CLIMATOLOGY

Past warming trend constrains future warming in CMIP6 models

Katarzyna B. Tokarska[1]*[†], Martin B. Stolpe[1]*, Sebastian Sippel[1], Erich M. Fischer[1], Christopher J. Smith[2], Flavio Lehner[1], Reto Knutti[1]

Future global warming estimates have been similar across past assessments, but several climate models of the latest Sixth Coupled Model Intercomparison Project (CMIP6) simulate much stronger warming, apparently inconsistent with past assessments. Here, we show that projected future warming is correlated with the simulated warming trend during recent decades across CMIP5 and CMIP6 models, enabling us to constrain future warming based on consistency with the observed warming. These findings carry important policy-relevant implications: The observationally constrained CMIP6 median warming in high emissions and ambitious mitigation scenarios is over 16 and 14% lower by 2050 compared to the raw CMIP6 median, respectively, and over 14 and 8% lower by 2090, relative to 1995–2014. Observationally constrained CMIP6 warming is consistent with previous assessments based on CMIP5 models, and in an ambitious mitigation scenario, the likely range is consistent with reaching the Paris Agreement target.

INTRODUCTION

Both international climate assessments [e.g., Intergovernmental Panel on Climate Change (IPCC) Assessment Reports (1)] and national climate scenarios rely heavily on results from multiple climate model simulations collected in model intercomparisons. Hence, the reliability of and confidence in these model intercomparisons have a wide-ranging influence on science and ultimately policy-targeted science communication. Model intercomparisons have always featured diverging model projections, for example, for the question of how much warming to expect for a doubling of global atmospheric CO_2 concentration. However, the spread across such ad hoc model ensembles of opportunity is challenging to interpret (2). This is because not all models are equally plausible (3), and the multimodel spread may be partly inconsistent with evidence from observations, theory, or process understanding. The range of models may be too wide when unrealistic models are included or too narrow when models underestimate uncertainties from processes that are not or poorly represented. The multimodel mean may be biased high or low when many models are biased in the same way or when near-duplicate models are included (4). It is therefore essential to relate and, when necessary, recalibrate (e.g., by reweighting models) the raw spread of such model ensembles, based on other constraints from process evidence, past trends, climatology, or probabilistic estimates from perturbed physics ensembles, to produce projections (including robust uncertainty estimates) of future climate that are consistent with our understanding and with observations of the current climate.

The long-term warming range of the Coupled Model Intercomparison Project Phase 5 (CMIP5) (5) models was interpreted in the IPCC Fifth Assessment Report (AR5) (1) to be unbiased in its raw mean, but the 5 to 95% ranges in global temperature projections were interpreted as "likely" (>66% probability) to account for structural model uncertainties. Phase 6 of the Coupled Model Intercomparison Project (CMIP6) will inform much of the physical science

basis for the upcoming Sixth Assessment Report (AR6) of the IPCC (6). It includes the latest generation of comprehensive Earth system models (ESMs), driven by historical greenhouse gas concentrations, and followed by different future greenhouse gas and aerosol concentrations according to the Shared Socioeconomic Pathways (SSP) scenarios (7). The first models submitted to the archive suggest that CMIP6 will span a wider range of warming responses than CMIP5. Several ESMs submitted to CMIP6 have equilibrium climate sensitivity (ECS) values (table S1) higher than any of the CMIP5 models (8), and a third of CMIP6 models submitted to date (10 of 29 models; table S1) exceed the range of 1.5° to 4.5°C for ECS assessed as likely (17 to 83% range) in the IPCC AR5 report. Note that for simplicity we use the term "equilibrium climate sensitivity," although the values are derived from nonequilibrium conditions and rather represent the "effective climate sensitivities" [i.e., a measure of the feedbacks during the transient regime that is extrapolated to equilibrium (9)]. As a result of higher climate sensitivity values, future climate projections from these models show stronger future global mean warming than the warming previously reported in AR5, although a direct comparison is challenging due to a novel generation of emission scenarios used to drive the models (10). Some models, for instance, project warming of 2.5° to 3°C for scenarios that were designed to be consistent with the Paris temperature target of well below 2°C (7). Therefore, the critical question arises whether projections of such models with high future warming are realistic. If they are, that would result in much higher risks and costs of future climate change than previously assessed and imply even faster mitigation to achieve climate targets. If the models, on the other hand, are biased high, that would imply that climate assessments need to recalibrate the raw ensemble.

A more near-term (transient) global warming that arises after 70 years of a 1% per year increase in atmospheric CO_2 concentration is referred to as the transient climate response (TCR). TCR and ECS metrics are often used to develop and calibrate simple climate model emulators, which are used with integrated assessment models and provide policy-relevant information regarding emission pathways and related climate responses (11). Estimates of TCR also affect the allowed carbon emissions for the Paris Agreement climate target (12) and are important for climate projections and risk assessment

[1]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland. [2]Priestley International Centre for Climate, University of Leeds, Leeds LS2 9JT, UK.
*These authors contributed equally to this work.
†Corresponding author. Email: kasia.tokarska@env.ethz.ch

(13), with substantial economic benefits resulting from narrowing down the TCR range (14). Therefore, consistency of the simulated TCR range with observational evidence is crucial and potentially narrowing the spread of TCR benefits not only the climate science community but also many other sectors.

Here, we make use of an emergent relationship between the simulated warming trend in recent decades and projected future warming in different emission scenarios, as well as between the simulated warming and the more idealized metric of future warming (TCR). On the basis of these correlations across models, we constrain the ranges of TCR and future warming projections.

RESULTS

For an emergent constraint to be robust, there needs to be an underlying physical explanation of why the correlation between the two quantities should exist in the first place (15). Here, we use the simulated historical global mean temperature over recent decades as an emergent constraint for the future warming in response to increasing CO_2 concentrations. To first order, the global temperature response is proportional to the radiative forcing, and the ratio of global mean warming to forcing is equivalent to TCR. As long as the forcing increases, stronger feedbacks imply more warming in both the past and the future. Dozens of studies have used this concept to constrain ECS, TCR, or future warming from past surface warming, forcing, and ocean heat uptake [see (16) for a review]. The relationship between past and future warming is often obscured through compensation of climate feedbacks and uncertain aerosol forcing, especially in the historical period (see below). However, this correlation becomes stronger when the greenhouse gas attributable warming (17) dominates the observed warming over recent decades, and it therefore constrains future warming, which is also dominated by greenhouse gases. The TCR or temperature projections to 2100 may appear to be more complex than ECS because they additionally involve ocean heat uptake (as the ocean is not in equilibrium). However, because such transient warming metrics specifically relate to a time scale of about a century, they suffer much less from the change in climate feedbacks on longer time scales (9). TCR is, therefore, better constrained by the observed warming, by using either energy balance arguments or detection and attribution studies. A discussion of the mechanisms explaining the correlation between the recent warming trends and TCR or ECS is also provided in (18), based on CMIP5 models and observed warming trends in an earlier historical period (1970–2005). While there are a number of other potential challenges such as the dependence of the transient response depending on the base state, the magnitude and the type of forcing, and the feedbacks being different for very short and long time scales [see (16) for a review], all of these contributing factors would weaken or destroy the emergent constraint rather than improve it, as long as the number of models is sufficiently large to avoid spurious correlation.

Ideally, the past period used as an emergent constraint on the future warming should thus be as representative as possible for the warming response to CO_2 and should thus fulfill the three criteria: (i) The period is sufficiently long such that the importance of internal variability is small; (ii) known modes of lower frequency variability in the Pacific and Atlantic Oceans that might influence observed global temperatures show small or compensating trends such that the observed warming is close to the forced response; and (iii) changes in other forcings such as aerosol forcing are small (note that the aerosol forcing does not need to be zero; it only needs to be approximately constant such that the warming is dominated by the change in the greenhouse gas forcing). Using the complete historical record from 1850 to present day fulfills (i) and (ii), but not (iii). Thus, the simulated global mean surface air temperature (GSAT) increase is only weakly correlated with TCR and ECS, as the relationship between warming and climate sensitivity is masked by large uncertainty in aerosol cooling (Fig. 1A) (19). Overestimating both the climate feedbacks and the aerosol forcing can result in a historical warming to present day that is similar as observed, but the temporal agreement with observations is poor, with little simulated warming until 1980 and too strong warming after [e.g., E3SM1 (20), UKESM1 (21), or GFDL-CM4 (22)].

