ARTICLE   **OPEN**

Constraint on regional land surface air temperature projections in CMIP6 multi-model ensemble

Jie Zhang [1,2,3 ✉], Tongwen Wu [1,2,3 ✉], Laurent Li [4], Kalli Furtado[5,6], Xiaoge Xin[1,2,3], Chengjun Xie[7,8], Mengzhe Zheng[7], He Zhao[1,2,3] and Yumeng Zhou[7]

The reliability of the near-land-surface air temperature (LSAT) projections from the state-of-the-art climate-system models that participated in the Coupled Model Intercomparison Project phase six (CMIP6) is debatable, particularly on regional scales. Here we introduce a method of constructing a constrained multi-model-ensemble (CMME), based on rejecting models that fail to reproduce observed LSAT trends. We use the CMME to constrain future LSAT projections under the Shared Socioeconomic Pathways 5–8.5 (SSP5–8.5) and 2–4.5 (SSP2–4.5), representing the high and intermediate scenarios. In comparison with the "raw" (unconstrained) CMIP6 multi-model ensemble (MME) mean, the impact of the observation-based constraint is less than 0.05°C 100 years^{-1} at a global scale over the second half of 21st century. However, the regional results show a wider range of positive and negative adjustments, from -1.0°C 100 years^{-1} to 1°C 100 years^{-1} under the SSP5–8.5 scenario. Although amplitude under SSP2–4.5 is relatively smaller, the CMME adjustment is similar to that under SSP5–8.5, indicating the scenario independency of the CMME impact. The ideal 1pctCO2 experiment suggests that the response of LSAT to carbon dioxide (CO_2) forcing on regional scales is responsible for the MME biases in the historical period, implying the high reliability of CMME in the 21st century projections. The advantage of CMME is that it goes beyond the idea of "model democracy" assumed in MME. The unconstrained CMIP6 MME may be overestimating the risks of future warming over North America, but underestimating the risks over Asia.

INTRODUCTION

The CMIP6 program is the most recent effort in coordinating the design and distribution of a large number of the state-of-the-art climate model simulations of the past, present, and future climates[1]. It serves as the basis for the sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC AR6)[2]. The observed global warming trend can be generally well captured by the state-of-the-art climate system models participating in CMIP6[3]. Almost two-thirds of the impacts related to atmospheric and ocean temperature changes can be confidently attributed to anthropogenic forcing[4]. There is a linear relationship between global mean surface air temperature and the equivalent-CO_2 concentrations[5]. The equivalent-CO_2 is calculated by converting amounts of all the greenhouse gases to the equivalent amount of carbon dioxide with the same global warming potential. The contribution from equivalent-CO_2 to historical global mean surface air temperature is about 70% on multi-decadal or longer timescales.

Greenhouse gases (GHG), CO_2 in particular, are considered to be the major external forcing in the 20th century and will also have a crucial impact in the 21st century projection. A high correlation between global mean near-surface air temperature and the CO_2 concentrations has also been found in future projections[6]. However, many of the latest CMIP6 models have larger climate sensitivities than the previous CMIP5 generation[7–9]. Higher climate sensitivity indicates the 'hot model' problem in CMIP6 models[10], and the larger climate sensitivity range suggests a wider range of warming responses to CO_2-forcing and larger model uncertainty in future warming projections[11].

There have been several studies trying to narrow uncertainty on estimates of past and future human-induced warming based on detection and attribution techniques or metrics of climate sensitivity[12,13]. There is a strong correlation between the recent global warming trend and transient climate response (TCR), and the past warming trend therefore can be used to constrain future warming projections in climate models on global scales[14]. But most of the previous studies focus on the projections on a global scale[15]. However, it is regional-scale changes that are of vital importance for the impacts of climate change. Under global warming, climate changes are more extreme regionally and induce more severe impacts. For example, the rate of heatwaves increases since the mid-twentieth century, but trend magnitudes are not globally uniform. Decadal trends in the frequency of heatwaves are biggest over northern South America, the Middle East, and the Maritime Continent at 50% per decade, but range between 10% and 30% per decade over most of the other regions[16]. In this study, we try to constrain future warming projections but focus on regional land surface air temperature (LSAT) trend projection under the high-emission SSP5–8.5 scenario and intermediate-emission SSP2–4.5 scenario[17]. The SSP5–8.5 and SSP2–4.5 update the RCP8.5 and RCP4.5 in CMIP5 and have the most participating climate models. We only focus on surface air temperature trend over land since Atlantic multidecadal oscillation (AMO) and Pacific decadal oscillation (PDO) account for most of the decadal variability over the oceans[5,18,19].

