

26. We chose these values of the target factors to produce our final results because we have concluded that they are the most likely to be free of errors. They are calculated from oceanic observations to reduce errors from uncorrected diurnal variations, and we use unweighted MSU channel 2 data (T2 in SOM) to avoid additional noise due to the differencing procedure used to calculate TLT. The values of the intersatellite offsets needed to be recalculated to remove obvious intersatellite differences. In the supporting online material, we discuss the impact of using different data subsets to determine the target factors. This information is used to help determine the structural uncertainty.
27. We obtain this estimate of the tropical TLT trend when we recalculate the intersatellite offsets to optimize them for tropical data. If this reoptimization is not performed, as it is not in producing maps such as those shown in Fig. 3, we obtain a smaller trend value of 0.164 K per decade.
28. T. M. Smith, R. W. Reynolds, *J. Clim.* **18**, 2021 (2005).
29. J. W. Hurrell, K. E. Trenberth, *J. Clim.* **11**, 945 (1998).
30. K. E. Trenberth, J. Fasullo, L. Smith, *Clim. Dyn.*, in press; published online 11 May 2005 (10.1007/s00382-005-0017-4).
31. J. Lanzante, S. Klein, D. Seidel, *J. Clim.* **16**, 224 (2003).
32. J. Lanzante, S. Klein, D. Seidel, *J. Clim.* **16**, 241 (2003).
33. P. W. Thorne et al., *J. Geophys. Res.*, in press.
34. F. J. Wentz, M. Schabel, *Nature* **394**, 661 (1998).
35. This work was supported by the NOAA Climate and

Global Change Program. We thank J. Christy and R. Spencer for providing numerical values for their diurnal adjustment.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1114772/DC1
SOM Text
Figs. S1 to S4
Tables S1 to S3
References and Notes

12 May 2005; accepted 27 July 2005

Published online 11 August 2005;

10.1126/science.1114772

Include this information when citing this paper.

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 G. A. Schmidt⁹

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

Tropospheric warming is a robust feature of climate model simulations that include 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), con-

sistent with model results (8). The observed tropical warming is partly due to 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 data sets 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 ob-

served 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 disparity would point toward the existence of fundamental deficiencies in current climate models (and/or in the forcings used 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) with 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, 25).

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 data sets: HadAT2 (Hadley Centre Atmospheric Temperatures, version 2) and RATPAC (Radiosonde Atmospheric Temperature Products for Assessing Climate). These were compiled (respectively) by the UK 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 (26). Observed trends and variability in tropical surface temperatures (T_s) were obtained from the NOAA (27) and HadCRUT2v data sets (28, 29).

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

*To whom correspondence should be addressed. E-mail: santer1@llnl.gov

Our second method for estimating scaling ratios uses the weighted-average temperatures of deep atmospheric layers (12, 17). These temperatures 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_{2LT}). 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_{2LT} data are also computed from the HadAT2 and RATPAC radiosonde data sets (25).

T_2 receives a contribution from the cooling stratosphere (30). This hampers its use for estimating the amplification of surface temperature changes in the free troposphere. We therefore focus on T_{2LT} , 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} data set (14).

Another strategy for removing stratospheric influences on T_2 relies on a linear combination of T_4 and T_2 (15, 25). This procedure yields T_{Fu} [named for the first author

of (15)], which is representative of temperatures in the bulk troposphere. Relative to T_{2LT} , T_{Fu} receives more of its signal from higher regions of the troposphere. On the basis of simple moist adiabatic lapse rate (MALR) theory (31), we expect scaling ratios in the deep tropics to increase with increasing height and to peak at roughly 200 mbar. 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).

