# MSU Tropospheric Temperatures: Dataset Construction and Radiosonde Comparisons

JOHN R. CHRISTY

*ESSC/GHCC, University of Alabama in Huntsville, Huntsville, Alabama*

ROY W. SPENCER

*NASA Marshall Space Flight Center, Huntsville, Alabama*

WILLIAM D. BRASWELL

*Nichols Research Corporation, Huntsville, Alabama*

(Manuscript received 14 May 1999, in final form 13 October 1999)

ABSTRACT

Two deep-layer tropospheric temperature products, one for the lower troposphere ($T_{2LT}$) and one for the midtroposphere ($T_2$, which includes some stratospheric emissions), are based on the observations of channel 2 of the microwave sounding unit on National Oceanic and Atmospheric Administration (NOAA) polar-orbiting satellites. Revisions to version C of these datasets have been explicitly applied to account for the effects of orbit decay (loss of satellite altitude) and orbit drift (east–west movement). Orbit decay introduces an artificial cooling in $T_{2LT}$, while the effects of orbit drift introduce artificial warming in both $T_{2LT}$ and $T_2$. The key issues for orbit drift are 1) accounting for the diurnal cycle and 2) the adjustment needed to correct for spurious effects related to the temperature of the instrument. In addition, new calibration coefficients for *NOAA-12* have been applied. The net global effect of these revisions (version D) is small, having little impact on the year-to-year anomalies. The change in global trends from C to D for 1979–98 for $T_{2LT}$ is an increase from +0.03 to +0.06 K decade$^{-1}$, and a decrease for $T_2$ from +0.08 to +0.04 K decade$^{-1}$.

## 1. Introduction

Scientists face many challenges when attempting to produce data with long-term stability from sequentially launched, polar-orbiting satellites whose original missions were to support operational forecasting. This paper describes the completely revised adjustments to the Microwave Sounding Unit (MSU) deep-layer tropospheric temperature products first reported in Spencer and Christy (1990). These data originate from nine different satellites, the first being launched in late 1978, and their periods of operation varied from about a year (TIROS-N) to over six years (*NOAA-11* and *-12*). The version presented here is termed version D, and is thus the third major revision to these datasets. For details on the background of the MSU data, the reader is referred to Spencer et al. (1990), Christy (1995), and Christy et al. (1998).

Two deep-layer tropospheric temperature products, one for the lower troposphere ($T_{2LT}$ surface to about 8 km) and one for the midtroposphere ($T_2$ surface to about 15 km, thus including some stratospheric emissions), are based on the observations of channel 2 of the MSU (Fig. 1). The basic measurement utilized is the intensity of the oxygen emissions near the 60-GHz absorption band, which is proportional to atmospheric temperature. Details of the satellite characteristics, scan pattern, and frequency distribution are found in the literature cited above.

Version A of these products was constructed by a simple merging procedure in which biases were calculated and removed from the individual satellites (Spencer and Christy 1992a,b). We updated version A after discovering that the eastward drift of *NOAA-11* over its 6-yr life span caused a spurious warming effect to develop due, as we believed, to the fact the satellite was sampling the earth at later times during the local diurnal cycle (version B, Christy et al. 1995). The net effect of the correction in version B caused the overall trend of $T_{2LT}$ to be more negative by 0.03 K decade$^{-1}$ over version A.[1] (Note: The changes brought about by

*Corresponding author address:* John R. Christy, ESSC/GHCC, University of Alabama in Huntsville, Huntsville, AL 35899.
E-mail: christy@atmos.uah.edu

[1] Values for versions A and B have not been updated or archived for comparisons with version D here.

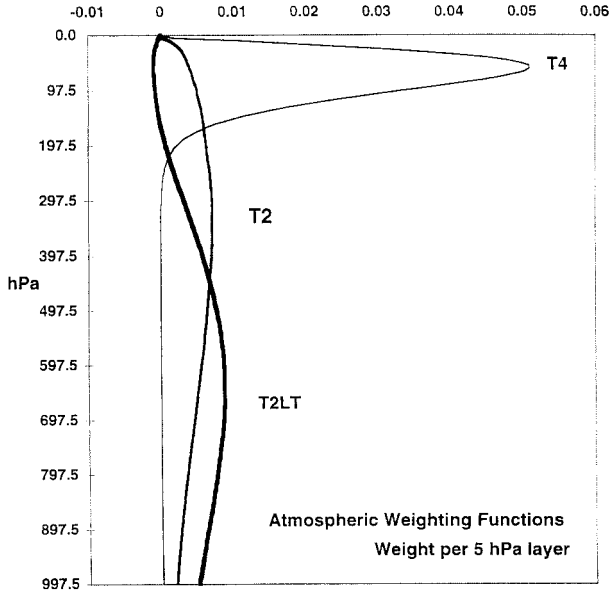


FIG. 1. Atmospheric weighting functions of three MSU products per 5-hPa layer.

each revision are very minor in terms of interannual variations, but the trend of the time series is sensitive to such small changes and thus is a useful metric in identifying the impact of revisions. So, except in a few instances throughout this paper, we shall report the impact of the changes in terms of the global trend: a linear best fit by the minimization of least square differences through the annual anomalies. The trend is not used here as a predictive tool for understanding climate variability and change.)

We subsequently noticed that *NOAA-7* had also drifted eastward enough to warrant a correction. In addition, spurious variations between *NOAA-12* and the other satellites, phase-locked with the annual cycle, indicated something unusual. We eventually identified the MSU's instrument body temperature as the source of these cyclic variations, and they were calculated and removed. (These intraannual cyclic errors were determined from the intersatellite comparisons, not from the actual instrument body temperatures as will be applied here.) Thus an estimated correction for *NOAA-7*'s drift and corrections for these spurious annual harmonics were applied with a net effect in version C (called c1 in the earlier publication) of making the trend more positive over version B by about 0.03 K decade$^{-1}$ for $T_{2LT}$ (Christy et al. 1998). Several tests to determine the precision of the intersatellite biases were included in that study, which are applicable to version D as well.

