Error Estimates of Version 5.0 of MSU–AMSU Bulk Atmospheric Temperatures

JOHN R. CHRISTY, ROY W. SPENCER, WILLIAM B. NORRIS, AND WILLIAM D. BRASWELL

Earth System Science Center, University of Alabama in Huntsville, Hunstville, Alabama

DAVID E. PARKER

Hadley Centre, Met Office, Berkshire, United Kingdom

(Manuscript received 16 May 2002, in final form 14 November 2002)

ABSTRACT

Deep-layer temperatures derived from satellite-borne microwave sensors since 1979 are revised (version 5.0) to account for 1) a change from microwave sounding units (MSUs) to the advanced MSUs (AMSUs) and 2) an improved diurnal drift adjustment for tropospheric products. AMSU data, beginning in 1998, show characteristics indistinguishable from the earlier MSU products. MSU–AMSU error estimates are calculated through comparisons with radiosonde-simulated bulk temperatures for the low–middle troposphere (TLT), midtroposphere (TMT), and lower stratosphere (TLS.) Monthly (annual) standard errors for global mean anomalies of TLT satellite temperatures are estimated at 0.10°C (0.07°C). The TLT (TMT) trend for January 1979 to April 2002 is estimated as +0.06° (+0.02°) ±0.05°C decade^{-1} (95% confidence interval). Error estimates for TLS temperatures are less well characterized due to significant heterogeneities in the radiosonde data at high altitudes, though evidence is presented to suggest that since 1979 the trend is −0.51° ± 0.10°C decade^{-1}.

1. Introduction

The time series beginning in 1979 of bulk atmospheric temperatures derived from microwave emissions near the 60-GHz oxygen absorption band now incorporate measurements from 11 National Oceanic and Atmospheric Administration (NOAA) polar-orbiting spacecraft and two types of instruments. In constructing these datasets we have always attempted to satisfy climate research requirements so as to provide the maximum climate utility even though the original purpose of these measurements was to initialize short-term weather forecast models (NRC 1999; 2000a,b). There have been five revisions of these products as nonclimatic influences were discovered and removed (see Christy et al. 2000, hereafter CSB). In version A, a simple bias removal was performed between the satellite records (Spencer and Christy 1990). In version B, a linear approximation for *NOAA-11* diurnal drift, at the time only recently discovered, was applied (along with other minor adjustments; Christy et al. 1995.) In Christy et al. (1998; version C) further diurnal adjustments were applied to the other satellites, a filtering procedure employed, and error analyses were performed. A full description of the construction techniques applied to version D is found in CSB where adjustments for orbital decay and instrument calibration were applied. The satellite products described below will be referred to as version 5.0 (rather than ''E'').

Every orbiting instrument utilized in our products has experienced at least some overlapping observations with another, a key feature necessary for merging. Most overlapping periods are 2 yr or longer, allowing precise determination of instrument biases and relative drifts to be assessed. The early spacecraft were plagued by shorter overlap periods (Christy et al. 1998), but the influence of those early potential errors on the time series is continually reduced as the overall period of observations lengthens. Also, as more records are merged, random errors tend to have a smaller net influence on trends.

We have two main goals in this study. First, we shall present the updated version 5.0 of the temperature products produced at the University of Alabama in Huntsville (UAH). In particular we shall describe and test 1) the methodology for improving the diurnal drift correction and 2) the procedure for continuing the time series using data from the new microwave instruments. Second, we shall investigate the error characteristics of the time series on scales ranging from monthly anomalies of a single grid to annual anomalies of global averages. In so doing we show a discovery of a discontinuity in U.S. radiosonde data related to a slight change in instrumentation in 1997.

It is important to note that this study addresses only

Corresponding author address: Dr. John R. Christy, Earth System Science Center, University of Alabama in Huntsville, Huntsville, AL 35899.
E-mail: john.christy@nsstc.uah.edu

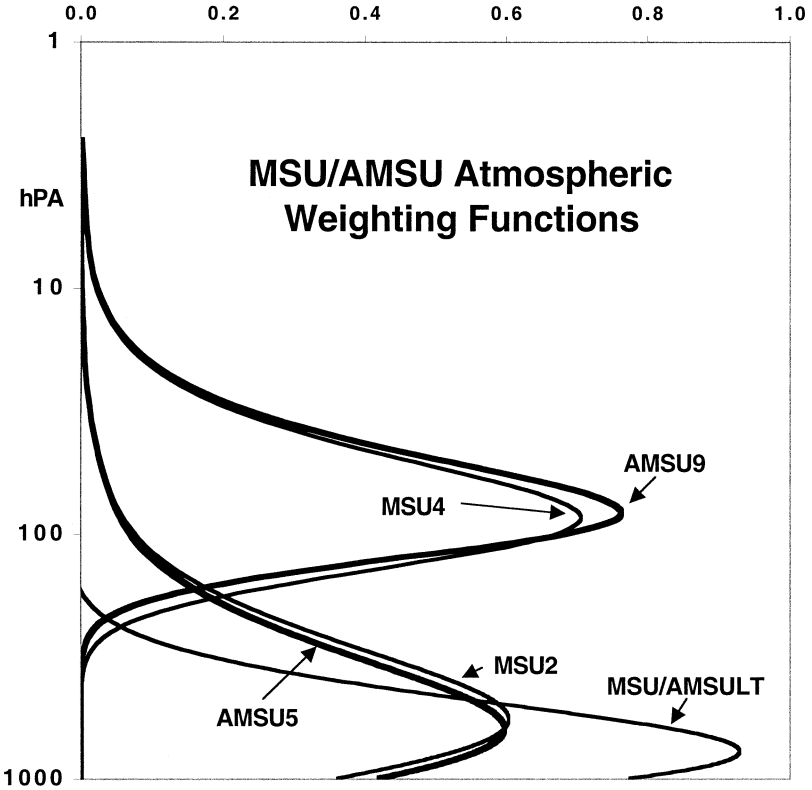

FIG. 1. Static atmospheric weighting function profile for the MSU and AMSU products. Not shown is the surface contribution factors, which for land (ocean) for TLT are 0.20 (0.10) and for TMT are 0.10 (0.05) of the total weighted profile. The land surface contribution increases for higher surface altitudes.

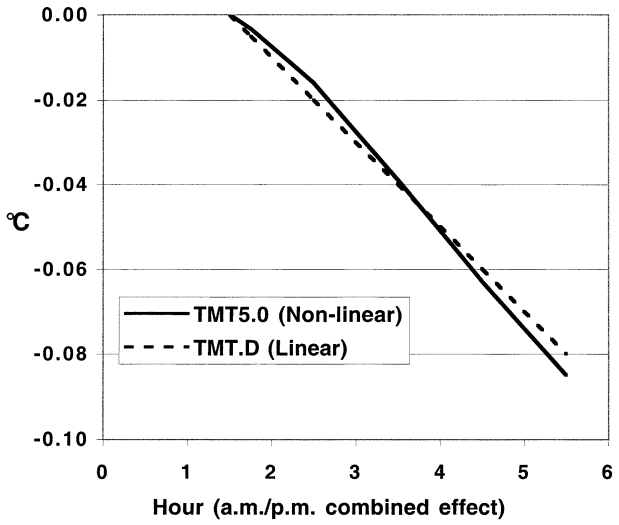

FIG. 2. A comparison of the annual, global mean temperature error for MSU2 due to diurnal drifting of the afternoon satellites (see text). The values are the net drift effect from 0130 to 0530 and 1330 to 1730 combined.

one type of error in the satellite products: measurement error. We do not delve into the issue of sampling error, that is, whether a time series of a given length answers questions of trends for longer periods (Folland et al. 2001).

2. Recent developments in UAH microwave products

The three bulk atmospheric temperature products generated since our effort began were known as microwave sounding unit (MSU) data (vertical profiles in Fig. 1). Each of the four MSU channels represents a frequency that detects emissions from a vertical atmospheric profile dependent upon the frequency and view angle from nadir (there are 11 view angles per scan). However, *NOAA-14*, placed in service in April 1995, was the last to carry an MSU. Since the commissioning of *NOAA-15* in September 1998, with its advanced MSU (AMSU), there are new data streams to accommodate and explore. The AMSU-A instrument has considerably increased the monitoring capability, using 15 channels with 30 view angles per scan.[1] The final products examined here are produced on a global, 2.5° rectangular grid.

[1] The other component of AMSU, AMSU-B, monitors five channels with 90 view angles per scan.

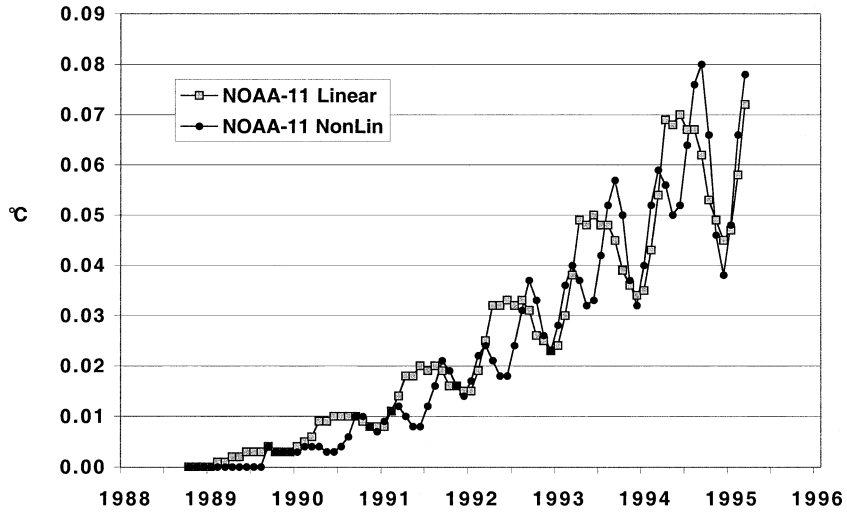

FIG. 3. Monthly global mean errors for MSU2 on *NOAA-11* as it drifted from a 0130/1330 local crossing time in 1988 to almost 0530/1730 local crossing time in 1995.