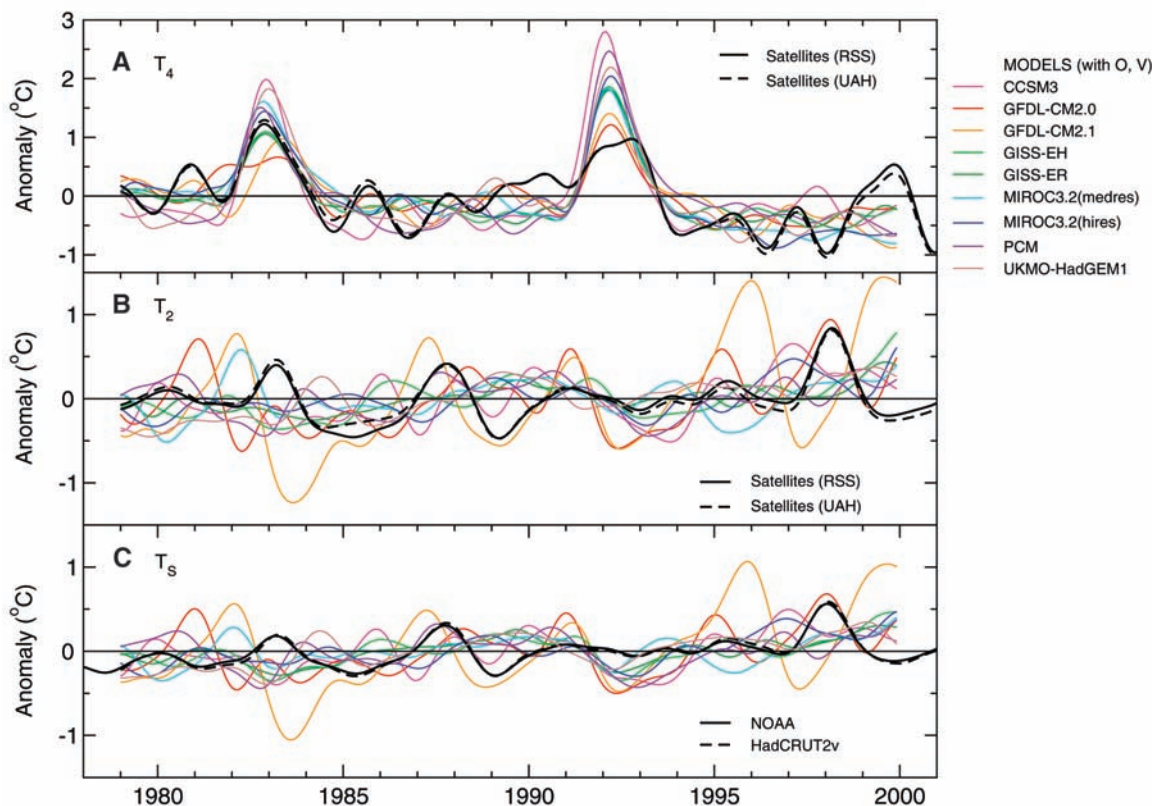
Time series of tropical T_4 changes in UAH, RSS, and the IPCC simulations are characterized by overall cooling trends and volcanically induced stratospheric warming signals (Fig. 1A). High-frequency variability associated with the quasi-biennial oscillation is evident in the observations but not in the model simulations (5, 25). 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 caused by 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 (Fig. 1, B and C). Because 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 data sets (0.12 and 0.14°C per decade, respectively) are very similar to \bar{X} , the average warming over all model simulations (Fig. 2E) (32). In the troposphere, however, model-observed trend agreement is sensitive to the atmospheric layer examined and the choice of observational data set. In both radiosonde data sets used here, T_2 cools over the years 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). The use of T_{Fu} removes most of the stratospheric influence on T_2 and yields positive temperature trends in all observed data sets (Fig.

Fig. 1. Time series of monthly-mean tropical temperature anomalies in (A) T_4 , (B) T_2 , and (C) T_s . Observed T_4 and T_2 data are from UAH (12) and RSS (17). Observed T_s results are from the NOAA (27) and HadCRUT2v datasets (28). 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 (25). Results shown are restricted to those models that included forcing by both stratospheric ozone depletion (O) and volcanic aerosols (V). All data were spatially averaged over 20°N to 20°S, expressed as anomalies relative to climatological monthly means over the years 1979 to 1999, and low-pass filtered. To facilitate model observation and model-model variability comparisons involving models with different ensemble sizes, only the first realization is plotted from each model.



2C) (5, 15, 30). All observed T_{Fu} trends are within the envelope of model values.

In the tropical lower troposphere, all data sets except UAH have positive T_{2LT} trends (Fig. 2D). The difference between the UAH and RSS trends ($\sim 0.13^\circ\text{C}$ per 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, whereas the latter warms) is not evident in any other data set (14, 15, 30).