### *a. Previous studies*

Several studies, most using version B, have been published that in some fashion have examined the utility of the MSU datasets. Trenberth et al. (1992), Christy (1995), and Hurrell and Trenberth (1996) focused on the relationship between surface temperatures and those of the deep layer troposphere as measured by the MSU. These studies indicated, among other things, that there is low correlation of anomalies between the surface and the tropospheric temperature in tropical and subtropical oceanic regions, at times even falling below zero. However, over midlatitude continents, in which the atmosphere is subject to greater vertical mixing, regional correlations were often above 0.9.

Other studies relying on measurements of the tropospheric temperature from radiosondes and/or global weather analyses (rather than surface measurements) have demonstrated excellent agreement between those datasets and MSU $T_{2LT}$. By excellent agreement we mean annual global anomalies differ by less than 0.10 K and global trends are within 0.06 K decade$^{-1}$ (Nicholls et al. 1996; Basist and Cheliah 1997; Parker et al. 1997; Stendel and Bengtsson 1997; Pielke et al. 1998). These studies indicated that the surface–troposphere relationship should *not* be viewed as a rigidly connected system over 20-yr time periods.

In spite of the excellent agreement between various datasets of tropospheric temperature in the studies noted above, a very legitimate motivation for closer examination of the MSU datasets (versions A, B, and C) dealt with the presence of a trend in the midtroposphere $T_2$, which was more positive than that of the lower troposphere $T_{2LT}$ (Hurrell and Trenberth 1997, Wentz and Schabel 1998). Though the measurement error range associated with the trends of each layer allowed for the values of both quantities to be equal (or even for $T_2 \cdot C$ to be more negative than $T_{2LT}$. C by up to 0.03 K decade$^{-1}$), the fact the stratosphere has experienced a strongly negative trend provided evidence that something required correction in either $T_2 \cdot C$ or $T_{2LT}$. C or both. Results of the discoveries described below demonstrate that both datasets required reconstruction from the digital radiance counts. With version D this unusual trend behavior is reconciled resulting in a systematic reduction of trends from $T_{2LT}$ to $T_2$ to $T_4$ (lower stratosphere from MSU channel 4).

### *b. New discoveries*

Following the release of version C in mid-1996 there was the typical delay in the appearance of the published results (August 1998), during which we discovered a temporal component to the instrument body temperature effect (discussed later) that was interannual, not just intraannual as documented in version C. This effect appeared to introduce an artificial warming in the time series of both $T_2$ and $T_{2LT}$. Elsewhere, Wentz and Schabel (1998) discovered that the vertical height of the satellites was a critical parameter affecting $T_{2LT}$ and kindly shared their results with us before their paper was published (also August 1998) and just before our version C galley proofs were returned to the printers

(thus it is mentioned but not applied to version C in Christy et al. 1998). Their important finding is that altitude losses of only 1 km cause artificial cooling in $T_{2LT}$ while having virtually no effect on $T_2$. The accumulated downward fall of the satellites over the 1979–98 period was over 15 km, and thus became a rather substantial factor requiring attention. In addition, corrected NESDIS nonlinear calibration coefficients for *NOAA-12* became available in this period (between release of version C and publication) and were needed for any further versions.

In version D, presented here, we apply the new NESDIS calibration coefficients to *NOAA-12* and then account for and remove the effects of orbit decay and the diurnal effect of orbit drift individually from the original satellite brightness temperatures (sections 2a and 2b). We finally calculate, by solving a system of over 4000 linear equations, the coefficients of the MSU's instrument body temperature needed for each satellite to eliminate this spurious effect (section 2c). Relative to version C, the global impact of version D is characterized by a more negative trend for 1979–98 of $T_2$ (from $+0.08$ to $+0.04$ K decade$^{-1}$) and a more positive trend of $T_{2LT}$ (from $+0.03$ to $+0.06$ K decade$^{-1}$). We estimate the 95% measurement error range of these trends as $\pm 0.06$ K decade$^{-1}$. These results now show that the stratospheric portion of $T_2$ influences its trend to be more negative than that of $T_{2LT}$, though the error range may imply other possibilities. In section 3 we discuss accuracy and comparisons with version C and in section 4 we offer concluding remarks.

## 2. Adjustments prior to merging

The basic problem of this research is to determine how to merge data from nine instruments to produce a useful time series of deep-layer atmospheric temperatures. In constructing the previous versions of the MSU data (A, B, and C) we relied exclusively on the observations obtained as two satellites monitored the earth simultaneously, that is, as a coorbiting pair, to adjust the data for errors. Corrections were applied which eliminated major differences between the various pairs (e.g., intersatellite difference trends and annual cycle perturbations; Christy et al. 1998). In general, when data differences between two satellites were found, a decision was made as to which satellite was correct and which was in error, based on local equatorial crossing time variations or other factors. Some aspects of the temperature differences (trend and annual cycle) of the one deemed in error were then removed, forcing a good (but somewhat contrived) match with the one deemed to be correct.

We take a different approach here in which we calculate and remove two effects, 1) orbit decay and 2) diurnal sampling due to orbit drift, for each satellite prior to any intersatellite comparison. A third source of error, also resulting from orbit drift due to the variations in the temperature of the instrument, is calculated by solving a system of linear equations in which the differences of global temperatures from coorbiting pairs and their instrument body temperatures are utilized. The solutions provide the linear coefficients of the instrument body temperatures which explain the most variance in the temperature difference time series; that is, we solve for the cause of the differences in earth-viewed temperatures measured simultaneously by a coorbiting pair. This third source of error, based on the instrument body temperature, is then removed from the time series of each satellite. After this, the time series of the nine satellites are merged taking into account the simple biases among them as in Christy et al. 1998.

