

LETTER • OPEN ACCESS

Climate sensitivity indices and their relation with projected temperature change in CMIP6 models

To cite this article: Linnea L Huusko *et al* 2021 *Environ. Res. Lett.* **16** 064095

View the [article online](#) for updates and enhancements.

You may also like

- [A re-evaluation of the Earth's surface temperature response to radiative forcing](#)
Peter C Young, P Geoffrey Allen and John T Bruun
- [Converging evolution leads to near maximal junction diversity through parallel mechanisms in B and T cell receptors](#)
Jennifer I C Benichou, Jeroen W J van Heijst, Jacob Glanville et al.
- [Influence of lipid rafts on pattern formation during T-cell adhesion](#)
Long Li, Jinglei Hu, Bartosz Róycki et al.

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

Climate sensitivity indices and their relation with projected temperature change in CMIP6 models

OPEN ACCESS

RECEIVED
23 October 2020REVISED
20 May 2021ACCEPTED FOR PUBLICATION
2 June 2021PUBLISHED
17 June 2021

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.

Linnea L Huusko¹ , Frida A-M Bender^{1,*} , Annica M L Ekman¹ and Trude Storelvmo² ¹ Department of Meteorology and Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden² Department of Geosciences, University of Oslo, Oslo, Norway

* Author to whom any correspondence should be addressed.

E-mail: frida@misu.su.se**Keywords:** equilibrium climate sensitivity, transient climate response, temperature projectionSupplementary material for this article is available [online](#)**Abstract**

Equilibrium climate sensitivity (ECS) and transient climate response (TCR) are both measures of the sensitivity of the climate system to external forcing, in terms of temperature response to CO₂ doubling. Here it is shown that, of the two, TCR in current-generation coupled climate models is better correlated with the model projected temperature change from the pre-industrial state, not only on decadal time scales but throughout much of the 21st century. For strong mitigation scenarios the difference persists until the end of the century. Historical forcing on the other hand has a significant degree of predictive power of past temperature evolution in the models, but is not relevant to the magnitude of temperature change in their future projections. Regional analysis shows a superior predictive power of ECS over TCR during the latter half of the 21st century in areas with slow warming, illustrating that although TCR is a better predictor of warming on a global scale, it does not capture delayed regional feedbacks, or pattern effects. The transient warming at CO₂ quadrupling (T140) is found to be correlated with global mean temperature anomaly for a longer time than TCR, and it also better describes the pattern of regional temperature anomaly at the end of the century. Over the 20th century, there is a weak correlation between total forcing and ECS, contributing to, but not determining, the model agreement with observed warming. ECS and aerosol forcing in the models are not correlated.

1. Introduction

The climate system of the Earth responds to a perturbation to the top of atmosphere (TOA) radiative balance through a change in temperature. This imbalance constitutes a radiative forcing of the climate system, and the magnitude of the response is determined by the strength of the forcing and the net radiative feedback. The climate sensitivity, quantified as a change in global mean temperature resulting from a given forcing, is a quantity of central importance to future climate projection, but it is also elusive as it depends on the time scale, forcing agent and state of the climate system (e.g. Collins *et al* 2013, Marvel *et al* 2016, Stevens *et al* 2016, Pfister and Stocker 2017, Richardson *et al* 2019, Rugenstein *et al* 2020).

A number of measures of climate sensitivity are used in the literature, and here we focus primarily on equilibrium climate sensitivity (ECS) and transient

climate response (TCR), two different measures of the sensitivity of the climate system to external forcing, in terms of temperature response to doubling of atmospheric CO₂ concentration. The idealised equilibrium response ECS is often approximated by extrapolation from non-equilibrium simulations, resulting in an effective climate sensitivity (or EffCS) that in fact underestimates the true ECS (e.g. Armour *et al* 2013, Rugenstein *et al* 2020). In the absence of a large set of equilibrated simulations, we follow common practice (e.g. Grose *et al* 2018, Flynn and Mauritsen 2020, Meehl *et al* 2020) and refer to the effective climate sensitivity simply as ECS.

Regardless of what measure or method is used, quantifying climate sensitivity poses a challenge. A recent assessment (Sherwood *et al* 2020) narrows the long-standing uncertainty range of sensitivity, but emphasizes that further constraining its value remains an important goal. Estimates of climate

sensitivity from coupled models diverge, and for instance current-generation models (from the sixth phase of the coupled model intercomparison project, CMIP6 (Eyring *et al* 2016)) as a group have been found to have higher sensitivities than the previous generation models, CMIP5 (Forster *et al* 2020, Meehl *et al* 2020, Nijssen *et al* 2020, Zelinka *et al* 2020). This difference holds for both ECS and TCR. For current as well as previous model generations, ECS and TCR are positively correlated, so that a model with high sensitivity on short time scale (high TCR) generally also has high sensitivity on long time scale (high ECS) (Yoshimori *et al* 2014, Meehl *et al* 2020). TCR is lower than ECS as ocean heat uptake delays surface warming (e.g. Hansen *et al* 1985, Nijssen *et al* 2020).

For TCR or any other transient measure of sensitivity (see also T140 used by Gregory *et al* 2015, Grose *et al* 2018, Sanderson 2020), the uncertainty is dependent not only on uncertainty in forcing and feedbacks but also on uncertainty in the rate of heat transfer to the deep ocean (Lutsko and Popp 2019). Whereas ECS refers to an idealised state of equilibrium, TCR, with its transient nature, has been argued to be of greater relevance for the prediction of climate change over the nearer future of decades to centuries, and thereby for policy decisions on mitigation strategies (Frame *et al* 2005, Allen and Frame 2007, Knutti *et al* 2017, Tokarska *et al* 2020). Analysis of CMIP5 models has indicated, however, that ECS overall explains more of the model spread in the global mean temperature trend over the 21st century than TCR, and is thereby a more useful measure for describing model spread in projected global temperature change (Gregory *et al* 2015, Grose *et al* 2018). As pointed out by Sanderson (2020), a single number cannot be expected to describe future temperature change, and ensembles of simulated future warming under given scenarios are our most complete way of depicting and communicating future change. Still, the spread in response among models is related to their sensitivities, and the question remains which sensitivity measure better explains the variability in future temperature change on a given time scale.

The uncertainty in sensitivity and forcing, particularly the contribution from changes in aerosol, allows for different combinations of forcing and sensitivity to be compatible with a given temperature evolution. For earlier generation models, a relation has been highlighted between ECS and total radiative forcing, consistent with a compensation between sensitivity and aerosol forcing magnitude (Kiehl 2007). However, no robust correlation between aerosol forcing and ECS has been found in subsequent model generations (Forster *et al* 2013, Meehl *et al* 2020, Smith *et al* 2020), and the question of whether models display a correlation between total forcing and sensitivity, and whether such a relation would be indicative of tuning, has remained open.

In this study, we investigate the relation between sensitivity metrics, forcing, and past and projected temperature change in the latest coupled climate models. Following the methods of studies on earlier generation models, we address the questions of which sensitivity metric is more closely related to simulated temperature anomaly (section 3.1), how regional temperature change relates to the global mean sensitivity metrics (section 3.2), and how forcing and sensitivity compensate (section 3.3).

2. Methods

The present analysis is based on the CMIP6 models listed in table S1 (available online at stacks.iop.org/ERL/16/064095/mmedia), as described in the following.

