Amplification of Surface Temperature Trends and Variability in the Tropical Atmosphere

B.D. Santer,[1]* T.M.L. Wigley,[2] C. Mears,[3] F.J. Wentz,[3] S.A. Klein,[1] D.J. Seidel,[4] K.E. Taylor,[1] P.W. Thorne,[5] M.F. Wehner,[6] P.J. Gleckler,[1] J.S. Boyle,[1] W.D. Collins,[2] K.W. Dixon,[7] C. Doutriaux,[1] M. Free,[4] Q. Fu,[8] J.E. Hansen,[9] G.S. Jones,[5] R. Ruedy,[9] T.R. Karl,[10] J.R. Lanzante,[7] G.A. Meehl,[2] V. Ramaswamy,[7] G. Russell,[9] and G.A. Schmidt[9]

[1] Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA; [2] National Center for Atmospheric Research, Boulder, CO 80303, USA; [3] Remote Sensing Systems, Santa Rosa, CA 95401, USA; [4] NOAA/Air Resources Laboratory, Silver Spring, MD 20910, USA; [5] Hadley Centre for Climate Prediction and Research, U.K. Met. Office, Exeter, EX1 3PB, UK; [6] Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA; [7] NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, NJ 08542, USA; [8] Department of Atmospheric Sciences, University of Washington, Seattle, WA 98195, USA; [9] NASA/Goddard Institute for Space Studies, New York, NY 10025, USA; [10] NOAA/National Climatic Data Center, Asheville, NC 28801, USA.

Submitted to *Science*, May 13, 2005. Revised: July 19, 2005

* To whom correspondence should be addressed. Email: santer1@llnl.gov

The month-to-month variability of tropical temperatures is larger in the troposphere than at the Earth's surface. This amplification behavior is similar in a range of observations and climate model simulations, and is consistent with basic theory. On multi-decadal timescales, tropospheric amplification of surface warming is a robust feature of model simulations, but occurs in only one observational dataset. Other observations show weak or even negative amplification. These results suggest that either different physical mechanisms control amplification processes on monthly and decadal timescales, and models fail to capture such behavior, or (more plausibly) that residual errors in several observational datasets used here affect their representation of long-term trends.

Tropospheric warming is a robust feature of climate model simulations driven by historical increases in greenhouse gases (*1–3*). Maximum warming is predicted to occur in the middle and upper tropical troposphere. Atmospheric temperature measurements from radiosondes also show warming of the tropical troposphere since the early 1960s (*4–7*), consistent with model results (*8*). The observed tropical warming arises in part from a step-like change in the late 1970s (*5*, *6*).

Considerable attention has focused on the shorter record of satellite-based atmospheric temperature measurements (1979 to present). In both models and observations, the tropical surface warms over this period. Simulated surface warming is amplified in the tropical troposphere, corresponding to a decrease in lapse rate (*2*, *3*, *9*). In contrast, a number of radiosonde and satellite datasets suggest that the tropical troposphere has warmed *less* than the surface, or even cooled, which would correspond to an increase in lapse rate (*4–12*).

This discrepancy may be an artifact of residual inhomogeneities in the observations (*13–19*). Creating homogeneous climate records requires the identification and removal of non-climatic influences from data that were primarily collected for weather forecasting purposes. Different analysts have followed very different data adjustment pathways (*4–7*, *12*, *14*, *17*). The resulting 'structural uncertainties' in observed estimates of tropospheric temperature change (*20*) are as large as the model-predicted climate-change signal that should have occurred in response to combined human and

natural forcings (*16*).

Alternately, there may be a real disparity between modeled and observed lapse-rate changes over the satellite era (*9–11*, *21*). This would point towards the existence of fundamental deficiencies in current climate models (and/or in the forcings employed in model experiments), thus diminishing our confidence in model predictions of climate change.

This scientific puzzle provides considerable motivation for revisiting comparisons of simulated and observed tropical lapse-rate changes (*10*, *13*, *21*, *22*) using more comprehensive estimates of observational uncertainty and a wide range of recently-completed model simulations. The latter were performed in support of the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), and involve 19 coupled atmosphere-ocean models developed in nine different countries. Unlike previous model intercomparison exercises involving idealized climate change experiments (*23*), these new simulations incorporate estimated historical changes in a variety of natural and anthropogenic forcings (*24*) (see Supporting Online Material; "SOM").

Our focus is on the amplification of surface temperature variability and trends in the free troposphere. We study this amplification behavior in several different ways. The first is to compare atmospheric profiles of "scaling ratios" in the IPCC simulations and in two new radiosonde datasets: HadAT2 (Hadley Centre Atmospheric Temper-

atures, version 2) and RATPAC (Radiosonde Atmospheric Temperature Products for Assessing Climate). These were compiled (respectively) by the United Kingdom Met. Office (UKMO) (*6*) and the National Oceanic and Atmospheric Administration (NOAA) (*7*). The scaling factor is simply the ratio between the temperature variability (or trend) at discrete atmospheric pressure levels and the same quantity at the surface (*25*). Observed trends and variability in tropical surface temperatures (T_S) were obtained from the NOAA (*26*) and HadCRUT2v datasets (*27*).

Our second method for estimating scaling ratios uses the weighted-average temperatures of deep atmospheric layers (*12*, *17*). These are available from the satellite-based Microwave Sounding Unit (MSU), which monitors atmospheric microwave emissions from the lower stratosphere (T_4) and the troposphere (T_2). MSU T_2 data have also been used to retrieve lower tropospheric temperatures ($T_{2\mathrm{LT}}$). We calculate synthetic MSU temperatures from the IPCC simulations, and then compare these with actual MSU temperatures produced by research groups at the University of Alabama in Huntsville (UAH) (*12*) and Remote Sensing Systems (RSS) in California (*14*, *17*). Synthetic T_4, T_2, and $T_{2\mathrm{LT}}$ data are also computed from the HadAT2 and RATPAC radiosonde datasets (see SOM).

T_2 receives a contribution from the cooling stratosphere (*28*). This hampers its use for estimating the amplification of surface temperature changes in the free troposphere. We therefore focus on $T_{2\mathrm{LT}}$, which is relatively unaffected by the stratosphere

(*15*). Until recently, only UAH provided a satellite-based T_{2LT} product (*12*). The RSS group has now independently derived a second T_{2LT} dataset (*14*).

Another strategy for removing stratospheric influences on T_2 relies on a linear combination of T_4 and T_2 (*15*) (see SOM). This procedure yields T_{Fu}, which is representative of temperatures in the bulk troposphere (1000 to 100 hPa). Relative to T_{2LT}, T_{Fu} receives more of its signal from higher regions of the troposphere (*15*). Based on simple moist adiabatic lapse rate (MALR) theory (*29*), we expect scaling ratios in the deep tropics to increase with increasing height, and to peak at roughly 200 hPa. Comparison of the amplification factors estimated with T_{2LT} and T_{Fu} data allows us to verify whether models and observations confirm this theoretical expectation.

Before discussing the scaling ratio results, it is instructive to examine the variability and trends in layer-averaged atmospheric temperatures and T_S. Our analysis period (January 1979 through December 1999) is constrained by the start date of observed satellite data and the end date of the IPCC historical forcing experiment. A total of 49 realizations of this experiment were available (*24*).

Time series of tropical T_4 changes in UAH, RSS, and the IPCC simulations are characterized by overall cooling trends and volcanically-induced stratospheric warming signals (Fig. 1A). High-frequency variability associated with the Quasi-Biennial Oscillation (QBO) is evident in the observations, but not in the model simulations (*5*) (see SOM). Satellite T_4 trends lie within the range of model results, but the larger

cooling trends estimated from radiosondes do not (Fig. 2A). Part of this discrepancy may be due to residual stratospheric and upper tropospheric cooling biases in the tropical radiosonde data (*18*, *19*).