[1]CMA Earth System Modeling and Prediction Centre, China Meteorological Administration, 100081 Beijing, China. [2]State Key Laboratory of Severe Weather, 100081 Beijing, China. [3]Key Laboratory of Earth System Modeling and Prediction China Meteorological Administration, 100081 Beijing, China. [4]Laboratoire de Météorologie Dynamique, IPSL, CNRS, Sorbonne Université, Ecole Normale Supérieure, Ecole Polytechnique, 75005 Paris, France. [5]Centre for Climate Research Singapore, 537054 Singapore, Singapore. [6]Met Office Hadley Centre, EX1 3PB Exeter, UK. [7]Chinese Academy of Meteorological Sciences, China Meteorological Administration, 100081 Beijing, China. [8]University of Chinese Academy of Sciences, 100049 Beijing, China. ✉email: jiezhang@cma.gov.cn; twwu@cma.gov.cn
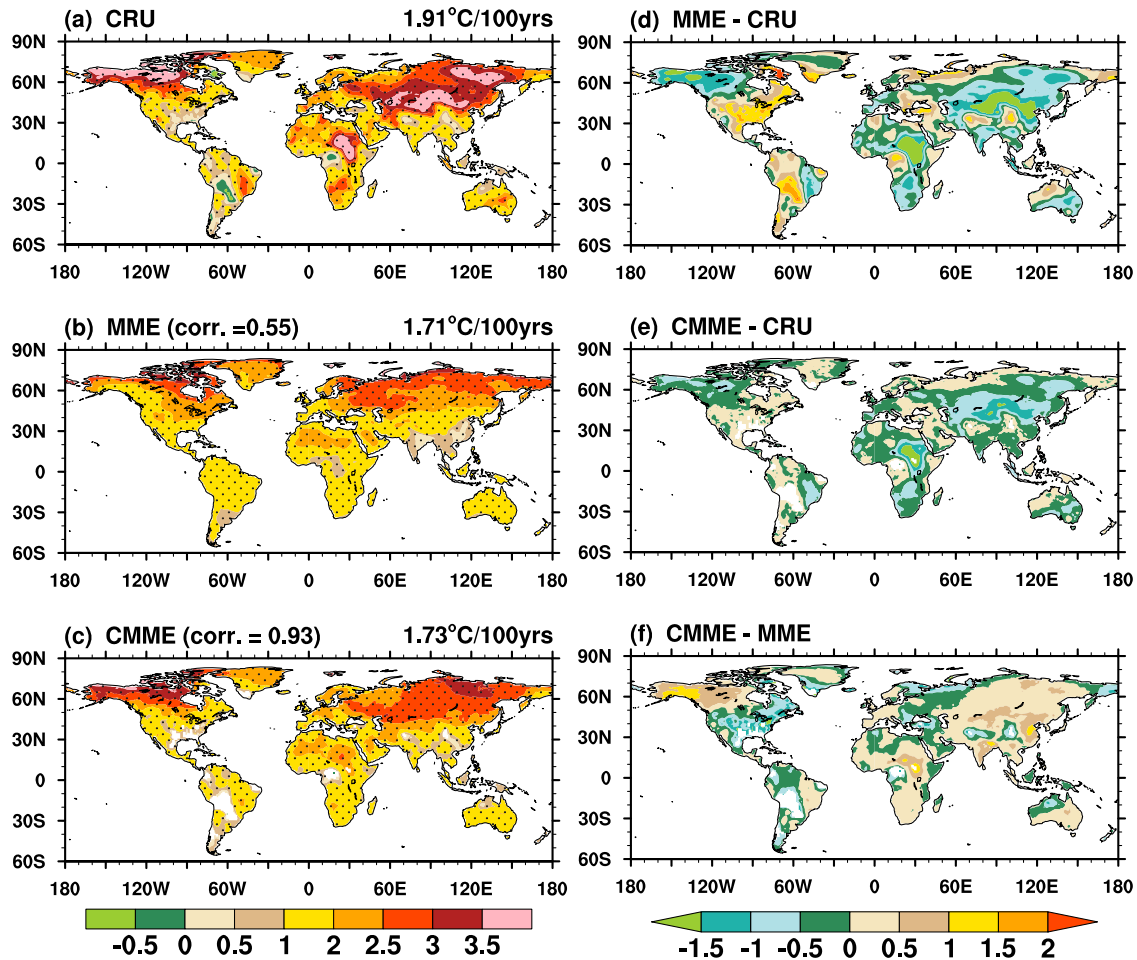
Fig. 1 Land surface air temperature (LSAT) trends from 1950 to 2014 in. a CRU, **b** raw (MME), and **c** constrain (CMME) ensemble under 0.5 ~ 1.5 OBT of 33 CMIP6 couple models. Global mean LSAT trends are marked at the top-right corner of Figs. 1a–c. **d** and **e** are the trend biases in MME and CMME, respectively. **f** The trend differences between CMME and MME. Units: °C 100 years⁻¹.

RESULTS

Historical LSAT trend: observation and CMIP6 simulations

Figure 1a shows the LSAT trend from 1950 to 2014 in the CRU dataset. Almost the entire global land surface has experienced significant surface warming. The warming trend is remarkable over the mid- and high-latitudes of the Northern Hemispheric continent and northeast Africa, weakest over the Tibetan Plateau and its downstream area, and the Andes Mountains along the western edge of South America. The global warming features can be generally well captured by the ensemble mean of the 33 CMIP6 models (MME, Fig. 1b). The spatial correlation coefficient between MME and CRU is 0.55, higher than the correlations for 32 of the 33 models. However, the LSAT trend in MME is spatially much smoother than in CRU. As shown in Fig. 1d, regions with LSAT trend underestimated by more than 0.5 °C 100 years⁻¹ in MME are where the observed warming rate higher than 1°C 100 years⁻¹, e.g., over Asia and Alaska. And the overestimation in MME is where the observed warming rate is lower than 1°C 100 years⁻¹, e.g., over the Continental United States (U.S.) and South America. The amplitude of the underestimation over Asia and the overestimation over the U.S. is almost half of the regional warming rate in the observation. That is, there is a big challenge for recent models to reasonably reproduce the LSAT trend in Asia and the U.S. The longitudinal gradient between Europe and northern Asia shown in CRU is also much weaker in MME.

The standard MME tends to be a better estimate of the forced climate change of the real system than the result from a particular model by allowing model errors to cancel each other out for a