Both model and satellite data indicate that variability in T_S is amplified in the tropical troposphere (Fig. 1, B and C). Amplification of surface warming is a direct result of moist thermodynamic processes (31). 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 multidecadal trends in these quantities, where z denotes a height coordinate (pressure in mbars). Because most of the monthly time-scale variability in tropical surface and tropospheric temperatures is driven by interannual fluctuations in ENSO, $R_S(z)$ largely reflects amplification processes acting on annual time scales (fig. S1) (33).

Figure 3A shows $R_S(z)$ values in models and radiosondes. The theoretically expected

profile is also displayed (34). In all cases, $R_S(z)$ increases above the boundary layer, with maximum amplification at ~ 200 mbar. Below ~ 400 mbar, there is close agreement between the scaling ratios in models, radiosondes, and theory. Between 400 and 150 mbar, the theoretical scaling ratios are consistently larger than they are 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 multidecadal linear trends, both radiosonde data sets 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 demonstrates 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. 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 (25).

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 multidecadal time scales, whereas the radiosonde data are only consistent with theory on monthly time scales.

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 $s(T_{2LT})$, the temporal standard deviations of monthly-mean tropical T_S and T_{2LT} 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 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 (Fig. 4, A and B). The regression between the model $s(T_S)$ and $s(T_{2LT})$ values has a slope of 1.3, in accord with the theoretically expected scaling ratio at the peak of the T_{2LT} 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) (25). All model and observational results in Fig. 4, A and B, are tightly clustered around the fitted (red) regression lines, which is consistent with the close agreement between

Fig. 2. Simulated and observed least-squares linear trends in tropical (A) T_4 , (B) T_2 , (C) T_{Fu} , (D) T_{2LT} , and (E) T_S . Red bars represent \bar{X} , the mean of the model results (32). The black lines that encompass \bar{X} are the maximum and minimum values from 49 realizations of the IPCC historical forcing experiment (25). Asterisks identify observational trends outside the range of model results. All trends were calculated from spatially-averaged (20°N to 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).

