MIT Libraries | DSpace@MIT

MIT Open Access Articles

Comparing Tropospheric Warming in Climate Models and Satellite Data

Massachusetts Institute of Technology

Comparing Tropospheric Warming in Climate Models and Satellite Data

Benjamin D. Santer,[a] Susan Solomon,[b] Giuliana Pallotta,[a] Carl Mears,[c]
Stephen Po-Chedley,[d] Qiang Fu,[d] Frank Wentz,[c] Cheng-Zhi Zou,[e] Jeffrey Painter,[a]
Ivana Cvijanovic,[a] and Céline Bonfils[a]

[a] *Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory,
Livermore, California*
[b] *Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts*
[c] *Remote Sensing Systems, Santa Rosa, California*
[d] *Department of Atmospheric Sciences, University of Washington, Seattle, Washington*
[e] *Center for Satellite Applications and Research, NOAA/NESDIS, Camp Springs, Maryland*

ABSTRACT

Updated and improved satellite retrievals of the temperature of the mid-to-upper troposphere (TMT) are used to address key questions about the size and significance of TMT trends, agreement with model-derived TMT values, and whether models and satellite data show similar vertical profiles of warming. A recent study claimed that TMT trends over 1979 and 2015 are 3 times larger in climate models than in satellite data but did not correct for the contribution TMT trends receive from stratospheric cooling. Here, it is shown that the average ratio of modeled and observed TMT trends is sensitive to both satellite data uncertainties and model–data differences in stratospheric cooling. When the impact of lower-stratospheric cooling on TMT is accounted for, and when the most recent versions of satellite datasets are used, the previously claimed ratio of three between simulated and observed near-global TMT trends is reduced to approximately 1.7. Next, the validity of the statement that satellite data show no significant tropospheric warming over the last 18 years is assessed. This claim is not supported by the current analysis: in five out of six corrected satellite TMT records, significant global-scale tropospheric warming has occurred within the last 18 years. Finally, long-standing concerns are examined regarding discrepancies in modeled and observed vertical profiles of warming in the tropical atmosphere. It is shown that amplification of tropical warming between the lower and mid-to-upper troposphere is now in close agreement in the average of 37 climate models and in one updated satellite record.

1. Introduction

Reliable thermometer measurements of large-scale changes in Earth's surface temperature are available for over a century. These measurements document warming of roughly 0.85°C since 1880, with the three warmest decades in the most recent portion of the record (IPCC 2013). In global average terms, 2015 was the warmest year in surface temperature datasets (Tollefson 2016). Satellite-based estimates of trends in

tropospheric temperature cover a shorter period of time (from late 1978 to the present) but also provide independent confirmation of planetary-scale warming (Zou et al. 2006; Christy et al. 2007; Mears et al. 2011; Po-Chedley et al. 2015; Mears and Wentz 2016).

Although observational and model temperature data provide compelling evidence for the existence of a "discernible human influence" on global climate (Santer et al. 1995; Karl et al. 2006; Hegerl et al. 2007; Bindoff et al. 2013), studies of temperature change continue to yield interesting and important scientific puzzles. Examples of such puzzles include apparent differences between surface and tropospheric warming rates in observational records (Yulaeva and Wallace 1994; Hurrell and Trenberth 1998; National Research Council 2000; Gaffen et al. 2000; Santer et al. 2000; Hegerl and Wallace 2002; Karl et al. 2006) and differences between modeled and observed warming trends (National Research Council 2000; Gaffen et al. 2000; Hegerl and Wallace 2002; Karl

et al. 2006; Easterling and Wehner 2009; Fu et al. 2011; Santer et al. 2011; Po-Chedley and Fu 2012b). The causes of such differences remain the subject of both scientific interest (IPCC 2013; Fyfe et al. 2016; Lewandowsky et al. 2016) and political attention (U.S. Senate 2015).

The present study focuses on differences between satellite- and model-based estimates of tropospheric temperature change. We assess the validity of two highly publicized claims: that modeled tropospheric warming is a factor of 3–4 larger than in satellite and radiosonde observations (Christy 2015) and that satellite tropospheric temperature data show no statistically significant warming over the last 18 years (U.S. Senate 2015). We also address long-standing concerns regarding differences in the vertical structure of tropospheric warming in models and satellite data. Such differences are particularly pronounced in the tropics (Santer et al. 2000; Gaffen et al. 2000; Hegerl and Wallace 2002; Fu and Johanson 2005; Johanson and Fu 2006; Karl et al. 2006; Fu et al. 2011; Po-Chedley and Fu 2012b). We rely exclusively on satellite measurements of atmospheric temperature; we do not compare model results with radiosonde-based atmospheric temperature measurements, as has been done in a number of previous studies (Gaffen et al. 2000; Hegerl and Wallace 2002; Thorne et al. 2007, 2011; Santer et al. 2008; Lott et al. 2013).

2. Satellite and model temperature data

Since late 1978, satellite-based microwave temperature sounders have measured the microwave emissions from oxygen molecules. These emissions are proportional to the temperature of broad layers of the atmosphere (Mears et al. 2011). The two claims mentioned above (Christy 2015; U.S. Senate 2015) focused on trends in the temperature of the mid-to-upper troposphere (TMT), which extends to approximately 18 km above Earth's surface (Karl et al. 2006). Here, we analyze TMT data from four different research groups: Remote Sensing Systems (RSS; Mears and Wentz 2016), the Center for Satellite Applications and Research (STAR; Zou et al. 2006), the University of Alabama at Huntsville (UAH; Christy et al. 2007), and the University of Washington (UW; Po-Chedley et al. 2015). We also consider satellite estimates of the temperature of the lower stratosphere (TLS) and the temperature of the lower troposphere (TLT), which span approximate altitude ranges from 14 to 29 km and from the surface to 8 km (respectively).

Previous scientific assessments (National Research Council 2000; Karl et al. 2006; IPCC 2013) have highlighted

the large structural uncertainties in satellite estimates of tropospheric temperature change. The major uncertainties arise because the satellite TMT record is based on measurements made by more than 10 different satellites; over their lifetimes, most of these satellites experience orbital decay (Wentz and Schabel 1998) and orbital drift (Mears and Wentz 2005). These orbital changes affect the measurements of microwave emissions, primarily because of gradual shifts in the time of day at which measurements are made. Adjustments for such shifts in measurement time are large and involve many subjective decisions (Mears and Wentz 2005, 2016; Mears et al. 2011; Karl et al. 2006; Zou et al. 2006, 2009; Zou and Wang 2011; Christy et al. 2007; Po-Chedley et al. 2015). Further adjustments to the raw data are necessary for drifts in the onboard calibration of the microwave measurements (Mears et al. 2003; Po-Chedley and Fu 2012a; Zou et al. 2009; Zou and Wang 2011) and for the transition between earlier and more sophisticated versions of the microwave sounders (Mears and Wentz 2016).

Multiple dataset versions are available for the temperature records produced by RSS, UAH, and STAR (see the supplemental material). Newer dataset versions incorporate adjustments for problems identified after public release of earlier datasets and are likely to represent improved estimates of atmospheric temperature change. Use of multiple dataset versions highlights the evolutionary nature of satellite temperature datasets—an evolution paced by advances in identifying and correcting the complex nonclimatic factors affecting these measurements.[1] This corrective process is ongoing.

Satellite TMT measurements receive a contribution from the stratosphere (Spencer and Christy 1992; Fu et al. 2004; Fu and Johanson 2004, 2005; Johanson and Fu 2006). Large, anthropogenically driven cooling of the lower stratosphere (Solomon 1999; Karl et al. 2006; Ramaswamy et al. 2006; IPCC 2013; Santer et al. 2013b) can contribute significantly to TMT trends (Fu et al. 2004; Fu and Johanson 2005; Fu et al. 2011; Po-Chedley and Fu 2012b; Po-Chedley et al. 2015). A regression-based method has been used to correct TMT data for this contribution (Fu et al. 2004; Fu and Johanson 2005). The efficacy of this approach was validated with both observed and model atmospheric temperature data (Fu and Johanson 2004; Gillett et al. 2004; Kiehl et al. 2005). We employ the same regression approach here to derive corrected tropospheric

---

[1] For RSS, UAH, and STAR, the newer TMT versions used here only became available in 2016; currently available model-versus-data comparisons relied exclusively on older dataset versions (Fu et al. 2011; Po-Chedley and Fu 2012b; Santer et al. 2013a,b).

temperatures (TMT_{cr}) from satellite and model TMT datasets (see appendix A).

Model atmospheric temperatures were available from phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012). We analyzed simulations of externally forced climate change performed with 37 different CMIP5 models. The simulations have estimated historical changes in natural and anthropogenic external forcing from the mid-1800s to 2005. From 2006 to the end of the twenty-first century, changes in anthropogenic greenhouse gases and aerosols are prescribed according to the representative concentration pathway 8.5 (RCP8.5), which has radiative forcing of roughly $8.5 \, W \, m^{-2}$ by 2100. We also used preindustrial control runs (with no changes in external influences on climate) from 36 models to obtain information on natural internal climate variability. To facilitate the direct comparison of satellite data with model output, "synthetic" satellite temperatures were calculated from all model simulations (Santer et al. 2013b). The model atmospheric temperature data analyzed here are fully described in the supplemental material and in Tables 1–4 of the supplemental material, together with information on the forcings used in the simulations of historical climate change.[2]

To avoid truncating comparisons between modeled and observed atmospheric temperature trends in December 2005, we spliced together synthetic satellite temperatures from the historical simulations and the RCP8.5 runs.[3] Splicing allows us to compare actual and synthetic temperature changes over the full 37-yr length of the satellite record. We use the label "ALL+8.5" to identify these spliced simulations.

3. Atmospheric temperature time series

We consider first the time series of changes in simulated and observed atmospheric temperature over the satellite era (Fig. 1). Our focus is on temperatures averaged over a near-global domain and over the tropics. In the lower stratosphere (Figs. 1a,b), the ALL+8.5 simulations and the satellite data are both characterized by overall cooling in response to human-caused decreases in stratospheric ozone and increases in carbon dioxide (Solomon 1999; Karl et al. 2006; Ramaswamy et al. 2006). This long-term stratospheric cooling trend is punctuated by short-term (1–2 yr) lower-stratospheric

warming arising from the eruptions of El Chichón in 1982 and Mount Pinatubo in 1991 (Robock 2000; Ramaswamy et al. 2006; Santer et al. 2013b). The size of this short-term warming is very similar in the satellite data and the multimodel average of the ALL+8.5 simulations, but this apparent agreement arises from compensating errors (see the supplemental material).

Since 1979, mid- to upper-tropospheric temperature has increased in both the observations and the ALL+8.5 integrations, with larger warming in the simulations (Figs. 1c–f). Another prominent feature of the TMT and TMT_{cr} time series is cooling caused by the eruptions of El Chichón and Mount Pinatubo (Robock 2000; Santer et al. 2001; Wigley et al. 2005; Thompson et al. 2009; Santer et al. 2013b, 2014). Volcanic cooling of the troposphere is noticeably less noisy in the multimodel average than in the observations for well-understood reasons (see the supplemental material).

Correction of TMT for lower-stratospheric cooling is expected to increase overall trends in mid-to-upper-tropospheric temperature (Fu et al. 2004, 2011; Fu and Johanson 2005; Po-Chedley and Fu 2012b). Simple visual comparison of the TMT and TMT_{cr} temperature time series for near-global averages (cf. Figs. 1c and 1e) and for tropical averages (cf. Figs. 1d and 1f) does not reveal how this correction affects the level of consistency between model and observed tropospheric warming trends. In the following, we provide a quantitative assessment of the impact of TMT correction on model–data trend consistency.

4. Trend ratios

A key aspect of our analysis framework is that we consider the sensitivity of linear temperature trends (and of model–data trend ratios) to different choices of the start date and the trend length L (Santer et al. 2011). Rather than focusing on one limited subset of the temperature time series in Fig. 1, such as the last 18 years of TMT records (U.S. Senate 2015), our strategy here is to examine all possible 18-yr temperature trends during the satellite era (see appendix B). Since temperature trends on 18-yr time scales have no special diagnostic value, we vary L in increments of 1 year, from a minimum of 10 years to a maximum of 37 years (the full length of the satellite records). This allows us to compare the average values of modeled and observed temperature trends on a range of different time scales, while accounting for the effect of monthly and interannual variability on linear trend estimates. Our strategy reduces the chance of making incorrect statistical inferences based on analysis of a single arbitrarily selected trend.

Figures 2a and 2b show the averages of the sampling distributions of L-year trends for near-global TMT and

---

[2] Detailed information on the implementation of external forcings in the full ensemble of CMIP5 models is available in a limited number of cases only (e.g., for stratospheric ozone forcing; see Eyring et al. 2013).

