

Supporting Information

1 Observational atmospheric temperature data

We used satellite estimates of atmospheric temperature change produced by three different groups:

1. Remote Sensing Systems in Santa Rosa, California (RSS) (S1);
2. The University of Alabama at Huntsville (UAH) (S2);
3. The Center for Satellite Applications and Research, NOAA/National Environmental Satellite, Data, and Information Service, Camp Springs, Maryland (STAR) (S3).

All three groups provide MSU-based estimates of layer-average temperature changes for the lower stratosphere (TLS) and the mid- to upper troposphere (TMT). Currently, only RSS and UAH produce a satellite-based estimate of atmospheric temperature change for the lower troposphere (TLT). The approximate altitude range and pressure level boundaries associated with each of these temperature measurements is given in Table 2 in (S4).

All analyses reported on here rely on the following versions of these temperature data sets:

1. Version 3.3 of RSS TLS, TMT, and TLT data, downloaded from:

<http://www.remss.com/data/msu/data/netcdf> on 1/9/2012;

2. Version 5.4 of UAH TLS, TMT, and TLT data, downloaded from:

<http://vortex.nsstc.uah.edu/data/msu> on 1/13/2012;

3. Version 2.0 of STAR TLS and TMT data, downloaded from:

<ftp.orbit.nesdis.noaa.gov> on 3/15/2012.

All observed MSU data sets were in the form of monthly means on $2.5^\circ \times 2.5^\circ$ latitude/longitude grids, and span the 396-month analysis period considered here (January 1979 to December 2011). We analyzed complete years only; data available for the last several months of 1978 and the initial months of 2012 were not used.

There are differences in the spatial coverage of the MSU data sets produced by the three groups. While the UAH MSU data have global coverage, STAR TLS and TMT products extend from 87.5°N to 87.5°S . RSS data sets extend from 82.5°N to 82.5°S for TLS and TMT, and from 82.5°N to 70°S for TLT. The RSS TLT coverage is restricted to 82.5°N to 70°S because:

1. Poleward of 82.5° , there are virtually no MSU brightness temperature measurements from the central view angle of the satellite “swath”;
2. In the Southern Hemisphere, the reliable estimation of brightness temperatures is hampered by the large (and poorly-known) surface emissivity contribution

from snow- and ice-covered areas of the Antarctic continent which lie above 3,000 meters (S5).

To exclude any impact of spatial coverage differences on trend comparisons, we calculated all spatial averages of observed and simulated MSU temperatures over the area of common coverage in the RSS, UAH, and STAR data (82.5°N to 82.5°S for TLS and TMT, and 82.5°N to 70°S for TLT.)

2 Details of model output

2.1 General information

We used model output from phase 5 of the Coupled Model Intercomparison Project (CMIP-5) (S6). A full list of modeling groups participating in CMIP-5 is given at http://cmip-pcmdi.llnl.gov/cmip5/docs/CMIP5_modeling_groups.pdf. Eventually, results from simulations performed by *ca.* 27 research groups (using over 60 different numerical models) will be incorporated in the CMIP-5 multi-model archive.

At the time our research was performed, the CMIP-5 archive was not fully populated with model results. We analyzed results from 20 different CMIP-5 models, contributed by 15 different research groups (see Table S1). Information on the horizontal and vertical resolution of these 20 models is given in Table S2. As noted in

the main text, we analyzed three different types of simulation:

1. Simulations with estimated historical changes in human and (in most cases) natural external forcings (see Table S3);
2. Simulations with 21st century changes in greenhouse gases and anthropogenic aerosols prescribed according to the Representative Concentration Pathway 8.5 (RCP8.5), with radiative forcing of approximately 8.5 W/m^2 in 2100, eventually stabilizing at roughly 12 W/m^2 (S7, S8);
3. Pre-industrial control runs with no changes in external influences on climate, which provide information on internal climate noise.

Details of the start dates, end dates, and lengths of the historical simulations and RCP8.5 runs are given in Table S4. Corresponding information for the pre-industrial control runs is supplied in Table S5. Note that the RCP8.5 simulations were initiated from conditions of the climate system at the end of the historical run.

In the case of GISS-E2-R, simulation output was available from two slightly different model versions (p1 and p2). For the purposes of calculating multi-model average (MMA) quantities, it was necessary to decide whether atmospheric temperatures from the p1 and p2 historical/RCP8.5 runs should be treated as different realizations of historical climate change performed with a similar physical model, or as results from two different models of the climate system. In the former case, the number of models

employed in estimating the MMA would have been 19 (rather than 20).

There are important differences between these two versions of GISS-E2-R. Historical and future changes in aerosols and ozone are prescribed in **p1**, but are interactive in **p2** (S9). Such differences can have significant implications for the atmospheric temperature changes (and temperature variability) simulated in these two model versions.¹ We therefore decided to treat **p1** and **p2** as separate models.

This decision also affects estimates of internal variability. Model versions **p1** and **p2** have separate control integrations (Table S5). Our baseline (“BASE”) detection and attribution calculations relied on atmospheric temperatures from the last 200 years of both the **p1** and **p2** pre-industrial control runs. Similarly, temperature results from both control runs were available for selection as the “TOP-5” model noise estimates.

2.2 Selection of models for O3+V fingerprint calculations

The purpose of our “O3+V” sensitivity study was to calculate fingerprints of externally-forced atmospheric temperature change using a subset of CMIP-5 models with more reliable estimates of forcing by stratospheric ozone and volcanic aerosols. Only 12 of the 20 CMIP-5 models analyzed here were used in estimating the O3+V finger-

¹For example, the interannual variability of TLS is larger in version **p2** than in version **p1** (*c.f.* Figs. 1J and 1I).

print. The 8 excluded models were CCSM4, CNRM-CM5, GFDL-CM3, GISS-E2-R (p2), INM-CM4, IPSL-CM5A-LR, MIROC-ESM-CHEM, and NorESM1-M. The justification for excluding these models from the O3+V fingerprint calculations is given below.

2.2.1 Volcanic aerosol forcing

Compared with all other models analyzed here, both INM-CM4 and IPSL-CM5A-LR have noticeable deficiencies in their volcanic aerosol forcing. (see Fig. 1). Consider INM-CM4 first. Although the INM-CM4 historical run included stratospheric volcanic aerosols, only their scattering effects were simulated (see Fig. 1M). Since stratospheric volcanic aerosols in INM-CM4 do not absorb incoming solar radiation and outgoing long-wave radiation, the model cannot generate volcanically-induced stratospheric warming signals.²

In IPSL-CM5A-LR, volcanically-induced cooling of surface and tropospheric temperature was indirectly represented by tuning the solar irradiance. The absence of aerosol-induced absorption (see above) explains the lack of short-term lower stratospheric warming after the eruptions of El Chichón and Pinatubo. The IPSL-CM5A-LR model actually shows slight cooling of TLS after the El Chichón and Pinatubo eruptions (see Fig. 1N). This is due to the decrease in outgoing long-wave radiation caused

²This does not explain why INM-CM4 fails to produce noticeable volcanic cooling signals in the troposphere (see Fig. 2M).

by the solar irradiance-tuned cooling of the troposphere and surface.

2.2.2 Stratospheric ozone forcing

In the following, we group CMIP-5 models into two classes:

1. CHEM models with fully interactive or semi-offline ozone chemistry;
2. NOCHEM models with prescribed changes in ozone.

Most of the CHEM models have errors in their simulations of historical ozone changes. These errors are not surprising, particularly since CMIP-5 is the first phase of the Coupled Model Intercomparison Project to incorporate a number of models with interactive ozone chemistry. Although the inclusion of interactive ozone chemistry is essential for the prediction of stratospheric ozone recovery, accurate simulation of historical changes in ozone is scientifically challenging. Errors in the physical climate system affect chemical reaction rates, thereby affecting simulations of historical changes in ozone concentrations, which in turn influence stratospheric temperature.

There is considerable uncertainty in observed estimates of ozone changes (S10). This uncertainty hampers the screening of CMIP-5 CHEM models based solely on their performance in simulating observed ozone changes. Here, we chose the simpler approach of estimating the O₃+V fingerprint using only the NOCHEM models. Of the 20 models analyzed here, 13 are NOCHEM models (see Table S3). One of these

NOCHEM models, **INM-CM4**, was excluded because of the above-described problems in its treatment of volcanic aerosol forcing. The O3+V fingerprint, therefore, was calculated using simulation output from 12 CMIP-5 models. These 12 models are identified in Figs. 1 and 2.

2.3 Splicing of simulation output

2.3.1 Splicing of CNRM-CM5 synthetic MSU temperatures

For 19 of the 20 CMIP-5 models analyzed here, our estimates of atmospheric temperature changes over 1979 to 2011 are based on synthetic MSU temperatures from the spliced historical/RCP8.5 simulations. In the case of **CNRM-CM5**, however, we relied on results from the spliced historical/historicalExt simulation. The reasons for this decision are described below.

The **CNRM-CM5** model has explicit treatment of the interactions of volcanic aerosols with long- and short-wave radiation. In the historical run, changes in stratospheric aerosol optical depth were specified according to Gao *et al.* (2008) (S11). At the start of the RCP8.5 run in January 2006, the stratospheric aerosol optical depth was set to the Gao *et al.* estimate of average stratospheric optical depth over 850 to 1999. This produces a small discontinuity in the volcanic aerosol forcing at the splice point between the **CNRM-CM5** historical and RCP8.5 runs, and therefore leads to a small

‘jump’ in lower stratospheric temperature. This jump is evident in each of the three CNRM-CM5 historical/RCP.5 realizations.

To evaluate the impact of this forcing discontinuity on estimated changes in atmospheric temperature, we analyzed three realizations of the CNRM-CM5 “historicalExt” simulation. The forcings by well-mixed greenhouse gases and anthropogenic aerosols are very similar in the CNRM-CM5 RCP8.5 and historicalExt runs, as are the predicted changes in stratospheric ozone. The major difference between these two sets of simulations is in the volcanic aerosol forcing. In historicalExt, the stratospheric aerosol optical depth over January 2000 to December 2012 is set to the *Gao et al.* estimate of stratospheric aerosol optical depth in December 1999 – *i.e.*, volcanic forcing is unchanged (and close to zero) from December 1999 onwards. So unlike the CNRM-CM5 spliced historical/RCP8.5 runs, the spliced historical/historicalExt runs have no discontinuity in volcanic forcing.

The step-function change in volcanic forcing between the end of the CNRM-CM5 historical run and the start of the RCP8.5 integration introduces a warm bias in the simulated TLS changes over the satellite era. As a result, the ensemble-mean lower stratospheric cooling trend over 1979 to 2011 is relatively weak in the CNRM-CM5 spliced historical/RCP8.5 integrations ($-0.091^{\circ}\text{C}/\text{decade}$), and larger and closer to observations in the CNRM-CM5 spliced historical/historicalExt runs ($-0.160^{\circ}\text{C}/\text{decade}$).

To remove this discontinuity in volcanic aerosol forcing, all analyses involving

CNRM-CM5 estimates of externally-forced temperature changes relied on results from the spliced historical/historicalExt simulation.

2.3.2 Impact of splicing on S/N ratios

If we had not performed any splicing, and had focused on the period 1979 to 2005,³ the maximum record length available for model-versus-data comparisons would have been only 27 years. As noted in the main text, splicing of the historical and RCP8.5 simulations enables us to compare modeled and observed atmospheric temperature trends over the full 33-year satellite era (1979 to 2011).

This increase in record length (from 27 to 33 years) has some effect on our estimated S/N ratios. The effect is primarily due to the decrease in the amplitude of noise trends with an increase in the trend fitting period (see Figs. 7B and 8B). It is this decrease in the amplitude of the noise (rather than an increase in signal amplitude) that explains why S/N increases by roughly 10-15% as the analysis period increases from 27 to 33 years (see Figs. 7C and 8C).

For the stratospheric and tropospheric temperature changes considered here, S/N ratios in 2005 are still well above the 1% significance threshold for almost all model-versus-observed detection and attribution (“D&A”) tests with the global-mean signal

³This is the period of maximum overlap between observed MSU records and the CMIP-5 historical runs (recall from Table S4 that most CMIP-5 historical runs end in 2005).

included (see, *e.g.*, Figs. 7C and 8C). This means that our primary conclusions regarding signal detection do not depend on our decision to perform splicing.