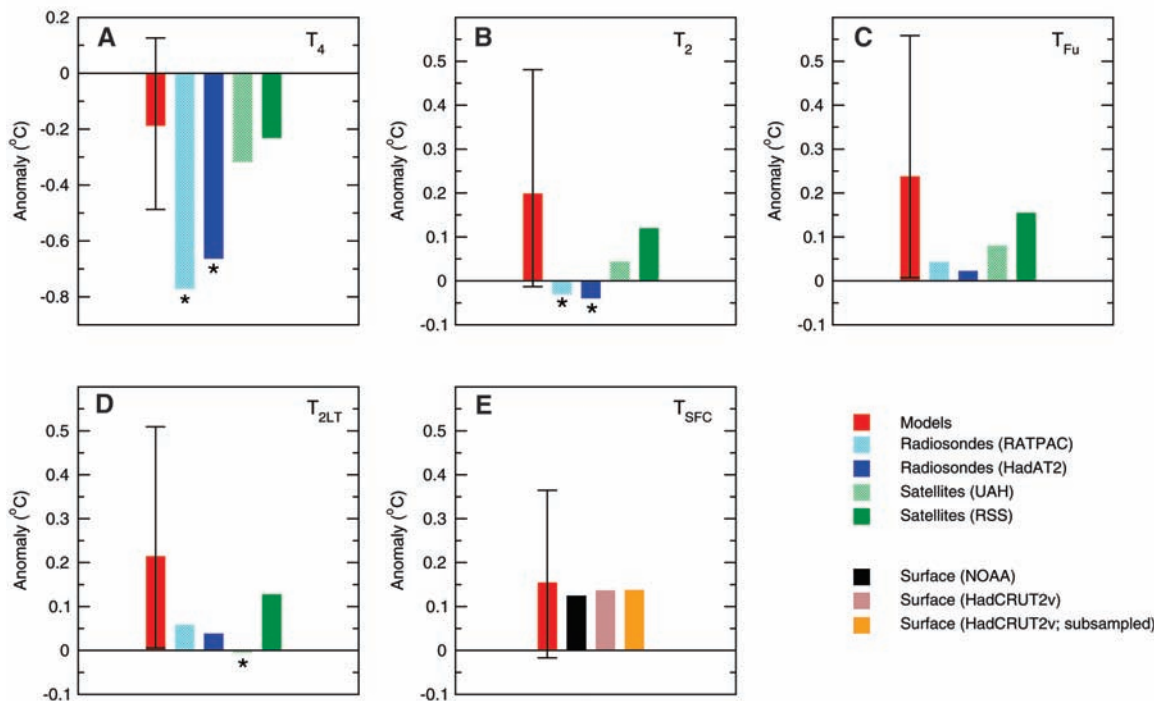
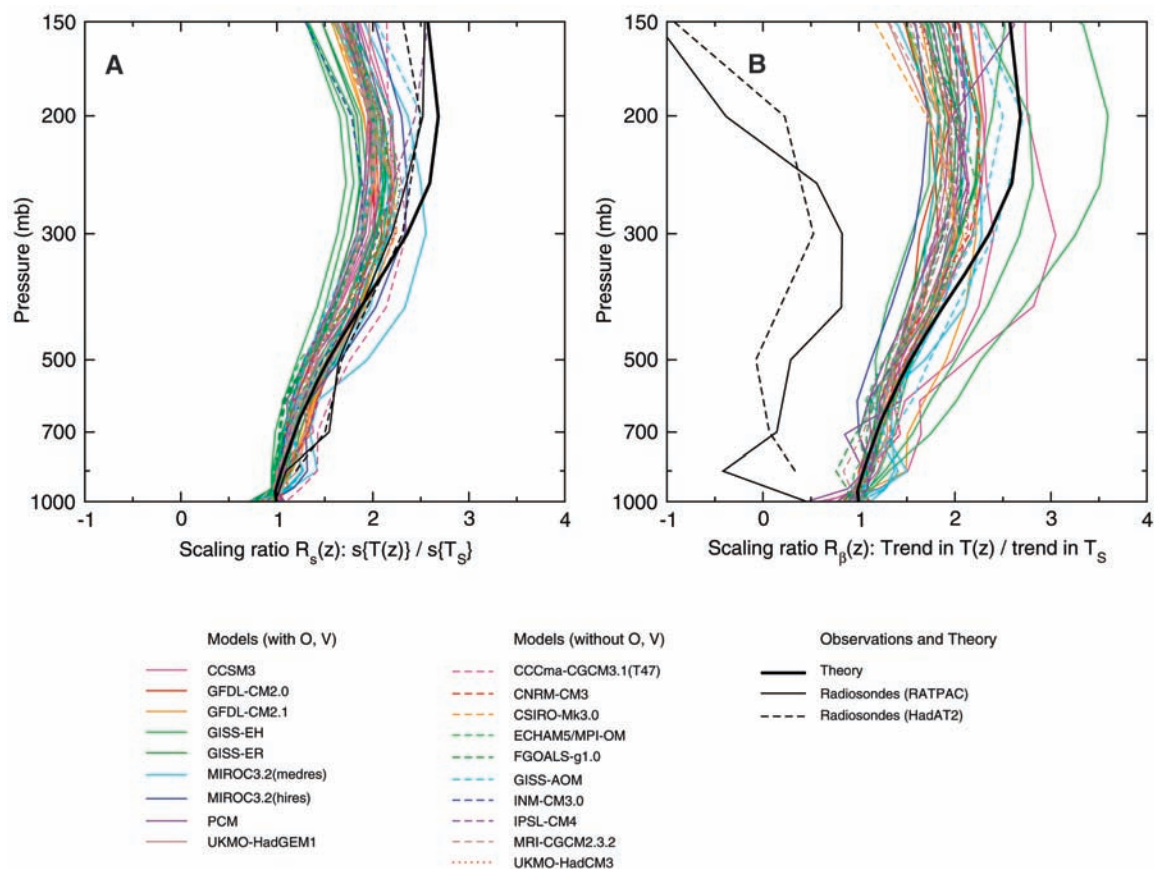


Fig. 3. Atmospheric profiles of temperature scaling ratios in models, theory, and radio-sonde data. (A) $R_S(z)$ is the ratio between the temporal standard deviations of $T(z)$, the temperature at discrete pressure levels, and the surface temperature T_S . (B) $R_B(z)$ is similarly defined, but for trends over 1979 to 1999. Model results are from 49 realizations of the IPCC historical forcing experiment. Radio-sonde scaling ratios were calculated with HadAT2 and RATPAC $T(z)$ data (6, 7). Scaling ratios for HadAT2 are based on unsampled 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 $R_B(z)$ are also shown (34). All standard deviations in panel (A) were calculated with linearly detrended data. $R_B(z)$ results in panel (B) are not plotted for three model realizations with surface warming close to zero (25). All results are for spatial averages over 20°N to 20°S. For anomaly definition, data sources, and further processing details, see Fig. 1 and (25).



modeled and observed $R_S(z)$ values in the lower troposphere (Fig. 3A).