[3] The RCP8.5 simulations were typically initiated from conditions of the climate system at the end of the historical run.
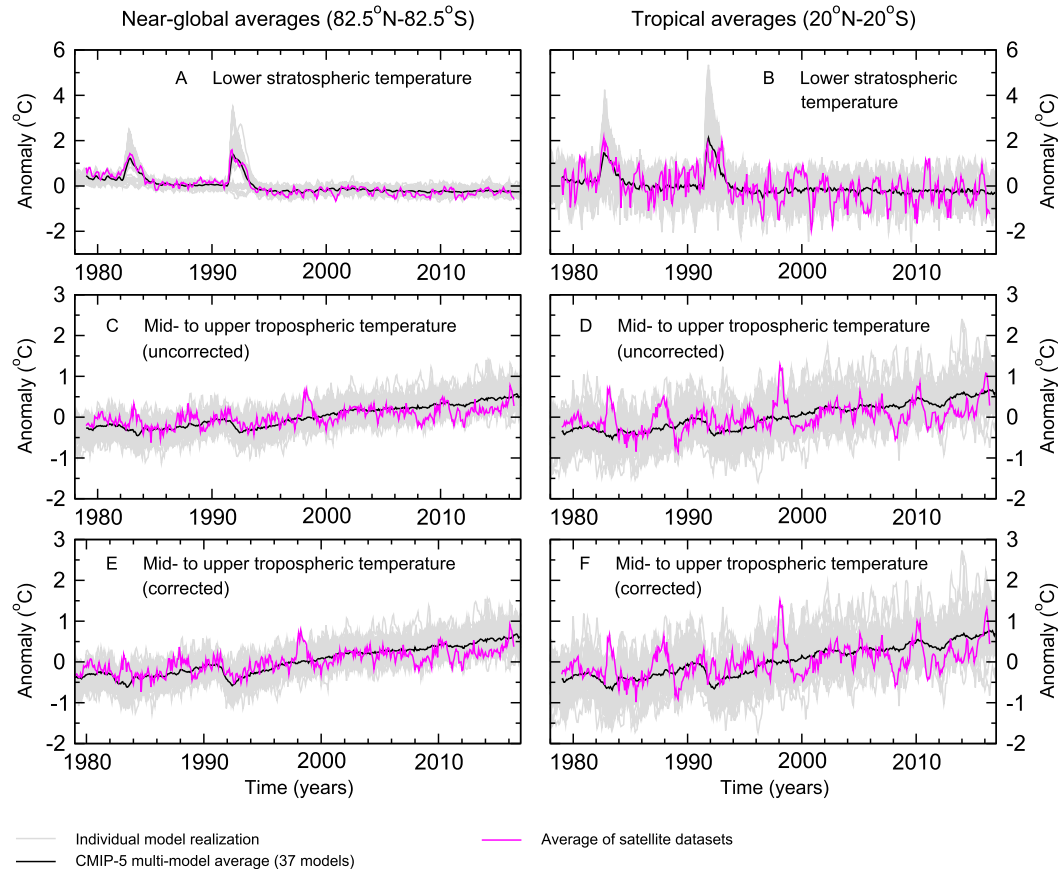
FIG. 1. Time series of monthly mean anomalies in atmospheric temperature over January 1979–December 2016 (simulations) and January 1979–June 2016 (observations). Results are spatially averaged over (left) a near-global domain, and (right) over the tropics. Temperature estimates for (a),(b) the lower stratosphere and (c),(d) the mid-to-upper troposphere were obtained from satellite-based microwave sounders. Correcting TMT for stratospheric cooling yields TMT$_{cr}$, which is more representative of true changes in (e),(f) mid-to-upper-tropospheric temperature. Model synthetic TLS and TMT data were computed using vertical weighting functions that approximate the satellite-based vertical sampling of the lower stratosphere and mid-to-upper troposphere. Synthetic satellite temperatures are from 49 simulations of externally forced climate change performed with 37 different CMIP5 models. All anomalies are relative to climatological monthly means calculated over January 1979–December 2015.

TMT$_{cr}$ data. The distribution average trends $\overline{b_o}(k, l)$ and $\overline{\overline{b_f}}(l)$ are for the observations and the forced ALL+8.5 simulations, respectively, where the indices k and l span the number of observational datasets and the number of values of L (see appendix B).[4] We consider first TMT results that have not been corrected for lower-stratospheric cooling (Fig. 2a). At all time scales considered, simulated TMT trends are larger than satellite TMT trends. Only the values of $\overline{b_o}(k, l)$ for RSS version 4.0 and STAR versions 3.0 and 4.0 are consistently

within the 5th–95th percentile range of model estimates of externally forced TMT trends.[5]

Correcting for lower-stratospheric cooling (Fig. 2b) increases the size of mid-to-upper-tropospheric warming

---

[4] Here, the single overbar in $\overline{b_o}(k, l)$ indicates the average of a distribution of L-year trends. The double overbar in $\overline{\overline{b_f}}(l)$ signifies a distribution average as well as an average over models and ALL+8.5 realizations.

[5] In a companion paper, we evaluate the statistical significance of differences between tropospheric temperature trends in individual satellite datasets and in the multimodel average of the ALL+8.5 simulations (B. Santer et al. 2016, unpublished manuscript). We show that the statistical significance of these trend differences is highly sensitive to the analysis time scale L and to the trend start date. Over the first 15 years of the twenty-first century, differences between modeled and observed tropospheric warming rates are highly significant and are unlikely to be explained by internal variability alone. In contrast, model-versus-observed trend differences in the last two decades of the twentieth century are generally consistent with internal variability.

Mid- to Upper Tropospheric Temperature Trends and Trend Ratios in Models and Data

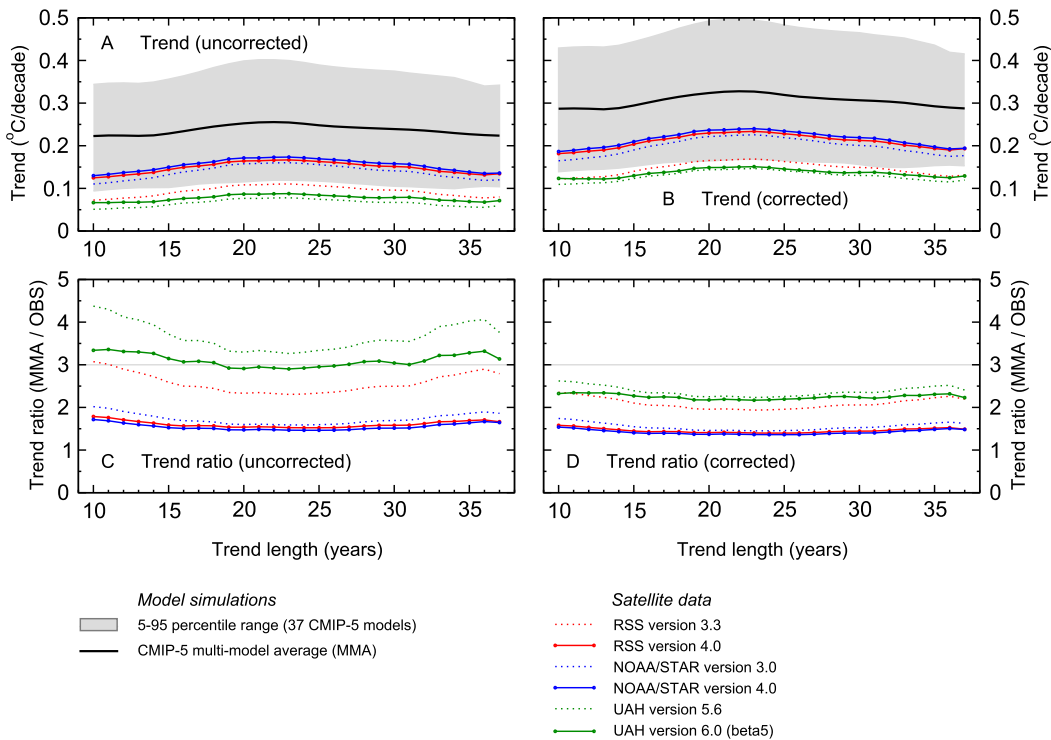### Results spatially averaged over a near-global domain (82.5°N-82.5°S)

FIG. 2. Comparison of average simulated and observed trends in near-global tropospheric temperature as a function of the trend length L. Results are for (a),(c) uncorrected TMT and (b),(d) TMT_{cr}. The average of all maximally overlapping L-year trends in a given observed time series (calculated separately for each satellite dataset and each L-year time scale of interest) is shown in (a) and (b), together with the multimodel average trend results from the spliced historical and RCP8.5 runs. Values of the trend ratio $R(k, l)$ in (c) and (d) provide information on the relative sizes of temperature trends in the externally forced simulations and the satellite observations (see appendix B; k and l are indices over satellite datasets and values of the time scale L). The gray line in (c) and (d) shows the model–observed trend ratio of 3 reported for near-global TMT comparisons (Christy 2015).

trends (Fu et al. 2004, 2011; Fu and Johanson 2004, 2005; Johanson and Fu 2006; Karl et al. 2006; Po-Chedley and Fu 2012b). TMT correction also systematically reduces $R(k, l)$, the ratio between modeled and observed L-year temperature trends (Figs. 2c,d). The reason for this reduction is that in all satellite datasets examined here, the observed lower-stratospheric cooling is larger than in the average of the ALL+8.5 simulations[6] (Figs. 3a,c).

This discrepancy between satellite and model TLS trends arises from multiple factors: the underestimation of

observed stratospheric ozone loss in many of the CMIP5 ALL+8.5 runs (Solomon et al. 2012; Hassler et al. 2013; Eyring et al. 2013; Young et al. 2014), model–data differences in stratospheric water vapor changes (Solomon et al. 2010; Gilford et al. 2016), and different phasing of stratospheric internal variability in the real world and the model simulations (Gilford et al. 2016). The systematic model–data differences in lower-stratospheric cooling in Fig. 3 hamper reliable estimation of the relative sizes of simulated and observed tropospheric warming. If tropospheric trend comparisons are the primary scientific focus, then the use of uncorrected TMT data (as in Christy 2015) leads to erroneous conclusions.

As a simple measure of overall consistency between model and satellite trends, we compute $\overline{R}(k)$, where the overbar indicates a model–data trend ratio that is averaged over all values of the trend length L. Using trends in near-global averages of uncorrected TMT data, we obtain $\overline{R}(k)$

---

[6] Lower-stratospheric cooling in radiosonde data is also larger than in the multimodel average of the externally forced simulations (Seidel et al. 2016). As expected, the observational datasets with the largest cooling of the lower stratosphere (UAH versions 5.6 and 6.0) show the largest decrease in trend ratios after TMT is corrected (see Table 5 in the supplemental material).

Lower Stratospheric Temperature Trends and Trend Ratios in Models and Data

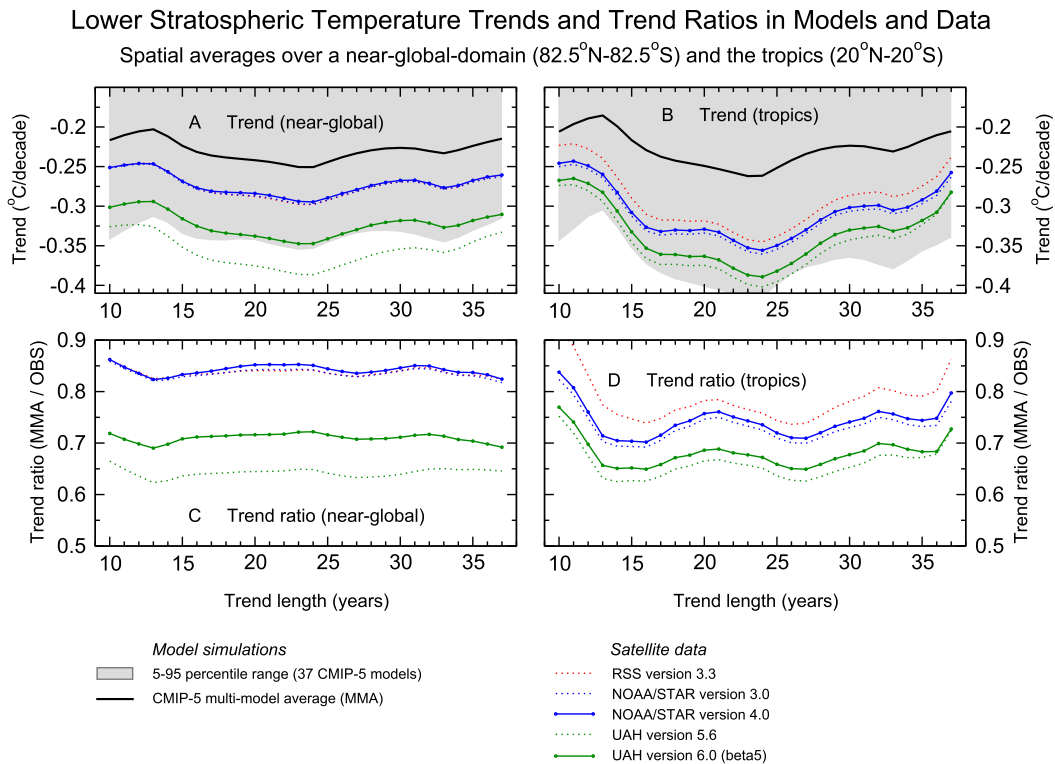Spatial averages over a near-global-domain (82.5°N-82.5°S) and the tropics (20°N-20°S)

FIG. 3. (a),(b) Average trends and (c),(d) trend ratios in the TLS. All results were calculated using simulated and observed monthly mean TLS time series spatially averaged (a),(c) over a near-global domain and (b),(d) over the tropics. The analysis is for temperature trends on time scales ranging from 10 to 37 yr. Refer to Fig. 2 and appendix B for analysis details.

values of 2.58 and 1.61 for RSS versions 3.3 and 4.0, 1.73 and 1.54 for STAR versions 3.0 and 4.0, and 3.67 and 3.10 for the earlier and most recent versions of the UAH temperature data (Fig. 2c and Table 5 of the supplemental material). Correcting TMT for stratospheric cooling reduces these $\overline{R}(k)$ values to 2.09, 1.46, 1.55, 1.42, 2.38, and 2.25, respectively, and brings simulated and satellite-inferred tropospheric warming trends into closer agreement (Fig. 2d). The impact of observational uncertainties on $\overline{R}(k)$ is also reduced.

