

Diagnosis of Radiosonde Vertical Temperature Trend Profiles: Comparing the Influence of Data Homogenization versus Model Forcings

JOHN R. LANZANTE

NOAA/Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, New Jersey

(Manuscript received 20 December 2006, in final form 7 March 2007)

ABSTRACT

Measurements from radiosonde temperatures have been used in studies that seek to identify the human influence on climate. However, such measurements are known to be contaminated by artificial inhomogeneities introduced by changes in instruments and recording practices that have occurred over time. Some simple diagnostics are used to compare vertical profiles of temperature trends from the observed data with simulations from a GCM driven by several different sets of forcings. Unlike most earlier studies of this type, both raw (i.e., fully contaminated) as well as adjusted observations (i.e., treated to remove some of the contamination) are utilized. The comparisons demonstrate that the effect of observational data adjustment can be as important as the inclusion of some major climate forcings in the model simulations. The effects of major volcanic eruptions critically influence temperature trends, even over a time period nearly four decades in length.

In addition, it is seen that the adjusted data show consistently better agreement than the unadjusted data, with simulations from a climate model for 1959–97. Particularly noteworthy is the fact that the adjustments supply missing warming in the tropical upper troposphere that has been attributed to model error in a number of earlier studies.

Finally, an evaluation of the fidelity of the model's temperature response to major volcanic eruptions is conducted. Although the major conclusions of this study are unaffected by shortcomings of the simulations, they highlight the fact that even using a fairly long period of record (~40 yr), any such shortcomings can have an important impact on trends and trend comparisons.

1. Introduction

Temperature data play a central role in assessing whether human activity has altered the climate. Temperature change in the form of both horizontal patterns at the surface as well as vertical patterns in the atmosphere (the focus of this work) has proven useful. While satellite data have been used in such endeavors, they lack the resolution needed for a more detailed examination of vertical structure and are limited in record to about the last 25 yr. Temperature data from radiosondes would seem to be an ideal candidate for this purpose. Such data resolve much more detail in the vertical and their record spans about the last half century. However, the fact that their time series are known to be contaminated by nonclimatic inhomogeneities,

specifically the effects of historical changes in instruments and recording practices (Parker and Cox 1995; Gaffen 1994), presents a significant complication. Linear trends derived from such data may exhibit considerable bias (Lanzante et al. 2003b).

A number of earlier studies compared vertical profiles of temperature change from climate models with those derived from radiosondes. The comparisons yielded a consistent discrepancy in that the warming of the tropical upper troposphere is greater in these atmosphere–ocean general circulation models (GCMs) than in the observations (Santer et al. 1996; Tett et al. 1996, 2002; Folland et al. 1998; Sexton et al. 2001; Hansen et al. 2002). In some instances, the discrepancy was attributed to error in the model or in the forcings used to drive the model (Tett et al. 1996; Sexton et al. 2001; Hansen et al. 2002). Only some of the radiosonde temperature data used in those studies were adjusted to account for nonclimatic changes over time, and then only in a very limited fashion (i.e., in the stratosphere starting in 1979). However, a subsequent diagnosis sug-

Corresponding author address: John R. Lanzante, NOAA/Geophysical Fluid Dynamics Laboratory, Princeton University, P.O. Box 308, Princeton, NJ 08542.
E-mail: John.Lanzante@noaa.gov

gested that nonclimatic changes have diminished the long-term warming in the tropical upper troposphere as deduced from radiosonde temperature data (Lanzante et al. 2003b), suggesting a possible explanation for the model–data discrepancy.

Only recently have radiosonde data that have been comprehensively adjusted to try to account for some of the artificial inhomogeneities become publicly available (Lanzante et al. 2003a; Thorne et al. 2005; Free et al. 2005). Santer et al. (2005) used these and other homogeneity adjusted data in an extensive comparison of temperatures in the tropical atmosphere with simulations from GCMs. Based on trend analyses, they found an inconsistency such that the amount of warming in the whole troposphere, as well as the vertical amplification of this warming (i.e., ratio of warming in the troposphere as compared to the surface) was smaller in the radiosonde data as compared to a wide array of model simulations. Stott et al. (2006) found that climate change simulations from one of the state-of-the-art climate models also exhibited a tropical upper-tropospheric maximum in warming that was not found in an adjusted radiosonde temperature dataset (Thorne et al. 2005). Further studies of another of the adjusted datasets (Lanzante et al. 2003a), involving diagnoses of radiosonde data alone (Sherwood et al. 2005), as well as comparisons with independent satellite data (Randel and Wu 2006) suggest that even the adjusted radiosonde temperature data are afflicted by spurious cooling trends.

In summary, evidence suggests that biases in the trends computed from radiosonde temperatures could account for discrepancies with models. However, attempts to remove these biases from the observations have yielded data whose trends, while in better agreement with models, still show important disagreements. For example, these data exhibit less tropospheric warming than the models (CCSP 2006). Although new interpretations of satellite records have improved understanding somewhat (Fu et al. 2004; Fu and Johanson 2005), unfortunately, satellite-derived temperatures offer no straightforward resolution of either the uncertainties in radiosonde trends or the inconsistencies with the GCMs. This follows from the fact that trends produced by different satellite adjustment teams diverge considerably (CCSP 2006).