Amplification factors estimated from multidecadal trends in T_S , T_{2LT} , and T_{Fu} also display considerable internal consistency in the 19 IPCC models (Fig. 4, C and D). This consistency occurs despite large intermodel 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 $R_B(z)$ ratios is robust to differences in the natural and anthropogenic forcings applied by each group (24, 25). Many of these forcings are heterogeneous in space and time (2, 3, 35). These differences in forcings and physics do not cause appreciable displacement of model results from the regression line in Fig. 4, C and D. The regression slopes are similar to those estimated from monthly-timescale variability, with T_{Fu} results again yielding a steeper slope than does T_{2LT} .

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_{2LT} and T_{Fu} trends have scaling factors that are in reasonable accord with

model results (Fig. 4, C and D) (25). Despite sustained warming of the tropical land and ocean surfaces, the UAH T_{2LT} trend is negative—i.e., $R_B(z) < 0$. The UAH $R_B(z)$ 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 an MALR.

In contrast to the model results and theoretical expectations, both radiosonde data sets used here have $R_B(z)$ ratios $\ll 1.0$ (Fig. 4, C and D). As in the case of the satellite data sets, 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.

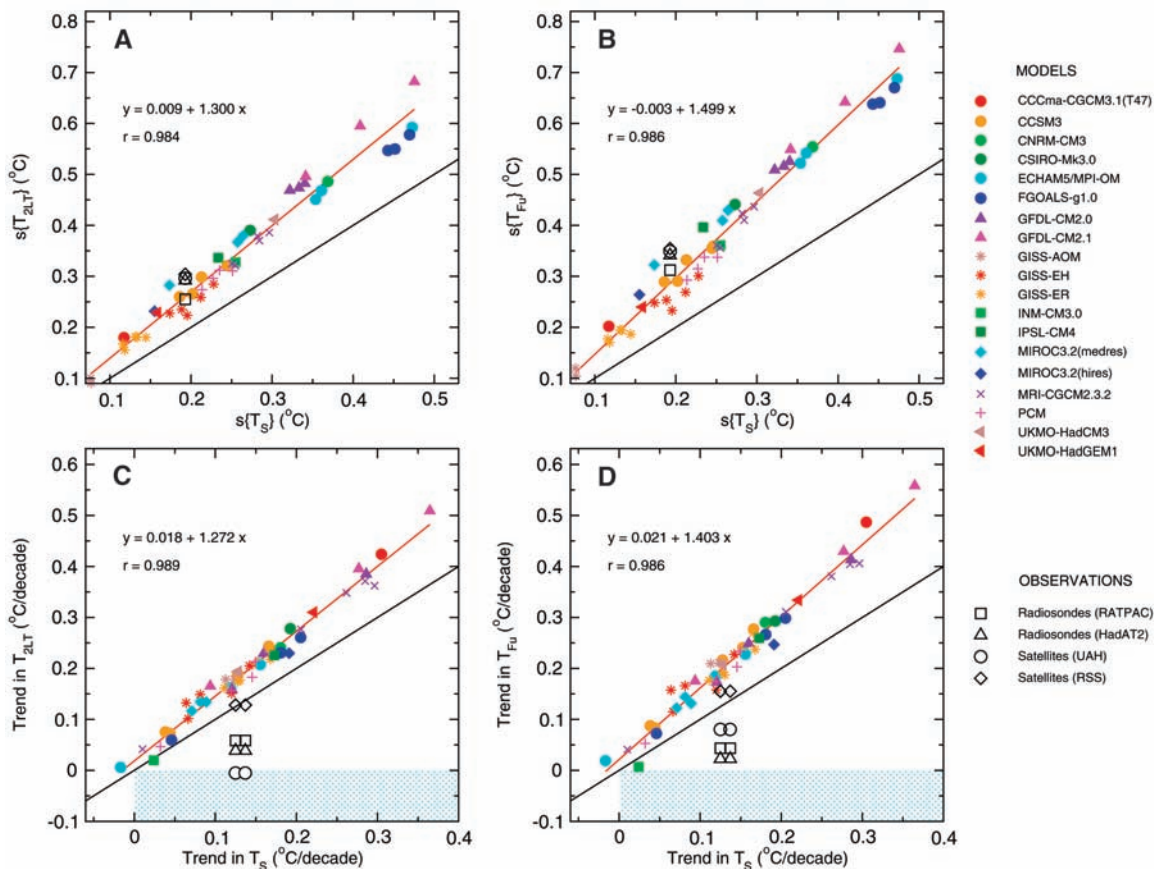
The existence of residual inhomogeneities in the observational data is likely. Current atmospheric observing systems were designed for real-time monitoring of weather rather than long-term monitoring of climate. The construction of reliable climate records