In observations, the tropical variability of tropospheric and surface temperatures is dominated by the large El Niño events in 1982/83, 1987/88, and 1997/98 (Figs. 1B,C). Since the IPCC runs are coupled model simulations, they cannot reproduce the time sequence of observed El Niño and La Niña events, except by chance (*2*, *16*). The range of simulated El Niño/Southern Oscillation (ENSO) variability spans an order of magnitude. Models with very strong ENSO variability have fluctuations in surface and tropospheric temperatures that are noticeably larger than observed.

The observed tropical T_S trends in the NOAA and HadCRUT2v datasets (0.12 and 0.14°C/decade, respectively) are very similar to $\overline{X}$, the average warming over all model simulations (Fig. 2E) (*30*). In the troposphere, however, model-observed trend agreement is sensitive to the atmospheric layer examined and the choice of observational dataset. In both radiosonde datasets used here, T_2 *cools* over 1979 to 1999, and trends are outside the spread of model results (Fig. 2B). Large stratospheric cooling biases in the radiosonde data probably contribute to this disparity (*18*, *19*). Use of T_{Fu} removes most of the stratospheric influence on T_2, and yields positive temperature trends in all observed datasets (Fig. 2C) (*5*, *15*, *28*). All observed T_{Fu} trends are within the envelope of model values.

In the tropical lower troposphere, all datasets except UAH have positive T_{2LT} trends (Fig. 2D). The difference between the UAH and RSS trends (*ca.* 0.13°C/decade) is a factor of two larger than the claimed 95% confidence interval for the UAH global T_{2LT} trend (*12*). This difference is primarily attributable to the different ways in which the two groups account for the effects of orbital drift on the sampling of the diurnal temperature cycle (*14*). The UAH T_{2LT} trend lies outside the range of model solutions. The disparate behavior of T_{2LT} and T_{Fu} in the UAH data (the former cools, while the latter warms) is not evident in any other dataset (*14*, *15*, *28*).

Both model and satellite data indicate that variability in T_S is amplified in the tropical troposphere (Figs. 1B,C). Amplification of surface warming is a direct result of moist thermodynamic processes (*29*). We examine two different amplification metrics: $R_S(z)$, the ratio between the temporal standard deviations of monthly-mean tropospheric and T_S anomalies, and $R_\beta(z)$, the ratio between the multi-decadal trends in these quantities. Since most of the monthly-timescale variability in tropical surface and tropospheric temperatures is driven by interannual fluctuations in ENSO, $R_S(z)$ largely reflects amplification processes acting on annual timescales (*31*) (see Fig. S1).

Figure 3A shows $R_S(z)$ values in models and radiosondes. The theoretically-expected profile is also displayed (*32*). In all cases, $R_S(z)$ increases above the boundary layer, with maximum amplification at roughly 200 hPa. Below roughly 400 hPa, there is close agreement between the scaling ratios in models, radiosondes and theory.

Between 400 and 150 hPa, the theoretical scaling ratios are consistently larger than in either the radiosondes or the IPCC simulations. Such departures may be due to the fact that MALR theory is applicable to regions of the tropical ocean experiencing deep convection. In contrast, the model and radiosonde temperature data used to calculate $R_S(z)$ include many convectively-inactive areas, where the surface air temperature change is not constrained by the moist adiabat set by the convectively-active regions. Furthermore, active moist convection does not always penetrate above 400 hPa, which would weaken the connection to a moist adiabat above this level.

When scaling ratios are calculated for multi-decadal linear trends, both radiosonde datasets are clear outliers. HadAT2 and RATPAC $R_\beta(z)$ values never exceed 0.82, indicating *damping* of the surface warming trend in the free atmosphere (Fig. 3B). None of the 49 model realizations demonstrate such behavior. The shapes of the radiosonde-based scaling ratio profiles also differ from model and theoretical results, with peak values at generally lower atmospheric levels. Note that subsampling the HadCRUT2v T_S data at the locations of the HadAT2 radiosonde stations has little impact on the observed $R_S(z)$ or $R_\beta(z)$ values (see SOM).

In the low- to mid-troposphere, model $R_\beta(z)$ results are in good agreement with theoretical expectations. Model scaling ratios are therefore consistent with theory on both monthly and multi-decadal timescales, while the radiosonde data are only consistent with theory on monthly timescales.

A qualitatively similar picture emerges from scatter plots of the individual components of $R_S(z)$ and $R_\beta(z)$ (Fig. 4). These display scaling behavior for layer-averaged atmospheric temperatures rather than for temperatures at discrete atmospheric levels. Figure 4A shows $s\{T_S\}$ and $s\{T_{2\mathrm{LT}}\}$, the temporal standard deviations of monthly-mean tropical T_S and $T_{2\mathrm{LT}}$ data. Both vary by a factor of ≥ 5 over the 19 IPCC models. Values of $s\{T_{Fu}\}$ span a comparable range (Fig. 4B). These large ranges are primarily dictated by model differences in the amplitude of El Niño/Southern Oscillation (ENSO) variability.

Despite this large spread of model variability estimates, the tropospheric amplification of $s\{T_S\}$ is internally consistent across a wide range of models and observed data (Figs. 4A,B). The regression between the model $s\{T_S\}$ and $s\{T_{2\mathrm{LT}}\}$ values has a slope of 1.3, in accord with the theoretically-expected scaling ratio at the peak of the $T_{2\mathrm{LT}}$ weighting function. The regression line for $s\{T_S\}$ and $s\{T_{Fu}\}$ is steeper (1.5). This is because the T_{Fu} weighting function peaks higher in the atmosphere, where scaling ratios are larger (Fig. 3A) (see SOM). All model and observational results in Figs. 4A,B are tightly clustered around the fitted (red) regression lines, consistent with the close model-observed agreement in lower tropospheric values of $R_S(z)$ (Fig. 3A).

Amplification factors estimated from multi-decadal trends in T_S, $T_{2\mathrm{LT}}$, and T_{Fu} also display considerable internal consistency in the 19 IPCC models (Figs. 4C,D).

This consistency occurs despite large inter-model differences in convective parameterizations, boundary layer formulation, and resolution, all of which affect the simulation of tropical convection and tropospheric lapse rates. Furthermore, the model-model consistency in $R_\beta(z)$ ratios is robust to differences in the natural and anthropogenic forcings applied by each group (*24*) (see SOM). Many of these forcings are heterogeneous in space and time (*2*, *3*, *33*). These differences in forcings and physics do not cause significant displacement of model results from the regression line in Figs. 4C and 4D. The regression slopes in Figs. 4C and D are similar to those estimated from monthly-timescale variability, with T_{Fu} results again yielding a steeper slope than $T_{2\mathrm{LT}}$ (*c.f.* Figs. 4A,B).

The real conundrum in Fig. 4 is the complex behavior of the observations. On monthly timescales, the amplification behavior of models and observations is consistent. On decadal timescales, however, only the RSS-based $T_{2\mathrm{LT}}$ and T_{Fu} trends have scaling factors that are in reasonable accord with model results (Figs. 4C,D) (see SOM). Despite sustained warming of the tropical land and ocean surfaces, the UAH $T_{2\mathrm{LT}}$ trend is *negative* – *i.e.*, $R_\beta(z) < 0$. The UAH $R_\beta(z)$ value seems physically implausible (*14*, *15*). Prolonged surface warming should destabilize tropical temperature profiles, thus enhancing conditions for moist convection and readjustment of atmospheric temperatures to a MALR.

