely TCR range, without significantly changing the central estimate.~~ The EBM-ϵ model performs poorly for large values of the ocean heat uptake pattern 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.

Explicitly simulating the two-layer model indicates that the steepness of the graph is overestimated: assuming no deep ocean temperature rise dampens the temperature response of the upper ocean. Geoffroy et al. (2013a) included a analytical solution to the two-box model under the weaker assumption of a linearly increasing forcing, which again showed a larger temperature rise. However, by using an decreased ocean heat uptake parameter e' and forcing, the two analytical solutions do overlap (Supplementary Fig. 1), which indicates 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 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) 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 ~~1.82~~1.68 K, which remains close to the centre of the likely range (~~1-2~~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 (~~1.5-2.2~~16-84%, 1.29–2.05 K).

We find a consistent emergent constraint from the CMIP5 models against observed global warming from ~~1970 to 2018 (1.31-2.22~~ 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)~~ 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~~ –2.16 K). In terms of the classification proposed by Hall et al. (2019), we therefore now have a *confirmed* emergent constraint on TCR, ~~implying an approximate likely range of 1.5 to 2.2K.~~ with consistency across generations and a sound theoretical framework.

Equilibrium climate sensitivity is likely between 1.9 and 3.4 K (16-83% 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; Gc . 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 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 our 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.

~~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. Regression line computed with orthogonal distance regression (ODR).~~ 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).

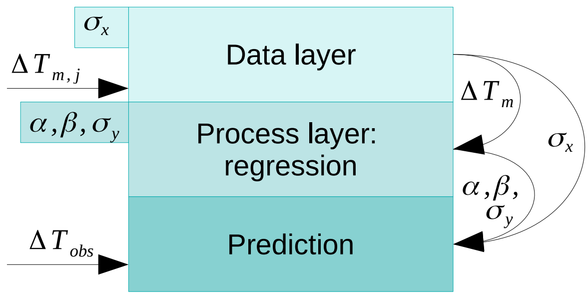
(b) *If so, do the CMIP6 models demand an upward revision to the IPCC likely ranges for climate sensitivity?*

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 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.

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.



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

$$\text{for}(j \text{ in } 1{:}N)\{ \quad \Delta T_{all[j]_{m,j}} | \Delta T_m, \sigma_x \quad \sim \text{normal}(\Delta T_{T[i]_m}, \sigma_x);$$

$$\text{TCR}_m | \alpha, \beta, \sigma_y \quad \}\text{for}(i \text{ in } 1{:}N_g)\{\text{TCR}[i] \sim$$

The probability density function for TCR is then sampled from the ~~observation of warming between 1970 and 2018 ΔT and~~ 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.

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

~~Here `rng` is a (pseudo) random number generator. The second for loop~~ The second layer corresponds with normal linear regression, while the first ~~for loop~~ layer makes an estimate of the true ~~ΔT_T~~ ΔT_m. Note that especially for models with only few initial value member, the "best" ~~ΔT_T~~ ΔT_m

does not necessarily correspond with ~~that the value of this only ensemble member~~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 ~~intercept α, the~~ slope β, the uncertainty of the regression σ_y and the internal variability σ_x:

$$\begin{aligned}
\alpha &\sim \cancel{\mathrm{normal}(0,5)}\mathrm{normal}(0,1);\\
\beta &\sim \cancel{\mathrm{normal}(0,10)}\mathrm{normal}(2,10);\\
\sigma_y &\sim \text{half-normal}(0.5,10);\\
\sigma_x &\sim \text{half-normal}(0.2,0.5);
\end{aligned}$$

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.

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., Schwinger, J., Bopp, L., Boucher, O., Cadule, P., Chamberlain, M. E., Christian, J. R., Delire, C., Fischer, R. A., Hajima, T., Ilyina, T., Joetzjer, E., Kawamiya, M., Koven, C., Krasting, J., Law, R. M., Lawrence, D. M., Lenton, A., Lindsay, K., Pongratz, J., Raddatz, T., Séférian, R., Tachiiri, K., Tjiputra, J. F., Wiltshire, A., Wu, T., and Ziehn, T.: Carbon-concentration and carbon-climate feedbacks in CMIP6 models, and their comparison to CMIP5 models, Biogeosciences Discussions, 2, 2019.