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 account for most of the variance of tropospheric temperature fluctuations in the deep tropics (Fig. 1B). However, systematic errors can have a pronounced effect on interdecadal variability. This helps to explain why model/data comparisons of $R_B(z)$ ratios are sensitive to observational uncertainty, whereas $R_S(z)$ ratios are not.

We have demonstrated that all observed data sets and model results are remarkably consistent in terms of their relation between monthly- and annual-time scale temperature variations at the surface and in the free troposphere. This is a strong verification of the

Fig. 4. Scatter plots of the individual components of the $R_s(z)$ and $R_p(z)$ scaling ratios. Results are for the deep tropics (20°N to 20°S). The two upper panels provide information on amplification of the monthly-timescale T_s variability in (A) T_{2LT} and (B) T_{Fu} . The two bottom panels show the relation between decadal-timescale trends in (C) T_s and T_{2LT} and in (D) T_s and T_{Fu} . 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 the black lines indicate tropospheric enhancement, and values below the black line indicate tropospheric 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 per decade, respectively). Because $s(T_s)$ (the temporal standard deviation of T_s) is very similar in the NOAA and HadCRUT2v data sets, 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 (25).

model physics that governs the amplification of tropical surface temperature changes. On decadal time scales, however, only one observed data set (RSS) shows amplification behavior that is generally consistent with model results. The correspondence between models and observations on monthly and annual time scales does not guarantee that model scaling ratios are valid on decadal time scales. 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 time scales.

The RSS T_{2LT} , T_2 , and T_{Fu} trends are physically consistent (all three layers warm as the surface warms), whereas the UAH data show trends of different sign in the lower- and midtroposphere. These results support the contention that the tropical warming trend in RSS T_{2LT} data is more reliable than T_{2LT} trends in other observational data sets. 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 (26, 36), which are in accord with warming but not cooling of the free troposphere.

We have used 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 data sets has been of great benefit to this research, and highlights the dangers inherent in drawing inferences on the agreement between models and observations without adequately accounting for uncertainties in both.

