Tropospheric temperature change since 1979 from tropical radiosonde and satellite measurements

John R. Christy,[1] William B. Norris,[1] Roy W. Spencer,[1] and Justin J. Hnilo[2]

[1] Temperature change of the lower troposphere (LT) in the tropics (20°S–20°N) during the period 1979–2004 is examined using 58 radiosonde (sonde) stations and the microwave-based satellite data sets of the University of Alabama in Huntsville (UAH v5.2) and Remote Sensing Systems (RSS v2.1). At the 29 stations that make both day and night observations, the average nighttime trend (+0.12 K decade^{-1}) is 0.05 K decade^{-1} more positive than that for the daytime (+0.07 K decade^{-1}) in the unadjusted observations, an unlikely physical possibility indicating adjustments are needed. At the 58 sites the UAH data indicate a trend of +0.08 K decade^{-1}, the RSS data, +0.15. When the largest discontinuities in the sondes are detected and removed through comparison with UAH data, the trend of day and night releases combined becomes +0.09, and using RSS data, +0.12. Relative to several data sets, the RSS data show a warming shift, broadly occurring in 1992, of between +0.07 K and +0.13 K. Because the shift occurs at the time NOAA-12 readings began to be merged into the satellite data stream and large NOAA-11 adjustments were applied, the discrepancy appears to be due to bias adjustment procedures. Several comparisons are consistent with a 26-year trend and error estimate for the UAH LT product for the full tropics of +0.05 ± 0.07, which is very likely less than the tropical surface trend of +0.13 K decade^{-1}.

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

1. Introduction

[2] Recent climate model hindcasts and forecasts are consistent in depicting a tropical lower troposphere that warms at a rate about 1.3 times that of the surface [*Santer et al.*, 2005]. The same study found only one observational tropospheric data set with a ratio near 1.0, the rest having ratios less than 1.0 for the period 1979–1999. The study suggested that the observations are likely to have the greater sources of error.

[3] As noted by *Santer et al.* [2005], confidence in long-term measurements of tropospheric temperature change in the tropics (20°S–20°N) is low because of the nature of the observing systems in the region. Radiosonde (sonde) stations are not distributed uniformly and suffer from numerous inhomogeneities [*Lanzante et al.*, 2003; *Seidel et al.*, 2004; *Sherwood et al.*, 2005; *Free et al.*, 2005, *Randel and Wu*, 2006]. Deep layer observations from satellites have excellent spatial coverage but suffer from intersatellite biases, calibration deficiencies, and drifting orbits [*Christy et al.*, 1998, *Prabhakara et al.*, 2000; *Christy et al.*, 2003; *Mears et al.*, 2003; *Grody et al.*, 2004]. Uncertainty in their corrections can be of the same magnitude as the long-term trends being sought [*Christy et al.*, 2003]. Thus observational uncertainty makes checking the variability of modeled vertical temperature more difficult.

[4] Recently, attention has also been brought to the issue of sonde trends determined from day and night releases separately. *Sherwood et al.* [2005] demonstrated significant differences between trends determined from day and night releases. The differences were fairly small in the lower troposphere and quite large in the stratosphere. *Randel and Wu* [2006] followed with a similar study, which included satellite comparisons that documented the day and night differences. They also noted trend differences between sondes deemed of "high" and "low" quality. On this latter topic, *Randel and Wu* [2006] demonstrated that the trends of high-quality sondes were more positive than those of low-quality sondes. Again, the largest discrepancies were in the stratosphere. A conclusion of both studies was that uncorrected, trends from sondes were likely too negative, even in the troposphere.

[5] Both studies noted the relative difference between trends of sondes divided according to release time or quality. However, neither study tried to determine the effects of more fundamental changes that could affect both day and night sondes or both high- and low-quality sondes. *Christy and Spencer* [2005] discussed this issue and indicated that both day and night releases of the Australian sondes (also of high quality), which were prominent in the data sets used by both *Sherwood et al.* [2005] and *Randel*

[1]Earth System Science Center, University of Alabama, Huntsville, Alabama, USA.
[2]Lawrence Livermore National Laboratory, Livermore, California, USA.
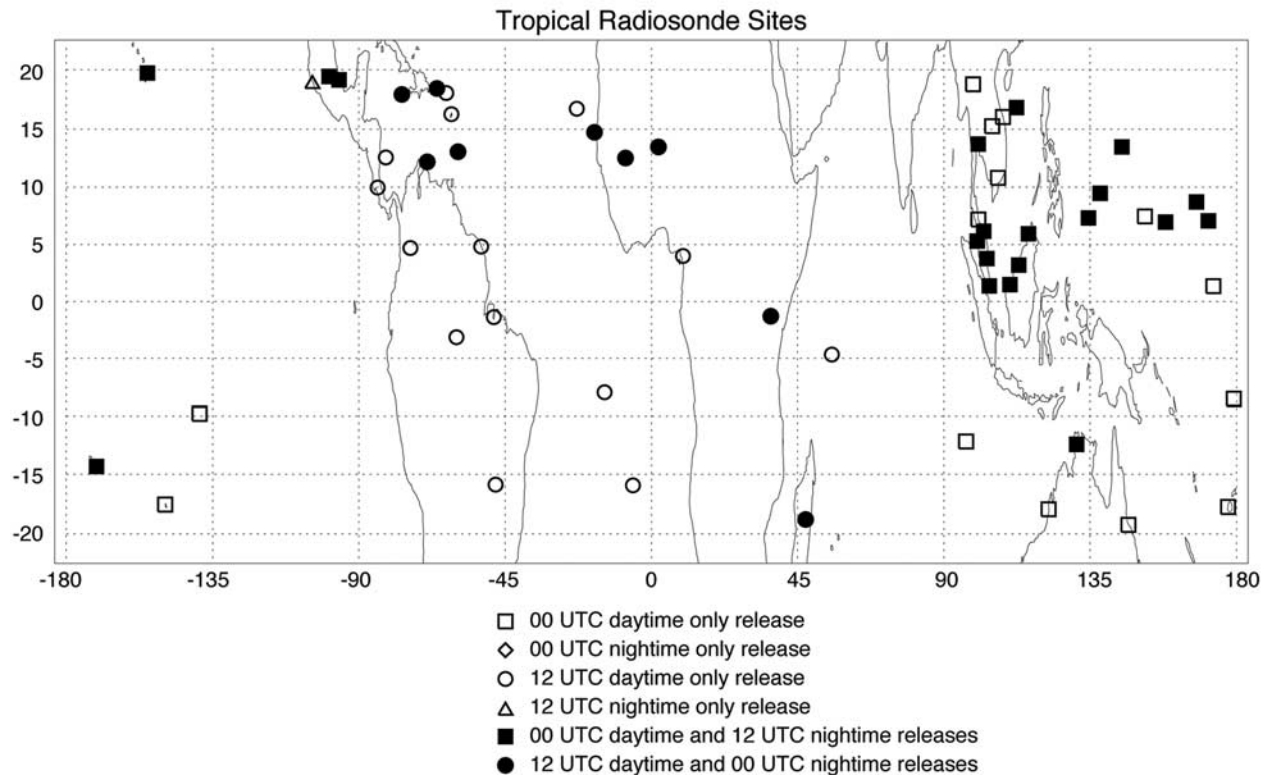
Figure 1. Location of the 58 tropical sonde stations whose temperature observations were used in this study. Symbols indicate the times sondes were released at each station and whether the releases occurred in the day or at night. Vertical scale is exaggerated to allow the reader to more clearly distinguish the stations.

and Wu [2006], likely contained spuriously warm tropospheric biases. Therefore absolute trends based only on differences of release time or quality designations could not be determined with high confidence. Because of the evidence for spuriously warm and cold biases in trends produced from sonde time series, it is important to understand the magnitude of these problems. Therefore we shall look at results from day and night releases separately and assess the impact of differences on overall trends.

[6] The purpose of this study is to use two satellite data sets in a series of experiments to identify the largest discontinuities in tropical sonde data relative to the two satellite data sets, and having eliminated them, to investigate changes in tropical tropospheric temperature over the past 26 years (1979–2004). In the process insight is gained into reasons for the differences between the satellite data sets and the sondes and whether the model ratio of troposphere-to-surface trends ($\sim$1.3) is supported by the observations.
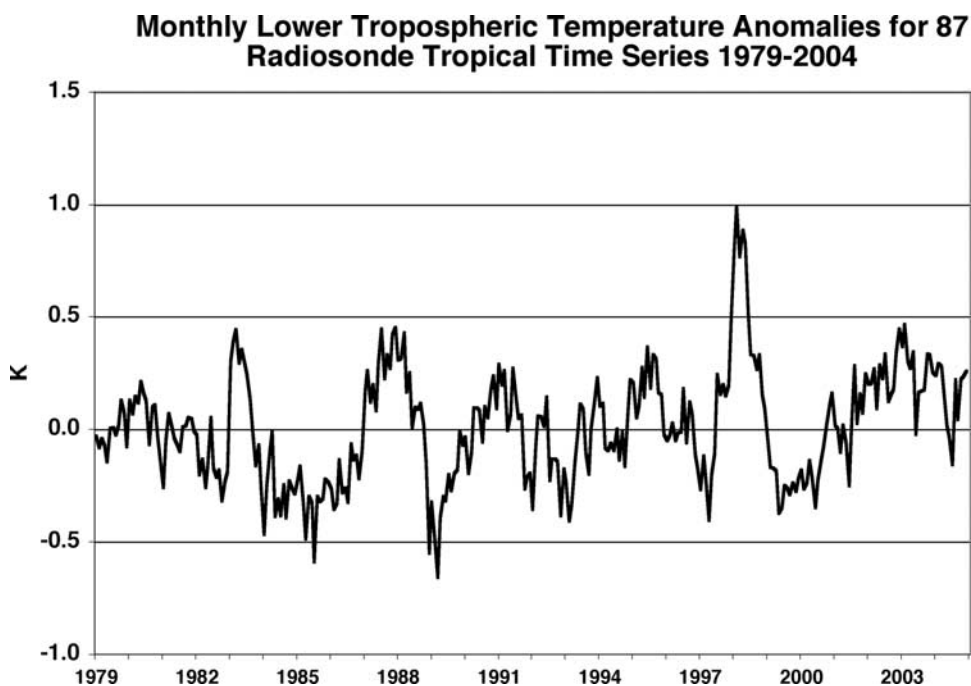
2. Satellite Data

[7] The University of Alabama in Huntsville (UAH) produces a lower-tropospheric (LT) temperature product based on a weighted difference of measurements from selected view angles of Microwave Sounding Unit (MSU) channel 2 and Advanced MSU channel 5 [*Spencer and Christy*, 1992; *Christy et al.*, 2003; *Spencer et al.*, 2006]. Version 5.2, used here, includes the updated diurnal cycle

adjustment to correct an error pointed out by C. Mears and F. Wentz (personal communication, May 2005). Remote Sensing Systems (RSS) produces a similar LT product, updated through 2004, which also uses MSU channel 2 [*Mears and Wentz*, 2005]. These products capture temperature variations in the layer from the surface to about 350 hPa. Over land, surface emissions can contribute up to 20 percent of the signal.