Version D becomes version 5.0

During its lifetime, a polar-orbiting spacecraft experiences some east–west drift relative to the sun (or local time); hence, it is not technically correct to refer to these as absolutely ''sun synchronous.'' A spacecraft initially crossing the equator at 1400 LT (and 0200 LT) typically drifts to 1700 LT over a 3–4-yr period. As such, the diurnal cooling experienced on earth between 1400 and 1700 LT (and 0200 and 0500 LT) is measured by the satellite, becoming a source of spurious trends. In version C we applied a linear latitudinal correction based solely on intersatellite comparisons (Christy et al. 1998.) In version D, a direct calculation of the effect using the cross-swath differences dependent upon latitude, land/ocean, and month of year was performed.

The diurnal adjustment in version D was determined as a linear (constant) change of temperature (T) per hour of drift (Δt), so that

$$\Delta T_{\text{adjustment}} = a_1(\Delta t_{\text{observed drift}}),$$

where a_1 represented a table of constants dependent on latitude, surface type, and month.[2] It is important to note, as explained in CSB, that the rate of change (a_1) was determined directly from the cross-swath data differences, which eliminated spurious effects from both intersatellite biases and interannual temperature variations (a significant advantage.)

In version 5.0 we use the same mapped cross-swath data though we perform a quadratic rather than a linear approximation to the data. For each given month, latitude, and surface type we generate a plot (or functional relationship) between local time versus the average change in temperature from the initial reference time. A quadratic equation is fitted producing constants a_1 and a_2 (see equation below). The ''constants'' are further modified, slightly, so that their values form a smooth seasonal cycle (to the second harmonic) for each latitude and surface type. Thus, for a given latitude, surface type, and orbit node (A.M. or P.M.) there are 12 values of a_1 and a_2, one for each month of the year. In other words, we have calculated a diurnal cycle for each latitude, surface type, orbit node, and month of the year:

$$\Delta T_{\text{adjustment}} = a_1(\Delta t_{\text{observed drift}}) + a_2(\Delta t_{\text{observed drift}})^2$$

In Fig. 2 we show the global, annual mean errors accounting for the diurnal drift for the ''afternoon'' satellites, *TIROS-N, NOAA-7, -9, -11,* and *-14.* The effect of applying a nonlinear approximation is quite small but does improve the error statistics for the tropospheric datasets. *NOAA-11* drifted significantly (from 1330 in late 1988 to 1730 local equatorial crossing time in early 1995), and we show in Fig. 3 the calculated adjustment for both linear and nonlinear monthly approximations. Near the end, the adjustments tend to be greater in the nonlinear approximation because cooling accelerates toward sunset. Since Fig. 3 represents the adjustment added to the data, this indicates that a compensation for a spurious cooling (seen in Fig. 2) is being applied to *NOAA-11.*

An analysis of version 5.0 products indicates slight

[2] Consider a northbound spacecraft looking down on the earth. The center of the swath (footprint 6 of the 1–11 views) observes the temperature at a local time, say 1500. The scanner sweeps left to right during its 25.6-s full rotation. Footprints to the right of 6 will represent later local times and those to the left, earlier. The almost-simultaneous difference between the temperatures of the right and left footprints (which may be from two adjacent swaths in high latitudes) will represent the change in the diurnal temperature over the time represented by the difference in local time of the footprint data used. The ''cross-swath'' differences are calculated from data after binning into grids by view angle and local time, so that the effect of swath angle rotation at higher latitudes is avoided, allowing data from adjacent swaths but of the same latitude to be used in the cross-swath difference at higher latitudes.

TABLE 1. Comparison metrics between version D and version 5.0 (nonlinear diurnal correction). Note that version 5.0 of TLS is shown as nonlinear (n.l.) though the actual TLS5.0 uses the linear approximation.

Jan 1979–Mar 2002	TLT	TMT	TLS
Version	D vs 5.0	D vs 5.0	D vs n.l.5.0
Trend °C decade^{-1}	+0.053, +0.061	+0.041, +0.013	0.509, −0.498
$\sigma_{\Delta\,\mathrm{DAY}}$ °C	0.069, 0.067	0.034, 0.033	0.047, 0.049
Signal/noise (DY)	37.9, 38.9	142, 149	187, 172
Rms satellite trend (> 1-yr overlap)	0.033, 0.028	0.010, 0.008	0.022, 0.027

changes in the intersatellite error statistics as summarized in Table 1. The three statistical metrics are 1) $\sigma_{\Delta\mathrm{DAY}}$, meaning the standard deviation of the difference in daily global temperature anomalies between every pair of co-orbiting satellites, 2) the ratio of the variance of the daily global anomalies $(\sigma_{\mathrm{DAY}})^2$ versus the variance of the intersatellite differences $(\sigma_{\Delta\mathrm{DAY}})^2$, and 3) the root-mean square (rms) of the intersatellite trend differences measured for each overlapping period of observations. For both tropospheric datasets, temperature of the low–middle troposphere (TLT; see below) and temperature of the midtroposphere (TMT) there are slight improvements in the intersatellite error characteristics, so this nonlinear approximation is adopted.[3] For temperature of the lower stratosphere (TLS), a slight degradation occurs, so we shall remain with the linear approximation there. Hence, TLS in versions D and 5.0 will be the same except for the addition of AMSU data.

1) T_{2LT} BECOMES TLT

The three products produced originally with MSUs were known as T_{2LT}, T_2, and T_4 (Fig. 1). Here T_{2LT} represents the low–middle troposphere and is derived from a linear combination of 8 of the MSU channel 2 (MSU2) 11 view angles to remove stratospheric and upper-tropospheric emissions (Spencer and Christy 1992b). Note that T_{2LT} differs from T_2 in that the temperatures of the differing view angles in T_2 data are not linearly combined to remove the stratosphere emissions [see section 2a(2)]. About 90% of the emissions originate below 400 hPa in T_{2LT} and we were able to reproduce the same weighting profile using AMSU channel 5 (AMSU5) with a linear combination of the 30 view angles' brightness temperatures.[4] Since MSU2 data will no longer be available once *NOAA-14* is decommissioned, we shall hereafter refer to this product as TLT.

One means of estimating errors in these data is to compare simultaneous observations from co-orbiting (overlapping) satellite pairs. In this way, the earth itself becomes the common calibration target observed by both instruments, as in $\sigma_{\Delta\mathrm{DAY}}$ described above. Comparing global TLT observations from *NOAA-14*'s MSU and *NOAA-15*'s AMSU we find $\sigma_{\Delta\mathrm{DAY}} = 0.059$°C, a value less than those calculated between the other MSU pairs (Table 2.) And $\sigma_{\Delta\mathrm{DAY}}$ of TLT between the two AMSUs for their 15 months of overlap (Feb 2001–Apr 2002) is 0.050°C, suggesting that there has not been a loss but an enhancement of precision with the new instrumentation.

2) T_2 BECOMES TMT

The vertical profile of T_2 (MSU2) is similar to AMSU5, though a bit higher in altitude, with 90% of the emissions originating in the layer below 120 hPa (Spencer and Christy 1990.) We have selected to use AMSU5 to continue the product represented by T_2. Due to its slightly lower profile, AMSU5's typical brightness temperature is 2.1°C warmer than T_2.[5] As shown in Table 2, this difference in profile is not evident in $\sigma_{\Delta\mathrm{DAY}}$, which from MSU2 (*NOAA-14*) versus AMSU5 (*NOAA-15*) is 0.040°C and is within the MSU2 versus MSU2 range. The two AMSU5s on *NOAA-15* and *NOAA-16* produce $\sigma_{\Delta\mathrm{DAY}}$ of 0.030°C. Though slightly different in profile,

[3] The difference between the linear and nonlinear diurnal adjustment has a slight dependence on view angle. TLT utilizes a difference between view angles near the limb, and the net effect was largely cancelled. TMT is the average of the middle view angles, the limb-most of which were affected but which were not compensated by the impact of view angles further to the limb as in TLT. Thus, the impact of the nonlinear correction on TMT is slightly greater than TLT, as seen in the slightly larger trend difference compared with version D.

TABLE 2. Comparison of $\sigma_{\Delta\,\mathrm{DAY}}$ (std dev of differences in global mean temperature anomaly) determined from co-orbiting classes of instruments.

	Values of global $\sigma_{\Delta\,\mathrm{DAY}}$ (°C)		
Atmospheric layer	Between MSUs	Between MSU and AMSU	Between AMSUs
LT	0.061–0.080	0.059	0.050
MT	0.026–0.047	0.040	0.030
LS	0.032–0.071	0.054	0.042

[4] We considered several multichannel representations of TLT, but in comparison tests with concurrently observed MSU TLT, we found significant increases in noise using multichannel retrievals versus the AMSU5/multiview-angle method.

[5] AMSU5 monitors two bands centered on 53.596 GHz while MSU2 monitors 53.74 GHz.

these results indicate that AMSU5 is sufficiently similar to MSU2 to continue the time series.[6]

The MSU2–AMSU5 product will be referred to as TMT rather than T_2. We note that the physical interpretation of variations of TMT is difficult to explain as emissions from the entire atmospheric column contribute to TMT. As troposphere and stratosphere vary, sometimes in strong anticorrelation (e.g., tropospheric cooling, stratospheric warming following volcanic events), it is unclear how to attribute responses for such a profile (CSB).

3) T_4 BECOMES TLS

The vertical profile of T_4 (MSU4) is very similar to AMSU9, with 90% of the emissions generated from the layer between 150 and 20 hPa.[7] About 20% of T4's emissions originate below the tropopause in the Tropics; thus, T_4 does not exclusively represent stratospheric temperature variations. The bias between AMSU9 (*NOAA-15*) and MSU4 (*NOAA-14*) is only 0.32°C, a value within the range of biases determined among MSU4s themselves. Again, the value of $\sigma_{\Delta DAY}$ calculated between MSU4 (*NOAA-14*) versus AMSU9 (*NOAA-15*) is 0.054°C, and that of the AMSU9s on *NOAA-15* versus *NOAA-16* is 0.042°C, both being in the range of MSU4 versus MSU4 (Table 2.) Thus, the error characteristics of MSU4 versus AMSU9 are indistinguishable from those of intercompared MSU4s. As noted, we shall designate this product as TLS.