Resolution of inconsistencies in trends of air temperatures of the free atmosphere remains a major outstanding issue in the science of climate change, although considerable progress has been made in the last several years (CCSP 2006). The complexity of data inhomogeneity issues insures that only a small segment of the climate community is likely to be well-versed in this

problem. But because of the impact, it is important for the broader community, to which this work is aimed, to appreciate these issues. Toward this end, this study is concerned with the explicit presentation of some uncertainties in temperature trends associated with inhomogeneities in radiosonde data in the context of comparisons with climate model simulations. More specifically, this work aims to illustrate the relative importance of data homogenization adjustment compared to the inclusion of some major climate forcings. This philosophy contrasts with that of most prior studies in which only the best (adjusted) estimate from the observations is shown, with no indication of the effect of adjustment.

A simple diagnostic approach is implemented, which should be considered a complement to formal detection–attribution (DA) analysis, rather than a competitor. While formal DA analysis could provide a more powerful alternative, some of its underlying assumptions may not always be met, and its greater complexity can make for more difficult interpretation, particularly for the nonexpert. Here the focus is on the vertical profile of linear trends. Trends, while a simplification of the expected signal, are nonetheless a convenient way to assess the low-order change in temperature. While temperature trends corresponding to a deep layer are examined, as was usually the case in prior studies, equal attention is paid to the shape of the trend profile, which has received much less attention. Since, in the course of evaluation, an interesting issue arose relating to the simulation of the climate response to major volcanic events, some diagnoses of this aspect are presented as well. These serve to illustrate the importance of volcanic effects even in a climate record nearly 40 yr in length.

2. Observed and model data

A radiosonde dataset that incorporates homogeneity adjustments intended to reduce the influence of the artificial, nonclimatic effects (Lanzante et al. 2003a) is used here in comparisons with climate model output from global change scenarios. These radiosonde temperature data, hereafter LKS, are derived from a global network of 87 stations and are available in both adjusted and raw (unadjusted) form. The value of the adjustments has been demonstrated via comparison with independent satellite-derived temperatures (Lanzante et al. 2003b). Nevertheless, there is ample evidence that biases remain even after adjustment (Sherwood et al. 2005; Santer et al. 2005; Randel and Wu 2006). The purpose of this work is diagnostic: to demonstrate the amount and manner in which data adjustment, as well the inclusion of particular climate forcings

in model simulations, make a difference in comparisons between observed and model trends. It should be kept in mind that the specific effects of data adjustment on the comparisons could be subject to change in the future as homogeneity adjustment procedures are further refined.

The climate model outputs used here (Delworth et al. 2002; Broccoli et al. 2003) were generated by a coupled ocean–atmosphere model developed at the Geophysical Fluid Dynamics Laboratory (GFDL) driven by some of the leading external climate forcings believed to have impacted the temperature record of the twentieth century. An ensemble of three simulations were available for each of four forcing scenarios: equivalent greenhouse gases only (G), G plus the direct effects of sulfate aerosols (GS), GS plus solar effects (GSS), and GSS plus volcanic effects (GSSV). The effects of stratospheric ozone depletion, as ascertained from a separate pair of simulations (Ramaswamy and Schwarzkopf 2002), have been incorporated as well. This was accomplished by assuming a simple linear change between separate equilibrium simulations based on 1979 or 1997 levels of stratospheric ozone and adding the resulting temperature change to the output from the coupled ocean–atmosphere model. This yields a set of four forcing scenarios, which are designated as GO, GSO, GSSO, and GSSVO. Since the primary output is taken from a model that, by design, has a more modest representation of the stratosphere, comparisons here focus on the troposphere.

For convenience, simulations from a single climate model have been used. One particularly appealing aspect of these experiments is the availability of an ensemble of runs for each of several different forcing scenarios. Furthermore, observations from a single observed dataset are used, although another homogeneity adjusted radiosonde temperature dataset, HadAT (Thorne et al. 2005) is available. The near universality of the amplification of warming in the upper tropical troposphere in response to anthropogenic forcing evidenced in the studies cited above, and particularly in the very comprehensive study by Santer et al. (2005), leads to the expectation that findings related to this issue will not be overly sensitive to the use of any particular model. More generally, simulations from any major climate model will suffice in illustrating the general degree of sensitivity of comparisons of temperature trend profiles, between model and radiosonde observations, to homogeneity adjustment of the observed data. In addition, the general similarity between LKS and HadAT regarding trends and the effect of adjustment (CCSP 2006) leads to the expectation that results will not be particularly sensitive to the observed dataset

that is employed. Indeed, work in progress finds that the results from the two observed datasets, as well as different climate models, are rather consistent. Those findings will be reported elsewhere as the focus of that study is somewhat different than this one.