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.

Flato, G., Marotzke, J., B. Abiodun, P. B., Chou, S., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M.: Evaluation of Climate Models, 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.

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, 2020.

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 CO2 from post-1970s global warming, Nature Geoscience, 12, 902–905, https://doi.org/10.1038/s41561-019-0463-y, 2019.

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.

Nijsse, F. J., Cox, P. M., Huntingford, C., and Williamson, M. S.: Decadal global temperature variability increases strongly with climate sensitivity, Nature Climate Change, 9, 598–601, https://doi.org/10.1038/s41558-019-0527-4, 2019.

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.

Po-Chedley, S., Proistosescu, C., Armour, K. C., and Santer, B. D.: Climate constraint reflects forced signal, Nature, 563, E6–E9, https://doi.org/10.1038/s41586-018-0640-y, 2018.

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.

Rypdal, M., Fredriksen, H. B., Rypdal, K., and Steene, R. J.: Emergent constraints on climate sensitivity, Nature, 563, E4–E5, https://doi.org/10.1038/s41586-018-0639-4, 2018.

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.

Response to Reviewer 1

Interactive comment on "An emergent constraint on Transient Climate Response from simulated historical warming in CMIP6 models" by Femke J. M. M. Nijsse et al.

Reviewer comments are listed in italics below, followed by our responses in normal font.

Anonymous Referee #1
Received and published: 16 January 2020

Review of "An emergent constraint on transient climate response from simulated historical warming in CMIP6 models" by F. J. M. M. Nijsse and co-authors. In this paper the authors apply a recently proposed emergent constraint on transient climate response (TCR) on a new set of climate models (CMIP6). The emergent constraint uses warming since the 1970's which is a period that has aerosol forcing which doesn't change too much, and so even if there is uncertainty in the absolute magnitude shouldn't affect the warming rate too much. A best estimate TCR of 1.82 K is obtained, which is about 10 percent higher than that found in other studies. These other studies are Jimenez-de-la-Cuesta and Mauritsen (2019), Tokarska et al. (submitted), and implicitly Winton et al. (2020, JAMES, see their Fig. 14).

It is obviously useful to test a method on new model ensembles, as there have been several cases of emergent constraints found in one ensemble that does not work in another. However, the authors have made a series of choices that are different from the original study which hinders a direct comparison. Again, it is useful that choices regarding the statistics are explored, but it is not currently possible to see whether the shift is related to these new methods or something more fundamental. Other problems were that the authors have not used too many models and some of the writing was less insightful. I suggest the authors undertake major revisions.

Response: We thank the reviewer for their thoughtful comments on our paper. Their suggestions with regards to ECS were especially important for improving the paper. The revised paper now contains an additional emergent constraint on ECS.

<u>*Major issues*</u>

1) *When a study obtains different quantitative estimates compared to previous studies (see above), then I expect to be able to understand why. It is not sufficient to say that this is within the error-bounds because the input data is in principle the same.*

Response: We have included further comparison with the Jimenez-de-la-Cuesta and Mauritsen study (henceforth JM19). We have identified several differences:

a) JM19 had slightly lower values for TCR compared to the IPCC values. We have now computed TCR values directly (instead of using AR5 values) for CMIP5 to ease comparison with JM19 and CMIP6. Our calculations are very close to the standard IPCC values;
b) JM19 compared different periods to those used in our draft paper;
c) a minor programming error was found in the analysis software of JM19;
d) JM19 used a different statistical method that assumes error solely in the independent variable. This reduces regression dilution. However, for an observation that is less than the average model warming, this method produces lower values (by about 0.1K).

Despite these differences, we get a very similar emergent constraint on TCR if we use a similar method (JM19 had a best estimate of 1.67K, and we get 1,66K using similar methods for CMIP5). We have now included the ODR statistical method in a revised figure 3. We also provide details of our attempts to reproduce JM19 in our revised results section, and have added a table of CMIP5 model values to the appendix.