The statistical tests performed on the MSU versus AMSU data indicate that the time series of these three products, TLT, TMT, and TLS, are sufficiently homogeneous to justify their continuance. However, additional and independent evaluation of these products is necessary, which will also assess the impact of the accumulation of all sources of measurement error and nonthermometric effects (e.g., humidity, surface emission changes, etc.; see Spencer et al. 1990).

3. Radiosonde comparisons

Intercomparing data records observed by separate coorbiting microwave instruments provides one means of error testing (e.g., Tables 1, 2.) However, because one of the correction factors (the instrument heating adjustment) required information gleaned from the overlapping observations themselves, the measured intersatellite agreement is in some sense contrived (CSB). What is needed is a completely independent set of observations with which to evaluate the satellite products.

Radiosonde observations provide pressure-level temperature readings from which a weighted average may be calculated to match the microwave vertical profiles. We use a static weighting function (Fig. 1) applied to the radiosonde data. In the first two analyses below, Minqin (China) and the 28-station composite (appendix), we utilized the Comprehensive Aerological Reference Data Set (CARDS),[8] which consists of individual soundings from several hundred stations distributed worldwide (Eskridge et al. 1995). Temperatures from each acceptable sounding were interpolated to a 5-hPa vertical grid and the appropriate static weighting function was then applied. Spencer and Christy (1992a) compared static weighting functions versus the full radiation model (which includes humidity), showing a degradation in the monthly standard errors of only 0.02°C, which is well within the magnitude of the error statistics discussed below.

A radiosonde-simulated temperature was generated if the sounding contained at least the mandatory levels. Anomalies were calculated for each sounding (0000 and 1200 UTC separately) based on the 22-yr mean, daily annual cycle for each layer and observed time. For each observing time 10 observations were required to calculate a monthly average from the daily anomalies. Linear interpolation was applied to fill temporal gaps not longer than 20 days for some levels, though missing data periods of this length were rare. Several stations reported consistently only once per day, 0000 or 1200 UTC, so their anomalies were based on the appropriately observed time.[9]

Unfortunately, virtually all radiosonde datasets suffer from some type of discontinuity due to changes in instrumentation, software processing, and station moves (Gaffen 1994; Gaffen et al. 2000; NRC 2000b). Below, we shall note the possible impacts of some of these problems. Radiosonde-simulated temperatures will be referred to as RLT, RMT, and RLS whose profiles are designed to match those of TLT, TMT, and TLS, respectively. The comparison studies shall be divided into three categories: 1) a single station, 2) multistation composites, and 3) global averages from independent research groups.

a. Minqin, China: A single station comparison

We use Minqin, China (38.6°N, 103.1°E, elevation 1367 m), to introduce the comparison method because

[6] We attempted to create a multichannel AMSU retrieval to match MSU2 from various combinations of AMSU channels (specifically 5–8), but the noise characteristics were greater than those of AMSU5 alone. The AMSU5 linear coefficients for view angles 1 (limb) to 15 (symmetric with 16 to 30) are −1.74, −0.57, 0.00, 0.42, 0.40, 0.56, 0.54, 0.39, 0.44, 0.26, 0.00, 0.26, 0.00, 0.00, and 0.04.

[7] AMSU9 monitors two bands centered on 57.290344 GHz while MSU4 monitors 57.95 GHz.

[8] We used the first version of CARDS, though another version is now being processed (I. Durre 2002, personal communication).

[9] A full description of the Climatological Averaging of Temperature Soundings (CATS) code and user's guide, which produced the radiosonde-simulated satellite temperatures authored by W.B. Norris and J.R. Christy, is available online at http://vortex.nsstc.uah.edu/atmos/cats/cats.html.

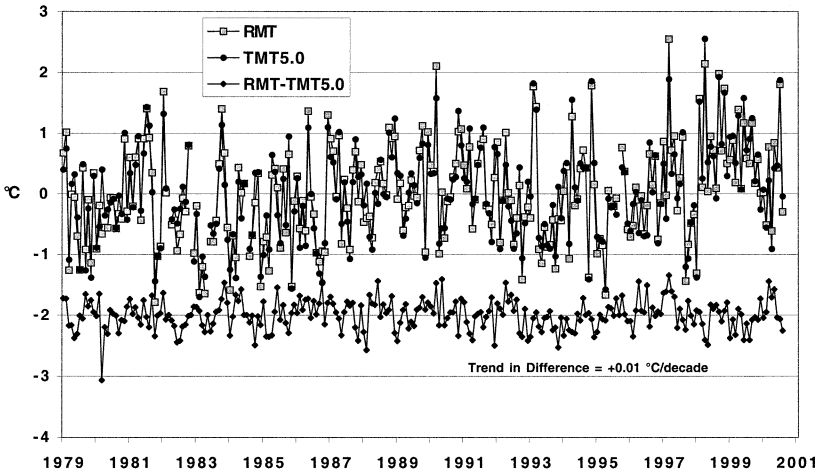

FIG. 4. Radiosonde and satellite comparison of monthly temperature anomalies for the midtroposphere (RMT and TMT) at Minqin, China. Monthly differences are shown offset 2°C in the lower time series for clarity.

Minqin is almost unique among radiosonde stations. Minqin radiosonde data are virtually complete for both 0000 and 1200 UTC and utilize consistent instrumentation for the 22-yr period of comparison (1979–2000.) Minqin was chosen independent of this study, being selected by NOAA as 1 of 12 stations for examination at the upper-air workshop in Asheville, North Carolina, October 2000 (Free et al. 2002). To our knowledge, Minqin temperatures had, previous to the workshop, not been compared with satellite products.

The basic statistical metrics used are monthly and annual correlation of anomalies (r_{MON}, r_{ANN}), trend per decade (b, the slope of the regression line of best fit), standard deviation of the monthly and annual anomalies (σ_{MON}, σ_{ANN}), and standard deviation of the monthly and annual difference of anomalies between two temperature datasets ($\sigma_{\Delta\mathrm{MON}}$, $\sigma_{\Delta\mathrm{ANN}}$). The Minqin results are shown in Fig. 4 for TMT versus RMT and Table 3 for all three layers.

The agreement shown in Table 3 between the two independent systems is quite remarkable. In previous work we performed many other individual comparisons over shorter periods (Spencer and Christy 1992a,b; Christy et al. 1998; CSB) and were not anticipating this level of agreement. At Minqin, both measuring systems show that substantial warming occurred in the low–middle troposphere, TLT and RLT, while the lower stratosphere, TLS and RLS, experienced substantial cooling. As indicated above, it is always difficult to interpret variations of the ''midtroposphere,'' TMT and RMT, as the weighting function spans portions of the troposphere and lower stratosphere. As expected, however, the trends for TMT and RMT are intermediate between TLT and TLS.

TABLE 3. Basic statistics utilized in comparisons between the radiosonde at Minqin, China, and microwave temperature products (satellite). The comparison period is 1979–2000. Here LT, MT, and LS refer to low–middle troposphere, midtroposphere, and lower stratosphere, respectively.

	LT	MT	LS
r_{MON}	0.90	0.95	0.95
r_{ANN}	0.97	0.97	0.96
b, Minqin (°C decade^{-1})	+0.48	+0.30	−0.78
b, satellite (°C decade^{-1})	+0.47	+0.29	−0.75
σ_{MON} Minqin, satellite (°C)	1.20, 1.19	0.81, 0.79	1.20, 1.25
$\sigma_{\Delta\ \mathrm{MON}}$ (°C)	0.53	0.25	0.38
N_{eff} of monthly δ's (N = 264)	212	174	99
σ_{ANN} Minqin, satellite (°C)	0.52, 0.56	0.34, 0.35	0.69, 0.72
$\sigma_{\Delta\ \mathrm{ANN}}$ (°C)	0.14	0.09	0.19
N_{eff} of annual Δ's (N = 22)	10	17	15
$\sigma_{b\ \mathrm{ERR}}/\sqrt{25}$ (global)	0.0140	0.0069	0.0155

The satellite data are processed without regard for any specific location, so that Minqin is truly a random selection in space. Given that satellite and radiosonde (or simply sonde) provide two independently measured time series for each layer, we now make a ''thought experiment'' to deduce more general results. CSB conservatively (i.e., as a minimum) estimated that the global atmosphere represented by the temperatures of the satellite products is characterized by about 26 spatial degrees of freedom (DOFs). Thus, assuming that Minqin represents 1 DOF, the variations of global average deep layer temperatures have 26 times the DOF of the Minqin sample. We may therefore infer standard errors for global averages assuming a normal distribution of random errors. From the standard errors of the annual Minqin anomalies we also estimate standard errors in the global trend (last entries in Table 3) as described below.

Let us further assume that the temperatures from the Minqin radiosonde are perfect and that all error resides with the satellite data. Upscaling the results of Table 3 by $\sqrt{(26\text{-}1)}$ to global error estimates, we would estimate that the 95% confidence interval (CI) for global mean, monthly values of TLT, TMT, and TLS are ±0.24°,

TABLE 4. As in Table 3 but for the composite of 10 stations between 0° and 30°N using VIZ radiosondes.

10 stations 0°–30°N	LT	MT	LS
r_{MON}	0.91	0.93	0.96
r_{ANN}	0.97	0.93	0.96
b, sondes (°C decade^{-1})	+0.07	−0.01	−0.65
b, satellite (°C decade^{-1})	+0.05	+0.04	−0.42
σ_{MON} sondes, satellite (°C)	0.35, 0.35	0.34, 0.33	0.76, 0.71
$\sigma_{\Delta\ MON}$ (°C)	0.15	0.12	0.21
σ_{ANN} sondes, satellite (°C)	0.23, 0.21	0.26, 0.24	0.59, 0.49
$\sigma_{\Delta\ ANN}$ (°C)	0.06	0.10	0.19

±0.12°, and ±0.25°C, respectively, assuming an additional adjustment to the magnitude of σ_Δ by $\sqrt{(N/N_{eff})}$, where N_{eff} equals the effective number of independent samples and is derived from the time series of the Δ's.[10] For annual global anomalies, the 95% CI for TLT, TMT, and TLS becomes ±0.08°, ±0.04°, and ±0.09°C, respectively.