3. Radiosonde Data

[8] We examined the data records of the 183 tropical stations available in the Integrated Global Radiosonde Archive (IGRA) database at the National Climatic Data Center [*Durre et al.*, 2005]. To use a sonde profile to simulate LT, we required that it reach at least 100 hPa. To include a station, we required that it have at least 180 of the possible 312 months of data. Enforcing these criteria reduced the number of stations to 73. Comparing the sonde and satellite series for consistency we eliminated the 43000 block (India) as being unacceptably noisy (as in *Parker et al.* [1997] and *Lanzante et al.* [2003]). We were left with the 58 stations shown in Figure 1 and described in Tables S1 and S2[1]. Of these, 29 provided observations for both day and night, 28 for day only, and one for night only. Because

---

Monthly Lower Tropospheric Temperature Anomalies for 87 Radiosonde Tropical Time Series 1979–2004

Figure 2. Monthly average lower troposphere (LT) temperature anomalies of the 87 time series derived from the radiosonde observations at the locations in Figure 1. These include 57 daytime series and 30 nighttime series.

day and night soundings are processed with different algorithms to account for the effects of contaminating solar radiation, we considered a station with both day and night releases as two stations. This gave 87 times series of tropospheric tropical temperatures of which 57 were day only and 30 were night only. Later we refer to these three sets as "all," "day," and "night."

[9] Sonde and satellite data can be compared by using the sonde observations to simulate the satellite brightness temperatures. This is done by taking the sonde temperatures at each pressure level and proportionally weighting them according to the contribution of that level to the microwave product. Full radiation code is applied to all reporting levels so that the effects of humidity and of surface emission and reflection are included. (Humidity is often missing, and in such cases we employed the climatological value. In addition, humidity is often a poorly observed variable, but its variation has a very small effect on microwave brightness temperatures; e.g., a 20% change in monthly mean humidity would impact the tropical brightness temperature by less than 0.02 K [*Spencer et al.*, 1990]).

4. Radiosonde and Satellite Trends

[10] Figure 2 shows the averaged monthly LT anomalies of the 87 sonde time series. When matched station for station and month for month, the UAH and RSS anomaly time series are visually almost identical to Figure 2 (differences will be displayed later). The trends of the three series, as calculated by least-squares regression and expressed in K decade^{-1}, are +0.07 (sondes), +0.09 (UAH), and +0.15 (RSS). For the tropics as a whole, the satellite data indicate the trends are +0.05 (UAH) and +0.15 (RSS). (Trends from

satellite products will vary depending on the application in this paper. Here we mention trends (K decade^{-1}) based on the 87-sonde grids (i.e., double counting those grids for which the station had both day and night observations). These values are +0.09 (+0.15) for UAH (RSS). This will also be subdivided into trends based on the 57 day station grids or 30 night station grids shortly. As mentioned here, however, the trend for the full tropics, representing all grids, is +0.05 (+0.15) for UAH (RSS). In the Abstract and later, the trends calculated from the 58 grids where stations reside will also be used (i.e., not double counting the grids where both day and night releases occur) and those trends are +0.08 (+0.15) for UAH (RSS).)

[11] One focus of this paper is the determination of measurement errors of trends, a concept which differs from "statistical" or "temporal sampling" errors. The latter two seek to answer the question, "How well does this current period represent other similar periods?" For statistical errors, the magnitude of interannual variability is highly important, and even a perfect measurement system will have statistical error bars [*Folland et al.*, 2001]. In other words, a time series with large interannual variability will have a large statistical error because sampling such a time series over differing periods can lead to differing trends. "Statistical" error will not be reported in this paper.

[12] The question for this paper rather is, "How well do current estimates of the temperature trend agree with the true tropical trend?" This is a different question and aims to understand errors in the data sets which arise from events such as changing instrumentation. UAH (see section 10 below) and RSS [*Mears and Wentz*, 2005] estimate their tropical measurement errors as ±0.07 and ±0.09 K decade^{-1} respectively. Though there is some overlap of error bars
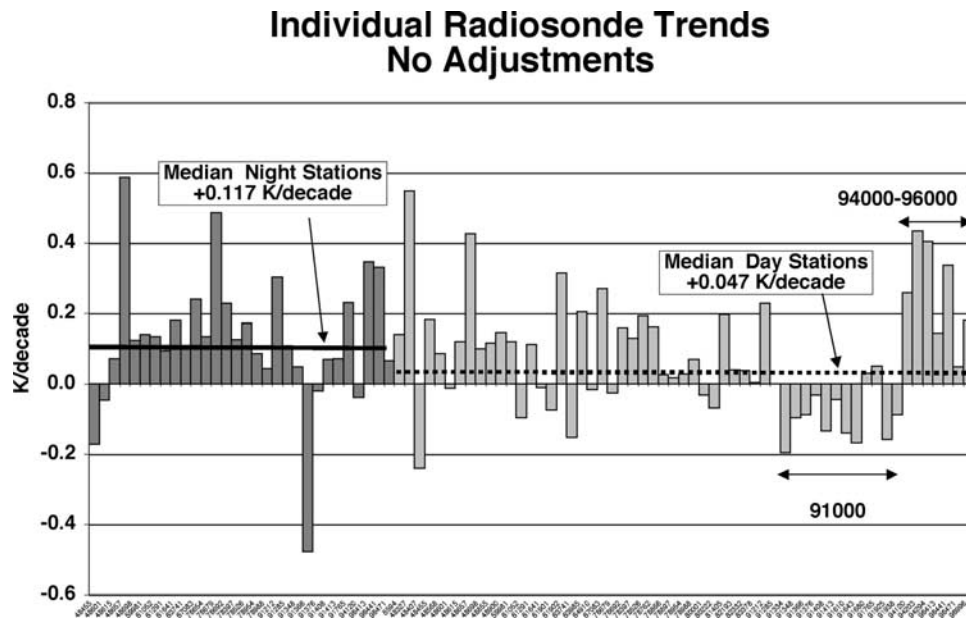
Figure 3. Individual trends (K decade^{-1}) of the 87 sonde LT time series, obtained by least squares linear regression. Dark gray bars (on the left) represent trends at the 30 stations with nighttime observations. Light gray bars represent trends at the 57 stations with daytime observations. Numbers on the right identify WMO blocks: 91000 (western Pacific), 94000–95000 (primarily Australia), 96000 (southwest Asia).

between the UAH and RSS tropical trends, the time series of the differences between the two is indeed significant, as will be shown later.

[13] Figure 3 shows the individual trends at the sonde stations, with the 30 night stations on the left and the 57 day stations on the right. Although some of the scatter can be explained by geographical variations or missing data, most is attributable to other causes. For determining a mean trend for the tropics, the 58-station distribution should be adequate since there are only about eight degrees of freedom in the tropical belt [*Hurrell and Trenberth*, 1996]. The 58 stations are not distributed in a geographically optimal fashion for tropical averages, however. Our study will rely mostly on the direct, site-by-site comparisons for error estimation which may then, through the geographically complete satellite data, be upscaled to the tropics as a whole.

[14] We rely on the median of the individual trends to reduce the influence of outliers and consistently characterize the central tendency of the sample. The median trends of the night and day stations (+0.12 K decade^{-1} and +0.05 K decade^{-1}) are shown as horizontal lines in Figure 3. The daytime trends of the 91000 block (western Pacific) are consistently lower than the tropics as a whole. The 94000 block (primarily Australia) and the 96000 block (Southeast Asia) are consistently higher. These results raise the possibility of regional changes in instrumentation since the two latter blocks are physically adjacent.

## 5. Day and Night Differences

[15] The differences in day and night trends at a given station arise from changes in algorithms and instrumentation which through the years have sought to reduce the effects of direct daytime solar heating and spurious nighttime cooling of the temperature sensor [*Sherwood et al.*, 2005]. Early sensors generally reported temperatures that were spuriously warm because the sensor itself, heated by direct sunlight, became warmer than the ambient air. As the sondes ascended, the problem became worse so that values reported for the stratosphere could be several degrees in error. As a result trends computed from daytime sondes are more prone to radiation errors than those at night.

[16] An estimate of the corrected daytime trends can be made by considering the day-minus-night (DMN) time series. To overcome the problem of day and night stations not being fully coincident in space and time and not always using similar instruments, we only compare the individual trend differences for the 29 stations having both day and night observations (Figure 4). For these sites the median DMN trend is −0.046 K decade^{-1}. If we assume that this trend captures a systematic difference in day and night sondes, we simply subtract this value from the day station median trend of 0.047 K decade^{-1} (based on all 58 stations) to obtain a new median daytime trend for all day stations of 0.093 K decade^{-1}. This crude result is actually quite close to results determined later using more complex techniques.