2.1. Climate sensitivity metrics

The values of ECS listed in table S1 are taken from Meehl *et al* (2020) and represent approximations made from simulations of abrupt quadrupling of the atmospheric CO₂ concentration. The method, initially suggested by Gregory *et al* (2004), is based on the radiative budget of the climate system, described as:

$$N = F - \alpha \Delta T, \quad (1)$$

where N is the TOA radiative imbalance, F is the imposed constant forcing, α is the climate feedback parameter and ΔT is the change in global mean surface air temperature (GMSAT) caused by the imposed forcing (compared to some reference temperature in a system with no imposed forcing). In this framework, the feedback parameter is assumed to be time-invariant, and a linear regression of $N(t)$ against $T(t)$ from model output then gives the values for α and F , and the ECS is given by $0.5F/\alpha$, where the factor 0.5 is applied because the *abrupt-4xCO2* experiment is used rather than an experiment with doubled CO₂ amounts. Dividing the response to quadrupled CO₂ forcing by two introduces an error, as it assumes a linearity that has been found not to hold, especially for high sensitivities (Jonko *et al* 2013, Bloch-Johnson *et al* 2015, Tsutsui 2017, Rugenstein *et al* 2020). More important to note, however, is the assumption of a constant feedback parameter, which is invalidated by pattern effects, i.e. that the time evolving spatial patterns of surface warming trigger different combinations of regional feedbacks at different times, causing discrepancies between regression-based estimates of ECS depending on the time period used. Recent temperature evolution has for instance induced more damping cloud feedback than is expected as warming continues, and hence underestimates ECS (Armour *et al* 2013, Stevens *et al* 2016, Tsutsui 2017, Andrews *et al* 2018, Silvers *et al* 2018, Dong *et al* 2020, Gregory *et al* 2020). Following common practice (e.g. Grose *et al* 2018, Flynn and Mauritsen 2020,

Meehl *et al* 2020, Smith *et al* 2020, Wang *et al* 2021), however, we take the value derived from regression of 150 years of an *abrupt-4xCO2* simulation, divided by two, to represent ECS.

The TCR values are taken from Meehl *et al* (2020) and are diagnosed from experiments in which CO₂ concentrations are increased by 1% per year (*1pctCO2*). The TCR is calculated as the temperature difference between pre-industrial conditions and the mean of a 20-year period centred on the 70th year of the experiment; the time at which the CO₂ concentration has doubled (Flato *et al* 2013). Similarly, the transient temperature response at the time of quadrupling of CO₂ in *1pctCO2* simulations is used e.g. by Grose *et al* (2018) and Sanderson (2020), and included here for comparison, referred to as T140.

2.2. Explaining variance in temperature evolution

The method for investigating the relationship between model simulated temperature change and different sensitivity indices in different forcing scenarios follows that of Grose *et al* (2018).

A running mean was calculated for the temperature change from 1861 to 2100 for each model and each scenario, for global mean and for specified regions respectively. Each individual year was represented by the mean of the 20 years centred on the year in question. The data were extrapolated 10 years past the year 2100 to allow for averages to be calculated in the last 9 years of the simulations, using coefficients obtained through a linear regression of the last 20 years of the data set. A similar extrapolation was performed at the beginning of the time series. The running mean minimizes the impact of interannual variability of the climate system on the warming (Grose *et al* 2018). For each year a linear regression was then performed between the GMSAT change since 1861 and the ECS, TCR, and T140, respectively, using the data from all available models.

In addition to the analysis of the different sensitivity indices, the same method was used to examine if historical forcing in models is related to their simulated past and future temperature evolution. The forcings used were the 1850–2014 global mean effective radiative forcing (ERF) from aerosols (F_{aer}) and well-mixed greenhouse gases (F_{GHG}), and the total anthropogenic ERF (F_{ant}). These were calculated from 30-year time slice experiments with sea surface temperatures (SSTs) fixed to pre-industrial conditions, with present-day aerosols, greenhouse gases and total anthropogenic forcing, respectively, compared to a pre-industrial control simulation, as in Smith *et al* (2020). Hence, like the sensitivity metrics, these forcing quantifications are not time dependent. Table S1 lists the forcings for the CMIP6 models providing the necessary output (a subset of which are given in Smith *et al* 2020). As for the sensitivity metrics, correlations between the forcing measures and the GMSAT change are used to quantify and

compare the skill of the different measures in explaining model spread in simulated past and future temperature change.

Globally gridded correlations between sensitivity indices and 1850–1869 to 2080–2099 temperature change under shared socioeconomic pathways (SSP5-8.5) were also calculated, and for this model data were interpolated to a common $1.5^\circ \times 1.5^\circ$ grid.

2.3. Future scenarios

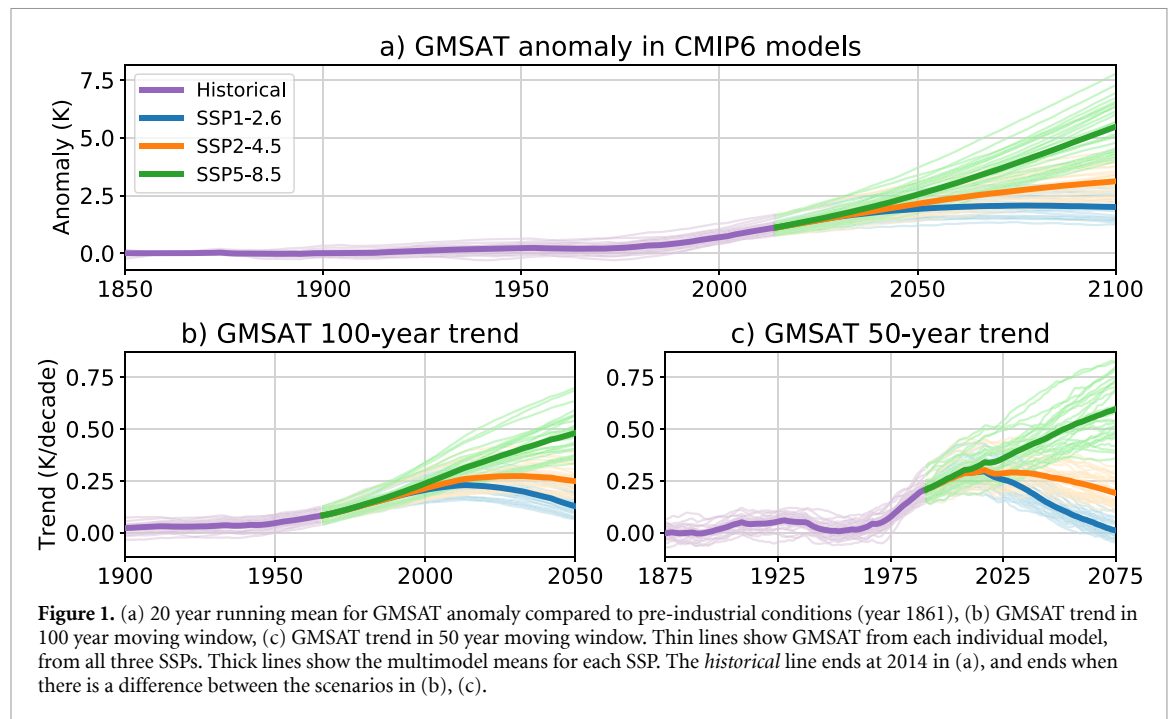
Grose *et al* (2018) studied the future temperature projections following the representative concentration pathways (RCPs) of CMIP5. The RCPs exemplify different scenarios of future warming, characterized by different pathways for the climate forcing. For example, RCP4.5 has a climate forcing of approximately 4.5 W m^{-2} by year 2100 (van Vuuren *et al* 2011). In CMIP6 the RCPs are combined with so called shared SSPs to create more complex scenarios, taking societal difficulties in mitigation of and adaption to climate change into account (O'Neill *et al* 2014). The scenarios that are available from most modelling centres are SSP1-2.6 (denoting the combination of SSP1 with RCP2.6), SSP2-4.5 and SSP5-8.5. The availability of SSP runs in the CMIP6 models considered is indicated in table S1.