Because the trends between satellite and radiosonde are essentially identical in Table 3, our estimate of the satellite trend error will be based on $\sigma_{\Delta ANN}$. We use the formulation for standard error of the slope (b) as follows:

$$\sigma^2_{b\,ERR} = N(\sigma^2_{\Delta ANN})(N/N_{eff}) \Big/ \left[N\left(\sum x^2\right) - \left(\sum x\right)^2\right].$$

Though the time series are 22 years in length (N), the values of $\sigma_{b ERR}$ and the confidence intervals require N_{eff} of the Δ's. For global estimates, the assumption again is 26 spatial DOFs. Thus random errors are reduced as the virtual sample size increases. For the present case where $N = 22$ and (DOF − 1) = 25 we have (units of °C^2 decade^{-2}):

$$\sigma^2_{b\,ERR} = (\sigma^2_{\Delta ANN})(N/N_{eff})(0.004\ 517).$$

The resulting 95% CI for global trends becomes ±0.031° ($N_{eff} = 10$), ±0.015° ($N_{eff} = 17$), and ±0.033°C decade^{-1} ($N_{eff} = 15$) for TLT, TMT, and TLS, respectively. We will place these remarkably small values in context once other independent evaluations (listed in Tables 9, 10) are completed. This will help us determine whether Minqin, though independent in its selection, presented a fortuitously optimistic result.

[10] The estimate of N_{eff} or the effective number of independent samples is $N_{eff} = N(1 - r_1)/(1 + r_1)$, where r_1 = autocorrelation at lag one of N, Δ's.

b. Multiple, mostly consistent radiosonde stations

In the United States and areas under its authority, 28 radiosonde stations (of the 100+ total) have utilized VIZ instrumentation since 1979 (appendix) and are therefore potentially excellent choices for another comparison study. As we shall see, some changes have apparently impacted the homogeneity of the observations, preventing us from assuming that all error lies with the satellite data when computing error estimates.

Tables 4–6 display the same statistics as calculated for Minqin for three multistation groupings. We divided the 28 stations into those south of 30°N (10 stations; Table 4) and those north of 30°N (18 stations; Table 5.) This was done so that a few high-latitude stations with high variance would not dominate the results. For the 28-station composite (Table 6) we simply averaged the composites from the 10 low-latitude and the 18 high-latitude averages.

Though there were some unique problems with some stations, overall there is good agreement, except that the satellite data produce more positive trends, especially in the lower stratosphere. Since TLS is entirely independent of TLT and TMT, we may hypothesize that inhomogeneities, increasing with altitude, caused these radiosonde time series to produce trends more negative than observed by the satellites.

Upon further examination, we find three possible causes for relative negative trends due to radiosonde issues. The first problem, noted in anecdotal accounts of radiosonde operators and reported in Oort and Liu

TABLE 5. As in Table 3 but for the composite of 18 stations between 30° and 90°N using VIZ radiosondes.

18 stations 30°–90°N	LT	MT	LS
r_{MON}	0.95	0.95	0.98
r_{ANN}	0.93	0.87	0.98
b, sondes (°C decade^{-1})	+0.15	+0.01	−0.80
b, satellite (°C decade^{-1})	+0.30	+0.14	−0.64
σ_{MON} sondes, satellite (°C)	0.71, 0.69	0.50, 0.47	1.09, 1.13
$\sigma_{\Delta\ MON}$ (°C)	0.22	0.15	0.22
σ_{ANN} sondes, satellite (°C)	0.31, 0.34	0.21, 0.20	0.70, 0.62
$\sigma_{\Delta\ ANN}$ (°C)	0.12	0.11	0.14

TABLE 6. As in Table 3 but using the composite of the 28 VIZ stations (sondes) utilized in Tables 4 and 5.

28 stations 0°–90°N	LT	MT	LS
r_{MON}	0.94	0.93	0.97
r_{ANN}	0.95	0.90	0.98
b, sondes (°C decade^{-1})	+0.11	+0.00	−0.73
b, satellite (°C decade^{-1})	+0.17	+0.09	−0.53
σ_{MON} sondes, satellite (°C)	0.41, 0.39	0.33, 0.30	0.81, 0.77
$\sigma_{\Delta\ MON}$ (°C)	0.15	0.12	0.19
σ_{ANN} sondes, satellite (°C)	0.24, 0.24	0.22, 0.19	0.63, 0.53
$\sigma_{\Delta\ ANN}$ (°C)	0.07	0.10	0.16

(1993), is due to the changing quality of radiosonde balloons over the years. On the coldest days, the earlier balloons tended to burst at lower altitudes so that cold readings were preferentially missing and monthly average temperatures were spuriously warm. Over time, there appears to have been an improvement so that more soundings on the colder days are successfully completed. This characteristic would appear as a relative negative trend in the radiosonde minus satellite record, with increasing magnitude with height.

We tested this notion by computing the difference trend between RLS and TLS for winter (December–February, hereafter DJF) and summer (June–August, hereafter JJA) separately for the 10 highest-latitude stations (poleward of 45°N.) The result is consistent with the hypothesis in that the DJF difference trend (RLS minus TLS) was −0.35°C decade^{-1} and for JJA, −0.12°C decade^{-1}. Similarly, for high-latitude stations, the DJF TMT difference trend was also more negative (−0.16 vs −0.12°C decade^{-1}).

We then looked at the number of observations reported at the 50-hPa pressure level at Barrow, Alaska. Over the past 22 years in the six months from October to March, the trend in the number of 0000 UTC observations per month is positive, indicating an increase over this period of 5.9 more daily observations per month in 2000 than in 1979 (about 24 to 30). This represents a highly significant increase in the number of observations taken and is consistent with the trends observed and the anecdotal information of early balloon failure in the coldest atmospheres.

A second VIZ issue was discovered in the comparison of TMT and RMT seen clearly in the 10 0°–30°N stations (Fig. 5), that being a statistically significant break between May and June 1997. We later learned that for most VIZ stations, the instrumentation was changed from VIZ-B to VIZ-B2 on 1 June 1997. The new model carried an improved pressure sensor that apparently is the source of the discontinuity discovered by the comparison shown in Fig. 5. Of the 28 stations used here, 21 switched to VIZ-B2 on 1 June 1997 and the remaining seven ''when supplies of VIZ-B are depleted'' (see http://www.ua.nws.noaa.gov/Equip.htm). We tested Annette Island and Hilo, two stations in the latter category, but could only determine that their break occurred anytime from June 1997 to mid-1998. Single-

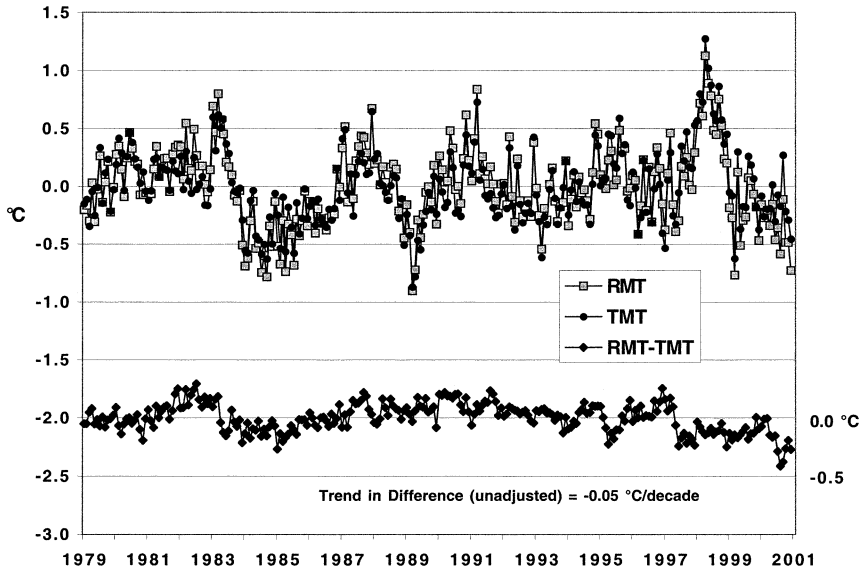

FIG. 5. Comparison of RMT and TMT for the 10 U.S. controlled radiosonde stations between 0° and 30°N. Note the difference time series showing relative shifts in early 1983 and mid-1997. Monthly differences are shown offset 2°C in the lower time series for clarity.

TABLE 7. As in Table 6 but after adjustment for VIZ-B to VIZ-B2 transition to RLT and RMT.

28 stations 0°–90°N	LT adj.	MT adj.	LS
r_{MON}	0.95	0.95	0.97
r_{ANN}	0.98	0.96	0.98
b, sondes (°C decade^{-1})	+0.16	+0.05	−0.73
b, satellite (°C decade^{-1})	+0.17	+0.09	−0.53
rms ($b_{sondes} - b_{sat}$) (°C decade^{-1})	0.105	0.095	0.230
σ_{MON} sondes, satellite (°C)	0.42, 0.39	0.34, 0.30	0.81, 0.77
$\sigma_{\Delta\ MON}$ (°C)	0.15	0.10	0.19
N_{eff} of mon Δ's ($N = 264$)	115	90	64
σ_{ANN} sondes, satellite (°C)	0.26, 0.24	0.22, 0.19	0.63, 0.53
$\sigma_{\Delta\ ANN}$ (°C)	0.051	0.068	0.16
N_{eff} of annual Δ's ($N = 22$)	21	8	6
$\sigma_{b\ ERR}/\sqrt{5}$ (global)	0.0078	0.0169	0.0460

station results are less robust due to background noise in the difference time series, so we assumed that all stations changed from VIZ-B to VIZ-B2 on 1 June 1997.