3 Calculation of TTT

TMT receives a substantial contribution from the cooling of the lower stratosphere (S12). Fu *et al.* (2004) developed a regression-based approach for removing the bulk of this stratospheric cooling component from observational estimates of tropospheric temperature changes (S12). This enabled Fu *et al.* to calculate TTT, which they refer to as the “temperature of the tropical troposphere” (S13). Fu *et al.* note that TTT represents “the entire troposphere from the surface to the tropopause.” Although concerns have been expressed about the efficacy of this approach (S14), tests with climate model simulation output suggest that the Fu *et al.* method is reasonably successful in capturing the true large-scale, multi-decadal trends in bulk tropospheric temperature (S15, S16).

Since the observed pattern of stratospheric cooling shows pronounced latitudinal and vertical structure, it would seem reasonable to generate geographical patterns of TTT trends with a regression-based approach which uses latitudinally varying regression coefficients. Given large differences in the latitudinal and altitudinal structure of stratospheric cooling in the observational and model TLS data sets analyzed here (see Fig. 3A), many different estimates of the latitudinal dependence of such regression

coefficients could be derived.

Here, we calculate TTT in two different ways: with regression coefficients that do not vary as a function of latitude (TTT), and with coefficients that differ in the tropics and extratropics (TTT^{*}). This provides us with information on the sensitivity of our TTT D&A results to details of the statistical method used for removing stratospheric influence on TMT. For reasons described below, we show only TTT results in the main text.

TTT is computed with the same regression coefficient applied by Fu *et al.* (S13) in their equation 1b:

$$\text{TTT} = a_{24}\text{TMT} + (1 - a_{24})\text{TLS} \quad (1)$$

where $a_{24} = 1.1$. For TTT^{*}, $a_{24} = 1.1$ between 30°N and 30°S, and $a_{24} = 1.2$ poleward of 30°. Relative to TTT, TTT^{*} removes more of the high-latitude stratospheric cooling influence on TMT, and thus has larger tropospheric temperature increases at high latitudes (see Fig. S5).

Regression was performed locally, at each model and observational grid-point. We note that model and observational temperature data were processed in exactly the same way – *i.e.*, model-versus-observed differences in total tropospheric temperature changes are not attributable to differences in the applied regression coefficients.

Gridded data sets of monthly-mean TTT and TTT^{*} were produced using the

actual TLS and TMT temperatures from the RSS, UAH, and STAR groups. They were also generated with the synthetic TLS and TMT temperatures calculated from the CMIP-5 pre-industrial control runs and spliced historical/RCP8.5 runs.

In the main text, we focus on the discussion of TTT, primarily because:

- TTT and TTT* yield very similar D&A results;
- Due to the way in which TTT* is defined, patterns of changes in TTT* have a small discontinuity at 30°N and 30°S (see Fig. S5). TTT trend patterns do not have this discontinuity.

4 RSS percentile realizations

Our D&A analysis relied on both the publicly-available version of the RSS atmospheric temperature data sets and the RSS “percentile realizations.” These realizations provide valuable information on the impact of different processing choices that are made during the construction of MSU-based data sets. As described in Mears *et al.* (2011), a Monte Carlo approach was used to estimate uncertainties in atmospheric temperature change arising from “sampling error, premerge adjustments to each individual satellite, and the merging procedure” (S1). For each of the four atmospheric layers considered here (TLS, TMT, TLT, and TTT), a 400-member ensemble of gridded, monthly-mean temperature data sets was generated.

Given the very large number of sensitivity tests which we performed, it was not feasible to use the full 400-member RSS observational ensembles in our D&A analysis. To reduce the dimensionality of the problem, we ranked the 400 realizations by their global-mean temperature trends over 1979 to 2011, and then chose 11 realizations for each atmospheric layer. Ranking was performed separately for TLS, TMT, TLT, and TTT. The selected realizations were the 5th, 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th, and 95th percentiles of the ranked distributions.

In our D&A analysis, each of the RSS percentile realizations was projected onto the searched-for fingerprint. This procedure adequately samples the uncertainty in the RSS estimates of global-scale atmospheric temperature change, but does not fully capture RSS-based uncertainty estimates in the spatial patterns of the temperature changes. It is likely that the true 5-95 percentile range of RSS-based D&A results is underestimated.⁴

5 Regridding of model and observational data

Model results were available on different grids (Table S2). In order to calculate fingerprints from the multi-model averages of the atmospheric temperature changes

⁴Note, however, that for the purposes of comparing modeled and observed trends in zonal-mean atmospheric temperature, the RSS 5 to 95 percentile range is based on the full 400-member RSS observational ensembles (see Figs. 3 and S5).

in the spliced historical/RCP8.5 runs, to compare these fingerprints with observations, and to obtain ‘pooled’ noise estimates from the concatenated control simulations, it was necessary to regrid temperature data from the observations and the native grids of all 20 CMIP-5 models to a common grid.

We transformed all data sets to a regular $10^\circ \times 10^\circ$ latitude/longitude grid using an area-weighted regridding algorithm (S17). Transformation to a relatively coarse-resolution grid reduces the spatial dimensionality of the input data sets, which is of benefit in the estimation of Empirical Orthogonal Functions (EOFs) used in the fingerprint analysis. Because changes in atmospheric temperature tend to be smoothly varying, regridding does not lead to appreciable loss of information on the spatial structure of the leading signal or noise modes.

To ensure that differences in observational coverage have no impact on the D&A analysis, we used the coverage common to the RSS, UAH, and STAR data sets (see Section 1 of *SI*). After regridding to the target $10^\circ \times 10^\circ$ grid, observational and model temperature data were extracted for 80°N - 80°S (TLS, TMT, TTT, and TTT*) and 80°N - 70°S (TLT).

6 Fingerprint analysis

6.1 Definition of fingerprint

Detection methods generally require an estimate of the true but unknown climate-change signal in response to an individual forcing or set of forcings (S18). This is often referred to as the fingerprint, $F(x)$. The fingerprint is what we search for in observational records. $F(x)$ may be defined in a variety of different ways. A common strategy, which we employ here, is to use the first EOF of the multi-model average change in synthetic MSU temperature as the fingerprint.⁵

Let $S(i, j, x, t)$ represent annual-mean synthetic MSU temperature data at grid-point x and time t from the i^{th} realization of the j^{th} model's spliced historical/RCP8.5 simulation, where:

$i = 1, \dots, N_r(j)$ (the number of realizations for the j^{th} model).

$j = 1, \dots, N_m$ (the number of models used in fingerprint estimation).

$x = 1, \dots, N_x$ (the total number of grid-points).

$t = 1, \dots, N_t$ (the time in years).

⁵Other possible fingerprint choices include the mean change from a large ensemble of initial condition realizations, the mean change or leading EOF from an equilibrium response experiment with large changes in forcing, *etc.*

Here, N_r ranges from 1 to 5 realizations (see Table S4); $N_m = 20$ models in the BASE case, and 12 models in the O3+V case; $N_x = 576$ grid-points⁶ for TLS, TMT, TTT, and TTT*, and 540 grid-points for TLT; and N_t is either 151 or 33 years (see below).

The multi-model average atmospheric temperature change, $\overline{\overline{S}}(x, t)$, was calculated by first averaging over an individual model's historical/RCP8.5 realizations (where multiple realizations were available), and then averaging over models.⁷ The fingerprint $F(x)$ is the first EOF of $\overline{\overline{S}}(x, t)$. It primarily reflects the large-scale patterns of cooling of the lower stratosphere and warming of the troposphere.

To explore the sensitivity of our S/N results to different plausible choices in the fingerprint estimation process, we calculated $F(x)$ using input atmospheric temperature data for two different periods: 1861 to 2011, and 1979 to 2011. Our primary goal was to examine whether the geographical pattern of the fingerprint exhibits important changes over time. Use of the longer, 151-year analysis period involves a larger overall change in anthropogenic forcing. If the normalized spatial pattern of the fingerprint⁸

⁶After transforming synthetic MSU temperature data from each model's native grid to the common $10^\circ \times 10^\circ$ latitude/longitude grid.

⁷The double overbar denotes two separate averaging steps (over realizations and models).

⁸For information on fingerprint normalization, refer to the description of the subroutine TQLI in (S19). Normalization reduces the fingerprint's sensitivity to differences in global-mean radiative forcing (and global-mean temperature change) over different time intervals.

is relatively stationary, use of a longer analysis period is advantageous, and reduces the impact of internal variability on the estimate of $F(x)$.

If the spatial pattern of response to external forcing had changed markedly over the past 151 years (S20),⁹ it would be preferable to calculate $F(x)$ over the same time period used for estimating observed changes in atmospheric temperature (1979 to 2011). Use of this shorter, 33-year period for fingerprint estimation still involves the implicit assumption that there are not pronounced changes over time in the large-scale structure of the climate response to external forcing. In cases where such changes exist, it is more appropriate to use full space-time detection methods, which explicitly incorporate the time dimension in the detection scheme (S21, S22).

For the 1861 to 2011 historical analysis period (“HIST”), the beginning and end dates were dictated by the start date of the GFDL-ESM2G and GFDL-ESM2M historical runs (see Table S4) and the end of the observational satellite record. All synthetic MSU temperatures in the HIST case were expressed as annual-mean anomalies relative to the climatological annual mean of the 151-year simulation. For the 1979 to 2011 analysis period (“SAT-ERA”), anomalies were calculated with respect to climatological annual means over this 33-year period. All fingerprints and S/N results shown in the main text are for the HIST case; fingerprints for the SAT-ERA are given in Fig. S8 of the *SI*. The sensitivity of our findings to the choice of analysis period is

⁹In response, for example, to low-frequency changes in the pattern of net aerosol forcing.

discussed in Section 7.3.2.

The O3+V fingerprints for the HIST case are given in row 1 of Fig. 6. The O3+V fingerprints are very similar to the BASE estimates of $F(x)$ (not shown). The total variance explained by the O3+V and BASE fingerprints always exceeds 85%.

Many D&A strategies seek to rotate $F(x)$ in a direction that maximizes the signal strength relative to the control run noise (S18, S21, S22). Optimization of $F(x)$ generally leads to enhanced detectability of the signal. In most cases we considered, we achieved detection of an externally-forced fingerprint in observations without any optimization of $F(x)$ (see Fig. 9 and Fig. S9). We therefore report only on S/N ratios obtained with non-optimized fingerprints.

6.2 Calculation of concatenated noise data sets

As discussed later in Section 6.3, we need to determine whether the pattern similarity between $F(x)$ and the time-varying observations shows a statistically significant increase over time. To address this question, we require control run estimates of internally-generated variability, in which we know *a priori* that there is no expression of the fingerprint, except by chance.

In our multi-model D&A framework, we obtain such variability estimates from the control runs performed with multiple models. Because the length of the 20 control

runs analyzed here varies by a factor of up to 4 (see Figs. S1 and S2 and Table S5), models with longer control integrations could have a disproportionately large impact on the BASE noise estimates. To guard against this possibility, the BASE noise estimates rely on only the last 200 years of each model’s pre-industrial control run,¹⁰ yielding 4,000 years of concatenated control run data.

For the TOP-5 model noise estimates, only the control runs of a five-model subset are concatenated. Selection of the TOP-5 models was performed separately for TLS, TMT, TLT, TTT, and TTT*, and is based on values of s_{LOW} , the temporal standard deviation of detrended, band-pass filtered atmospheric temperature data (see Fig. 5). In contrast to the BASE case, we use the full length of each of the five model control runs. This ensures that we have reasonable sample sizes when we rely on TOP-5 noise information to assess the significance of multi-decadal signal trends.¹¹

In both the BASE and TOP-5 noise cases, annual-mean synthetic MSU temperatures from individual model control runs are regridded to the same $10^\circ \times 10^\circ$ target grid used for fingerprint estimation. After regridding, anomalies are defined relative to climatological annual means over the full length of each control run.

Visual inspection of the control run TLS and TLT anomalies shows that most of the 20 CMIP-5 models exhibit little residual drift in globally-averaged lower strato-

¹⁰Use of the last 200 years reduces the contribution of any initial residual drift to noise estimates.

¹¹The total number of years of concatenated control run data for the TOP-5 case ranges from 2,881 (for TLT) to 3,027 (for TLS).

spheric and lower tropospheric temperature (Figs. S1 and S2). The MIROC-ESM and HadGEM2-CC models are notable exceptions. In MIROC-ESM, the TLS and TLT trends over the entire 531-year control run are 0.017 and 0.075°C/century, respectively. The TLS and TLT trends over the 240-year HadGEM2-CC control run are -0.014 and -0.071°C/century.