In contrast to the model results and theoretical expectations, both radiosonde

datasets used here have $R_\beta(z)$ ratios $\ll 1.0$ (Figs. 4C,D). As in the case of the satellite datasets, there are large structural uncertainties in radiosonde estimates of tropospheric temperature change (*4–7*). Comparisons of tropical temperature data from day- and night-time radiosonde ascents suggest that the error arising from solar heating of temperature sensors has decreased over time (*18*, *19*). Inadequate correction for this effect may account for a residual cooling bias in tropospheric temperature changes.

The existence of residual inhomogeneities in the observational data is entirely likely. Current atmospheric observing systems were designed for real-time monitoring of weather rather than long-term monitoring of climate. The construction of reliable climate records from radiosondes is hampered by the above-noted changes in instrumentation (*18*, *19*), along with changes in observing practices and network density (*4–7*, *11*, *13*). Similar concerns apply to satellite data, which are influenced by intersatellite biases, orbital drift and decay, and uncertainties in instrument calibration coefficients (*11–14*, *17*).

Adjustments for these and other effects are applied at discrete points in an observational time series, such as times of transition to a new satellite. None of these corrections is precisely known. Small errors in adjustments can introduce systematic errors in the time series. These errors have little impact on monthly and interannual variability, which accounts for most of the variance of tropospheric temperature

fluctuations in the deep tropics (Fig. 1B). However, systematic errors can have a significant effect on interdecadal variability. This helps to explain why model/data comparisons of $R_\beta(z)$ ratios are sensitive to observational uncertainty, while $R_S(z)$ ratios are not.

In summary, we have demonstrated that all observed datasets and model results are remarkably consistent in terms of their relationship between monthly- and annual-timescale temperature variations at the surface and in the free troposphere. This is a strong verification of the model physics that governs the amplification of tropical surface temperature changes. On decadal timescales, however, only one observed dataset (RSS) shows amplification behavior that is generally consistent with model results. The correspondence between models and observations on monthly and annual timescales does not guarantee that model scaling ratios are valid on decadal timescales. However, given the very basic nature of the physics involved, this high-frequency agreement is suggestive of more general validity of model scaling ratios across a range of timescales.

The RSS $T_{2\mathrm{LT}}$, T_2, and T_{Fu} trends are physically consistent (all three layers warm as the surface warms), while the UAH data show trends of different sign in the lower- and mid-troposphere. These results support the contention that the tropical warming trend in RSS $T_{2\mathrm{LT}}$ data is more reliable than $T_{2\mathrm{LT}}$ trends in other observational datasets. This conclusion does not rest solely on comparisons with climate models.

It is independently supported by the empirical evidence of recent increases in tropospheric water vapor and tropopause height (*25*, *34*), which are accord with warming but not cooling of the free troposphere.

We have employed basic physical principles, as represented in current climate models, for interpreting and evaluating observational data. Our work illustrates that progress toward an improved understanding of the climate system can best be achieved by combined use of observations, theory, and models. The availability of a large range of model and observational surface and atmospheric temperature datasets has been of great benefit to this research, and highlights the dangers inherent in drawing inferences on agreement between models and observations without adequately accounting for uncertainties in both.

References and Notes

1. B. D. Santer *et al.*, *Nature* **382**, 39 (1996).

2. J. E. Hansen *et al.*, *J. Geophys. Res.* **107**, ACL-2, doi:10.1029/2001JD001143 (2002).

3. S. F. B. Tett *et al.*, *J. Geophys. Res.* **107**, doi 10.1029/2000JD000028 (2002).

4. J. R. Lanzante, S. A. Klein, D. J. Seidel, *J. Climate* **16**, 241 (2003).

5. D. J. Seidel *et al.*, *J. Clim.* **17**, 2225 (2004).

6. P. W. Thorne *et al.*, *J. Geophys. Res.* (accepted).

7. M. Free, D. J. Seidel, J. K. Angell, J. Lanzante, I. Durre, T. C. Peterson, Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC): A new dataset of large-area anomaly time series. *J. Geophys. Res.* (submitted).

8. P. W. Thorne *et al.*, *Geophys. Res. Lett.* **29**, doi 10.1029/2002GL015717 (2002).

9. B. D. Santer *et al.*, *Science* **287**, 1227 (2000).

10. D. J. Gaffen *et al.*, *Science* **287**, 1242 (2000).

11. J. M. Wallace *et al.*, National Research Council, Board on Atmospheric Sciences and Climate, *Reconciling Observations of Global Temperature Change*, National Academy Press, Washington D.C., 85 pp. (2000).

12. J. R. Christy, R. W. Spencer, W. B. Norris, W. D. Braswell, *J. Atmos. Ocean. Tech.* **20**, 613 (2003).

13. J. W. Hurrell, K. E. Trenberth, *J. Clim.* **11**, 945 (1998).

14. C. A. Mears, F. J. Wentz, The effect of diurnal correction on satellite-derived lower tropospheric temperature. *Science* (submitted).

15. Q. Fu and C. M. Johanson, *Geophys. Res. Lett.* **32**, L10703, doi:10.1029/2004GL022266 (2005).

16. B. D. Santer *et al.*, *Science* **300**, 1280 (2003).

17. C. A. Mears, M. C. Schabel, F. W. Wentz, *J. Clim.* **16**, 3650 (2003).

18. S. C. Sherwood, J. Lanzante, C. Meyer, Radiosonde daytime biases and late 20th century warming. *Science* (submitted).

19. W. J. Randel, F. Wu, Biases in stratospheric temperature trends derived from historical radiosonde data. *J. Climate* (submitted).

20. P. W. Thorne, D. E. Parker, J. R. Christy, C. A. Mears, Uncertainties in climate trends: Lessons from upper-air temperature records. *Bull. Amer. Met. Soc.* (in press).

21. G. C. Hegerl, J. M. Wallace, *J. Clim.* **15**, 2412 (2002).

22. N. P. Gillett, M. R. Allen, S. F. B. Tett, *Clim. Dyn.* **16**, 49 (2000).

23. G. A. Meehl, G. J. Boer, C. Covey, M. Latif, R. J. Stouffer, *Bull. Amer. Met. Soc.* **81**, 313 (2000).

24. While all 19 modeling groups used very similar changes in well-mixed greenhouse gases, the changes in other forcings were not prescribed as part of the experimental design. In practice, each group employed different combinations of 20th century forcings, and often used different datasets for specifying individual forcings. Ends dates for the experiment varied between groups, and ranged from 1999 to 2003. Some modeling centers performed ensembles of the historical forcing simulation (see SOM). An ensemble contains multiple realizations of the same experiment, each initiated from slightly different initial conditions, but with identical changes in external forcings (*2*). This yields many different realizations of the climate "signal" (the response to the imposed forcing changes) plus climate noise. Averaging over multiple realizations reduces noise and facilitates signal estimation.

25. F. J. Wentz, M. Schabel, *Nature* **403**, 414 (2000).

26. T. M. Smith, R. W. Reynolds, *J. Clim.*, in press (2005).

27. P. D. Jones, A. Moberg, *J. Clim.* **16**, 206 (2003). HadCRUT2v is the designation for version 2 of the (variance-corrected) Hadley Centre/Climatic Research Unit surface temperature dataset.

28. Q. Fu, C. M. Johanson, S. G. Warren, D. J. Seidel, *Nature* **429**, 55 (2004).

29. P. H. Stone, J. H. Carlson, *J. Atmos. Sci.* **36**, 415 (1979).

30. Here, we define $\overline{X}$ as the arithmetic mean of the ensemble means, *i.e.*, $\overline{X} = \frac{1}{N}\sum_{j=1}^{N}\overline{X}_j$, where N is the total number of models in the IPCC archive, and $\overline{X}_j$ is the ensemble mean signal of the j^{th} model. This weighting avoids undue emphasis on results from a single model with a large number of realizations.