3. Results

The linear trend for the time period 1959–97 is used as a metric of model–observation comparison. The ending date is dictated by the termination of the LKS dataset, which, due to the complex and laborious nature of the homogenization process, has not been extended any further. Separate vertical trend profiles for three nonoverlapping latitude zones were computed: the Tropics (TP), 30°N–30°S, and the extratropics of the Northern (NH) and Southern (SH) Hemispheres. The trend profiles shown at the top of Figs. 1a–c display results for both observations and model simulations. Although the characteristic shapes of the trend profiles vary by latitude zone, there is a consistent relationship amongst the model profiles such that the most complete scenario (GSSVO) yields considerably less tropospheric warming than the other three. Thus, the inclusion of volcanic forcing is seen to bring the model results into much better agreement with observations for the NH and TP. Another, more subtle difference is that compared to the others, GSSVO has a weaker rate of decrease with height above the surface for the NH, and a weaker rate of increase with height for the TP, both of which also appear to result in better agreement with observations. Volcanic cooling of the upper tropical troposphere relative to the surface, with more mixed effects in the extratropics, has been noted in observations (Free and Angell 2002). The inclusion of volcanic effects in model simulations has been shown to bring climate models into better agreement with observations (Santer et al. 2000, 2001; Broccoli et al. 2003).

The effects of data adjustment can be seen by comparing the unadjusted and adjusted LKS profiles with those from the model, especially for GSSVO, which uses the most realistic forcing. For the NH, adjustment results in warming, which leads to better agreement in the mid- and upper troposphere. Adjustment also enhances agreement for the TP by cooling the lower troposphere and surface, and especially by warming the upper troposphere. While adjustment appears to improve the model–observation agreement in the SH, the overall level of correspondence is poorer than in the other zones. However, results for the SH are less reliable and representative of that zone because of the much smaller number of stations (see Fig. 1 caption).

Although examination of vertical trend profiles is instructive, a more quantitative approach to profile com-

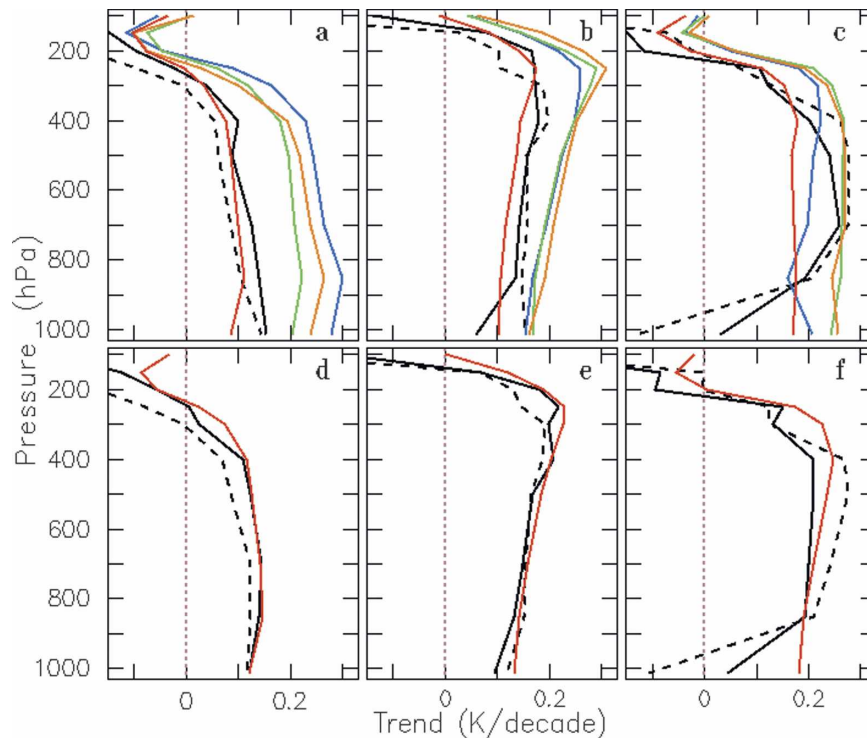


FIG. 1. Trend profiles for 1959–97 from model (GCM) and observations (LKS) for (a) the extratropics of the Northern Hemisphere (NH), 30°–90°N, (b) the Tropics (TP), 30°N–30°S, and (c) the extratropics of the Southern Hemisphere (SH), 30°–90°S. The four GCM scenarios are represented as follows: GO (blue), GSO (green), GSSO (orange), and GSSVO (red). The LKS data are represented by the black curves: unadjusted (dashed) and adjusted (solid). (d)–(f) Corresponding profiles that were computed after eliminating some volcanic time periods for LKS and GSSVO. The time periods that were eliminated are as follows: 3 yr after the Mt. Pinatubo eruption for tropospheric levels (surface–250 hPa for NH and SH; surface–150 hPa for TP) and 2 yr after each major volcano for higher levels (200 hPa and above for NH and SH; 100 hPa and above for TP). These are time periods for which the volcanic signature of the model departs most from observations. To facilitate a fair comparison, the model output (Delworth et al. 2002; Broccoli et al. 2003) has been sampled only at the LKS station locations, and for a given station only at levels for which a sufficiently representative record of observed data are available (Lanzante et al. 2003b). Trends have been computed using the median of pairwise slopes method (Lanzante 1996; Lanzante et al. 2003b). For the GCM, for each ensemble member the median of all available station trends at a given level was computed (Lanzante et al. 2003b); the values displayed are the average of the three ensemble medians. The number of station trends varies slightly by level, averaging ~34, 23, and 7 for the NH, TP, and SH, respectively.

parisons is desirable. Such comparisons are given in Fig. 2 using a bivariate scheme. Each symbol represents a comparison between trend profiles for one model forcing scenario and one observed data type (either unadjusted or adjusted). The horizontal coordinate is a relative bias term such that positive values indicate that, averaged over the troposphere, the model produces more warming (or less cooling) than the observations. The vertical coordinate measures the agreement in the shape of the vertical profiles in the troposphere. Significance testing of the statistics displayed in Fig. 2 is addressed in the appendix.