Apparently, the VIZ-B to VIZ-B2 change is detectable in the troposphere but not the stratosphere. This is likely because a pressure (P) bias translates into a temperature (T) bias at a given pressure in the troposphere where there is a strong T–P relationship, but not in the nearly isothermal stratosphere (Parker and Cox 1995). By examining the 36 months on either side of the May–June 1997 break, we calculated a spurious RMT cooling of 0.13°C for the 28-station composite. Similarly, for RLT, the break identified a shift of 0.16°C. Table 6 has been recalculated for RLT and RMT, using the adjustments applied on 1 June 1997, and the new statistics are presented in Table 7. The adjustment on 1 June 1997 has increased the radiosonde trends, which are now insignificantly different from the satellite trends. Though the trend is dependent on the adjustment applied, we have no other means to determine the obvious temperature impact due to the instrument change. What is important here, from the satellite perspective, is that the discontinuity was discovered prior to any knowledge of the sonde instrument change.

A third potential problem is seen in the RMT and RLS differences prior to early 1983 and evident in the 10 stations south of 30°N for RMT and all 28 stations in RLS (Figs. 5 and 6). The differences portray RMT and RLS as relatively warmer than TMT and TLS with a peak in the difference in 1982. This has proven to be a puzzling result. This is a period observed by two independent spacecraft, *NOAA-6* and *NOAA-7*, which have very minor corrections applied, that is, for TLS much less than 0.05°C (CSB). We also note that MSU2 and MSU4 on these spacecraft are essentially independent instruments and would not be expected to produce nearly identically timed errors in comparison with the radiosondes.

The differences are greatest in 1982 and their values for the three products are 0.06°, 0.17°, and 0.43°C for (RLT minus TLT), (RMT minus TMT), and (RLS minus TLS), respectively. This evidence suggests that the radiosondes are the main source of the error, but see below for an explanation due to satellite merging. A possible source of heterogeneity is that beginning in 1982 the U.S. National Weather Service began the transition from on-site, baseline-box calibration of sondes with mercury-in-glass thermometers to factory calibration. It is possible that this transition led to a net change in the absolute calibration of the U.S. sondes. Unfortunately, the transition varied from station to station and was not documented (Spencer and Christy 1993). Because we are unable to explain these early differences, we shall assume that they represent a problem with the satellite data in the calculations of monthly and annual statistical error below.

We investigated three possibilities that might cause satellite TLS to contain a spuriously positive trend shown in these results. First, we looked at the magnitude of the adjustments applied to the raw satellite data. The magnitudes in total are very minor, on the order of a few hundredths of a degree Celsius both positive and negative over the span of a few years. The individual satellite records indicate a high level of agreement during overlapping observations and so required very little adjustment. By contrast, the DJF RLS minus TLS trend difference alone accumulates to a total of 0.77°C over the 22-yr period. We could not propose a mechanism that would cause a significantly larger satellite trend error in winter versus summer related to the satellite adjustments or observations.

Second, it is possible that the static weighting function applied to the radiosonde data might be misrepresentative of the true profile, perhaps being more elevated than in reality. This would produce a more negative sonde-simulated trend versus the satellite. This possibility was examined in Spencer and Christy (1993), who varied weighting functions but found the best correlation with a profile only slightly lower in the very highest altitudes.

Third, we note that in Figs. 5 and 6 there is a shift or drift in the difference time series in early 1983. In Fig. 6, a similar, though less noticeable, shift occurs between February and March 1991. Both of these events occur close to transitions in spacecraft, *NOAA-6* to *NOAA-8* in April 1983 and *NOAA-10* to *NOAA-12* in

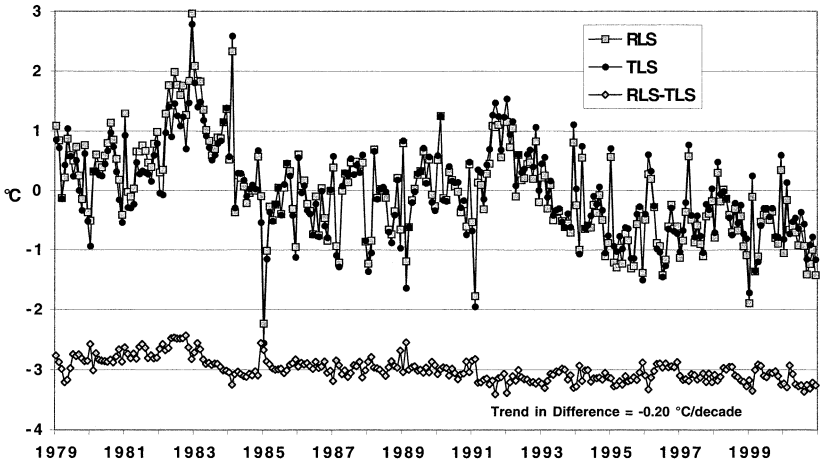

FIG. 6. Comparison of RLS and TLS for 28 U.S. controlled radiosonde stations. Note the difference time series showing the relative warmth of the radiosondes around 1982. Monthly differences are shown offset 3°C in the lower time series for clarity.

April 1991. In both cases, however, another spacecraft was on-station (*NOAA-7* and *NOAA-11*, respectively), providing data across these boundaries. However, the merging of the data from a new satellite may have perturbed the regional fields around these grids creating local heterogeneities, though the same feature is not found in other comparisons (e.g., Minqin above or global analyses below). The discrepancies around 1981–83 are particularly curious since *NOAA-6* and *NOAA-7* reveal exceptional agreement in this period with virtually no adjustments required to the raw data for either TMT or TLS. Nevertheless, because we assume that the difference time series is due to satellite issues, these discrepancies will increase the calculated satellite error.

We shall calculate global error estimates based on error statistics of the composite time series assuming the entire globe contains 6 times the DOF represented by the area monitored by the 28 radiosondes, which stretch from the western tropical Pacific, to Barrow (Alaska), to San Juan (Puerto Rico). Again, we shall assume all error is contained within the satellite data.

In general, the results in Tables 4–7 show that the low-latitude atmosphere is characterized by less variability and less error than in high latitudes, especially on shorter timescales. The 95% CI ranges for the global (i.e., generalized) trend of each product, TLT ($N_{\text{eff}} = 21$), TMT ($N_{\text{eff}} = 8$), and TLS ($N_{\text{eff}} = 6$) are $\pm 0.016°$, $\pm 0.039°$, and ± 0.113°C decade^{-1} (see bottom row of Table 7 and summary in Table 10). A trend in the difference time series is likely associated with an autocorrelation of Δ's, reducing N_{eff} and increasing the CI. Note that the absolute trend difference between RLS and TLS is 0.20°C decade^{-1} ($-0.73°$ vs -0.53°C decade^{-1}; Table 7), while the generalized result (Table 10) suggests an error range of ± 0.11°C decade^{-1}. This result, where the generalized error range does not capture the trend difference of the sample, occurs only here. There is a possibility that a systematic trend error is contained within the satellite data that affects all grids and thus is incorrectly reduced by the generalization method. However, because we have presented evidence that a systematic negative trend exists in the RLS VIZ data, and no other comparison produces this potentially inconsistent result, we shall remain with the error range calculated above (± 0.11°C decade^{-1}).

c. Global means of upper air temperatures from radiosondes and reanalyses

Our final comparisons will be performed using four datasets of tropospheric and stratospheric global anomalies, described in the Intergovernmental Panel on Climate Change (IPCC) 2001 report (Folland et al. 2001; though the NCEP reanalyses have since been upgraded; see below.) Our comparisons will utilize only annual anomalies of RLT and RLS, but will include an additional year (1979–2001).

Numerous changes have occurred since 1979 in virtually all of the stations that supply data to these compilations. An example is shown in Fig. 7 for the U.S. station at Chuuk, Federated States of Micronesia, where the VIZ-B instrument package was replaced with Vaisala RS80 on 1 Dec 1995 revealing an obvious cooling in RLS versus the satellite TLS. Indeed, many stations replaced their instrumentation with Vaisala equipment during this period with similar results. Such changes are more noticeable in the stratosphere where the effects of biases owing to faulty calibration, solar heating, and longwave radiation (and changing adjustments to attempt to compensate for these) tend to accumulate (Parker et al. 1997). In addition to the balloon bursting problem noted earlier, radiosonde trends in the stratosphere of these datasets are likely more negative than in reality given the general move toward Vaisala equipment.

As before we shall calculate the 95% CI of the trends

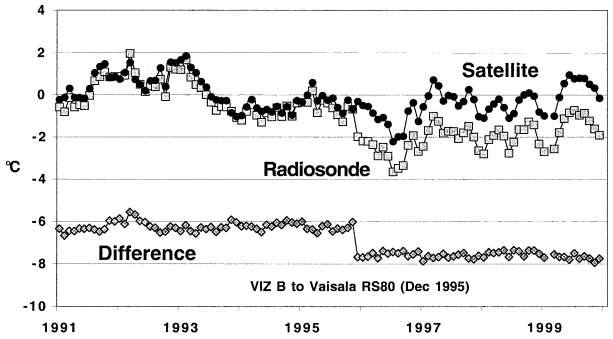

FIG. 7. Comparison of RLS and TLS for Chuuk, a west Pacific island. The change from VIZ-B to Vaisala RS80 in December 1995 gave significant, spurious cooling (differences offset by 6°C for clarity).

based on the magnitude of $\sigma_{\Delta ANN}$, which for global datasets (with no reduction of noise due to geographic upscaling) becomes

$$\sigma^2_{bERR} = (\sigma^2_{\Delta ANN})(N/N_{eff})(0.09881).$$

1) HADLEY CENTRE, MET OFFICE

The Hadley Centre Radiosonde Temperature (HadRT) products designed for comparison with microwave instrumentation were introduced by Parker et al. (1997). HadRT data rely on monthly average temperature reports from CLIMAT TEMP data (upper air climatological data) supplied by balloon station operators. A weighting function profile was developed and applied to the mandatory pressure level data to simulate temperature anomalies of TLT and TLS. Approximately 400 stations provide input for these averages, though large regions of the globe are not sampled. Three major comparison studies have been performed (Parker et al. 1997; Santer et al. 1999; Hurrell et al. 2000) in addition to the IPCC 2001 compilations (Folland et al. 2001). RLT values from HadRT2.0 have no adjustments applied to them. However, we use HadRT2.1s for RLS values in which significant stratospheric temperature heterogeneities worldwide, where supported by known instrumental or operational changes, have been eliminated by TLS comparisons (Parker et al. 1997).