31. One measure of ENSO variability is $s\{\mathrm{T}_{\mathrm{NIÑO\text{-}3.4}}\}$, the standard deviation of sea-surface temperatures in the Niño 3.4 region of the equatorial Pacific. Values of $s\{T_S\}$ in the 49 IPCC realizations are closely correlated with $s\{\mathrm{T}_{\mathrm{NIÑO\text{-}3.4}}\}$ ($r = 0.92$).

32. The theoretical expectation plotted in Fig. 3 was computed by taking the difference of two pseudo-adiabats calculated from surface air parcels with temperatures of 28.0 and 28.2°C and 80% relative humidity. These are conditions typical of deep convective regions over the tropical oceans. The pseudo-adiabats correspond to equivalent potential temperatures of 353.2 and 354.1 K. The assumed temperature difference of 0.2°C corresponds approximately to the total change in tropical ocean temperature over 1979 to 1999. Theoretical scaling ratios are relatively insensitive to reasonable variations in the baseline values of surface air temperature and relative humidity, as well as to the magnitude of the surface air temperature increase.

33. V. Ramaswamy *et al.*, in *Climate Change 2001: The Scientific Basis*, J. T.

Houghton *et al.*, Eds. (Cambridge Univ. Press, Cambridge, 2001), pp. 349-416.

34. B. D. Santer *et al.*, *Science* **301**, 479 (2003).

35. Work at Lawrence Livermore National Laboratory (LLNL) was performed under the auspices of the U.S. Dept. of Energy, Environmental Sciences Division, contract W-7405-ENG-48. A portion of this study was supported by the U.S. Dept. of Energy, Office of Biological and Environmental Research, as part of its Climate Change Prediction Program. TMLW was supported by NOAA Office of Climate Programs ("Climate Change Data and Detection") grant NA87GP0105. PWT and GJ were funded by the U.K. Dept. of the Environment, Food, and Rural Affairs. The authors acknowledge the international modeling groups for providing their data for analysis, the JSC/CLIVAR Working Group on Coupled Modelling (WGCM) and their Coupled Model Intercomparison Project (CMIP) and Climate Simulation Panel for organizing the model data analysis activity, and the IPCC WG1 TSU for technical support. The IPCC Data Archive at Lawrence Livermore National Laboratory is supported by the Office of Science, U.S. Department of Energy. The static MSU weighting functions and UAH MSU data were provided by John Christy (UAH). The authors thank Isaac Held, Tom Delworth (both GFDL), Dave Easterling (NCDC), Bruce Hicks (NOAA ARL), and two anonymous reviewers for useful comments. Olivier Boucher (Hadley Centre), Greg Flato (Canadian Climate Centre) and

Erich Roeckner (Max-Planck Institute for Meteorology) supplied information on the historical forcings used by CNRM-CM3, CCCma-CGCM3.1(T47), and ECHAM5/MPI-OM.

Figure 1: Time series of monthly-mean tropical temperature anomalies in T_4 (**A**), T_2 (**B**), and T_S (**C**). Observed T_4 and T_2 data are from UAH (*12*) and RSS (*17*). Observed T_S results are from the NOAA (*26*) and HadCRUT2v datasets (*27*). The latter were subsampled at the locations of HadAT2 radiosonde data (*6*). Model T_S results and synthetic MSU temperatures are from the IPCC historical forcing runs (see SOM). Results shown are restricted to those models that included forcing by both stratospheric ozone depletion and volcanic aerosols. All data were spatially-averaged over 20°N-20°S, expressed as anomalies relative to climatological monthly means over 1979 to 1999, and low-pass filtered. To facilitate model-data and model-model variability comparisons involving models with different ensemble sizes, only the first realization is plotted from each model.

Figure 2: Simulated and observed least-squares linear trends in tropical T_4 (**A**), T_2 (**B**), T_{Fu} (**C**), $T_{2\mathrm{LT}}$ (**D**), and T_S (**E**). Red bars represent $\overline{X}$, the mean of the model results (*30*). The black lines that encompass $\overline{X}$ are the maximum and minumum values from 49 realizations of the IPCC historical forcing experiment (see SOM). Asterisks identify observational trends outside the range of model results. All trends were calculated from spatially-averaged (20°N-20°S) anomaly data over the 252-month period January 1979 to December 1999. For anomaly definition and data sources, see Fig. 1. The orange bar in panel **E** is the T_S trend based on HadCRUT2v T_S data that were subsampled at the locations of HadAT2 radiosonde data (*6*).

Figure 3: Atmospheric profiles of temperature scaling ratios in models, theory, and radiosonde data. $R_S(z)$ is the ratio between the temporal standard deviations of $T(z)$, the temperature at discrete pressure levels, and the surface temperature $T(s)$ (**A**). $R_\beta(z)$ is similarly defined, but for trends over 1979 to 1999 (**B**). Model results are from 49 realizations of the IPCC historical forcing experiment. Radiosonde scaling ratios were calculated with HadAT2 and RATPAC $T(z)$ data (*6*, *7*). Scaling ratios for HadAT2 are based on unsubsampled HadCRUT2v $T(s)$ data. HadCRUT2v $T(s)$ data subsampled with HadAT2 coverage yield virtually identical scaling ratios (not shown). RATPAC-derived scaling ratios use spatially-complete NOAA $T(s)$ data. Theoretically-expected values of $R_S(z)$ and $R_\beta(z)$ are also shown (*32*). All standard deviations in panel **A** were calculated with linearly detrended data. $R_\beta(z)$ results in panel **B** are not plotted for three model realizations with surface warming close to zero (see SOM). All results are for spatial averages over 20°N-20°S. For anomaly definition, data sources, and further processing details, see Fig. 1 and SOM.

Figure 4: Scatter plots of the individual components of the $R_S(z)$ and $R_\beta(z)$ scaling ratios. Results are for the deep tropics (20°N-20°S). The two upper panels provide information on amplification of the monthly-timescale T_S variability in $T_{2\mathrm{LT}}$ (**A**) and T_{Fu} (**B**). The two bottom panels show the relationship between decadal-timescale trends in T_S and $T_{2\mathrm{LT}}$ (**C**) and in T_S and T_{Fu} (**D**). Each scatter plot has 49 pairs of model results. The fitted regression lines (in red) are based on model data only. The black lines denote a slope of 1. Values above (below) the black lines indicate

tropospheric enhancement (damping) of surface temperature changes. There are two columns of observational results in **C** and **D**. These are based on the NOAA and HadCRUT2v T_S trends (0.12 and 0.14°C/decade, respectively). Since $s\{T_S\}$ (the temporal standard deviation of T_S) is very similar in the NOAA and HadCRUT2v datasets, observed results in **A** and **B** use NOAA $s\{T_S\}$ values only. The blue shading in the two bottom panels defines the region of simultaneous surface warming and tropospheric cooling. For anomaly definition, analysis period, and data sources, refer to Fig. 1 and SOM.

Supporting Online Material

Calculation of synthetic MSU temperatures

We employed a static vertical weighting function to calculate synthetic satellite temperatures from model and radiosonde data. For large area averages, this approach yields decadal-timescale temperature changes similar to those obtained with a full radiative transfer code (*S1*). The static weighting function was applied to vertical profiles of monthly-mean, zonally-averaged atmospheric temperature anomalies (for radiosonde data) and to temperature profiles at individual grid-points (for model data). Results were then spatially averaged over the deep tropics (20°N-20°S).

Actual MSU temperatures

We used version 5.1 of the UAH MSU T_4, T_2, and T_{2LT} data and version 1.3 of the RSS MSU data. A T_2 product independently produced by a group at the University of Maryland (UM) shows tropospheric warming exceeding that in RSS and UAH (*S2*). The UM dataset is currently available in global mean form only, and so could not be used in our investigation of tropical temperature changes.