Since this control run drift can bias D&A results, its removal is advisable. Here, we assume that the drift behavior of synthetic MSU temperature can be well-approximated by a least-squares linear trend, and drift is removed at each grid-point. For TOP-5 noise estimates, we remove the overall linear trend from each control run prior to concatenation. In the BASE case, drift removal is performed over the last 200 control run years (since only the last 200 years are concatenated).

Figures 6 and S8 show the leading noise EOFs of $C(x, t)$ for the TOP-5 and BASE cases (respectively). In the former case, the total variance explained by the first EOF of $C(x, t)$ ranges from 12.0% for TLT to 29.6% for TLS. The corresponding explained variances for the BASE case range from 17.7% for TLT to 28.8% for TLS. The leading O3+V and BASE noise EOFs are remarkably similar. In the lower troposphere, however, use of the TOP-5 model subset yields EOF 2 and 3 noise patterns that are different from their BASE counterparts.

6.3 Estimating signal-to-noise ratios and detection time

We begin with regrided annual-mean observational data, $O(x, t)$, from RSS, UAH, and STAR. We also use regrided annual-mean data from the 11 RSS percentile realizations. Observed data are expressed as anomalies relative to climatological annual means over the entire 33-year period for which we have MSU data (1979 to 2011). The observations are projected onto $F(x)$, the time-invariant fingerprint:

$$Z_o(t) = \sum_{x=1}^{N_x} O(x, t) F(x) \quad (2)$$

where $F(x)$ is defined as described in Section 6.1. This projection is equivalent to a spatially uncentered covariance between the patterns $O(x, t)$ and $F(x)$ at time t (S23). The signal time series $Z_o(t)$ provides information on the strength of the fingerprint in the observational data. If the observed patterns of lower stratospheric temperature change are becoming increasingly similar to $F(x)$, $Z_o(t)$ should increase over time.

There are two approaches that may be used to assess the significance of the changes in $Z_o(t)$: direct comparison of test statistic values with some null distribution (S20), or comparison of trends in $Z_o(t)$ with a null distribution of trends (S24). We use the trend approach here. To assess trend significance, we require a case in which $O(x, t)$ is replaced by a record in which we know *a priori* that there is no expression of the fingerprint, except by chance. Here, we use the concatenated noise data set, $C(x, t)$, which has been regrided and detrended as described in Section 6.2. The noise time

series $N(t)$ is the projection of $C(x, t)$ onto the fingerprint:

$$N(t) = \sum_{x=1}^{N_x} C(x, t) F(x) \quad (3)$$

As in our previous work (S24), we estimate S/N ratios by fitting least-squares linear trends of increasing length L years to $Z_o(t)$, and then comparing these with the standard error of the distribution of non-overlapping L -length trends in $N(t)$. Signal detection is stipulated to occur when the trend in $Z_o(t)$ exceeds and remains above some stipulated significance level (typically either 1% or 5%). The test is one-tailed, and we assume a Gaussian distribution of trends in $N(t)$. The start date for fitting linear trends to $Z_o(t)$ is 1979, the first complete year of the observational MSU data. We use a minimum trend length of ten years, so the first S/N ratio (and the earliest possible detection time) is for 10-year trends ending in 1988. Full details of the detection method are given elsewhere (S24).

7 Further discussion/interpretation of Figures

7.1 Figure 5

7.1.1 Model bias in decadal variability

In the lower stratosphere, the multi-model average value of s_{LOW} (the temporal standard deviation of detrended, band-pass filtered atmospheric temperature data) is

virtually identical to the corresponding value for RSS v3.3 (see Fig. 5A). In the troposphere, however, the model average s_{LOW} value is between 1.55 to 1.69 times larger than the RSS v3.3 s_{LOW} values (for TMT and TLT, respectively; see Figs. 5B-D). The latter finding is qualitatively consistent with results obtained using estimates of TLT and sea-surface temperature variability from CMIP-3 models (S25).

Intriguingly, the CMIP-3 results showed that models overestimate the observed variability of band-pass filtered TLT data by a factor of 1.1 to 1.2. The much larger overestimate for CMIP-5 models may reflect true changes (from CMIP-3 to CMIP-5) in model simulations of internal variability on decadal timescales. Alternately, it may be due to the fact that a number of CMIP-5 models appear to overestimate volcanically-induced tropospheric cooling, and the percentage of models with volcanic forcing in their historical runs is much higher in CMIP-5 than in CMIP-3.

7.1.2 Model relationships between interannual and decadal variability

In the troposphere, the CMIP-5 models do not exhibit clear relationships between the amplitudes of interannual and decadal variability (Figs. 5B-D). A similar finding was obtained with CMIP-3 models, for both TLT and sea-surface temperature variability (S25). If this result also applies to observational data, it suggests that the well-quantified amplitude of observed interannual variability may not help us to constrain the more uncertain observational estimates of multi-decadal variability (or

to constrain model variability errors on these longer timescales).

In the lower stratosphere, there appears to be a strong relationship between CMIP-5 values of s_{LOW} and s_{HIGH} (Fig. 5A). However, this apparent relationship primarily reflects inter-model differences in the TLS response to volcanic forcing (see Fig. 1), and not a close scaling relationship between interannual and decadal variability.

7.2 Figure 8

7.2.1 Discussion of S/N ratios for 32-year TLT trends

In the present pattern-based S/N study, the RSS v3.3 and UAH v5.4 TLT data sets both have “model-observed” S/N ratios of *ca.* 5 for TLT changes over 1979 to 2010 (see Fig. 8C). In an earlier study (S25), model-observed S/N ratios for global-mean TLT changes over the same 32-year period (and for the same two observational data sets) were approximately 4. One important question is whether the incorporation of pattern information contributes to this 25% enhancement of S/N.

As described in the main text, both the model-predicted fingerprint in response to combined anthropogenic and natural forcing (Fig. 6) and the observed patterns of TLT change over the satellite era (Fig. 4) show large-scale warming of the lower troposphere. In contrast, the dominant modes of natural internal variability in the TOP-5 and BASE control runs (Fig. 6 and Fig. S8, respectively) have much smaller

spatial scales, and are characterized by areas of warming and cooling. Because of this difference in the spatial coherence of TLT changes in the fingerprint and leading noise modes, $F(x)$ acts as a pattern filter – *i.e.*, the spatially-coherent observed warming projects well onto $F(x)$, but the smaller-scale internal variability does not.

The current study relies on internal variability estimates from 20 different CMIP-5 models, whereas an earlier study by Santer *et al.* (2011) used internal variability information from 22 CMIP-3 models (S25). If the CMIP-5 models analyzed here had much smaller levels of multi-decadal variability, this could explain the larger TLT S/N ratios obtained in the current study. As discussed in Section 7.1.1, we find no evidence that this is the case.

In fact, the TLT variability on 5- to 20-year timescales is (on average) larger in the CMIP-5 models used here than in the CMIP-3 models analyzed by Santer *et al.* (2011). So even though the CMIP-5 noise estimates are larger (which should damp S/N ratios), S/N ratios in the current study are higher than in Santer *et al.* (2011). The most plausible explanation of this result involves the pattern filtering described above: the spatial structure of the observations, the fingerprint, and internal variability provide information that is useful in discriminating between externally-forced and internally-generated TLT changes.¹²

¹²Note that the enhancement of TLT S/N ratios (relative to those obtained in Santer *et al.*, 2011) occurs for both the BASE and TOP-5 noise estimates.

7.3 Figure 9

7.3.1 Discussion of model-model S/N results in Figure 9

The model-model S/N ratios in Fig. 9 have no direct relevance for model-versus-observed comparisons: they simply indicate whether the searched-for O3+V fingerprint is statistically identifiable in each individual model’s estimate of atmospheric temperature changes over 1979 to 2011. When the global-mean change is included, the O3+V fingerprint can be discriminated from the internal variability in the TOP-5 models in all 80 model-model comparisons.¹³

If the global-mean is removed, however, positive detection of $F(x)$ in individual models occurs in only 57 out of 80 cases. In the lower troposphere, for example, the “global-mean removed” O3+V fingerprint cannot be detected in the HadGEM2-CC and INM-CM4 models, because both models lack a prominent feature of the fingerprint (the pronounced warming of the Arctic relative to the Antarctic; see Fig. S7).

Figure 9 also shows that in the lower stratosphere, we detect the “global-mean removed” O3+V fingerprint in 4 of 20 models, but in none of the observational data sets. Models with positive detection results, like NorESM1-M, have patterns of TLS change that are similar to the fingerprint. In these four cases, positive detection indicates some pattern similarity between the sub-global features of $F(x)$ and an in-

¹³Four atmospheric layers \times 20 CMIP-5 models.

dividual model’s pattern of TLS change – it does not imply significant correspondence with observations.

7.3.2 Sensitivity of S/N results to different processing choices

We performed several sensitivity tests to explore the robustness of the S/N results presented in Fig. 9. In the first test, we computed model fingerprints and noise estimates with synthetic MSU temperatures from all 20 CMIP-5 models (the BASE case). S/N results for 33-year trends are qualitatively similar to those obtained with the O3+V fingerprints and the TOP-5 noise estimates (compare Fig. S9 and Fig. 9). There are, however, some quantitative differences.

In the “global-mean included” results, for example, S/N ratios are generally larger for the O3+V fingerprint/TOP-5 noise combination than for BASE fingerprints and BASE noise. This increase in S/N is primarily due to the exclusion from the TOP-5 of models with decadal TLS variability that is much larger than observed. The increase in S/N ratios ranges from 1.22 for TLS to 1.32 for TTT.

In the second sensitivity test, we retained the same O3+V/TOP-5 fingerprint and noise configuration, but used the period 1979 to 2011 (rather than 1861 to 2011) for estimating the fingerprint. Use of a different fingerprint estimation period has little impact on estimated TMT, TTT, and TLT fingerprints (compare the top rows of Fig. 6 and Fig. S8), and thus has little effect on S/N ratios for TMT, TTT, and TLT.

Note, however, that the TLS fingerprint pattern is sensitive to the length of record used for estimating $F(x)$. TLS S/N ratios are roughly 10% smaller when $F(x)$ is calculated over the satellite era – but in the “global-mean included” case, there is still ubiquitous detection of the O3+V TLS fingerprint in observations.

The sensitivity of the TLS fingerprint sensitive to the choice of analysis period requires some explanation. On multi-decadal timescales, the major external influences on lower stratospheric temperature are the decreases in stratospheric ozone, the volcanically-induced changes in stratospheric aerosol optical depth, and the increases in well-mixed greenhouse gases (S26, S27). Each of these external influences on TLS has a different forcing pattern and a different time history. The non-stationarity of the TLS fingerprint is caused by the complex spatio-temporal changes in these three external forcing mechanisms. As noted above, however, this non-stationarity does not hamper identification of the model-predicted TLS fingerprint in observational data.

7.4 Figure S3

The grey shaded envelopes in Figs. S3A and B provide information on the uncertainty in estimates of the TLS and TLT changes in the 12 O3+V models. Uncertainty estimates were calculated from the time series of monthly-mean, spatially averaged TLS and TLT anomalies in the spliced historical/RCP8.5 runs. After first averaging over

individual realizations (for models with multiple realizations of the historical/RCP8.5 run), we computed $s(t)$, the “between-model” standard deviation of the 12 ensemble-mean time series. The envelope is the O3+V multi-model average temperature change $\pm 2 \times s(t)$.

