

What may we conclude about global tropospheric temperature trends?

John R. Christy and William B. Norris

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

Received 23 December 2003; revised 13 February 2004; accepted 3 March 2004; published 31 March 2004.

[1] Three realizations of the atmospheric temperature representing the layer from the surface to about 18 km generated from microwave emissions were published in 2003. Their 1979–2002 linear trends were stated as $+0.24 \pm 0.02$, $+0.12 \pm 0.02$ and $+0.03 \pm 0.05^\circ\text{C decade}^{-1}$. Because the upper portion of this layer includes the stratosphere, the opportunity to utilize radiosonde measurements as an independent assessor of these trend values is diminished. However, the University of Alabama in Huntsville (UAH) also produces a lower tropospheric temperature (LT) product (0–8 km, trend of $+0.08^\circ\text{C decade}^{-1}$) for which radiosondes are much more suitable. Comparisons of this UAH product with radiosonde-simulated layer temperatures show no significant difference in LT trends, lending support for the least positive trend of the three deeper layer values ($+0.03 \pm 0.05^\circ\text{C decade}^{-1}$) as it was constructed in the same manner as LT. **INDEX TERMS:** 0394 Atmospheric Composition and Structure: Instruments and techniques; 1610 Global Change: Atmosphere (0315, 0325); 1620 Global Change: Climate dynamics (3309); 1640 Global Change: Remote sensing. **Citation:** Christy, J. R., and W. B. Norris (2004), What may we conclude about global tropospheric temperature trends?, *Geophys. Res. Lett.*, 31, L06621, doi:10.1029/2003GL019361.

1. Introduction

[2] In 2003, two new versions of microwave-based temperatures as well as an updated version of an older dataset produced by the University of Alabama in Huntsville (UAH) were published [Vinnikov and Grody, 2003; Mears *et al.*, 2003; and Christy *et al.*, 2003]. We shall refer to the datasets of Vinnikov and Grody as VG, of Mears *et al.* (from Remote Sensing Systems) as RSS and of Christy *et al.* as UAH. One temperature product common to all three groups is called the mid-troposphere (hereafter MT) because the peak of the emissions occurs around 5 km. However, because of its very broad vertical extent, about 15% of the emissions actually originate from the lower stratosphere.

[3] All three groups begin their processing stream with the same basic satellite data observed by NOAA polar orbiting satellites. However, when the metric of trend by least squares regression is applied, the results are significantly different. For 1979–2002, the published trends with 95% error ranges in $^\circ\text{C decade}^{-1}$ are $+0.24 \pm 0.02$, $+0.12 \pm 0.02$ and $+0.03 \pm 0.05$ for VG, RSS and UAH respectively (Figure 1). Only UAH developed their error estimates in two ways (a) by applying error ranges for each process in the construction sequence and (b) comparisons with northern hemisphere radiosonde data.

[4] These disparate trend values create uncertainty for readers who try to interpret long term trends over the past 24 years. The goal of this study is to utilize independent radiosonde data to come to at least some defensible conclusion about long term trends for one of these datasets. (Note that we use “trend” as a metric to discriminate datasets, not as a quantity applied to understand the physical climate system.)

2. The Problem

[5] The reasons for these trend differences are fairly well understood as they arise from the differing methods used to account for errors before merging the time series of eleven satellites into a single, homogeneous time series. The reader is encouraged to consult the cited publications for background on the issue of microwave-based thermometry from space. The details of the differences have been and will be addressed in other publications and only briefly mentioned here.

[6] RSS and UAH differ principally in two ways. First, the source of the largest difference concerns the adjustment for the calibration shift which occurs as the sensor itself is exposed to changing temperature. The changing conditions are related to the east-west drift of the spacecraft from their initial meridional, quasi-sun-synchronous orbital paths. The changing exposures to sunlight and shadows during the slow drift (about 45° longitude in 4 years) cause differential heating and cooling of the components. This effect upon measured temperature was discovered when the global mean temperatures of two co-orbiting satellites displayed differences that were highly correlated with the temperature of the instrument’s hot target plate.

[7] This issue may relate to the pre-launch laboratory tests which define the non-linear coefficient which accounts for the slight non-linearity in the relationship between the actual reference temperature and the temperature reported by the instrument [Mo *et al.*, 2001]. There is uncertainty in this pre-launch equation, but the physical basis for explaining a possible change in the equation after the instrument is launched into space is at present unknown. The cause may be related to something other than an error in the calibration equation, so further research is needed. However, that there is a problem, related in some way to the instrument temperature, is clear from the intercomparison of data from co-orbiting spacecraft.