2.4. Relation between forcing and sensitivity

The investigation of the relation between forcing and sensitivity follows Kiehl (2007) and Forster *et al* (2013), who surveyed that relationship in previous generation models. Consistent with these studies, the forcing is here estimated directly from the *historical* simulations rather than from single-forcing timeslice experiments as in section 2.2. Historical forcing estimates are based on equation (1), from which Kiehl (2007) argued that for a given energy imbalance and temperature change, such as those observed for the 20th century, there is an inverse relationship between forcing and ECS. The ECS is the temperature change after equilibrium has been reached following a doubling in CO₂, so at equilibrium, equation (1) becomes: $F_{2 \times \text{CO}_2} = \alpha \Delta T_{2 \times \text{CO}_2} = \alpha \text{ECS}$, meaning that equation (1) can be generally rewritten as:

$$F = N + \frac{F_{2 \times \text{CO}_2} \Delta T}{\text{ECS}}. \quad (2)$$

Using the method described by Forster *et al* (2013), the forcing in the *historical* simulation of each model was calculated through a two-step procedure based on equation (1). In the first step, the *abrupt-4xCO2* experiment was used in the same way as in the calculation of the ECS, as described in section 2.1. N was regressed against T to obtain α , where N and T are both defined as differences from pre-industrial conditions. In the second step, it was assumed that α is constant in time and independent of forcing, and equation (1) was applied on the temperature change and difference in TOA radiative imbalance since 1850



from the *historical* simulation to obtain the total forcing F_{total} at year 2003 (as the 2001–2005 average). Following Forster, the temperature and TOA radiation in the *abrupt-4xCO2* and *historical* experiments were first corrected for TOA radiative imbalance in the pre-industrial control simulation (*piControl*, see figure S1) and for any drift, by fitting and subtracting a linear trend for each model in the *abrupt-4xCO2* and *historical* experiments. Table S1 lists the adjusted ERF (F_{total}) for 2003 in the models where *abrupt-4xCO2* simulations were made available for the calculation. Again, the assumption that α is constant is in fact not valid, so although the method is consistent with previous studies, and across models, it introduces errors in the forcing estimate. For a smaller number of models in RFMIP (Radiative Forcing Model Intercomparison Project), it is possible to calculate the transient total forcing F from pre-industrial control simulations with all forcing agents, as described by Pincus *et al* (2016) and also used by Gregory *et al* (2020). We refer to this more correctly estimated total forcing as F_{RFMIP} , and list it in table S1, noting that it agrees within 10% with the F_{total} estimates in four out of five models, while for IPSL-CM6A-LR, F_{RFMIP} is 40% greater than F_{total} .

3. Results

3.1. Relating forcing and sensitivity metrics to past and projected temperature change

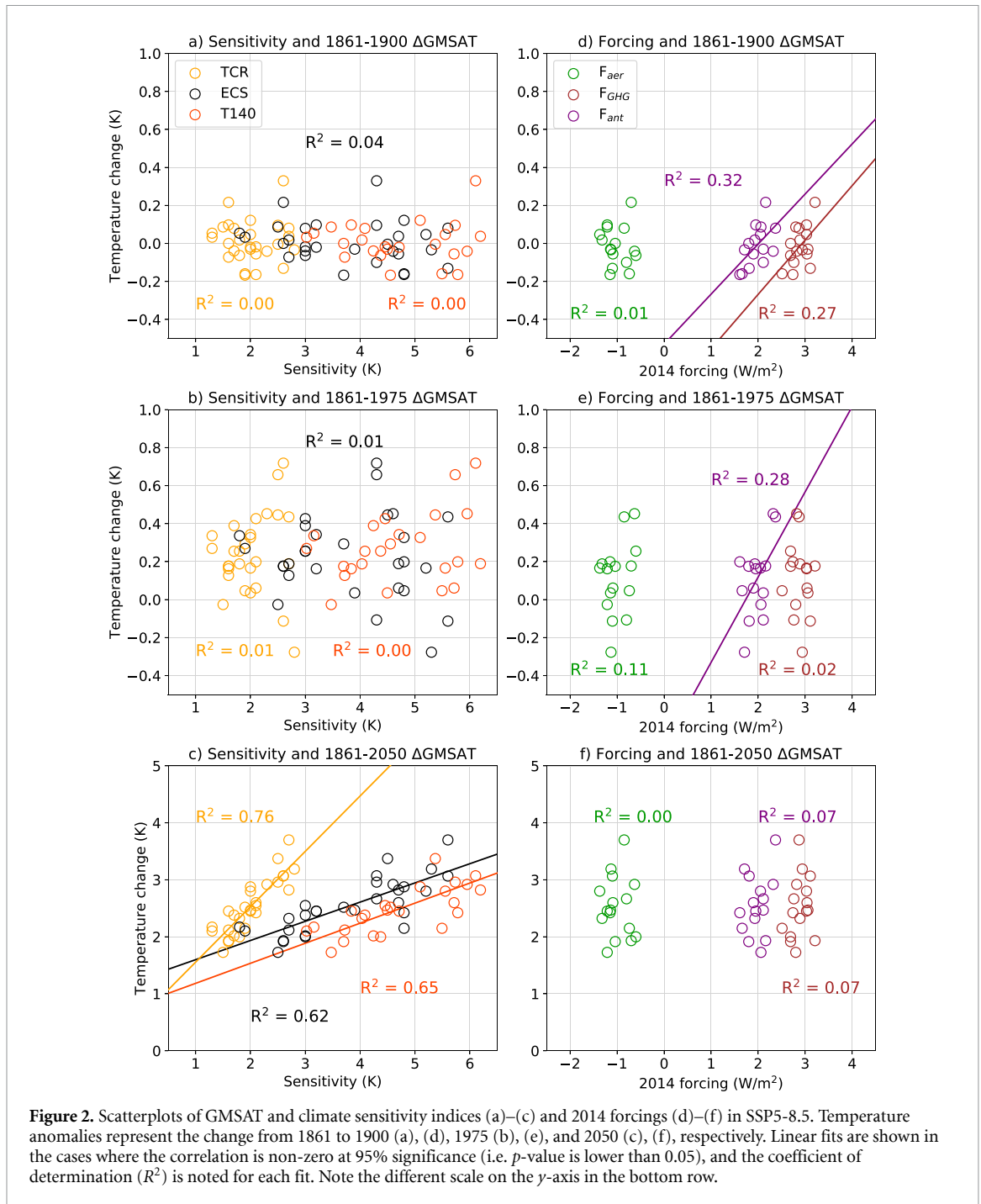
Figure 1 shows the development of the GMSAT anomalies in the models listed in table S1, in historical simulations and under three different future scenarios. In SSP1-2.6, the temperature stabilizes during the 21st century and the 50- and 100-year trends reach their maximum in the early 2000 s, while in SSP2-4.5

trends are reduced later and GMSAT is still increasing in 2075, and in SSP5-8.5 the 50- and 100-year trends in temperature continue to increase and the temperature anomalies grow increasingly towards the end of the 21st century.