[17] The variability in the station DMN trends is large relative to the median trend (−0.046 K decade^{-1}), having a 95 percent confidence interval of ±0.08 K decade^{-1}. When this trend is tested against the hypothesis that it is not statistically different from zero, the hypothesis cannot be disproved at the 95 percent confidence level. However, the fact that stations in the 91000 block consistently show more negative trends in the day observations than they do in those
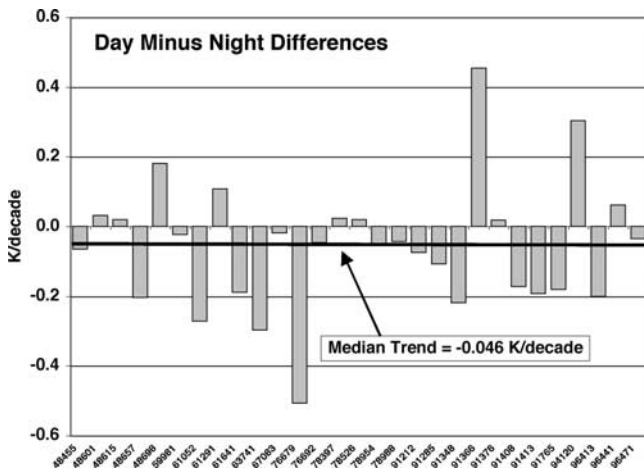
Figure 4. Trends (K decade^{-1}) of day-minus-night (DMN) LT observations at the 29 stations having both daytime and nighttime releases.

at night suggests that we should investigate stations individually.

6. Error Adjustment Experiments

[18] Sonde data can be contaminated by significant shifts in temperature due to instrument changes [*Gaffen*, 1994; *Gaffen et al.*, 2000; *Lanzante et al.*, 2003]. In 1979 only 11 of the 58 stations used Vaisala sondes, RS-18 or RS-21, [*Gaffen*, 1996]. By 2004 about 40 were using Vaisala products (RS-80 and successors) with none still using the RS-18 or RS-21 models (World Meteorological Organization, 2006, WMO Catalogue of Radiosondes and Upper Air Wind Systems, available at http://www.wmo.ch/index-en.html). All stations experienced sonde changes including those that used Vaisala or VIZ instruments

consistently [*Christy and Norris*, 2006]. The trends for blocks 91000, 94000, and 96000 (Figure 3) give evidence that these changes caused shifts of differing signs in some of the 58 stations.

[19] To detect the magnitude of possible nonclimatic changes, we follow the procedure of *Christy and Norris* [2006]. The simulated LT time series from individual sondes are compared to corresponding satellite series that are made consistent with the sonde series in space and time [*Christy et al.*, 2003; *Christy and Norris*, 2006]. From these we associate with each station the difference of these two series (sonde minus satellite). The 87 resulting difference series can be examined for significant shifts that may indicate station moves or changes in station operation, instrumentation, or data processing.

[20] In this approach the satellite data are considered to be the reference. Unfortunately, satellite data can have spurious local shifts because of the insertion or deletion of observations from a new or decommissioned spacecraft. For example, *Christy and Norris* [2006] report instances when local shifts were detected between the two satellite data sets at the point in time when a new satellite was incorporated, suggesting that differing merging methods were responsible. Tests of the UAH and RSS midtropospheric (MT) satellite products (surface to about 100 hPa) against sonde data, for which shifts were known, showed that the magnitude of the satellite error shift could be as much as 0.10 K [*Christy and Norris*, 2006]. Thus shifts greater than this amount are more likely to be due to inhomogeneities in the sonde data. For the LT product the satellite error could be about twice this amount [*Christy and Norris*, 2006]. Therefore the difference series would not be expected to detect discontinuities in the sonde data for shifts much less than 0.2 K at a single station with high confidence that the shift was due to sonde changes.

[21] Breakpoints can be discovered by transforming the difference time series as follows. For a target month the difference series is first averaged for the 36-month period
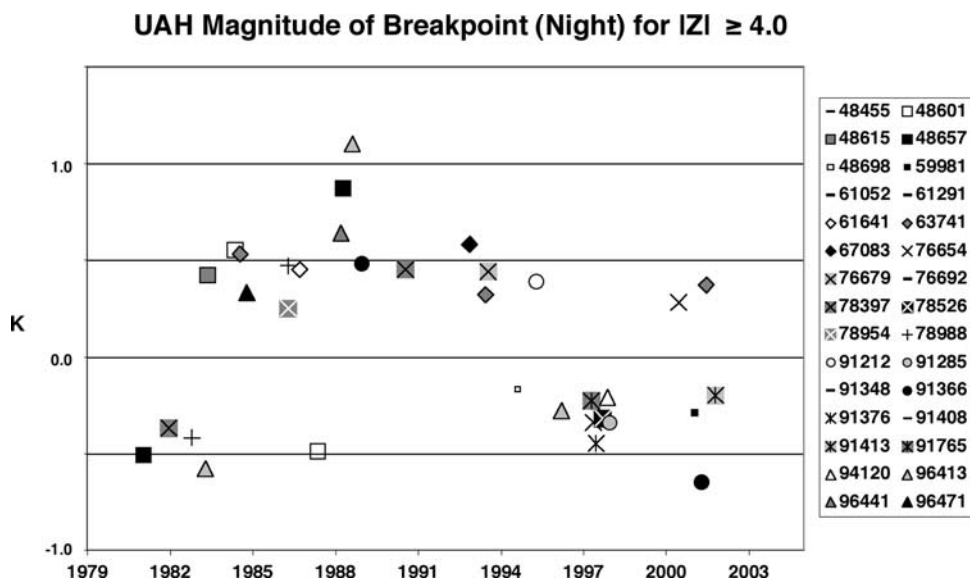
Figure 5a. Size of the breakpoints (K) for the 30 stations with nighttime releases whose breakpoints had absolute z-scores ($|z|$) equaling or exceeding 4.0. Breakpoints were identified with the University of Alabama in Huntsville (UAH) data.
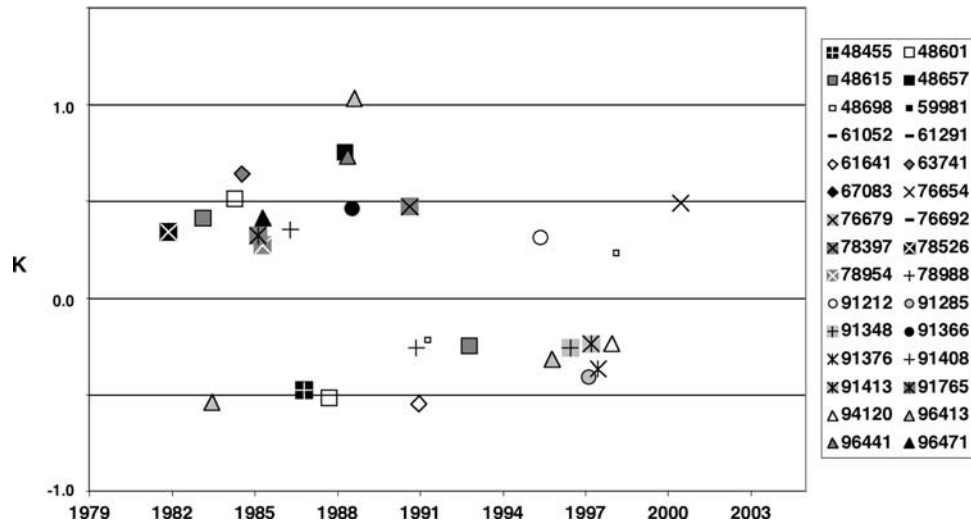
Figure 5b. As in Figure 5a except breakpoints based on Remote Sensing Systems (RSS) data.

before the target and then for the subsequent 36 months. The value associated with the target month is the difference of the two averages. If a shift does not occur during the target month and any existing trends are small, the difference is approximately zero. To serve as a metric for the discovery of breakpoints, the z-score is computed on the basis of the difference of 36-month averages on either side of each target month. The z-score is the ratio of a certain quantity divided by its standard error adjusted for loss of degrees of freedom. Here the z-score is derived from the difference of the two consecutive 36-month means of the difference times series. The standard error used in the computations has been adjusted for the loss of degrees of freedom due to autocorrelation.

[22] The threshold for breakpoint detection should be large enough to identify breakpoints outside the level of satellite error and thus not completely compromise the independence of the satellite and sonde data. If $|z| \geq 5.00$, the magnitude of the breakpoint is almost always greater than 0.3 K; if $|z| \geq 4.00$, almost always greater than 0.2 K. (If the satellite data were without error, a z-score of ± 2.90 would be significant at the 99-percent confidence level.) Thus $|z| = 4.0$ was taken as the lowest threshold for LT series used in these experiments.

[23] The experiment found many instances of $|z| \geq 4.00$ in the 87 times series, even up to $|z| = 17$. As examples, Figures 5a (UAH) and 5b (RSS) show the breakpoints for the night stations when the criterion is $|z| \geq 4.00$. A positive value indicates that a sonde experienced a positive shift relative to the satellites. An adjustment to the sonde would therefore cause the trend for that station to become more negative. Of the 30 night stations only six had no shifts relative to UAH and only seven relative to RSS.

7. Results for $|z| \geq 5.00$

[24] We began a detailed analysis of breakpoint identification and removal by considering the more extreme case of $|z| \geq 5.00$. The results are summarized in the rows of Table 1 where 5.00 appears in column 2.

[25] Relative to the UAH experiment, discontinuities with $|z| \geq 5.00$ were detected in about half of the series, and the signs were split evenly between positive and negative. This was true for "day", "night", and "all" sondes and suggests that there are no large, systematic changes either in the UAH data or in any of the three sets of sondes. The RSS data set generated three times more positive shifts than negative in the night sondes, but three times more negative than positive in the day sondes. These differences affect the trends of the adjusted sonde data sets.

[26] The median trend of the unadjusted night sondes is +0.117 K decade^{-1}. After adjustment, the median trend is +0.093 K decade^{-1} relative to UAH and +0.083 K decade^{-1} relative to RSS—an absolute difference of only 0.010 K decade^{-1}. RSS would be expected to produce a cooler trend than the one produced by UAH since RSS removed more positive and fewer negative shifts than UAH. Because both UAH and RSS produced cooler trends for the night sondes, the unadjusted night trend is likely to be spuriously warm on the basis of these experiments.

[27] The median trend of the unadjusted day sondes is +0.047 K decade^{-1}. After adjustment, the median trend is +0.086 K decade^{-1} relative to UAH and +0.120 K decade^{-1} relative to RSS—an absolute difference of 0.034 K decade^{-1}. Again, RSS would be expected to produce a warmer trend than the one produced by UAH since RSS removed more negative and fewer positive shifts than UAH. Because both UAH and RSS produced warmer trends for the day sondes, this unadjusted trend is likely to be spuriously cool. Similar results hold for the set of all sondes: (unadjusted: +0.073; UAH: +0.092; RSS: +0.118; absolute difference: 0.026).