### *a. Orbit decay*

Polar-orbiting satellites are placed at altitudes near 850 km at which there exists a very slight drag due to the thin atmosphere. During solar activity maxima, the upper atmosphere expands vertically, thus placing the satellite in an atmosphere with greater density and greater drag. As a result, the NOAA satellites tend to drop in altitude about 7 or 8 km over a 3- to 4-yr period during the enhanced solar activity.

The MSU observes 11 views per 26-s cross-track scan, with view 6 being at nadir and views 1 and 11 at the left and right limbs (47° from nadir; Spencer et al. 1990). Wentz and Schabel (1998) examined the impact on the earth-viewed brightness temperature (Tb) of these individual view angle positions as a function of satellite altitude. They discovered that as the satellite's orbit decays, there is a differential effect on the observed Tb as a function of view angle: the outer view angle observations will show greater warming than the inner view angles (see Wentz and Schabel 1998 for details). This result occurs for any atmosphere in which there is a decrease in temperature with height, and is thus greater in a tropical atmosphere than a polar atmosphere (i.e., there would be no impact in an isothermal atmosphere). The $T_{2LT}$ retrieval includes as part of its calculation the difference of the inner (3, 4, 8, and 9) and outer (1, 2, 10, and 11) view temperatures in order to remove the effect of emissions from the stratosphere (Fig. 1). Thus, any differential change in the inner and outer Tb will impact the calculated value of $T_{2LT}$ as illustrated with the retrieval formula:

$$T_{2LT} = T_{inner} + 3(T_{inner} - T_{outer}). \quad (1)$$

Because the outer views will respond with greater warming than the inner views during orbit decay, the net impact on the 1979–98 time series of $T_{2LT}$ is to introduce an artificial cooling of almost $-0.10$ K decade$^{-1}$.[2] The

[2] Wentz and Schabel calculate 0.12 K decade$^{-1}$ for 1979–95. However, since little decay occurred during 1996–98 and the fact we include seasonal and latitudinal corrections, the overall trend is slightly less than 0.10 when considering 1979–98.

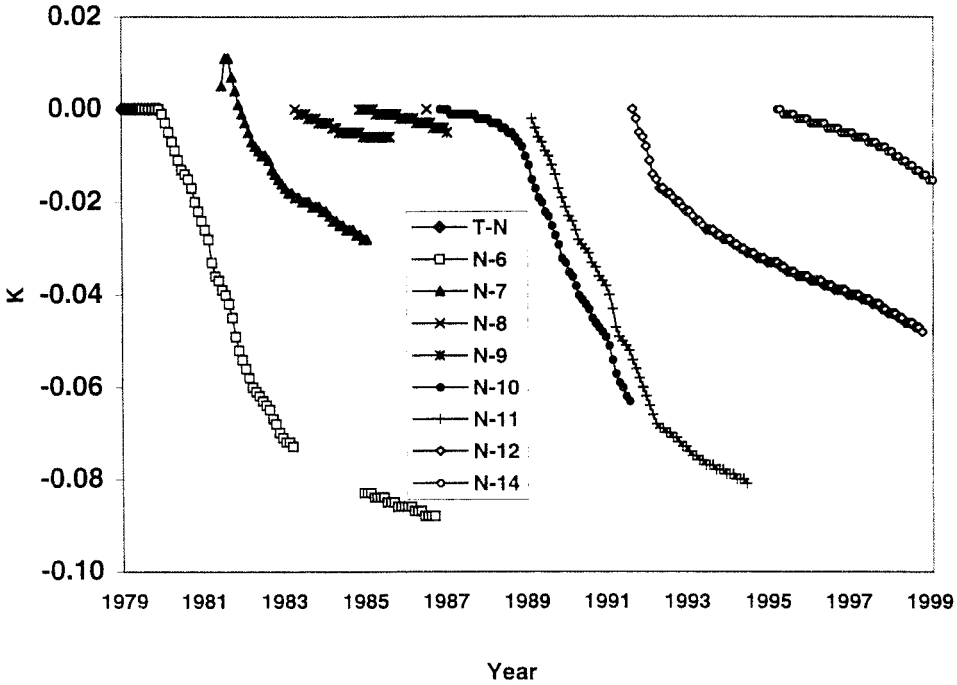


FIG. 2. Monthly, globally averaged temperature effect on the MSU temperatures of $T_{2LT}$ due to orbit decay (loss of altitude) for each satellite identified (e.g., N-6 is *NOAA-6*).

effects for each satellite individually are shown in Fig. 2, but note that this is an accumulating effect. (This figure also indicates the period of operation of the respective satellites.) The effect has a seasonal variation in the extratropics as the average lapse rate changes slightly, so the corrections are latitudinally and seasonally dependent and are based on standard atmospheric profiles for the Tropics, midlatitudes, and the subarctic. The corrections are calculated for each latitude band with there being two transition zones: 1) from latitudes 22.5° to 37.5° for tropical to midlatitude, and 2) from 60.0° to 70.0° for midlatitude to subarctic. We employed one standard profile for the Tropics, and two each for the midlatitudes and subarctic (winter and summer). For a given Julian day, the proximity to midsummer and midwinter was determined and a profile generated based on an interpolated average in time and space. From the temperature profile, a radiation model requiring each spacecraft's altitude and view angle orientation, determined the orbit decay effect. Variations in this interpolation procedure were tested but had only minor impact (less than 0.01 K decade$^{-1}$) on the global average.