However, after around the year 1980, the global mean aerosol cooling trend is very small and consistent across all available CMIP6 models (−0.01°C per decade from 1981 to 2014 for the CMIP6 ensemble mean; Fig. 1B), and the observed warming trend is therefore
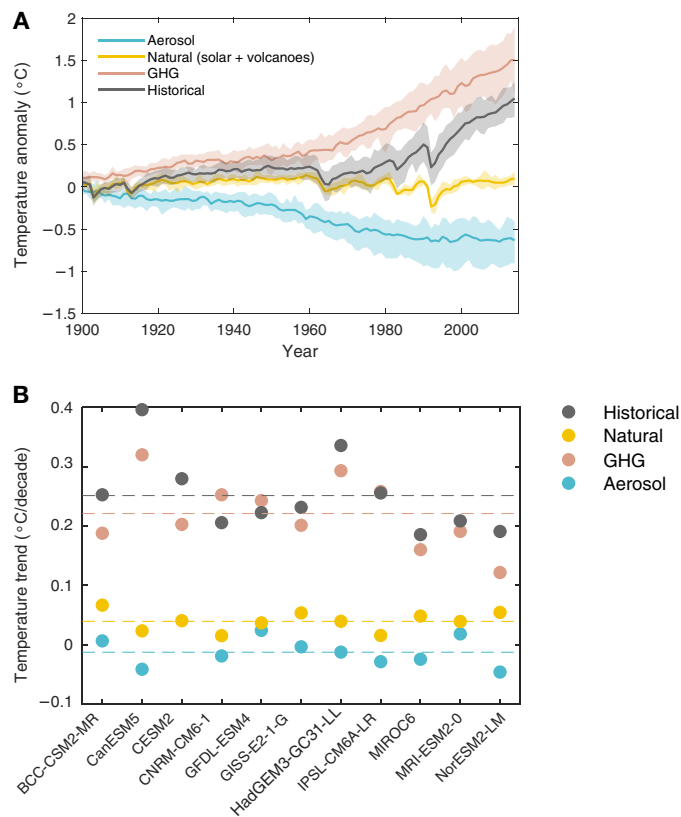

Fig. 1. Global mean temperature anomaly and its decadal trend in CMIP6 models in response to different radiative forcings. (**A**) Simulated global mean surface air temperature (GSAT) anomaly relative to 1850–1900 in CMIP6 models forced with different forcings during the historical period: anthropogenic aerosols (blue), natural forcing (solar irradiance and stratospheric aerosol; yellow), well-mixed greenhouse gases (GHG; red), and all natural and anthropogenic forcings (historical; gray). The shaded area indicates the likely range (17 to 83% percentile). Note that the ensemble sizes differ for the experiments, and in particular, the historical experiment is available for a larger set of CMIP6 models. (**B**) Trend in GSAT from 1981 to 2014 (as not all models have simulations available until the year 2017), using the same set of simulations with different sets of forcings as in (A). The dashed horizontal lines indicate multimodel mean decadal trends for each simulation type. Note that for CESM2, the aerosol-only simulation was not available.

expected to be strongly associated with the greenhouse gas warming dominated by CO_2 and thus the TCR and climate sensitivity (*18*). Internal variability can also affect the warming rate. Pacific variability has temporarily slowed short-term warming during the "global warming hiatus" from the late 1990s up to around year 2012 (*23*, *24*). Estimates of the internal component of Atlantic multidecadal variability depend largely on how the forced signal is estimated and removed (*25*). It has been argued that decadal Atlantic sea surface temperature (SST) variability might largely reflect a forced signal in the period after 1980 [e.g., in (*26*)]. We estimate a Pacific internal variability contribution to GSAT of about −0.02°C per decade and an Atlantic contribution of about +0.01°C per decade from 1981 to 2017 (see Materials and Methods; and fig. S1). Compared to the observed global mean temperature increase of about 0.19°C per decade, the 1981–2017 warming is therefore unlikely to be strongly influenced by lower frequency variability, which is further estimated to partly cancel out between the contributions from the Atlantic and Pacific Oceans. The warming during this period may therefore act as an emergent constraint on future warming (by mid-century, 2041–2060, and end of the century, 2081–2100). Because not all CMIP6 models provide simulations up to 2017, we also use a second period, 1981–2014. Pacific variability dampened the rate of global warming to a somewhat greater extent during this period (fig. S1), and the observed warming might therefore slightly underestimate the forced trend. We use the latter period to constrain the more idealized TCR metric. Sensitivity analysis to different methodological choices is included in the Supplementary Materials (figs. S2 and S3 and tables S3 and S4).

We use the mean of two observational datasets [Cowtan and Way v2 (*27*) and GISTEMPv4 (*28*, *29*)], which are both spatially nearly complete during the examined periods for the emergent constraint. The quoted uncertainty range in the emergent constraint is estimated by randomly sampling from the observed distribution (including the uncertainties of the trend from internal variability, of structural data uncertainty, and of the blending effect that accounts for the model-observation differences in the temperature metric they simulate or report; see Materials and Methods) and the associated future warming estimated by linear regression and the prediction error of the fit. This approach of quantifying the uncertainty of the constraint warming is similar to that used in (*30*).

Constraints on the TCR

We find that the recent warming trend (1981–2017) is strongly correlated with TCR across CMIP6 models ($R = 0.82$) and a joint distribution of CMIP6 and CMIP5 models ($R = 0.71$; fig. S2). A similar correlation ($R = 0.74$) holds for the period 1981–2014, for which more CMIP6 models and ensemble members are available, as it only covers the "historical" scenarios as defined in CMIP6. Given the theoretical arguments discussed above, this strong correlation (Fig. 2, A to C) can serve as an emergent constraint on the TCR. High ECS models (defined here as ECS > 4.5°C; shown in dark red color) have difficulties reproducing the observed warming trend (Fig. 2, A to C) (*20*–*22*). The observationally constrained likely ranges of TCR estimates based on CMIP6, CMIP5, or both combined (Fig. 2, A to C, blue rectangle, and D, blue boxes) are consistent but substantially narrower than those reported by AR5 of 1.0° to 2.5°C (*1*), regardless of the set of models used (Fig. 2D). The two likely ranges are, however, not fully comparable, as different lines of evidence were combined in AR5, leading to a broader uncertainty range. The observationally constrained TCR likely range (17 to 83%), based on CMIP6 models

alone, of 1.20° to 1.99°C with a median of 1.60°C is narrower and lower than the raw CMIP6 likely range of 1.55° to 2.55°C with a median of 1.95°C (Fig. 2D, gray CMIP6 bar, and table S3). Our results from CMIP6 observationally constrained TCR (of 1.60°C) are consistent with a recent median TCR estimate of 1.67°C derived from CMIP5 models (*18*).

Pacific variability has a larger cooling effect over 1981–2014 than from 1981 to 2017 (fig. S1), and accordingly, the observed global mean temperature increase is weaker during the first period. The forced trend, in turn, is expected to be very similar in both periods based on the CMIP6 multimodel mean. Using the 1981–2017 period therefore results in a slightly higher observationally constrained median TCR of 1.71°C compared to the constraint based on the 1981–2014 warming (1.60°C) and in a narrower observationally constrained TCR likely range of 1.38° to 2.04°C, though based on a smaller set of CMIP6 models that had SSP scenario simulations available (fig. S2). Depending on the period used, the median TCR based on raw CMIP6 models is approximately 16 to 22% higher than the observationally constrained median TCR based on CMIP6 models (Fig. 2D and table S3). We also obtain consistent observationally constrained TCR estimates when we instead use the CMIP5 ensemble, or a joint distribution of CMIP5 and CMIP6 models, or if we use an alternative observational dataset (Fig. 2D and fig. S3B) (*31*).

Emergent constraints on ECS based on past warming are less straightforward because TCR also depends on ocean heat uptake, thus making the relationship between TCR and ECS nonlinear (*18*, *32*). In addition, there are large uncertainties associated with how the feedbacks change in the future, mostly as a result of changing warming patterns (*33*). Therefore, the correlations of recent warming with ECS (Fig. 3) are weaker than with TCR. We do not provide a formally constrained range for ECS here, but note that 7 of the 10 CMIP6 models with ECS larger than 4.5°C simulate recent warming that is inconsistent with observed warming trends (outside the ±2σ range, light gray rectangle in Fig. 3). While that does not strictly rule out high ECS values, it suggests that these high ECS values are unlikely.