Figure 2 shows snapshots of the relation between the sensitivity metrics (ECS, TCR and T140) and 1850–2014 forcings (F_{aero} , F_{GHG} and F_{ant}), respectively, and temperature change since 1861 for 20-year averages centred around the years 1900, 1975, and 2050 (under SSP5-8.5). These three years were selected to illustrate the contrast between the historical period, when F_{ant} dominates the correlation to temperature change with varying contribution from F_{GHG} and F_{aero} , and the future period when sensitivity dominates (see further figure 3). In each case the coefficient of determination, R^2 , and statistical significance at the 95% level is indicated in figure 2.

The 1861–1900 GMSAT anomaly is small, and sometimes even negative (indicating that the trend is negligible compared to the internal variability in temperature), and correlations with sensitivity metrics are negligible. For the period 1861–1975 when the temperature anomaly is still small, this remains the case, but for the future projection period 1861–2050 (SSP5-8.5) the sensitivity metrics display a positive correlation with the GMSAT anomaly, and TCR more so than ECS. The spread in sensitivity among models is also smaller for TCR. T140 has a similar correlation as ECS with GMSAT anomaly at this time.

The correlation between total anthropogenic forcing and GMSAT anomaly, on the other hand, is only significant for the historical snapshots (1900 and 1975) and near zero for the future projection snapshot (figure 2). The correlation with individual forcing

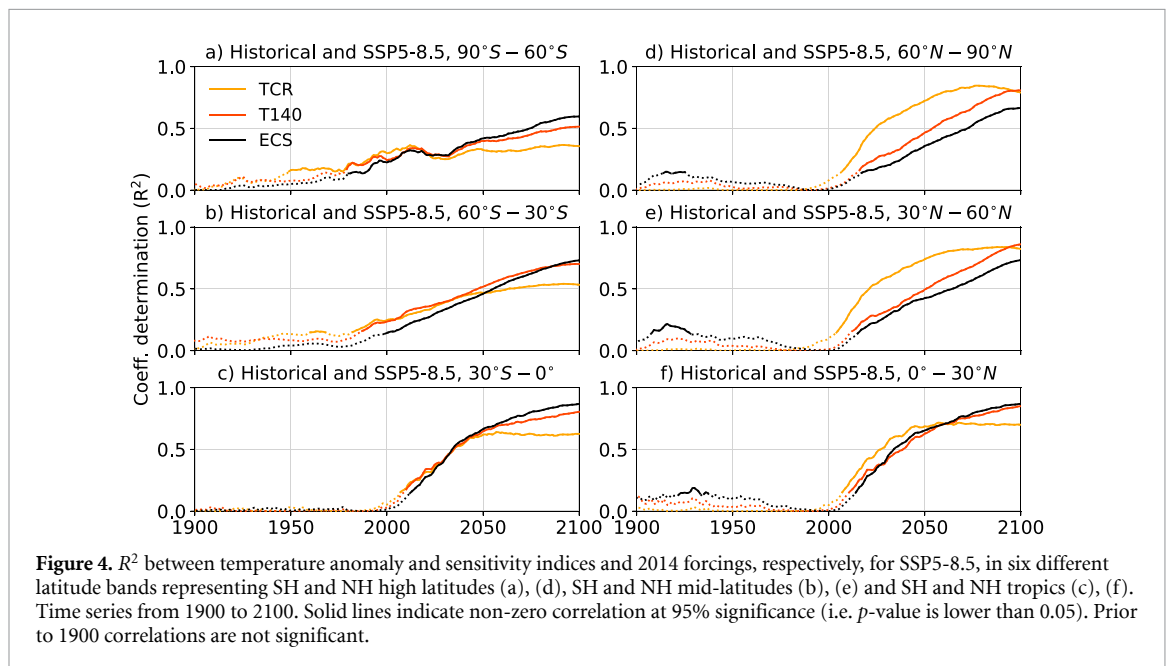
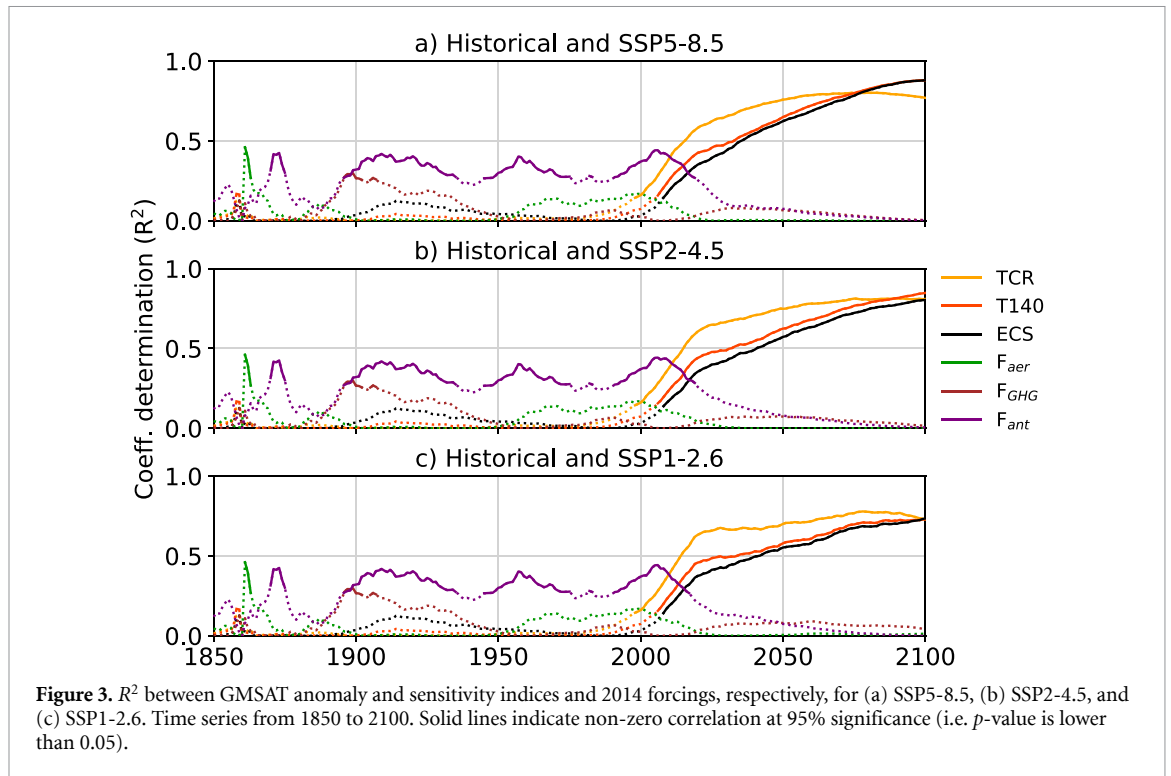


components (F_{aer}, F_{GHG}) is for all cases small, and statistically significant only for F_{GHG} at 1900.