A recent study by Christy (2015) reported that global warming of the mid-to-upper troposphere is a factor of 3 larger in models than in observations; that is, $\overline{R}(k) \approx 3$ for TMT trends over the full satellite era. This finding is only supported by model–data comparisons relying on uncorrected UAH TMT data. It is not supported by comparisons involving uncorrected STAR or RSS TMT data (Fig. 2c). After correcting TMT for stratospheric cooling, the claim that $\overline{R}(k) \approx 3$ does not hold for any model–data trend comparisons. If the observational average of $\overline{R}(k)$ [denoted here by $\overline{\overline{R}}$, where the double overbar denotes averaging of $R(k, l)$ over both the number of observational datasets and the number of trend lengths considered] is calculated with all six

versions of the near-global TMT_{cr} time series, then $\overline{\overline{R}} = 1.86$. The average trend ratio is even lower if only the three most recent TMT_{cr} versions are used in this calculation ($\overline{\overline{R}} = 1.71$; see Fig. 2d and Table 5 of the supplemental material). Values of $\overline{\overline{R}}$ are relatively insensitive to different reasonable processing choices, such as the exclusion of ALL+8.5 simulations lacking explicit treatment of the radiative effects of stratospheric volcanic aerosols (see section 1.2.2 of the supplemental material).

We obtain qualitatively similar results for TMT data averaged over the tropics (Fig. 4). As in the case of near-global averages, correcting tropical TMT for stratospheric cooling (Figs. 3b,d) systematically reduces $\overline{R}(k)$ (Figs. 4c,d and Table 5 in the supplemental material). The statement that model tropical TMT trends over the satellite era are a factor of 4 larger than in observations (Christy 2015) holds only for the uncorrected UAH TMT data (Fig. 4c). All model–data trend comparisons with corrected tropical-average TMT datasets yield $\overline{R}(k)$ values less than 4: 2.39 and 1.72 for RSS versions 3.3 and 4.0, 1.84 and 1.52 for STAR versions 3.0 and 4.0, 3.73 and 3.24 for the earlier and most recent UAH dataset versions, and approximately 1.96 for UW

Mid- to Upper Tropospheric Temperature Trends and Trend Ratios in Models and Data

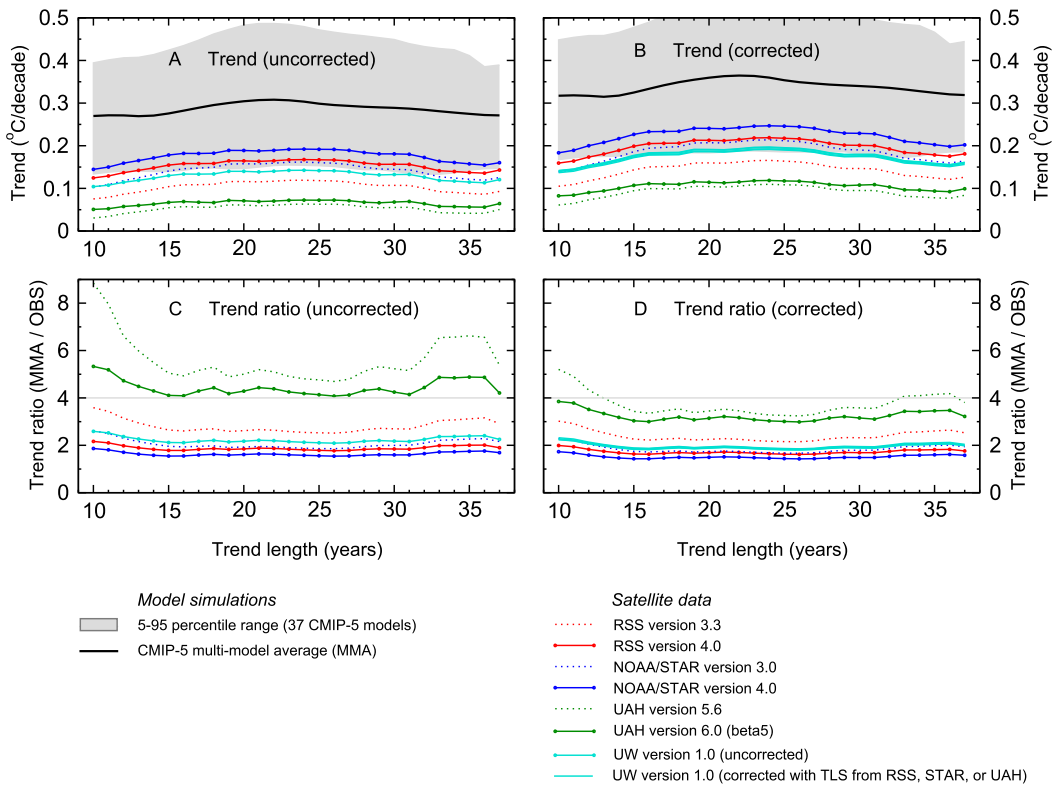Results spatially averaged over the tropics (20°N-20°S)

FIG. 4. As in Fig. 2, but for TMT and TMT$_{cr}$ data spatially averaged over a tropical domain. The gray line in (c) and (d) shows the model–observed trend ratio of 4 reported by Christy (2015) for tropical TMT comparisons.

(Fig. 4d). Averaging these ratios yields $\overline{\overline{R}} = 2.29$ if all observational datasets are used and $\overline{\overline{R}} = 2.11$ if only the most recent dataset versions are employed.

Although accounting for stratospheric cooling effects on TMT brings modeled and observed tropospheric warming trends into better agreement, values of $\overline{R}(k)$ in Figs. 2d and 4d are still sufficiently large to be of scientific concern. These concerns are not new. Differences in the size of simulated and observed warming trends—both in the troposphere and at Earth's surface—have been the subject of scientific attention since the late 1990s (National Research Council 2000; Hegerl and Wallace 2002; Hegerl et al. 2007; Karl et al. 2006; Fu et al. 2004, 2011; Santer et al. 2011, 2013b, 2014; Solomon et al. 2011; Po-Chedley and Fu 2012b; IPCC 2013; Fyfe et al. 2013a, 2016).

The message from this large body of research is that temperature trend differences have multiple explanations. These explanations are not mutually exclusive. They include model errors in the response to external forcing (Trenberth and Fasullo 2010), systematic model errors in the forcings themselves (Solomon et al. 2010,

2011, 2012; Kopp and Lean 2011; Shindell et al. 2013; Hassler et al. 2013; Eyring et al. 2013; Young et al. 2014; Santer et al. 2014; Smith et al. 2016), residual errors in satellite temperature records (Wentz and Schabel 1998; Mears and Wentz 2005, 2016; Mears et al. 2003, 2011; Zou et al. 2006, 2009; Zou and Wang 2011; Po-Chedley and Fu 2012a; Po-Chedley et al. 2015) and in surface temperature data (Morice et al. 2012; Cowtan and Way 2014; Karl et al. 2015), and differences in the phasing of internal climate variability in the "many worlds" of the simulations and the single world of the observations (Fyfe et al. 2013a, 2016; Kosaka and Xie 2013; Meehl et al. 2014; England et al. 2014; Risbey et al. 2014; Steinman et al. 2015; Trenberth 2015; Gilford et al. 2016). It is incorrect to assert that a large model error in the climate sensitivity to greenhouse gases is the only or most plausible explanation for differences in simulated and observed warming rates (Christy 2015).

We also compare satellite and model trends for TLT (Fig. 5). Unlike TMT, TLT is far less contaminated by lower-stratospheric cooling (Spencer and Christy 1992; Fu et al. 2011) and is thus relatively unaffected by

Lower Tropospheric Temperature Trends and Trend Ratios in Models and Data

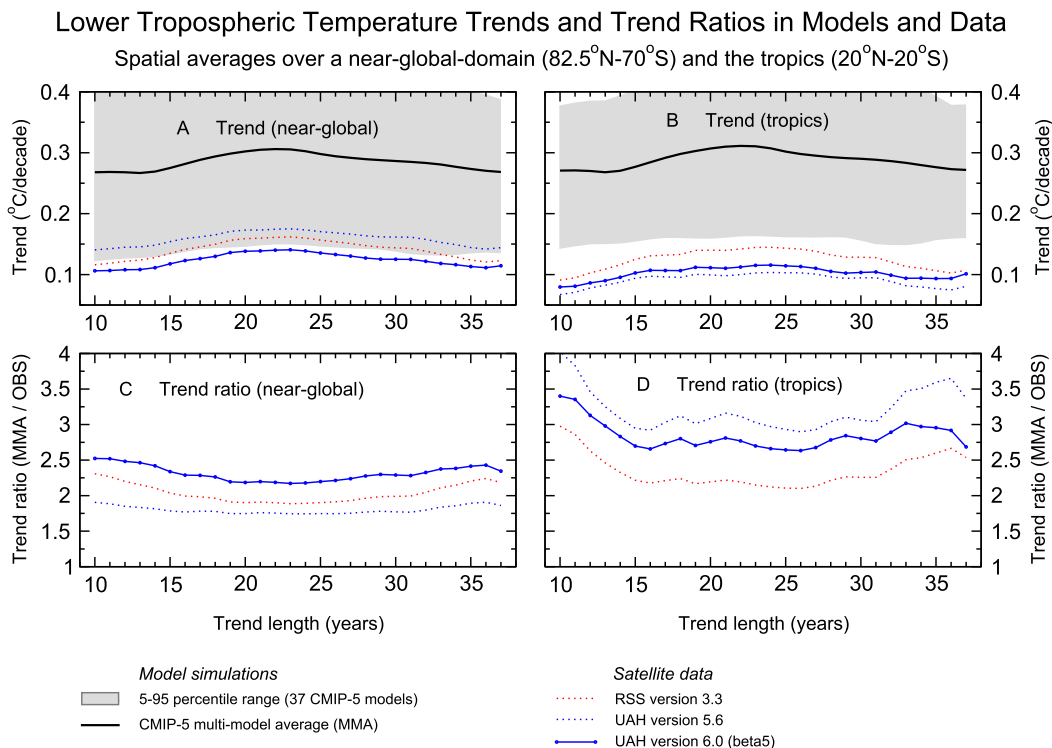Spatial averages over a near-global-domain (82.5°N-70°S) and the tropics (20°N-20°S)

FIG. 5. As in Fig. 2, but for average trends and trend ratios in TLT. See appendix B for further details. Note that the NOAA/STAR and UW research groups do not produce a TLT dataset. RSS provides two versions of their TMT dataset (versions 3.3 and v4.0), but for TLT only one dataset is currently available (version 3.3).

differences between modeled and observed lower-stratospheric temperature trends. For near-global TLT data, values of the average model–data trend ratio $\overline{R}(k)$ range from 1.80 to 2.31 (Fig. 5c). This is well below the ratio of 3 claimed for near-global TMT trends (Christy 2015). Similarly, tropical TLT data (Fig. 5d) yield $\overline{R}(k)$ values ranging from 2.35 to 3.21, which are consistently below the ratio of 4 reported by Christy (2015) for tropical TMT trends. As in the case of TMT, differences between simulated and observed TLT trends are due to multiple factors (see above).

5. Significance of tropospheric warming trends

Next, we examine the statement that "according to the satellite data, there has been no significant global warming for the past 18 years" (U.S. Senate 2015). Our concern is with two specific issues arising from this statement: whether a single, arbitrarily selected 18-yr period is statistically representative of all possible 18-yr periods in the full satellite record, and whether the claim of no significant warming over the last 18 years is valid.

Consider first the representativeness of a single temperature trend calculated over the last 18 years. One of the dominant modes of internal climate variability is El

Niño–Southern Oscillation. El Niño is the warm phase of this mode of variability. Large El Niño events are characterized by warming of the eastern equatorial Pacific Ocean, followed by global-scale warming of the troposphere after a lag of roughly 4–6 months (Santer et al. 2001; Wigley et al. 2005; Thompson et al. 2009). One of the largest El Niño events of the twentieth century occurred during the winter–spring season of 1997/98, with peak global-mean tropospheric warming in April 1998 and a gradual decay to more normal conditions by the fall of 1998 (Fig. 6). For a selected trend length of 18 years (216 months) and a trend start date of January 1998, the trend end date is in December 2015. A time horizon of the last 18 years, therefore, yields an anomalously warm trend start point because of the unusually large 1997/98 El Niño.