Because the magnitude of the lapse rate of the vertical layer sampled by the MSU declines as one moves poleward, the orbit decay effect is reduced. For example, the decay effect on *NOAA-11* from beginning of service (October 1989) through March 1995 was 0.096, 0.073, and 0.061 K for the tropical, midlatitude, and subarctic regions, respectively.

As a geometric problem, it is straightforward to calculate and apply the corrective adjustments for orbit decay, and this is now the first correction performed on the raw satellite temperatures (which for *NOAA-12* has new calibration coefficients) for each product. Because orbit decay had a miniscule effect on inner view angles, the product $T_2$ (which is the average of positions 4–8) was negligibly affected, though the small corrections are also now applied.

### *b. Diurnal effect of earth emissions*

A NOAA polar orbiter is nominally ''sun synchronous,'' meaning whenever it observes a particular spot on the earth at nadir, the local time on the earth is constant from year to year, usually being referenced to the crossing time over the equator [i.e., local equatorial crossing time (LECT)]. In practice, however, all of the spacecraft experienced an east–west drift away from their initial LECT. The morning satellites (about 1930/0730 UTC; *NOAA-6, -8, -10, -12*) remained close to their original LECTs, but after a few years would drift westward to earlier LECTs, for example from 1930/0730 to 1900/070.[3] The afternoon satellites (about 1400/0200 - *TIROS-N*, *NOAA-7, -9, -11,* and *-14*) were purposefully given a small nudge to force them to drift eastward to later LECTs to avoid backing into local solar noon. *NOAA-11,* for example, drifted from 1400/0200 to about

[3] The first value, in 24-h time, represents the northbound LECT and the second the southbound LECT 12 h later (or earlier).

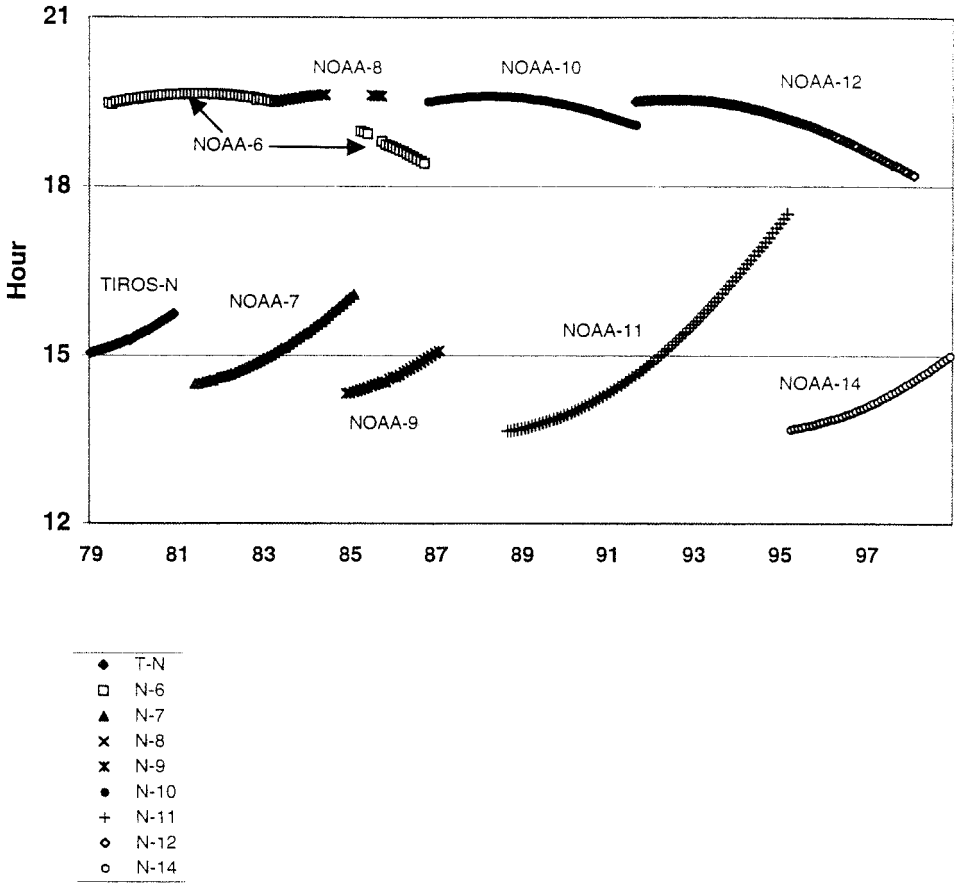


FIG. 3. Local equatorial crossing time for each satellite for the northbound (ascending node) pass in local time.

1800/0600 during six years, becoming essentially a morning satellite. Figure 3 displays the LECTs for the northbound (ascending) pass of each of the spacecraft during their operational service.[4]

As a satellite drifts through new LECTs, it consequently samples the emissions from the earth at changing local times, in effect allowing local diurnal cycle variations to appear in the time series as spurious trends. This is particularly true for the afternoon spacecraft since the temperature change is greater as the afternoon (northbound) pass drifts to new times than the nighttime (southbound) pass. Thus there is a net trend in the daily average of the measured temperature.

For $T_2$, the net effect of the drift is to introduce small artificial changes. For example, over oceans, Tb tends to rise to a peak in late afternoon as the troposphere warms due to the combination of mechanisms affecting the vertical transport of heat, that is, convection which transports sensible and latent heat combined with direct solar heating of the atmosphere. However, over bare ground, Tb may decrease as the skin temperature, which contributes more to Tb over land than ocean, becomes cooler after local noon. Over vegetated regions, the effect on Tb of an eastward drift is a combination of tropospheric warming and surface cooling and is difficult to detect for a few hours of orbit drift in the daily average. Only in land regions such as the Sahara Desert do we see a systematic drop in Tb shortly after solar noon. Globally, these effects are very small for the inner views (i.e., $T_2$) of the MSU. We find, however, that Tb of the outer view positions used in $T_{2LT}$ cool at a greater rate during the drift than the inner view positions. The net impact is to introduce an artificial warming trend almost everywhere in $T_{2LT}$.