Spatial pattern information supports TCR constraint

Next, we address the question of whether the spatial pattern of regional temperature trends for the 1981–2014 period, with global mean information removed, might further pinpoint model structural differences related to TCR. Such correlation of TCR with the magnitude of regional variations in the warming signal in each model would support the argument that the constraint on TCR (Fig. 2C) arises from the strength of climate feedbacks that also have an imprint on regional warming patterns, rather than from a spurious combination of other effects. To do so, we first subtract from each model's 1981–2014 spatial trend pattern its respective global mean trend (i.e., trend values in Fig. 2C). This yields for each CMIP5 and CMIP6 model a spatial trend pattern that only contains regional deviations from the model's global mean warming trend, but no information about the global mean warming trend itself. Second, we calculate from this set of individual model patterns a multimodel mean pattern, which results in a "fingerprint of spatial trend variations." This multimodel fingerprint highlights Arctic amplification and land-sea warming contrast as two characteristic features, where regional warming across models deviates from the global mean trend (Fig. 4A). Last, we project each model's spatial trend pattern onto the multimodel mean fingerprint. This last step measures the spatial congruence between the multimodel fingerprint and each
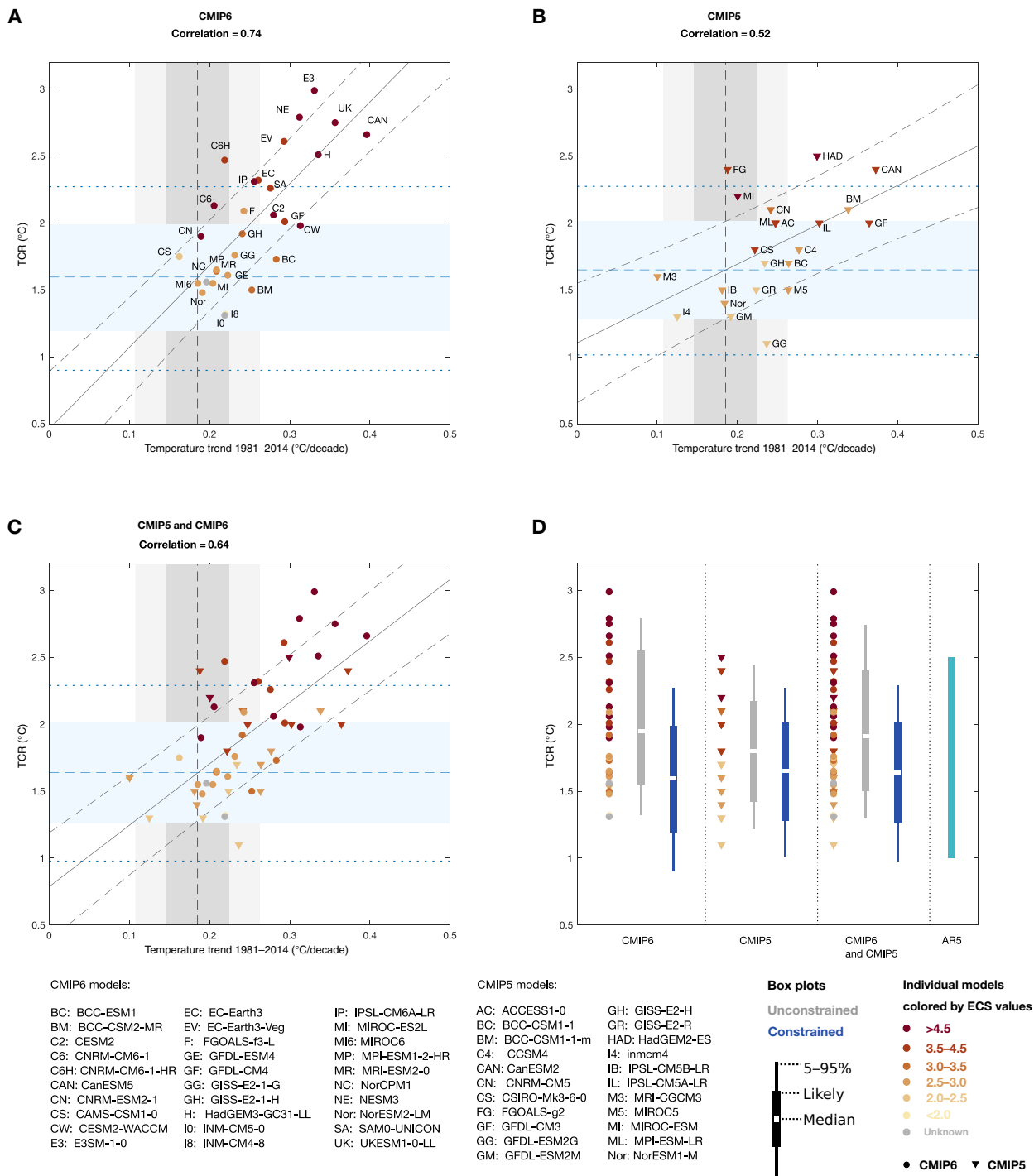
Fig. 2. Correlation of the simulated warming trend for the period 1981–2014 with TCR. (**A**) Correlation based on CMIP6 models, (**B**) based on CMIP5 models, and (**C**) based on the joint distribution of CMIP6 models (circles) and CMIP5 models (triangles). The emergent constraint is based on the mean of two observational datasets [Cowtan and Way (*27*) and GISTEMP (*28*, *29*)], adjusted for the blending effects (gray vertical line). If a model had more than one ensemble member, its ensemble mean is shown and was used in the regression. On (A) to (C), the dark gray rectangle shows the ±1σ uncertainty range in the observed trends for the period 1981–2014 (with the uncertainty range encompassing effects of internal variability, blending, and structural uncertainties), and the light gray rectangle shows the ±2σ range (see Materials and Methods). The blue rectangle indicates the likely range (>66%) of the emergent constraint on future warming (TCR). The median value is shown by dashed blue line, and dotted blue lines indicate the 5 to 95% uncertainty range (see Materials and Methods on how the uncertainty range on constrained TCR was derived). (**D**) Constrained and unconstrained ranges of TCR based on CMIP6 and CMIP5 models [following from (A) to (C)], compared with the IPCC AR5 likely range. Unconstrained ranges (gray box plots) are based on raw CMIP models, shown to the left of each box plot by individual dots. Constrained ranges (blue box plots) are based on the emergent constraint (as in top panels). The last box plot in (D) shows the IPCC AR5 likely (>66% probability; equivalent to 17 to 83% range) range. Each box plot shows 5 to 95% range, likely range, and median value, as illustrated in the legend.
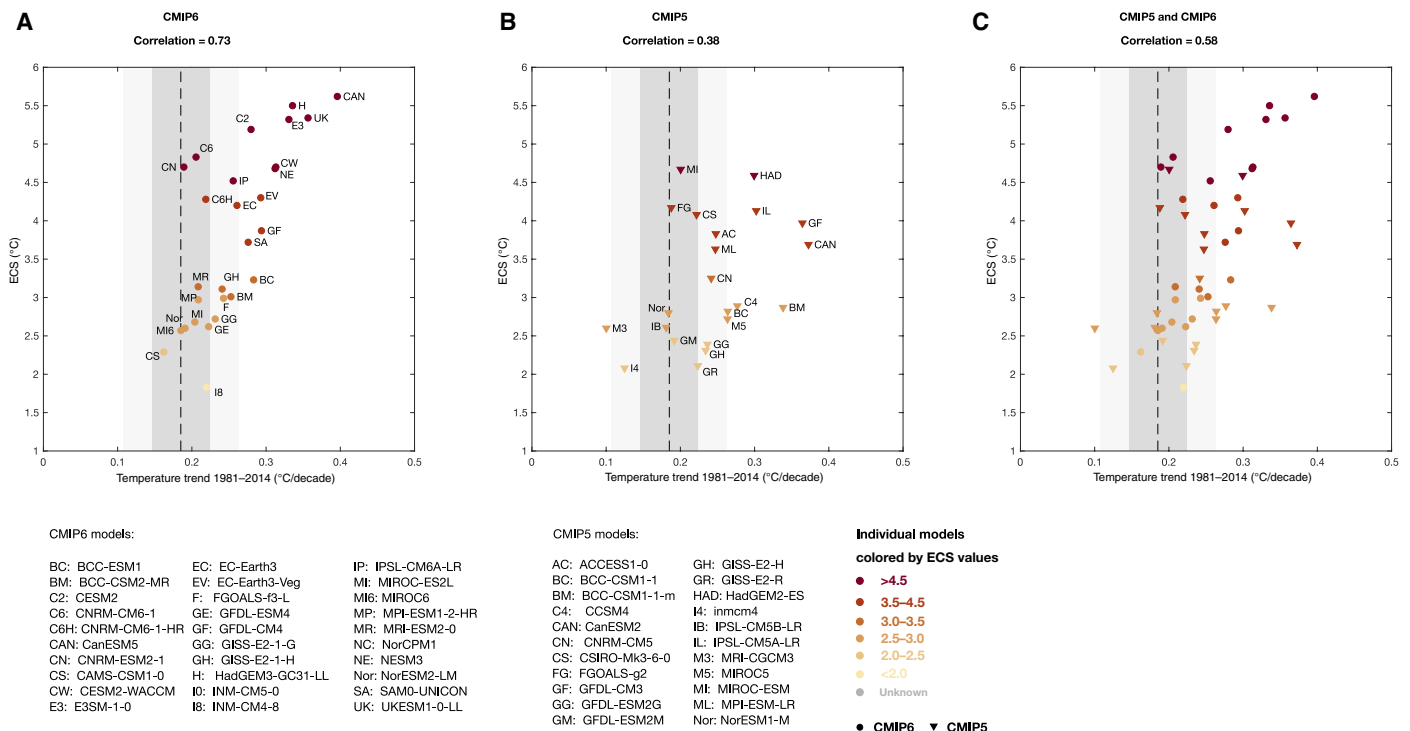
Fig. 3. Correlation of the simulated warming trend for the period 1981–2014 with ECS. (**A**) Based on CMIP6 models, (**B**) based on CMIP5 models, and (**C**) based on the joint distribution of CMIP6 models (circles) and CMIP5 models (triangles). Gray rectangles show the ±1σ and ±2σ ranges of uncertainty in the observed trend for the period 1981–2014, based on the mean of the Cowtan and Way (27) and GISTEMP (28, 29) datasets (as in Fig. 2).

model's 1981–2014 regional trend map based only on pattern covariance, similar to standard detection and attribution methods (34), and is independent of any global mean warming trends.

This approach of "fingerprint of spatial trend variation" yields a correlation of each model's pattern covariance with TCR ($R = 0.59$ in Fig. 4B, compared with $R = 0.64$ for the correlation of TCR with recent global mean warming in Fig. 2C; using a joint sample of CMIP6 and CMIP5 models; and if based on CMIP6 models alone, correlation coefficient of $R = 0.71$, compared with $R = 0.74$ in Fig. 2A). The correlation arises because models with the highest TCR produce a larger magnitude of the spatial trend pattern shown in Fig. 4A in the 1981–2014 period. This, in turn, suggests that the constraint on TCR shown in Fig. 2 results from spatial trend patterns that are consistent across models and that the models with the highest TCR differ not only in global mean warming but also in the magnitude of the simulated regional trend pattern from most other models. The spatial consistency might arise at least partly due to physical feedbacks that are common across the models, related, for instance, to temperature, surface albedo, and cloud feedbacks that contribute to Arctic amplification. The projection of observations (global mean trend removed) onto the fingerprint (vertical line in Fig. 4B) indicates consistently with the global mean analysis in Fig. 2C that models with very high TCR (table S1) seem less likely. This is because high-sensitivity models simulate a too large magnitude of 1981–2014 regional warming patterns that is not supported by the observations (Fig. 4B). Uncertainty due to internal variability is expected to be higher on regional scales; therefore, more models fall within the observational estimate (Fig. 4B). In contrast, on global mean scale, this uncertainty from internal variability is reduced (Fig. 2), and thus, global temperature gives a narrower constraint on response of forcing

to pattern. Overall, the pattern-based analysis provides important complementary and independent evidence that the emergent constraint based on the global mean warming trend derived in Fig. 2 does not emerge from the models because of spurious compensation of effects but instead has its origin in a consistent pattern-based signal across models. As a note of caution, however, we do not recommend to derive any emergent constraint to be based only on mean-removed trend patterns, because a crucial piece of information (the global mean warming trend) is disregarded.