2) I found the discussion of ECS somewhat problematic; several detailed comments are provided below. The culmination, however, is at the beginning of section 4, when Figure 4 is discussed, plotting ECS against TCR. From this plot it is claimed that, contrary to earlier studies the post-1970s warming does not constrain ECS. However, the plot uses the posterior TCR to make this claim, not observed warming, and furthermore the authors do not provide a statistical analysis to support the claim. It is furthermore claimed that a straight line is superior to any other more physically based model, which is clearly not right. A physical constraint is that ECS -> 0 as TCR -> 0, and this linear fit is far from crossing the origin. Any curve looks linear if you zoom in far enough.

Response: Our aim in including Figure 4 was to demonstrate that a good constraint on TCR does not imply a good constraint on ECS. However, we accept the reviewer's criticism that our discussion of this point was not well justified by the analysis that we presented. In the revised manuscript, we now attempt to fit the non-linear function between ECS and warming trend, as proposed by JM19. Diego Jimenez de la Cuesta kindly provided us with the python code he used in JM19, which we were able to compare to our own code. We conclude the following: (a) there was a minor coding error in the JM19 code (errors switched between x and y variables), and also some arbitrary adjustments made to compensate for said error; (b) when these are corrected, and we use the warming to 2005 (as JM19), we find no useful constraint on ECS; (c) however, when we use the warming out to 2019, the corrected code produces a constraint on ECS which is similar to that reported by JM19 (95% ranges of ECS: JM19=1.72-4.12 K; our study=1.76-4.52 K).
We have also investigated whether the functional form proposed by JM19 is supported by the relationship between ECS and the warming trend across the models:

$ECS = DT/(s' - e'DT)$.

In order to turn this theoretical relationship into an emergent constraint on ECS, JM19 used the ocean heat uptake parameter *e'* and the radiative forcing parameter *s'* as fitting parameters. As these values are highly variable between models, the residual of the DT-ECS emergent relationship should be at least partially explained by model differences in ocean heat uptake and forcing. - if this function is indeed theoretically sound. Yet, $DT/(s'-e'DT)$ correlated better with ECS if the fitted s' and e' are used, instead of model-specific s' and/or e', suggesting that the theory used by JM19 is not fully consistent with the results from the CMIP5 and CMIP6 models.

We discuss these issues concerning the relationship between ECS and the warming trend in a revised Section 3.

3) A perhaps somewhat less important point is that the authors first apply smoothing, then average over periods and ensembles, which is effectively the same thing. I mention this because it bothered me that the authors would add an unnecessary layer of complexity, and also because it was unclear what is done with the running-mean smoothing when you approach the end of the time-series in year 2018. For the early period, nominally 1970-1980, it simply means there is some weighing of years outside the interval, out to 1965-1985 for an 11-year filter. But for the late period, which years are then included? All in all, though, there is no reason to do the smoothing at all, averaging over periods as well as ensemble members is a filter.

Response: In fact, there is only one step of time-averaging (the wording smoothing and running mean referred to the same step), and the years 1965-1970 were not used at all. We have made changes in the text and figure captions to make this clearer (e.g. not using a central year but instead writing the period out explicitly in the x-label of Fig 2).

<u>Detailed comments</u>

5, Please report what range is given.

Response: this would make the abstract less readable, as it already contains quite a few numbers. The full range is clear from Table 1, and is now also reiterated in the introduction.

23, It is well-known that the TCR/ECS ratio is not a constant, but decreases with ECS (Hansen et al. 1985, Science). We now understand that the ratio is dependent on the feedback, heat uptake coefficient and pattern effects (e.g. Armour 2017).

Response: we have removed this rule-of-thumb and added a discussion about the reduced ratio.

31, 'climate trends, variability or other observables

Response: added

32-34, This sentence left me with an impression that the debate over the value of ECS revolves only around Cox et al. (2018a). Please remove or rewrite

Response: We removed two of the citations, and added two more recent papers by others.

35, I suggest adding more relevant references, e.g. Gregory and Forster (2008), Otto et al. (2013) and Bengtsson and Schwartz (2013) etc.

Response: we've added the reference to Gregory and Bengtsson.

40-41, What is this claim based on? Please explain and/or provide references.