To explore the trend dependence on the trend length L and the trend start and end dates, we show maximally overlapping near-global TMT_{cr} trends for seven different values of L (15–21 yr). Values of $b_o(i, k, l)$ (the linear trend for the ith overlapping L-year segment of the kth observed TMT_{cr} time series, and for the lth value of L) are plotted in Fig. 7, left. For each overlapping trend, satellite dataset, and trend length L, we use CMIP5 control runs to calculate the probability $\overline{p_c}(i, k, l)'$ that
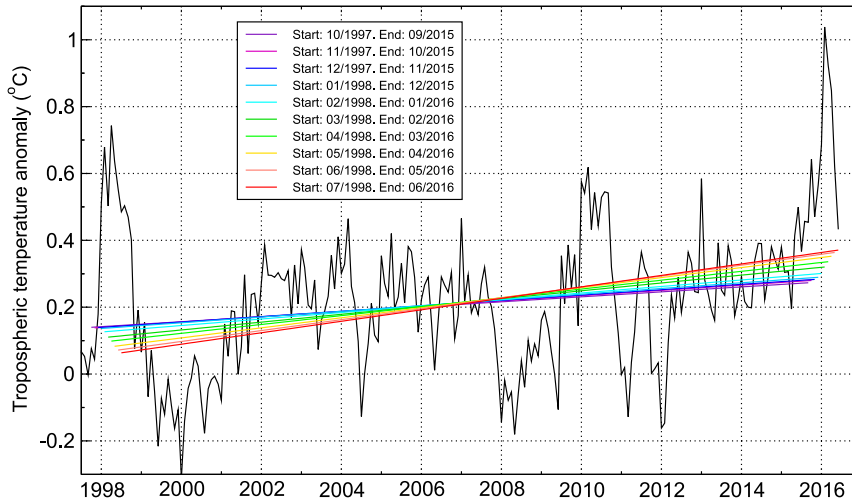
FIG. 6. Recent 18-yr (216-month) trends in TMT$_{cr}$. Results are for RSS version 4.0. Trends were calculated from the time series of monthly mean anomalies of near-global TMT$_{cr}$ data. The most recent 18-yr trend ends in June 2016.

the observed warming trend could have been caused by internal climate variability alone (see appendix B). Values of $\overline{p_c}(i, k, l)'$ are given in Fig. 7, right.

As expected, shorter trends are more affected by interannual variability and thus yield a wider range of trend values (Santer et al. 2011). Even for the short, 15-yr trends in Fig. 7a, however, it is difficult to obtain periods with tropospheric cooling. The occurrence of cooling periods is related to the length of L relative to the phasing of specific events. There are two groups of negative 15-yr TMT$_{cr}$ trends: the first group of trends has end points close to the maximum cooling caused by the Mount Pinatubo eruption; the second group of trends has start points close to the warming "spike" associated with the 1997/98 El Niño. As L increases beyond 15 yr, the influence from Mount Pinatubo on trend end points diminishes, and the first group of negative trends disappears. In the second group, negative trends persist out to trend lengths of 17 yr, but are highly unusual for $L = 18$ yr, and occur in only two of the six satellite dataset versions[7] (Fig. 7g).

For $L \geq 19$ yr, all near-global TMT$_{cr}$ trends are positive in every satellite dataset. At these longer time scales, the impact of seasonal and interannual temperature anomalies is damped, and gradual tropospheric warming is more reliably sampled. For values of $L = 21$ yr, almost all observed warming trends are significantly larger (at the 10% level or better) than 21-yr warming trends inferred from model estimates of internal variability (Fig. 7n).

Figure 7 illustrates that it is no longer valid to claim that satellite TMT data show "no significant global warming for the past 18 years" (U.S. Senate 2015). In five of the six versions of the satellite TMT$_{cr}$ time series, the most recent 216-month warming trends attain significance at the 10% level or better.[8] Trend significance is partly due to the fact that these recent periods sample warming associated with the 2015/16 El Niño event, which contributed to the record-breaking annual global-mean surface temperature in 2015 (Pidcock 2016; Tollefson 2016). Significance also arises because the start point of the most recent 216-month trend is less influenced by the anomalous warmth of the 1997/98 El Niño and is beginning to sample the cooler conditions caused by the La Niña in 1999/2000 (see Fig. 6).

Other claims of "no significant warming over the last X years" are also sensitive to the choice of starting point and analysis time scale. For example, statements made in 2013 (2014) that satellite data show no significant global warming over the last 16 (17) years would be incorrect if made today. In four (six) satellite TMT$_{cr}$ datasets, the most recent 16 (17)-yr warming trends are now significantly larger (at the 10% level or better) than the estimated warming from natural internal climate variability (Figs. 7d,f). Furthermore, a possible 2017 claim

---

[7] Both are older dataset versions (RSS version 3.3 and STAR version 3.0).

[8] Significance is attained for 216-month trends ending in the following months: June 2016 (RSS version 3.3), January 2016 through June 2016 inclusive (RSS version 4.0), March 2016 through June 2016 inclusive (STAR version 4.0 and UAH version 5.6), and May 2016 and June 2016 (UAH version 6.0). None of the most recent near-global TMT$_{cr}$ trends show significant warming in the older version (version 3.0) of the STAR dataset.

Observed Near-Global Corrected TMT Trends: Tests Against Internal Climate Variability
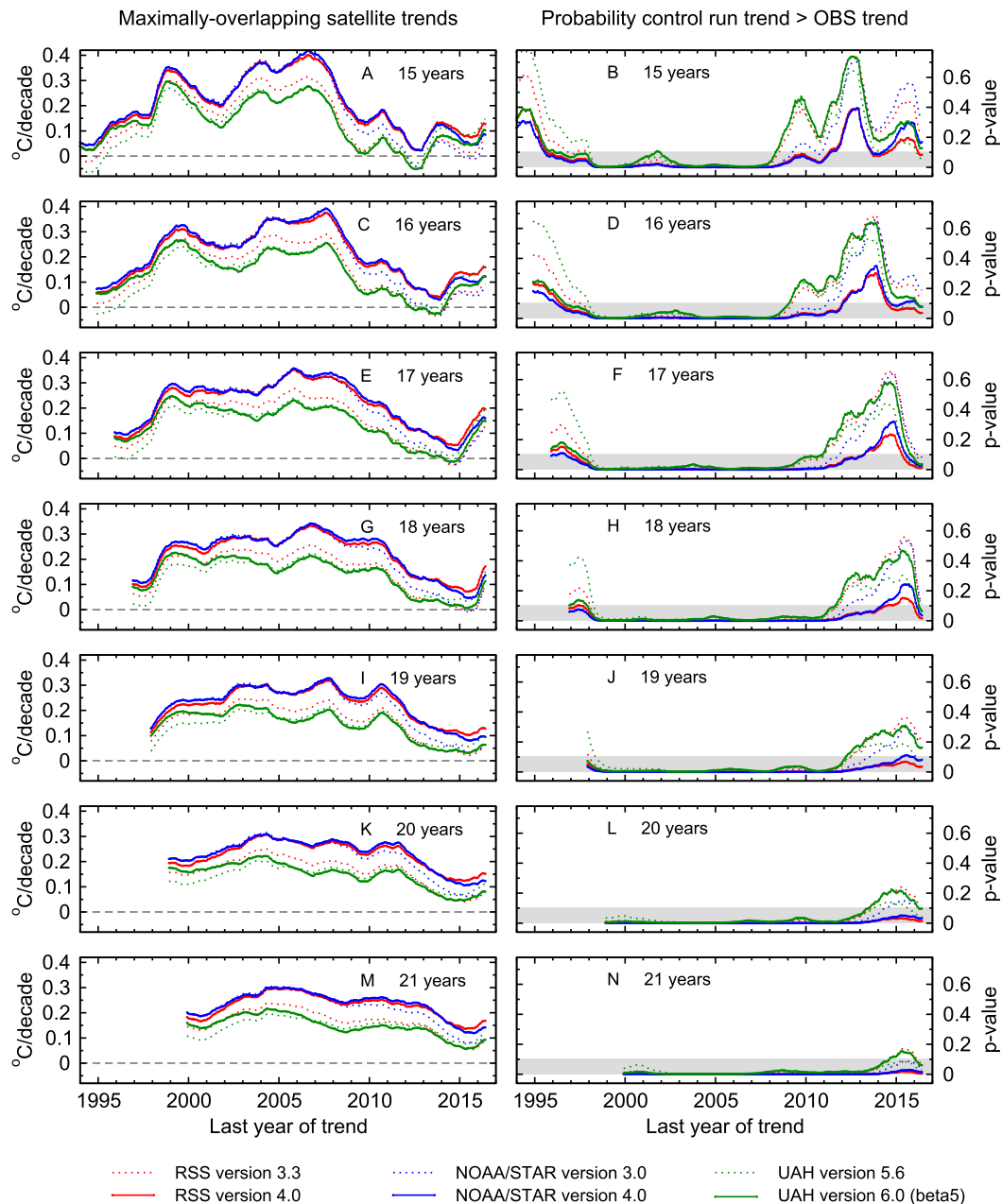

FIG. 7. Satellite trends in near-global averages of (left) TMT_{cr} and (right) associated p values. (top)–(bottom) Results are for trend lengths of $L = 15$–21 yr. Trends are from a "sliding window" analysis, where the L-year window advances by one month at a time. The final month of each satellite TMT_{cr} time series is June 2016. Results are plotted on the trend end date. The p values are for tests of the null hypothesis that an individual observed L-year warming trend could have been generated by internal variability alone (see appendix B). The rejection region for this hypothesis (at a stipulated 10% significance level) is shaded in gray. The y-axis range was extended to -0.06 to facilitate visual display of p values at or close to zero.

of no significant warming over the last 19 years would not be supported by three of the six satellite datasets (Fig. 7j).

While Fig. 7 shows estimates of the significance of tropospheric warming for individual L-year observed

trends, it is also useful to consider the mean significance levels. For the kth satellite dataset and lth trend length L, we simply average the individual $\overline{p_c}(i, k, l)'$ values over i, the index of maximally overlapping L-year trends

in the observations (see appendix B). This yields the average probability $\overline{\overline{p_c}}(k,l)'$ that the warming trends in a particular satellite dataset (and for a selected L-year time scale) could be due to internal variability alone (Santer et al. 2011). Values of $\overline{\overline{p_c}}(k,l)'$ identify the time scale at which we might expect an observed warming trend to surpass (and remain above) the level of model-estimated internal variability. We refer to this subsequently as the detection time scale. It is assessed here at a stipulated significance level of 10%.

In the uncorrected near-global TMT data, this time scale is 19 and 16 yr for RSS versions 3.3 and 4.0, 18 and 16 yr for STAR versions 3.0 and 4.0, and 22 and 20 yr for UAH versions 5.6 and 6.0 (Fig. 8a). Correcting TMT for stratospheric cooling generally yields shorter detection time scales for the tropospheric warming trends estimated from these satellite datasets (18, 15, 17, 15, 18, and 18 yr, respectively; Fig. 8b). It also reduces the range of observational uncertainty in the detection time scale.

In tropical-mean TMT data, internal climate noise is larger than for near-global averages of TMT (not shown). Detection time scales for uncorrected tropical TMT data are therefore longer than for uncorrected near-global TMT, ranging from 18 yr for STAR version 4.0 to 37 yr for UAH version 5.6 (Fig. 8c). As in the case of the near-global results, correcting tropical TMT for stratospheric cooling leads to systematically shorter (and more similar) detection time scales, which range from 17 to 24 yr (Fig. 8d).

The credibility of these detection time scales [and of the $\overline{p_c}(i,k,l)'$ values in Fig. 7] is critically dependent on the reliability of model-based estimates of the natural variability of tropospheric temperature, particularly on multidecadal time scales. In previous work, we found no evidence that current climate models systematically underestimate the amplitude of observed tropospheric temperature variability on 5–20-yr time scales (Santer et al. 2011, 2013b). In fact, our results suggest that CMIP5 models overestimate observed temperature variability on these time scales (Santer et al. 2013b), which implies that our statistical significance estimates are conservative. If the results from such variability comparisons are confirmed, the true $\overline{p_c}(i,k,l)'$ values may be lower than in Fig. 7, and the true detection time scales may be shorter than in Fig. 8.

6. Amplification of tropical warming with increasing altitude

Finally, it is of interest to examine how well current climate models perform in capturing observed relationships between trends in TLT and TMT (Fu et al. 2004, 2011; Fu and Johanson 2004, 2005; Po-Chedley and Fu 2012b). In the tropics, moist thermodynamic processes amplify surface warming, yielding peak warming at roughly

200 hPa (Yulaeva and Wallace 1994; Hegerl and Wallace 2002; Stone and Carlson 1979; Santer et al. 2005). We expect, therefore, that after correcting TMT for lower-stratospheric cooling, the warming of the tropical mid-to-upper troposphere should exceed the warming in the tropical lower troposphere. Such tropical amplification occurs for any surface warming; it is not a unique signature of greenhouse gas (GHG)-induced warming, as has been incorrectly claimed (Christy 2015).