large enough ensemble and has been widely used in studies on climate change simulations and projections[2]. However, it may fail to capture the real LSAT trend since some heavily-biased simulations may diminish the real signals or lead to unreliable results. The capabilities of individual models in reproducing the LSAT trend may also vary with regions, e.g., models that can reasonably produce the LSAT trend in Asia may fail to simulate the LSAT trend in the U.S. or other areas. In this study, we build ensemble at each grid-point based on certain OBS-Based Threshold (OBT), which we call a "constrained multi-model ensemble" (noted hereafter as CMME, see Methods section for details). General features of LSAT trend in CMME (Fig. 1c), constrained by the chosen OBT, are similar to those in MME. The blank areas over land, such as the Andes Mountains, are where no model can capture the LSAT trend under the OBT. Although the global mean LSAT trend barely change (1.71 °C 100 years⁻¹ in MME versus 1.73 °C 100 years⁻¹ in CMME), the spatial correlation coefficient with CRU increases to 0.93 in CMME. The mean CMME biases are 0.3°C 100 years⁻¹ (Fig. 1e), less than half of that in MME (0.71°C 100 years⁻¹). As expected, the differences between CMME and MME are similar to the MME biases but in opposite sign over 93% of the total area (Fig. 1f). Regions with the largest differences between CMME and MME are the areas where the MME shows the largest biases, i.e., Asia and the U.S. Therefore, we conclude that the warming amplitude may be largely biased on regional scales, although the geographic distribution of recent warming in MME is better than most individual models. The CMME result further
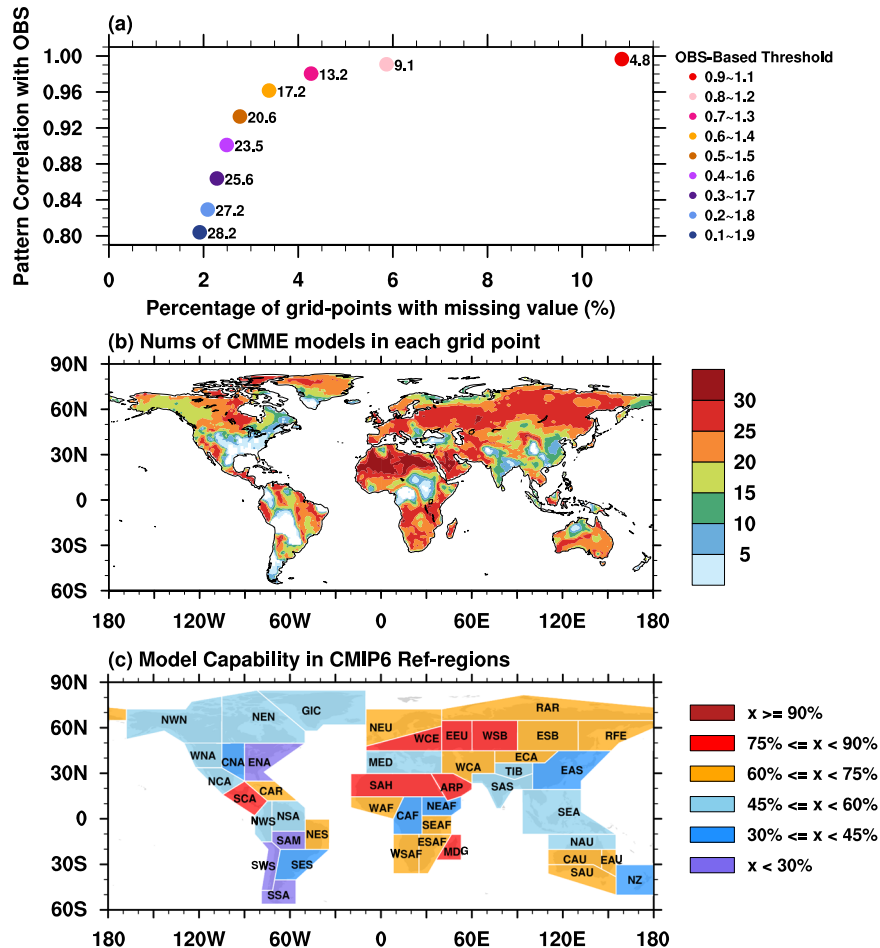
Fig. 2 Capability of CMIP6 models in reproducing LSAT trend from 1950 to 2014. a Percentage of grid-points that the LSAT trend is missing in CMME (x-axis) versus pattern correlation between CMME and OBS (y-axis) under different OBS-Based Threshold (OBT). The global mean model numbers in CMME with different OBT are marked at the right-hand side of the corresponding color dot. **b** Numbers of the CMIP6 models that contribute to the CMME under 0.5 ~ 1.5 OBT at each grid-point. **c** Percentage of models that contribute to CMME in each IPCC AR6 WG1 reference regions, defined as model capability in this study. The acronym of each region in c are introduced in Table 1.

improves the geographic distribution of LSAT trend in MME and its amplitude at a regional scale, because by construction the CMME is closer to the observed trends.

Model performances in reproducing LSAT trends vary with regions. As shown in Fig. 2b, more than 20 models are able to reproduce the LSAT trend over North Africa and the high latitude of Asia under the OBT. However, model performances are relatively lower over the U.S., western South America, India, and East Asia, where the LSAT trend is smaller than 1°C 100 years^{-1} (Fig. 1a). We further quantify model capability in reproducing regional LSAT trend in the 44 IPCC AR6 WG1 regions (Fig. 2c). Model capability is defined as the fraction of models that contribute to CMME under the OBT for each region. Region definitions follow the IPCC AR6 Working Group 1 (WG1) reference regions over land[20]. Region names and their acronym are introduced in Table 1, as well as the model capability in each region. Model capability is relatively high over Central America, the Central Eurasian continent, the northern and southern Africa, but relatively less good over North and South America, East and South Asia.

The adjustments in CMME upon MME in Future LSAT projections

The impact of constrained ensemble under the high scenario SSP5–8.5 is assessed (Fig. 3a–c). Since the SSP5–8.5 is intended to explore an extremely high-risk future and the projections perhaps not the most realistic[21], we also examine the impact under a more

optimistic scenario SSP2–4.5 (Fig. 3d–f). The LSAT trends from 2050 to 2100 are examined. The growing rates of CO_2 concentration in SSP5–8.5 and SSP2–4.5 are about 11.7 ppmv year^{-1} and 2.0 ppmv year^{-1}, respectively. The spatial patterns of the warming under these two different scenarios are similar and close to those featured in the 20th century. Globally, the LSAT trend in MME is about 7.62°C 100 years^{-1} under the SSP5–8.5 scenario (Fig. 3a) and 2.40 °C 100 years^{-1} under the SSP2–4.5 scenario (Fig. 3d). Consistent with what we observed in the historical period, the global mean LSAT trend projections in CMME are also close to those in MME (Fig. 3b and e). However, the regional differences are pronounced and generally grow with the growing rate of CO_2 concentration. Left out low model capability over the Andes Mountains, the constraint of LSAT projections under both scenarios generally resemble the effect of constraint ensemble in the historical period (Fig. 1f) over more than 79% and 71% of the land surface under the SSP5–8.5 scenario (Fig. 3c) and the SSP2–4.5 scenario(Fig. 3f), respectively: regions that show exaggerated historical warming trend are also the regions where CMME tends to reduce the projected warming rate, and vice versa. According to CMME adjustments, warming projection over the Eurasian continent may be more intense than the raw ensemble (MME), whereas the warming risk over the U.S. may be lower than expected. The similarities between the CMME impacts under different scenarios indicate the scenario independency of CMME effect and the essential role of the CO_2-concentration trend.

Table 1. Model capability in reproducing the LSAT trend from 1950 to 2014 in 44 IPCC AR6 WGI reference regions (units: %) and LSAT trend in 2050–2100 under SSP5–8.5 and SSP2–4.5 scenarios in MME and the adjustment (CMME-MME, °C 100 years^{-1}).