References and Notes

- B. D. Santer et al., *Nature* **382**, 39 (1996).
- J. E. Hansen et al., *J. Geophys. Res.* **107**, ACL-2, 10.1029/2001JD001143 (2002).
- S. F. B. Tett et al., *J. Geophys. Res.* **107**, 10.1029/2000JD000028 (2002).
- J. R. Lanzante, S. A. Klein, D. J. Seidel, *J. Clim.* **16**, 241 (2003).
- D. J. Seidel et al., *J. Clim.* **17**, 2225 (2004).
- P. W. Thorne et al., *J. Geophys. Res.* in press.
- M. Free et al., *J. Geophys. Res.*, in press.
- P. W. Thorne et al., *Geophys. Res. Lett.* **29**, 10.1029/2002GL015717 (2002).
- B. D. Santer et al., *Science* **287**, 1227 (2000).
- D. J. Gaffen et al., *Science* **287**, 1242 (2000).
- J. M. Wallace et al., *Reconciling Observations of Global Temperature Change* (National Academy Press, Washington DC, 2000).
- J. R. Christy, R. W. Spencer, W. B. Norris, W. D. Braswell, *J. Atmos. Ocean. Tech.* **20**, 613 (2003).
- J. W. Hurrell, K. E. Trenberth, *J. Clim.* **11**, 945 (1998).
- C. A. Mears, F. J. Wentz, *Science* **309**, 1548 (2005); published online 11 August 2005 (10.1126/science.1114772).
- Q. Fu, C. M. Johanson, *Geophys. Res. Lett.* **32**, L10703, 10.1029/2004GL022266 (2005).
- B. D. Santer et al., *Science* **300**, 1280 (2003).
- C. A. Mears, M. C. Schabel, F. W. Wentz, *J. Clim.* **16**, 3650 (2003).
- S. C. Sherwood, J. R. Lanzante, C. L. Meyer, *Science* **309**, 1556 (2005); published online 11 August 2005 (10.1126/science.1115640).
- W. J. Randel, F. Wu, in preparation.
- P. W. Thorne, D. E. Parker, J. R. Christy, C. A. Mears, *Bull. Am. Met. Soc.*, in press.
- G. C. Hegerl, J. M. Wallace, *J. Clim.* **15**, 2412 (2002).
- N. P. Gillett, M. R. Allen, S. F. B. Tett, *Clim. Dyn.* **16**, 49 (2000).
- G. A. Meehl, G. J. Boer, C. Covey, M. Latif, R. J. Stouffer, *Bull. Am. Met. Soc.* **81**, 313 (2000).
- Whereas 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 applied different combinations of 20th century forcings and often used different data sets for specifying individual forcings. End dates for the experiment varied between groups and ranged from 1999 to 2003. Some modeling centers performed ensembles of the historical forcing simulation (25). 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. Materials and methods are available as supporting material on Science Online.
 26. F. J. Wentz, M. Schabel, *Nature* **403**, 414 (2000).
 27. T. M. Smith, R. W. Reynolds, *J. Clim.* **18**, 2021 (2005).
 28. P. D. Jones, A. Moberg, *J. Clim.* **16**, 206 (2003).
 29. HadCRUT2v is the designation for version 2 of the (variance-corrected) Hadley Centre/Climatic Research Unit surface temperature data set.
 30. Q. Fu, C. M. Johanson, S. G. Warren, D. J. Seidel, *Nature* **429**, 55 (2004).
 31. P. H. Stone, J. H. Carlson, *J. Atmos. Sci.* **36**, 415 (1979).
 32. Here, we define \bar{X} as the arithmetic mean of the ensemble means, i.e., $\bar{X} = \frac{1}{N} \sum_{j=1}^N \bar{X}_j$, where N is the total number of models in the IPCC archive and \bar{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.
 33. One measure of ENSO variability is $s(T_{\text{Niño-3.4}})$, the standard deviation of sea-surface temperatures in the Niño 3.4 region of the equatorial Pacific. Values of $s(T_j)$ in the 49 IPCC realizations are closely correlated with $s(T_{\text{Niño-3.4}})$ (correlation coefficient $r = 0.92$).
 34. 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 the years 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.
 35. 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.
 36. B. D. Santer *et al.*, *Science* **301**, 479 (2003).
 37. Work at Lawrence Livermore National Laboratory (LLNL) was performed under the auspices of the U.S. Department of Energy (DOE), Environmental Sciences Division, contract W-7405-ENG-48. A portion of this study was supported by the U.S. DOE, Office of Biological and Environmental Research, as part of its Climate Change Prediction Program. T.M.L.W. was supported by NOAA Office of Climate Programs (Climate Change Data and Detection) grant NA87GP0105. P.W.T. and G.J. were funded by the UK Department of the Environment, Food, and Rural Affairs. We acknowledge the international modeling groups for providing their data for analysis, the Joint Scientific Committee/

Climate Variability and Predictability Working Group on Coupled Modeling and their Coupled Model Inter-comparison Project and Climate Simulation Panel for organizing the model data analysis activity, and the IPCC WG1 TSU for technical support. The IPCC Data Archive at LLNL is supported by the Office of Science, U.S. DOE. The static MSU weighting functions and UAH MSU data were provided by J. Christy (UAH). We thank I. Held, T. Delworth (both Geophysical Fluid Dynamics Laboratory), D. Easterling (National Climatic Data Center), B. Hicks (NOAA Air Resources Laboratory), and two anonymous reviewers for useful comments. O. Boucher (Hadley Centre), G. Flato (Canadian Climate Centre), and E. Roeckner (Max-Planck Institute for Meteorology) supplied information on the historical forcings used by CNRM-CM3, CCCma-CGCM3.1(T47), and ECHAM5/MPI-OM.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1114867/DC1

Materials and Methods

Fig. S1

Table S1

References and Notes

16 May 2005; accepted 27 July 2005

Published online 11 August 2005;

10.1126/science.1114867

Include this information when citing this paper.