Supporting Material References

- S1. Mears C, Wentz FJ, Thorne P, Bernie D (2011) Assessing uncertainty in estimates of atmospheric temperature changes from MSU and AMSU using a Monte-Carlo technique. *J Geophys Res* 116, D08112, doi:10.1029/2010JD014954.
- S2. Christy JR, Norris WB, Spencer RW, Hnilo JJ (2007) Tropospheric temperature change since 1979 from tropical radiosonde and satellite measurements. *J Geophys Res* 112, D06102, doi:10.1029/2005JD006881.
- S3. Zou CZ, et al. (2006) Recalibration of microwave sounding unit for climate studies using simultaneous nadir overpasses. *J Geophys Res* 111, D19114, doi:10.1029/2005JD006798.
- S4. Karl TR, Hassol SJ, Miller CD, Murray WL (eds) (2006) *Temperature Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences*. A Report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research. National Oceanic and Atmospheric Administration, National Climatic Data Center, Asheville, NC.
- S5. Swanson RE (2003) Evidence of possible sea-ice influence on Microwave Sounding Unit tropospheric temperature trends in polar regions, *Geophys Res Lett* 30, 2040, doi:10.1029/2003GL017938.
- S6. Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the

- experiment design. *Bull Amer Meteor Soc* 93, 485-498.
- S7. Hibbard KA, Meehl GA, Cox P, Friedlingstein P (2007) A strategy for climate change stabilization experiments. *EOS* 88, 217, 219, 221.
- S8. Meinshausen M, et al. (2011) The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim Change*, doi:10.1007/s10584-011-0156-z.
- S9. Shindell DT, et al. (2012) Interactive ozone and methane chemistry in GISS-E2 historical and future simulations. *Atmos Chem Phys* (in press).
- S10. Solomon S, Young PJ, Hassler B (2012) Uncertainties in the evolution of stratospheric ozone and implications for recent temperature changes in the tropical lower stratosphere. *Geophys Res Lett* (in press).
- S11. Gao C, Robock A, Ammann C (2008) Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models. *J Geophys Res* 113, D23111, doi:10.1029/2008JD010239.
- S12. Fu Q, Johanson CM, Warren SG, Seidel DJ (2004) Contribution of stratospheric cooling to satellite-inferred tropospheric temperature trends. *Nature* 429:55-58.
- S13. Fu Q, Johanson CM (2005) Satellite-derived vertical dependence of tropical temperature trends. *Geophys Res Lett* 32, L10703, doi:10.1029/2004GL022266.
- S14. Tett S, Thorne P (2004) Tropospheric temperature series from satellites. *Nature* 432, doi:10.1038/nature03208.

- S15. Gillett NP, Santer BD, Weaver AJ (2004) Stratospheric cooling and the troposphere. *Nature* 432, doi:10.1038/nature03209.
- S16. Kiehl JT, Caron JM, Hack JJ (2005) On using global climate model simulations to assess the accuracy of MSU retrieval methods for tropospheric warming trends. *J Clim* 18:2533-2539.
- S17. Taylor KE (1996) EzGet: A library of Fortran subroutines to facilitate data retrieval. *Program for Climate Model Diagnosis and Intercomparison*, Report No. 34, Lawrence Livermore National Laboratory, Livermore, CA.
- S18. Hasselmann K (1979) On the signal-to-noise problem in atmospheric response studies. In *Meteorology of Tropical Oceans*, Shaw DB (ed), pp. 251-259, Roy Met Soc London, U.K.
- S19. Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1986) *Numerical Recipes*. Cambridge University Press, New York, 818 pp.
- S20. Wigley TML, Jaumann PJ, Santer BD, Taylor KE (1998) Relative detectability of greenhouse-gas and aerosol climate change signals. *Cli Dyn* 14:781-790.
- S21. Stott PA, Tett SFB, Jones GS, Allen MR, Mitchell JFB, Jenkins GJ (2000) External control of 20th century temperature by natural and anthropogenic forcings. *Science* 290:2133-2137.
- S22. Tett SFB, et al. (2002) Estimation of natural and anthropogenic contribu-

- tions to twentieth century temperature change. *J Geophys Res* 107, D16, doi:10.1029/2000JD000028.
- S23. Preisendorfer RW (1988) *Principal Component Analysis in Meteorology and Oceanography*. Elsevier, Amsterdam, 425 pp.
- S24. Santer BD, Mikolajewicz U, Brüggemann W, Cubasch U, Hasselmann K, Höck H, Maier-Reimer E, Wigley TML (1995) Ocean variability and its influence on the detectability of greenhouse warming signals. *J Geophys Res* 100:10693-10725.
- S25. Santer BD, et al. (2011) Separating signal and noise in atmospheric temperature changes: The importance of timescale. *J Geophys Res* 116, D22105, doi:10.1029/2011JD016263.
- S26. Ramaswamy V, et al. (2001) Stratospheric temperature trends: Observations and model simulations. *Rev Geophys* 39:71-122.
- S27. Santer BD, et al. (2003b) Contributions of anthropogenic and natural forcing to recent tropopause height changes. *Science* 301:479-483.

Figure S1: Time series of simulated monthly-mean, near-global anomalies in the temperature of the lower stratosphere (TLS). Results are from pre-industrial control simulations performed with 20 different CMIP-5 models. Anomalies were averaged over 82.5°N - 82.5°S (the latitudinal extent of RSS TLS data), and are defined with respect to climatological monthly means over the entire control run. The y -axis range is identical in each panel. To facilitate visual comparison of results, the nominal start date for each control run is assumed to be 1850.

Figure S2: As for Fig. S1, but for time series of monthly-mean, near-global anomalies in the temperature of the lower troposphere (TLT). Model anomalies are averaged over 82.5°N - 70°S (the latitudinal extent of RSS TLT data).

Figure S3: Comparison of observed near-global changes in TLS (panel A) and TLT (panel B) with CMIP-5 multi-model average TLS and TLT changes. The comparison period is January 1979 to December 2011. Satellite-based estimates of changes in TLS and TLT are from RSS, UAH, and STAR. In addition to version 3.3 of the RSS temperature data, 11 RSS “percentile realizations” are also shown. These provide information on the sensitivity of RSS estimates of atmospheric temperature change to data set construction uncertainties. The CMIP-5 multi-model average (MMA) TLS and TLT changes were computed in two ways: 1) using results from all 20 models analyzed here (BASE); and 2) using a subset of 12 NOCHEM models with prescribed historical changes in stratospheric ozone, and with more realistic treatment of volcanic

forcing (O3+V). The grey envelope is the ± 2 standard deviation uncertainty of the O3+V MMA temperature change, calculated as described in the *SI*. For information on anomaly definition and spatial averaging, refer to Figs. 1 and 2 of the main text. Because the CMIP-5 MMA results were averaged over different realizations of the spliced historical/RCP8.5 run (where multiple realizations were available) and over different models, they are much smoother than the single realization of the observations. The apparent oscillations in the MMA TLS anomalies are a residual seasonal cycle effect, which arises from the impact of the El Chichón and Pinatubo eruptions on climatological monthly means.

Figure S4: Zonal-mean trends in observed and synthetic TLS, TMT, and TLT (panels A-C) over 1979 to 2011. Observational results are the 5-95 percentile ranges of the full 400-member RSS observational ensembles (see *SI*). Model synthetic MSU temperatures are from three selected CMIP-5 models (CCSM4, HadGEM2-CC, and MPI-ESM-LR). Each of these models has multiple realizations of the spliced historical/RCP8.5 simulation. Note the large between-realization variability in zonal-mean trends, particularly at high latitudes in the lower stratosphere. This variability hampers reliable estimation of the true response to the imposed external forcing changes.

Figure S5: Zonal-mean trends in the observed and simulated temperature of the total troposphere (TTT) over 1979 to 2011. TTT is a linear combination of MSU TLS and TMT data, which reduces the substantial contribution TMT receives from

the cooling of the lower stratosphere (see Section 3 of the *SI*). The regression coefficients used in calculating TTT do not vary with latitude. Observational TTT results are from UAH, STAR, version 3.3 of the RSS data set, and the 11 RSS percentile realizations. The 5 to 95 percentile range of the RSS results was computed as described in Section 4 of the *SI*. Model synthetic MSU temperatures are from the spliced historical/RCP8.5 simulations performed with 20 different CMIP-5 models. Only the first realization is shown for each model. The CMIP-5 multi-model average is the O3+V case. The O3+V multi-model average trends are also plotted for TTT*, which is an alternative version of bulk tropospheric temperature calculated with latitudinally-varying regression coefficients (see Section 3 of the *SI*).

Figure S6: Geographical patterns of trends in lower stratospheric temperature (TLS) over 1979 to 2011 in 20 individual CMIP-5 models. Model trends (in °C/decade) are from spliced historical/RCP8.5 runs. For models with multiple historical/RCP8.5 realizations (see Table S4), results are shown for the first realization only.

Figure S7: As for Fig. S6, but for trends in TLT.

Figure S8: Leading signal and noise modes from the pattern-based signal-to-noise (S/N) analysis, together with the percentage variance explained by each mode. Results are for TLS, TMT, TTT, and TLT (columns 1-4, respectively). The O3+V fingerprints (row 1) were calculated using multi-model average synthetic MSU temperature changes over 1979-2011. The leading noise modes (rows 2-4) are from the

BASE case, and were estimated from the control simulations performed with 20 different CMIP-5 models. Only the last 200 years of each control run were used for noise estimation. Full details of fingerprint and noise mode calculations are given in Sections 6.1 and 6.2 of the *SI*.

Figure S9: Sensitivity of S/N ratios to uncertainties in fingerprint and noise estimates. Results are for TLS (panel A), TTT (panel B), TMT (panel C), and TLT (panel D). S/N ratios are directly comparable to the results shown in Fig. 9 in the main text, which were based on use of the O3+V fingerprint, and relied on noise estimates from the TOP-5 models. Here, the searched-for fingerprint and internally generated noise were estimated with all 20 CMIP-5 models (the BASE case). As in Fig. 9, the fingerprint was computed using multi-model average atmospheric temperature changes over 1861 to 2011. For further details, refer to Fig. 9 and Section 6.3 of the *SI*.

Unforced Lower Stratospheric Temperature Changes in 20 CMIP-5 Models Synthetic MSU temperatures (TLS). Pre-industrial control runs

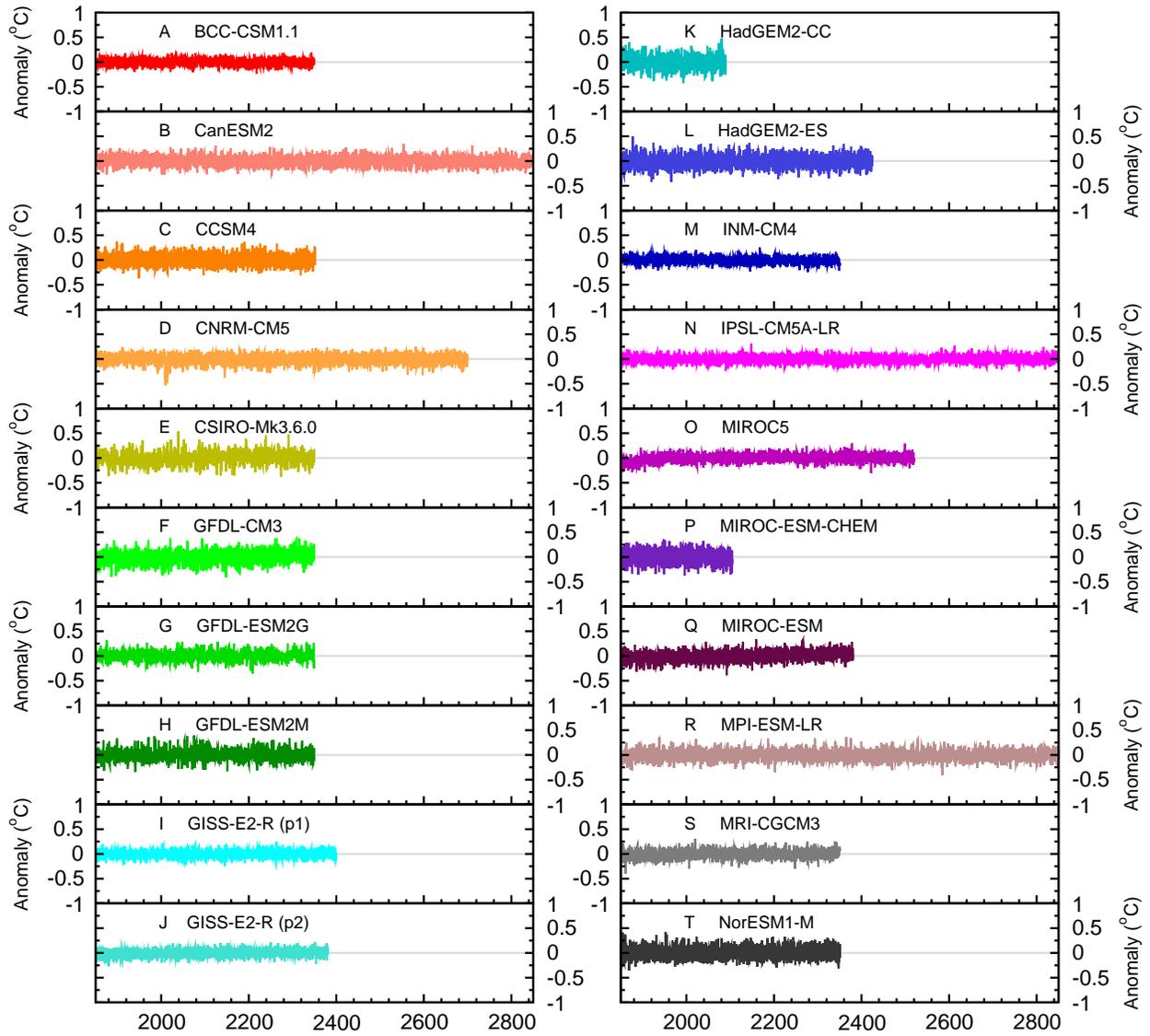


Figure S1: Santer et al.