In version C of the MSU products we estimated this diurnal trend by direct comparison of an eastward drifting afternoon satellite versus a slowly drifting morning satellite. Not known to us at the time version C was released, however, was another consequence of the drifting LECTs, namely, the interannual relationship of the observed Tb and the temperature of the radiometer. These two effects were not separately dealt with in version C. Thus we assumed the estimated artificial warm-

[4] *NOAA-6* was replaced by *NOAA-8* in 1983. However, *NOAA-8* developed problems and was eventually shut down. *NOAA-6* was placed back into service to monitor the 1930/0730 orbit slot.

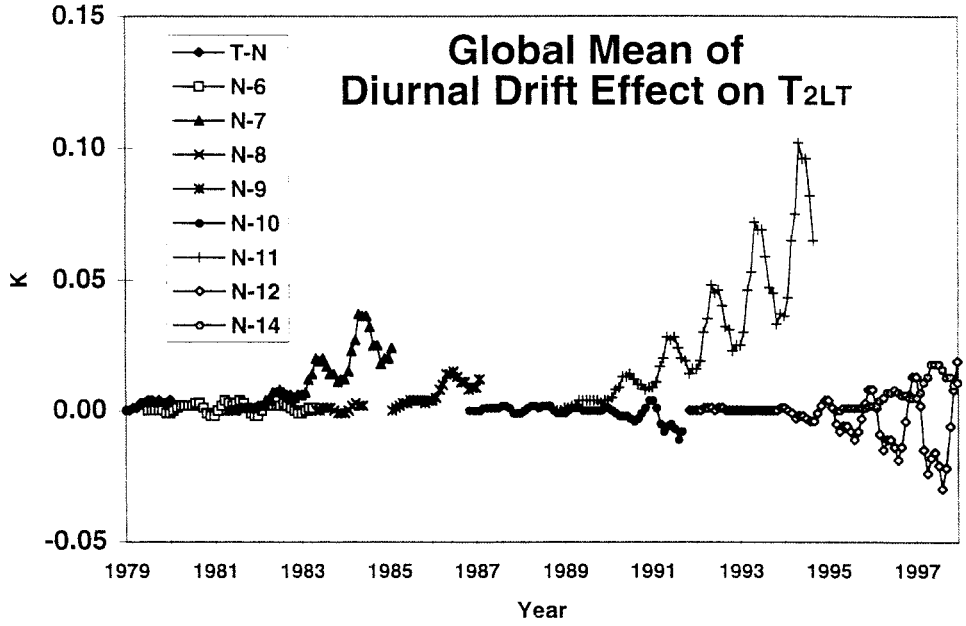


FIG. 4. Monthly, globally averaged temperature errors of $T_{2LT}$ due to variations in the local time of observation (diurnal variations of earth emissions) because of east–west drift of the spacecraft.

ing in *NOAA-11* (a drifting satellite) versus *NOAA-10* and *-12* was due entirely to the diurnal effect when in fact it was a combination of the expected diurnal warming of *NOAA-11* from changing earth emissions and the unaccounted-for instrument body temperature effect in all three instruments (see the next section).

To determine the diurnal effect in isolation, we examined the average temperature by view angle position across the swath as a function of crossing time, latitude, season, and surface type (land or ocean). At the equator, position 11 represents a local time about 80 min later than position 1. Similarly, position 10 observes a local time about 50 min later than position 2, and so on. Cross-swath temperature differences (i.e., 11 minus 1, etc.) were found to be very systematic from one satellite to the next. The cross-scan Tb differences calculated here occur over the time period of a single scan, and thus are essentially instantaneous differences (~15 s). This measurement is thus unaffected by slow changes in effects such as the instrument body temperature (discussed below), which varies on the order of weeks to years and is dependent on other factors.

By accumulating millions of such ''almost-instantaneous'' cross-swath differences, we were able to estimate the temperature change that would result as the satellite drifted through LECTs for $T_2$ and $T_{2LT}$. Values for ascending and descending nodes were calculated separately, then applied with appropriate proportions for land and ocean, to the zonal mean temperatures accordingly. The monthly, global mean values of those corrections for $T_{2LT}$ are shown in Fig. 4. The net effect on the 1979–98 time series of $T_{2LT}$ is to remove an artificial positive trend of about 0.03 K decade$^{-1}$.

### c. Instrument body temperature effect

As mentioned in Christy et al. (1998), we discovered a spurious influence on the calculation of Tb due to the temperature of the instrument itself. This was readily apparent in systematic, intraannual differences of coorbiting satellites and was calculated and removed. However, the interannual component to this effect was not considered in constructing version C when released in 1996. The basic idea is that the temperature of the instrument itself varies in accordance with its exposure to sunlight, the shadowing effects of instrument configuration and the position of the louvers which regulate the loss of heat from the MSU's radiator plates. In Fig. 5 we show the monthly averages of the warm target plate temperature ($T_W$), a surrogate for the temperature of the radiometer, as monitored by the platinum resistance thermometers (PRTs) embedded in the plate. Note that the morning satellites tend to have annual cycle oscillations, while the afternoon satellites tend to show systematic warming as they drift to later LECTs, exposing the instrument to more sunlight.