Figure 3 shows how R^2 for the correlation between GMSAT, sensitivity metrics and forcing measures changes over time for the three SSPs, from the year 1850 to 2100. The relationship between warming and climate sensitivity is negligible in the historical period, increases strongly in the early 2000s and then remains high in all three future scenarios. The total anthropogenic forcing instead correlates with the modelled temperature anomaly during most of the 20th century, but does not determine model spread in the future projections.

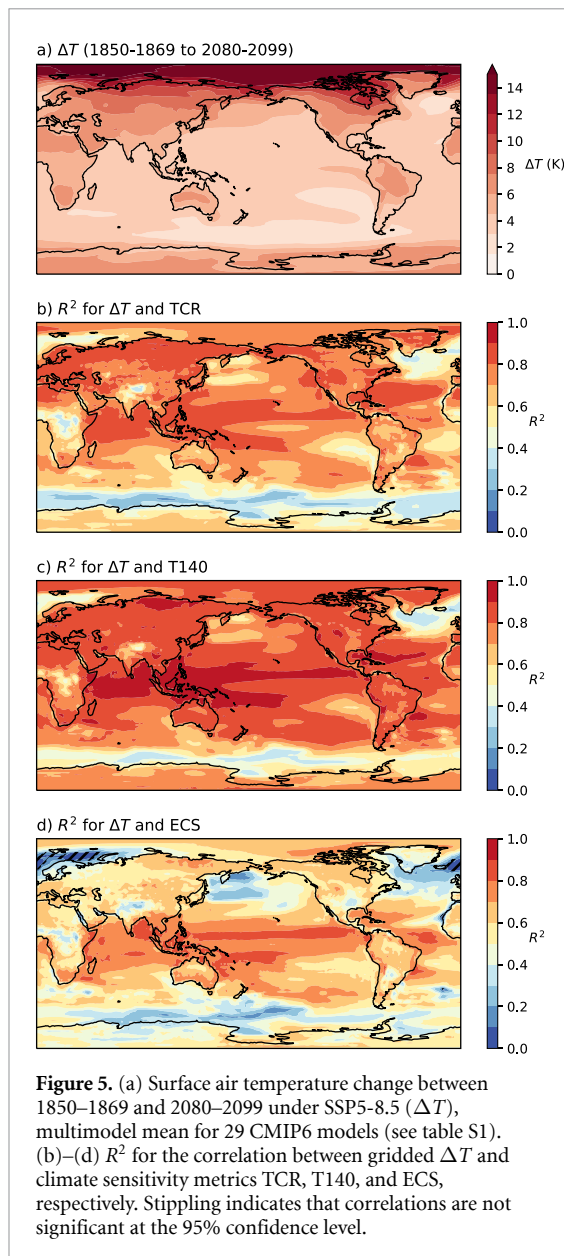
This confirms the picture of warming during the historical period being determined by model forcing strength while feedback strength, and hence sensitivity metrics, play a more prominent role in predicting temperature change in the future, when CO_2 forcing dominates and total forcing is less uncertain (Crook and Forster 2011, Forster *et al* 2013, Grose *et al* 2018, Lutsko and Popp 2019). The individual components of anthropogenic forcing, compared to pre-industrial conditions, remain largely uncorrelated with temperature anomaly throughout the period.

For comparison of the different sensitivity indices, it is useful to look first at the SSP5-8.5 scenario



(figure 3(a)), which is most similar to the $1pctCO_2$ simulations that TCR and T140 are derived from. Here, R^2 for TCR levels out at ca. 0.75 by the end of the 21st century, meaning that the inter-model spread in TCR explains 75% of the variance in the inter-model spread in projected change in GMSAT. For T140 and ECS, the degree of explanation becomes even larger, with an R^2 of 0.9 by 2100. Considering the effects of warming patterns and time-dependent feedbacks (section 2.1), this is consistent with TCR representing a sensitivity that includes only the feedbacks that dominate in early stages of the warming, whereas delayed feedbacks are to a greater extent captured by

T140, taken 70 years later, and of course also by ECS that includes the adjustment of ocean heat uptake to equilibrium, that neither of the transient measures can capture. In the mitigation scenarios SSP5-4.5 and SSP1-2.6 (figures 3(b) and (c)), where the temperature is more stabilized, ECS does not surpass the transient measures in predictive power within the 21st century. In the CMIP6 *abrupt-4xCO2* simulations, from which the ECS is estimated, there is a shift in the feedback parameter after around 3–5 K warming (on average 4 K, see figure 3 of Meehl *et al* 2020). This warming is reached in SSP5-8.5 around the same time as R^2 for ECS surpasses those for the transient



measures (figures 1 and 3), while in the mitigation scenarios an average warming of 4 K is not reached within the 21st century. Hence, the weaker correlation between ECS and GMSAT change in the mitigation scenarios could relate to the weaker warming in those scenarios compared to the stronger forcing scenario. These results differ slightly from those of Grose *et al* (2018), who in their figure 2 showed the ECS R^2 in the CMIP5 models to meet or surpass the TCR R^2 earlier for all scenarios.

For correlations with temperature trends (figure S2), the R^2 value for ECS becomes larger than that for TCR by the year 2000 in all scenarios. This confirms the greater relevance of ECS on longer time scales, beyond 2100, seen in figure 3, as the later and larger GMSAT anomalies increasingly contribute to the temperature trend.

The climate sensitivity is greater, and the global mean temperature trend is overall larger in CMIP6 than in CMIP5 in all three scenarios reviewed

(comparing figure 1 with a and b of figure 3 of Grose *et al* 2018). This may partly explain the differences in predictive power of ECS and TCR between CMIP5 and CMIP6, for both temperature anomalies (figure 3) and temperature trends (figure S2). A stronger positive temperature trend and higher sensitivity indicates a climate with a longer response time, which is farther from equilibrium, and for which ECS is a less suitable measure of the temperature evolution. A relation between response time and sensitivity follows from equation (1) with assumed time-invariant α (see Hansen *et al* 1985). It is also manifest as a positive correlation between ECS and response time scale for the *abrupt-4xCO2* simulations in the studied model ensemble (figure S3). Dividing the CMIP6 models into two subsets based on their sensitivity indeed suggests that ECS performs better for the lower sensitivity models, and less well for the higher sensitivity models, but the robustness of this analysis is limited by the small sample size (figure S4).