[8] Christy *et al.* [2000] developed an empirically-based technique which determines a linear coefficient which when multiplied by the temperature of the sensor itself corrects for this problem. (The “sensor temperature” is the temperature of the hot target plate. Other sensor temperatures were less correlated.) Their technique requires heavy smoothing of the antecedent data to produce robust coefficients. However,

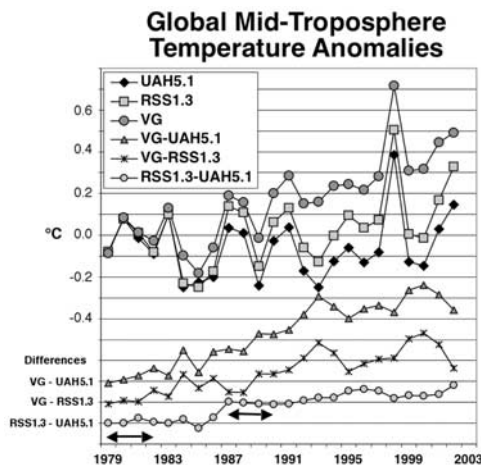


Figure 1. Annual anomalies of global mean, mid-tropospheric temperature from three versions based on microwave emissions for 1979 to 2002 relative to the mean of 1979–1980. Paired differences are shown below. Trend values are $+0.24$, $+0.12$ and $+0.03^{\circ}\text{C decade}^{-1}$ for VG, RSS and UAH respectively. The two arrows delineate the two periods which are intercompared with radiosonde data to assess the NOAA-9 period.

if the antecedent data do not correlate well with the temperature errors, no corrective coefficient is applied.

[9] RSS uses a similar technique, however they do not smooth the antecedent data (other than pentad averages) and determine coefficients for all satellites without a constraint based on effectiveness or magnitude of error reduction. These two different approaches lead to a significantly different correction coefficient for one satellite in particular - NOAA-9. NOAA-9 happens to form a critical link between pre-1984 and post-1986 data. Thus, the correction applied to NOAA-9 determines the relative relationship between the early and later data, and therefore has a significant impact on the metric of trend.

[10] The second main reason for RSS and UAH differences deals with the way the eleven satellite records are patched together. UAH uses a “backbone” technique in which there is a unique, single path on which the basic structure of the time series depends. The reasoning is that a single, best path is chosen from one satellite to the next which contains the lowest error magnitude [Christy *et al.*, 1998]. RSS uses a “unified” or “consensus” approach in which a statistically best path is found which links all of the satellites together using all of the overlapping satellite data available [Mears *et al.*, 2003]. The difference is seen in the way the time series is built from NOAA-6 to NOAA-9 (1979 to 1987). UAH uses a direct path from NOAA-6 to NOAA-9, so that a single bias between these two is calculated and removed. However, there were also brief overlaps between NOAA-9 and both NOAA-7 and NOAA-8. RSS incorporates two additional paths to NOAA-9 (1) NOAA-6 to NOAA-7, then NOAA-7 to NOAA-9 and (2) NOAA-6 to NOAA-7, NOAA-7 to NOAA-8 then NOAA-8 to NOAA-9 as well as (3) NOAA-6 to NOAA-9 directly. It becomes a matter of choice as to whether one has confidence in a single direct path (UAH) or in the consensus of multiple, but more noisy paths (RSS). All in all, however, the general techniques of UAH and RSS are quite similar.

[11] The differences between VG and the other two are more profound and structural. VG assume all error due to satellite drift may be directly projected onto a 24-hour diurnal cycle resolved by only two harmonics. No effort was made to characterize the temperature-dependent calibration shifts which were independently verified in UAH and RSS publications and which varied from satellite to satellite. Apparently, VG desired a more physical explanation of the observed intersatellite differences before accepting their variations as being separate from the diurnal cycle. Too, the disagreement between UAH and RSS on the NOAA-9 coefficient also raised a question as to the ability to quantify the calibration shift.