Estimation of T_{Fu}

This was calculated as $T_{Fu} = a\ T_2\ - b\ T_4$, with coefficients $a = 1.1$ and $b = -0.1$. We applied these coefficients to the monthly-mean, spatially-averaged (20°N-20°S) T_2 and T_4 anomalies from satellites (UAH and RSS), radiosondes (HadAT2 and RATPAC), and the IPCC historical forcing simulations. Note that a and b were derived mathematically (from the T_2 and T_4 weighting functions) rather than empirically (*S3*). Trends in T_{Fu} receive only a small contribution ($< \pm 0.005$°C/decade) from the cooling stratosphere. The peak of the broad effective weighting function for T_{Fu} occurs between *ca.* 300 and 400 hPa. The peak weight for T_{2LT} is lower in the atmosphere, between *ca.* 600 and 700 hPa (*S3*).

Calculation of temporal standard deviations

All temporal standard deviations used for the calculation of $R_S(z)$ results in Figs. 3 and 4 were estimated from linearly detrended data. This was done because some of the model simulations examined here have large decadal trends in surface and atmospheric temperature, which inflate the temporal variance of the raw T_S and $T(z)$ data. This aliasing effect makes it difficult to use standard deviations estimated from raw data to separate amplification behavior on monthly and decadal timescales.

Subsampling observed T_s data with radiosonde coverage

Subsampling surface temperature data at radiosonde locations actually degrades r, the correlation between time series of monthly-mean $T_{2\mathrm{LT}}$ and T_S anomalies. For example, $r = 0.82$ for tropical average anomalies calculated from the unsubsampled HadCRUT2v T_S data and the HadAT2 radiosonde $T_{2\mathrm{LT}}$ data. The correlation is decreased ($r = 0.73$) if the same radiosonde data are correlated with the HadCRUT2v T_S data subsampled at the radiosonde locations. This result must be related to the fact that surface temperature fluctuations have smaller correlation length scales than temperature fluctuations in the free troposphere (*S4*).

Modeling groups contributing to IPCC database

At the time this research was conducted, 19 modeling groups had performed a wide range of simulations in support of the IPCC Fourth Assessment Report. Climate data from these simulations were made available to the scientific community through the U.S. Dept. of Energy's Program for Climate Model Diagnosis and Intercomparion (PCMDI).

One of the simulations (the so-called "20c3m" run) involved historical changes in a number of anthropogenic and natural forcings. The following modeling groups provided multiple realizations of the 20c3m run (the text in parentheses gives the

official model designation): the National Center for Atmospheric Research in Boulder (CCSM3, five; PCM, four); the Max-Planck Institute for Meteorology in Germany (ECHAM5/MPI-OM, three); the Institute for Atmospheric Physics in China (FGOALS-g1.0, three), the Geophysical Fluid Dynamics Laboratory in Princeton (GFDL-CM2.0, three; GFDL-CM2.1, three), the Goddard Institute for Space Studies in New York (GISS-AOM, two; GISS-EH, five; GISS-ER, five); the Center for Climate System Research, National Institute for Environmental Studies, and Frontier Research Center for Global Change in Japan (MIROC-CGCM2.3.2(medres), three; MIROC-CGCM2.3.2(hires), one); and the Meteorological Research Institute in Japan (MRI-CGCM2.3.2, five). Individual realizations were supplied by the Canadian Centre for Climate Modelling and Analysis (CCCma-CGCM3.1(T47)); Météo-France/Centre National de Recherches Météorologiques (CNRM-CM3); the Institute for Numerical Mathematics in Russia (INM-CM3.0); the Institute Pierre Simon Laplace in France (IPSL-CM4); and the Hadley Centre for Climate Prediction and Research in the U.K. (UKMO-HadCM3 and UKMO-HadGEM1). Some groups provided results for several different model configurations.

Forcings used in 20c3m runs

Details of the natural and anthropogenic forcings used by differing modeling groups in their IPCC "historical forcing" simulations are given in Table 1. This Table was com-

piled using information that participating modeling centers provided to the PCMDI (see http://www-pcmdi.llnl.gov/ipcc/model_documentation). Model acronyms are defined in the previous section.

A total of 11 different forcings are listed in Table 1. A letter 'Y' denotes inclusion of a specific forcing. As used here, 'inclusion' signifies the specification of time-varying forcings, with changes on interannual and longer timescales. Forcings that were varied over the seasonal cycle only, or not at all, are identified with a dash. A question mark indicates a case where there is uncertainty regarding inclusion of the forcing.

Stratification of 20c3m simulations

In Fig. 3, simulations were stratified according to model groups that included both stratospheric ozone depletion and volcanic aerosols (O+V) and groups that omitted these forcings. Only 9 (12) of the 19 IPCC models included forcing by volcanic aerosols (stratospheric ozone). There is no indication that the $R_S(z)$ results in Fig. 3A reflect this stratification. However, some of the largest $R_\beta(z)$ values in the lower troposphere occur in realizations performed with CCSM3, GFDL-CM2.1, and GISS-EH (Fig. 3B). All three of these models included changes in carbonaceous aerosols, which may act to warm the lower troposphere relative to the surface, thus increasing $R_\beta(z)$ (*S5*).

Simulated and observed changes in T_4

The overall stratospheric cooling during the satellite era is related to stratospheric ozone depletion (*S6*). The warming of T_4 after the eruptions of El Chichón in April 1982 and Pinatubo in June 1991 is due to the absorption of incoming solar radiation and upwelling terrestrial radiation by volcanic aerosols (*S6*, *S7*). A change in the phase of the QBO during the Pinatubo eruption induced cooling of the equatorial stratosphere (*S8*), thus damping the stratospheric warming response to Pinatubo. QBO variability is not well-simulated by most models in the IPCC archive, which partly explains why the simulated T_4 response to Pinatubo is larger and less attenuated in the model simulations.

Anomalous $R_\beta(z)$ results

Two of the 49 model realizations (INM-CM3.0) and (ECHAM5/MPI-OM, run 2) have $R_\beta(z)$ values < 1.0 for trends in $T_{2\mathrm{LT}}$ and T_{Fu}. These anomalous results are due to either very weak surface warming (INM-CM3.0) or very weak surface cooling (ECHAM5/MPI-OM, run 2). Neither case is analogous to the radiosonde- or UAH-based $R_\beta(z) \ll 1.0$ results, which occur in conjunction with pronounced surface warming. The ECHAM result arises from the chance occurrence of a large La Niña event near the end of the 1979-1999 period. Run 1 of MRI-CGCM2.3.2 has very large $R_\beta(z)$ values, which are associated with T_S, $T_{2\mathrm{LT}}$, and T_{Fu} trends that are all close to

zero (Figs. 4C,D).

Caption for Figure S1

Figure S1: Scatter plots of the individual components of the $R_S(z)$ scaling ratio. Results are for the deep tropics (20°N-20°S). Panel **A** is identical to Fig. 4A, and provides information on the monthly-timescale variability in T_S and T_{2LT} in a wide range of model and observational datasets. The standard deviations of T_S and T_{2LT} in panel **B** are based on annual-mean rather than monthly-mean anomalies. All standard deviations were calculated from linearly detrended data. Observed $s\{T_S\}$ values are from NOAA (*S9*). The fitted regression lines (in red) are based on model data only. The black lines denote a slope of 1. For further details of datasets and analysis periods, refer to Fig. 4. Note the close correspondence between the monthly-mean and annual-mean results, which is due to the fact that monthly-timescale variability in T_S and T_{2LT} is dominated by interannual fluctuations in ENSO.

Supporting References and Notes

S1. B. D. Santer *et al.*, *J. Geophys. Res.* **104**, 6305 (1999).