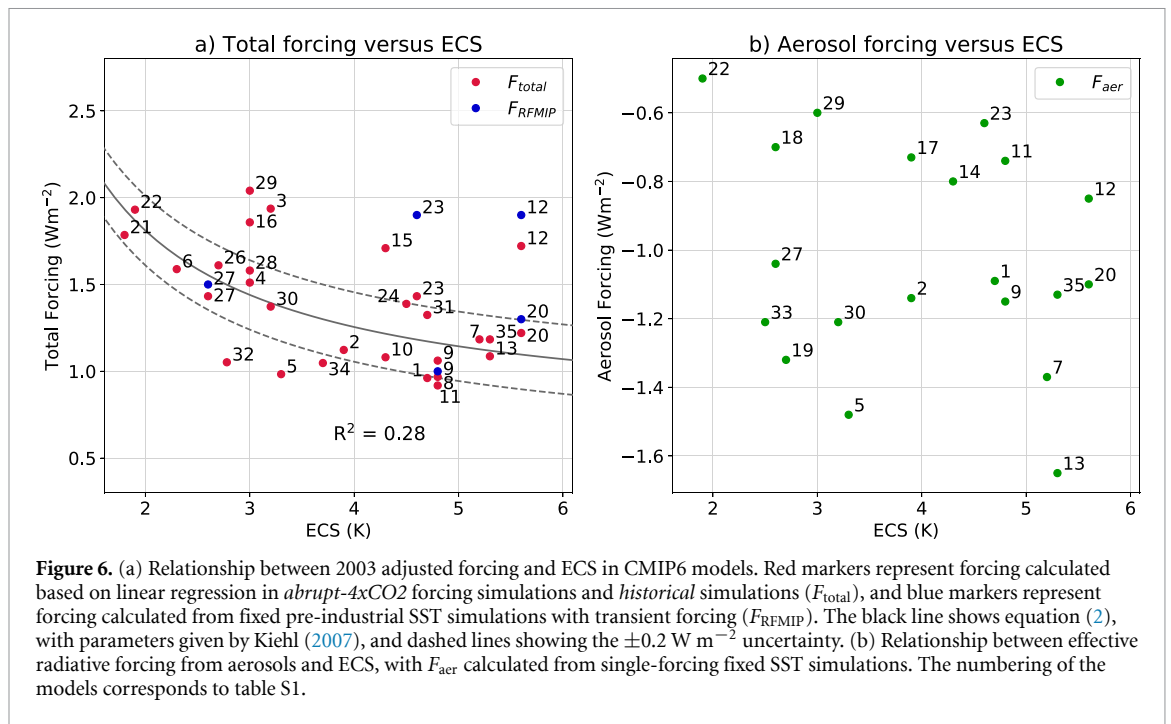
3.2. Regional variations

Further insight can be gained from separating the GMSAT into land and ocean, which yields higher correlation with ECS earlier for the ocean only case (see figure S5). This is in line with ECS better describing slow or delayed warming, and shifts in feedback dominance from evolving temperature patterns occurring over the ocean.

An even greater difference, however, is seen from separating the correlations with temperature anomaly into northern (NH) and southern hemisphere (SH) averages. In the SH, ECS clearly dominates the degree of explanation of projected temperature anomaly (see figure S6), which can only partly be explained by the ocean dominance of the SH. A more detailed separation of the temperature anomaly into latitude bands, shown in figure 4, points specifically at the tropics in SH and NH, and SH mid-latitudes as regions where the correlation between temperature anomaly and ECS becomes quite high. In the NH mid- and high latitudes, where the early transient warming is large, TCR explains more variation in model spread in temperature.

The relation between regional temperature change and global mean sensitivity metrics is further illustrated in figure 5 which shows R^2 for the correlation between 1850–1869 and 2080–2099 warming and the three sensitivity indices, respectively, for each grid point on the map. This figure can be compared with figure 4 in Grose *et al* (2018), except that it uses a different reference period for the warming.

The correlation overall increases with stronger forcing scenario (not shown), and figure 5 shows SSP5-8.5, where the temperature change is most prominent at the end of the century. T140 has the overall highest correlation values, i.e. this sensitivity metric best describes the pattern of regional temperature anomaly at the end of the



century. Compared to CMIP5 (figure 4 of Grose *et al* 2018), the pattern for correlation with ECS is similar, but both TCR and T140 are in CMIP6 better correlated with regional temperature anomalies, particularly in the NH. In agreement with Grose *et al* (2018), however, correlations are weak in the Southern Ocean and the North Atlantic for all three sensitivity indices. These are areas where the relative standard deviation in temperature is large (not shown) and the low correlation indicates that the regional temperature anomaly in those areas is not directly coupled to the global mean temperature change, neither transient (TCR, T140) nor approximately equilibrated (ECS).

Even though the gridded correlations in figures 5(b)–(d) are almost exclusively statistically significant, we cannot assign statistical significance to the geographical distribution of their differences, with the small sample size.

3.3. Are forcing and sensitivity correlated?

Using model spread in sensitivity and historical forcing to explain variability in past and future simulated temperature also leads to the question whether forcing strength and sensitivity in the models are related, so that high-sensitivity models are also low-forcing models, and vice versa.

Figure 6(a) shows adjusted ERF and ECS for each of the CMIP6 models where total forcing could be calculated (see section 2.4) with the method of Forster *et al* (2013), and in addition those from the RFMIP pre-industrial SST simulations. The theoretical line from Kiehl (2007), valid for the 20th century temperature change and change in ocean heat content, is included in the figure. A majority of the models fall within the given uncertainty range of

$\pm 0.2 \text{ W m}^{-2}$, but there is significant spread around the line. The value for R^2 is low (0.28 when calculated with respect to the theoretical line and 0.31 when calculated with respect to the least-squares best fit), comparable to what Forster *et al* (2013) found for CMIP5 models, and not large enough to suggest a direct compensation between forcing and sensitivity. Figure 6(b) in turn shows the ERF from aerosol forcing only (F_{aer}) in relation to ECS, as also shown by Smith *et al* (2020) and Meehl *et al* (2020) for two different and partly overlapping subsets of CMIP6 models. Our figure includes those models, and an additional six models, as described in section 2.4, and no significant correlation is found ($R^2 = 0.00$).

In figure 6(a) the forcing is taken as an average for five years (2001–2005 for 2003), but interannual variability, not least in the natural component of the total forcing, makes the analysis sensitive to the choice of year, and for example 2008 (2006–2010) and 2012 (2010–2015) have greater spread, and R^2 values of 0.19 and 0.17 respectively, calculated for the best-fit lines in each case (not shown).

4. Discussion and conclusion

We have presented relations between climate sensitivity indices, forcing and temperature change in an ensemble of CMIP6 models. Comparing the behavior of these models to that of earlier generation models provides both confirmation and contradiction of previous findings.

As expected, models with higher sensitivity display greater warming, but compared to the findings of Grose *et al* (2018) for CMIP5, we see more clearly for CMIP6 that the transient sensitivity metrics (TCR

and/or T140) remain similar or superior to ECS in terms of degree of explanation of projected global mean temperature anomaly throughout the 21st century, across three future scenarios: SSP1-2.6, SSP2-4.5 and SSP5-8.5. However, in the SSP5-8.5 scenario in particular, the predictive power of TCR becomes smaller relative to that of T140 and ECS before the end the century. This is consistent with evolving SST patterns changing the balance of feedbacks, and TCR (that is defined as the warming at year 70 of a transient simulation) not being able to capture delayed feedbacks, the effects of which T140, and of course ECS, can better incorporate.

The relative and absolute increase in ECS correlation, over TCR, is indeed seen particularly in regions where delayed feedbacks due to changing SST patterns are expected to occur, such as the Southern Ocean and the equatorial Pacific (see Dong *et al* 2020). Hence, contrary to Grose *et al* (2018), we argue that in areas where warming is delayed compared to the global mean (like the Southern Ocean and North Atlantic), TCR does not well represent the evolution of the regional mean temperature, and ECS instead gains predictive power. It is also clear that these slowly warming areas are those where the local temperature anomaly at the end of the century is least related to the global mean warming, and global mean measures of sensitivity.

A possible explanation for the overall greater and longer lasting correlations of GMSAT with TCR in the current analysis, compared to that of Grose *et al* (2018), may be the greater sensitivity and warming trend in the CMIP6 model ensemble compared to the CMIP5, rendering a temperature evolution farther from equilibrium that makes ECS less relevant for a longer time.