Constraints on future warming in SSP scenarios

The robust correlation between the recent warming trend and TCR (Fig. 2) justifies further tests whether such an emergent constraint also arises between the simulated recent warming trend and future warming in different SSP scenarios. Directly applying the emergent constraint (based on recent warming trends) on future warming in SSP scenarios is not straightforward because the models account differently for changes in non-CO_2 forcings. Nevertheless, we find that the recent warming trend is strongly correlated with warming by the mid-century and end of the century (Fig. 5, with respect to 1850–1900 baseline, and figs. S4 and S5, with respect to the 1995–2014 baseline) particularly for a high-emission scenario (i.e., SSP5-8.5), which is dominated by greenhouse gas forcing ($R = 0.92$ and $R = 0.86$ for mid-century and end of century, respectively; Fig. 5, A and C). We also find that TCR is highly correlated with the warming in SSP5-8.5 and the ambitious mitigation scenario SSP1-2.6 across the CMIP6 models ($R > 0.8$ in either scenario; fig. S6; with respect to the 1995–2014 baseline). This justifies using the present-day observational trend estimates to constrain future projections.

The observationally constrained likely range of future warming (blue rectangles; Fig. 5, B and D) in response to an ambitious mitigation
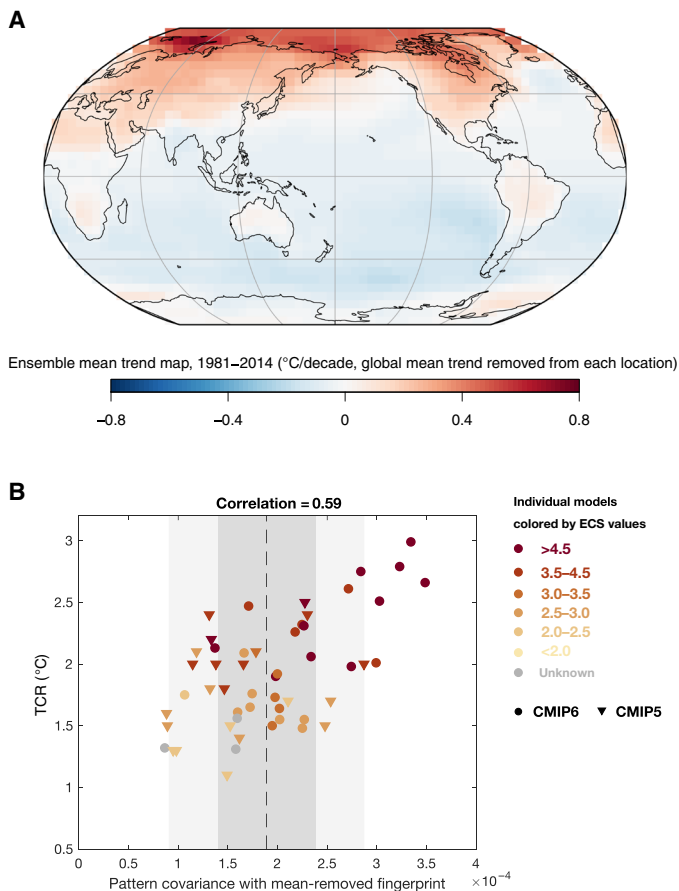
A

Ensemble mean trend map, 1981–2014 (°C/decade, global mean trend removed from each location)

B

Fig. 4. Pattern covariance between each model's trend map (global mean trend removed) and multimodel mean fingerprint. (**A**) Multimodel mean deviation of regional warming trends from the global mean warming trend (fingerprint of regional trend variation). (**B**) Correlation of the pattern covariance metric [that is, the covariance of each model's regional trend pattern (global mean removed) with the multimodel mean fingerprint shown in (A)] with each model's TCR. The dashed black line in (B) indicates an observational estimate, based on the mean of the observational datasets [Cowtan and Way (27) and GISTEMP (28, 29)], and the gray rectangles indicate estimate of uncertainty in the observations due to internal variability at 1σ and 2σ levels (based on the large ensembles simulations listed in table S2). Spatial pattern information reveals that high TCR models simulate a large magnitude of a regional warming pattern without global mean information.

SSP1-2.6 scenario is 1.36° to 1.86°C by mid-century and 1.33° to 1.99°C by the end of the century with respect to the 1850–1900 baseline. These results are generally in line with the Paris Agreement target of limiting warming to well below 2°C above preindustrial temperatures [yellow lines in Fig. 5, B and D, using Paris Agreement–consistent temperature metrics in Fig. 5; (35)]. Most of the models with climate sensitivity values outside the AR5 likely range exceed the 2°C warming in this scenario. However, their past trend also falls outside the observationally constrained range (Fig. 5, B and D) and might thus be considered less likely. Similarly, the emergent constraint indicates that the strong warming of high ECS models under the high-emission scenario is less likely (Fig. 5, A and C), thereby constraining the end-of-century warming relative to 1850–1900 to a lower median level of 4.15°C than the unconstrained CMIP6 median warming of 4.69°C. The observationally constrained future warming in other scenarios, SSP2-4.5 and SSP3-7.0, is also lower

than raw (unconstrained) CMIP6 warming in those scenarios (fig. S5 and table S4).

Comparing warming projections in CMIP6 and CMIP5 models

Comparing future warming projections in Representative Concentration Pathway (RCP) and SSP scenarios is not straightforward because of differences in the radiative forcings in both the historical and future periods in corresponding scenarios (7, 36). To allow an approximately like-for-like comparison between CMIP5 and CMIP6 future warming, we use a simple approach for estimating the CMIP5 responses for SSP scenarios, by scaling the future warming in the RCP 2.6 and RCP 8.5 scenarios (7, 36) by the ratio of the total anthropogenic forcing in corresponding scenarios (i.e., we scale RCP 2.6 warming by the SSP1-2.6 to RCP 2.6 anthropogenic forcing ratio, and we scale RCP 8.5 warming by the SSP5-8.5 to RCP 8.5 forcing ratio, calculated for the period of interest: mid-century or end of the century). This is based on the fact that the global temperature response is approximately proportional to the forcing during the transient phase (37). Such scaling results in an approximate range of the SSP scenarios had they been simulated by the CMIP5 models. However, because the ratio of total anthropogenic forcing for SSP and RCP scenarios is close to one, the resulting ranges are very similar to the original CMIP5 warming (10) and are also close to observationally constrained CMIP6 future warming in each corresponding scenario (SSP5-8.5 and SSP1-2.6, respectively; table S4), shown in Fig. 6. The results suggest that most of the difference between the median CMIP5 RCP 8.5 in IPCC AR5 and the median CMIP6 SSP5-8.5 is due to the CMIP6 models simulating stronger warming for a given forcing or scenario, (10), however, it is less certain how much influence the effective radiative forcing in CMIP6 compared to CMIP5 has changed.

DISCUSSION

Our results show that most models with high climate sensitivity (outside the AR5 likely range) or high transient response overestimate recent warming trends, with differences that cannot be explained by internal variability. This probably leads to future warming projections being biased high. Thus, the raw ensemble median and spread of future warming in CMIP6 (and therefore most other variables that scale to first order with global mean temperature) are not representative of a distribution constrained by observed trends, even if some of those models show a more realistic representation of processes in individual components than their CMIP5 predecessors (20–22). Conversely, CMIP6 models with climate sensitivity values that are within the IPCC AR5 likely range show warming trends much more consistent with the observations.