Of the three radiosonde-only global analyses, we anticipate the best satellite agreement from HadRT products due to the application of the appropriate weighting function to mimic the satellite profile. The following two datasets [NOAA–Angell and the Russian Research Institute of Hydrometeorological Information (RIHMI)] are simple averages of layer temperatures and will be included only for illustrative purposes and will not be included in the cumulative statistics due to the mismatch between layer averages and satellite profiles.

2) NOAA–ANGELL

NOAA–Angell is described in Angell (1988) with an update in Angell (2000) in which some comparisons with our satellite data were made. The NOAA–Angell dataset is built from 63 stations with long and fairly continuous records, though it contains some significant heterogeneities (Trenberth and Olson 1991; Gaffen et al. 2000.) Global anomalies are determined from a geographically weighted sum of anomalies in seven broad latitude bands. The quantity measured is the thickness temperature determined by the physical distance (i.e., volume) between two pressure levels. For comparison purposes we use the 850–300-hPa layer for RLT and the 100–50-hPa layer for RLS. In addition to the mismatch in layer averages versus satellite profile, the implicit inclusion of the influence of moisture in layer thickness will also lead to differences (CSB).

3) RIHMI

The Russian Research Institute for Hydrometeorological Information maintains a set of global analyses of pressure level temperature based on over 800 radiosonde stations from the CARDS and telecommunicated datasets (Sterin 1999). The data are objectively interpolated to all unobserved regions where relaxation to climatology (i.e., zero anomaly) is assumed. Due to the relatively stiff interpolation procedures, Folland et al. (2001) found the variance of RIHMI much reduced. As with all radiosonde-based datasets, RIHMI suffers from heterogeneities in the individual station records. The temperature of the 850–300-hPa layer is calculated using all mandatory level information (i.e., 850, 700, 500, 400, and 300 hPa) and so is more consistent with a mass-weighted value than NOAA–Angell's 850–300-hPa thickness temperature. We shall use RIHMI's 850–300- and 100–50-hPa layer temperatures for comparison with TLT and TLS.

4) NCEP REANALYSES

The U.S. National Centers for Environmental Prediction (NCEP) have generated an historical set of global weather analyses using a single global circulation model with observations from in situ and satellite sources (Kalnay et al. 1996). Pressure level temperatures from these analyses have been vertically averaged, proportionally, to create RLT and RLS as was done with HadRT data (Stendel et al. 2000). Some concerns about continuity have been discussed in Basist and Chelliah (1997), Pielke et al. (1998), and Stendel et al. (2000) who also compared NCEP and satellite data, finding generally good agreement. The data utilized here include a replacement of the post-January 1997 data, which alleviated a problem with land surface temperatures (see http://wesley.wwb.noaa.gov/reanalysis.html).

The reanalyses do not use the microwave datasets as

TABLE 8. Statistics computed from global upper air datasets 1979–2001.

Global analyses	LT	LS
	r_{ANN}	r_{ANN}
HadRT, NOAA, RIHMI	0.94, 0.89, 0.92	0.99, 0.94, 0.96
NCEP	0.96	0.96
	b, °C decade^{-1}	b, °C decade^{-1}
HadRT, NOAA, RIHMI	+0.04, −0.02, +0.00	−0.64, −1.20, −0.46
NCEP	+0.05	−0.82
Satellite	+0.06	−0.52
	σ_{ANN} °C	σ_{ANN} °C
HadRT, NOAA, RIHMI	0.15, 0.21, 0.11	0.57, 0.89, 0.35
NCEP	0.17	0.67
Satellite	0.16	0.47
	$\sigma_{\Delta\ ANN}$ °C	$\sigma_{\Delta\ ANN}$ °C
HadRT, NOAA, RIHMI	0.05, 0.10, 0.07	0.13, 0.47, 0.17
NCEP	0.05	0.25
	N_{eff} of Δ's (N = 23)	N_{eff} of Δ's (N = 23)
HadRT, NCEP	6, 7	11, 1
	$\sigma_{b\ ERR}$	σ_{bERR}
HadRT, NCEP	0.0308, 0.0285	0.0578, ?

adjusted and merged by UAH but do incorporate the NOAA satellite sounding profiles of which the MSU and AMSU radiances are a part. Thus, in simple terms, the global weather analyses are based on the spatial pattern of temperatures provided mostly by satellite information, which is simultaneously anchored to radiosonde soundings and made consistent with the model dynamics. Because the sounding coefficients used to create the satellite profiles are updated every week by direct comparison with radiosondes, the long-term trends in NCEP are dependent on radiosondes. The NCEP reanalyses are quite different from the satellite and radiosonde datasets discussed above, but in the strict sense they are not independent from the sondes.

d. Results

The comparison results for these global analyses are shown in Table 8. The RLT and TLT comparisons indicate very good agreement with the largest differences found between NOAA–Angell RLT and TLT as expected. Excellent correlations are shown for RLS and TLS, but the differences in trends are quite large. We would expect the trends of NOAA–Angell RLS and RIHMI RLS to be more negative because the 100–50-hPa layer avoids the tropical upper troposphere and its positive trend. However, RIHMI's low variance seems to have dominated the otherwise anticipated trend difference.

The 95% CI of the TLT global trend becomes ±0.075° and ±0.067°C decade^{-1} for HadRT2.0 and NCEP, respectively (Fig. 8, Table 10). The wider range of confidence intervals than determined earlier is due to fewer DOF in the Δ time series (Table 8). For HadRT2.1s RLS, N_{eff} is 11, but because the difference trend of the NCEP RLS versus TLS is pronounced and systematic, its value of N_{eff} is 1. The NCEP RLS minus TLS difference time series is essentially a line with zero slope to 1992, then

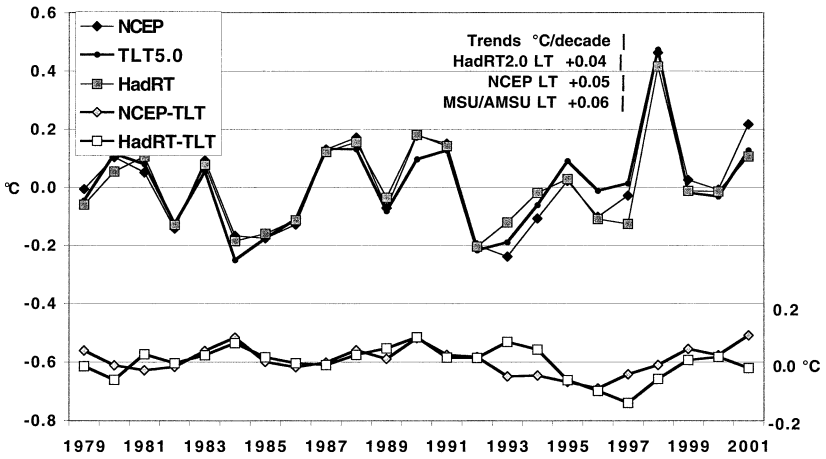

FIG. 8. Time series of annual global mean lower-tropospheric temperature anomalies from satellite, HadRT2.1s, and NCEP reanalyses. Differences (HadRT minus satellite and NCEP minus satellite) are offset 0.6°C in the lower time series.

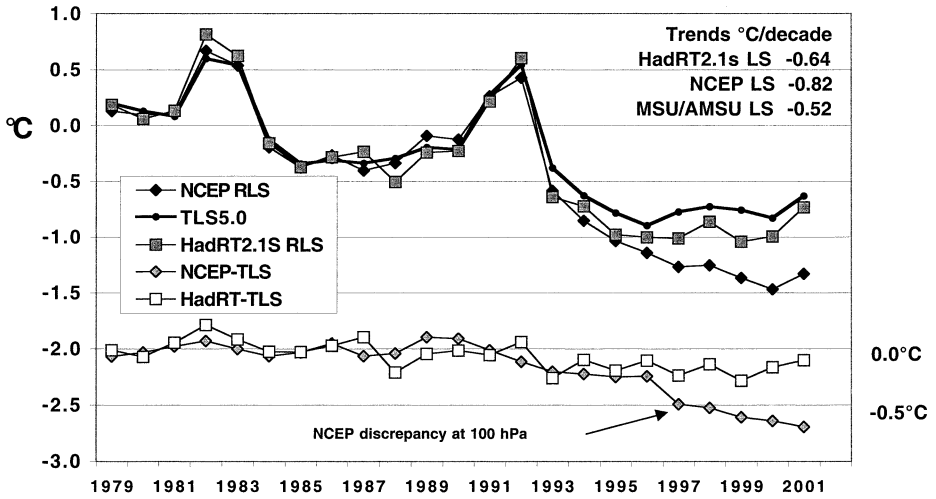

FIG. 9. As in Fig. 8 but for the lower stratosphere: HadRT2.1s is used instead of HadRT2.0. Differences (HadRT minus satellite and NCEP minus satellite) are offset 2°C in the lower time series. Note the divergence between the satellite data and NCEP in 1992 and 1997 in the NCEP minus satellite time series.

a negative slope from 1992 onward, hence only 1 DOF. Consequently, the 95% CI varies drastically for these two estimates, being ±0.13° (HadRT2.1s) and ±4.79°C decade^{-1} (NCEP). The trends for 1979–91 from TLS and NCEP RLS are −0.03° and −0.04°C decade^{-1}, respectively, being virtually identical; but then the systematic and pronounced divergence begins (Fig. 9.) We have since discovered that a significant discrepancy in the post-1996 100-hPa temperatures between the two main versions of the NCEP reanalyses has appeared, so values for the appropriate NCEP parameters in Table 8 (and Table 10) are replaced with ''?'' (see http://wesley.wwb.noaa.gov/tovs_problem/index.html).