Unforced Lower Tropospheric Temperature Changes in 20 CMIP-5 Models Synthetic MSU temperatures (TLT). Pre-industrial control runs

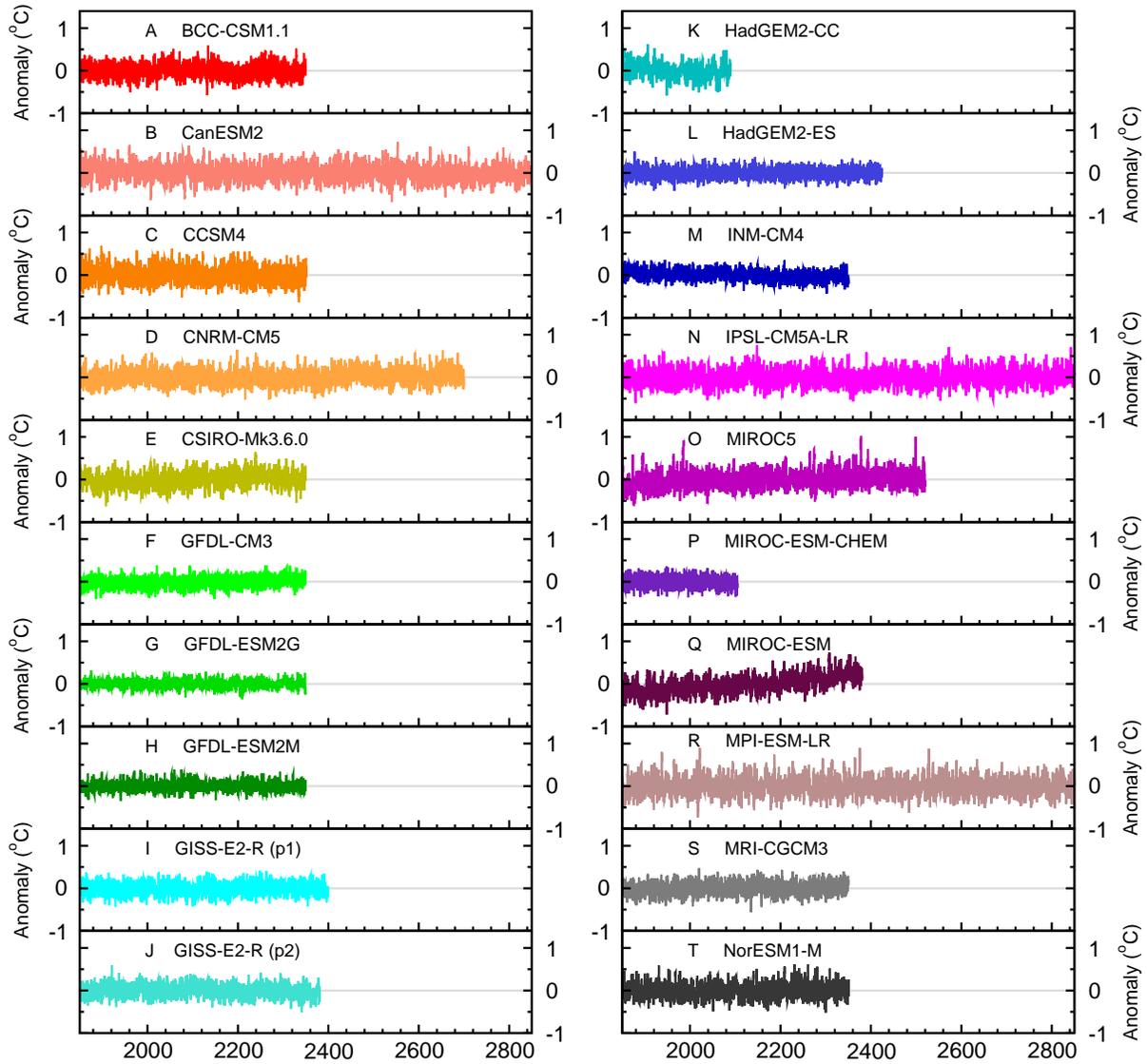


Figure S2: Santer *et al.*

Atmospheric Temperature Changes in Models and Observations

Spatial averages over 82.5°N-82.5°S (TLS) and 82.5°N-70°S (TLT)

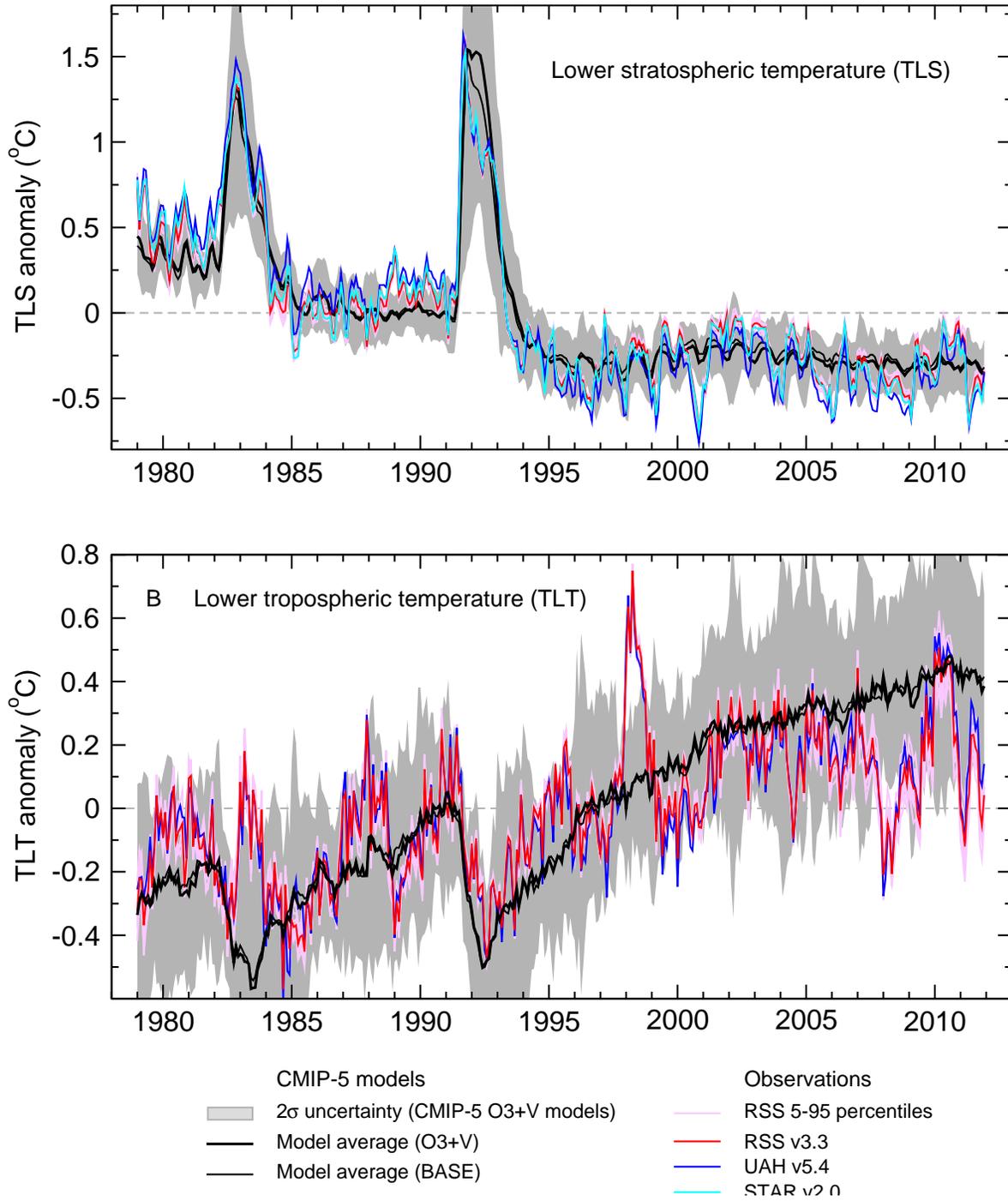


Figure S3: Santer et al.