We demonstrate that the observed recent warming trends from 1981–2014 and 1981–2017 (see the Supplementary Materials for sensitivity analysis) are highly correlated with TCR across CMIP6 as well as CMIP5. Given the theoretical background (18) and robust correlations across two generations of ESMs, we provide an estimate of the observationally constrained likely range for TCR based on CMIP6 models of 1.20° to 1.99°C (17 to 83% range). The constrained CMIP6 median TCR (1.60°C) is substantially lower than the raw CMIP6 median (1.95°C) and is consistent with other recently published TCR estimates (18, 38). We also show that the observational constraint on TCR remains robust, with high-TCR CMIP6 models being consistently different from the remainder of CMIP5 and CMIP6 models, even
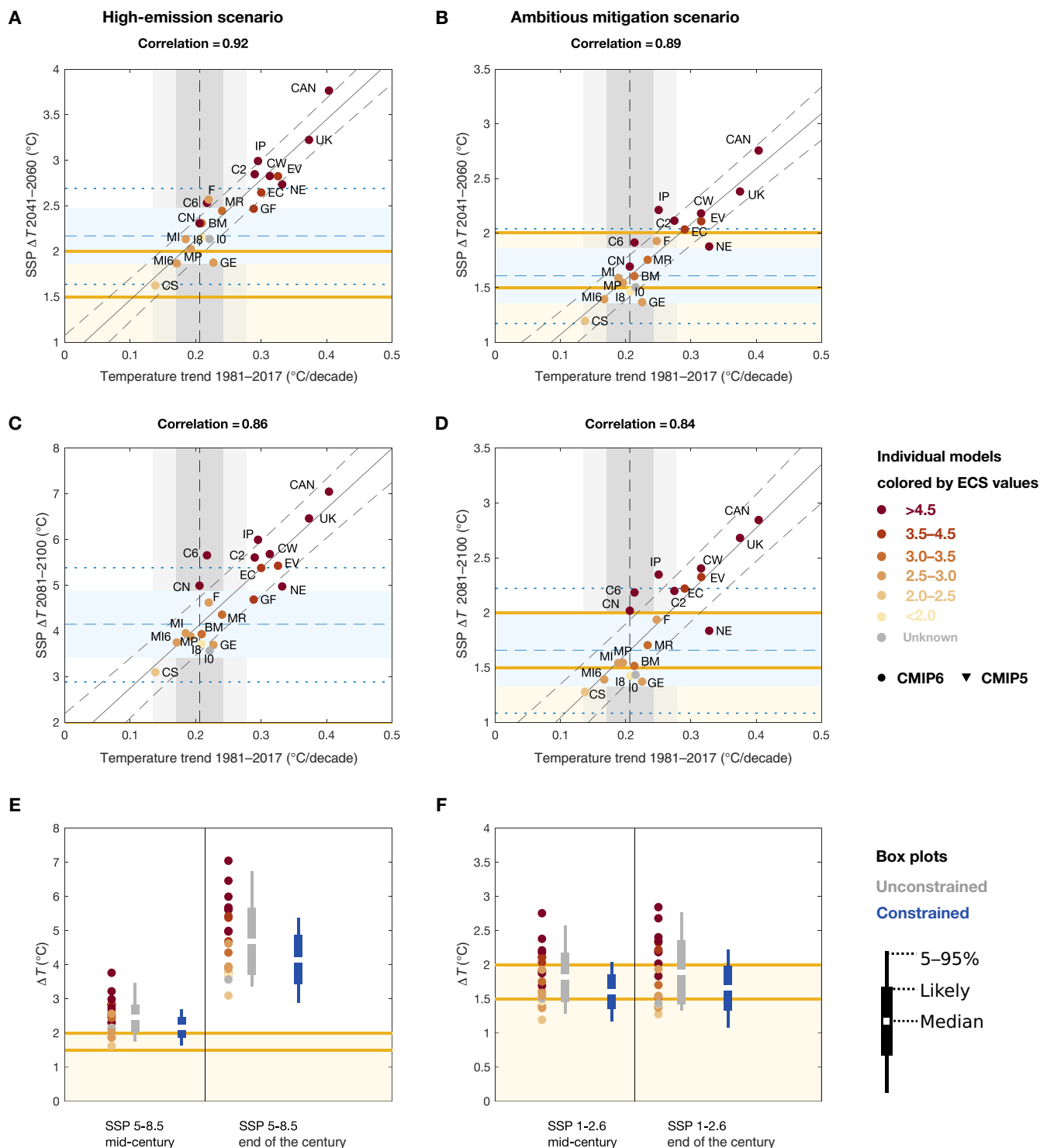
Fig. 5. Future warming constrained by the observed warming trend in comparison with the Paris Agreement target. (**A**) Future constrained warming by mid-century (years 2041–2060) in the high-emission SSP5-8.5 scenario. (**B**) As (A) but in the ambitious mitigation SSP1-2.6 scenario. (**C**) Constrained warming by the end of the century (2081–2100) in SSP5-8.5 scenario. (**D**) As (C) but in SSP1-2.6 scenario. (**E** and **F**) Resulting constrained and unconstrained (raw) ranges, as labeled. Future warming is with respect to the 1850–1900 baseline in all panels. Gray rectangles show observed warming trends for the period 1981–2017, using the mean of the observational datasets [Cowtan and Way (*27*) and GISTEMP (*28*, *29*)], with ±1σ and ±2σ uncertainty ranges. Blue rectangle indicates the likely range (>66%) of the emergent constraints on future warming. The median value is shown by dashed blue lines, and dotted blue lines indicate 5 to 95% uncertainty range. Yellow lines indicate the Paris Agreement thresholds of 1.5° and 2.0°C, and the yellow shaded area indicates warming interval consistent with achieving the Paris Agreement. Note: Future GSAT warming was adjusted for each model to make simulated warming consistent with the definition of a Paris Agreement temperature metric (*35*). For full model names, see Fig. 2.

Fig. 6. Future warming in CMIP5 and CMIP6 models (with respect to 1995–2014 baseline), constrained by the observed warming trend (1981–2017). (**A**) Constrained warming in SSP5-8.5 scenario (based on CMIP6 ensemble), in RCP 8.5 scenario, and estimated CMIP5 response to SSP5-8.5 scenario (i.e., CMIP5 scaled by the total forcing ratio, for a like-for-like comparison of responses to SSP and RCP scenarios). (**B**) In SSP1-2.6 scenario. Colored dots on each panel show the full CMIP6 simulated range by mid-century (years 2041–2060) and by the end of the century (years 2081–2100), with respect to the 1995–2014 baseline. The panels have different vertical axis limits. Note: The baseline for the future warming (ΔT with respect to 1995–2014) is different than in Fig. 5 (1850–1900). See fig. S4 for scatter plots and correlations and fig. S5 for constrained warming of the SSP2-4.5 and SSP3-7.0 scenarios. Constrained warming is based on the mean of the observational datasets [Cowtan and Way (*27*) and GISTEMP (*28*, *29*)] as in fig. S4.

if only the spatial warming pattern is considered (with the global mean temperature trend removed). We emphasize that our goal is to provide a defensible constraint on future warming (i.e., TCR or future warming in SSP scenarios), acknowledging that additional predictors might yield an even more robust constraint [e.g., using ocean heat content (*16*, *39*)]. Therefore, the past warming trend is only one of many possible ways of constraining future warming in climate models.

The emergent constraints derived here may underrepresent uncertainty from the statistical assumption of interpreting the observed trend as a random sample from the same distribution as the simulated trends (*40*). Some processes that are not represented in CMIP6 models, but are present in reality, and potential systematic biases in the models, could therefore contribute to a wider uncertainty range (*40*). On the other hand, the estimated uncertainty may be too large if the relationship is weakened by models that are unrealistic in aspects unrelated to the constraint. The fact that the relationships between the past and future global mean warming (and TCR) hold over two generations of models and are supported by theoretical arguments provides evidence that the emergent constraints derived here are robust.

Correlations are similarly high between the recent warming and future warming in the SSP scenarios, thus suggesting that future warming in the SSP scenarios simulated by models with high climate sensitivity is also likely to be biased high. Observationally constrained future warming in the SSP5-8.5 scenario, with respect to the 1995–2014 baseline, by the mid-century (years 2041–2060) is estimated at 1.01° to 1.90°C (5 to 95% range), and by the end of the century (years 2981–2100) is estimated at 2.26° to 4.60°C (5 to 95% range). The constrained median warming is 16% lower by mid-century and 14% lower by the end of the century than the unconstrained warming simulated by the CMIP6 ensemble (table S4). For comparison, the observationally constrained warming of the CMIP5 ensemble is essentially unchanged from its unconstrained warming, which justifies the use of the CMIP5 raw mean in AR5.

Despite the expectation that the constraint should be weaker in emission scenarios where non-CO_2 forcings such as aerosol reduction have a substantial contribution to the future temperature evolution, the SSP1-2.6 warming is also highly correlated with warming during the past decades. Constrained warming in SSP1-2.6, with respect to the 1850–1900 baseline consistent with the Paris Agreement (*35*), by mid-century (years 2041–2060) is estimated at 1.36° to 1.86°C (likely range), and by the end of the century (years 2081–2100) is estimated at 1.33° to 1.99°C (likely range). Our results thus suggest that this ambitious mitigation scenario is consistent with meeting the Paris Agreement target based on the observationally constrained CMIP6 models, while the Paris Agreement target would be exceeded by several high ECS models.

Last, we show that the CMIP6 projections are consistent with the CMIP5 projections after observationally constraining the CMIP6 ensemble and accounting for scenario differences, in this case through a simple rescaling CMIP5 warming by the ratio of the anthropogenic radiative forcings in the respective SSP and RCP scenarios. The difference of about 0.83°C between the raw CMIP5 RCP 8.5 and raw CMIP6 SSP5-8.5 warming by the end of the century (with respect to the 1995–2014 baseline) is primarily due to the higher TCR values in CMIP6. Given the constraint from past warming, the CMIP6 raw model ensemble is therefore likely biased high and is not representative of the constrained distribution, while the observationally constrained

CMIP6 ensemble is generally consistent with the raw and constrained CMIP5 estimates.

The high ECS models that are outside of the observationally constrained range may still provide very useful information regarding earth system behavior at high levels of warming, such as exploring climate and carbon cycle feedbacks for large deviations from present-day climate, for estimating pattern scaling of extreme events (per degree of warming), or a basis of storylines relevant for high impacts (*13*). It also remains important to improve our understanding of the regional responses to global warming across the full range of models. However, the clustering of models at the high end of global mean warming in the ensemble of opportunity needs to be accounted for (e.g., through model weighting or rescaling the ensemble) to avoid projections that are biased high.

MATERIALS AND METHODS

We make use of available CMIP6 ESMs (*6*) (table S1) driven by historical forcings for the period 1850–2014 and extended by different SSP scenarios (SSP1-2.6 and SSP5-8.5 in the main text; and SSP2-4.5 and SSP3-7.0 in the Supplementary Materials) until the year 2100. We use the 1981–2014 period in Figs. 2 to 4 (for which more model simulations were available; table S1) and the 1981–2017 period for Figs. 5 and 6, which are based on fewer models that had SSP simulations available. These periods are chosen such that there is little trend in aerosol cooling (Fig. 1) and that they are only weakly influenced by known modes of internal variability (see below).