Wide-ranging changes in instrumentation and methods have their greatest impact on stratospheric temperatures. Consequently, our error estimates for TLS will be large and likely will not represent the true precision of the satellite data. This has proven especially true for the NCEP reanalyses where many factors affect the determination of temperature, including unadjusted radiosonde data (e.g., Chuuk), top-of-atmosphere radiation balance constraints, model assimilation techniques, intersatellite bias adjustments (or lack thereof), etc.

4. Combined results

The simple error statistics for each of the test samples are compiled in Table 9 for grid points and Table 10 for global statistics (where evaluation is possible). The monthly gridpoint errors appear reasonable as the errors determined independently by Minqin (38.6°N) are intermediate between those of the low latitudes and high latitudes of the VIZ network. In fact, the composited monthly values of all 28 VIZ stations agree remarkably well with Minqin. (Gridpoint results from the multistation VIZ network are derived from the rms of each individual station's $\sigma_{\Delta MON}$ and $\sigma_{\Delta ANN}$.)

A different result emerges for the annual anomalies where the Minqin results are consistently smaller than those calculated from the VIZ network. The magnitude

TABLE 9. Summary of 95% confidence intervals for gridpoint (2.5°) statistics. The values in parentheses assume that the differences between radiosondes and satellite data are due equally to errors with each system rather than due to satellites alone.

	TLT (°C)	TMT (°C)	TLS (°C)
95% CI monthly gridpoint anomaly			
Minqin	±1.01	±0.48	±0.72
U.S. VIZ radiosondes low lat	±0.67	±0.41	±0.71
U.S. VIZ radiosondes high lat	±1.28	±0.69	±0.90
U.S. VIZ radiosondes (all)	±1.02	±0.57	±0.81
95% CI annual gridpoint anomaly			
Minqin	±0.27	±0.17	±0.36
U.S. VIZ radiosondes low lat	±0.27 (±0.19)	±0.27 (±0.19)	±0.53 (±0.37)
U.S. VIZ radiosondes high lat	±0.51 (±0.36)	±0.34 (±0.24)	±0.52 (±0.37)
U.S. VIZ radiosondes (all)	±0.41 (±0.29)	±0.31 (±0.22)	±0.52 (±0.37)

TABLE 10. Summary of 95% confidence interval estimates from calculations for global temperature statistics based on datasets defined.

	TLT °C decade^{-1}	TMT °C decade^{-1}	TLS °C decade^{-1}
Global 95% CI trend			
Minqin	±0.031	±0.015	±0.033
U.S. VIZ sondes, composite $\sigma_{\Delta\ ANN}$	±0.016	±0.039	±0.113
U.S. VIZ sondes, rms Δb	±0.043	±0.039	
HadRT	±0.075		±0.127
NCEP	±0.067		?
Rms consensus	±0.051	±0.033	±0.100
95% CI monthly global anomalies	°C	°C	°C
Minqin	±0.24	±0.12	±0.25
U.S. VIZ radiosondes	±0.20	±0.15	±0.35
NCEP	±0.16		±0.41
Rms consensus	±0.20	±0.14	±0.35
95% CI annual global anomalies	°C	°C	°C
Minqin	±0.08	±0.04	±0.09
U.S. VIZ radiosondes	±0.04	±0.10	±0.27
HadRT, NCEP	±0.20, ±0.18		±0.38, ?
Rms consensus	±0.14	±0.08	±0.27

of annual anomalies is more easily impacted by systematic errors of the type suggested earlier for the U.S. VIZ sondes, and this is likely to be what is causing the increased error estimates. The values shown in Table 9 essentially assume that all of the difference between radiosonde and satellite is due to satellite error. It is reasonable to conclude that at least half of the difference is due to VIZ radiosonde inaccuracies, and we have demonstrated the existence of such without mentioning errors due to interpolation of data or instrument noise. Thus, a more accurate result is likely obtained by reducing the error ranges shown for U.S. VIZ sondes by a factor of 0.707 ($1/\sqrt{2}$). These are shown in parentheses in Table 9 for the annual values and now, for the 28-station composite, agree remarkably well with those of Minqin.

Table 10 lists the estimates of precision for global quantities from the four (mostly) independent and certainly very dissimilar datasets. Recall that the four realizations arise from

1) a single radiosonde with consistent instrumentation (Minqin),
2) 28 U.S. radiosondes with mostly consistent instrumentation,
3) 400 global radiosondes (HadRT) with less consistency, and
4) NCEP reanalyses using satellite and radiosonde data in a data assimilation project.

The consensus error statistics are simply the rms values of the individual results. We assume that each of the estimates is of equal value in the consensus calculations. Each has its strengths and weaknesses. In terms of instrument consistency, the preference is essentially the order given in the above list. In terms of the ability to generalize to the global average, the order would be reversed. Thus weighing both consistency and geographic coverage together, we assume that the datasets are equivalent in terms of their ability to provide useful information on satellite errors.

For a randomly selected monthly gridpoint anomaly, the results suggest the 95% CIs are ±1.0°, ±0.6°, and ±0.8°C for TLT, TMT, and TLS, respectively (Table 9). Annual gridpoint CI values for the same three layers are likely about ±0.3°, ±0.2°, and ±0.4°C, respectively. Global mean monthly anomalies for TLT, TMT, and TLS likely are known to within ±0.20°, ±0.15°, and ±0.35°C. Likewise, for annual, global mean anomalies we estimate ±0.15°, ±0.10°, and ±0.30°C for TLT, TMT, and TLS, respectively, though it is conceivable that these are overestimates given the problems, especially in the stratosphere, with radiosonde data.

For the trend of TLT and TMT, the 95% CI is likely about ±0.05°C decade^{-1}. CSB (version D) estimated this as ±0.06 based on shorter time periods and fewer comparisons. However, our present result assumes that all of the difference between the satellite and the various datasets is due entirely to satellite error, a very conservative assumption. Further arguments could be made regarding radiosonde-based errors that would reduce the portion of error attributed to TLT and TMT, and thus reduce the magnitude of their error statistics, but we shall not make those arguments at this time. In any case, as of this writing, we note that the global TLT (version 5.0) trend is +0.06°C decade^{-1} ±0.05.[11]

As with CSB, we are less confident in estimating trend errors for TLS. There simply are no significant groups of radiosonde stations (on which much of the NCEP

[11] A separate error estimate for TLT and TMT is based on the magnitude of the rms of the individual U.S. slope differences (Table 7, and given in Table 10 as rms Δb). For example, assuming 26 spatial DOF of the globe and a TLT slope error of 0.105°C decade^{-1} for an individual grid (or spatial DOF), the 95% CI for the error trend becomes ±0.043°C decade^{-1}. TLS is not included because the mean composite slope difference was significantly different from zero.

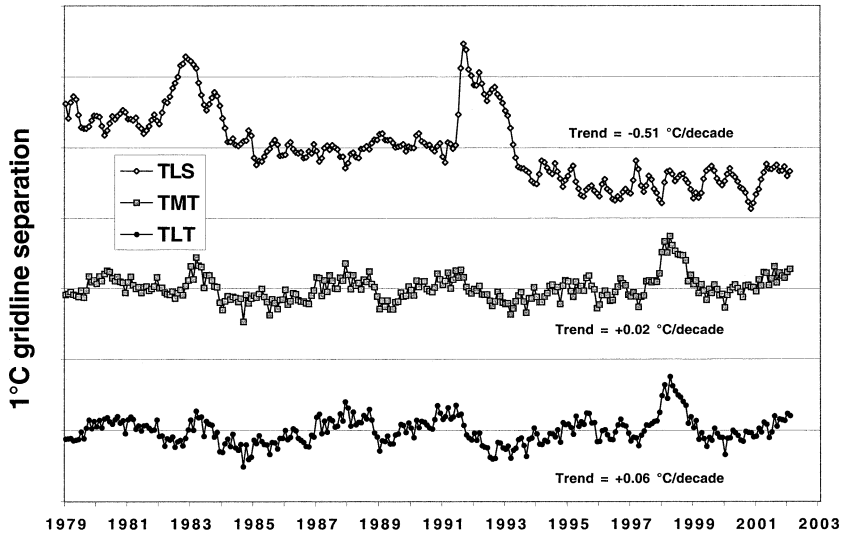

FIG. 10. Monthly temperature anomaly time series of TLS, TMT, and TLT.

reanalyses are also based) that have good records with long-term stability. The satellite adjustments required for TLS are smaller than those of TLT, so one would expect the performance of TLS to contain less error, yet the radiosonde comparisons (except the generalized Minqin results) imply greater error. The 95% CI for the precision of the TLS global trend is likely less than the value listed as consensus (± 0.10°C decade^{-1}).

5. Conclusions

Advances in technology have lead to the need for merging data from older instruments (MSU) with newer instruments (AMSU) to continue the time series of bulk atmospheric temperatures derived from microwave radiometers. Though the AMSU channel frequencies do not exactly match the MSU frequencies in the case of MSU2, we have demonstrated that approximations generated from AMSU data do indeed reproduce the type of information formerly provided from the MSUs. We have found from comparisons of MSU versus AMSU and AMSU versus AMSU that in all cases the new data met or improved upon the error characteristics of MSU versus MSU. The homogeneity of these data is therefore retained to the level of error characteristic of the MSU-only time series.

We have upgraded the diurnal drift correction to the tropospheric datasets, TLT and TMT, now applying a slightly nonlinear adjustment. Our method, which uses the cross-swath satellite data for the diurnal adjustment, has the advantage of eliminating intersatellite biases and interannual variability of the temperature field. The monthly time series of TLT, TMT, and TLS anomalies are shown in Fig. 10.

To assess the measurement error of the satellite datasets we have performed direct comparisons with radiosonde and radiosonde-based data. These are independent comparisons. Unfortunately, no instrumentation provides a ''perfect'' measurement of the bulk atmospheric temperature, especially over a 23-yr period, so there is difficulty in determining the error of one system by comparison with another, which itself is subject to error. Our use of four ''independent'' groups of data (Minqin, 28 U.S. controlled radiosondes, about 400 radiosondes of HadRT, and NCEP reanalyses), attempted to provide multiple estimates of the errors sought.