Grose *et al* (2018) did not include any comparison to different forcing agents, but confirmed that historical warming is related to model forcing while feedback strength plays a greater role in predicting future temperature change (Crook and Forster 2011, Forster *et al* 2013). This is reaffirmed by the results of this study, both by the fact that the relationship between warming and climate sensitivity is stronger in future scenarios than in the past, and that there is a positive correlation between warming and total anthropogenic forcing during most of the 20th century. The model spread in historical forcing (difference between 1850 and 2014) provides some degree of explanation for the spread in historical evolution of GMSAT in the models (20%–40%), i.e. models with larger total forcing (related to less negative aerosol forcing) have warmed more throughout the 20th century. The model spread in historical forcing however rapidly loses predictive power towards the end of the 20th century, as the importance of sensitivity increases. Correlations for individual anthropogenic forcing components (from GHG and aerosol) are low for the whole historical period and a separation

in time of their relevance as predictor of temperature anomaly cannot be made. Aerosol forcing does indeed reach a maximum during the second half of the 20th century (Shindell *et al* 2013), but the single measure of forcing at 2014 does not capture this temporal variation. Compared to the sensitivity metrics, the model spread in forcing is small, and cannot explain the spread in temperature anomalies, which also remains small over the historical period compared to the future projections. The lack of correlation between historical forcing and future projection may be seen as a reassurance that models are not tuned too heavily to the historical forcing, as that would have affected their feedbacks, and simulated warming (see Lutsko and Popp 2019).

Forcing and climate sensitivity are key factors for determining the Earth's temperature evolution, and it is possible that intermodel variation in ECS is compensated by intermodel variation in forcing, resulting in models with very different ECS replicating a similar historical warming (Kiehl 2007, Knutti 2008). While Kiehl (2007) initially demonstrated such an inverse relation between ECS and forcing in nine coupled models, prior to CMIP3, Forster *et al* (2013) performed the same analysis on CMIP5 models and found that only a subset of models, within the 90% uncertainty range of the observed 100 year linear temperature trend, fit the expected inverse relationship suggested by Kiehl (2007). Accordingly, as expected for models reasonably reproducing the observed warming, our results indicate some adherence among the CMIP6 models to the theoretical relation described by Kiehl (2007), but with significant spread and a weak correlation ($R^2 = 0.28$ for 2003, and even lower for other time periods).

The explanation for the partial compensation between sensitivity and forcing has particularly been sought in aerosol forcing, which indeed contributes the greatest uncertainty to the total forcing, but with inconclusive results. While Smith *et al* (2020) find a weak non-significant positive correlation between F_{aer} and ECS among a set of CMIP6 models, which they argue suggests that models are not tuning present day aerosol forcing to reproduce observed warming, Meehl *et al* (2020), using a different subset of CMIP6 models, actually find a weak negative correlation. Contrary to Chylek *et al* (2016) they are not able to relate differences to model aerosol representation complexity. Wang *et al* (2021) take yet another step, and show specifically that models with more negative aerosol-cloud interaction have more positive cloud feedback, which is reasonable given that these are in turn the greatest contributors to uncertainty in F_{aer} and ECS, respectively (e.g. Bellouin *et al* 2020, Zelinka *et al* 2020). The difference between F_{aer} and ECS correlations found by Meehl *et al* (2020) and Smith *et al* (2020) is suggested by Wang *et al* (2021) to be due to the set of models by Smith *et al* (2020) being less consistent with the observed temperature

record, but this is not a fully adequate explanation: the index used by Wang *et al* (2021) to quantify deviation from observed temperature is in fact distributed very similarly across the model subsets (median and standard deviation is 0.24 ± 0.10 for all models listed by Wang *et al* (2021), 0.24 ± 0.09 for those used by Smith *et al* (2020) and 0.24 ± 0.11 for those in Meehl *et al* (2020)).

With our larger set of CMIP6 models, we find the covariation between F_{aer} and ECS to be negligible. This shows that temperature change is not being used systematically as a direct or single constraint on the balance between anthropogenic forcing and climate sensitivity in the models, and alludes to the greater complexity of model tuning: neither aerosol forcing, ECS, cloud feedback nor aerosol-cloud interaction are single tunable parameters, and global mean temperature is not the only tuning target (Bender 2010, Mauritsen *et al* 2012, Hourdin *et al* 2017, Schmidt *et al* 2017, Mauritsen and Roeckner 2020).

Our results are of use in the continued work on emergent constraints on climate sensitivity, as a guidance to what measure of sensitivity is actually the most meaningful to constrain, depending on the time period in focus. Based on the CMIP6 model ensemble studied, TCR has more predictive power than ECS for the temperature evolution during the remainder of the century, in mitigation scenarios. For the most business-as-usual like scenario, ECS takes over with higher correlation with model spread in temperature projection around 2075. In both cases, the longer-term transient measure T140 remains at a high correlation with GMSAT. TCR also already varies less among the models than ECS, although model agreement is of course only a necessary and not a sufficient condition for a correct estimate. The spread among models in a given scenario, determined largely by the model sensitivities, is the range of future climate evolution available to the community and to policy makers, and our results give an indication as to how that range can be narrowed by constraints not least on transient climate sensitivity metrics.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://esgf-node.llnl.gov/search/cmip6/>.

Acknowledgments

F B and L H acknowledge financial support from the Swedish Research Council for Sustainable Development (Formas), Grant No. 2018-01797, and the Swedish e-Science Research Centre (SeRC). Funding for F B, A E and T S is also partly provided by the FORCeS project funded by the European Union's Horizon 2020 programme, Grant Agreement No. 821205. We acknowledge the World Climate Research

Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. The data handling was enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at the National Supercomputer Centre (NSC) partially funded by the Swedish Research Council through Grant Agreement No. 2016-07213.

ORCID iDs

Linnea L Huusko  <https://orcid.org/0000-0003-3298-6153>

Frida A-M Bender  <https://orcid.org/0000-0003-4867-4007>

Annica M L Ekman  <https://orcid.org/0000-0002-5940-2114>

Trude Storelvmo  <https://orcid.org/0000-0002-0068-2430>

References

- Allen M R and Frame D J 2007 Call off the quest *Science* **318** 582–3
- Andrews T *et al* 2018 Accounting for changing temperature patterns increases historical estimates of climate sensitivity *Geophys. Res. Lett.* **45** 8490–9
- Armour K C, Bitz C M and Roe G H 2013 Time-varying climate sensitivity from regional feedbacks *J. Clim.* **26** 4518–34
- Bellouin N *et al* 2020 Bounding global aerosol radiative forcing of climate change *Rev. Geophys.* **58** e2019RG000660
- Bender F A-M 2010 A note on the effect of GCM tuning on climate sensitivity *Environ. Res. Lett.* **3** 014001
- Bloch-Johnson J, Pierrehumbert R T and Abbot D S 2015 Feedback temperature dependence determines the risk of high warming *Geophys. Res. Lett.* **42** 4973–80
- Collins M *et al* 2013 Long-term climate change: projections, commitments and irreversibility *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* ed T F Stocker *et al* (Cambridge: Cambridge University Press)
- Crook J A and Forster P M 2011 A balance between radiative forcing and climate feedback in the modeled 20th century temperature response *J. Geophys. Res.* **116** D17
- Crook J A and Forster P M 2016 Indirect aerosol effect increases CMIP5 models' projected arctic warming *J. Climate* **29** 1417–28
- Dong Y, Armour K C, Zelinka M D, Proistosescu C, Battisti D S, Zhou C and Andrews T 2020 Intermodel spread in the pattern effect and its contribution to climate sensitivity in CMIP5 and CMIP6 models *J. Clim.* **33** 7755–75
- Eyring V, Bony S, Meehl G A, Senior C A, Stevens B, Stouffer R J and Taylor K E 2016 Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization *Geosci. Model Dev.* **9** 1937–58
- Flato G *et al* 2013 Evaluation of climate models *climate change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* ed T F Stocker *et al* (Cambridge: Cambridge University Press)
- Flynn C and Mauritsen T 2020 On the climate sensitivity and historical warming evolution in recent coupled model ensembles *Atmos. Chem. Phys.* **20** 7829–42