The ratio between tropical TMT_{cr} and TLT trends $R_{MT/LT}$ has been used to assess model performance in capturing observed amplification behavior (Fu and Johanson 2005; Fu et al. 2011; Po-Chedley and Fu 2012b). From theory (Stone and Carlson 1979) and basic physical principles (Santer et al. 2005; Held and Soden 2006), we expect that models and satellite observations should have values of $R_{MT/LT} > 1$. Using corrected TMT data, two previous studies confirmed this expectation (Fu et al. 2011; Po-Chedley and Fu 2012b). However, this earlier work also found that $R_{MT/LT}$ was significantly smaller in satellite data than in three different multimodel ensembles. The two investigations were unable to determine whether discrepancies between modeled and satellite-based $R_{MT/LT}$ values were due to systematic errors in model amplification behavior, residual errors in the satellite TMT_{cr} and TLT data, or a combination of these factors.

Here, we calculate $\overline{R}_{MT/LT}$, the time scale–average TMT_{cr}/TLT trend ratio. Since STAR and UW do not produce TLT datasets, and version 4.0 of the RSS TLT dataset is not yet available, only three satellite datasets analyzed here can be used to compute internally consistent[9] values of $\overline{R}_{MT/LT}$. As in Fu et al. (2011) and Po-Chedley and Fu (2012b), the use of corrected TMT data increases $\overline{R}_{MT/LT}$ in these three satellite datasets (cf. Figs. 9a and 9b). But in contrast to the results from the two earlier studies, RSS version 3.3 now yields $\overline{R}_{MT/LT} = 1.149$, which is within 2% of the CMIP5 value of $\overline{R}_{MT/LT} = 1.172$. For RSS, therefore, we no longer find evidence of a serious mismatch between simulated and observed amplification behavior in the tropical troposphere. Since Po-Chedley and Fu (2012b) also relied on CMIP5 simulations and on version 3.3 of the corrected RSS tropospheric temperature data, the fact that we obtained closer agreement between RSS and model average $\overline{R}_{MT/LT}$ values appears to be primarily related to the availability of a longer observational record.

In contrast, UAH-based $\overline{R}_{MT/LT}$ values of 1.013 and 1.030 (for UAH versions 5.6 and 6.0, respectively) are now even

---

[9] Internally consistent denotes use of the same dataset versions of TLS, TMT, and TLT for calculating ratios between tropical TMT_{cr} and TLT trends. Internally consistent amplification ratios can be calculated with temperature data from RSS version 3.3 and UAH versions 5.6 and 6.0.

Tests of Observed Temperature Trends Against Internal Variability
Results are for the mid- to upper troposphere

Satellite data

RSS version 3.3
RSS version 4.0
NOAA/STAR version 3.0
NOAA/STAR version 4.0
UAH version 5.6
UAH version 6.0 (beta5)

Satellite data

UW version 1.0 (uncorrected)
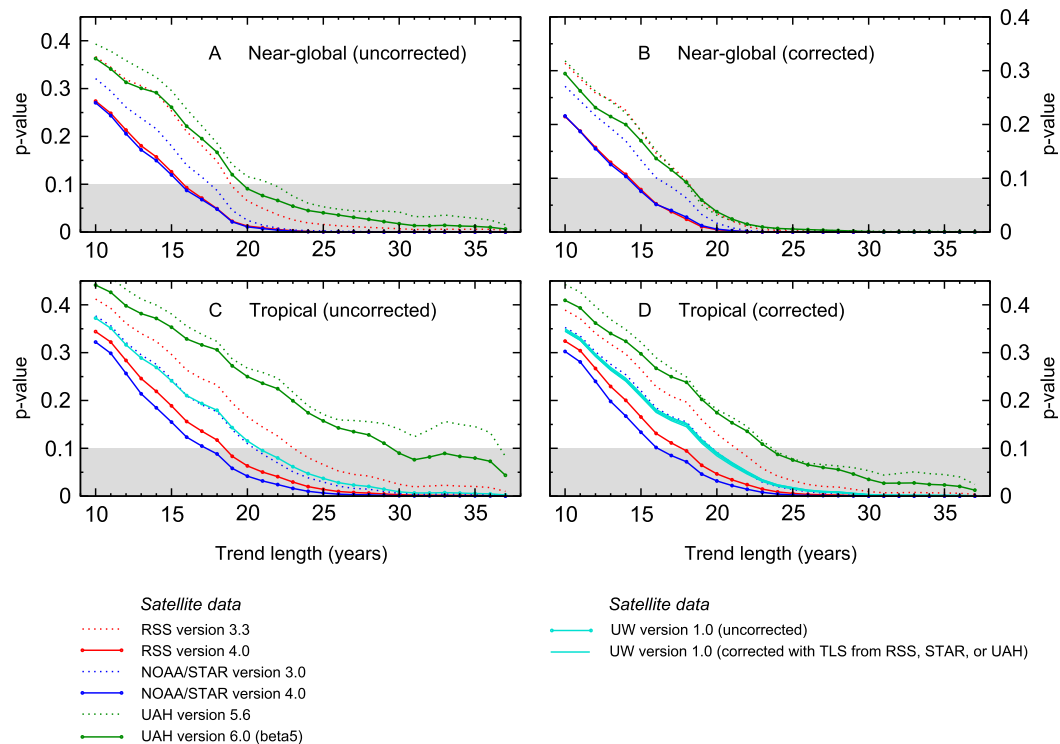UW version 1.0 (corrected with TLS from RSS, STAR, or UAH)

FIG. 8. Determination of the trend length L (yr) at which observed trends in TMT are significantly larger than trends arising from natural internal variability. Results are for temperature data spatially averaged (a),(b) over a near-global domain (82.5°N–82.5°S) and (c),(d) over the tropics (20°N–20°S). Tropospheric temperature data are (left) uncorrected and (right) corrected for stratospheric cooling effects. Rather than assessing the statistical significance of individual observed L-year trends, the results displayed here are (separately for each satellite dataset) the average p values for all maximally overlapping L-year trends. For example, the results shown in (a) (for time scales of 15–21 yr) are simply averages over the index i of the results in Fig. 7, right. Gray shading denotes the rejection region (at a stipulated 10% significance level) for the null hypothesis that the average observed L-year trend could have been generated by internal variability alone. See appendix B for further details.

lower than the UAH results in Fu et al. (2011) and Po-Chedley and Fu (2012b) and are 13%–14% smaller than the CMIP5 $\overline{R}_{MT/LT}$ value. On the longest time scales (35–37 yr), version 6.0 of the UAH TMT_{cr} and TLT datasets yields tropical TMT_{cr}/TLT trend ratios <1 (Fig. 9b). Such behavior is difficult to reconcile with basic physical principles (Stone and Carlson 1979), with model simulations (Po-Chedley and Fu 2012b; Held and Soden 2006; Thorne et al. 2007; Flannaghan et al. 2014), or with satellite estimates of tropical amplification on monthly to interannual time scales (Yulaeva and Wallace 1994; Hegerl and Wallace 2002; Santer et al. 2005; Karl et al. 2006). Taken together, these results suggest that residual errors in the UAH TMT_{cr} and TLT datasets[10] are the most likely

explanation for UAH $\overline{R}_{MT/LT}$ values close to unity, as well as for UAH TMT_{cr} trends that are smaller than surface temperature trends over tropical oceans (Po-Chedley et al. 2015).

7. Summary

We have provided a detailed, updated comparison of atmospheric temperature trends in satellite observations and model simulations. Our study explores the sensitivity of these comparisons to current uncertainties in a number of different factors: climate model simulations of internal variability and the response to external forcing; the satellite datasets chosen; the selected time scale, start, and end dates of temperature trends; and the correction of TMT data for stratospheric cooling. We also examined three issues that have been the focus of scientific attention (National Research Council 2000; Karl et al. 2006;

---

[10] If UAH data were excluded from the calculation of satellite- and time-scale-average trend ratios, $\overline{\overline{R}}$ would be 1.63 for near-global averages of TMT_{cr} and 1.90 for tropical averages of TMT_{cr}.
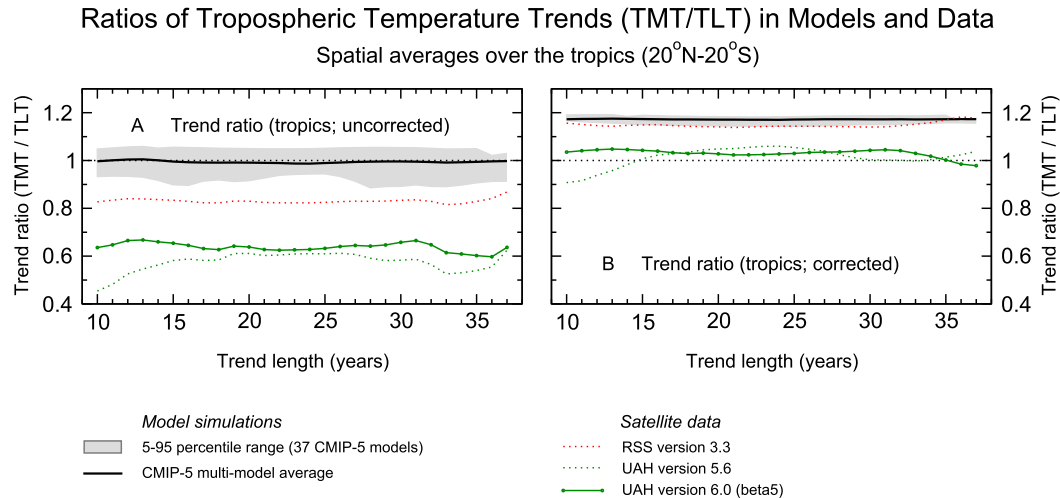
FIG. 9. Values of $R_{MT/LT}$ are calculated for satellite datasets and model externally forced simulations and for (a) TMT and (b) TMT_{cr}. The ratios are simply the uncorrected and corrected tropical TMT trends in Figs. 4a and 4b divided by the tropical TLT trends in Fig. 5b (which do not require correction for lower-stratospheric cooling). The value of $R_{MT/LT}$ is given as a function of the trend length L. As in the case of results shown in Figs. 2, 3, 4, and 5, trends were calculated using all maximally overlapping L-year trends rather than a single L-year trend only. The dotted black line indicates an $R_{MT/LT}$ value of 1. Values above the line denote amplification of lower-tropospheric temperature trends in the mid-to-upper troposphere.

IPCC 2013) and political inquiry (U.S. Senate 2015): 1) the relative sizes of tropospheric warming trends in model simulations and satellite data; 2) the statistical significance of recent tropospheric warming trends; and 3) whether current climate models are capable of capturing the observed amplification of warming in the tropical atmosphere.

With regard to the first issue, we have shown that $\overline{\overline{R}}$ (the ratio between simulated and observed TMT trends)[11] is sensitive to current uncertainties in satellite TMT data and to systematic model–data differences in the size of lower-stratospheric cooling trends. When the impact of lower-stratospheric cooling on TMT is accounted for, and the most recent versions of satellite datasets are used, the previously claimed ratio of 3 between simulated and observed near-global TMT trends (Christy 2015) is reduced to approximately 1.7. In the tropics, correcting for stratospheric cooling and using recent satellite data reduces the reported trend ratio from 4 (Christy 2015) to approximately 2.1. Potential explanations for the remaining model–data differences in warming rates include the combined effects of model response errors (Trenberth and Fasullo 2010), model forcing errors (Solomon et al. 2010, 2011,

2012; Kopp and Lean 2011; Shindell et al. 2013; Hassler et al. 2013; Eyring et al. 2013; Young et al. 2014; Santer et al. 2014; Smith et al. 2016), errors in satellite temperature data (Wentz and Schabel 1998; Mears and Wentz 2005, 2016; Mears et al. 2003, 2011; Zou et al. 2006, 2009; Zou and Wang 2011; Po-Chedley and Fu 2012a; Po-Chedley et al. 2015), and different phasing of internal climate variability in simulations and the observations (Fyfe et al. 2013a, 2016; Kosaka and Xie 2013; Meehl et al. 2011, 2014; England et al. 2014; Risbey et al. 2014; Steinman et al. 2015; Trenberth 2015; Gilford et al. 2016).

The second issue relates to the claim that satellite data show "no significant global warming for the past 18 years" (U.S. Senate 2015). The last 18 years are strongly influenced by the anomalous warmth at the beginning of the period and are not representative of the full 37-yr TMT dataset. In all satellite datasets analyzed here, most 18-yr periods show significant tropospheric warming. But even in the context of the last 18 years, the "no significant warming" claim is invalid: five out of six satellite TMT datasets that have been corrected for stratospheric cooling now yield significant global-scale warming for the most recent 216-month trends.