Response. We've added a reference to Tanaka & O'Neill, NCC, (2018).

42, The questions are also science-relevant, why deprive them to being only policy-relevant?

Response. Added.

53, Here, and in several other places, the authors refer to the emergent constraint as theirs ("our constraint"). I suggest rewriting.

Response: done

58-59, a 1 percent per year increase is also exponential. Table 1, please add number of simulations and the temperature change.

Response: We've added exponential before 1% to make this point clearer. We'll add more information to Table 1.

94-95, Why omit so many years of data, 1980-2008 is not used but contains information as well.

Response: Figure 3 shows a sensitivity to parameter choice, of which one is the running mean. We've extended this to 20 years, and conclude not much extra information is obtained if more years are used.

97-99, I didn't understand this, see also major point above.

Response: we have clarified this by adjusting the figure captions. We now state the difference between period A and B more clearly.

106, if using 2014 does not significantly change the results, then I suggest to stick with 2014 which would allow including many more models and alleviates concerns that stitching together two experiments could lead to biases (e.g. from missing volcanoes in scenarios).

Response: With the addition of more models and an extra year, the difference between ending in 2014 and 2019 is more pronounced. There is a trade-off between using a shorter period, that was also dominated by the 'global warming hiatus', and biases from a discrepancy between real and modelled forcings. We have now investigated the sensitivity to forcings by comparing the various SSPs. A difference of only 1% was found between DT for SSP126 and SSP585.

113, what is "post-aerosol"? Figure 1, it would seem that more than 13 lines are plotted.

Response: post-aerosol changed into post-1970. In Figure 1, up to ten lines per model were shown. To emphasize this, the sentence stating this has been brought forward in the caption.

131, this type of information belongs in Methods.

Response: moved.

133, I would like to see CMIP5 models tabulated as well.

Response: we will do this.

153, I was confused over this sentence, do the authors mean to refer to 3a instead, and the case where end and start year are so close that there is no signal?

Response: The sentence referred to 3c. However, with more data analysed, we have adjusted the paragraph. The uncertainty is now minimum between 1970 and 1976. The second part of the confusing sentence has been deleted, as uncertainty is not that big anymore with the larger set of models.

175, I am not sure Rugenstein et al. (2019) said this.

Response: We removed the sentence, and replaced it with a discussion of the nonlinear relationship.

176, likewise, I don't think Jimenez-de-la-Cuesta and Mauritsen (2019) said this.

Response: Likewise deleted.

Appendix A, I struggled to understand this. Would it be possible to provide an illustration of how the method works?

Response: we have added an illustration, replaced the pseudo code with equations and simplified the text.

Geoffroy, O., Saint-Martin, D., Bellon, G., Voldoire, A., Olivié, D. J. L., & Tytéca, S. (2013). 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*(6), 1859–1876. https://doi.org/10.1175/JCLI-D-12-00196.1

Williamson, M. S., Cox, P. M., & Nijsse, F. J. M. M. (2018). Theoretical foundations of emergent constraints: relationships between climate sensitivity and global temperature variability in conceptual models. *Dynamics and Statistics of the Climate System*, *3*(1). https://doi.org/10.1093/climsys/dzy006

Response to Reviewer 2

Interactive comment on "An emergent constraint on Transient Climate Response from simulated historical warming in CMIP6 models" by Femke J. M. M. Nijsse et al.

Reviewer comments are listed in italics below, followed by our responses in normal font.

Anonymous Referee #2
Received and published: 5 March 2020

In this paper authors apply the concept of "emergent constraints" to new CMIP6 model data aiming to restrict a possible range of Earth climate system sensitivity to CO2 doubling. The topic of the paper is of considerable importance especially in the light that many of CMIP6 model demonstrate increased sensitivity to CO2 forcing (5K+/2xCO2). The paper fits well within the scope of the journal. I recommend the paper for publishing in general but I think some aspects of the paper should be improved.