Low–middle tropospheric comparisons suggest that globally the satellite system provides monthly and annual anomalies within a 95% CI range of error of approximately $\pm 0.20°$ and ± 0.15°C, respectively (Table 10), though conceivably these ranges could be smaller depending on the magnitude of error that may be assigned to radiosondes. Stratospheric measurements from radiosondes are fraught with time-dependent biases and the estimates for global monthly (annual) satellite errors are likely to be smaller than the consensus values of ± 0.35°C (± 0.27°C) derived in Table 10.

The evidence presented herein suggests that the global, low–middle tropospheric and midtropospheric trends are known to within about ± 0.05°C decade^{-1} and possibly better. An argument could be made that each of the comparison tropospheric datasets used in this study has systematic, long-term error that by chance happens to agree with the UAH satellite products. This is highly unlikely as we have deliberately performed the analyses with multiple networks, most being completely independent of the others. This is not the case in the stratosphere, however, where systematic changes, which have very little impact on the tropospheric portion of the sounding, have considerable impact on the stratospheric readings (Parker et al. 1997). With this as cautionary note, we estimate the global TLS trend as -0.51°C decade^{-1} with a 95% CI of ± 0.10°C decade^{-1} (Table 10).

In this paper, we have at least identified probable

bounds of the satellite error statistics. With additional work on applying radiosonde corrections independent from microwave data, it may be possible to better characterize these errors in the future (Free et al. 2002; Lanzante et al. 2003). More generally, this study emphasizes the need for an observing system that meets the requirements of climate science regarding upper air observations, especially for a vertical resolution that is finer than provided by satellites alone and particularly for altitudes above the troposphere (NRC 1999; 2000a,b).

Acknowledgments. Christy and Norris were supported by the U.S. Department of Transportation (DTFH61-99-X-00040). Spencer and Braswell received funding from NOAA's Office of Global Programs (NA00AANRG0264). Parker's support was provided by the U.K. Government Meteorological Research Program. Through the work of the U.K. author, this paper is Crown Copyright. We thank James Angell (NOAA), Alexander Sterin (RIHMI), and Martin Stendel (DMI) for their aid in securing datasets used in this study. We thank three anonymous reviewers for their comments.

REFERENCES

Angell, J. K., 1988: Variations and trends in tropospheric and stratospheric global temperatures. 1958–87. *J. Climate,* **1,** 1296–1313.

——, 2000: Difference in radiosonde temperature trend for the period 1979–1998 of MSU data and the period 1959–1998 twice as long. *Geophys. Res. Lett.,* **27,** 2177–2180.

Basist, A. N., and M. Chelliah, 1997: Comparison of tropospheric temperatures derived from the NCEP/NCAR reanalysis, NCEP operational analysis, and the Microwave Sounding Unit. *Bull. Amer. Meteor. Soc.,* **78,** 1431–1447.

Christy, J. R., R. W. Spencer, and R. T. McNider, 1995: Reducing noise in the MSU daily lower-tropospheric global temperature dataset. *J. Climate,* **8,** 888–896.

——, ——, and E. S. Lobl, 1998: Analysis of the merging procedure for the MSU daily temperature time series. *J. Climate,* **11,** 2016–2041.

——, ——, and W. D. Braswell, 2000: MSU tropospheric temperatures: Dataset construction and radiosonde comparisons. *J. Atmos. Oceanic Technol.,* **17,** 1153–1170.

Eskridge, R. E., O. A. Alduchov, I. V. Chernykh, Z. Panmao, A. C. Polansky, and S. R. Doty, 1995: A Comprehensive Aerological Reference Data Set (CARDS): Rough and systematic errors. *Bull. Amer. Meteor. Soc.,* **76,** 1759–1775.

Folland, C. K., and Coauthors, 2001: Observed climate variability and change. *Climate Change 2001: The Scientific Basis,* J. T. Houghton et al., Eds., Cambridge University Press, 99–181.

Free, M., and Coauthors, 2002: Creating climate reference datasets: CARDS workshop on adjusting radiosonde temperature data for climate monitoring. *Bull. Amer. Meteor. Soc.,* **83,** 891–899.

Gaffen, D. J., 1994: Temporal inhomogeneities in radiosonde temperature records. *J. Geophys. Res.,* **99,** 3667–3676.

——, M. A. Sargent, R. E. Habermann, and J. R. Lanzante, 2000: Sensitivity of tropospheric and stratospheric temperature trends to radiosonde data quality. *J. Climate,* **13,** 1776–1796.

Hurrell, J., S. J. Brown, K. E. Trenberth, and J. R. Christy, 2000: Comparison of tropospheric temperatures from radiosondes and satellites: 1979–1998. *Bull. Amer. Meteor. Soc.,* **81,** 2165–2177.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Lanzante, J. R., S. A. Klein, and D. J. Seidel, 2003: Temporal homogenization of monthly radiosonde temperature data. Part II: Trends, sensitivities, and MSU comparison. *J. Climate,* **16,** 241–262.

NRC, 1999: *Adequacy of Climate Observing Systems.* National Academy Press, 51 pp.

——, 2000a: *Issues in the Integration of Research and Operational Satellite Systems for Climate Research. Part II: Implementation.* National Academy Press, 82 pp.

——, 2000b: *Reconciling Observations of Global Temperature Change.* National Academy Press, 85 pp.

Oort, A., and H. Liu, 1993: Upper-air temperature trends over the globe. *J. Climate,* **6,** 292–307.

Parker, D. E., and D. I. Cox, 1995: Towards a consistent global climatological rawinsonde data-base. *Int. J. Climatol.,* **15,** 473–496.

——, M. Gordon, D. P. N. Cullum, D. M. H. Sexton, C. K. Folland, and N. Rayner, 1997: A new global gridded radiosonde temperature data base and recent temperature trends. *Geophys. Res. Lett.,* **24,** 1499–1502.

Pielke, R. A., Sr., J. Eastman, T. N. Chase, J. Knaff, and T. G. F. Kittle, 1998: 1973–1996 trends in depth-averaged tropospheric temperature. *J. Geophys. Res.,* **103,** 16 927–16 933.

Santer, B. D., J. J. Hnilo, T. M. L. Wigley, J. S. Boyle, C. Doutriaux, M. Fiorino, D. E. Parker, and K. E. Taylor, 1999: Uncertainties in observationally based estimates of temperature change in the free atmosphere. *J. Geophys. Res.,* **104** (D6), 6305–6333.

Spencer, R. W., and J. R. Christy, 1990: Precise monitoring of global temperature trends from satellites. *Science,* **247,** 1558–1562.

——, and ——, 1992a: Precision and radiosonde validation of satellite gridpoint temperature anomalies. Part I: MSU channel 2. *J. Climate,* **5,** 847–857.

——, and ——, 1992b: Precision and radiosonde validation of satellite gridpoint temperature anomalies. Part II: A tropospheric retrieval and trends during 1979–90. *J. Climate,* **5,** 858–866.

——, and ——, 1993: Precision lower stratospheric temperature monitoring with the MSU: Technique, validation, and results 1979–91. *J. Climate,* **6,** 1194–1204.

——, ——, and N. C. Grody, 1990: Global atmospheric temperature monitoring with satellite microwave measurements: Methods and results 1979–84. *J. Climate,* **3,** 1111–1128.

Stendel, M., J. R. Christy, and L. Bengtsson, 2000: Assessing levels of uncertainty in recent temperature time series. *Climate Dyn.,* **16,** 587–601.

Sterin, A. M., 1999: An analysis of linear trends in the free atmosphere temperature series for 1958–1997. *Meteor. Gidrol.,* **5,** 52–68.

Trenberth, K. E., and J. G. Olson, 1991: Representativeness of a 63-station network for depicting climate changes. *Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations,* M. E. Schlesinger, Ed., Elsevier Science, 249–260.

APPENDIX

Radiosonde Station Information

TABLE A1. Information on individual stations used in the comparisons is provided in this appendix. Observation times (UTC) indicate the time that contained the most complete data throughout the 22-yr period (99 indicates 0000 and 1200 UTC were both well observed and averaged together).

Station	WMO no.	Lat (°N)	Lon (°E)	Elevation (m)	Obs time (UTC)
Minqin	52681	38.6	103.1	1367	99
10 VIZ 0–30°N					
Barbados	78954	13.07	−59.50	58	12
Brownsville, TX	72250	25.90	−97.42	7	00
Corpus Christi, TX	72251	27.77	−97.50	13	99
Del Rio, TX	72261	29.37	−100.92	313	99
Grand Cayman	78384	19.30	−81.36	3	12
Guam	91217	13.55	144.83	111	00
Hilo, HI	91285	19.72	−155.07	11	99
Key West, FL	72201	24.56	−81.72	6	12
Majuro, Marshall Islands	91376	7.08	171.38	3	00
San Juan, Puerto Rico	78526	18.43	−66.00	19	12
18 VIZ 30°90°N					
Albuquerque, NM	72365	35.05	−106.61	1620	00
Annette Island, AK	70398	55.03	−131.57	34	00
Barrow, AK	70026	71.30	−156.78	4	00
Bismarck, ND	72764	46.77	−100.75	506	00
Caribou, ME	72712	46.87	−68.02	190	12
Dodge City, KS	72451	37.77	−99.97	790	00
Glasgow, MT	72768	48.21	−106.62	700	00
Great Falls, MT	72776	47.47	−111.37	1130	99
Green Bay, WI	72645	44.49	−88.12	214	99
Jackson, MS	72235	32.32	−90.07	101	99
Medford, OR	72597	42.37	−122.87	405	00
Mercury, NV	72387	36.62	−116.02	1009	00
Nashville, TN	72327	36.25	−86.57	184	12
Oakland, CA	72493	37.75	−122.22	3	12
Pittsburgh, PA	72520	40.53	−80.23	370	12
Quillayute, WA	72797	47.95	−124.55	62	00
Spokane, WA	72786	47.63	−117.54	728	00
St. Paul Island, AK	70308	57.15	−170.22	9	00
VIZ B to Vaisala					
Chuuk, Federated States of Micronesia	91344	7.45	151.84	3	00