Trends in Zonal-Mean Actual and Synthetic MSU Temperature

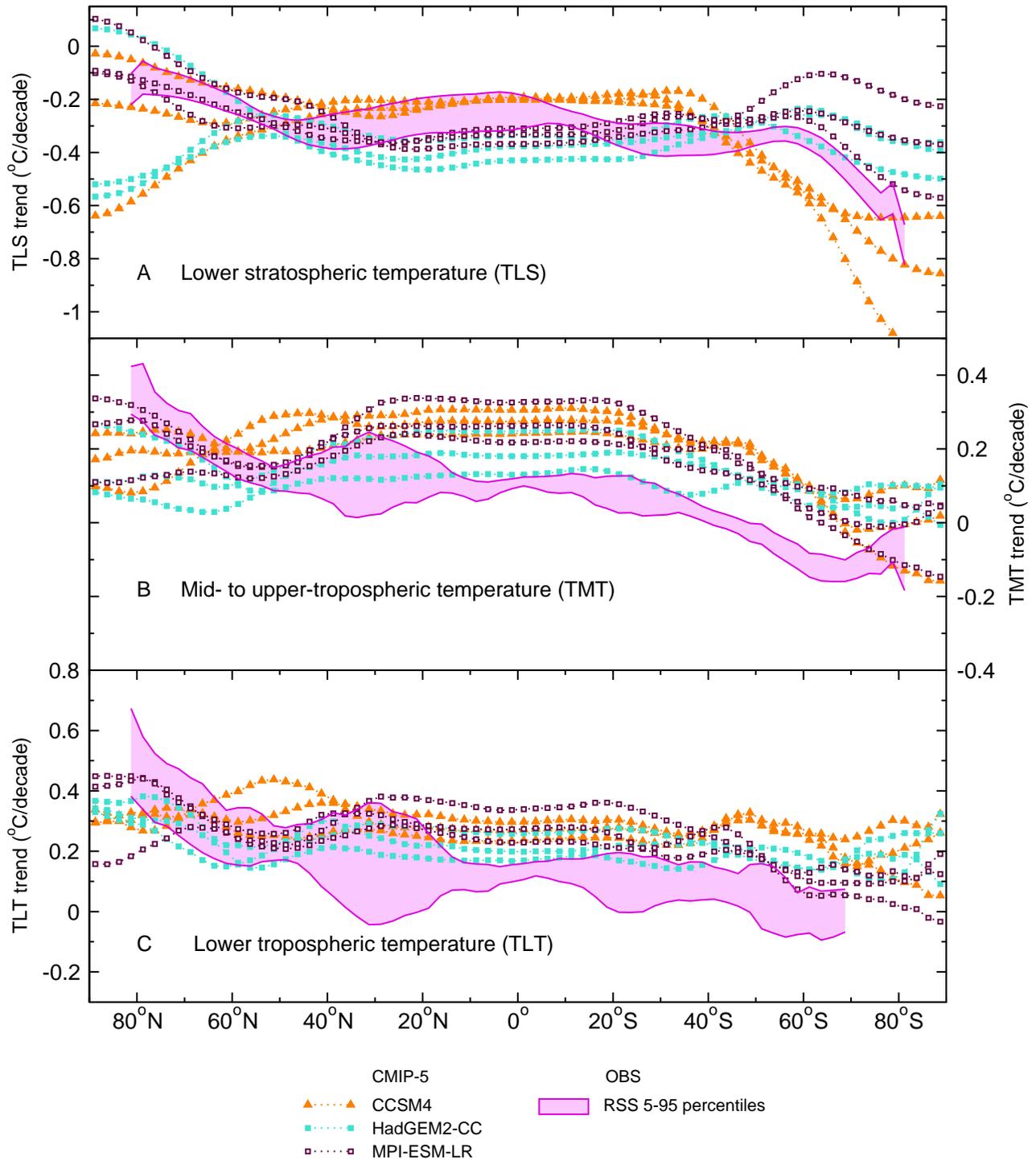


Figure S4: Santer *et al.*

Trends in Zonal-Mean Actual and Synthetic MSU Temperature

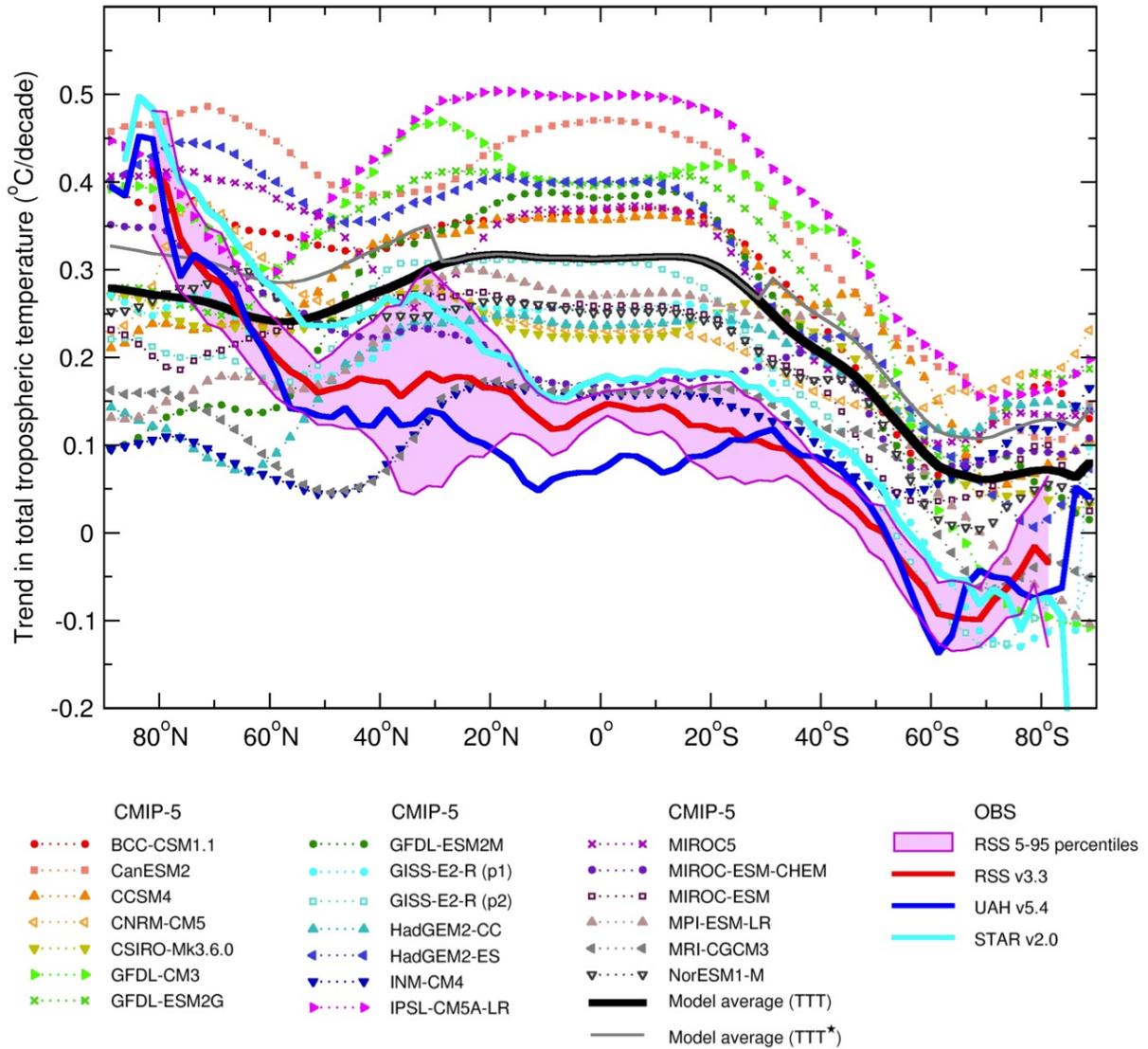


Figure S5: Santer et al.

Lower Stratospheric Temperature Trends (1979 to 2011) in 20 CMIP-5 Models

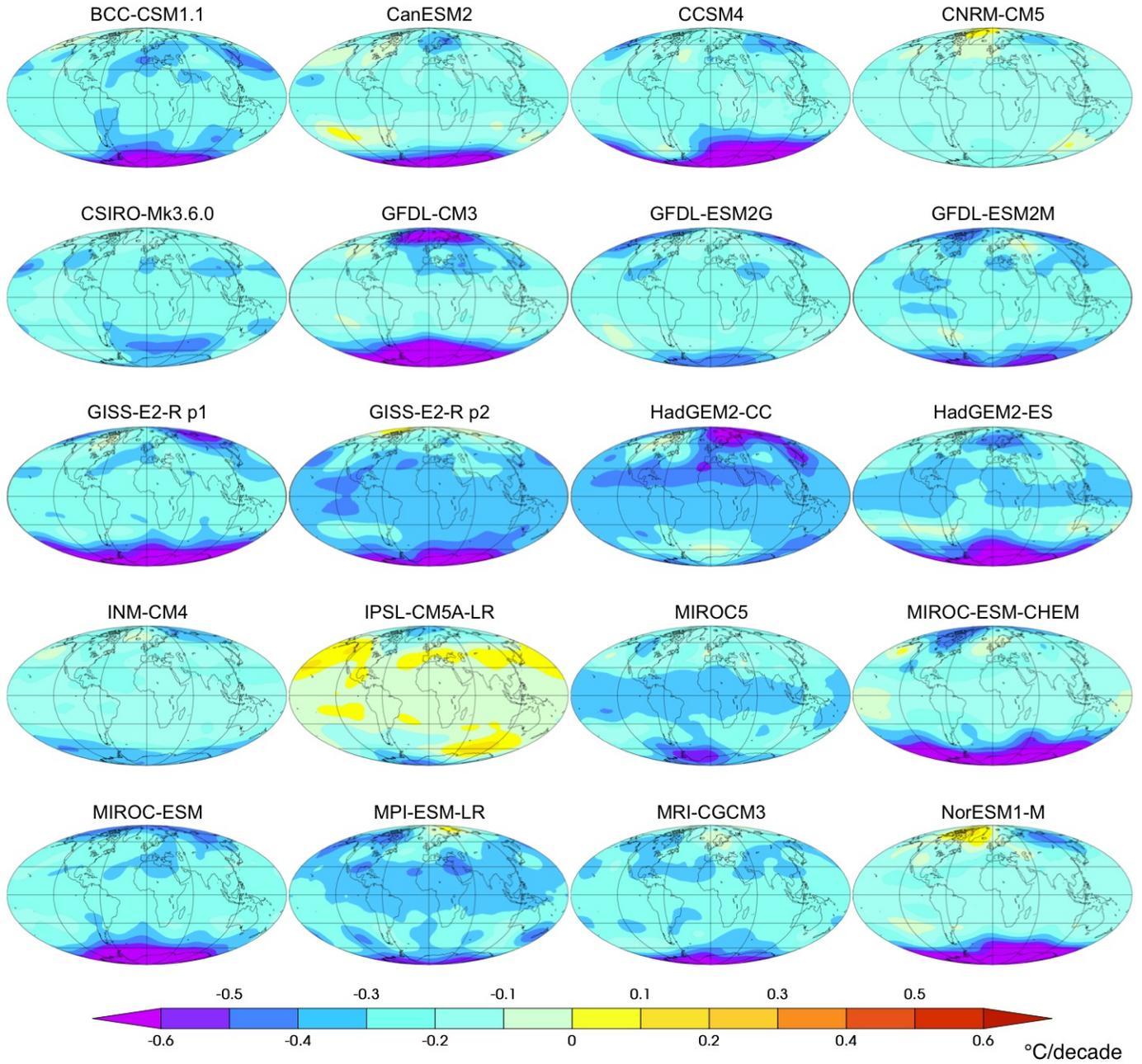


Figure S6: Santer *et al.*

Lower Tropospheric Temperature Trends (1979 to 2011) in 20 CMIP-5 Models

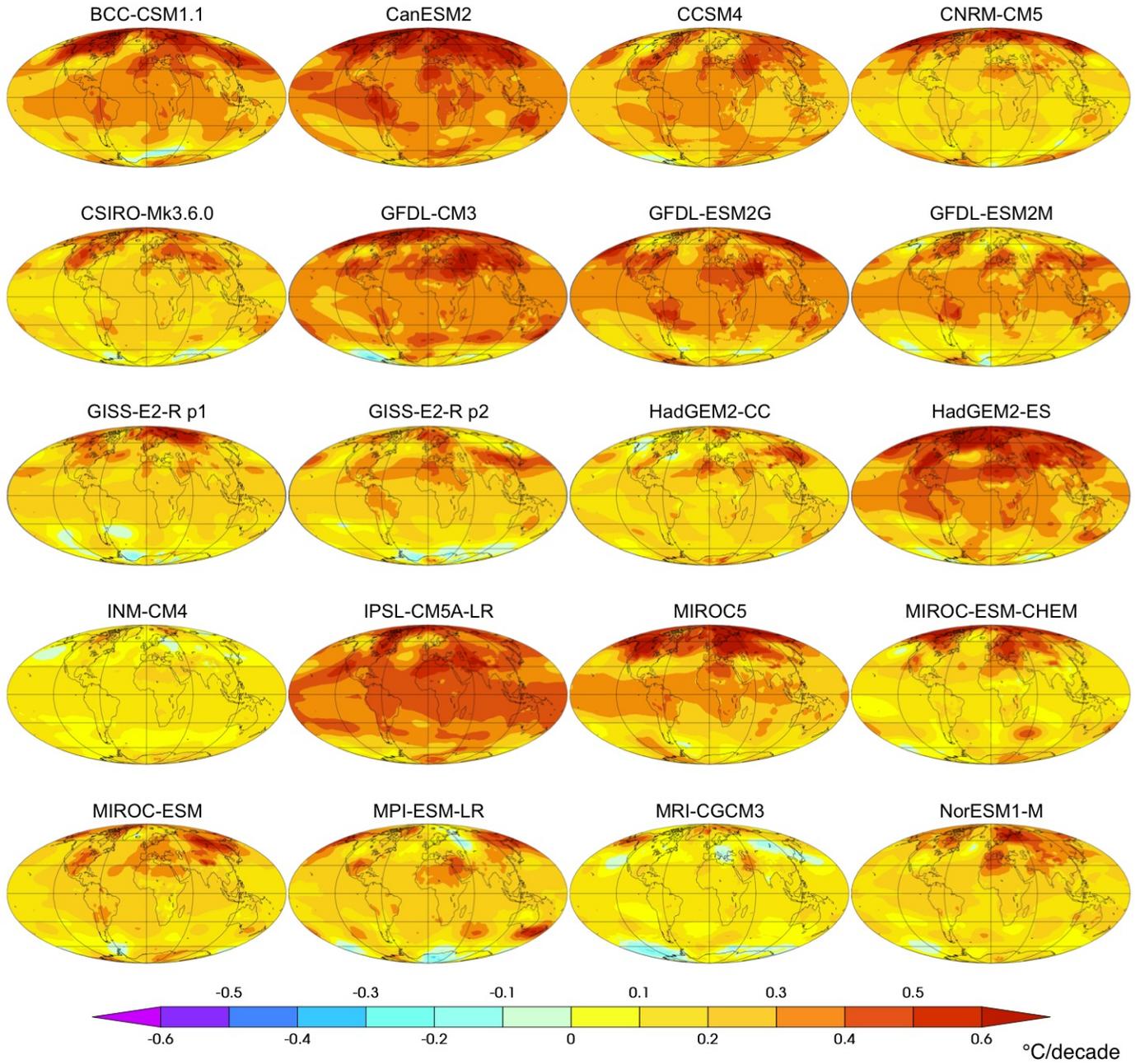


Figure S7: Santer et al.