Examining Fig. 2 it is seen that for a given adjustment

type (i.e., symbol size), addition of a model forcing usually results in a better agreement; solar forcing is an exception. Addition of sulfate aerosols noticeably improves agreement in shape, due to dramatic improvements in the SH. Addition of volcanic forcing results in a large bias improvement by reducing the warming. The effect of data adjustment (i.e., going from small to large symbol for a particular color) can be as large as the addition of a model forcing and always improves agreement, slightly reducing bias (by enhancing data warming) and improving agreement in shape (most dramatically for the TP), often by a considerable amount.

Although volcanic forcing has improved the model–

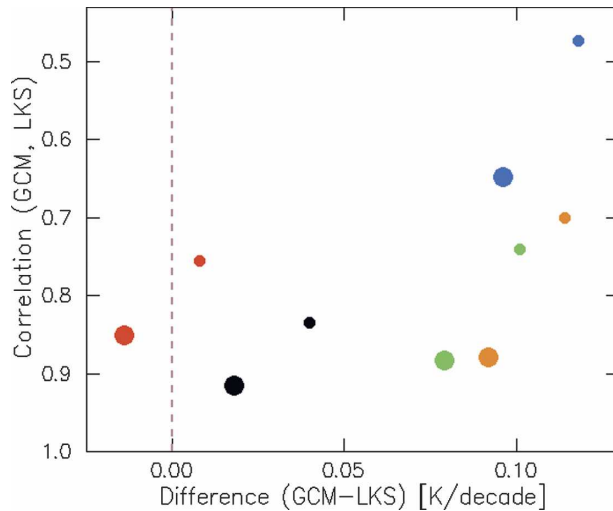


FIG. 2. Bivariate comparison statistics of pairs of model-observation trend profiles for 1959–97. The horizontal coordinate is the difference, GCM minus LKS, of the tropospheric-averaged temperature trend while the vertical coordinate is the correlation between the GCM and LKS profiles over the troposphere. The statistics are computed by aggregating over separate profiles for the NH, TP, and SH, weighting each zone by its vertically averaged number of station values. Perfect model-observation agreement would be indicated by a coordinate of (0, 1). Each circle designates a bivariate coordinate for the troposphere (surface–250 hPa for NH and SH; surface–150 hPa for TP). Smaller (larger) symbols correspond to unadjusted (adjusted) LKS data. Symbol color indicates model scenario: GO (blue), GSO (green), GSSO (orange), GSSVO (red), and DELVOL (black), which is GSSVO with deletion of some volcanic periods, as indicated in the caption for Fig. 1.

observation agreement considerably, its realism may be criticized because of the simplicity of the method of inclusion (Delworth et al. 2002; Broccoli et al. 2003), as well as evidence suggesting that the forcings employed (Andronova et al. 1999) may have been too strong, leading to excessive cooling (Broccoli et al. 2003). The global tropospheric time series shown in Fig. 3 indicate that the model and observations agree qualitatively in the sense that both show cooling after each major volcanic eruption. However, for the third event, Pinatubo, the model exhibits considerably more cooling than the observed data.

A simple diagnostic analysis aimed at quantifying the GCM–LKS agreement with regard to volcanic effects has been devised. A vertical profile was generated by computing the difference in temperature averaged some time interval after minus before each major volcanic event. This provides an estimate of the temperature change associated with the volcano. Next, the difference of these profiles between the GCM (GSSVO) and LKS was computed. A perfect agreement between model and observations would be indicated by a profile

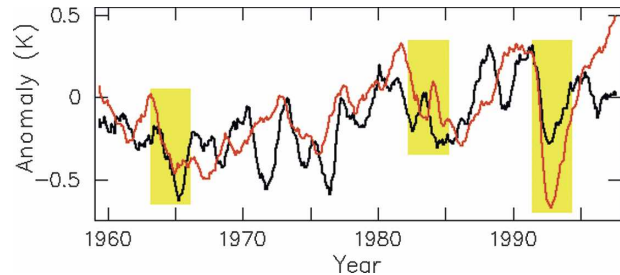


FIG. 3. Global time series of tropospheric temperature anomaly. The curves displayed (black for adjusted LKS and red for GSSVO) are smoothed, based on nine-point running averages of the monthly values. Yellow boxes, 3 yr after each major eruption, indicate approximately the period of largest volcanic response. To compute the time series, for a given month and vertical level, the median temperature anomaly was computed from all stations (Lanzante et al. 2003b). These monthly medians were then averaged over all levels in the troposphere (surface–250 hPa for NH and SH; surface–150 hPa for TP). Temperature anomalies were computed with respect to 1980–92 and the same station levels were used for the GCM and LKS.

identically equal to zero. Such profiles, as shown in Fig. 4, indicate that in the troposphere the agreement between model and observations is reasonable for the first two volcanoes (Agung and El Chichon); however, for Pinatubo the model produces noticeably more cooling than the observations. Significance testing of vertical averages of the profiles shown in Fig. 4 is addressed in the appendix.