[28] Comparing Figures 6a and 6b with Figure 3 shows that adjusting by either UAH or RSS reduces the variability in the daytime trends, but adjusting by RSS does little to reduce the variability in the nighttime trends. As shown in Table 1 the difference in the medians of the night and day trends is 0.070 K decade^{-1} for the unadjusted sonde data. When data are adjusted with UAH data for the $|z| \geq 5.00$ breakpoints, the difference between the 57 day and 30 night

Table 1. Statistics of Lower-Troposphere (LT) Tropical Time Series

Series[a]	z-Score[b]	Stations[c]	Breakpoints	Positive Breakpoints	Negative Breakpoints	Median Trend[d]	Correlation Versus Sonde[e]	RMS Trend Difference[f]
				NIGHT				
Sondes	None	(30)				+0.117		
UAH	None					+0.091	0.894	0.201
RSS	None					+0.148	0.920	0.214
Sondes (UAH)	5.00	13	16	8	8	+0.093	0.934	0.175
Sondes (RSS)	5.00	11	12	9	3	+0.083	0.948	0.204
Sondes (UAH)	4.00	24	34	18	16	+0.095	0.955	0.146
Sondes (RSS)	4.00	23	29	16	13	+0.130	0.951	0.205
				DAY				
Sondes	None	(57)				+0.047		
UAH	None					+0.076	0.968	0.178
RSS	None					+0.147	0.942	0.188
Sondes (UAH)	5.00	33	46	23	23	+0.086	0.969	0.142
Sondes (RSS)	5.00	30	45	11	34	+0.120	0.964	0.157
Sondes (UAH)	4.00	46	83	39	44	+0.086	0.971	0.126
Sondes (RSS)	4.00	45	81	30	51	+0.145	0.971	0.148
				ALL[g]				
Sondes	None	(87)				+0.073		
UAH	None					+0.090	0.966	0.187
RSS	None					+0.146	0.958	0.196
Sondes (UAH)	5.00	46	60	31	31	+0.092	0.972	0.154
Sondes (RSS)	5.00	41	57	20	37	+0.118	0.971	0.175
Sondes (UAH)	4.00	70	117	58	59	+0.091	0.976	0.134
Sondes (RSS)	4.00	68	115	48	67	+0.136	0.973	0.170
				DAY-NIGHT[h]				
Sondes	None	(29)				−0.046		0.195
Sondes (UAH)	5.00					−0.022		0.195
Sondes (RSS)	5.00					+0.046		0.222
Sondes (UAH)	4.00					−0.020		0.180
Sondes (RSS)	4.00					+0.087		0.250

[a]An entry of the form A refers to an unadjusted series. An entry of the form A (B) indicates that series A has been adjusted by the removal of breakpoints detected by B. The University of Alabama in Huntsville (UAH) and Remote Sensing Systems (RSS) series come from sampling the full tropical LT data sets at the locations and releases of the sonde stations.

[b]The z-score is defined in the text. For adjusted series, this column gives the breakpoint detection threshold z_0 in the criterion $|z| \geq z_0$.

[c]An entry in parentheses indicates the total number of stations in a category (night, day, etc.). Otherwise, an entry indicates the number of stations in the category in which breakpoints were detected.

[d]The median of the trends of all series in a category. In most cases the trends come from a mixture of adjusted and unadjusted series; e.g., for sondes having nighttime releases, there are 13 adjusted series and 17 unadjusted series when breakpoints are detected by $|z| \geq 5.00$.

[e]The correlation coefficient of the composite unadjusted sonde series in a category with the composite of the other series named in Column 1 in the same category.

[f]The root mean square of a set of trend differences. Within a category, satellite trends are always subtracted from sonde trends. When none of the sonde series has been adjusted, the RMS value is entered in the row labeled UAH or RSS in column 1 according to the satellite data set used. When sonde series have been adjusted by UAH (RSS), then UAH (RSS) trends are subtracted from sonde trends.

[g]The 29 stations having both daytime and nighttime releases are double counted.

[h]Day-night difference series were computed only for the 29 stations having both daytime and nighttime releases and after being matched month for month. Adjustments to series, if any, were made before the differences were computed.

sondes is reduced to 0.007 K decade^{-1}. When the same is done with the RSS data, the difference is reduced to 0.037 K decade^{-1}. For the 29 sondes with both day and night releases, the median difference of −0.046 K decade^{-1} is reduced to −0.022 using UAH and changed to +0.046 using RSS (bottom entries of Table 1). Recall that day and night corrections are performed independently, so an expected outcome of each experiment, if corrections are realistic, should be a reduction in the day and night differences. In the case of RSS experiment, the 29 night sondes received adjustments which produced a median trend now less than that of the 29 day sondes, hence a positive DMN value.

8. Results for $|z| \geq 4.00$

[29] As shown earlier, the breakpoints identified by the UAH data using a $|z| \geq 4.00$ criterion at night stations are given in Figure 5a; those for RSS, in Figure 5b. The results are similar for the early years, especially around 1988 when several Philips Mark III sondes were changed to Vaisala RS-80. These changes likely account for significant positive shifts in the nighttime readings. The cluster of negative shifts around 1997 are likely associated with changes from VIZ-B to either VIZ-B2 or Vaisala [*Elliott et al.*, 2002; *Christy and Norris*, 2006]. Of special interest are the negative shifts in Figure 5b around 1991 that do not appear in Figure 5a (also true for daytime releases, not shown). These shifts are important because a shift in the center of a time series has a relatively large impact on the overall trend. (If a relatively flat time series is subjected to a positive shift, the trend increases. For a time series where such a shift occurs, the linear trend may incorrectly portray the underlying variability [*Seidel and Lanzante*, 2004; *Thorne et al.*, 2005a].) The changes in the trends resulting from

Radiosonde Trends Adjustments |Z| ≥ 5.00
UAH Adjustments
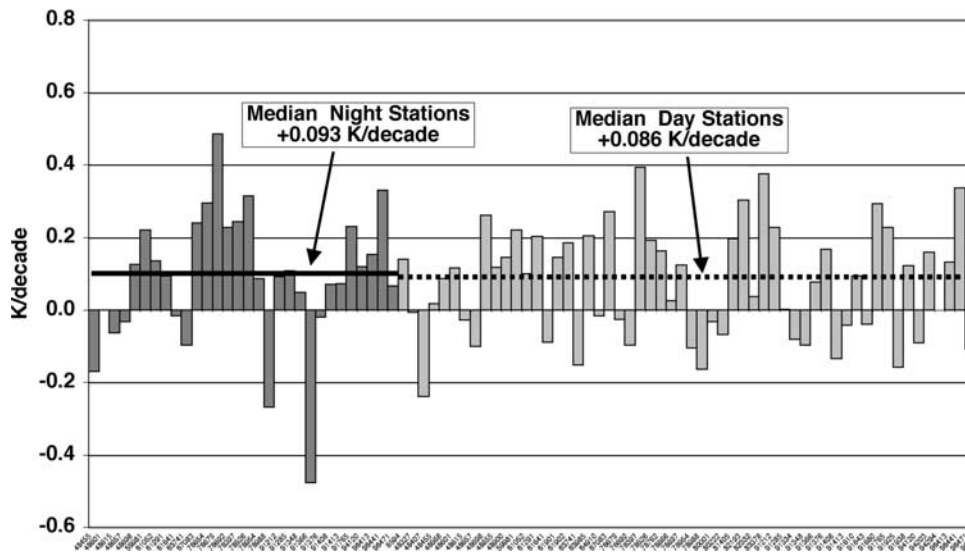
Figure 6a. The same as Figure 3 except that trends are of sonde series adjusted to remove breakpoints having *z*-scores satisfying $|z| \geq 5.0$. Adjustments were based on breakpoints identified with UAH data.

UAH adjustments are no greater than 0.002 K decade^{-1} between the $|z| = 4.00$ and $|z| = 5.00$ thresholds, but for RSS adjustments the nighttime trends change by 0.047 K decade^{-1} and the daytime trends, by 0.025 K decade^{-1}. (Not listed in Table 1 is the median correlation of the individual trends, which in every case, is higher for UAH relative to RSS.)

[30] Table 1 shows that when the nighttime sondes are adjusted, the sonde vs. satellite variability (i.e., r.m.s. of difference trends) declines from 0.201 to 0.175 K decade^{-1}

relative to UAH but changes comparatively less (0.214 to 0.204 K decade^{-1}) relative to RSS.