For the simulated warming from 1981–2017, we extend the CMIP6 historical simulations by the SSP5-8.5 scenario and the CMIP5 (*5*) historical simulations by the RCP 8.5 scenario. The warming trend until the year 2017 should, however, be very similar across the scenarios (*41*). The CMIP5 scenarios also deviate slightly from observed changes (e.g., in stratospheric aerosol or solar variability) (*42*). As the CMIP6 models were forced with updated external drivers up to 2014, this is less of a concern for the CMIP6 ensemble. Both the CMIP5 and CMIP6 ensembles, however, lead to consistent constrained TCR estimates (table S3), suggesting that the results are not strongly influenced by the differences in radiative forcing. For the models' output, we take ensemble means from models that provide multiple ensemble members, which reduce noise due to internal variability in the models.

The observed warming trends are calculated as the mean of two spatially interpolated datasets: Cowtan and Way (*27*) v2 updated with HadSST4 (*43*) and GISTEMP (v4) (*28*, *29*). We also examined the Berkeley Earth Surface Temperature (BEST) (*31*) dataset, but it shows nearly identical warming as the Cowtan and Way dataset over the two periods considered (fig. S3B). We did not include the BEST dataset into the observational mean as it is structurally similar to the Cowtan and Way dataset, and both use SST datasets based on HadSST. On the contrary, GISTEMP uses a more independent SST dataset. We quantify structural data uncertainty of the observed trend by the standard deviation (SD) across the 100 members of the Cowtan and Way v2 (with HadSST3) dataset.

Some of the model-observation mismatches can be explained by the differences in global mean temperature definitions (*44*). The models' output is the global mean near-surface air temperature (GSAT), while observation-based datasets report a blend of land near-surface air and sea surface temperatures [here referred to as global blended surface temperature (GBST)], which on average have been warming slightly slower than GSAT only (*44*). However, for future climate

projections and impact assessments, the GSAT temperature metric is more relevant (*35*). To quantify the blending bias, we use data of (*44*) and compare simulated GSAT with simulated GBST (constructed from temperature anomalies). The difference from 1981 to 2017 (or 1981 to 2014) is an estimate of the blending bias in a model simulation during this period. To allow a like-for-like comparison among models and observations, we add an estimate of the blending effect (difference between GSAT and GBST) to the GBST observations to make them GSAT-like. We regress the simulated GBST increase over the examined period against the blending effect over the same period across the CMIP5 ensemble. Models that simulate greater warming also tend to show a larger blending effect. Using this relationship, we estimate the blending effect for GBST observations and use the prediction error of the linear fit as an estimate of uncertainty. For an observed warming trend of about 0.19°C per decade (for the period 1981–2017), the blending effect is estimated at 0.014° ± 0.005°C per decade (1σ). For 1981–2014, the observed warming is slightly lower and accordingly also the estimated blending effect (0.013° ± 0.005°C per decade). Both observational datasets considered [Cowtan and Way (*27*) and GISTEMP (*28, 29*)] are interpolated to near-full coverage, and we therefore compare them with the simulated temperature field averaged over the whole Earth.

To quantify the role of unforced internal variability to a potential difference between observed and simulated trends, we make two independent estimates: one based on climate model simulations and one based on observed GBST. For the first estimate, we use a mean estimate of the SD across the warming trends for the period 1981–2014 in 12 large initial condition ensembles of CMIP5 and CMIP6 ESMs, resulting in a noise estimate of 0.035°C per decade due to internal variability (ranging from 0.023° to 0.049°C per decade between the models; table S2). Under the assumption that internal variability and the forced signal are independent, which is likely the case for relatively weak radiative forcing but may break down under larger climate change (*45*), we estimate internal variability from 32 CMIP6 control simulations (from each simulation separately). The mean SD of 34-year-long trends is with 0.037°C per decade similar to the smaller set of large ensembles. For the second estimate, we subtract both the raw and the scaled CMIP5 and CMIP6 GBST ensemble means from the observations from 1900 to 2018 [the multimodel means are scaled towards the observations (*46*)]. These residuals from different combinations of the simulated and observed GBST are an estimate of internal variability (*46*), but due to observational and forcing uncertainties (*26*), we interpret them as an upper estimate. Based on this, we estimate an SD of 0.038°C per decade for 34-year-long trends, slightly higher, but consistent with the model simulations in agreement with the findings in (*46*). As a conservative choice, we use this last estimate throughout the paper for the 1981–2014 period. For the 1981–2017 period, the internal variability estimates are slightly lower (cf. table S2), and we again use a conservative estimate based on the difference between observed and simulated GBST of 0.035°C per decade for the analyses in that period. The overall observational uncertainty is calculated as the sum in quadrature of the above three effects: structural uncertainty, internal variability, and uncertainty of the blending effect. Uncertainty from internal variability dominates the trend uncertainty.

The presence of internal variability in the observed GBST may bias the central value of the constrained climate response. We estimate the contribution of Pacific and Atlantic low-frequency variability to GBST using variability analogues (*35*). To quantify the influence of

Pacific variability (fig. S1), we search for simulated 40-month-long periods from the CMIP5 and CMIP6 control simulations that follow the observed (ERSSTv5 and COBE-SST2) SST evolution in the tropical Pacific (15°N to 15°S, 180° to 90°W). In addition, we search for analogues that follow the observed (ERA5, MERRA2, and JRA55) wind stress evolution over the western tropical Pacific (150°E to 150°W, 10°S to 10°N). The observational datasets are introduced and described in (*47–51*). For the contribution of Atlantic variability, we smooth (with 13-month-long running mean) the observed extratropical North Atlantic (30° to 60°N) SST before selecting 120-month-long analogues. Thereby, we remove some of the high-frequency variability and highlight the role of the Atlantic variability on a multidecadal time scale. Before selecting the best matching variability analogues (based on the root mean square deviation), we remove the CMIP5 and CMIP6 multimodel means from the observed tropical Pacific and extratropical North Atlantic SST to obtain estimates of the internal variability component in these regions. In addition, we estimate the forced signal in these two regions by scaling the CMIP6 multimodel mean GBST time series against the observations from 1900 to 2018 to reduce biases in the simulated warming and also remove these scaled multimodel means from the observations (*46*). The models do not simulate substantial trends in wind stress over the western tropical Pacific, and therefore, we directly use the observed wind stress variability. For the Pacific SST, we further estimate the forced signal with the method in (*52*). Different to the North Atlantic SST and tropical Pacific wind stress, we standardize the time series of equatorial Pacific SST before selecting analogues. Standardization favors models that under- or overestimate observed variability, but it has only a small influence on the results. We interpret the results from estimating the forced signal by scaling the GBST as a best estimate but show the range from the other approaches in fig. S1. Pacific variability has contributed a cooling over both examined periods, but less so over 1981–2017, consistent with other studies (fig. S1). As the Atlantic contribution is weak and similar in both periods, the observed 1981–2017 warming period is probably less influenced by internal variability. The central estimate of TCR constrained by 1981–2014 warming might therefore be slightly underestimated (see the "Constraints on the TCR" section).

We use ordinary least squares (OLS) regression for the relationship between the recent simulated warming rate, which consists of a forced signal and noise (depending on the ensemble sizes, the noise is smaller or larger for individual models) and future warming or TCR. The presence of noise in the predictor biases the OLS regression slope toward zero, i.e., we underestimate the relationship between forced signal and future warming. Errors-in-variables regression models, such as total least squares (TLS), allow to account for that. Forcing the regression line to intercept with the origin (0,0), as in fig. S3A, is based on an assumption of strict linearity between the simulated forced trend and the future warming (or TCR). This results in a similar TCR estimate of 1.45°C as obtained by TLS, but is slightly lower than the OLS estimate without fixing the intercept of 1.60°C (using 1981–2014; Fig. 2; constrained TCR; table S3). The assumption of strict linearity, however, is not satisfied in ESMs, due to imperfect representation of different feedbacks and simulated response to forcing, and due to the presence of internal variability. Because the observed trend is also influenced by internal variability as discussed above, we argue that we are rather interested in estimating the relationship between simulated warming and future warming than the relationship between forced warming and future warming,

and OLS results in an unbiased estimate of the former. It is therefore generally accepted for predictive modeling (*53*). A concern, however, is that depending on the ensemble size, amounts of variability present in the models and in the observations greatly differ. An alternative approach would be to take one ensemble member per model, which, however, neglects a lot of the available data, or to use all ensemble members and weight them such that each model receives the same weight. This approach results in a slightly higher TCR estimate of 1.71°C (fig. S3D) than the OLS regression on ensemble means. Given the simplicity of OLS and that the results do not depend strongly on the regression approach, we use OLS for the main analysis.

The dotted lines around the linear regressions in Figs. 2 and 5 show the prediction error for the fit. The gray rectangles in Figs. 2 to 5 represent the observed GSAT trend (i.e., GBST with our estimate of the blending effect) and its combined uncertainty from internal variability, structural uncertainties in the observational datasets, and the blending effect as introduced above (shown are the ±1σ and ±2σ ranges; for Fig. 4, only the effect of internal variability is included). The blue rectangles on Figs. 2 and 5 represent the uncertainty (likely range; 17 to 83%) in the observationally constrained future warming, and the blue dashed lines show the 5 to 95% ranges. We obtain this uncertainty by randomly sampling from the distribution of observed warming (gray square) and its associated future warming given by the linear regression and its prediction error.