To assess how these deficiencies may have an impact, sensitivity analyses were performed. Trends were recomputed while excluding, for tropospheric levels, the 3-yr time period after the Pinatubo eruption, and for stratospheric levels, two years after all three major volcanic eruptions, in both model and observations. The excluded periods represent times when the model is most suspect in simulating volcanic effects. While there is some evidence that the effect of volcanic eruptions on temperature may last longer than 3 yr in the troposphere (Santer et al. 2001), a major fraction of the cooling occurs by this time. This allows for an assessment of the sensitivity of the results to this effect, even if the effect has not been eliminated entirely.

An assessment of the potential effects of the model volcanic simulation deficiency can be made through examination of trend profiles by latitude zone based on volcanic exclusion (Figs. 1d–f). In comparing the top and bottom panels of Fig. 1 it is seen that exclusion leads to a warming of the model (GSSVO) relative to LKS and that the model-observation agreement improves. In the Tropics, the prior assertion that data adjustment can virtually eliminate the upper-tropospheric discrepancy between model and observations

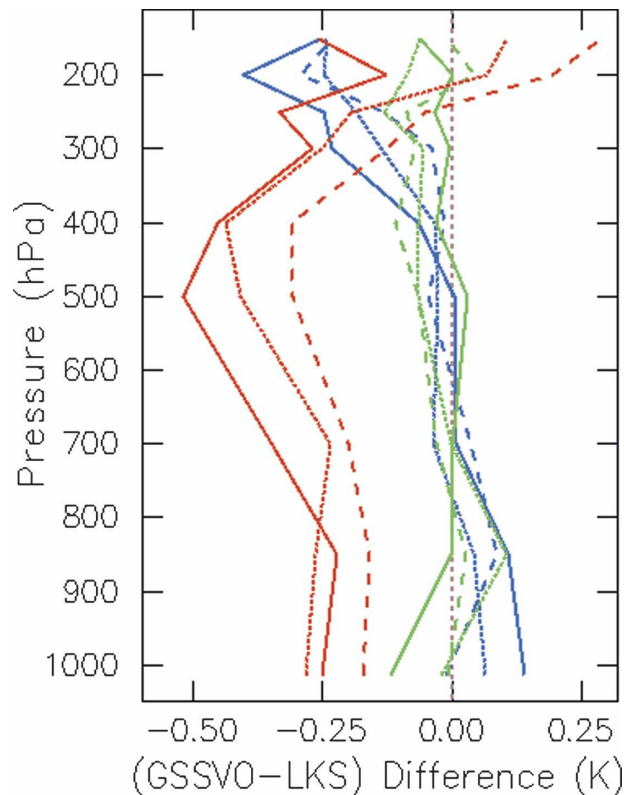


FIG. 4. Difference in volcanic profiles, GSSVO minus adjusted LKS. Differences are shown for Agung (blue), El Chichon (green), and Pinatubo (red) for window lengths of 24 (solid), 36 (dotted), and 48 (dashed) months. First, volcanic profiles were constructed for each station by computing, at each level, the difference in temperature anomaly for a certain interval (window), after minus before the month of eruption. Next, for each level, a median was computed over the values from all stations. For the GCM, the procedure was followed for each ensemble member and then an ensemble mean was formed. The profiles displayed in this figure are differences in the profiles described above, GSSVO minus adjusted LKS. These profiles were computed after the statistical removal of the effects of the El Niño–Southern Oscillation (ENSO) from tropospheric time series and quasi-biennial oscillation (QBO) from stratospheric time series using linear regression (Free and Angell 2002).

still holds, and if anything, appears stronger. Looking more closely it is seen that for the TP and SH, volcanic exclusion tilts the profiles by warming the upper relative to the lower troposphere, consistent with earlier work (Santer et al. 2000). These findings are quantified in Fig. 2, in which the bivariate statistics based on volcanic exclusion (black) can be compared to those based on the full time period (red). It is seen that volcanic exclusion changes the sign of the bias and, when considering the adjusted LKS, the magnitude hardly changes. However, volcanic exclusion increases agreement in the shape of the profiles.

4. Conclusions

It was found that homogeneity adjustment of atmospheric temperature time series from radiosondes leads to better agreement with regard to long-term trends from global change simulations of a particular climate model. Data adjustment appears to explain some of the discrepancy between climate models and observations that was previously attributed to model error. These conclusions are unaffected by a model deficiency in the simulation of volcanic effects. Nevertheless, the fidelity with which model simulations reproduce short-term volcanic effects can have a considerable impact on long-term trends.