Region	Acronym	Full Name	Model Capability	SSP5–8.5 (Adjustment)	SSP2–4.5 (Adjustment)
North America	NWN	North-Western North America	62.2	9.87 (0.29)	3.67 (0.06)
	NEN	North-Eastern North America	60.2	11.53 (−0.60)	**3.86 (−0.31)**
	WNA	Western North America	56.4	7.16 (−0.38)	**2.34 (−0.22)**
	CNA	Central North America	39.8	**7.68 (−0.81)**	**2.44 (−0.46)**
	ENA	Eastern North America	45.0	**7.75 (−0.90)**	**2.59 (−0.68)**
Central America	NCA	Northern Central America	**74.6**	6.62 (−0.40)	**1.97 (−0.26)**
	SCA	Southern Central America	**82.6**	6.04 (−0.01)	1.76 (0.03)
	CAR	Caribbean	**77.3**	4.84 (0.16)	1.54 (0.08)
South America	NWS	North-Western South America	46.4	6.27 (−0.08)	1.87 (0.04)
	NSA	Northern South America	46.1	**6.96 (−0.57)**	**1.99 (−0.23)**
	NES	North-Eastern South America	64.7	6.27 (0.12)	1.98 (0.10)
	SAM	South American Monsoon	*24.0*	7.48 (−0.05)	2.32 (0.01)
	SWS	South-Western South America	*26.9*	**6.11 (−0.41)**	**1.91 (−0.17)**
	SES	South-Eastern South America	42.3	5.60 (−0.18)	1.76 (−0.10)
	SSA	Southern South America	*28.9*	4.46 (−0.23)	1.41 (0.06)
Europe	GIC	Greenland/Iceland	49.7	**8.70 (−0.62)**	3.18 (−0.11)
	NEU	Northern Europe	57.3	7.04 (0.02)	2.38 (0.02)
	WCE	Western and Central Europe	**69.5**	7.24 (0.04)	2.26 (−0.04)
	EEU	Eastern Europe	**70.4**	8.79 (0.15)	3.06 (0.07)
	MED	Mediterranean	59.7	7.19 (−0.15)	2.12 (−0.11)
Africa	MED	Mediterranean	59.7	7.19 (−0.15)	2.12 (−0.11)
	SAH	Sahara	**86.0**	7.65 (0.04)	2.26 (0.00)
	WAF	Western Africa	**70.1**	6.44 (0.03)	2.05 (0.00)
	CAF	Central Africa	44.0	6.09 (0.09)	1.99 (−0.03)
	NEAF	North Eastern Africa	51.7	5.86 (0.33)	1.94 (−0.01)
	SEAF	South Eastern Africa	**77.8**	5.63 (0.05)	1.84 (−0.02)
	WSAF	West Southern Africa	**75.8**	7.06 (0.05)	2.23 (−0.01)
	ESAF	East Southern Africa	**74.5**	6.67 (−0.09)	2.09 (−0.03)
	MDG	Madagascar	**78.8**	5.40 (−0.06)	1.62 (−0.01)
Asia	RAR	Russian Arctic	**69.3**	11.95 (−0.15)	3.97 (−0.07)
	WSB	West Siberia	**79.8**	9.72 (0.38)	3.02 (0.06)
	ESB	East Siberia	**72.7**	9.48 (0.58)	2.78 (0.17)
	RFE	Russian Far East	**71.3**	9.95 (0.15)	3.20 (−0.03)
	WCA	West Central Asia	**68.8**	8.00 (0.18)	2.28 (0.05)
	ECA	East Central Asia	**70.1**	**8.40 (0.59)**	2.33 (0.13)
	TIB	Tibetan Plateau	54.2	7.98 (−0.25)	2.30 (−0.09)
	EAS	East Asia	48.0	6.99 (0.40)	2.27 (0.08)
	ARP	Arabian Peninsula	**82.5**	7.71 (−0.07)	2.25 (−0.07)
	SAS	South Asia	54.4	6.64 (0.26)	2.04 (−0.01)
	SEA	South East Asia	65.3	4.98 (−0.06)	1.61 (−0.02)
Australasia	NAU	Northern Australia	48.5	5.91 (−0.32)	**1.88 (−0.15)**
	CAU	Central Australia	58.7	7.08 (−0.21)	2.08 (−0.08)
	EAU	Eastern Australia	**71.0**	6.23 (0.19)	1.82 (−0.10)
	SAU	Southern Australia	**71.7**	5.64 (0.19)	1.72 (0.05)
	NZ	New Zealand	*29.5*	**4.72 (−0.47)**	**1.39 (-0.35)**

Note: Model capability larger than 66% (22 of 33 models) and adjustment larger than 7% of the MME trend are bolded.

Regions with the most considerable adjustments in future projections

The regional LSAT projections in the raw and constrained ensembles are further detailed over the 44 landscape reference regions. Considering the similarities between the adjustments under the different scenarios, we mainly focus on the adjustments under the SSP5–8.5 scenario. As shown in Fig. 4a, LSAT increases monotonically over all regions but the warming rates are quite
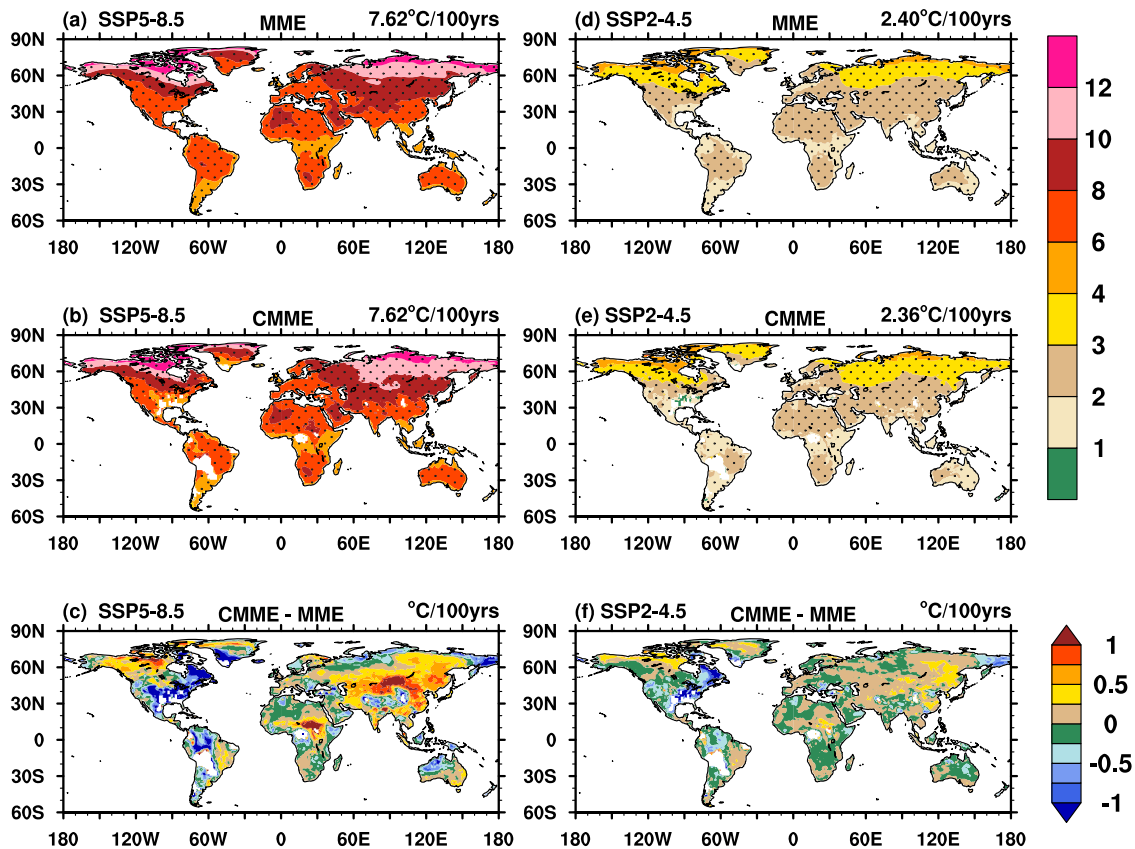
Fig. 3 Patterns of LSAT trend under the SSP5–8.5 and SSP2–4.5 scenarios from 2050 to 2100. (**a**) and (**d**) are results in MME. (**b**) and (**e**) are results in CMME. Global mean LSAT trend is marked at the top-right corner of the corresponding plot. (**c**) and (**f**) are the differences between CMME and MME. Units: °C 100 years^{-1}.