The third issue—model–data differences in the vertical structure of atmospheric temperature change in the deep tropics—is a long-standing scientific concern (National Research Council 2000; Gaffen et al. 2000;

---

[11] Here, $\overline{\overline{R}}$ represents an average over 1) different analysis time scales and trend start dates; 2) different CMIP5 models and different initial condition realizations of the ALL+8.5 simulation; and 3) different satellite datasets.

Hegerl and Wallace 2002; Santer et al. 2000, 2005; Fu and Johanson 2005; Karl et al. 2006; Held and Soden 2006; Johanson and Fu 2006; Thorne et al. 2007, 2011; Fu et al. 2011; Po-Chedley and Fu 2012b; Flannaghan et al. 2014). Because of moist thermodynamic processes, warming of the tropical ocean surface is amplified aloft, with peak warming in the upper troposphere (Yulaeva and Wallace 1994; Hegerl and Wallace 2002; Santer et al. 2005; Held and Soden 2006). Previous work with shorter temperature records investigated warming of TMT$_{cr}$ relative to the lower troposphere and identified statistically significant differences between simulated and observed amplification behavior in the tropics (Fu et al. 2011; Po-Chedley and Fu 2012b). Such statistically significant differences no longer exist in one updated satellite dataset.

Based on the information presented here, prospects appear to be favorable for reconciling remaining differences in simulated and observed tropospheric temperature trends. Errors in model estimates of key anthropogenic and natural influences are now better understood (Solomon et al. 2010, 2011, 2012; Kopp and Lean 2011; Vernier et al. 2011; Neely et al. 2013; Shindell et al. 2013; Hassler et al. 2013; Eyring et al. 2013; Young et al. 2014). This improved understanding has led to simulations of historical climate with improved representation of forcings (Solomon et al. 2011; Fyfe et al. 2013b; Haywood et al. 2014; Santer et al. 2014; Schmidt et al. 2014). There is also better understanding of the role of different realizations of internal variability in the real world and the "model world" (Fyfe et al. 2013a, 2016; Kosaka and Xie 2013; Meehl et al. 2011, 2014; England et al. 2014; Risbey et al. 2014; Huber and Knutti 2014; Marotzke and Forster 2015; Steinman et al. 2015; Trenberth 2015; Gilford et al. 2016).

On the data side, encouraging progress has been made in identifying nonclimatic artifacts in satellite temperatures and in understanding why different research groups have divergent trend estimates (Wentz and Schabel 1998; Mears and Wentz 2005, 2016; Mears et al. 2003, 2011; Zou et al. 2006, 2009; Zou and Wang 2011; Po-Chedley and Fu 2012b; Po-Chedley et al. 2015). There is real potential to reconcile these "between group" trend differences by applying physically based constraints. Examples of such constraints include adherence to theoretically predicted tropical amplification behavior (Stone and Carlson 1979; Fu and Johanson 2005; Santer et al. 2005; Held and Soden 2006; Karl et al. 2006; Po-Chedley et al. 2015), consistency of amplification ratios across a range of time scales (Yulaeva and Wallace 1994; Hegerl and Wallace 2002; Wentz and Schabel 2000; Santer et al. 2005), and the covariability between tropospheric temperature and independently monitored water vapor (Wentz and Schabel 2000; Mears et al. 2007; Mears and Wentz 2016). The challenge in such complex science is to ensure that the best

scientific understanding is accurately represented to all stakeholders.

APPENDIX A

Method Used for Correcting TMT Data

Trends in TMT estimated from microwave sounders receive a substantial contribution from the cooling of the lower stratosphere (Fu et al. 2004; Fu and Johanson 2004, 2005; Johanson and Fu 2006). Fu et al. (2004) developed a regression-based approach for removing the bulk of this stratospheric cooling component of TMT. Here, we refer to this "corrected" version[A1] of TMT as TMT$_{cr}$. The Fu et al. (2004) correction method has been validated with both observed and model atmospheric temperature data (Fu and Johanson 2004; Gillett et al. 2004; Kiehl et al. 2005).

[A1] In other publications (Fu and Johanson 2005; Po-Chedley et al. 2015), TMT$_{cr}$ is designated as the temperature of the tropical troposphere (TTT) or as T_{24} (since it is generated using brightness temperatures estimated with the emissions measurements obtained from channels 2 and 4 of microwave sounders).

Correction was performed locally at each model and observational grid point. Corrected gridpoint data were then spatially averaged over tropical and near-global domains. For calculating tropical averages of TMT_{cr}, we employed the same regression coefficients used by Fu and Johanson (2005) in their Eq. (1b):

$$TMT_{cr} = a_{24}TMT + (1 - a_{24})TLS, \qquad (A1)$$

where $a_{24} = 1.1$. Subsequent work by Johanson and Fu (2006) obtained very similar estimates[A2] of a_{24} for calculations involving tropical-average TLS and TMT data.

For a near-global domain, TMT trends receive a larger contribution from high-latitude stratospheric cooling,[A3] so a_{24} is larger (Fu et al. 2004; Johanson and Fu 2006). In Fu et al. (2004) and Johanson and Fu (2006), $a_{24} \approx 1.15$ was applied directly to near-global averages of TMT and TLS. Since we are performing corrections on local (gridpoint) data, we used $a_{24} = 1.1$ between 30°N and 30°S and $a_{24} = 1.2$ poleward of 30° in both hemispheres. This is approximately equivalent to using $a_{24} = 1.15$ for globally averaged data. The main text discusses results obtained with this correction method (referred to as M_{lat} in Table 5 of the supplemental material).

As a sensitivity test, we also performed corrections of satellite and model TMT data with $a_{24} = 1.1$ at all latitudes (i.e., with removal of less stratospheric cooling in the extratropics). This has relatively small impact on values of $\overline{R}(k)$, the time-scale average of the model-versus-observed temperature trend ratios for the kth observational dataset (see Table 5 in the supplemental material). These results suggest that the $\overline{R}(k)$ values shown in the main text are robust to different plausible choices of a_{24}.

Finally, we note that model and observational temperature data were processed in exactly the same way; that is, model-versus-observed differences in TMT_{cr} trends are not attributable to differences in the applied regression coefficients.

APPENDIX B

Statistical Analysis

a. Terminology

1) ACRONYMS

MMA Multimodel average.
MMSD Multimodel sampling distribution.

CTL CMIP5 control run with no year-to-year changes in external forcings.
ALL+8.5 CMIP5 historical simulation (with combined natural and anthropogenic forcings) spliced with RCP8.5 run.

2) SUBSCRIPTS

o Satellite observations.
c Output from model control runs.
f Output from model forced experiments.

3) INDICES

i Index over number of maximally overlapping trends in observations.
j Index over number of models (for control run analyses) or over number of models and forced run realizations (for ALL+8.5 analyses).
k Index over number of observed satellite datasets.
l Index over number of selected values of the trend length L (10, 11, . . . , 37).

4) SAMPLE SIZES

L Length of trend-fitting period (yr).
N_L Number of values of L considered.
$N_o(l)$ Number of overlapping trends in observed dataset for lth value of trend length L.
$N_c(l)$ Number of overlapping trends in control run MMSD for lth value of trend length L.
$N_f(l)$ Number of overlapping trends in ALL+8.5 MMSD for lth value of trend length L.
$N_c(j, l)$ Number of overlapping trends in jth model control run for lth value of trend length L.
$N_f(j, l)$ Number of overlapping trends in jth model ALL+8.5 run for lth value of trend length L.
N_{obs} Number of observational datasets (varies according to atmospheric layer considered).
N_{model} Number of models (36 for control runs, 37 for ALL+8.5 runs).

5) SUMMATION VARIABLES

$K_c(i, k, l)$ For ith overlapping L-year segment of time series, kth observational dataset, and lth value of the trend length L, the number of overlapping L-year trends in control run MMSD greater than $b_o(i, k, l)$.

[A2] See Table 1 in Johanson and Fu (2006).
[A3] This is due to two effects: the tropopause is lower at mid-to-high latitudes than in the tropics, and stratospheric cooling over the satellite era is larger at high latitudes than in the tropics (Santer et al. 2013b).

$K_c(i, j, k, l)$ For ith overlapping L-year segment of time series, kth observational dataset, and lth value of the trend length L, the number of overlapping L-year trends in jth model control run greater than $b_o(i, k, l)$.

6) LINEAR TRENDS

$b_o(i, k, l)$ Least-squares linear trend for ith overlapping L-year segment of time series, kth observational dataset, and lth value of the trend length L.

$b_f(i, j, l)$ Least-squares linear trend for ith overlapping L-year segment of time series, jth model ALL+8.5 time series, and lth value of the trend length L.

$\overline{b_o}(k, l)$ Average (over index i) of $b_o(i, k, l)$.

$\overline{b_f}(j, l)$ Average (over index i) of $b_f(i, j, l)$.

$\overline{\overline{b_f}}(l)$ Average (over combined realization and model index j) of $\overline{b_f}(j, l)$.

7) STATISTICS FOR MODEL-VERSUS-OBSERVED TREND COMPARISONS

$p_c(i, k, l)$ Unweighted p value for comparison of $b_o(i, k, l)$ and control run MMSD.

$p_c(i, j, k, l)$ The p value for comparison of $b_o(i, k, l)$ and jth model control run.

$\overline{p_c}(i, k, l)'$ Weighted p value, model average of $p_c(i, j, k, l)$.

$\overline{\overline{p_c}}(k, l)'$ Weighted p value, average over index i of $\overline{p_c}(i, k, l)'$.

8) TREND RATIOS

$R(k, l)$ Model–observed trend ratio, $\overline{\overline{b_f}}(l)/\overline{b_o}(k, l)$.

$\overline{R}(k)$ Average (over all N_L values of the trend length L) of $R(k, l)$.

$\overline{\overline{R}}$ Average (over all N_o observational datasets) of $\overline{R}(k)$.

b. Introduction

We compare trends in spatial averages of model and satellite temperature data.[B1] Although all trends are calculated with monthly mean data, we simplify the discussion by referring to L-year trends (rather than to L-month trends).[B2] Trend comparisons are on time scales ranging from 10 to 37 yr, in increments of 1 yr.

As used here and subsequently, "maximally overlapping trends" indicates that an L-year sliding window is being used for trend calculations, with the window advancing in increments of one month until the end of the current window reaches the final month of the time series. For $L = 10$ yr, for example, the first trend is over January 1979–December 1988, the second trend is over February 1979–January 1989, etc.

Statistical analyses are performed separately for each of the four temperature variables of interest (TLS, TMT, TMT$_{cr}$, and TLT). We do not explicitly include the selected layer-average temperature in our notation. We employ another notational simplification for analysis of the ALL+8.5 simulations: we specify that j is a combined index over models and over realizations of the ALL+8.5 run.[B3] For the control runs, each model analyzed here has one realization of the preindustrial control run, so j is an index over models only.

Anomalies in the ALL+8.5 runs were defined relative to climatological monthly means over the 444-month period from January 1979 to December 2015. Control run anomalies were defined relative to climatological monthly means over the full length of each model's control integration (see Table 4 in the supplemental material).

c. Calculation of p values

We seek to determine whether a selected satellite temperature trend is unusually large relative to model-based estimates of temperature trends arising from natural internal climate variability. Internal variability estimates are obtained from CMIP5 control runs. As in our previous work (Santer et al. 2011), we assess the significance of observed warming trends using both unweighted and weighted p values. Weighted p values are distinguished by the use of prime notation and account for intermodel differences in the length of the control run.

Consider first the unweighted p value $p_c(i, k, l)$:

[B1] For any given trend length L and for each selected analysis period, it is assumed that the externally forced component in a temperature time series is well represented by a linear trend.

[B2] This avoids the less transparent use of 432-month trends, 444-month trends, etc.

[B3] For example, the CCSM4 model has three different realizations of the spliced ALL+8.5 run (see Table 3 in the supplemental material). In the $L = 10$ yr case, and for maximally overlapping trends calculated over January 1979–December 2015, CCSM4 provides 325×3 samples of forced temperature trends for a given atmospheric layer, and $N_f(j, l) = 975$. All 975 trends were used in computing the average of CCSM's sampling distribution of 120-month trends.

$$p_c(i,k,l) = K_c(i,k,l)/N_c(l) \quad \text{for} \quad i = 1,\ldots,N_o(l);$$
$$k = 1,\ldots,N_{\text{obs}}; \quad l = 1,\ldots,N_L, \quad \text{(B1)}$$

where i, k, and l are, respectively, indices over the number of maximally overlapping observed trends, the number of satellite datasets, and the number of selected values of the trend length L. The summation variable $K_c(i,k,l)$ is the number of trends in the MMSD of control run trends that are larger than $b_o(i,k,l)$, the ith overlapping trend for the kth observed dataset and the lth value of the trend length L. The sample sizes $N_c(l)$ and $N_o(l)$ are, respectively, the total number of overlapping trends in the MMSD of control run trends and the total number of overlapping observed trends in the 444-month analysis period. Both $N_c(l)$ and $N_o(l)$ are a function of the selected trend length L. For $L = 10\,\text{yr}$, $N_c(l) = 168\,758$, and $N_o(l) = 325$.