[12] Whereas UAH and RSS find the calibration shift as a strong function of sensor temperature, VG require these errors to be a function of hour-of-day in a single diurnal cycle adjustment which then is applied systematically to all satellites. Thus, a spurious warming or cooling in a single satellite, which is then removed individually in RSS and UAH data, will be accepted as a real expression of the diurnal cycle and removed from all satellites in the VG methodology. This leads to significant differences in the time series and a diurnal cycle of temperature in which there is a maximum in local temperature at 11:30 a.m., a decline to a relative minimum at 3:30 p.m., then a rise to another relative maximum at 9:00 p.m. [Figure 4b, Vinnikov and Grody, 2003]. We are aware of no observations [e.g. Christy *et al.* 2003] or modeling results [e.g., Mears *et al.*, 2003] which corroborate such a finding.

3. Method

[13] Is there a path forward given the structural and philosophical differences in merging methods? One way to check these results is to develop layer temperatures from completely independent data and compare. Radiosonde data provide the possibility of such a check because the discrete levels of information they provide may be weighted in such a fashion as to produce a simulated microwave temperature of the layer in question. (We use the static weighting function method to generate the simulated satellite temperatures from radiosondes, Spencer and Christy [1992].) The effectiveness of such a test is diminished in this case because the layer of interest, MT, requires information from at least 50 hPa and preferably higher. Numerous studies report that radiosonde data become less reliable for time series construction as the altitude increases [Parker *et al.*, 1997; Gaffen *et al.*, 2000; Lanzante *et al.*, 2003]. In particular as new radiosonde instrumentation and new correction tables are brought into the mix over time, the adjustments required at altitudes 100 hPa and above tend to be quite large, due to corrections for solar heating, lag etc. In the lower stratosphere, shifts of 1 to 3°C were evident for many stations when they switched manufacturers to Vaisala RS-80 from Philips or VIZ-B [Parker *et al.*, 1997; Christy *et al.*, 2003]. However, for altitudes below 200 hPa, the temperature shifts were much smaller, thus rendering these lower levels more susceptible to time series construction. Because the stratospheric levels require corrections that are larger than the signals being sought, we defer such comparisons for other researchers [e.g., Lanzante *et al.*, 2003].

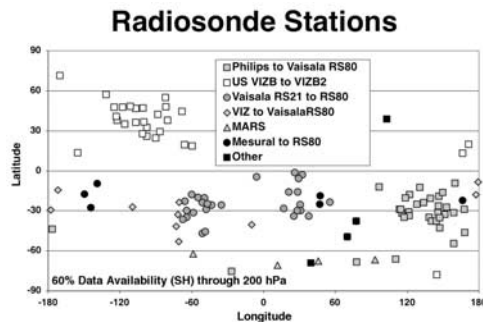


Figure 2. Location of stations and their instrument type used in this study.

[14] We have one path to explore here. UAH also produces a lower tropospheric product (LT) which is easily simulated from radiosonde temperatures from the surface to 200 hPa. This product avoids the upper altitudes where correction tables and changes in manufactures have considerable effect. Too, UAH LT is generated with the identical merging pathway as is UAH MT. For the 25-year period Dec 1978 to Nov 2003, the trends for MT and LT respectively were $+0.04$ and $+0.08^{\circ}\text{C decade}^{-1}$ the latter being more positive due to the absence of stratospheric influences affecting the former.

[15] *Christy et al.* [2003] utilized data from 29 stations in the NH which were characterized by instrument consistency. The stations were distributed from the western tropical Pacific to northern Alaska. From these and other comparisons, it was determined that the 95% confidence interval for the metric of trend in LT and MT was $\pm 0.05^{\circ}\text{C decade}^{-1}$.

[16] In comparing RSS and UAH, we find a slightly larger trend discrepancy in the SH than the NH. Therefore, if UAH data are subject to problems of which RSS data are not, then the discrepancy should appear most readily in the SH data. (VG gridded data are unavailable.) Because the adjustments for diurnal drift and intersatellite bias are independent of latitude in the UAH data, the following will be a new and different test than that of *Christy et al.* [2003] for which only 29 NH locations were used.

[17] We emphasize the independence of the radiosonde and UAH data. Only after the current version (5.1) of the UAH data was finalized, published and made available on the web, did we contact the National Climatic Data Center for the radiosonde records of the SH stations.

[18] We discovered there were over 300 SH stations, but quickly found many unsuitable. Two thresholds were established to provide a reasonable dataset. We selected those sites for which at least 60% of the months could be generated from daily soundings, calling this dataset *A*. A subset of *A* for which at least 75% of the months could be constructed we termed dataset *B*. We constructed monthly anomalies for 89 stations in *A* and 72 in *B* covering the 271-month period Jan 1979 to Jul 2001 (Figure 2). In a few analyses below we will add the 29 NH radiosonde stations (called *C*) from *Christy et al.* [2003].