different. The three regions with the highest increasing rate (higher than 9.9°C 100 years^{-1}) in MME are the Russian Arctic (RAR), North-Eastern North America (NEN), and Russian Far East (RFE), all of which locate at high latitudes in the Northern Hemisphere. The three least warming regions (lower than 4.9°C 100 years^{-1}) are the Caribbean (CAR), New Zealand (NZ), and Southern South America (SSA), surrounded by ocean.

The LSAT trend during the second half of the 21st century (2050–2100) in each region is also listed in Table 1. The CMME adjustments vary from -0.90 °C 100 years^{-1} in Eastern North America (ENA) to 0.59 °C 100 years^{-1} in East Central Asia (ECA). Accounting for the adjustment larger than 0.2 °C/ 100 years^{-1}, warming trends are intensified only over 7 regions but suppressed over 13 regions. The amplification of warming is evident in Asia and Africa. Suppression of warming is mostly located over the American continents. Relative to the regional LSAT trend in MME, the warming adjustment is generally weaker and most significant over ECA (7%); the cooling adjustment prevails over ENA (-11.6%) and Central North America (CNA, -10.5%) in North America, Northern South America (NSA, -8.2%) in South America, Greenland/Iceland (GIC, -7.1%), and NZ (-10%). The adjustment may be also partly attributed to forcings other than GHG. Previous observational and model-based studies have found a cooling trend over the southeast and central United States, i.e., the United States 'warming hole'. This is suggested to be attributed to anthropogenic aerosol forcing or internal climate variability with dominant variation by season, region, and time[22]. Although the amplitude is relatively smaller, the LSAT trends in MME and adjustments in CMME under the SSP2–4.5 scenario among the 44 regions vary linearly with those under the SSP5–8.5 scenario, with correlation coefficients of 0.96 and 0.87, respectively.

Generally, as suggested by CMME, the warming challenge in future projections may be more intense over Asia and less intense over American continents than suggested by the raw CMIP6 MME projections. The physical mechanism of the CMME adjustment is investigated and discussed in the next subsection.

LSAT-response to CO_2-forcing and its impact on the LSAT trend reproduction

Global warming in the 20th century is mainly attributable to the increase in well-mixed greenhouse gases (WMGHGs), mainly CO_2. The 1pctCO2 experiment is a transient climate simulation, in which the CO_2 is the only anthropogenic external forcing. It is an idealized CO_2-forced experiment that resembles the CO_2 forcing during the Industrial Era. Here we compare the LSAT-response to CO_2-forcing in MME and CMME in 1pctCO2 to quantify the effect of our bias-correction procedure, i.e., the CMME, in eliminating CO_2-forcing response biases in climate models.

Figure 5a and b show the LSAT linear trend from 1850 to 1910 in the 1pctCO2 experiments for MME and CMME. The great similarity between the LSAT-response to CO_2-forcing (Fig. 5a) and the LSAT trend with all forcings (Fig. 1b) with a correlation coefficient of 0.91, confirms the dominant role of CO_2 in global warming. The warming trend in CMME generally resembles the MME results, but with different amplitude of regional warming (Fig. 5b). Figure 5c shows the differences between CMME and MME in LSAT-response to CO_2-forcing in the 1pctCO2 experiment, which allows us to compare the CO_2-response in MME to what the response would be if the historical LSAT biases were smaller. Interestingly, their differences closely resemble the historical trend biases in the MME with opposite signs (Fig. 1d). Over more than 15,000 model grids, the corresponding spatial correlation
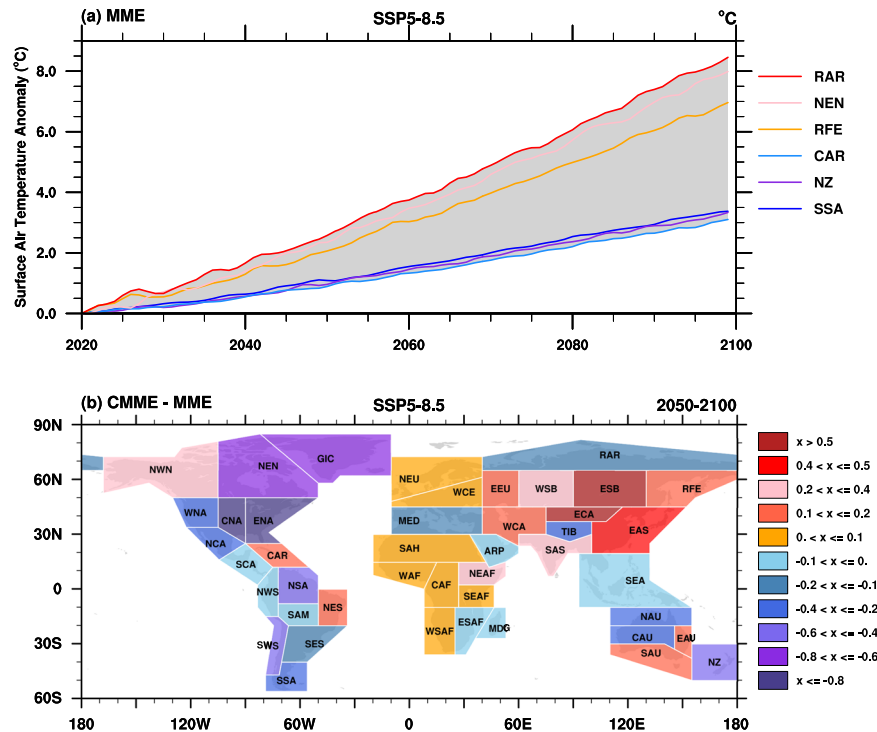
Fig. 4 LSAT trend projections under the SSP5–8.5 scenario. a The range of annual mean LSAT anomalies in the 44 CMIP6 reference regions in MME from 2020 to 2100 (shading) and the evolutions of LSAT anomalies in the regions with the highest and least warming rates. The anomalies are relative to the year 2020. **b** The LSAT trend differences from 2050 to 2100 between CMME and MME in the 44 CMIP6 regions. For a reasonable comparison, values over the areas where the values are missing in CMME are not considered in MME.