[31] As mentioned, adjustments should cause the DMN trend to move closer to zero from the value of -0.046 K decade^{-1} since the day and night time series are both being adjusted against the same satellite time series. This is the case for the UAH experiments. For the RSS experiments, the 29 night sonde series were adjusted so that their median trend became more negative than the 29 day sonde series
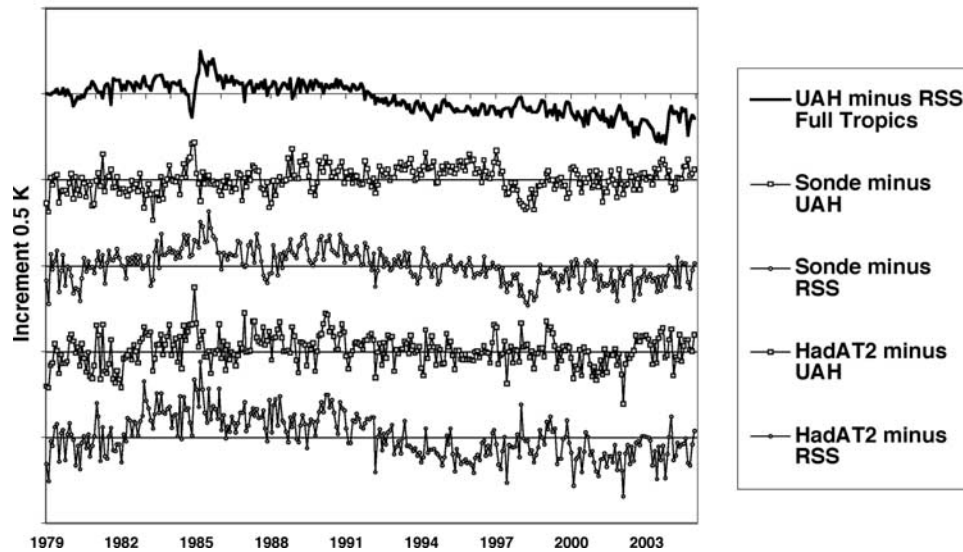
Radiosonde Trends Adjustments |Z| ≥ 5.00
RSS Adjustments
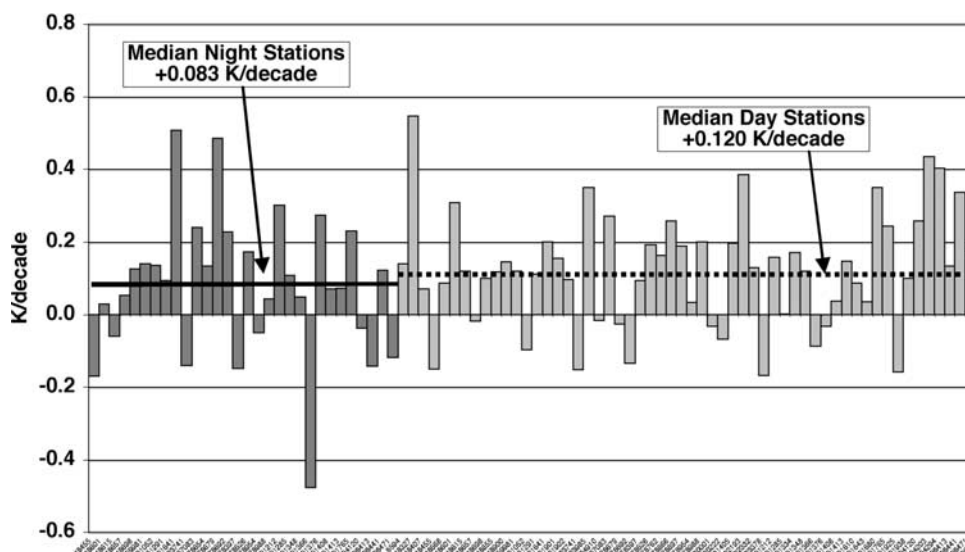
Figure 6b. As in Figure 6a except breakpoints based on RSS data.

Monthly Differences 1979-2004

Figure 7. Differences of monthly anomalies for pairs of tropical LT time series. The "sonde" data sets are unadjusted. When satellite data sets are compared to each other, the full tropical area is used. For the sondes (and for satellites when they compared with the sondes), 87 series are composited, which means that the 29 stations with both day and night releases are double counted. HadAT2 is a gridded, homogenized data set of sonde temperatures [*Thorne et al.*, 2005a]. For intercomparisons of satellite data and for HadAT2 comparisons with satellite data, the satellite data sets represent the full tropics.

resulting in a median DMN trend which changed sign and for which the magnitude increased.

[32] From the entire set of experiments, it appears that the unadjusted daytime sonde series are likely too negative in trend while trends of the nighttime sondes are likely two positive. We note that relative to both satellite data sets several sondes consistently indicate positive shifts in the 1980s and negative shifts in the late 1990s. The detection of shifts by UAH and RSS is somewhat different in the early 1990s with RSS detecting more negative sonde shifts. This difference motivates the following section in which UAH and RSS are directly compared.

9. Comparison of UAH and RSS Data Sets

[33] To compare the UAH and RSS data sets, we take the difference of their anomaly series. We also subtract them from the anomaly series of the sondes and HadAT2, an independently constructed and homogenized sonde data set [*Thorne et al.*, 2005a]. Figure 7 displays the five resulting difference series. If any of these data sets are consistent, the trends of these difference series should not be significantly different from zero. The trends are: -0.099 K decade^{-1} (UAH-RSS); 0.018 K decade^{-1} (sonde-UAH); -0.054 K decade^{-1} (sonde-RSS); 0.003 K decade^{-1} (HadAT2-UAH); and -0.078 K decade^{-1} (HadAT2-RSS). The 95-percent confidence intervals range from ± 0.020 to ± 0.039 K decade^{-1}. Thus for sonde-RSS or HadAT2-RSS, the trends are significantly different from zero. The corresponding UAH differences are not.

9.1. The 1985–1987 Period

[34] The most obvious feature in the UAH-RSS series of Figure 7 is the large negative and positive spikes during 1985 and 1986. The satellite differences during 1985–1987, influenced considerably by the NOAA-9 instrument and its interaction with NOAA-7 and NOAA-8, are the basis for a large part of the lack of agreement between RSS and UAH in other time series [*Christy and Norris*, 2004]. In the LT comparison presented here, this appears as the up-and-down shift.

[35] The shift has little impact on the long-term trend. For the period 1979–1991, the RSS trend (-0.084 K decade^{-1}) is almost identical to the UAH trend (-0.064 K decade^{-1}). The negative spike in 1985 in the UAH-RSS difference series is repeated in both sonde-UAH comparisons (as a positive spike) and is probably related to a problem in the UAH data. The broader positive spike in 1985 appears in the sonde-RSS comparisons and is probably related to a problem in the RSS data. The resolution of this issue will have little effect on the overall trends.

9.2. The 1991/1992 Breakpoint
9.2.1. Examination of Difference Series

[36] The largest and most significant breakpoint in the UAH-RSS difference series occurs between October 1991 and March 1992 (Figure 7). The NOAA-12 spacecraft began adding data into the data stream around October 1991. In addition, adjustments applied to NOAA-11, which have their greatest effect in the latter half of the satellite's operation, i.e., after October 1991, may also be a factor. The schematic in Figure 8 shows how a several month shift could occur in one of the data sets. This appears to have
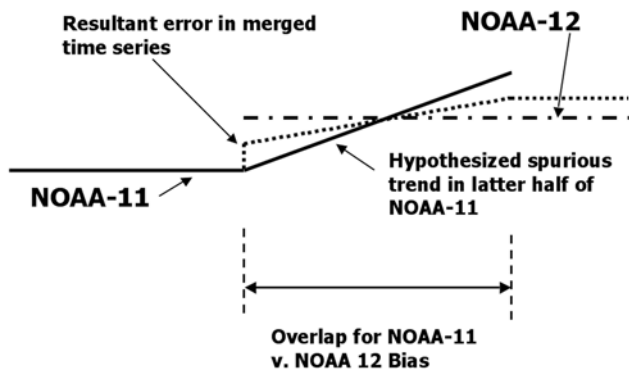
Figure 8. Schematic of hypothesized cause of the difference between UAH and RSS tropical LT temperatures. Suppose a satellite data set shows an upward trend of temperatures for NOAA-11 during the period of overlap with NOAA-12. When bias adjustments are made to merge NOAA-11 and NOAA-12, the new temperatures appear as an upward shift followed by a trend (dotted line). If a second satellite data set does not show a trend in NOAA-11 in the overlap, the difference series of the two will contain a breakpoint that reaches its full magnitude at the end of the overlap.

occurred in the RSS data set relative to UAH's and is reflected in the RSS-UAH difference series in Figure 7, although the diagram is in the opposite direction.

[37] Further, UAH and RSS determine the relative bias of NOAA-12 to NOAA-11 differently. UAH uses a simple "backbone" method in which direct, latitude by latitude biases are calculated and removed first between NOAA-10 and NOAA-11, then between NOAA-11 and NOAA-12. RSS applies a consensus method in which a single global mean bias per satellite is determined as part of the method which calculates the adjustment parameters for the temperature fluctuations of the instrument [*Mears et al.*, 2003]. For RSS, this calculation includes influences of the short overlap between NOAA-10 and NOAA-12. UAH calculates the adjustment parameters separately from the biases and does not include information from the overlap between NOAA-10 and NOAA-12. Thus differences in the merging methodology from the end of NOAA-10 through the beginning of NOAA-12 likely explain this highly significant temperature difference between UAH and RSS at this time.