S2. K. Y. Vinnikov, N. C. Grody, *Science* **302**, 269 (2003).

S3. Q. Fu and C. M. Johanson, *Geophys. Res. Lett.* **32**, L10703, doi:10.1029/2004GL022266 (2005).

S4. J. W. Hurrell, K. E. Trenberth, *J. Clim.* **11**, 945 (1998).

S5. V. Ramanathan *et al.*, *Proc. Nat. Acad. Sci.* **102**, doi: 10.1073/pnas.0500656102 (2005).

S6. V. Ramaswamy *et al.*, *Rev. Geophys.* **39**, 71 (2001).

S7. J. E. Hansen *et al.*, *J. Geophys. Res.* **107**, ACL-2, doi:10.1029/2001JD001143 (2002).

S8. L. Bengtsson, E. Roeckner, M. Stendel, *J. Geophys. Res.* **104**, 3865 (1999).

S9. T. M. Smith, R. W. Reynolds, *J. Clim.*, in press (2005).

Table 1: Forcings used in IPCC simulations of 20th century climate change

	Model	G	O	SD	SI	BC	OC	MD	SS	LU	SO	VL
1	CCCma-CGCM3.1(T47)	Y	-	Y	-	-	-	-	-	-	-	-
2	CCSM3	Y	Y	Y	-	Y	-	-	-	-	Y	Y
3	CNRM-CM3	Y	Y	Y	-	Y	?	-	-	-	-	-
4	CSIRO-Mk3.0	Y	-	Y	-	?	?	?	?	?	?	?
5	ECHAM5/MPI-OM	Y	Y	Y	Y	-	-	-	-	-	-	-
6	FGOALS-g1.0	Y	-	Y	?	-	-	-	-	-	-	-
7	GFDL-CM2.0	Y	Y	Y	-	Y	Y	-	-	Y	Y	Y
8	GFDL-CM2.1	Y	Y	Y	-	Y	Y	-	-	Y	Y	Y
9	GISS-AOM	Y	-	Y	-	-	-	-	Y	-	-	-
10	GISS-EH	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
11	GISS-ER	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
12	INM-CM3.0	Y	-	Y	-	-	-	-	-	-	Y	a
13	IPSL-CM4	Y	-	Y	?	-	-	-	-	-	-	-
14	MIROC3.2(medres)	Y	Y	Y	?	Y	Y	Y	Y	Y	Y	Y
15	MIROC3.2(hires)	Y	Y	Y	?	Y	Y	Y	Y	Y	Y	Y
16	MRI-CGCM2.3.2	Y	-	Y	-	-	-	-	-	-	Y	a
17	PCM	Y	Y	Y	-	-	-	-	-	-	Y	Y
18	UKMO-HadCM3	Y	Y	Y	Y	-	-	-	-	-	-	-
19	UKMO-HadGEM1	Y	Y	Y	Y	Y	Y	-	-	Y	Y	Y