The time series of spatially averaged temperature anomalies from individual models are not concatenated prior to trend calculation (which could spuriously inflate trends spanning the "splice point" between two different model control runs). Instead, overlapping trends are calculated separately for each individual model control run, and each model's temperature trends are then accumulated in a multimodel trend distribution.

In the weighted version, individual $p_c(i,j,k,l)$ values are first calculated separately for each model control run, then summed over all models, and finally averaged:

$$\overline{p_c}(i,k,l)' = \sum_{j=1}^{N_{\text{model}}} p_c(i,j,k,l)/N_{\text{model}} \quad \text{for}$$
$$i = 1,\ldots,N_o(l); \quad k = 1,\ldots,N_{\text{obs}};$$
$$l = 1,\ldots,N_L, \quad \text{(B2)}$$

where j is the index over N_{model}, the number of CMIP5 models with preindustrial control runs from which synthetic MSU temperatures could be calculated (here, $N_{\text{model}} = 36$). The individual $p_c(i,j,k,l)$ values for each model are calculated as follows:

$$p_c(i,j,k,l) = K_c(i,j,k,l)/N_c(j,l) \quad \text{for}$$
$$i = 1,\ldots,N_o(l); \quad j = 1,\ldots,N_{\text{model}};$$
$$k = 1,\ldots,N_{\text{obs}}; \quad l = 1,\ldots,N_L. \quad \text{(B3)}$$

Here, $K_c(i,j,k,l)$ is the number of L-year trends in the preindustrial control run that are larger than $b_o(i,k,l)$ (for the ith overlapping observed trend, the jth model control run, the kth observational dataset, and the lth value of the trend length L).

Values of $p_c(i,k,l)$ and $\overline{p_c}(i,k,l)'$ are very similar, indicating that intermodel differences in control run

length do not distort our estimates of whether observed atmospheric temperature trends are large relative to trends arising from internally generated variability. We only discuss weighted $\overline{p_c}(i,k,l)'$ values in the main text.

Figure 8 displays time scale–average p values. For each of the selected L-year time scales of interest, we simply average the $N_o(l)$ individual values of $\overline{p_c}(i,k,l)'$ over the index i:

$$\overline{\overline{p_c}}(k,l)' = \sum_{i=1}^{N_o(l)} \overline{p_c}(i,k,l)'/N_o(l) \quad \text{for}$$
$$k = 1,\ldots,N_{\text{obs}}; \quad l = 1,\ldots,N_L. \quad \text{(B4)}$$

Our use of maximally overlapping trends has the advantage of reducing the impact of seasonal and interannual noise on underlying atmospheric temperature trends, both in the observations and in the model control runs. However, it has the disadvantage of decreasing the statistical independence of trend samples. While nonindependence of samples is an important issue in formal statistical significance testing, it is not a serious concern here. This is because $\overline{p_c}(i,k,l)'$ is not used as a basis for formal statistical tests. Instead, it simply provides useful information on whether observed atmospheric temperature trends are unusually large relative to model-based estimates of unforced trends. Furthermore, we process observed temperature data and model output in identical ways, with the same overlap between successive L-year trends.

The key point is that whether we employ overlapping or nonoverlapping control run trends has very small impact on estimates of $\overline{p_c}(i,k,l)'$. This suggests that the sample sizes of nonoverlapping trends in the CMIP5 control runs are adequate for obtaining reasonable estimates of p values.[B4]

d. Calculation of trend ratios

Our $R(k,l)$ statistic measures the similarity between temperature trends in externally forced simulations and satellite data. For each observational dataset and L-year time scale of interest, we form the ratio between the model and time scale average of ALL+8.5 trends, and the time scale average of observed trends:

[B4] In the case of the observations, however, the use of nonoverlapping segments of satellite records does not adequately capture the large impact of interannual variability on relatively short trends. For example, the 1979–2015 analysis period contains only three nonoverlapping trends $\geq 10\,\text{yr}$ and $\leq 12\,\text{yr}$, two nonoverlapping trends $\geq 13\,\text{yr}$ and $\leq 18\,\text{yr}$, and only one nonoverlapping trend $\geq 19\,\text{yr}$ and $\leq 37\,\text{yr}$. This is why we focus on maximally overlapping observed trends.

$$R(k, l) = \overline{\overline{b}_f}(l)/\overline{b}_o(k, l) \quad \text{for}$$
$$k = 1, \ldots, N_{\text{obs}}; \quad l = 1, \ldots, N_L. \quad \text{(B5)}$$

The double overbar in $\overline{\overline{b}_f}(l)$ denotes two separate averaging operations. The first averaging step is over the index i (where i runs over the number of maximally overlapping L-year trends in an individual ALL+8.5 realization). This yields $\overline{b}_f(j, l)$, where j is the joint index over ALL+8.5 realizations and CMIP5 models. The second averaging step is first over the number of realizations (for CMIP5 models with more than one ALL+8.5 realization) and then over the number of models with spliced ALL+8.5 runs. The observed mean L-year trend $\overline{b}_o(k, l)$ is defined similarly but only involves averaging over the index i. Both $\overline{\overline{b}_f}(l)$ and $\overline{b}_o(k, l)$ are calculated from temperature time series spanning the same 444-month period (January 1979–December 2015).

Results for $R(k, l)$ are shown in Figs. 2c,d, 3c,d, 4c,d, and 5c,d. In the main text, we also discuss $\overline{R}(k)$. This is simply the average (over all 28 time scales considered)[B5] of the individual $R(k, l)$ values. The observational average of $\overline{R}(k)$ is $\overline{\overline{R}}$. Table 5 of the supplemental material provides values of $\overline{\overline{R}}$ calculated with all observational dataset versions and with newer satellite data only.

REFERENCES

Bindoff, N. L., and Coauthors, 2013: Detection and attribution of climate change: From global to regional. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker, et al., Eds., Cambridge University Press, 867–952. [Available online at https://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5_Chapter10_FINAL.pdf.]

Christy, J. R., 2015: Testimony. Data or dogma? Promoting open inquiry in the debate over the magnitude of human impact on Earth's climate. Hearing in front of the U.S. Senate Committee on Commerce, Science, and Transportation, Subcommittee on Space, Science, and Competitiveness, 8 December 2015. [Available online at https://www.commerce.senate.gov/public/_cache/files/fcbf4cb6-3128-4fdc-b524-7f2ad4944c1d/80931BD995AF75BA7B819A51ADA9CE99.dr.-john-christy-testimony.pdf.]

——, W. B. Norris, R. W. Spencer, and J. J. Hnilo, 2007: Tropospheric temperature change since 1979 from tropical radiosonde and satellite measurements. *J. Geophys. Res.*, **112**, D06102, doi:10.1029/2005JD006881.

Cowtan, K., and R. G. Way, 2014: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quart. J. Roy. Meteor. Soc.*, **140**, 1935–1944, doi:10.1002/qj.2297.

Easterling, D. R., and M. F. Wehner, 2009: Is the climate warming or cooling? *Geophys. Res. Lett.*, **36**, L08706, doi:10.1029/2009GL037810.

England, M. H., and Coauthors, 2014: Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nat. Climate Change*, **4**, 222–227, doi:10.1038/nclimate2106.

Eyring, V., and Coauthors, 2013: Long-term ozone changes and associated climate impacts in CMIP5 simulations. *J. Geophys. Res.*, **118**, 5029–5060, doi:10.1002/jgrd.50316.

Flannaghan, T. J., S. Fueglistaler, I. M. Held, S. Po-Chedley, B. Wyman, and M. Zhao, 2014: Tropical temperature trends in atmospheric general circulation model simulations and the impact of uncertainties in observed SSTs. *J. Geophys. Res. Atmos.*, **119**, 13 327–13 337, doi:10.1002/2014JD022365.

Fu, Q., and C. M. Johanson, 2004: Stratospheric influences on MSU-derived tropospheric temperature trends: A direct error analysis. *J. Climate*, **17**, 4636–4640, doi:10.1175/JCLI-3267.1.

——, and ——, 2005: Satellite-derived vertical dependence of tropical tropospheric temperature trends. *Geophys. Res. Lett.*, **32**, L10703, doi:10.1029/2004GL022266.

——, ——, S. G. Warren, and D. J. Seidel, 2004: Contribution of stratospheric cooling to satellite-inferred tropospheric temperature trends. *Nature*, **429**, 55–58, doi:10.1038/nature02524.

——, S. Manabe, and C. M. Johanson, 2011: On the warming in the tropical upper troposphere: Models versus observations. *Geophys. Res. Lett.*, **38**, L15704, doi:10.1029/2011GL048101.

Fyfe, J. C., N. P. Gillett, and F. W. Zwiers, 2013a: Overestimated global warming over the past 20 years. *Nat. Climate Change*, **3**, 767–769, doi:10.1038/nclimate1972.

——, K. von Salzen, J. N. S. Cole, N. P. Gillett, and J.-P. Vernier, 2013b: Surface response to stratospheric aerosol changes in a coupled atmosphere–ocean model. *Geophys. Res. Lett.*, **40**, 584–588, doi:10.1002/grl.50156.

——, and Coauthors, 2016: Making sense of the early-2000s warming slowdown. *Nat. Climate Change*, **6**, 224–228, doi:10.1038/nclimate2938.

Gaffen, D. J., B. D. Santer, J. S. Boyle, J. R. Christy, N. E. Graham, and R. J. Ross, 2000: Multidecadal changes in the vertical structure of the tropical troposphere. *Science*, **287**, 1242–1245, doi:10.1126/science.287.5456.1242.

Gilford, D., S. Solomon, and R. W. Portman, 2016: Radiative impacts of the 2011 abrupt drops in water vapor and ozone in the tropical tropopause layer. *J. Climate*, **29**, 595–612, doi:10.1175/JCLI-D-15-0167.1.

Gillett, N. P., B. D. Santer, and A. J. Weaver, 2004: Quantifying the influence of stratospheric cooling on satellite-derived tropospheric temperature trends. *Nature*, **432**, doi:10.1038/nature03209.

Hassler, B., P. J. Young, R. W. Portmann, G. E. Bodeker, J. S. Daniel, K. H. Rosenlof, and S. Solomon, 2013: Comparison of three vertically resolved ozone data sets: Climatology, trends and radiative forcings. *Atmos. Chem. Phys.*, **13**, 5533–5550, doi:10.5194/acp-13-5533-2013.

Haywood, J. M., A. Jones, and G. S. Jones, 2014: The impact of volcanic eruptions in the period 2000–2013 on global mean temperature trends evaluated in the HadGEM2-ES climate model. *Atmos. Sci. Lett.*, **15**, 92–96, doi:10.1002/asl2.471.

Hegerl, G. C., and J. M. Wallace, 2002: Influence of patterns of climate variability on the difference between satellite and surface temperature trends. *J. Climate*, **15**, 2412–2428, doi:10.1175/1520-0442(2002)015<2412:IOPOCV>2.0.CO;2.

——, and Coauthors, 2007: Understanding and attributing climate change. *Climate Change 2007: The Physical Science Basis*, S. Solomon, et al., Eds., Cambridge University Press, 663–745. [Available online at https://www.ipcc.ch/pdf/assessment-report/ar4/wg1/ar4-wg1-chapter9.pdf.]

[B5] In other words, over the $R(k, l)$ values for $L = 10, 11, \ldots, 37$.

Held, I. M., and B. J. Soden, 2006: Robust responses of the hydrological cycle to global warming. *J. Climate*, **19**, 5686–5699, doi:10.1175/JCLI3990.1.

Huber, M., and R. Knutti, 2014: Natural variability, radiative forcing and climate response in the recent hiatus reconciled. *Nat. Geosci.*, **7**, 651–656, doi:10.1038/ngeo2228.

Hurrell, J. W., and K. E. Trenberth, 1998: Difficulties in obtaining reliable temperature trends: Reconciling the surface and satellite Microwave Sounding Unit records. *J. Climate*, **11**, 945–967, doi:10.1175/1520-0442(1998)011<0945:DIORTT>2.0.CO;2.

IPCC, 2013: Summary for policymakers. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker, et al., Eds., Cambridge University Press, 1–29. [Available online at https://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WGIAR5_SPM_brochure_en.pdf.]

Johanson, C. M., and Q. Fu, 2006: Robustness of tropospheric temperature trends from MSU channels 2 and 4. *J. Climate*, **19**, 4234–4242, doi:10.1175/JCLI3866.1.

Karl, T. R., S. J. Hassol, C. D. Miller, and W. L. Murray, Eds., 2006: Temperature trends in the lower atmosphere: Steps for understanding and reconciling differences. U.S. Climate Change Science Program and the Subcommittee on Global Change Research Synthesis and Assessment Product 1.1, 164 pp.

——, and Coauthors, 2015: Possible artifacts of data biases in the recent global surface warming hiatus. *Science*, **348**, 1469–1472, doi:10.1126/science.aaa5632.

Kiehl, J. T., J. Caron, and J. J. Hack, 2005: On using global climate model simulations to assess the accuracy of MSU retrieval methods for tropospheric warming trends. *J. Climate*, **18**, 2533–2539, doi:10.1175/JCLI3492.1.