Radiosonde Daytime Biases and Late-20th Century Warming

Steven C. Sherwood,^{1*} John R. Lanzante,² Cathryn L. Meyer¹

The temperature difference between adjacent 0000 and 1200 UTC weather balloon (radiosonde) reports shows a pervasive tendency toward cooler daytime compared to nighttime observations since the 1970s, especially at tropical stations. Several characteristics of this trend indicate that it is an artifact of systematic reductions over time in the uncorrected error due to daytime solar heating of the instrument and should be absent from accurate climate records. Although other problems may exist, this effect alone is of sufficient magnitude to reconcile radiosonde tropospheric temperature trends and surface trends during the late 20th century.

Atmospheric models and simple thermodynamic arguments indicate that tropospheric and surface temperature changes should be closely linked (*1*). Radiosonde data during the late 20th century, however (*2–5*), have not shown warming commensurate with that reported for the surface (*1, 6, 7*). The main discrepancy is in the Tropics during the last two decades of the 20th century.

A number of design changes to radiosonde systems over the years may have affected trends (*8*). Indeed, the spread of trends among stations well exceeds that implied by satellite data (*9*), suggesting that trends in the observation bias typically exceed those of the actual temperature at individual stations.

Among the most serious known problems is bias due to solar heating of the temperature sensor (*10*). For many radiosonde designs this can elevate the temperature several °C above ambient during daylight, an effect that must be removed via an estimated correction. For other designs no correction is standard even though the effect may not be completely absent. Adjustment of climate records for instrument changes using their documented histories is problematic (*8, 11*).

One can try to remove undocumented artifacts by careful examination of the data itself. Several such efforts have detected hundreds or thousands of apparent artifacts (*3–5, 12*). Their net effect on trends was found to be large only in the stratosphere. Revised trends were still lower than those indicated by the Microwave Sounding Unit (MSU) in both the troposphere and stratosphere (*13*). Because empirical separation of artificial discontinuities from genuine variability is extremely challenging in correlated time series (*14, 15*), especially as changes can probably occur in many small steps (*16*), it is

not clear how successful the above efforts may have been in detecting discontinuities—or avoiding false adjustments—of amplitudes well below 1°C.

Here we adopt a strategy for quantifying trend errors that does not require identifying specific change events. The strategy applies only to the solar heating error and does not detect other errors. It relies on the fact that the diurnal temperature range in the free troposphere, hence its expected trend, is small and has known characteristics that differ from those expected from a radiation error.

The diurnal temperature variation in Earth's atmosphere is a tide arising from its direct solar heating and from diurnal variations of convective heating driven by the diurnal variation of surface temperature. Atmospheric heating, which occurs primarily in the stratosphere via ozone absorption, drives migrating resonant oscillations that cause temperature fluctuations of several °C in the upper stratosphere. In the troposphere, weaker solar heating occurs due mainly to near-infrared absorption by water with a contribution from dark aerosols. These influences produce diurnal temperature fluctuations of 1°C or less in the free troposphere (*17*). Near the land surface, variations of 5° to 15°C occur due to surface diurnal heating (*18*); over oceans, variations are ~1°C.

Because atmospheric tides are a linear phenomenon (*19*), the diurnal variation of temperature is proportional to that of the heating, though the two need not be in phase. Trends of ~ −0.2°C per decade are evident in the land surface diurnal temperature range (DTR) (*20*), which amount to about 2% of the mean DTR per decade. Tropospheric water vapor and stratospheric ozone changes do not exceed a few percent per decade in recent decades (*21, 22*), and absorption increases

¹Department of Geology and Geophysics, Yale University, New Haven, CT 06520, USA. ²National Oceanic and Atmospheric Administration/Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, NJ 08542, USA.

*To whom correspondence should be addressed. E-mail: ssherwood@alum.mit.edu