One important point that was demonstrated is that homogeneity adjustment of the observed data can have as much influence as the addition of a major forcing to a climate model. This finding argues for the direction of much more effort toward understanding and reducing remaining temporal inhomogeneities in atmospheric temperature data. In addition, adjustment seems to unambiguously improve agreement with the climate model. The fact that the comparisons were limited to the examination of output from only a single climate model and one observed dataset would seem to warrant caution. However, similar comparisons, based on collaborative research in progress, involving output from other climate models, and an additional observational dataset render comparable findings. The results, which will be reported elsewhere, suggest a general insensitivity to the choice of model or observed datasets.

Acknowledgments. The G, GS, GSS, and GSSV model simulations were generated by the GFDL Climate Dynamics and Prediction group. Assistance and advice pertaining to these simulations were provided by Keith Dixon and Tony Broccoli. The stratospheric ozone depletion model simulations were generated by the GFDL Atmospheric Physics and Chemistry group. Assistance and advice pertaining to these simulations were provided by Dan Schwarzkopf and V. Ramaswamy. Thanks to reviewers Qiang Fu, Kostya Vinnikov, Ron Stouffer, and Tom Delworth for their comments.

APPENDIX

Statistical Significance

a. One-sample significance of trend profile comparison statistics

One-sample t tests have been used to determine if a given component (difference or correlation) of the bivariate statistics displayed in Fig. 2 is significant. One set of tests assesses whether the difference is zero, or

TABLE A1. Significance levels for two-tailed, one-sample t tests for components of selected statistics shown in Fig. 2. Each statistic (difference or correlation) is based on a comparison of trend profiles for an LKS data type (unadjusted or adjusted) and GCM scenario, as defined in the caption for Fig. 2, for the tropospheric layer. This tests if one coordinate of a symbol in Fig. 2 is significant. Values are significances $\times 100$, rounded to the nearest whole number.

	GCM	LKS	Statistic	
1	GO	Unadjusted	Difference	1
2	GO	Adjusted	Difference	1
3	GSO	Adjusted	Difference	11
4	GSSO	Adjusted	Difference	10
5	GSSVO	Adjusted	Difference	52
6	DELVOL	Adjusted	Difference	19
7	GO	Unadjusted	Correlation	10
8	GO	Adjusted	Correlation	8
9	GSSVO	Adjusted	Correlation	2
10	DELVOL	Adjusted	Correlation	<1

equivalently whether there is a difference in the layer-mean trend between the GCM and LKS. The other set of tests aims to determine if there is a significant correlation between GCM and LKS trend profiles. For both types of tests, the statistic of interest is calculated separately for each ensemble member; then a standard t test is applied to the sample of 3. In testing the correlation, the small-sample, bias-corrected Fisher's z transformation has been used (Zar 1996). Although caution is advised in interpreting the results given the small sample size, the test results help to place the differences in perspective.

Rather than performing the large number of all possible tests, test statistics for a selected number of cases have been computed for illustrative purposes and are shown in Table A1. Based on rows 1–2, for the simplest forcing scenario (greenhouse gases) it can be concluded that irrespective of data adjustment, the model is significantly warmer than observations in the troposphere. Rows 3–6 of Table A1 show that the addition of sulfate aerosols renders the difference nonsignificant. For the correlation metric (rows 7–10), GO approaches significance while the scenarios that include volcanic forcing are highly significant. Taken together, Table A1 and Fig. 2 suggest that the simplest scenario (GO) is not consistent with observations for the troposphere, but with the inclusion of sulfate aerosols and other forcings, it becomes more difficult to claim a significant difference between model and observations.

b. Two-sample significance of trend profile comparison statistics

Two-sample t tests have been used to determine if, for a given component (difference or correlation) of the

TABLE A2. Significance levels for two-tailed, two-sample t tests for component statistics of selected pairs of cases shown in Fig. 2. Each case represents a comparison of trends for a GCM scenario (as defined in the caption for Fig. 2) and an LKS data type [unadjusted (unad) or adjusted (adj)]. The t test tests the difference between the component statistic (difference or correlation of trend profiles between GCM and LKS) for the troposphere. This tests if the coordinates of a pair of symbols from Fig. 2 are significantly different. Note that for the difference statistic, significance implies a difference between the two cases, whereas for the correlation statistic, significance implies that the two cases agree. Values are significances $\times 100$, rounded to the nearest whole number.

	Case 1	Case 2	Diff	Corr
1	GO-Unad	GO-Adj	16	41
2	GSO-Unad	GSO-Adj	62	5
3	GSSO-Unad	GSSO-Adj	66	<1
4	GSSVO-Unad	GSSVO-Adj	43	26
5	DELVOL-Unad	DELVOL-Adj	17	3
6	GO-Adj	GSO-Adj	59	11
7	GSO-Adj	GSSO-Adj	77	94
8	GSSO-Adj	GSSVO-Adj	4	40
9	GSSVO-Adj	DELVOL-Adj	18	10
10	GSSVO-Unad	DELVOL-Unad	17	3
11	GSSO-Adj	DELVOL-Adj	8	7

bivariate statistics displayed in Fig. 2, there is a significant difference between two cases (each of which is defined by a GCM forcing scenario and an LKS adjustment type). By varying one factor between the two cases it is possible to determine if this factor results in a significant change. As for the one-sample tests, the statistics are calculated for each case separately for each ensemble member, and then a standard two-sample t test is applied with a sample size of 3. In testing the correlation, the small-sample, bias-corrected Fisher's z transformation has been used (Zar 1996). Again caution is advised in interpreting the results given the small sample size.