Kopp, G., and J. L. Lean, 2011: A new, lower value of total solar irradiance: Evidence and climate significance. *Geophys. Res. Lett.*, **38**, L01706, doi:10.1029/2010GL045777.

Kosaka, Y., and S.-P. Xie, 2013: Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature*, **501**, 403–407, doi:10.1038/nature12534.

Lewandowsky, S., J. S. Risbey, and N. Oreskes, 2016: The "pause" in global warming: Turning a routine fluctuation into a problem for science. *Bull. Amer. Meteor. Soc.*, **97**, 723–733, doi:10.1175/BAMS-D-14-00106.1.

Lott, F. C., P. A. Stott, D. M. Mitchell, N. Christidis, N. P. Gillett, L. Haimberger, J. Perlwitz, and P. W. Thorne, 2013: Models versus radiosondes in the free atmosphere: A new detection and attribution analysis of temperature. *J. Geophys. Res.*, **118**, 2609–2619, doi:10.1002/jgrd.50255.

Marotzke, J., and P. M. Forster, 2015: Forcing, feedback and internal variability in global temperature trends. *Nature*, **517**, 565–570, doi:10.1038/nature14117.

Mears, C. A., and F. J. Wentz, 2005: The effect of diurnal correction on satellite-derived lower tropospheric temperature. *Science*, **309**, 1548–1551, doi:10.1126/science.1114772.

——, and ——, 2016: Sensitivity of satellite-derived tropospheric temperature trends to the diurnal cycle adjustment. *J. Climate*, **29**, 3629–3646, doi:10.1175/JCLI-D-15-0744.1.

——, M. C. Schabel, and F. J. Wentz, 2003: A reanalysis of the MSU channel 2 tropospheric temperature record. *J. Climate*, **16**, 3650–3664, doi:10.1175/1520-0442(2003)016<3650:AROTMC>2.0.CO;2.

——, B. D. Santer, F. J. Wentz, K. E. Taylor, and M. Wehner, 2007: The relationship between temperature and precipitable water changes over tropical oceans. *Geophys. Res. Lett.*, **34**, L2470, doi:10.1029/2007GL031936.

——, F. J. Wentz, P. Thorne, and D. Bernie, 2011: Assessing uncertainty in estimates of atmospheric temperature changes from MSU and AMSU using a Monte-Carlo estimation technique. *J. Geophys. Res.*, **116**, D08112, doi:10.1029/2010JD014954.

Meehl, G. A., J. M. Arblaster, J. T. Fasullo, A. Hu, and K. E. Trenberth, 2011: Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods. *Nat. Climate Change*, **1**, 360–364, doi:10.1038/nclimate1229.

——, H. Teng, and J. M. Arblaster, 2014: Climate model simulations of the observed early-2000s hiatus of global warming. *Nat. Climate Change*, **4**, 898–902, doi:10.1038/nclimate2357.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophys. Res.*, **117**, D08101, doi:10.1029/2011JD017187.

National Research Council, 2000: *Reconciling Observations of Global Temperature Change*. National Academy Press, 169 pp.

Neely, R. R., and Coauthors, 2013: Recent anthropogenic increases in SO_2 from Asia have minimal impact on stratospheric aerosol. *Geophys. Res. Lett.*, **40**, 999–1004, doi:10.1002/grl.50263.

Pidcock, R., 2016: Analysis: How much did El Niño boost global temperature in 2015? [http://www.carbonbrief.org/analysis-how-much-did-el-nino-boost-global-temperature-in-2015.]

Po-Chedley, S., and Q. Fu, 2012a: A bias in the midtropospheric channel warm target factor on the *NOAA-9* Microwave Sounding Unit. *J. Atmos. Oceanic Technol.*, **29**, 646–652, doi:10.1175/JTECH-D-11-00147.1.

——, and ——, 2012b: Discrepancies in tropical upper tropospheric warming between atmospheric circulation models and satellites. *Environ. Res. Lett.*, **7**, 044018, doi:10.1088/1748-9326/7/4/044018.

——, T. J. Thorsen, and Q. Fu, 2015: Removing diurnal cycle contamination in satellite-derived tropospheric temperatures: Understanding tropical tropospheric trend discrepancies. *J. Climate*, **28**, 2274–2290, doi:10.1175/JCLI-D-13-00767.1.

Ramaswamy, V., M. D. Schwarzkopf, W. J. Randel, B. D. Santer, B. J. Soden, and G. L. Stenchikov, 2006: Anthropogenic and natural influences in the evolution of lower stratospheric cooling. *Science*, **311**, 1138–1141, doi:10.1126/science.1122587.

Risbey, J. S., S. Lewandowsky, C. Langlais, D. P. Monselesan, T. J. O'Kane, and N. Oreskes, 2014: Well-estimated global surface warming in climate projections selected for ENSO phase. *Nat. Climate Change*, **4**, 835–840, doi:10.1038/nclimate2310.

Robock, A., 2000: Volcanic eruptions and climate. *Rev. Geophys.*, **38**, 191–219, doi:10.1029/1998RG000054.

Santer, B. D., T. M. L. Wigley, T. P. Barnett, and E. Anyamba, 1995: Detection of climate change and attribution of causes. *Climate Change 1995: The Science of Climate Change*, J. T. Houghton, et al., Eds., Cambridge University Press, 407–443.

——, and Coauthors, 2000: Interpreting differential temperature trends at the surface and in the lower troposphere. *Science*, **287**, 1227–1232, doi:10.1126/science.287.5456.1227.

——, and Coauthors, 2001: Accounting for the effects of volcanoes and ENSO in comparisons of modeled and observed temperature trends. *J. Geophys. Res.*, **106**, 28 033–28 059, doi:10.1029/2000JD000189.

——, and Coauthors, 2005: Amplification of surface temperature trends and variability in the tropical atmosphere. *Science*, **309**, 1551–1556, doi:10.1126/science.1114867.

——, and Coauthors, 2008: Consistency of modelled and observed temperature trends in the tropical troposphere. *Int. J. Climatol.*, **28**, 1703–1722, doi:10.1002/joc.1756.

——, and Coauthors, 2011: Separating signal and noise in atmospheric temperature changes: The importance of timescale. *J. Geophys. Res.*, **116**, D22105, doi:10.1029/2011JD016263.

——, and Coauthors, 2013a: Human and natural influences on the changing thermal structure of the atmosphere. *Proc. Natl. Acad. Sci. USA*, **110**, 17 235–17 240, doi:10.1073/pnas.1305332110.

——, and Coauthors, 2013b: Identifying human influences on atmospheric temperature. *Proc. Natl. Acad. Sci. USA*, **110**, 26–33, doi:10.1073/pnas.1210514109.

——, and Coauthors, 2014: Volcanic contribution to decadal changes in tropospheric temperature. *Nat. Geosci.*, **7**, 185–189, doi:10.1038/ngeo2098.

Schmidt, G. A., D. T. Shindell, and K. Tsigaridis, 2014: Reconciling warming trends. *Nat. Geosci.*, **7**, 158–160, doi:10.1038/ngeo2105.

Seidel, D. J., and Coauthors, 2016: Stratospheric temperature changes during the satellite era. *J. Geophys. Res. Atmos.*, **121**, 664–681, doi:10.1002/2015JD024039.

Shindell, D. T., and Coauthors, 2013: Radiative forcing in the ACCMIP historical and future climate simulations. *Atmos. Chem. Phys.*, **13**, 2939–2974, doi:10.5194/acp-13-2939-2013.

Smith, D. M., and Coauthors, 2016: Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nat. Climate Change*, **6**, 936–940, doi:10.1038/nclimate3058.

Solomon, S., 1999: Stratospheric ozone depletion: A review of concepts and history. *Rev. Geophys.*, **37**, 275–316, doi:10.1029/1999RG900008.

——, K. H. Rosenlof, R. W. Portman, J. S. Daniel, S. M. Davis, T. J. Sanford, and G.-K. Plattner, 2010: Contributions of stratospheric water vapor to decadal changes in the rate of global warming. *Science*, **327**, 1219–1223, doi:10.1126/science.1182488.

——, J. S. Daniel, R. R. Neely, J.-P. Vernier, E. G. Dutton, and L. W. Thomason, 2011: The persistently variable "background" stratospheric aerosol layer and global climate change. *Science*, **333**, 866–870, doi:10.1126/science.1206027.

——, P. J. Young, and B. Hassler, 2012: Uncertainties in the evolution of stratospheric ozone and implications for recent temperature changes in the tropical lower stratosphere. *Geophys. Res. Lett.*, **39**, L17706, doi:10.1029/2012GL052723.

Spencer, R. W., and J. R. Christy, 1992: Precision and radiosonde validation of satellite gridpoint temperature anomalies. Part II: A tropospheric retrieval and trends during 1979–1990. *J. Climate*, **5**, 858–866, doi:10.1175/1520-0442(1992)005<0858:PARVOS>2.0.CO;2.

Steinman, B. A., M. E. Mann, and S. K. Miller, 2015: Atlantic and Pacific multidecadal oscillations and Northern Hemisphere temperatures. *Science*, **347**, 988–991, doi:10.1126/science.1257856.

Stone, P. H., and J. H. Carlson, 1979: Atmospheric lapse rate regimes and their parameterization. *J. Atmos. Sci.*, **36**, 415–423, doi:10.1175/1520-0469(1979)036<0415:ALRRAT>2.0.CO;2.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Thompson, D. W. J., J. M. Wallace, P. D. Jones, and J. J. Kennedy, 2009: Identifying signatures of natural climate variability in time series of global-mean surface temperature: Methodology and insights. *J. Climate*, **22**, 6120–6141, doi:10.1175/2009JCLI3089.1.

Thorne, P. W., and Coauthors, 2007: Tropical vertical temperature trends: A real discrepancy? *Geophys. Res. Lett.*, **34**, L16702, doi:10.1029/2007GL029875.

——, J. R. Lanzante, T. C. Peterson, D. J. Seidel, and K. P. Shine, 2011: Tropospheric temperature trends: History of an ongoing controversy. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 66–88, doi:10.1002/wcc.80.

Tollefson, J., 2016: 2015 breaks heat record. *Nature*, **529**, 450.

Trenberth, K. E., 2015: Has there been a hiatus? *Science*, **349**, 691–692, doi:10.1126/science.aac9225.

——, and J. T. Fasullo, 2010: Simulation of present-day and twenty-first-century energy budgets of the southern oceans. *J. Climate*, **23**, 440–454, doi:10.1175/2009JCLI3152.1.

U.S. Senate, 2015: Data or dogma? Promoting open inquiry in the debate over the magnitude of human impact on Earth's climate. Archived webcast of hearing before the U.S. Senate Committee on Commerce, Science, and Transportation, Subcommittee on Space, Science, and Competitiveness, 8 December 2015. [Available online at http://www.commerce.senate.gov/public/index.cfm/2015/12/data-or-dogma-promoting-open-inquiry-in-the-debate-over-the-magnitude-of-human-impact-on-earth-s-climate.]

Vernier, J.-P., and Coauthors, 2011: Major influence of tropical volcanic eruptions on the stratospheric aerosol layer during the last decade. *Geophys. Res. Lett.*, **38**, L12807, doi:10.1029/2011GL047563.

Wentz, F. J., and M. Schabel, 1998: Effects of orbital decay on satellite-derived lower-tropospheric temperature trends. *Nature*, **394**, 661–664, doi:10.1038/29267.

——, and ——, 2000: Precise climate monitoring using complementary satellite data sets. *Nature*, **403**, 414–416, doi:10.1038/35000184.

Wigley, T. M. L., C. M. Ammann, B. D. Santer, and S. C. B. Raper, 2005: The effect of climate sensitivity on the response to volcanic forcing. *J. Geophys. Res.*, **110**, D09107, doi:10.1029/2004JD005557.

Young, P. J., S. M. Davis, B. Hassler, S. Solomon, and K. H. Rosenlof, 2014: Modeling the climate impact of Southern Hemisphere ozone depletion: The importance of the ozone data set. *Geophys. Res. Lett.*, **41**, 9033–9039, doi:10.1002/2014GL061738.

Yulaeva, E., and J. M. Wallace, 1994: The signature of ENSO in global temperature and precipitation fields derived from the Microwave Sounding Unit. *J. Climate*, **7**, 1719–1736, doi:10.1175/1520-0442(1994)007<1719:TSOEIG>2.0.CO;2.

Zou, C.-Z., and W. Wang, 2011: Intersatellite calibration of AMSU-A observations for weather and climate applications. *J. Geophys. Res.*, **116**, D23113, doi:10.1029/2011JD016205.

——, M. D. Goldberg, Z. Cheng, N. C. Grody, J. T. Sullivan, C. Cao, and D. Tarpley, 2006: Recalibration of microwave sounding unit for climate studies using simultaneous nadir overpasses. *J. Geophys. Res.*, **111**, D19114, doi:10.1029/2005JD006798.

——, M. Gao, and M. Goldberg, 2009: Error structure and atmospheric temperature trends in observations from the Microwave Sounding Unit. *J. Climate*, **22**, 1661–1681, doi:10.1175/2008JCLI2233.1.