The ECS of each CMIP6 model is here estimated by regressing the top-of-atmosphere radiative imbalance against the GSAT change during the first 150 years in a CO_2-only simulation that quadruples the amount of atmospheric CO_2 (*8*). This estimate is scaled by a factor of 2 (we neglect that CO_2 forcing rises slightly faster than logarithmic (*54*)]. The so-obtained ECS is an effective sensitivity and underestimates the actual equilibrium climate response for most models (*16*), but it is consistent with the ECS values reported for the CMIP5 ensemble (*8*). TCR is calculated from the CO_2-only simulation, where the atmospheric CO_2 concentration increases at a rate of 1% per year, centered on the time of doubling of the atmospheric CO_2, which occurs during simulation year 70 (we use the mean of the years 61 to 80). Note that in the GISS-E2-1-G simulations, the CO_2 concentration only increases until year 70. Therefore, TCR of this model is slightly underestimated. To estimate the forced change in each idealized CO_2-only simulation, we subtract a linear fit to the corresponding segment of the unforced control simulation. For INM-CM5-0, no control simulation was available at the time of writing, and we therefore estimate TCR with respect to the first 5 years of its +1% CO_2 per year experiment. Its control experiment became available after the revisions, and estimating warming with respect to the control climate indicates a slightly higher TCR of 1.39°C instead of 1.31°C, which does not change our conclusions. The ECS and TCR values of the CMIP6 ensemble are reported in table S1. The ECS and TCR values for CMIP5 models can be found in table 1 of (*8*).

CMIP6 models used in this paper are listed in table S1. (Note: Not all models had SSP data available. Also, simulations with CAMS-CSM1-0 run only to the year 2099, so instead of the change for the 2081–2100 period, the change for 2081–2099 was calculated in this model only.) We make use of the following CMIP5 models (historical scenario, followed by RCP 2.6 and RCP 8.5 scenario): ACCESS1-0, bcc-csm1-1, bcc-csm1-1-m, CCSM4, CNRM-CM5, CSIRO-Mk3-6-0, CanESM2, FGOALS-g2, GFDL-CM3, GFDL-ESM2G, GFDL-ESM2M, GISS-E2-H, GISS-E2-R, HadGEM2-ES, inmcm4, IPSL-CM5A-LR,

IPSL-CM5B-LR, MIROC-ESM, MIROC5, MPI-ESM-LR, MRI-CGCM3, and NorESM1-M. For CMIP5 models, we use all available ensemble members in the "p1"-only variant.

SUPPLEMENTARY MATERIALS

REFERENCES AND NOTES

1. 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*, T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, P. M. Midgley, Eds. (Cambridge Univ. Press, 2013), pp. 1535.
2. C. Tebaldi, R. Knutti, The use of the multi-model ensemble in probabilistic climate projections. *Philos. Trans. R. Soc. A* **365**, 2053–2075 (2007).
3. R. Knutti, The end of model democracy? *Clim. Change* **102**, 395–404 (2010).
4. R. Knutti, D. Masson, A. Gettelman, Climate model genealogy: Generation CMIP5 and how we got there. *Geophys. Res. Lett.* **40**, 1194–1199 (2013).
5. K. E. Taylor, R. J. Stouffer, G. A. Meehl, An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.* **93**, 485–498 (2011).
6. V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, K. E. Taylor, Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).
7. M. Meinshausen, Z. Nicholls, J. Lewis, M. J. Gidden, E. Vogel, M. Freund, U. Beyerle, C. Gessner, A. Nauels, N. Bauer, J. G. Canadell, J. S. Daniel, A. John, P. Krummel, G. Luderer, N. Meinshausen, S. A. Montzka, S. Rayner, S. Reimann, S. J. Smith, M. van den Berg, G. J. M. Velders, M. Vollmer, H. J. Wang, The SSP greenhouse gas concentrations and their extensions to 2500. *Geosci. Model Dev. Discuss* **2019**, 1–77 (2019).
8. P. M. Forster, T. Andrews, P. Good, J. M. Gregory, L. S. Jackson, M. Zelinka, Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *J. Geophys. Res. Atmos.* **118**, 1139–1150 (2013).
9. R. Knutti, M. A. A. Rugenstein, Feedbacks, climate sensitivity and the limits of linear models. *Philos. Trans. R. Soc. A* **373**, 20150146 (2015).
10. P. M. Forster, A. C. Maycock, C. M. McKenna, C. J. Smith, Latest climate models confirm need for urgent mitigation. *Nat. Clim. Chang.* **10**, 7–10 (2020).
11. C. J. Smith, P. M. Forster, M. Allen, N. Leach, R. J. Millar, G. A. Passerello, L. A. Regayre, FAIR v1.3: A simple emissions-based impulse response and carbon cycle model. *Geosci. Model Dev.* **11**, 2273–2297 (2018).
12. UNFCC, *FCCC/CP/2015/L.9/Rev.1: Adoption of the Paris Agreement* (UNFCCC, 2015), pp. 1–32.
13. R. T. Sutton, ESD Ideas: A simple proposal to improve the contribution of IPCC WGI to the assessment and communication of climate change risks. *Earth Syst. Dynam.* **9**, 1155–1158 (2018).
14. C. Hope, The $10 trillion value of better information about the transient climate response. *Philos. Trans. A Math. Phys. Eng. Sci.* **373**, 20140429 (2015).
15. P. M. Caldwell, M. D. Zelinka, S. A. Klein, Evaluating emergent constraints on equilibrium climate sensitivity. *J. Clim.* **31**, 3921–3942 (2018).
16. R. Knutti, M. A. A. Rugenstein, G. C. Hegerl, Beyond equilibrium climate sensitivity. *Nat. Geosci.* **10**, 727–736 (2017).
17. P. A. Stott, J. A. Kettleborough, Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature* **416**, 723–726 (2002).
18. D. Jiménez-de-la-Cuesta, T. Mauritsen, Emergent constraints on Earth's transient and equilibrium response to doubled CO_2 from post-1970s global warming. *Nat. Geosci.* **12**, 902–905 (2019).

19. J. T. Kiehl, Twentieth century climate model response and climate sensitivity. *Geophys. Res. Lett.* **34**, L22710 (2007).

20. J.-C. Golaz, P. M. Caldwell, L. P. V. Roekel, M. R. Petersen, Q. Tang, J. D. Wolfe, G. Abeshu, V. Anantharaj, X. S. Asay-, Davis, D. C. Bader, S. A. Baldwin, G. Bisht, P. A. Bogenschutz, M. Branstetter, M. A. Brunke, S. R. Brus, S. M. Burrows, P. J. Cameron-Smith, A. S. Donahue, M. Deakin, R. C. Easter, K. J. Evans, Y. Feng, M. Flanner, J. G. Foucar, J. G. Fyke, B. M. Griffin, C. Hannay, B. E. Harrop, M. J. Hoffman, E. C. Hunke, R. L. Jacob, D. W. Jacobsen, N. Jeffery, P. W. Jones, N. D. Keen, S. A. Klein, V. E. Larson, L. R. Leung, H.-Y. Li, W. Lin, W. H. Lipscomb, P.-L. Ma, S. Mahajan, M. E. Maltrud, A. Mametjanov, J. L. McClean, R. B. McCoy, R. B. Neale, S. F. Price, Y. Qian, P. J. Rasch, J. E. J. Reeves Eyre, W. J. Riley, T. D. Ringler, A. F. Roberts, E. L. Roesler, A. G. Salinger, Z. Shaheen, B. Shi, B. Singh, J. Tang, M. A. Taylor, P. E. Thornton, A. K. Turner, M. Veneziani, H. Wan, H. Wang, S. Wang, D. N. Williams, P. J. Wolfram, P. H. Worley, S. Xie, Y. Yang, J.-H. Yoon, M. D. Zelinka, C. S. Zender, X. Zeng, C. Zhang, K. Zhang, Y. Zhang, X. Zheng, T. Zhou, Q. Zhu, The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *J. Adv. Model. Earth Syst.* **11**, 2089–2129 (2019).

21. A. A. Sellar, C. G. Jones, J. P. Mulcahy, Y. Tang, A. Yool, A. Wiltshire, F. M. O'Connor, M. Stringer, R. Hill, J. Palmieri, S. Woodward, L. de Mora, T. Kuhlbrodt, S. T. Rumbold, D. I. Kelley, R. Ellis, C. E. Johnson, J. Walton, N. L. Abraham, M. B. Andrews, T. Andrews, A. T. Archibald, S. Berthou, E. Burke, E. Blockley, K. Carslaw, M. Dalvi, J. Edwards, G. A. Folberth, N. Gedney, P. T. Griffiths, A. B. Harper, M. A. Hendry, A. J. Hewitt, B. Johnson, A. Jones, C. D. Jones, J. Keeble, S. Liddicoat, O. Morgenstern, R. J. Parker, V. Predoi, E. Robertson, A. Siahaan, R. S. Smith, R. Swaminathan, M. T. Woodhouse, G. Zeng, M. Zerroukat, UKESM1: Description and evaluation of the U.K. Earth system model. *J. Adv. Model. Earth Syst.* **11**, 4513–4558 (2019).