The determination of earth-viewed Tb from the observed digital counts is based on an interpolation scheme between two temperature anchor points: cold space and the onboard warm target plate $T_W$. The MSU reports the intensity of microwave radiation as digital counts for the 11 earth views and for cold space and the warm target. The temperature for cold space is known (2.7 K) and that of the warm target is monitored by the two PRTs. Thus a relationship is then computed between counts and Tb given the digital counts and temperatures of the anchor points (Spencer et al. 1990). A key factor

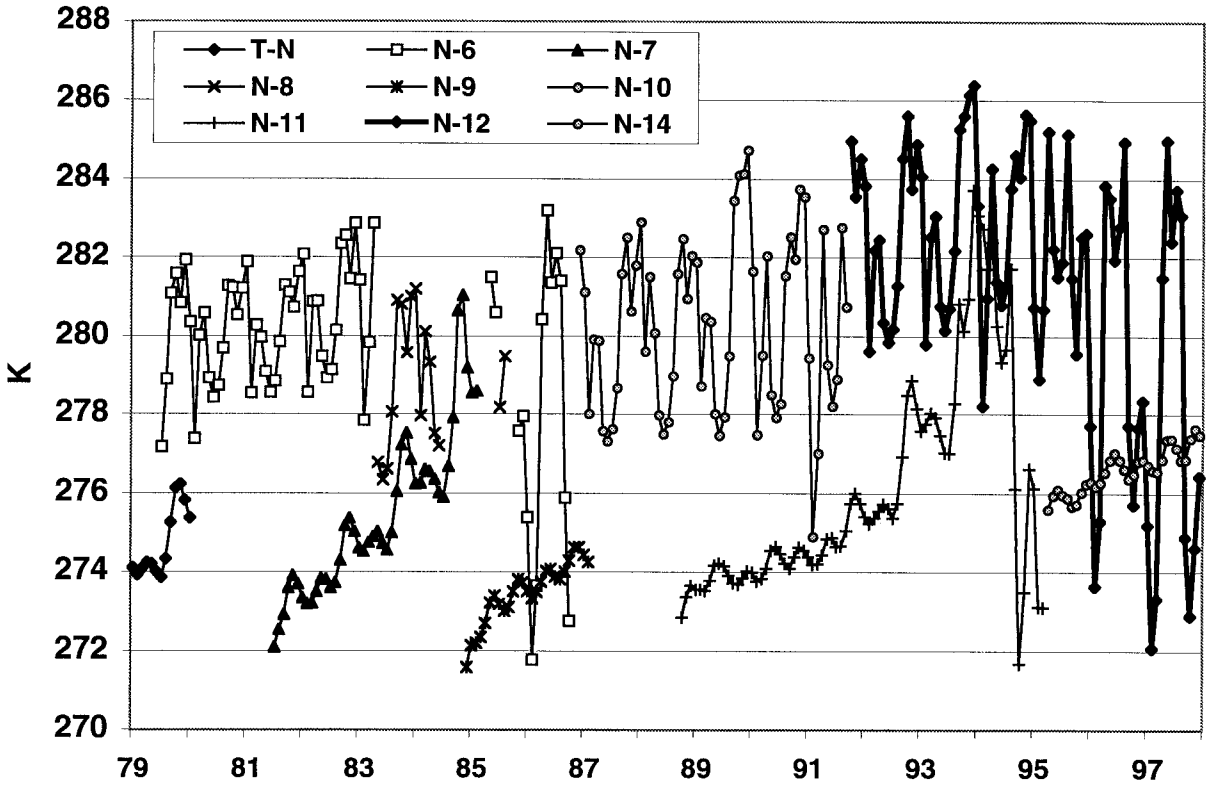


FIG. 5. Monthly average temperature ($T_w$) of the two platinum resistance thermometers (PRTs) embedded in the channel-2 warm target of the individual MSUs. This represents the temperature of the instrument itself.

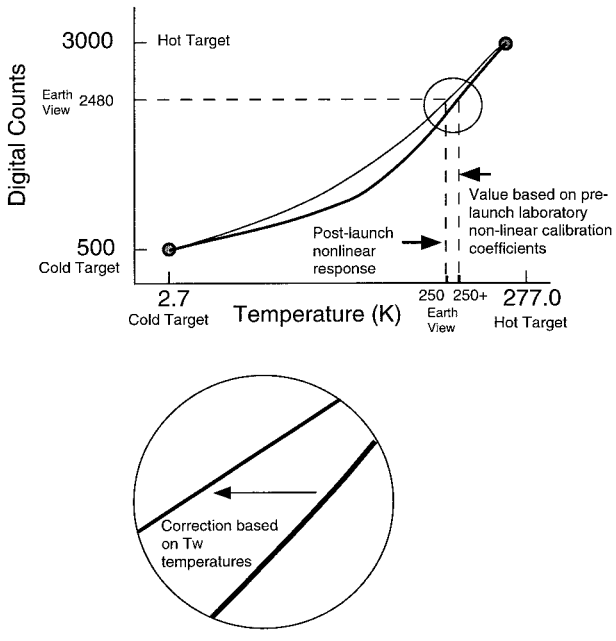


FIG. 6. Schematic diagram of nonlinear error correction that is partially a consequence of the temperature variations of the instrument. In general, as the instrument heats ($T_w$ increases) the calculated earth-viewed temperature based on laboratory nonlinear calibration coefficients becomes too warm.

here, however, is that the interpolation scheme is nonlinear, being additionally dependent on the difference between the scene that is viewed and the actual temperature of the radiometer (Mo 1995). The nonlinear adjustment is intended to account for this (and other effects) because the two anchor points are not sufficient in and of themselves to determine the magnitude of the nonlinear effect. A schematic of this situation is given in Fig. 6. (Note that over the values of the earth-viewed temperatures the adjustment is essentially linear, being a recalibration of the slope.)