G = Well-mixed greenhouse gases

O = Tropospheric and stratospheric ozone

SD = Sulfate aerosol direct effects

SI = Sulfate aerosol indirect effects

BC = Black carbon

OC = Organic carbon

MD = Mineral dust

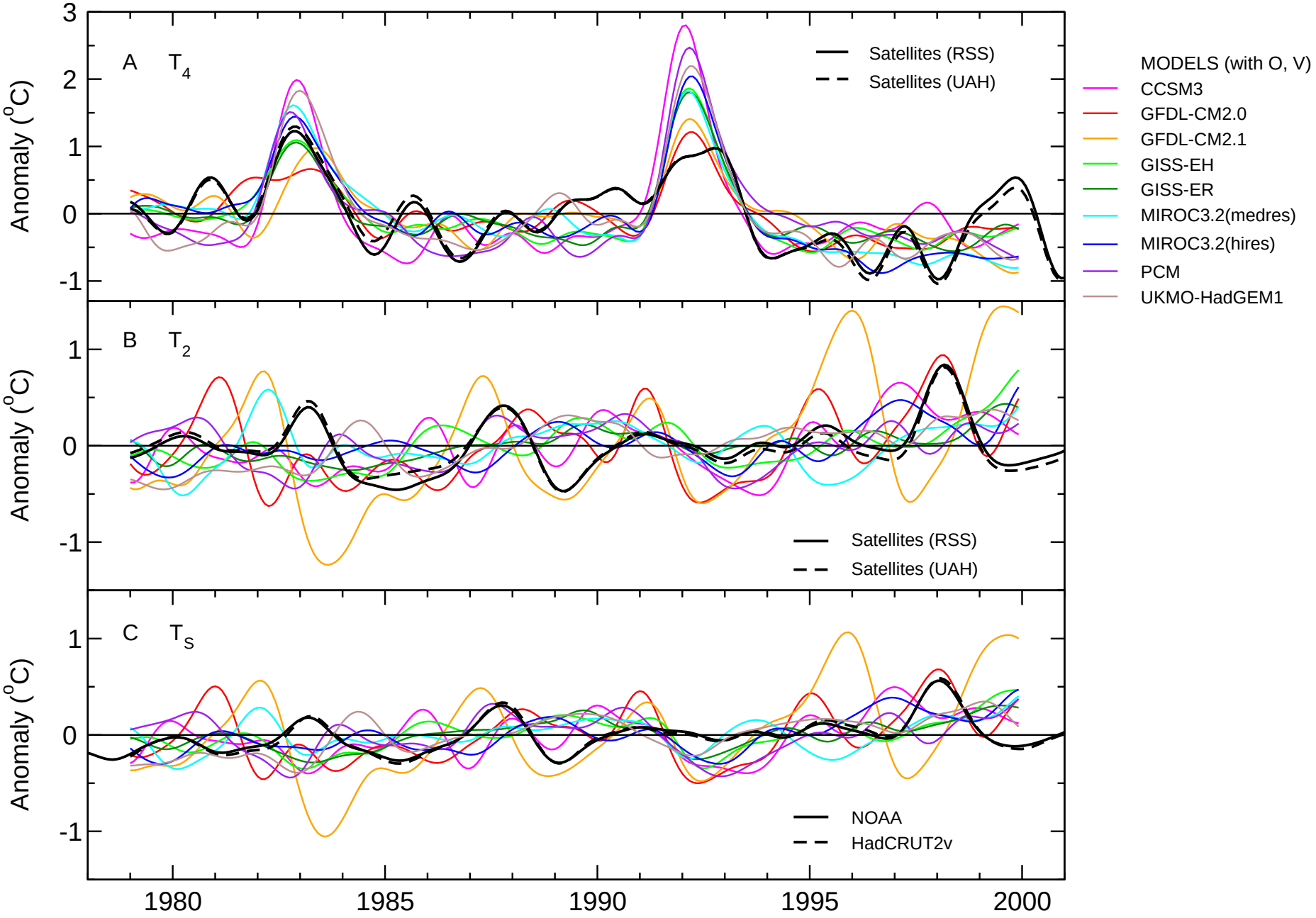
SS = Sea salt

LU = Land use change

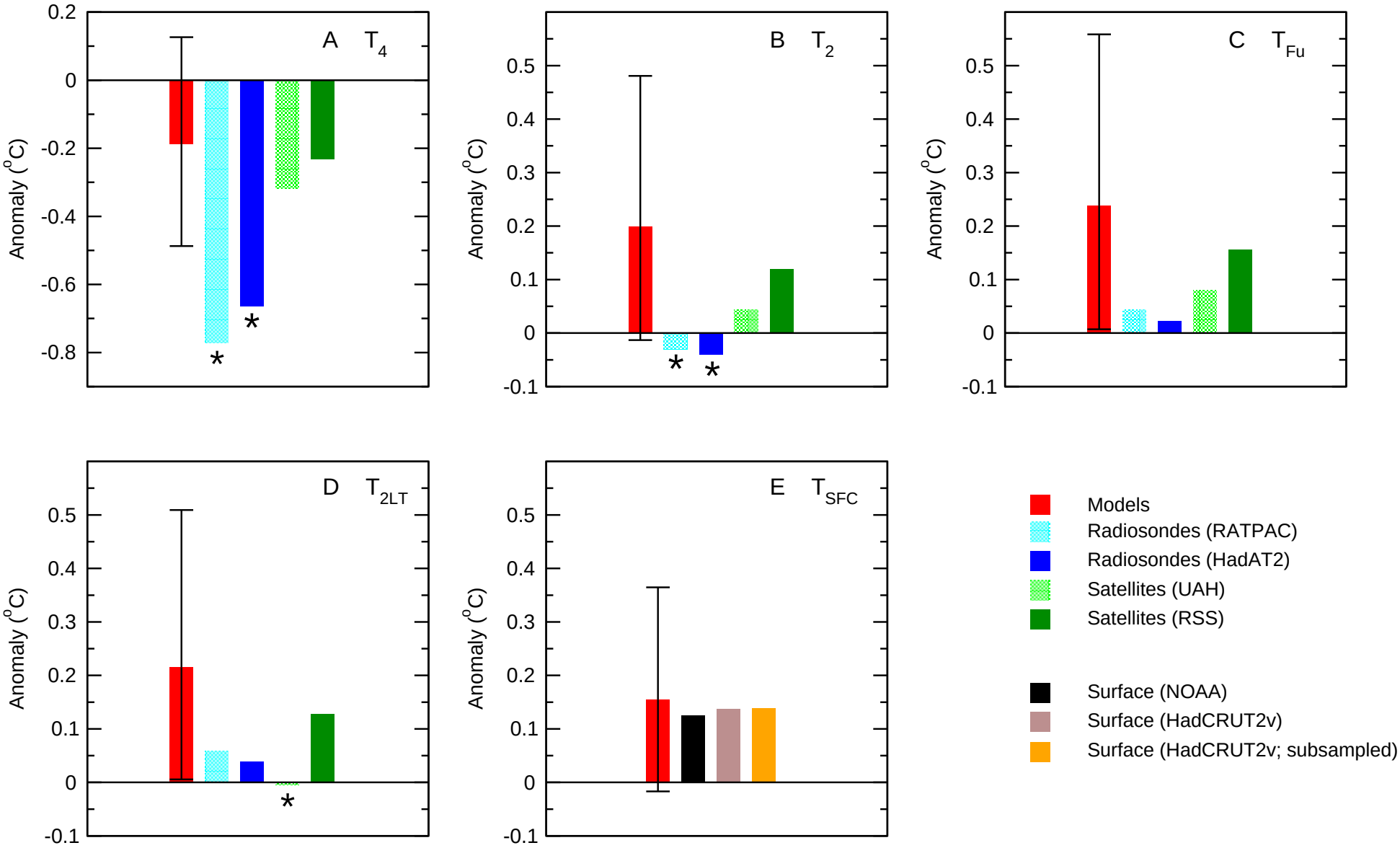
SO = Solar irradiance

VL = Volcanic aerosols.

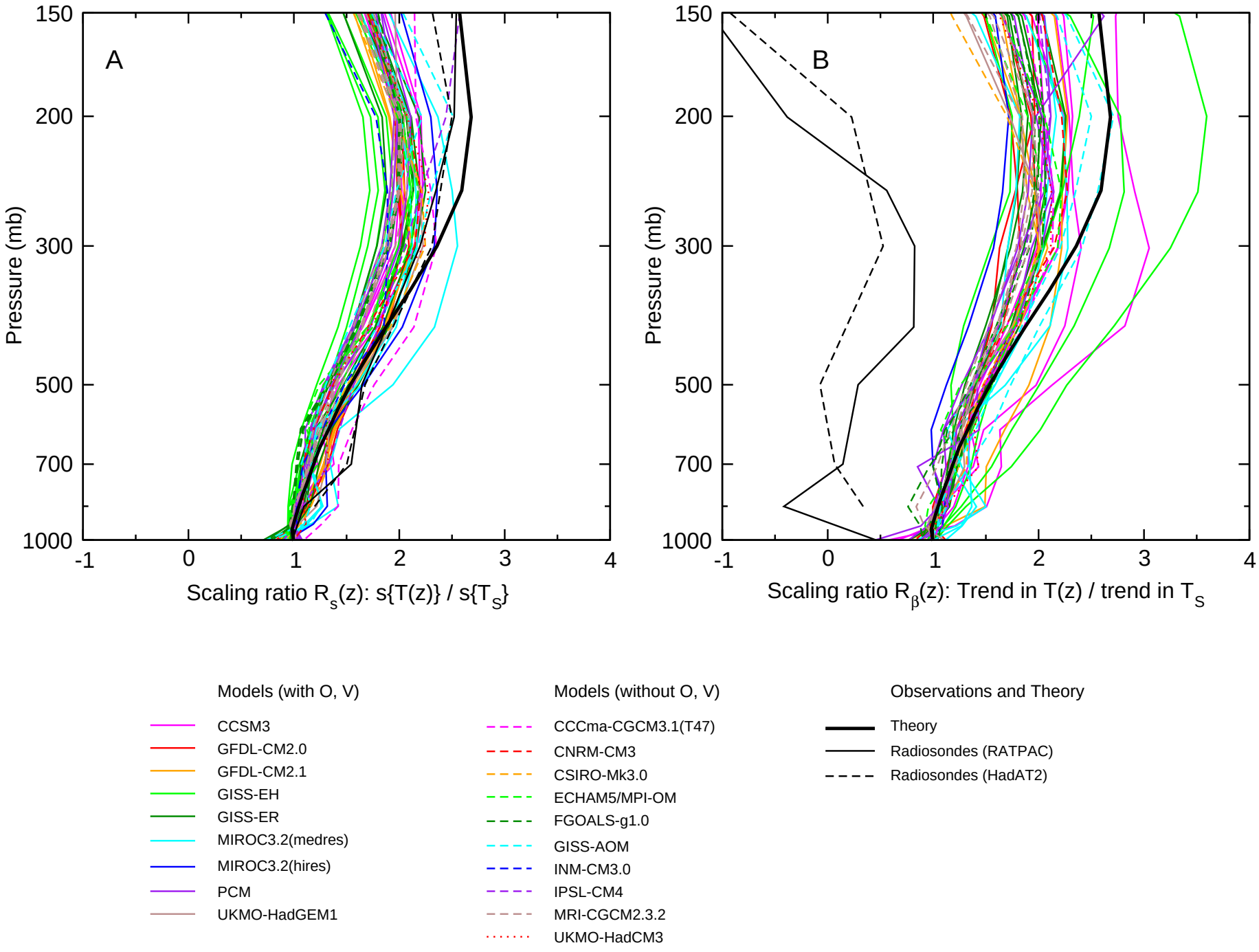
a = Documentation claims inclusion of volcanic aerosols, but there are no stratospheric warming responses in T_4.