[38] This reasoning explains the shift in the RSS-UAH difference series but does not identify the data set having the greater error. To aid in identifying the potential error we compare the RSS and UAH data sets with data from sonde compilations. The sondes allow for an independent assess-
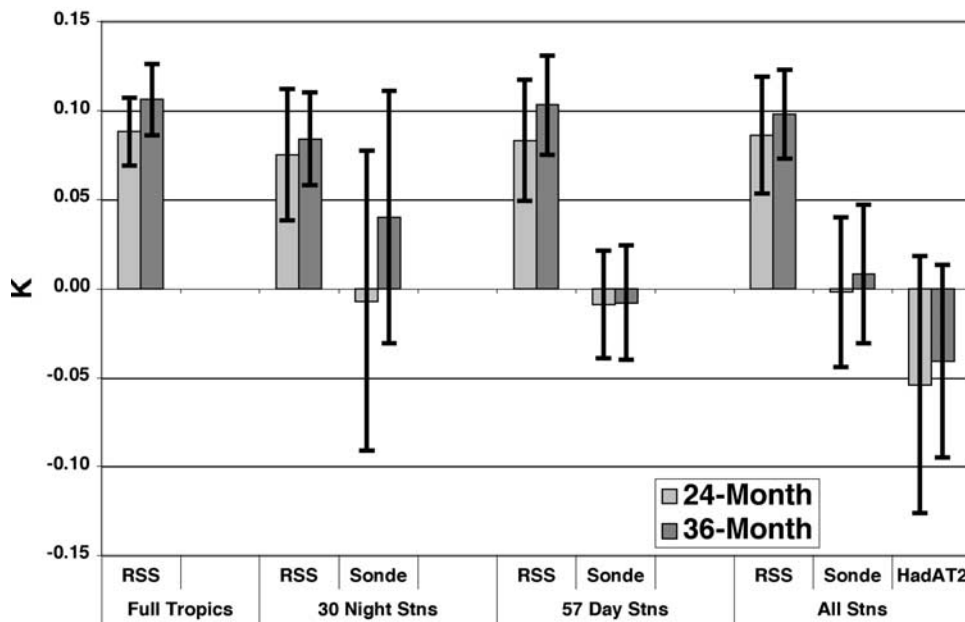
Figure 9a. Differences between averages ending with December 1991 and averages beginning with January 1992 in series of monthly anomaly differences. Light (dark) gray bars are the differences when the averaging time is 24 (36) months. The difference of the before-and-after averages represents the change in temperature caused by the inclusion of NOAA-12 data. In the figure, B = UAH and A ranges over the remaining data sets (refer to equations (1) and (2) in the text for an explanation of the notation). Thus the bars represent the shifts the various data sets detect in UAH. When UAH and RSS are compared, the full tropical satellite data sets are used; otherwise, they are sampled at station locations and release times. The sonde data are unadjusted. The HadAT2 data have undergone rigorous homogenization [*Thorne et al.*, 2005a].
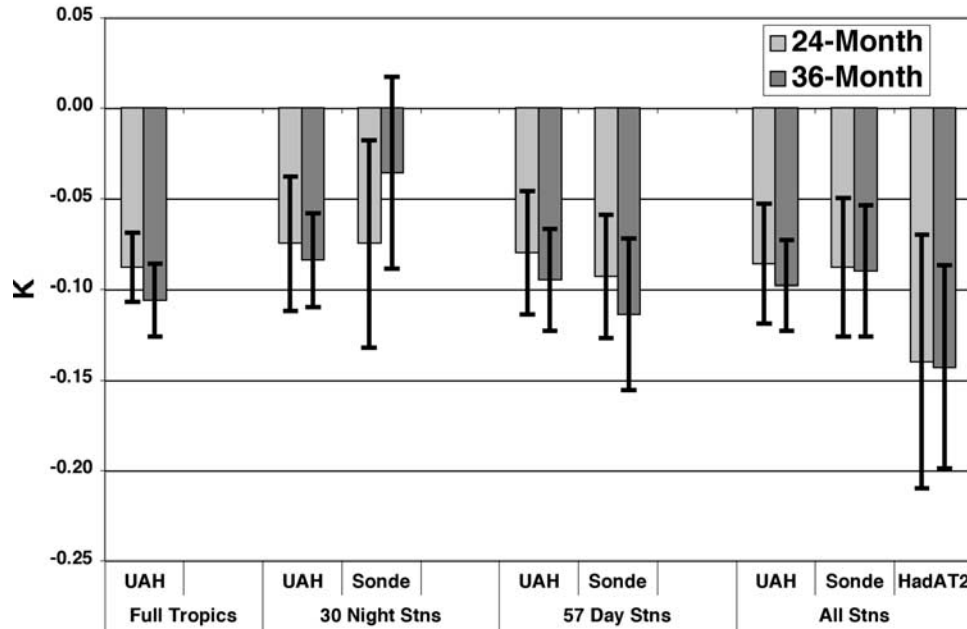
Figure 9b. As in Figure 9a except B = RSS and A ranges over the remaining data sets in equations (1) and (2) of the text.

ment since they have not yet been used in the RSS-UAH comparison. We also appeal to the HadAT2 data set of adjusted sondes as a further check. For these comparisons we shall use the average of the satellite data at the 58 station grids.

[39] Because the 1991/92 shift is not sudden, we tested several breakpoint periods by inserting gaps between the two periods being differenced. In all cases, the largest shift occurred for gaps beginning in January 1992. The corresponding z-scores ranged from 8 to 10, indicating a highly significant difference between UAH and RSS at this point in the time series.

[40] To clarify the approach for quantifying the break, let A and B be two time series, and let

$$\Delta(A, B) = B - A \qquad (1)$$

be the difference series such as those shown in Figure 7. Then $\Delta(A, B)[m]$ is the value of the series at month m. To minimize noise in $\Delta(A, B)$, compute averages before and after a gap of length g months. The pregap average ends with month m. The postgap average begins with month $m + g + 1$. If t is the averaging period, denote the pregap average ending with month m as $\overline{\Delta(A, B)}^{t-}[m]$. Similarly, denote the postgap average beginning with month m as $\overline{\Delta(A, B)}^{t+}[m]$. Then the before-and-after difference is given by

$$\overline{\Delta}^t(\Delta(A, B), m, g) = \overline{\Delta(A, B)}^{t+}[m + g + 1] - \overline{\Delta(A, B)}^{t-}[m]. \qquad (2)$$

Series A is assumed to provide ground truth in the search for a break in B beginning at m. This type of double differencing was discussed in section 6 for $t = 36$ without the use of the present notation.

[41] The results of applying (2) to the difference series of Figure 7 are shown in Figures 9a and 9b ($g = 0$) and Figures 10a and 10b ($g = 12$) for m = December 1991 and $t = 24$ and 36 months. The leftmost pair of bars in these figures are differences derived only from the satellite data sets. These bars have the same magnitude but opposite sign in Figure 9a (A = RSS, B = UAH) and 9b (A = UAH, B = RSS). The same is true in Figures 10a and 10b. In Figures 9a and 10a, the comparisons of UAH and the sonde data, including HadAT2, show that in all cases the differences are not significantly different from zero at the 95-percent confidence level. Thus the sondes indicate that relative to UAH there is no significant break at this time. In Figures 9b and 10b, the comparisons of RSS and both sonde data sets show that every difference except the one for night sondes is significantly different from zero.

[42] The 36-month difference relative to the night sondes is likely affected by a processing change in the night sonde measurements that occurred about 24 months before the overlap between NOAA-11 and NOAA-12. Around January 1990, most VIZ stations changed ground station computers from Mini-Art to Micro-Art. *Christy and Norris* [2006] show, in both UAH and RSS LT comparisons, that on average these VIZ sondes experienced a relative tropospheric warming of +0.17 K precisely when the software change was implemented. This change affected at least 10 of the 30 night stations in our sample and thus introduced a spurious shift of about 0.06 K in the composite night sonde series at that time. (The change in computers also affects the results for the daytime sondes but to a smaller extent since only 10 of the 57 daytime stations used VIZ sondes.) Because of its timing, this shift would impact the 36-month average, but not the 24-month average. If the shift is removed from the

Period Differences: 12-Month Gap
Relative to UAH
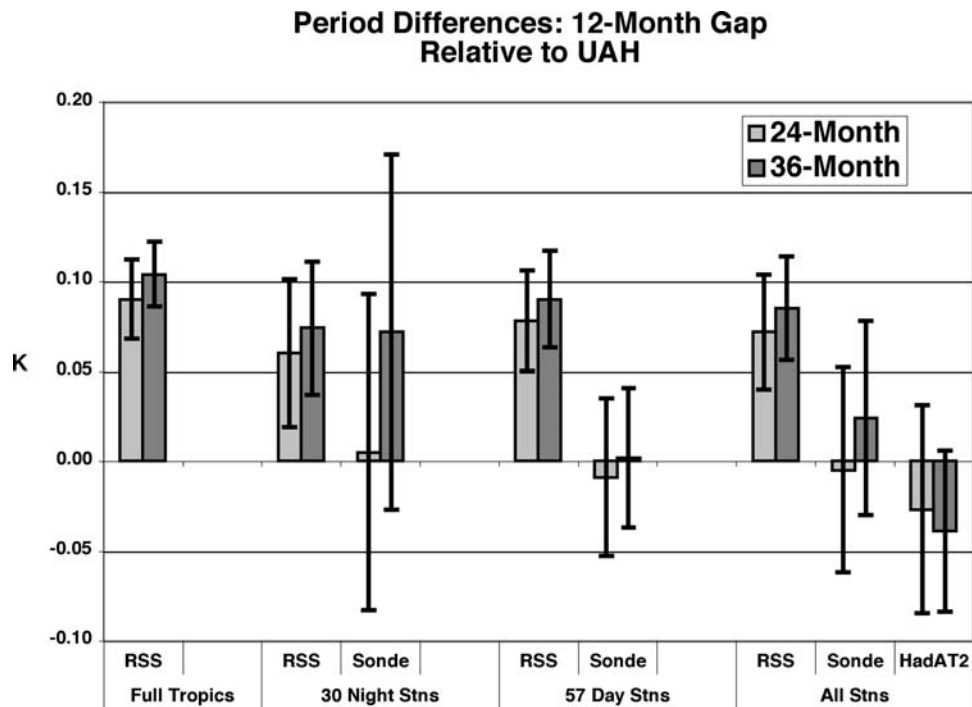
Figure 10a. As Figure 9a except that the "after" average begins with January 1993 instead of January 1992, thus creating a gap of 12 months between the "before" and "after" averages.

night sonde series, then the length of the 36-month difference bars for night sondes would become more negative in Figures 9 and 10, thus more nearly equaling the length of the 24-month bars. The sonde results therefore support the magnitude of the UAH satellite variations across January

1992. Also, HadAT2, with its own independent means of sonde adjustments, supports this result. Thus the evidence suggests that the difference in tropical temperatures across January 1992 for RSS is significantly different from that of the other data sets.

Period Differences: 12-Month Gap
Relative to RSS
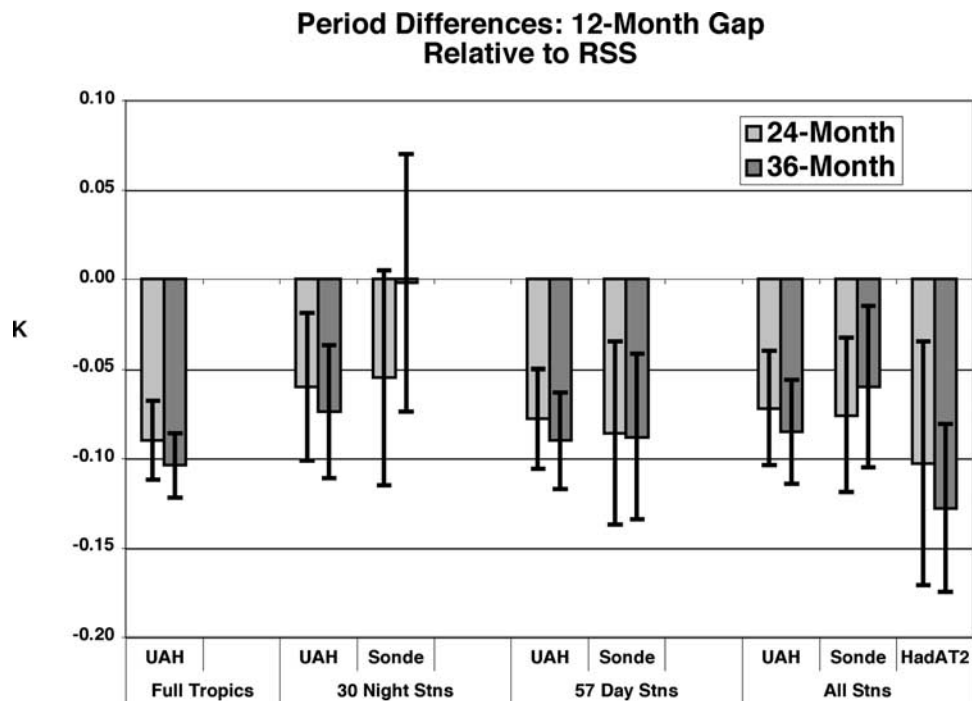
Figure 10b. As in Figure 9b except that the "after" average begins with January 1993 instead of January 1992.