Before launch, the instruments are tested for their response to various instrument body temperature regimes while viewing targets of known temperature in a laboratory chamber. An example is given in Mo (1995). In this way, the coefficients of the nonlinear equation are empirically estimated for each instrument in prelaunch laboratory conditions. However, once the instrument is integrated onto the spacecraft and launched into the environment of space, these coefficients tend to require readjustment due to changes, for example, in the instrument gain (i.e., ratio of $\Delta$counts to $\Delta$Tb; Mo 1995).

We noted in Christy et al. (1998), that *NOAA-12* had a significant systematic error response which was highly correlated (0.96) with its instrument body temperature. A clear example of this effect also occurs for *NOAA-11*. In Fig. 7 we show the instrument body temperatures,

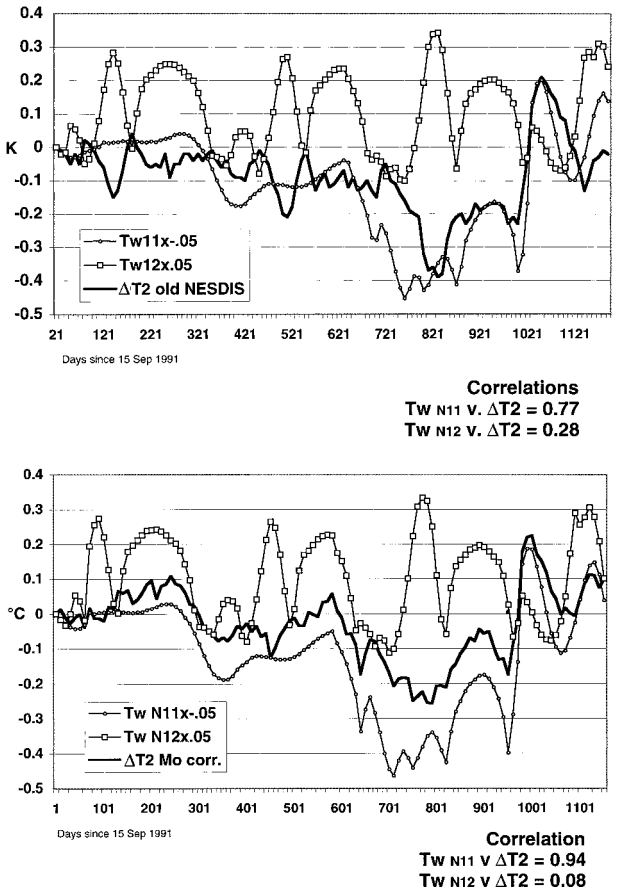


FIG. 7. (top) Instrument body temperature departures for *NOAA-11* and *-12* (Sep 1991–Mar 1994) and the difference in their observed global temperature ($\Delta T_2$) for 10-day averages (original NESDIS calibration coefficients). (bottom) As above except the global $T_2$ temperatures of *NOAA-12* in the calculation of $\Delta T_2$ utilize the updated calibration coefficients from Mo (1995).

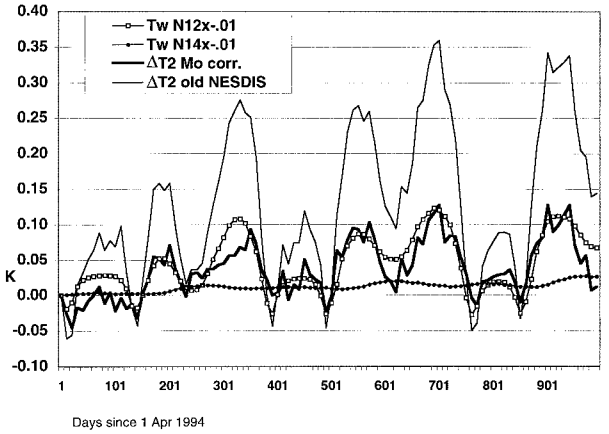


FIG. 8. As in Fig. 7 for a comparison on *NOAA-12* and *-14* from Apr 1994 to Dec 1997. Note the significant reduction in the global temperature differences ($\Delta T_2$) observed by the two spacecraft after the Mo-corrected nonlinear calibration coefficients are applied.

$T_W$, for both *NOAA-11* and *NOAA-12* as well as the difference in their daily global $T_2$ earth-viewed temperatures ($\Delta T_{11,12}$) averaged over 10-day periods. In the top set of time series, we show $\Delta T_{11,12}$ based on the original prelaunch NESDIS coefficients for *NOAA-12*, and there appears to be strong relationships between $T_W$ of each satellite and their observed global average temperature differences ($\Delta T_{11,12}$). Note, for example, the anticorrelated temperature spikes between *NOAA-12* $T_W$ and $\Delta T_{11,12}$ near days 100 and 500. If both instruments were perfectly calibrated, the time series of $\Delta T_{11,12}$ would be zero since both satellites observe the same earth over the same periods. It is clear there are some differences.

The laboratory-estimated coefficients of *NOAA-12* were in need of revision due to an error in their original derivation discovered by Mo (1995) and due to a significant change in the gain of channel 2 after launch. The original nonlinear coefficients were based on cold target counts of about 700, but once in space, the instrument was reporting cold target counts of about 1800. Mo recalculated the coefficients from the original laboratory data and the observed data, resulting in a substantial increase in the nonlinear term. We applied the new nonlinear coefficients giving the results in the second plot of Fig. 7. The essential consequence of the correction was to alter the relative slope over the nominal range of earth-viewed temperatures of *NOAA-12* by about 2.5% (in terms of $T_W$ coefficients described later this would be $-0.025$).