[19] The stations making up the two SH datasets reflect a variety of instrument designs and instrument changes over the 1979 to 2001 period. While this presents a challenge in terms of determining proper adjustments, it also provides a sampling effect of disallowing a single radiosonde type or change to dominate the results. As a consequence, this

offers a condition of randomness to the errors that is important for composite analysis. This, combined with the fact tropospheric temperature shifts due to radiosonde instrumental or procedural changes are small, grants us at least the opportunity to proceed.

[20] We classify the radiosonde data into two categories, unadjusted and adjusted. The unadjusted data are, as named, simply simulated microwave temperatures produced from the original radiosonde soundings. Adjusted data refer to the same stations for which a single adjustment was calculated and applied at the point in time when a change in instrumentation was recorded and entered into metadata files.

[21] To determine the adjustments for changing instrumentation, a time series of differences between the radiosonde and the corresponding UAH gridbox temperature was generated. For stations with a common type of instrument change, the difference time series were aligned so that the change point was common to all and the difference-time series composited. A difference between 36 month-averages prior to and subsequent to the change point was determined and then applied systematically to each appropriate station according to its individual change date. We also checked the results of *Durre et al.* [2002] who used the same technique but for NH stations, and of P. Thorne (Hadley Centre) who developed a correction based only on neighboring radiosondes.

[22] In Table 1 we show that the actual adjustments applied to the radiosonde data were not necessarily those calculated in the course of this present study. Because the Philips to Vaisala RS80 was potentially the largest, we selected the value determined by P. Thorne which is independent of any satellite data, though almost identical to that so calculated. Other selections were based on the largest number of observations. The Mesural change was only marginally significant because at least two other breakpoints of similar magnitude were found (one positive, one negative), so a correction was not applied. In the VIZ-B to VIZ-B2 correction, we found that the value for MT was identical whether using UAH or RSS data as the basis, so we have confidence that the value calculated here for LT is appropriate. For the Meisei70 to Meisei80 change, the difference was not significant. In any case, the choices of adjustment magnitude among the entries here turned out to be of little consequence to the outcome of the comparisons.

4. Results

[23] We first test the hypothesis that there may be a spurious shift in UAH data during the NOAA-9 era where

Table 1. Adjustments for Changes in Radiosonde Instrumentation ($^{\circ}\text{C}$)

Inst. Change	Applied Here	This study	Durre et al.	Thorne
P-V80	-0.27	-0.28 (29)	-0.18 (1)	-0.27 (11)
V21-V80	+0.24	+0.24 (28)	+0.22 (3)	
VZB-V80	+0.24	+0.11 (10)	+0.24 (14)	
M-V80	0	-0.11 (8)		
VZB-VZB2 NH	+0.16	+0.16 (28)*		
Me70-Me80	0	-0.05 (1)		

P = Philips, V21 = Vaisala RS21, V80 = Vaisala RS80, VZB = VIZ-B, VZB2 = VIZ-B2, M = Mesural, Me70 = Meisei70, Me80 = Meisei80. The values applied in this study are those in the "Applied Here" column while those calculated in this study are in "This Study". The number of stations used for each calculation is in parenthesis.

**Christy et al.* [2003].

Table 2. Trends of the Differences in the Composited Time Series of Monthly Anomalies for the Datasets Defined in the Text (1979–2001, °C decade⁻¹) With 95% Confidence Intervals

Grid set (no. of sites)	Trend Δ Sonde minus UAH LT	Trend Δ RSS minus UAH MT
A Adj (89)	+0.01 ± 0.04	+0.08 ± 0.02
A Unadj (89)	+0.02 ± 0.04	+0.08 ± 0.02
B Adj (72)	+0.00 ± 0.04	+0.09 ± 0.02
B Unadj (72)	+0.01 ± 0.04	+0.09 ± 0.02
C Adj (29)	+0.01 ± 0.03	+0.07 ± 0.03
A + C Adj (118)	+0.01 ± 0.03	+0.08 ± 0.02
B + C Adj (101)	+0.00 ± 0.03	+0.08 ± 0.02

The number of stations is given in parentheses.

the greatest difference between UAH and RSS occurs. For this, we select a suitable metric as the difference between the four years after and before NOAA-9's influence (i.e., average of 1987–90 minus average of 1979–1982, arrows Figure 1). Because missing data would have a greater chance to influence the outcome, we limit this test to sets *B* and *B + C*. When comparing MT at these stations, this metric produces a difference between UAH and RSS of 0.08 ± 0.03 and 0.07 ± 0.03 °C for *B* and *B + C* grids respectively. Comparing LT differences for UAH vs. radiosondes we have 0.01 ± 0.06 and 0.02 ± 0.05 for *B* and *B + C* respectively. Thus, for product MT, UAH and RSS are significantly different for this metric, while for product LT, UAH and adjusted radiosondes are not significantly different. This suggests there is no apparent problem with UAH data over the NOAA-9 period.