coefficient is -0.76, significant at 99% confidence level using student's t-test. That is, the historical LSAT trend biases in MME are mainly due to model capability in simulating CO_2-forcing response. Future constraint in SSP5–8.5 and SSP2–4.5 by CMME is therefore physically relative to its more reasonable CO_2-forcing response since CO_2 is also the dominant forcing in these scenarios.

DISCUSSION

The radiative forcing due to anthropogenic activity has increased continuously during the Industrial Era, with increases in atmospheric CO_2 playing a major role. By observational constraint, our study found that model response to CO_2-forcing is crucial for model performances in reproducing recent LSAT trends, especially on regional scales. Therefore, under future scenarios with increasing CO_2 emission, models with better capability in reproducing recent LSAT trends are expected to give better estimations of future LSAT change.

In this study, we introduced a method to constrain future LSAT projections based on observed constrained model ensemble (CMME), especially on regional scales. The CMME is generally an extension of the standard MME method but the effect of poorly performing models is eliminated from the mean and its estimates of the future are informed by present-day model performance. Under the SSP5–8.5 and SSP2–4.5 scenarios, the CMME is close to the raw CMIP6 ensemble (MME) on a global scale, but the regional warming amplitudes are quite different. CMIP6 models may underestimate future warming over Asia but overestimate it over the American continents. More specifically, the cooling adjustments over Central and Eastern North America (CNA and ENA) are more than 10% of the MME projection.

Therefore, we should take care when handling regional LSAT projections by CMIP6 models, since the CMIP6 model response to CO_2-forcing may be larger or smaller than expected and vary with region. Moreover, we should take great caution with projections

over the regions with larger projection adjustment and lower model capability, i.e., South-Western South America (SWS) and New Zealand (NZ), where factors other than CO_2 may also be important for the LSAT changes.

METHODS

The constrained ensemble

The constrained ensemble is a set of ensemble members, by selecting models for each grid-point based on LSAT trends from 1950 to 2014 under certain OBS-Based Threshold (OBT). The OBT is set to a certain range centered on the observed LSAT trend. A model only contributes to the statistics of the CMME at the grid-points for which it satisfies the observational constraint. As shown in Fig. 2a, with the narrowing of OBT, the pattern correlation coefficient with the observation and the invalid grid-points (grid-points with missing value) both increases in CMME. We choose the 0.5 to 1.5 range of the corresponding observed counterpart as the threshold for CMME. In CMME, about 21 models (62% of the models) on average are selected, pattern correlation is about 0.93, and less than 3% of grid-points are filled with missing values.

We further examine the distribution of selected grid-points in each model (Figures not shown). Under the chosen OBT, the percentage of valid grid-points in all land surface grids ranges from 41.2% to 71.3%, about 60.7% on average. There are also not noticeable discontinuities in the coverage of valid grid-points. Quantile mapping bias correction algorithms are commonly used to modify future model projections[23]. However, due to the limited number of model projections, the TAS trends are discrete-continuous distributed after quantile mapping (Figures not shown), which does not meet the nature of TAS trend mappings. That is, quantile mapping bias correction algorithms may be not suitable here.