Table 2. Comparison of UAH and RSS to Observed or Modeled Responses to the June 1991 Mt. Pinatubo Eruption[a]

(1) Data set	Difference, K, in Temperatures Before and After January 1992[b]		24-Month Averages		36-Month Averages	
	(2) 24-mon. avg.	(3) 36-mon. avg.	(4) |Col. 2 − UAH|	(5) |Col. 2 − RSS|	(6) |Col. 3 − UAH|	(7) |Col. 3 − RSS|
UAH[c]	−0.138	+0.025	—	—	—	—
RSS[c]	−0.051	+0.130	—	—	—	—
Models[d]	−0.470	−0.320	0.332	0.419	0.345	0.450
Sondes[c,e]	−0.167	−0.007	0.029	0.116	0.032	0.137
HadAT2[c]	−0.200	−0.044	0.062	0.149	0.069	0.174
Tropical land surface[f]	−0.102	−0.024	0.036	0.051	0.049	0.154
Tropical surface all[f]	−0.042	+0.025	0.096	0.009	0.000	0.105

[a]The purpose of this table is to compare the responses of the UAH and RSS data sets to the responses found in nonsatellite data sets. The responses themselves are listed in columns (2) and (3); the comparisons are listed in columns (4)–(7). Entries in which UAH and RSS would be compared are shown by "—". Columns (4)–(7) are absolute values of the differences between the responses of the nonsatellite data sets and the corresponding responses of UAH and RSS.

[b]Averaging periods of equal length (24 or 36 months) were taken on either side of January 1992. The earlier period ended with December 1991; the latter began with January 1992.

[c]LT products.

[d]Pressure-weighted averages of the layer 850–300 hPa composited from nine coupled climate model runs that included the effects of the eruption. The runs were made for the "Climate of the 20th Century" project [*Santer et al.*, 2005].

[e]A composite of the 87 unadjusted sonde series.

[f]Derived from HadCRUT3v, the gridded surface temperature data set described by *Brohan et al.* [2006]. Note that surface temperature residuals provide limited confidence in testing a tropospheric phenomenon as they may be affected by different processes and have different response characteristics.

[43] The averaging periods on either side of January 1992 represented in Figures 9 and 10 encompass a four to seven year span when few changes occurred in sonde instrumentation according to available metadata, except as noted above. We would not expect spurious shifts of more than 0.1 K in the 87-station composite series over these spans since the 95-percent confidence-level error bars in Figures 9 and 10 are only 0.04 to 0.05 K long for this series. The relative shift in RSS data for the no-gap comparison (Figure 9b) is from 0.07 to 0.11 K and is from 0.08 to 0.13 K in the 12-month gap comparison (Figure 10b). Thus the evidence supports the hypothesis that the RSS data set contains a spurious shift on the order of 0.1 K around 1992. Removing this from the RSS time series reduces the tropical LT trend by 0.04 K to 0.07 K decade^{-1}.

9.2.2. Post-Pinatubo Behavior

[44] Examining variability before and after January 1992 will incur the effects of the June 1991 eruption of Mt. Pinatubo. Besides the data sets already considered, nine coupled climate model runs for the "Climate of the 20th Century" project include these effects [*Santer et al.*, 2005]. A composite tropical average temperature for the layer 850–300 hPa (highly correlated with LT) was calculated from these runs. Also, tropical surface temperatures from the HadCRUT3v data set [*Brohan et al.*, 2006] were consulted. For each data set, we computed a 24-month average ending with December 1991 and a second 24-month average beginning with January 1992. The absolute difference of these two averages was taken as an indicator of the extent to which the effects of the eruption were observed in each data set. Similar computations were carried out for 36-month periods. The results are shown in Table 2. In every case but one (tropical surface, 24-month period), the UAH response, relative to the RSS response, is more consistent with the responses of the nonsatellite data. In other words, the values of columns (4) and (6) are closer to zero than are corresponding values in columns (5) and (7) except in one case.

[45] Neither UAH nor RSS agreed well with the model responses, and inferences from the models should not be overstated. The oceanic variability unique to each model and the relatively large magnitudes reported above are among the reasons that might render the comparisons inapplicable to our issue. At most we can say that under a variety of oceanic realizations, models indicate that the posteruption period should not be warmer than the preeruption period. From the relative cooling in the period following December 1991 suggested by the models, the sondes, and HadAT2, one might conclude that both UAH and RSS suffer from spurious warming in the posteruption period but with UAH suffering less so.

9.2.3. Further HadAT2 Comparisons

[46] The time series of differences between HadAT2 and the two satellite data sets (lower two time series of Figure 7) offer additional information. Above we compared HadAT2 tropical mean temperature to the satellite mean temperature as calculated at the 58 sonde station grids. If we use the full tropical mean of both satellite data sets in comparison with the HadAT2 full tropical mean (trend of +0.072 K decade^{-1}) the results are of interest.

[47] A statistical examination of the time series of HadAT2-UAH and HadAT2-RSS (analogous to the lower two times series in Figure 7) reveals two significant breakpoints. The first is in 1982 which for RSS (UAH) is +0.16 ± 0.05 (+0.10 ± 0.07) K (i.e., sondes become relatively warmer). This event is possibly related to the change from Philips Mark II to Mark III in the many Australian stations which may not have been fully captured by HadAT2. On the other hand, the uncorrected sondes do not show such a relative warming (2nd and 3rd time series of Figure 7) so the issue is uncertain.

[48] The second break, already noted when looking at satellite data on the 58 station grids, occurs in 1991 and relative to RSS (UAH) is −0.20 ± 0.06 (−0.08 ± 0.06) K. As indicated before, this appears to be largely related to a shift in RSS temperatures.

[49] These two shifts are of opposite sign and, if adjusted, would affect the HadAT2 mean trend by about +0.07 if

Table 3. Comparison of Topical Trends (1979–2004 Except Where Noted) in Terms of Differences Between the Data Set Identified and UAH LT

Data Set	Trend Difference Relative to UAH LT, K decade^{-1}	Data/Method	Reference
RATPAC	+0.03	Radiosondes	*Free et al.* [2005], *Lanzante et al.* [2003]
HadAT2	+0.02	Radiosondes	*Thorne et al.* [2005a]
NCEP/NCAR	+0.00	Reanalyses	*Kalnay et al.* [1996]
ERA-40	+0.00	Reanalyses (1979–2001)	*Uppala et al.* [2005]
JRA-25	−0.02	Reanalyses (1979–2002)	*Sakamoto et al.* [2006]
ECMWF	−0.03	Radiosondes + forecast model	*Haimberger* [2005]
RSS	+0.10	Satellite	*Mears and Wentz* [2005]

adjusted according to RSS and +0.02 K decade^{-1} if adjusted according to UAH. On the other hand, we might say HadAT2 has detected two significant errors in the satellite data sets which would alter them by the same magnitude but of the opposite sign. It is probable that the true trend lies somewhere in the midst of these realizations.

10. Other Data Sets

[50] *Christy et al.* [2003] determined that the 95-percent confidence interval of the UAH LT global mean trend was ±0.05 K decade^{-1}. We estimate from this that the corresponding error for the tropics is ±0.07 K decade^{-1}. Table 3 shows tropical lower-tropospheric trends produced from radiosonde and reanalyzes data sets which were developed with significantly different methodologies. Some used model assimilation or model guidance with satellite-based retrievals, and others used statistical techniques applied only to sondes. These data sets produce tropical tropospheric trends similar to that of UAH and differ by small amounts of either sign, thus being consistent with the stated UAH error estimate. The trend of the RSS LT data set for the full tropics differs from that of the UAH trend by 0.10 K decade^{-1} and is the only data set with a trend difference outside of UAH's ±0.07 K decade^{-1} error estimate.

[51] The additional data sets in Table 3 supply evidence consistent with the hypothesis that the RSS data may contain a spurious shift around 1992. If this hypothesis is true, the UAH error range could be reduced. Otherwise, the UAH error range should be larger, and significant, spurious features in the sonde records, and in the reanalyzes which partially depend on the sondes, need to be discovered. The implication from the evidence presented here is that the UAH tropical error range of ±0.07 K decade^{-1} is appropriate at this time.

11. Caveats

[52] We point out that data sets based on satellites undergo constant examination by the developers and users. These data are observed by complicated instruments which measure the intensity of the emissions of microwaves from atmospheric oxygen, requiring physical relationships to be applied to the raw satellite data to produce a temperature value. Further, the program under which these satellites were designed and operated was intended to improve weather forecasts, not to generate precise, long-term climate records.

[53] Since 1992 the UAH LT data set has been revised seven times or about once every 2 to 3 years. There is no

expectation that the current version (5.2, May 2005) will not continue to be revised similarly as better ways to account for known biases are developed and/or new biases are discovered and corrected. Thus the production of climate time series from satellites will continue to be a work-in-progress.

[54] Usually, developers of data sets underestimate the measurement error ranges of their products [*Morgan*, 1990]. To this point, we have relied on various assessments of radiosonde versus satellite differences at the station level and in multistation aggregates to aid in determining reasonable error ranges [e.g., *Christy and Norris*, 2006]. The magnitude of our trend error estimate is consistent with essentially all other trends (except RSS), but we recognize this does not provide the highest level of confidence as we lack absolute standards for error assessment; that is, the degree of veracity of the comparison data sets has not been established [*Thorne et al.*, 2005b].

[55] It may well be that significant and pervasive negative biases through time, such as suggested by *Sherwood et al.* [2005] and *Randel and Wu* [2006], afflict the radiosonde time series. In addition, the character of these biases may be too subtle and too difficult to completely remove. This would create a coincidence of agreement between data sets whose trends were spuriously negative and therefore false confidence.