[24] We next test the hypothesis that the 23-year trends are different between radiosondes and UAH. In Table 2 we have the composite results, with errors based on the estimated degrees of freedom accounting for autocorrelation of errors. Note that there is no significant difference in the results whether using stations with 60% or 75% of data present or whether the stations are adjusted or unadjusted. We conclude that there is no significant difference in LT trends between the radiosonde data and UAH. We also conclude there is a significant difference between UAH and RSS MT data for composite trends at those same gridpoints. However, because of the aforementioned stratospheric problems with radiosondes we cannot deliver a conclusion here regarding UAH or RSS MT data other than that stated above. (However, we note in *Christy et al.* [2003], the 29 NH stations do supply information for MT comparisons.) Though VG gridded data were unavailable, it is clear that the differences between VG and both UAH and RSS would be significant for any set of grids, being on the order of 0.20 °C decade⁻¹ vs. UAH.

5. Conclusion

[25] What we have not been able to do is to compare the mid-tropospheric (MT) temperature trends of three

recently published microwave-based products against that of an independent source which has a reasonable level of accuracy. At this point, the stratospheric levels of radiosonde soundings are sufficiently uncertain that we leave the issue for other studies [e.g., *Seidel et al.*, 2004]. However, we have shown that the UAH lower tropospheric (LT) data are highly consistent with the more robust lower elevation radiosonde data. These results support the conclusion of *Christy et al.* [2003] that for Dec 1978 to Nov 2003 (25 years) the global trend in LT is $+0.08 \pm 0.05$ °C decade⁻¹. It is likely that the same precision may be applied to UAH MT data as the procedures followed to produce MT are the same as those which produce LT [*Christy et al.*, 2003].

[26] **Acknowledgments.** This work was supported by the U.S. Department of Transportation (DTFH61-99-X-00040) and the National Oceanic and Atmospheric Administration (NA96GP0386). We express gratitude to I. Durre of NCDC for making the radiosonde data available and P. Thorne of the Hadley Centre for providing independent estimates of temperature changes due to radiosonde instrumentation changes.

References

- Christy, J. R., R. W. Spencer, and W. D. Braswell (2000), MSU Tropospheric temperatures: Data set construction and radiosonde comparisons, *J. Atmos. Oceanic Tech.*, *17*, 1153–1170.
- Christy, J. R., R. W. Spencer, and E. 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 Tech.*, *20*, 613–629.
- Durre, I., T. C. Peterson, and R. S. Vose (2002), Evaluation of the effect of the Luers-Eskridge radiation adjustments on radiosonde temperature homogeneity, *J. Clim.*, *15*, 1335–1347.
- 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.
- 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. Clim.*, *16*, 241–262.
- Mears, C. A., M. C. Schabel, and F. J. Wentz (2003), A reanalysis of the MSU channel 2 tropospheric temperature record, *J. Clim.*, *16*, 3650–3664.
- Mo, T., M. D. Goldberg, D. S. Crosby, and Z. Cheng (2001), Recalibration of the NOAA microwave sounding unit, *J. Geophys. Res.*, *106*(D10), 10,145–10,150.
- 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*(12), 1499–1502.
- Seidel, D. J., et al. (2004), Uncertainty in signals of large-scale climate variations in radiosonde and satellite upper-air temperature datasets, *J. Clim.*, *17*, in press.
- Spencer, R. W., and J. R. Christy (1992), Precision and radiosonde validation of satellite gridpoint temperature anomalies, Part II: A tropospheric retrieval and trends during 1979–90, *J. Clim.*, *5*, 858–866.
- Vinnikov, K. Y., and N. C. Grody (2003), Global warming trend of mean tropospheric temperature observed by satellites, *Science*, *302*, 269–272.

J. R. Christy and W. B. Norris, ESSC/NSSTC, University of Alabama in Huntsville, Huntsville, AL 35899, USA. (christy@nsstc.uah.edu; norris@nsstc.uah.edu)