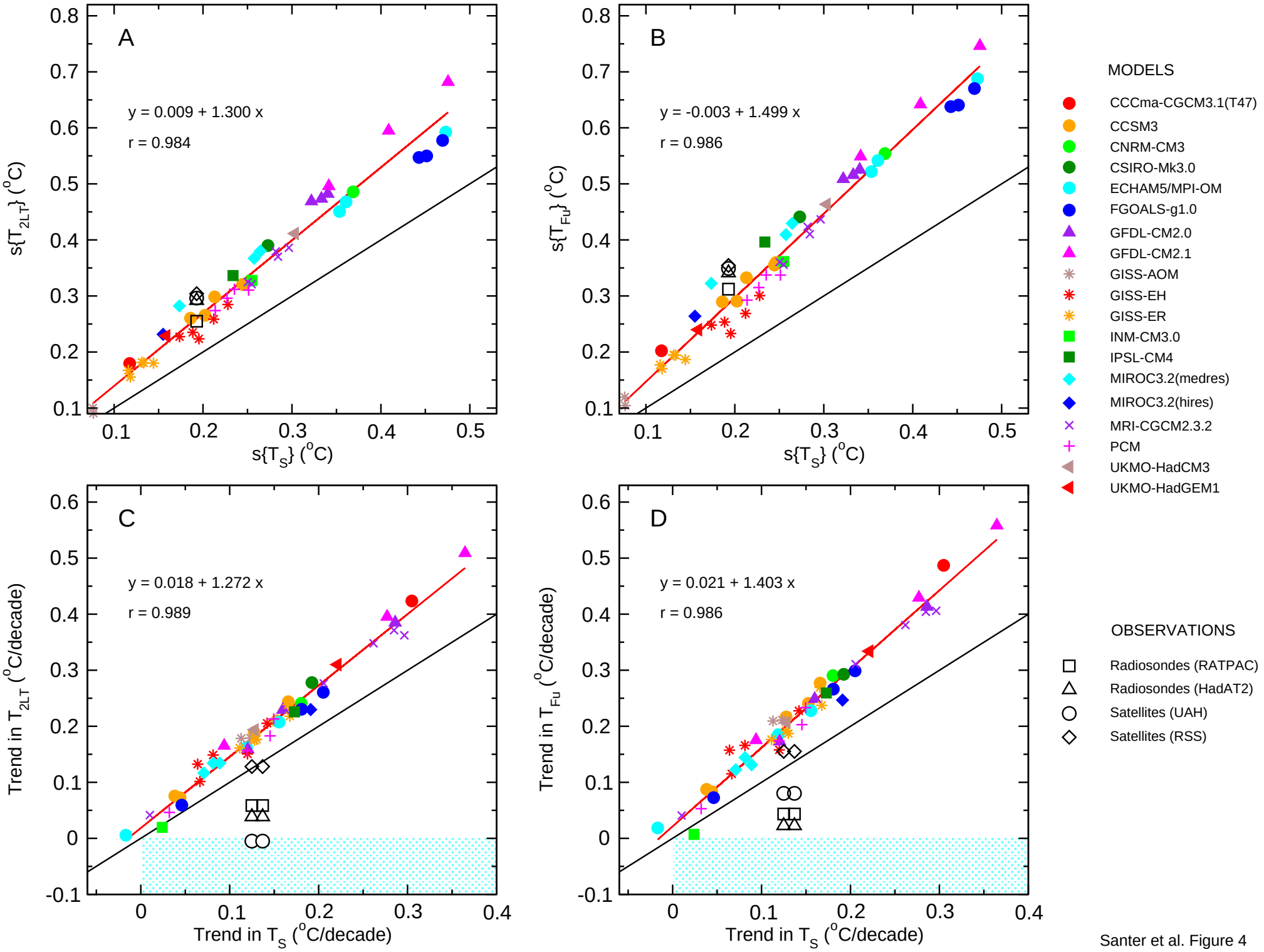
Santer et al. Figure 1



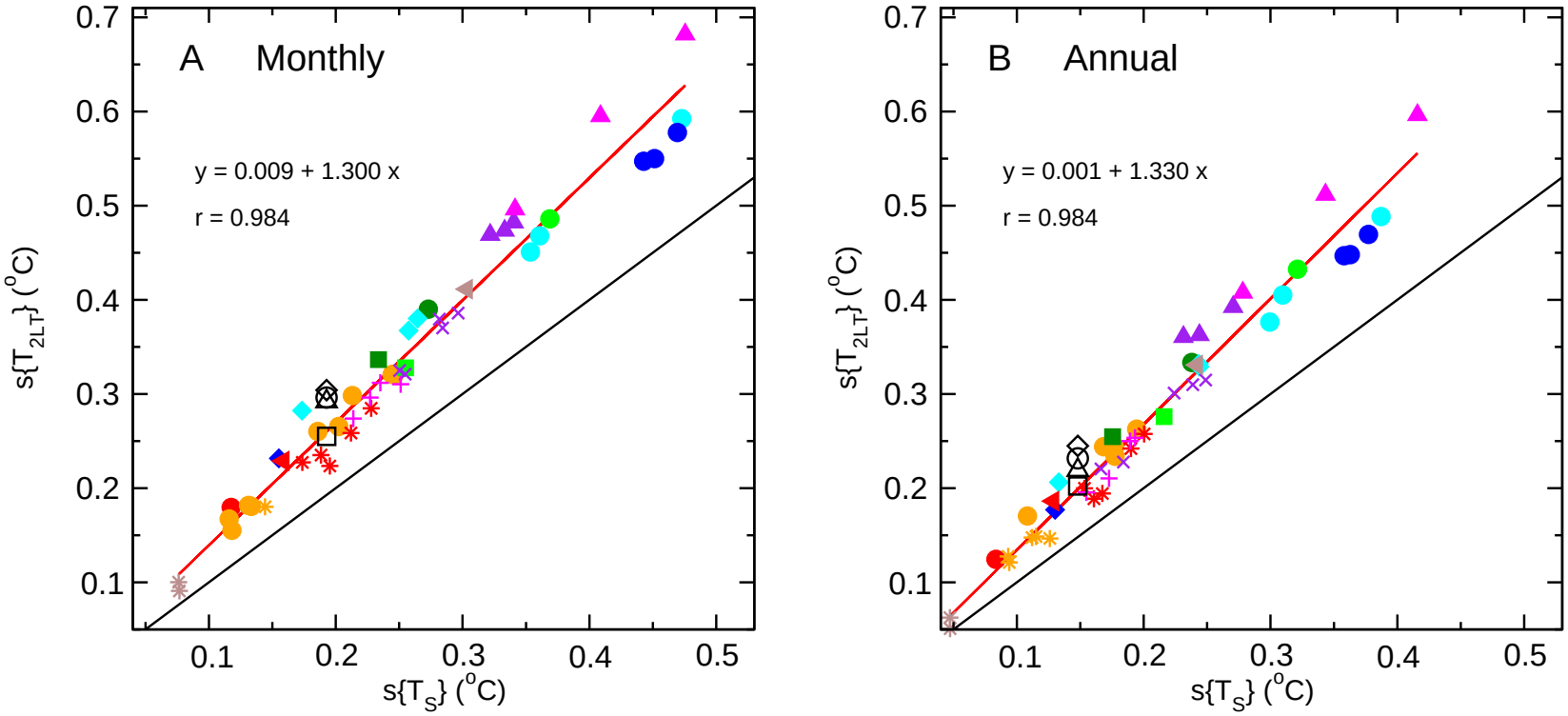
Santer et al. Figure 2



Santer et al. Figure 3



Santer et al. Figure 4



MODELS

- CCCma-CGCM3.1(T47)
- CCSM3
- CNRM-CM3
- CSIRO-Mk3.0
- ECHAM5/MPI-OM
- FGOALS-g1.0
- GFDL-CM2.0
- GFDL-CM2.1
- GISS-AOM
- GISS-EH
- GISS-ER
- INM-CM3.0
- IPSL-CM4
- MIROC3.2(medres)
- MIROC3.2(hires)
- MRI-CGCM2.3.2
- PCM
- UKMO-HadCM3
- UKMO-HadGEM1

OBSERVATIONS

- Radiosondes (RATPAC)
- Radiosondes (HadAT2)
- Satellites (UAH)
- Satellites (RSS)

Santer et al. Figure S1