[56] We have and will continue to examine various families of radiosondes to document inhomogeneities which create problems for time series analysis. To date, using a number of tools, we have discovered both positive and negative biases in many types of radiosondes [*Christy and Norris*, 2004, 2006]. As noted here, many shifts appear to be spuriously negative, but there are also many, including some of the largest in magnitude, which appear to be spuriously positive. Thus in total these would seem to have a relatively small impact on lower-tropospheric trends of large-scale averages. Given the results of the current versions of the data sets and experiments presented here, we see that all (except RSS and one RSS-adjusted sonde experiment) indicate trends for the tropical lower tropo-sphere that are less than that of the surface (+0.125 K decade^{-1}). This yields trend ratios of troposphere versus surface of less than 1.0, which is smaller than the ratio of 1.3 generated from climate model simulations for this time period.

12. Summary and Conclusion

[57] The individual LT trends (1979–2004) of 58 tropical sonde stations ranges from about −0.5 to +0.6 K decade^{-1} when considering day and night releases as separate data

sources (87 in total). The median trend of the 87 sonde time series is +0.073 K decade^{-1}. At the same locations (and double counting satellite grids where both day and night sonde releases are used) the median trend of the corresponding 87 time series for UAH satellite data is +0.090 K decade^{-1} and of the RSS data, +0.146 K decade^{-1}.

[58] Comparisons with satellite data suggested that the aggregate trends of the nighttime sonde observations were too positive and of the daytime, too negative. This was determined by identifying and adjusting for the largest sonde discontinuities through comparisons with the UAH and RSS data sets. When the largest breaks were removed, the trends of the UAH- and RSS-adjusted sondes converged to +0.092 and +0.118 K decade^{-1}. When the breakpoint criterion was tightened to force greater consistency with the satellite data sets, the trend of the UAH-adjusted sondes hardly changed: +0.091 K decade^{-1}, and the trend of the RSS-adjusted sondes increased to +0.136 K decade^{-1}. In all cases the RMS differences of the individual trend comparisons and the median of individual station correlations indicated greater consistency with the UAH- rather than the RSS-adjusted sondes.

[59] A key difference between the UAH and RSS data sets occurred around January 1992 when a significant positive shift occurred in the RSS data relative to UAH. This date coincides with the inclusion of data from the newly launched NOAA-12 satellite and the latter part of NOAA-11's time series when large corrections needed to be applied. Further comparisons with sonde and other data sets between the periods before and after January 1992 show consistency with the UAH data but a relative positive shift in the RSS data of 0.07–0.13 K. The upward shift in the RSS data relative to UAH and the other data sets cannot be explained by potential discontinuities in those data sets at this time. We speculate that the upward shift in RSS data likely relates to warming due to corrections applied to NOAA-11. Overall, the results presented here indicate consistency with the estimated UAH LT trend of +0.052 ± 0.07 K decade^{-1} for the entire tropics. With a corresponding surface trend of +0.125 K decade^{-1}, the ratios of the present versions of UAH, sonde and reanalyzes tropospheric warming trends versus the surface trend are less than 1.0 while for RSS the ratio is 1.2.

References

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.*, *111*, D12106, doi:10.1029/2005JD006548.

Christy, J. R., and W. B. Norris (2004), What may we conclude about global tropospheric temperature trends?, *Geophys. Res. Lett.*, *31*, L06211, doi:10.1029/2003GL019361.

Christy, J. R., and W. B. Norris (2006), Satellite and VIZ-radiosonde intercomparisons for diagnosis on non-climatic influences, *J. Atmos. Oceanic Technol.*, *23*, 1181–1194.

Christy, J. R., and R. W. Spencer (2005), Correcting temperature data sets, *Science*, *310*, 972.

Christy, J. R., R. W. Spencer, and E. S. Lobl (1998), Analysis of the merging procedure for the MSU daily temperature time series, *J. Clim.*, *11*, 2016–2041.

Christy, J. R., R. W. Spencer, W. B. Norris, W. D. Braswell, and D. E. Parker (2003), Error estimates of Version 5.0 of MSU/AMSU bulk atmospheric temperatures, *J. Atmos. Oceanic Technol.*, *20*, 613–629.

Durre, I., R. S. Vose, and D. B. Wuertz (2005), Overview of the integrated global radiosonde archive, *J. Clim.*, *19*, 53–68.

Elliott, W. P., R. J. Ross, and W. H. Blackmore (2002), Recent changes in NWS upper-air observations with emphasis on changes from VIZ to Vaisala radiosondes, *Bull. Am. Meteorol. Soc.*, *83*, 1003–1017.

Folland, C. K., et al. (2001), Observed climate variability and change, in *Climate Change 2001: The Scientific Basis*, edited by Houghton et al., pp. 99–181, Cambridge Univ. Press, New York.

Free, M., 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 data set of large-area anomaly time series, *J. Geophys. Res.*, *110*, D22101, doi:10.1029/2005JD006169.

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

Gaffen, D. J. (1996), A digitized metadata set of global upper-air station histories, *NOAA Tech. Memo., ERL ARL-211*, 38 pp., Natl. Oceanic Atmos. Admin., Silver Spring, Md.

Gaffen, D. J., M. A. Sargent, R. E. Habermann, and J. R. Lanzante (2000), Sensitivity of tropospheric and stratospheric temperature trends to radiosonde data quality, *J. Clim.*, *13*, 1776–1796.

Grody, N. C., K. Y. Vinnikov, M. D. Goldberg, J. T. Sullivan, and J. D. Tarpley (2004), Calibration of multisatellite observations for climatic studies: Microwave Sounding Unit (MSU), *J. Geophys. Res.*, *109*, D24104, doi:10.1029/2004JD005079.

Haimberger, L. (2005), Homogenization of radiosonde temperature time series using ERA-40 analysis feedback information, *ERA-40 Proj. Rep. Ser., 23*, Eur. Cent. for Medium Range Weather Forecasts, Shinfield Park, Reading, UK.

Hurrell, J. W., and K. E. Trenberth (1996), Satellite versus surface estimates of air temperature since 1979, *J. Clim.*, *9*, 2222–2232.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-year Reanalysis Project, *Bull. Am. Meteorol. Soc.*, *77*, 437–471.

Lanzante, J. R., S. A. Klein, and D. J. Seidel (2003), Temporal homogenization of monthly radiosonde temperature data. part I: Methodology, *J. Clim.*, *16*, 224–240.

Mears, C. A., and F. W. Wentz (2005), The effect of diurnal correction on satellite-derived lower tropospheric temperature, *Science*, *11*, *309*, 1548–1551, doi:10.1126/science.1114772.

Mears, C. A., M. C. Schnabel, and F. J. Wentz (2003), A reanalysis of the MSU Channel 2 tropospheric temperature record, *J. Clim.*, *16*, 3650–3664.

Morgan, M. G. (1990), Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis, 332 pp., Cambridge Univ. Press, New York.

Parker, D. E., 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.

Prabhakara, C., R. Iacovazzi Jr., J.-M. Yoo, and G. Dalu (2000), Global warming: Evidence from satellite observations, *Geophys. Res. Lett.*, *27*, 3517–3520.

Randel, W. J., and Fei Wu (2006), Biases in stratospheric and tropospheric temperature trends derived from historical radiosonde data, *J. Clim.*, *19*, 2094–2104.

Sakamoto, M., K. Okanmoto, Y. Yoshizaki, H. Murata, S. Kobayashi, J. Tsutsui, S. Kadokura, K. Wada, and H. Hatsushika (2006), Reanalysis impact on quality controls of long-term satellite sounding observation paper presented at the 14th Conference on Satellite Meteorology and Oceanography, 86th AMS Annual Meeting, Am. Meteorol. Soc., Atlanta, Ga., 29 Jan. to 2 Feb.

Santer, B. D., et al. (2005), Amplification of surface temperature trends and variability in the tropical atmosphere, *Science*, *309*, 1551–1556, doi:10.1126/science.1114867.

Seidel, D. J., and J. R. Lanzante (2004), An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes, *J. Geophys. Res.*, *109*, D14108, doi:10.1029/2003JD004414.

Seidel, D. J., et al. (2004), Uncertainty in signals of large-scale climate variations in radiosonde and satellite upper-air temperature data sets, *J. Clim.*, *17*, 2225–2240.

Sherwood, S. C., J. Lanzante, and C. Meyer (2005), Radiosonde daytime biases and late 20th century warming, *Science*, *309*, 1556–1559, doi:10.1126/science.1115640.

Spencer, R. W., and J. R. Christy (1992), Precision and radiosonde valida-tion of satellite gridpoint temperature anomalies, part II: A tropospheric retrieval and trends during 1979–90, *J. Clim.*, *5*, 858–866.

Spencer, R. W., J. R. Christy, and N. C. Grody (1990), Global atmospheric temperature monitoring with satellite microwave measurements: Method and results 1979–1985, *J. Clim.*, *3*, 1111–1128.

Spencer, R. W., J. R. Christy, W. D. Braswell, and W. B. Norris (2006), Estimation of tropospheric temperature trends from MSU channels 2 and 4, *J. Atmos. Oceanic Technol.*, *23*, 417–423.

Thorne, P. W., D. E. Parker, S. F. B. Tett, P. D. Jones, M. McCarthy, H. Coleman, P. Brohan, and J. R. Knight (2005a), Revisiting radio-sonde upper-air temperatures from 1958 to 2002, *J. Geophys. Res.*, *110*, D18105, doi:10.1029/2004JD005753.

Thorne, P. W., D. E. Parker, J. R. Christy, and C. A. Mears (2005b), Uncertainties in climate trends: Lessons from upper-air temperature records, *Bull. Am. Meteorol. Soc.*, *86*, 1437–1442.

Uppala, S. M., et al. (2005), The ERA-40 reanalyses, *Q. J. R. Meteorol. Soc.*, *131*, 2961–3012.

————————————

J. R. Christy, W. B. Norris, and R.W. Spencer, ESSC/Cramer Hall, University of Alabama in Huntsville, Huntsville, AL 35899, USA.

J. J. Hnilo, PCMDI/CAPT, Lawrence Livermore National Laboratory, M.S. L-103, 7000 East Avenue, Livermore, CA 94550, USA.