Fingerprints and Leading Noise Modes in CMIP-5 Models

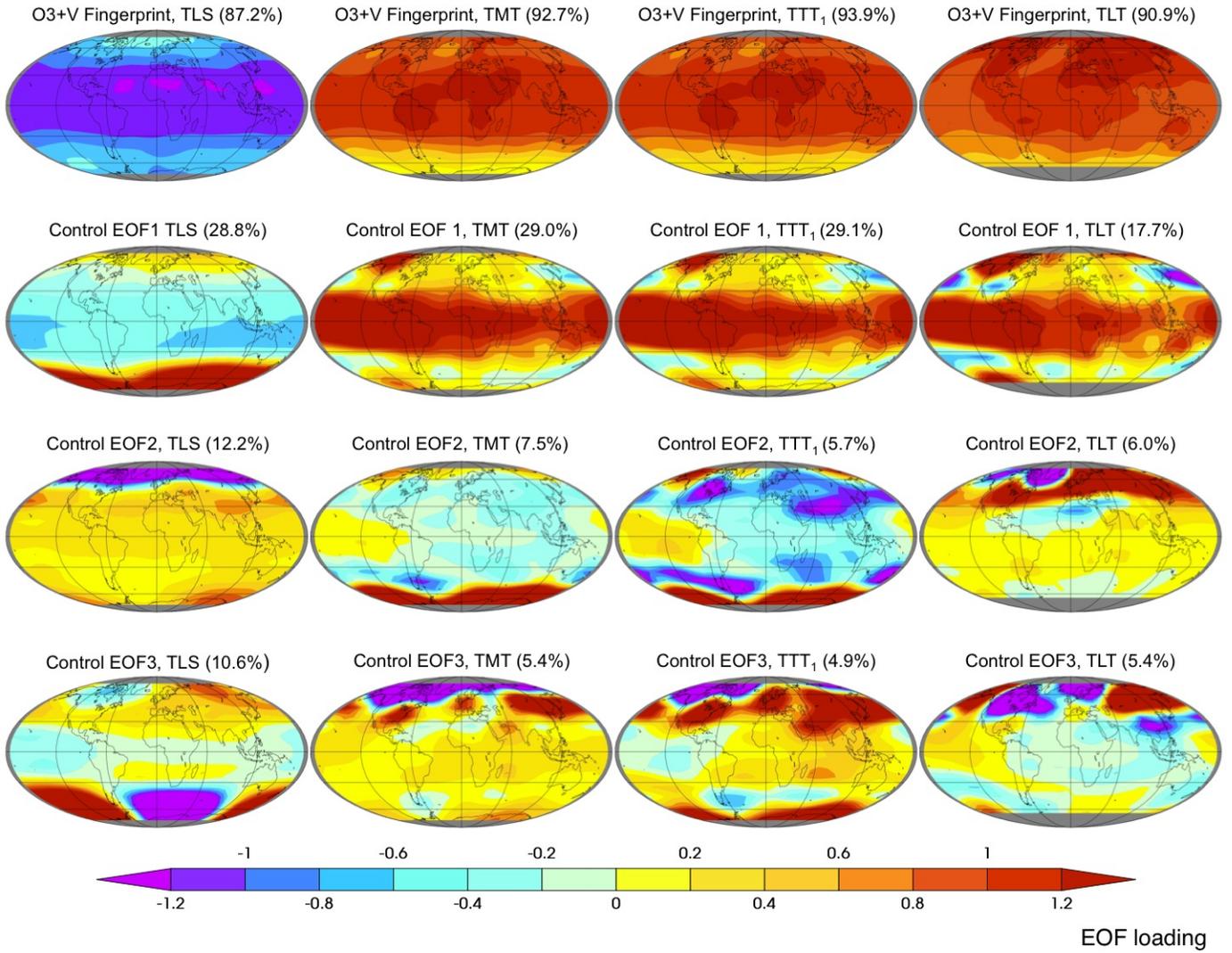


Figure S8: Santer et al.

Signal-to-Noise Ratios for Temperature Changes Over 1979 to 2011

Fingerprints (BASE case) estimated over 1861 to 2011. Noise estimates: BASE models

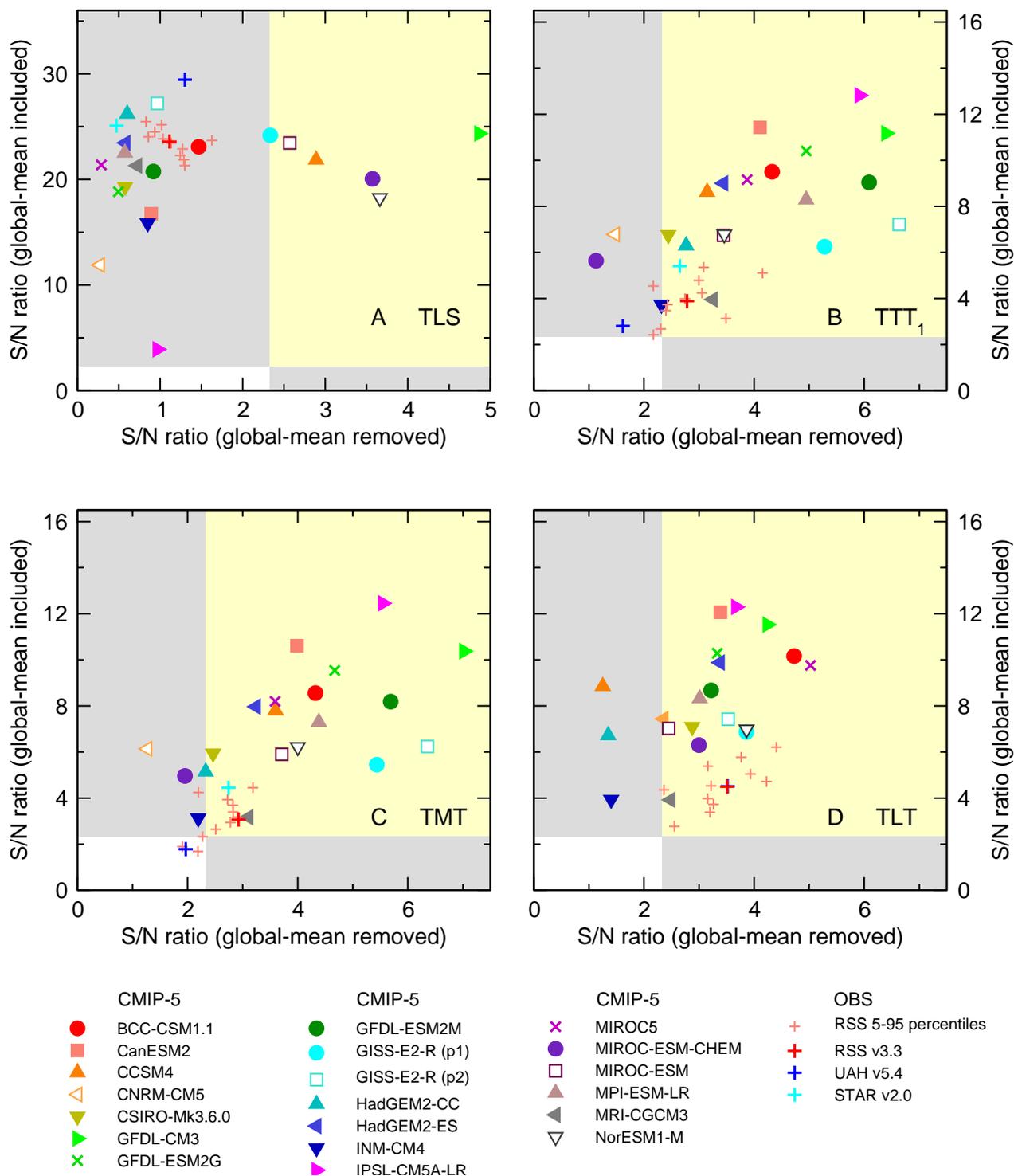


Figure S9: Santer et al.

Table S1: Modeling center information and official acronyms of the CMIP-5 models used in this study.

	Model	Country	Modeling center
1	BCC-CSM1.1	China	Beijing Climate Center, China Meteorological Administration, Climate System Model version 1.1
2	CanESM2	Canada	Canadian Centre for Climate Modelling and Analysis, Earth System Model version 2
3	CCSM4	USA	National Center for Atmospheric Research, Community Climate System Model version 4
4	CNRM-CM5	France	Centre National de Recherches Meteorologiques / Centre Européen de Recherche et Formation Avancées en Calcul Scientifique, Climate Model version 5
5	CSIRO-Mk3.6.0	Australia	Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence, model version Mark3.6.0
6	GFDL-CM3	USA	NOAA Geophysical Fluid Dynamics Laboratory, Climate Model version 3
7	GFDL-ESM2G	USA	NOAA Geophysical Fluid Dynamics Laboratory, Earth System Model version G
8	GFDL-ESM2M	USA	NOAA Geophysical Fluid Dynamics Laboratory, Earth System Model version M
9	GISS-E2-R (p1)	USA	NASA Goddard Institute for Space Studies, atmospheric ModelE with Russell ocean, physics version 1
10	GISS-E2-R (p2)	USA	NASA Goddard Institute for Space Studies, atmospheric ModelE with Russell ocean, physics version 2
11	HadGEM2-CC	UK	Met. Office Hadley Centre, Global Environment Model version 2, Carbon Cycle configuration
12	HadGEM2-ES	UK	Met. Office Hadley Centre, Global Environment Model version 2, Earth System configuration
13	INM-CM4	Russia	Institute for Numerical Mathematics, Climate Model version 4
14	IPSL-CM5A-LR	France	Institut Pierre-Simon Laplace, Climate Model version 5A, low resolution configuration

Table S1: Modeling center information and official acronyms of the CMIP-5 models used in this study (continued).

Model	Country	Modeling center
15 MIROC5	Japan	Atmosphere and Ocean Research Institute (the University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, version 5
16 MIROC-ESM-CHEM	Japan	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (the University of Tokyo), and National Institute for Environmental Studies, Earth System Model, configuration with atmospheric chemistry
17 MIROC-ESM	Japan	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (the University of Tokyo), and National Institute for Environmental Studies, Earth System Model
18 MPI-ESM-LR	Germany	Max Planck Institute for Meteorology, Earth System Model, low resolution configuration
19 MRI-CGCM3	Japan	Meteorological Research Institute, Coupled General Circulation Model, version 3
20 NorESM1-M	Norway	Norwegian Climate Centre, Earth System Model version 1, medium resolution configuration

Table S2: Information on horizontal and vertical resolution of the CMIP-5 models used in this study.

Model	Resolution (lat. \times long.)	No. levels (atmosphere) ¹	Top level (hPa)	No. levels above 100 hPa ²
1 BCC-CSM1.1	64 \times 128	17	10	5
2 CanESM2	64 \times 128	22	1	10
3 CCSM4	192 \times 288	17	10	5
4 CNRM-CM5	128 \times 256	17	10	5
5 CSIRO-Mk3.6.0	96 \times 192	18	5	6
6 GFDL-CM3	90 \times 144	23	1	10
7 GFDL-ESM2-G	90 \times 144	17	10	5
8 GFDL-ESM2-M	90 \times 144	17	10	5
9 GISS-E2-R (p1)	90 \times 144	21	0.4	9
10 GISS-E2-R (p2)	90 \times 144	17	10	5
11 HadGEM2-CC	144 \times 192	23	0.4	11
12 HadGEM2-ES	144 \times 192	17	10	5
13 INM-CM4	120 \times 180	17	10	5
14 IPSL-CM5A-LR	96 \times 96	17	10	5
15 MIROC5	128 \times 256	17	10	5
16 MIROC-ESM-CHEM	64 \times 128	35	0.03	20
17 MIROC-ESM	64 \times 128	35	0.03	20
18 MPI-ESM-LR	96 \times 192	25	0.1	13
19 MRI-CGCM3	160 \times 320	23	0.4	11
20 NorESM1-M	96 \times 144	17	10	5

¹This represents the total number of model levels at which atmospheric temperature information is archived. It is typically smaller than the vertical resolution of the atmospheric model itself. Similarly, the top level in the fourth column is the highest pressure level at which atmospheric temperature data is archived – it is not necessarily identical to the pressure of the model lid.

²Not including the 100 hPa level.

Table S3: Information on the external forcings which were included in the historical simulations of the CMIP-5 models used in this study. Information was extracted from the global attribute named “forcing” in the metadata of the relevant NetCDF files.¹

In the third column, models are stratified according to their treatment of ozone.