As seen in Table A2, rather than performing all possible tests, test statistics for a selected number of cases have been computed for illustrative purposes. Rows 1–5 of Table A2 show that the effect of data adjustment generally does not produce a significant change in the GCM–LKS difference, but often produces a significant change in the correlation. This indicates that in the troposphere adjustment has more of an effect on the shape of the profile than on the layer-averaged trend.

Rows 6–8 in Table A2 demonstrate the effect of incrementally adding a model forcing and in most cases this does not produce a significant change. One noteworthy exception is the addition of volcanic forcing, which as seen in Fig. 2 significantly reduces the tropospheric GCM–LKS difference by cooling the GCM. In synthesizing the results of rows 1–5 as compared to

results from rows 6–8 it is worth noting that it has been shown quantitatively that the effect of data adjustment is often more influential than that of the addition of a model forcing, which is the most important conclusion of this study.

Rows 9–10 in Table A2 illustrate the effect of deleting some volcanic time periods that have not been well simulated by the model. In the troposphere, where deletion is limited to the post-Pinatubo time period, we see that there is a borderline significant change in agreement in the shape of the profile. Figure 2 reminds us that the change corresponds to an improvement, consistent with the notion that the model response to Pinatubo is substandard. Finally, rows 8 and 11 of Table A2 indicate that irrespective of the model's shortcomings in simulating Pinatubo, the addition of volcanic forcing has a borderline significant effect in reducing the GCM–LKS difference in the troposphere by cooling the model. This points out that even using a fairly long data record (~40 yr), the effects of volcanic eruptions have a considerable impact on the temperature trend for the period.

c. Significance of difference in volcanic profiles

Tests of the statistical significance of layer-averaged temperatures corresponding to the volcanic difference profiles shown in Fig. 4 have been performed. The quantity of interest is the difference, GSSVO minus LKS, which was computed at each level, separately for each of the three ensemble members. Vertical averages were formed from these for the troposphere using the same levels as defined in the caption for Fig. 2. A one-sample t test was applied to the ensemble of three values. To assess sensitivity, tests were performed (a) using window lengths of 24, 36, and 48 months, (b) using unadjusted and adjusted LKS data, and (c) both with and without the removal of the ENSO signal using linear regression. The calculations were repeated for each of the three major volcanic eruptions. Given the small sample size some caution is required in the interpretation of the results. However, it is instructive to examine variations in the calculated t values among different cases in a relative sense. For a sample size of 3, a t value of about 4.3 is required for two-tailed significance at the 5% level.

The t values shown in Table A3 indicate that in the troposphere none of the tests even approach significance for Agung or El Chichon. However, for Pinatubo the magnitudes are much larger and about half of them approach or exceed the critical value, particularly for the shorter window lengths and with the removal of the ENSO signal. For Pinatubo the sign is always negative, indicating that the posteruption drop in temperature is

TABLE A3. Student's t statistics testing differences in tropospheric layer averages from Fig. 4. A separate test was performed for each volcano (Agung, El Chichon, and Pinatubo), for three window lengths (24, 36, and 48 months), for LKS data that has been adjusted (adj) or not adjusted (unad), and based on time series for which the ENSO signal was either removed (Free and Angell 2002) or not removed. For the sample size of 3 (ensemble members) the test statistic is significant at the 5% level for a two-tailed test if its magnitude exceeds 4.3.

Volcano	Unad			Adj		
	24	36	48	24	36	48
Agung						
Not Removed	+0.02	+0.08	+0.00	-0.11	-0.26	-0.53
Removed	-0.37	+0.00	+1.95	-0.59	-0.80	-0.39
El Chichon						
Not Removed	+0.02	-0.15	-0.95	+0.01	-0.19	-0.78
Removed	-0.21	-0.75	-1.57	+0.09	-0.80	-1.24
Pinatubo						
Not Removed	-4.14	-3.00	-1.57	-4.21	-3.06	-1.81
Removed	-5.20	-7.91	-3.17	-5.07	-8.30	-3.33

always greater in the model than in the observations. These results are in accord with the visual perception from Figs. 3 and 4 that the volcanic response of the model in the troposphere is consistent with observations for Agung and El Chichon, but not likely for Pinatubo. In the stratosphere (not shown), as expected, the model is inconsistent with the observations for all three volcanoes since it produces cooling, whereas observations indicate warming.