- Forster P M *et al* 2020 Latest climate models confirm need for urgent mitigation *Nat. Clim. Change* **10** 7–10
- Forster P M, Andrews T, Good P, Gregory J M Jackson L S and Zelinka M 2013 Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models *J. Geophys. Res. Atmos.* **118** L139–50
- Frame D J, Booth B B B, Kettleborough J A, Stainforth D A, Gregory J M, Collins M and Allen M R 2005 Constraining climate forecasts: the role of prior assumptions *Geophys. Res. Lett.* **32** L09702
- Gregory J M, Andrews T, Ceppi P, Mauritsen T and Webb M J 2020 How accurately can the climate sensitivity to CO₂ be estimated from historical climate change? *Clim. Dyn.* **54** 129–57
- Gregory J M, Andrews T and Good P 2015 The inconstancy of the transient climate response parameter under increasing CO₂ *Phil. Trans. R. Soc. A* **373** 20140417
- Gregory J M, Ingram W J, Palmer M A, Jones G S, Stott P A, Thorpe R B, Lowe J A, Johns T C and Williams K D 2004 A new method for diagnosing radiative forcing and climate sensitivity *Geophys. Res. Lett.* **31** 3
- Grose M R, Gregory J, Colman R and Andrews T 2018 What climate sensitivity index is most useful for projections? *Geophys. Res. Lett.* **45** 1559–66
- Hansen J, Russell G, Laci A, Fung I, Rind D and Stone P 1985 Climate response times: dependence on climate sensitivity and ocean mixing *Sci.* **229** 857–9
- Hourdin F *et al* 2017 The art and science of climate model tuning *Bull. Amer. Meteor. Soc.* **98** 1559–66
- Jonko A K, Shell K M, Sanderson B M and Danabasoglu G 2013 Climate feedbacks in ccs3 under changing CO₂ forcing. Part II: variation of climate feedbacks and sensitivity with forcing *J. Climate* **26** 2784–95
- Kiehl J T 2007 Twentieth century climate model response and climate sensitivity *Geophys. Res. Lett.* **34** L22710
- Knutti R 2008 Why are climate models reproducing the observed global surface warming so well? *Geophys. Res. Lett.* **35** L18704
- Knutti R, Rugenstein M and Hegerl G 2017 Beyond equilibrium climate sensitivity *Nat. Geosci.* **10** 727–36
- Lutsko N J and Popp M 2019 Probing the sources of uncertainty in transient warming on different timescales *Geophys. Res. Lett.* **46** 11,367–11,377
- Marvel K *et al* 2016 Implications for climate sensitivity from the response to individual forcings *Nat. Clim. Change* **6** 386–9
- Mautitsen T *et al* 2012 Tuning the climate of a global model *J. Adv. Model. Earth Syst.* **4** 3
- Mautitsen T and Roeckner E 2020 Tuning the MPI-ESM1.2 global climate model to improve the match with instrumental record warming by lowering its climate sensitivity *J. Adv. Model. Earth Syst.* **12** e2019MS002037
- Meehl G A, Senior C A, Eyring V, Flato G, Lamarque J-F, Stouffer R J, Taylor K E and Schlund M 2020 Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 *Sci. Adv.* **6** eaba1981
- Nijse F J M M, Cox P M and Williamson M S 2020 Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models *Earth Syst. Dynam.* **11** 737–50
- O'Neill B C *et al* 2014 The roads ahead: narratives for shared socioeconomic pathways describing world futures in the 21st century *Global Environ. Change* **42** 169–80
- Pfister P L and Stocker T F 2017 State-dependence of the climate sensitivity in earth system models of intermediate complexity *Geophys. Res. Lett.* **44** 20
- Pincus R, Forster P M and Stevens B 2015 The Radiative Forcing Model Intercomparison Project (RFMIP): experimental protocol for CMIP6 *Geosci. Model. Dev.* **9** 3447–60
- Richardson T B *et al* 2019 Efficacy of climate forcings in PDRMIP models *J. Geophys. Res.* **124** 12824–44
- Rugenstein M *et al* 2020 Equilibrium climate sensitivity estimated by equilibrating climate models *Geophys. Res. Lett.* **47** e2019GL083898
- Sanderson B 2020 Relating climate sensitivity indices to projection uncertainty *Earth Syst. Dyn.* **11** 721–35
- Schmidt G A, Bader D, Donner L J, Elsaesser G S, Golaz J-C, Hannay C, Molod A, Neale R B and Saha S 2017 Practice and philosophy of climate model tuning across six US modeling centers *Geosci. Model. Dev.* **10** 3207–23
- Sherwood S *et al* 2020 An assessment of Earth's climate sensitivity using multiple lines of evidence *Rev. Geophys.* **58** e2019RG000678
- Shindell D T *et al* 2013 Radiative forcing in the ACCMIP historical and future climate simulations *Atmos. Chem. Phys.* **13** 2939–74
- Silvers L G, Paynter D and Zhao M 2018 The diversity of cloud responses to twentieth century sea surface temperatures *Geophys. Res. Lett.* **45** 391–400
- Smith C J *et al* 2020 Effective radiative forcing and adjustments in CMIP6 models *Atmos. Chem. Phys.* **20** 9591–9618
- Stevens B, Sherwood S C, Bony S and Webb M J 2016 Prospects for narrowing bounds on Earth's equilibrium climate sensitivity *Earth's Future* **4** 512–22
- Tokarska K B, Stolpe M B, Sippel S, Fischer E M, Smith C J, Lehner F and Knutti R 2020 Past warming trend constrains future warming in CMIP6 models *Sci. Adv.* **6** 12
- Tsutsui J 2017 Quantification of temperature response to CO₂ forcing in atmosphere-ocean general circulation models *Climate Change* **140** 287–305
- van Vuuren D P *et al* 2016 The representative concentration pathways: an overview *Climate Change* **109** 5
- Wang C, Soden B, Yang W and Vecchi G 2021 Compensation between cloud feedback and aerosol-cloud interaction in CMIP6 models *Geophys. Res. Lett.* **48** 4
- Yoshimori M, Watanabe M, Shiogama H, Oka A, Abe-Ouchi A, Ohgaito R and Kamae Y 2014 A review of progress towards understanding the transient global mean surface temperature response to radiative perturbation *Prog. Earth Planet. Sci.* **3** 21
- Zelinka M D *et al* 2020 Causes of higher climate sensitivity in CMIP6 models *Geophys. Res. Lett.* **47** e2019GL085782