The constrained ensemble is used to calibrate the regional LSAT projection under the SSP5–8.5 and SSP2–4.5 scenarios in the 21st
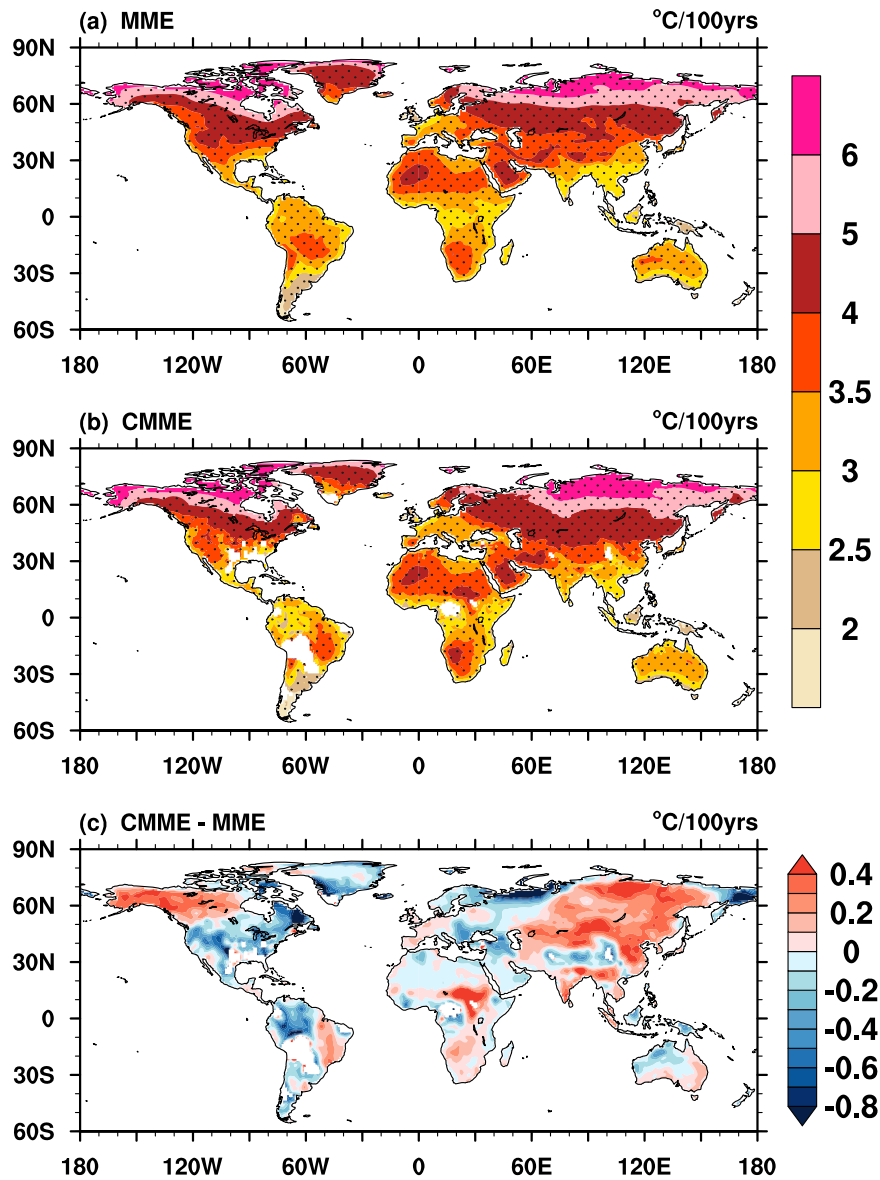
Fig. 5 Linear LSAT trend in the 1pctCO2 experiment from 1850 to 1910, when CO2 increases from pre-industrial value to about 500ppmv, in. (a) MME and (b) CMME. (c) Differences between CMME and MME. Units: °C/ 100 years^{-1}.

century in our study. The 3-point 1–2–1 low-pass filter is used to remove fluctuations less than 50 years in CRU and all the employed model outputs. It is aimed to minimize the impact of internal variability, especially the AMO and PDO. Further comparison shows that the effects of the band-pass filter on model selection are small, which suggests that the performance filter is selecting models mainly based on their forced responses, not on internal variability.

The observation and all the model outputs are interpolated to a regular 1° × 1° grid by bilinear interpolation.

Data

Grided observational dataset. The observational gauge-based gridded monthly mean LSAT from the Climatic Research Unit (CRU TS4) at University of East Anglia from 1950 to 2014 serves as the observation ref. [24]. The dataset was derived from historical weather stations and interpolated on a 0.5° grid over the land surface. It is highly reliable, especially in mid-latitudes where there is a dense observation network.

CMIP6 models datasets. Simulations of 33 models from the Coupled Model Intercomparison Project Phase 6 (CMIP6) were used. Four experiments are relevant to our study: all-forcing historical experiments from 1950–2014, idealized 1pctCO2 experiments from 1850 to 1910, and Shared Socioeconomic Pathways 5–8.5 (SSP5–8.5) and 2–4.5 (SSP2–4.5) experiments in the 21st century. Table 2 lists the models' names, institutions, and resolution of the atmospheric component. For the sake of equality, we use only the first member of each model if multiple runs were produced.

In the 1pctCO2 experiment, CO_2 increases at a prescribed rate of 1% per year from 285ppmv in 1850AD until the concentration doubles at model year 70. The 1pctCO2 experiment is an idealized configuration but can be used to identify the climate response to CO_2 increase as the experiment does not include any confounding effects from other climate forcings like aerosols and land-use change.

The SSP5–8.5 scenario represents the high end of the range of future pathways and produces a radiative forcing of 8.5 W m^{-2} in 2100. The SSP2–4.5 scenario represents the intermediate-emission scenario with a nominal 4.5 W m^{-2} radiative forcing level by 2100.

Table 2. Information of the 33 CMIP6 CGCM models used in this study.

Model	Institute	AGCM resolution
ACCESS-CM2	CSIRO-ARCCSS CSIRO and Austr. Res. Council Centre of Excellence for Climate System Science, Australia	140 km, 85 L, Top 85 km
ACCESS-ESM1–5	CSIRO Commonwealth Scientific and Industrial Research Organization, Australia	140 km, 38 L, Top 39 km
AWI-CM-1–1-MR	AWI Alfred Wegener Institute,Germany	80 km, 95 L,Top 80 km
BCC-CSM2-MR	BCC Beijing Climate Center, China	100 km, 46 L, Top 45 km
CAMS-CSM1–0	CAMS Chinese Academy of Meteorological Sciences, China	100 km, 31 L, Top 31.2 km
CanESM5	CCCMa Canadian Centre for Climate Modelling and Analysis, Canada	250 km, 49 L, Top 48 km
CAS-ESM2–0	CAS Chinese Academy of Sciences, China	140 km, 30 L, Top 42 km
CESM2-WACCM	NCAR National Center for Atmospheric Research, U.S.A.	100 km, 70 L, Top 80 km
CIESM	THU Department of Earth System Science, China	100 km, 30 L, Top 42 km
CMCC-CM2-SR5	CMCC Centro Euro-Mediterraneo sui Cambiamenti Climatici, Italy	100 km, 30 L, Top 43 km
CMCC-ESM2	CMCC	100 km, 30 L, Top 43 km
EC-Earth3-CC	EC-Earth consortium, Europe	80 km, 91 L, Top 80 km
EC-Earth3-Veg	EC-Earth	80 km, 91 L, Top 80 km
FGOALS-f3-L	CAS	90 km, 32 L, Top 42 km
FGOALS-g3	CAS	190 km, 26 L, Top 42 km
FIO-ESM-2-0	FIO-QNLM First Institute of Oceanography, and Pilot National Laboratory for Marine Science and Technology, China	100 km, 26 L, Top 43 km
GFDL-CM4	NOAA-GFDL National Oceanic and Atmospheric Administration, Geophysical Fluid Dynamics Laboratory, U.S.A.	100 km, 33 L, Top 48 km
GFDL-ESM4	NOAA-GFDL	100 km, 49 L, Top 80 km
IITM-ESM	CCCR-IITM Centre for Climate Change Research, Indian Institute of Tropical Meteorology, India	170 km, 64 L, Top 61 km
INM-CM4-8	INM Institute for Numerical Mathematics, Russia	150 km, 21 L, Top 31 km
INM-CM5-0	INM	150 km, 73 L, Top 61 km
IPSL-CM6A-LR	IPSL Institut Pierre-Simon Laplace, France	160 km, 79 L, Top 80 km
KACE-1-0-G	NIMS-KMA National Institute of Meteorological Sciences, Korea Meteorological Administration, Korea	140 km, 85 L, Top 85 km
KIOST-ESM	KIOST Korea Institute of Ocean Science & Technology, Korea	190 km, 32 L, Top 43 km
MIROC6	MIROC consortium JAMSTEC, AORI, NIES, RCCS, Japan	120 km, 81 L, Top 80 km
MIROC-ES2L	MIROC	250 km, 40 L, Top 40 km
MPI-ESM1-2-HR	MPI-M Max Planck Institute for Meteorology, Germany	80 km, 95 L, Top 80 km
MPI-ESM1-2-LR	MPI-M	170 km, 47 L, Top 80 km
MRI-ESM2-0	MRI Meteorological Research Institute, Japan	100 km, 80 L, Top 80 km
NESM3	NUIST Nanjing University of Information Science and Technology, China	170 km, 47 L, Top 48 km
NorESM2-LM	NCC NorESM Climate Modelling Consortium, Norway	190 km, 32 L, Top 40 km
NorESM2-MM	NCC	100 km, 32 L, Top 40 km
UKESM1-0-LL	MOHC Met Office Hadley Centre, U.K.	140 km, 85 L, Top 85 km