22. I. M. Held, H. Guo, A. Adcroft, J. P. Dunne, L. W. Horowitz, J. Krasting, E. Shevliakova, M. Winton, M. Zhao, M. Bushuk, A. T. Wittenberg, B. Wyman, B. Xiang, R. Zhang, W. Anderson, V. Balaji, L. Donner, J. Dunne, J. Durachta, P. P. G. Gauthier, P. Ginoux, J.-C. Golaz, S. M. Griffies, R. Hallberg, L. Harris, M. Harrison, W. Hurlin, J. John, P. Lin, S.-J. Lin, S. Malyshev, R. Menzel, P. C. D. Milly, Y. Ming, V. Naik, D. Paynter, F. Paulot, V. Rammaswamy, B. Reichl, T. Robinson, A. Rosati, C. Seman, L. G. Silvers, S. Underwood, N. Zadeh, Structure and performance of GFDL's CM4.0 climate model. *J. Adv. Model. Earth Syst.* **11**, 3691–3727 (2019).

23. Y. Kosaka, S.-P. Xie, Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature* **501**, 403–407 (2013).

24. I. Medhaug, M. B. Stolpe, E. M. Fischer, R. Knutti, Reconciling controversies about the 'global warming hiatus'. *Nature* **545**, 41–47 (2017).

25. M. E. Mann, B. A. Steinman, S. K. Miller, On forced temperature changes, internal variability, and the AMO. *Geophys. Res. Lett.* **41**, 3211–3219 (2014).

26. K. Haustein, F. E. L. Otto, V. Venema, P. Jacobs, K. Cowtan, Z. Hausfather, R. G. Way, B. White, A. Subramanian, A. P. Schurer, A limited role for unforced internal variability in twentieth-century warming. *J. Clim.* **32**, 4893–4917 (2019).

27. K. Cowtan, R. G. Way, Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Q. J. R. Meteorol. Soc.* **140**, 1935–1944 (2014).

28. GISTEMP Team, GISS Surface Temperature Analysis (GISTEMP), version 4. NASA Goddard Institute for Space Studies (2019); https://data.giss.nasa.gov/gistemp/.

29. N. J. L. Lenssen, G. A. Schmidt, J. E. Hansen, M. J. Menne, A. Persin, R. Ruedy, D. Zyss, Improvements in the GISTEMP uncertainty model. *J. Geophys. Res. Atmos.* **124**, 6307–6326 (2019).

30. P. M. Cox, C. Huntingford, M. S. Williamson, Emergent constraint on equilibrium climate sensitivity from global temperature variability. *Nature* **553**, 319–322 (2018).

31. R. Rohde, R. A. Muller, R. Jacobsen, E. Muller, S. Perlmutter, A. Rosenfeld, J. Wurtele, D. Groom, C. Wickham, A new estimate of the average earth surface land temperature spanning 1753 to 2011. *Geoinfor. Geostat.: An Overview* **1**, 1 (2013).

32. R. Knutti, F. Joos, S. A. Müller, G.-K. Plattner, T. F. Stocker, Probabilistic climate change projections for CO₂ stabilization profiles. *Geophys. Res. Lett.* **32**, L20707 (2005).

33. K. C. Armour, Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks. *Nat. Clim. Chang.* **7**, 331–335 (2017).

34. B. D. Santer, S. Po-Chedley, M. D. Zelinka, I. Cvijanovic, C. Bonfils, P. J. Durack, Q. Fu, J. Kiehl, C. Mears, J. Painter, G. Pallotta, S. Solomon, F. J. Wentz, C.-Z. Zou, Human influence on the seasonal cycle of tropospheric temperature. *Science* **361**, eaas8806 (2018).

35. K. B. Tokarska, C.-F. Schleussner, J. Rogelj, M. B. Stolpe, H. D. Matthews, P. Pfleiderer, N. P. Gillett, Recommended temperature metrics for carbon budget estimates, model evaluation and climate policy. *Nat. Geosci.* **12**, 964–971 (2019).

36. M. Meinshausen, S. J. Smith, K. Calvin, J. S. Daniel, M. L. T. Kainuma, J.-F. Lamarque, K. Matsumoto, S. A. Montzka, S. C. B. Raper, K. Riahi, A. Thomson, G. J. M. Velders, D. P. P. van Vuuren, The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim. Change* **109**, 213–241 (2011).

37. R. Knutti, M. R. Allen, P. Friedlingstein, J. M. Gregory, G. C. Hegerl, G. A. Meehl, M. Meinshausen, J. M. Murphy, G.-K. Plattner, S. C. B. Raper, T. F. Stocker, P. A. Stott, H. Teng, T. M. L. Wigley, A review of uncertainties in global temperature projections over the twenty-first century. *J. Climate* **21**, 2651–2663 (2008).

38. A. Schurer, G. Hegerl, A. Ribes, D. Polson, C. Morice, S. Tett, Estimating the transient climate response from observed warming. *J. Climate* **31**, 8645–8663 (2018).

39. M. Steinacher, F. Joos, Transient Earth system responses to cumulative carbon dioxide emissions: Linearities, uncertainties, and probabilities in an observation-constrained model ensemble. *Biogeosciences* **13**, 1071–1103 (2016).

40. D. B. Williamson, P. G. Sansom, How are emergent constraints quantifying uncertainty and what do they leave behind? *Bull. Amer. Meteor. Soc.* **100**, 2571–2588 (2019).

41. M. H. England, J. B. Kajtar, N. Maher, Robust warming projections despite the recent hiatus. *Nat. Clim. Chang.* **5**, 394–396 (2015).

42. M. Huber, R. Knutti, Natural variability, radiative forcing and climate response in the recent hiatus reconciled. *Nat. Geosci.* **7**, 651–656 (2014).

43. J. J. Kennedy, N. A. Rayner, C. P. Atkinson, R. E. Killick, An ensemble data set of sea surface temperature change from 1850: The Met Office Hadley CentreHadSST.4.0.0.0 data set. *J. Geophys. Res. Atmos.* **124**, 7719–7763 (2019).

44. K. Cowtan, Z. Hausfather, E. Hawkins, P. Jacobs, M. E. Mann, S. K. Miller, B. A. Steinman, M. B. Stolpe, R. G. Way, Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures. *Geophys. Res. Lett.* **42**, 6526–6534 (2015).

45. P. T. Brown, Y. Ming, W. Li, S. A. Hill, Change in the magnitude and mechanisms of global temperature variability with warming. *Nat. Clim. Chang.* **7**, 743–748 (2017).

46. J. B. Kajtar, M. Collins, L. M. Frankcombe, M. H. England, T. J. Osborn, M. Juniper, Global mean surface temperature response to large-scale patterns of variability in observations and CMIP5. *Geophys. Res. Lett.* **46**, 2232–2241 (2019).

47. M. Barratt, ERA5 reanalysis is in production (2016); https://www.ecmwf.int/en/newsletter/147/news/era5-reanalysis-production.

48. S. Hirahara, M. Ishii, Y. Fukuda, Centennial-scale sea surface temperature analysis and its uncertainty. *J. Clim.* **27**, 57–75 (2013).

49. S. Kobayashi, Y. Ota, Y. Harada, A. Ebita, M. Moriya, H. Onoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, K. Miyaoka, K. Takahashi, The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteorol. Soc. Jpn.* **93**, 5–48 (2015).

50. R. Gelaro, W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, C. A. Randles, A. Darmenov, M. G. Bosilovich, R. Reichle, K. Wargan, L. Coy, R. Cullather, C. Draper, S. Akella, V. Buchard, A. Conaty, A. M. da Silva, W. Gu, G.-K. Kim, R. Koster, R. Lucchesi, D. Merkova, J. E. Nielsen, G. Partyka, S. Pawson, W. Putman, M. Rienecker, S. D. Schubert, M. Sienkiewicz, B. Zhao, The Modern-Era retrospective analysis for research and applications, version 2 (MERRA-2). *J. Clim.* **30**, 5419–5454 (2017).

51. B. Huang, P. W. Thorne, V. F. Banzon, T. Boyer, G. Chepurin, J. H. Lawrimore, M. J. Menne, T. M. Smith, R. S. Vose, H.-M. Zhang, Extended reconstructed sea surface temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *J. Clim.* **30**, 8179–8205 (2017).

52. T. Turkington, B. Timbal, R. Rahmat, The impact of global warming on sea surface temperature based El Niño–Southern Oscillation monitoring indices. *Int. J. Climatol.* **39**, 1092–1103 (2019).

53. R. J. Smith, Use and misuse of the reduced major axis for line-fitting. *Am. J. Phys. Anthropol.* **140**, 476–486 (2009).

54. M. Etminan, G. Myhre, E. J. Highwood, K. P. Shine, Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing. *Geophys. Res. Lett.* **43**, 12,614–12,623 (2016).

55. G. J. Boer, D. M. Smith, C. Cassou, F. Doblas-Reyes, G. Danabasoglu, B. Kirtman, Y. Kushnir, M. Kimoto, G. A. Meehl, R. Msadek, W. A. Mueller, K. E. Taylor, F. Zwiers, M. Rixen, Y. Ruprich-Robert, R. Eade, The decadal climate prediction project (DCPP) contribution to CMIP6. *Geosci. Model Dev.* **9**, 3751–3777 (2016).

56. G. Foster, S. Rahmstorf, Global temperature evolution 1979–2010. *Environ. Res. Lett.* **6**, 044022 (2011).

Citation: K. B. Tokarska, M. B. Stolpe, S. Sippel, E. M. Fischer, C. J. Smith, F. Lehner, R. Knutti, Past warming trend constrains future warming in CMIP6 models. *Sci. Adv.* **6**, eaaz9549 (2020).

Science Advances

Past warming trend constrains future warming in CMIP6 models

Katarzyna B. Tokarska, Martin B. Stolpe, Sebastian Sippel, Erich M. Fischer, Christopher J. Smith, Flavio Lehner and Reto Knutti