Model	Forcing information from metadata	Ozone
1 BCC-CSM1.1	Nat, Ant, GHG, SD, Oz, Sl, Vl, SS, Ds, BC, OC	NOCHEM ²
2 CanESM2	GHG, Oz, SA, BC, OC, LU, Sl, Vl	NOCHEM
3 CCSM4	Sl, GHG, Vl, SS, Ds, SD, BC, MD, OC, Oz, AA, LU	CHEM, S-OFF ³
4 CNRM-CM5	GHG, SA, Sl, Vl, BC, OC ⁴	CHEM, INT ⁵
5 CSIRO-Mk3.6.0	Ant, Nat	NOCHEM
6 GFDL-CM3	GHG, SA, Oz, LU, Sl, Vl, SS, BC, MD, OC ⁶	CHEM, INT
7 GFDL-ESM2G	GHG, SD, Oz, LU, Sl, Vl, SS, BC, MD, OC ⁷	NOCHEM
8 GFDL-ESM2M	GHG, SD, Oz, LU, Sl, Vl, SS, BC, MD, OC ⁷	NOCHEM
9 GISS-E2-R (p1)	GHG, LU, Sl, Vl, BC, OC, SA, Oz ⁸	NOCHEM
10 GISS-E2-R (p2)	GHG, LU, Sl, Vl, BC, OC, SA, Oz ⁸	CHEM, INT
11 HadGEM2-CC	GHG, Oz, SA, LU, Sl, Vl, BC, OC	NOCHEM
12 HadGEM2-ES	GHG, SA, Oz, LU, Sl, Vl, BC, OC ⁹	NOCHEM
13 INM-CM4	GHG, Oz, Sl, SA, Vl	NOCHEM
14 IPSL-CM5A-LR	Nat, Ant, GHG, SA, Oz, LU, SS, Ds, BC, MD, OC, AA	CHEM, S-OFF
15 MIROC5	GHG, SA, Oz, LU, Sl, Vl, SS, Ds, BC, MD, OC ¹⁰	NOCHEM
16 MIROC-ESM-CHEM	GHG, SA, Oz, LU, Sl, Vl, MD, BC, OC	CHEM, INT
17 MIROC-ESM	GHG, SA, Oz, LU, Sl, Vl, MD, BC, OC	NOCHEM
18 MPI-ESM-LR	GHG, Oz, SD, Sl, Vl, LU	NOCHEM
19 MRI-CGCM3	GHG, SA, Oz, LU, Sl, Vl, BC, OC ¹¹	NOCHEM
20 NorESM1-M	GHG, SA, Oz, Sl, Vl, BC, OC	CHEM, S-OFF

Table S3: Information on the external forcings which were included in the historical simulations of the CMIP-5 models used in this study (continued).

¹Forcing abbreviations are described in Appendix 1.2 of the CMIP-5 Data Reference Syntax document. Nat = natural forcing (a combination, not explicitly defined); Ant = anthropogenic forcing (a combination, not explicitly defined); GHG = well-mixed greenhouse gases; SD = anthropogenic sulfate aerosol (direct effects only); SI = anthropogenic sulfate aerosol (indirect effects only); SA = anthropogenic sulfate aerosol direct and indirect effects; Oz = tropospheric and stratospheric ozone; LU = land-use change; Sl = solar irradiance; Vl = volcanic aerosol; SS = sea salt; Ds = dust; BC = black carbon; MD = mineral dust; OC = organic carbon; AA = anthropogenic aerosols (a mixture of aerosols, not explicitly defined).

²NOCHEM = Model with prescribed changes in ozone.

³CHEM, S-OFF = Model with semi-offline ozone chemistry.

⁴Although stratospheric ozone forcing is not listed in the forcing metadata, chlorine concentration is an input for the prognostic ozone scheme of the CNRM-CM5 model. So according to the CMIP-5 Data Reference Syntax document, stratospheric ozone should have been listed as an external forcing.

⁵CHEM, INT = Model with interactive ozone chemistry.

⁶GHG includes CO₂, CH₄, N₂O, CFC11, CFC12, HCFC22, and CFC113. Aerosol direct and indirect effects are included.

⁷GHG includes CO₂, CH₄, N₂O, CFC11, CFC12, HCFC22, and CFC113. “The direct effect of tropospheric aerosols is calculated by the model, but not the indirect effects.”

⁸Also includes orbital change, BC on snow, and nitrate aerosols.

⁹GHG = CO₂, N₂O, CH₄, CFCs.

¹⁰GHG includes CO₂, N₂O, CH₄, and CFCs; Oz includes OH and H₂O₂; LU excludes change in lake fraction.

¹¹GHG includes CO₂, CH₄, N₂O, CFC-11, CFC-12, and HCFC-22.

Table S4: Basic information relating to the start dates, end dates, and lengths of the CMIP-5 historical and RCP8.5 simulations used in this study. EM is the “ensemble member” identifier described in the CMIP-5 Data Reference Syntax document.¹

	Model	EM	Hist. Start	Hist. End	Hist. (months)	RCP8.5 Start	RCP8.5 End	RCP8.5 (months)
1	BCC-CSM1.1	r1i1p1	1850-01	2012-12	1956	2006-01	2300-12	3540
2	CanESM2	r1i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140
3	CanESM2	r2i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140
4	CanESM2	r3i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140
5	CanESM2	r4i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140
6	CanESM2	r5i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140
7	CCSM4	r1i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140
8	CCSM4	r2i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140
9	CCSM4	r3i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140
10	CNRM-CM5	r1i1p1	1850-01	2005-12	1872	2006-01	2300-12	3540
11	CNRM-CM5	r2i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140
12	CNRM-CM5	r4i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140
13	CSIRO-Mk3.6.0	r10i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140
14	GFDL-CM3	r1i1p1	1860-01	2005-12	1752	2006-01	2100-12	1140
15	GFDL-ESM2G	r1i1p1	1861-01	2005-12	1740	2006-01	2100-12	1140
16	GFDL-ESM2M	r1i1p1	1861-01	2005-12	1740	2006-01	2100-12	1140
17	GISS-E2-R (p1)	r1i1p1	1850-01	2005-12	1872	2006-01	2300-12	3540
18	GISS-E2-R (p2)	r1i1p2	1850-01	2005-12	1872	2006-01	2300-12	3540
19	HadGEM2-CC	r1i1p1	1859-12	2005-11	1752	2005-12	2099-12	1129
20	HadGEM2-CC	r2i1p1	1959-12	2005-12	553	2005-12	2099-12	1129
21	HadGEM2-CC	r3i1p1	1959-12	2005-12	553	2005-12	2099-12	1129
22	HadGEM2-ES	r1i1p1	1859-12	2005-11	1752	2005-12	2299-12	3529
23	HadGEM2-ES	r2i1p1	1859-12	2005-12	1753	2005-12	2100-11	1140
24	INM-CM4	r1i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140
25	IPSL-CM5A-LR	r2i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140

Table S4: Basic information relating to the start dates, end dates, and lengths of the CMIP-5 historical and RCP8.5 simulations used in this study (continued).

Model	EM	Hist. Start	Hist. End	Hist. (months)	RCP8.5 Start	RCP8.5 End	RCP8.5 (months)
26 MIROC5	r1i1p1	1850-01	2012-12	1956	2006-01	2100-12	1140
27 MIROC-ESM-CHEM	r1i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140
28 MIROC-ESM	r1i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140
29 MPI-ESM-LR	r1i1p1	1850-01	2005-12	1872	2006-01	2300-12	3540
30 MPI-ESM-LR	r2i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140
31 MPI-ESM-LR	r3i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140
32 MRI-CGCM3	r1i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140
33 NorESM1-M	r1i1p1	1850-01	2005-12	1872	2006-01	2100-12	1140

¹See <http://cmip-pcmdi.llnl.gov/cmip5/documents.html> for further details.

Table S5: Basic information relating to the start dates, end dates, and lengths of the CMIP-5 pre-industrial control runs used in this study. EM is the “ensemble member” identifier described in the CMIP-5 Data Reference Syntax document.¹

	Model	EM	Control Start	Control End	Control (months)
1	BCC-CSM1.1	r1i1p1	1800-01	2799-12	6000
2	CanESM2	r1i1p1	2015-01	3010-12	11952
3	CCSM4	r1i1p1	800-01	1300-12	6012
4	CNRM-CM5	r1i1p1	1850-01	2699-12	10200
5	CSIRO-Mk3.6.0	r1i1p1	1651-01	2150-12	6000
6	GFDL-CM3	r1i1p1	1-01	500-12	6000
7	GFDL-ESM2G	r1i1p1	1-01	500-12	6000
8	GFDL-ESM2M	r1i1p1	1-01	500-12	6000
9	GISS-E2-R (p1)	r1i1p1	3981-01	4530-12	6600
10	GISS-E2-R (p2)	r1i1p2	3590-01	4120-12	6372
11	HadGEM2-CC	r1i1p1	1859-12	2099-12	2881
12	HadGEM2-ES	r1i1p1	1859-12	2435-11	6912
13	INM-CM4	r1i1p1	1850-01	2349-12	6000
14	IPSL-CM5A-LR	r1i1p1	1800-01	2799-12	12000
15	MIROC5	r1i1p1	2000-01	2669-12	8040
16	MIROC-ESM-CHEM	r1i1p1	1846-01	2100-12	3060
17	MIROC-ESM	r1i1p1	1800-01	2330-12	6372
18	MPI-ESM-LR	r1i1p1	1850-01	2849-12	12000
19	MRI-CGCM3	r1i1p1	1851-01	2350-12	6000
20	NorESM1-M	r1i1p1	700-01	1200-12	6012

¹See <http://cmip-pcmdi.llnl.gov/cmip5/documents.html> for further details.

Table S6: Ensemble-mean trends in near-global atmospheric temperature. Trends were computed over the period January 1979 to December 2011, using information from spliced historical/RCP8.5 simulations performed with the 20 CMIP-5 models used in this study. All trends are in $^{\circ}\text{C}/\text{decade}$. R# is the number of realizations used in calculating each model's ensemble-mean trend. All statistics (final 5 rows) are for the BASE case.

	Model	R#	TLS trend	TMT trend	TLT trend	TTT trend
1	BCC-CSM1.1	1	-0.295	0.254	0.321	0.309
2	CanESM2	5	-0.215	0.322	0.386	0.376
3	CCSM4	3	-0.271	0.233	0.285	0.283
4	CNRM-CM5	3	-0.160	0.193	0.237	0.228
5	CSIRO-Mk3.6.0	1	-0.253	0.179	0.225	0.222
6	GFDL-CM3	1	-0.290	0.297	0.365	0.356
7	GFDL-ESM2G	1	-0.248	0.284	0.327	0.337
8	GFDL-ESM2M	1	-0.272	0.233	0.274	0.284
9	GISS-E2-R (p1)	2	-0.305	0.146	0.215	0.191
10	GISS-E2-R (p2)	2	-0.351	0.167	0.234	0.218
11	HadGEM2-CC	3	-0.354	0.154	0.216	0.205
12	HadGEM2-ES	2	-0.305	0.241	0.314	0.296
13	INM-CM4	1	-0.207	0.089	0.124	0.119
14	IPSL-CM5A-LR	1	-0.050	0.374	0.392	0.416
15	MIROC5	1	-0.280	0.247	0.307	0.299
16	MIROC-ESM-CHEM	1	-0.245	0.151	0.199	0.190
17	MIROC-ESM	1	-0.295	0.171	0.223	0.217
18	MPI-ESM-LR	3	-0.301	0.212	0.264	0.264
19	MRI-CGCM3	1	-0.278	0.086	0.121	0.122
20	NorESM1-M	1	-0.212	0.179	0.218	0.218
	Minimum	-	-0.354	0.086	0.121	0.119
	Maximum	-	-0.050	0.374	0.392	0.416
	Mean	-	-0.259	0.211	0.262	0.257
	Median	-	-0.275	0.202	0.250	0.246
	Std. deviation	-	0.068	0.074	0.076	0.079

Table S7: Observed trends in near-global atmospheric temperature over January 1979 to December 2011. All trends are in °C/decade. The RSS “percentile realizations” were not used in calculating the mean observed trends. To facilitate comparison with model results, the mean CMIP-5 trends for the 12 O3+V and 20 BASE models are also shown. The absolute value of the ratio between simulated and observed mean trends is given in rows 9 and 10 of the table (for the O3+V and BASE cases, respectively). Note that the STAR group does not produce a TLT data set.

Model	TLS trend	TMT trend	TLT trend	TTT trend
1 RSS 5 th percentile	-0.329	0.043	0.086	0.075
2 RSS 95 th percentile	-0.272	0.127	0.190	0.171
3 RSS v3.3	-0.302	0.083	0.139	0.122
4 STAR v2.0	-0.324	0.127	-	0.172
5 UAH v5.4	-0.381	0.047	0.140	0.090
6 Mean (OBS)	-0.336	0.086	0.140	0.128
7 Mean (CMIP-5; O3+V)	-0.283	0.211	0.266	0.260
8 Mean (CMIP-5; BASE)	-0.259	0.211	0.262	0.257
9 O3+V/OBS (row 7/row 6)	0.842	2.453	1.900	2.031
10 BASE/OBS (row 8/row 6)	0.771	2.453	1.871	2.008