DATA AVAILABILITY

CODE AVAILABILITY

All data processing codes are available if a request is sent to the corresponding authors.

REFERENCES

1. Eyring, V. et al. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).
2. IPCC. Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. (eds Pörtner,H.-O. et al.) (Cambridge University Press, Cambridge, 2022).
3. Papalexiou, S. M., Rajulapati, C. R., Clark, M. P. & Lehner, F. Robustness of CMIP6 Historical Global Mean Temperature Simulations: Trends, Long-Term Persistence, Autocorrelation, and Distributional Shape. *Earth's Future* **8**, e2020EF001667 (2020).
4. Hansen, G. & Stone, D. Assessing the observed impact of anthropogenic climate change. *Nat. Clim. Change* **6**, 532–537 (2016).
5. Wu, T., Hu, A., Gao, F., Zhang, J. & Meehl, G. New insights into natural variability and anthropogenic forcing of global/regional climate evolution. *npj Clim. Atmos. Sci.* **2**, 18 (2019).
6. Gao, F., Wu, T., Zhang, J., Hu, A. & Meehl, G. A. Shortened Duration of Global Warming Slowdowns with Elevated Greenhouse Gas Emissions. *J. Meteorol. Res-Prc* **35**, 225–237 (2021).
7. Flynn, C. M. & Mauritsen, T. On the climate sensitivity and historical warming evolution in recent coupled model ensembles. *Atmos. Chem. Phys.* **20**, 7829–7842 (2020).
8. Meehl, G. A. et al. Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. *Sci. Adv.* **6**, eaba1981 (2020).
9. Nijsse, F. J. M. M., Cox, P. M. & Williamson, M. S. Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models. *Earth Syst. Dynam.* **11**, 737–750 (2020).
10. Hausfather, Z., Marvel, K., Schmidt, G. A., Nielsen-Gammon, J. W. & Zelinka, M. Climate simulations: recognize the 'hot model' problem. *Nature* **605**, 26–29 (2022).
11. Zelinka, M. D. et al. Causes of Higher Climate Sensitivity in CMIP6 Models. *Geophys. Res. Lett.* **47**, e2019GL085782 (2020).
12. Stott, P. A. & Kettleborough, J. A. Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature* **416**, 723–726 (2002).
13. Knutti, R., Rugenstein, M. A. A. & Hegerl, G. C. Beyond equilibrium climate sensitivity. *Nat. Geosci.* **10**, 727–736 (2017).
14. Tokarska, K. B. et al. Past warming trend constrains future warming in CMIP6 models. *Sci. Adv.* **6**, eaaz9549 (2020).
15. Ribes, A., Qasmi, S. & Gillett, N. P. Making climate projections conditional on historical observations. *Sci. Adv.* **7**, eabc0671 (2021).
16. Perkins-Kirkpatrick, S. E. & Lewis, S. C. Increasing trends in regional heatwaves. *Nat. Commun.* **11**, 3357 (2020).
17. O'Neill, B. C. et al. The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geosci. Model Dev.* **9**, 3461–3482 (2016).
18. Chylek, P., Klett, J. D., Lesins, G., Dubey, M. K. & Hengartner, N. The Atlantic Multidecadal Oscillation as a dominant factor of oceanic influence on climate. *Geophys. Res. Lett.* **41**, 1689–1697 (2014).
19. Meehl, G. A., Hu, A. X., Santer, B. D. & Xie, S. P. Contribution of the Interdecadal Pacific Oscillation to twentieth-century global surface temperature trends. *Nat. Clim. Change* **6**, 1005–1008 (2016).
20. Iturbide, M. et al. An update of IPCC climate reference regions for subcontinental analysis of climate model data: definition and aggregated datasets. *Earth Syst. Sci. Data* **12**, 2959–2970 (2020).
21. Hausfather, Z. & Peters, G. Emissions – the 'business as usual' story is misleading. *Nature* **577**, 618–620 (2020).
22. Mascioli, N. R., Previdi, M., Fiore, A. M. & Ting, M. Timing and seasonality of the United States 'warming hole'. *Environ. Res. Lett.* **12**, 034008 (2017).
23. Cannon, A. J., Sobie, S. R. & Murdock, T. Q. Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? *J. Clim.* **28**, 6938–6959 (2015).
24. Harris, I., Osborn, T. J., Jones, P. & Lister, D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data* **7**, 109 (2020).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

The main ideas were formulated by J.Z. and T.W. J.Z. wrote the original draft. The results were supervised by T.W., L.L., and K.F. All the authors discussed the results and contributed to the final manuscript. J.Z. and T.W. are both corresponding authors of this work.

COMPETING INTERESTS

The authors declare that they have no conflict of interest.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Jie Zhang or Tongwen Wu.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.