REFERENCES

- Andronova, N. G., E. V. Rozanov, F. Yang, M. E. Schlesinger, and G. L. Stenchikov, 1999: Radiative forcing by volcanic aerosols from 1850 to 1994. *J. Geophys. Res.*, **104**, 16 807–16 826.
- Broccoli, A. J., K. W. Dixon, T. L. Delworth, T. R. Knutson, R. J. Stouffer, and F. Zeng, 2003: Twentieth-century temperature and precipitation trends in ensemble climate simulations including natural and anthropogenic forcing. *J. Geophys. Res.*, **108**, 4798, doi:10.1029/2003JD003812.
- CCSP, 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 Rep. 1.1, T. R. Karl et al., Eds., National Oceanic and Atmospheric Administration, National Climatic Data Center, 180 pp. [Available online at <http://www.climate-science.gov/Library/sap/sap1-1/finalreport/default.htm>.]
- Delworth, T. L., R. J. Stouffer, K. W. Dixon, M. J. Spelman, T. R. Knutson, A. J. Broccoli, P. J. Kushner, and R. E. Wetherald, 2002: Review of simulations of climate variability and change with the GFDL R30 coupled climate model. *Climate Dyn.*, **19**, 555–574.
- Folland, C. K., D. M. H. Sexton, D. J. Karoly, C. E. Johnson, D. P. Rowell, and D. E. Parker, 1998: Influences of anthropogenic

- and oceanic forcing on recent climate change. *Geophys. Res. Lett.*, **25**, 353–356.
- Free, M. and J. K. Angell, 2002: Effect of volcanoes on the vertical temperature profile in radiosonde data. *J. Geophys. Res.*, **107**, 4101, doi:10.1029/2001JD001128.
- , D. J. Seidel, J. K. Angell, J. Lanzante, I. Durre, and T. C. Peterson, 2005: Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC): A new dataset of large-area anomaly time series. *J. Geophys. Res.*, **110**, D22101, doi:10.1029/2005JD006169.
- Fu, Q., and C. M. Johanson, 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.
- Gaffen, D. J., 1994: Temporal inhomogeneities in radiosonde temperature records. *J. Geophys. Res.*, **99**, 3667–3676.
- Hansen, J., and Coauthors, 2002: Climate forcings in Goddard Institute for Space Studies SI2000 simulations. *J. Geophys. Res.*, **107**, 4347, doi:10.1029/2001JD001143.
- Lanzante, J. R., 1996: Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.*, **16**, 1197–1226.
- , S. A. Klein, and D. J. Seidel, 2003a, 2003b: Temporal homogenization of monthly radiosonde temperature data. Part I: Methodology. *J. Climate*, **16**, 224–240.
- , —, and —, 2003b: Temporal homogenization of monthly radiosonde temperature data. Part II: Trends, sensitivities, and MSU comparison. *J. Climate*, **16**, 241–262.
- Parker, D. E., and D. I. Cox, 1995: Towards a consistent global climatological rawinsonde data-base. *Int. J. Climatol.*, **15**, 473–496.
- Ramaswamy, V., and M. D. Schwarzkopf, 2002: Effects of ozone and well-mixed gases on annual-mean stratospheric temperature trends. *Geophys. Res. Lett.*, **29**, 2064, doi:10.1029/2002GL015141.
- Randel, W. J., and F. Wu, 2006: Biases in stratospheric and tropospheric temperature trends derived from historical radiosonde data. *J. Climate*, **19**, 2094–2104.
- Santer, B. D., and Coauthors, 1996: A search for human influences on the thermal structure of the atmosphere. *Nature*, **382**, 39–46.
- , and Coauthors, 2000: Interpreting differential temperature trends at the surface and in the lower troposphere. *Science*, **287**, 1227–1232.
- , and Coauthors, 2001: Accounting for the effects of volcanoes and ENSO in comparisons of modelled and observed temperature trends. *J. Geophys. Res.*, **106**, 28 033–28 060.
- , and Coauthors, 2005: Amplification of surface temperature trends and variability in the tropical atmosphere. *Science*, **309**, 1551–1556.
- Sexton, D. M. H., D. P. Rowell, C. K. Folland, and D. J. Karoly, 2001: Detection of anthropogenic climate change using an atmospheric GCM. *Climate Dyn.*, **17**, 669–685.
- Sherwood, S. C., J. R. Lanzante, and C. L. Meyer, 2005: Radiosonde daytime biases and late-20th century warming. *Science*, **309**, 1556–1559.
- Stott, P. A., G. S. Jones, J. A. Lowe, P. Thorne, C. Durman, T. C. Johns, and J.-C. Thelen, 2006: Transient climate simulations with the HadGEM1 climate model: Causes of past warming and future climate change. *J. Climate*, **19**, 2763–2782.
- Tett, S. F. B., J. F. B. Mitchell, D. E. Parker, and M. R. Allen, 1996: Human influence on the atmospheric vertical temperature structure: Detection and observations. *Science*, **274**, 1170–1173.
- , and Coauthors, 2002: Estimation of natural and anthropogenic contributions to twentieth century temperature change. *J. Geophys. Res.*, **107**, 4306, doi:10.1029/2000JD000028.
- Thorne, P. W., D. E. Parker, S. F. B. Tett, P. D. Jones, M. McCarthy, H. Coleman, and P. Brohan, 2005: Revisiting radiosonde upper air temperatures from 1958 to 2002. *J. Geophys. Res.*, **110**, D18105, doi:10.1029/2004JD005753.
- Zar, J. H., 1996: *Biostatistical Analysis*. 3d ed. Prentice Hall, 376 pp.