General comments:

1. *The concept of emergent constraints must be explained much better. Please expand your definition.* "*By definition, we expect..." of emergent constraints on line 55 to be understandable for inexperienced reader. Why TCR has to be correlated with GMST changes across a model ensemble? Models are different, some could have wrong dynamics and incorrect response (Green) function correspondent to CO2 forcing etc. From the paper conclusions it follows that some of the models have wrong TCR while other models are based on the same principles and use more or less the same parameterizations, so why one should believe that TCR/GMST change ratio should be the same for models and for real climate system?*

   Response: we have added further explanation of the concept and assumptions (see response to comment 2 below).

2. *2. Authors must be more careful with the use of definitions. As far as I understood, the TCR is defined as the change of model/climate system GMST (in K) from equilibrium conditions at the moment of CO2 doubling (1%/year forcing for CO2 only). Then what is TCR in "idealized" conditions? "Global warming" is a general concept, you cannot relate/correlate it with TCR in K (line 50, line 55 etc).*

   Response: the paragraph about TCR has been rewritten, making clear that there is only one definition of TCR, and that global warming is defined here as rise in GMST. We further explicitly acknowledge that emergent constraints assume there is no systematic error in the relationship, with a reference to Winkel, Myneni & Brovnik (Earth system dynamics, 2019) and how this emergent constraint is in line with theoretical expectation (JM19).

3. *It could be interesting to have CMIP5 model results for comparison on Fig.2a and Fig.4 as well.*

Response: We've added the CMIP5 models to Figure 2a, but Figure 4 became too crowded with both model ensembles.

4. *It should be pointed out that GMST changes are estimated with respect to the nonequilibrium state (1970-80 average). Will the green line at Fig. 2a cross TSR=0 near the out-of-equilibrium temperature-in-1975 (around -0.4K)?*

   Response: we have adjusted the limits of the figure so that the intercept is visible. The intercept location is highly dependent on the regression method, with those methods assuming the error to be solely in the y-variable (OLS, Hierarchical) getting a positive intercept, while methods assuming similar errors in x and y (orthogonal distance regression), portraying a negative intercept. Per the theoretical foundations of JM19, we expect an approximate zero-intercept in this non-equilibrium regime.

5. *Why 13 models only for CMIP6? Zelinka et al., GRL, 2020 analyzed 27 CMIP6 models...*

   Response: at the time of submission, there were only 13 models for which all necessary information was available, including future scenarios. Now a larger set of 24 models is available. We have also included the emergent constraint with 31 models that ends in 2014 for which scenarios runs are not required.

<u>Special comments:</u>

Lines 20-25. TCR and ECS are introduced for ESMs where they are well defined characteristics for each ESM. On the next line (line 26) paper says that "both TCR and ECS remain uncertain". What do you mean here?

Response: We have clarified that the TCR and ECS values that we seek relate to the real climate system. These real-world values are still poorly known, even though ECS and TCR are well-defined for each model.

Line 50. Relationship between historical warming (expressed in terms of GMST) and TCR?

Response: Added.

Line 55-60. "By definition, we expect...". What "definition" do you mean? What are "idealized" conditions? (Are they somehow different from the ones used in your definition of TCR on lines 20-25)?

Response: The paragraph has been rewritten. 'By definition' has been changed to 'from physical principles', and it has been made clear we use the normal definition of TCR.

Line 80. Could you please provide link to the data?

Response: In addition to a reference to the ESGF nodes, we will upload all the code, including the data, to Code Ocean.

Line 107-108. Can you illustrate the similarity between aerosol forcing in 1970-80 and 2010-2020?

Response: we have added a graph computing the spread in effective radiative forcing in the appendix. This graph shows that the spread is highest in the sixties and early seventies.

Line 120. "The major uncertainty....". This sentence falls out of the context.
Response: we have removed the sentence and moved the reference to the introduction.

Line 122. Can you give a number for correlation between TCR and deltaT?
Response: Yes, the correlation is 0.84 for CMIP6, and 0.63 for CMIP5. Added.

Line 129-131. Move this sentence upward to line 85 (definitions of the table 1)?
Response: done.

Line 200 (Appendix). Appendix does not clarify anything. Either remove or expand it.
Response: the Appendix has been rewritten completely and a figure has been added for extra clarity. The pseudo-code has been replaced by normal equations for easier understanding.