As indicated above, the adjustment for errors and gain changes in *NOAA-12* appear excellent for the period through 1993. However, the substantial solar illumination change as indicated by the PRT temperatures for *NOAA-12* after 1994 indicate a further minor adjustment is required for that period (Fig. 5). Between 1993 and 1995, the annual cycle of *NOAA-12* $T_W$ temperatures changed dramatically with, for example, October's average $T_W$ cooling by over 10 K, while May's increased by over 3 K due to changing shadowing effects. In Fig. 8 we show $\Delta T_{12,14}$ (*NOAA-12* vs *NOAA-14*) and their $T_W$ values. Note that a considerable reduction was made to $\Delta T_{12,14}$ by the new *NOAA-12* coefficients of Mo, but with the new orbiting regime from 1995 onward, a further small correction is required as the gain appears to have changed again (see below).

Returning to *NOAA-11* and *NOAA-12*, it is apparent that the variations observed in $\Delta T_{11,12}$ are essentially related to the instrument body temperature of *NOAA-11*'s radiometer. Fortunately, the relationship of the error ($\Delta T_{11,12}$) to $T_{W11}$ is basically linear (correlation of $\Delta T_{11,12}$ versus $T_{W11} = 0.94$ for 10-day averages, $r = 0.97$ for 91-day averages) so that an adjustment to the original temperature, in this case a reduction in response of about 3.5% for *NOAA-11*, will produce a difference time series of $\Delta T_{11,12}$, which is essentially noise.

The adjustment of the original values of Tb using $T_W$ information is in essence an exercise in overcoming the differences between the prelaunch laboratory coeffi-

cients and the actual performance of the instrument in space. To do this in an objective way, we set up an equation for each day in which two satellites, *m* and *n*, are simultaneously observing the earth,

$$\Delta T_{m,n} = \text{bias}_{m,n} + a_m T_{Wm} - a_n T_{Wn}, \quad (2)$$

where $\Delta T_{m,n}$ is the global temperature difference (i.e., earth views) between the two satellites, $\text{bias}_{m,n}$ is the constant offset, and $a_m$ is the linear coefficient of the $T_{Wm}$ time series that explains the variance in $\Delta T_{m,n}$. We required at least one year of overlapping data. Four satellites (*NOAA-6, -7, -11,* and *-12*) have overlapping observations with two other satellites, while five (*TIROS-N, NOAA-8, -9, -10,* and, *-14*) overlap with one.

A preliminary check using all data indicated that the $T_W$ variations of three instruments explained less than 1% of the variance in $\Delta T_{m,n}$ during their overlapping periods. These were *TIROS-N* in its overlap with *NOAA-6* (July–December 1979), *NOAA-10* in its overlap with *NOAA-11* (October 1988–August 1991), and *NOAA-12* with its overlap with *NOAA-11* (September 1991–March 1995, e.g., Fig. 7). This implies that the nonlinear NESDIS coefficients for these three instruments over these periods were well-calibrated since the $T_W$ variation did not explain any of the $\Delta T_{m,n}$ variance. For *TIROS-N* and *NOAA-10,* these calibration coefficients were the prelaunch values and for *NOAA-12* the Mo-corrected values. (The next lowest amount of explained variance by a single instrument's $T_W$ of $\Delta T_{m,n}$ was greater than 40%.) We therefore removed these three data periods from the matrix so as not to corrupt the solution. There were five overlapping periods remaining using seven satellites (*NOAA-12* after 1994 is included), so that each satellite's data does interact with at least two others in the solution process.

This exercise, then, becomes a system of equations that is solved simultaneously to determine the biases and the constant coefficients *a* for $T_W$. Thus, to the extent that the variations in $\Delta T_{m,n}$ are linear functions of $T_W$, we are able to determine the adjustment factors for each instrument. Fundamentally, this is a procedure designed to reduce the differences between coorbiting satellites based on $T_W$ variations. The highest level of variance reduction in $\Delta T_{m,n}$, as explained by $T_W$, occurred when $\Delta T_{m,n}$, $T_W$ and the biases were averaged over periods of 61 to 121 days (checked in 10-day increments). The daily values of $\Delta T_{m,n}$ and $T_W$ (i.e., nonaveraged in time) contained greater noise, while longer averaging periods had too few degrees of freedom, especially given the 370-day overlap of *NOAA-6* and *-9.*

The values shown in Table 1 for coefficients $a_m$, are averages for these coefficients produced from solving seven matrices, each representing the averaging period indicated above (61, 71, 81, 91, 101, 111, and 121 days). The standard deviation among all results was 0.0033 per coefficient [see section 3a(1) for a characterization of this potential error on the time series]. For each of the seven matrices, the rows were simply the values for the expression shown in Eq. (2) for each day. The values for a specific day, say, 30 June 1990, would change slightly from matrix to matrix as the averaging period centered on that day was lengthened for each of the $T_W$ and $\Delta T_{m,n}$ inputs.

TABLE 1. Coefficients of the warm target plate temperature ($T_W$) for each satellite determined by means described in the text. During overlap periods, daily temperatures are available for each of the two co-orbiting satellites and statistics are produced therefrom. The "$\sigma_{\Delta\text{Tdy}}$ of daily $\Delta T_{m,n}$" is the standard deviation of the differences in the daily global temperatures produced from all overlapping periods of co-orbiting satellites.

| | $T_2$ Matrix | $T_2$·D Adjusted | $T_{2LT}$·D |
|---|---|---|---|
| *TIROS-N* | | | |
| *NOAA-6* | −0.002 | −0.002 | −0.002 |
| *NOAA-7* | −0.021 | −0.018 | −0.018 |
| *NOAA-8* | −0.039 | −0.036 | −0.036 |
| *NOAA-9* | −0.096 | −0.095 | −0.095 |
| *NOAA-10* | | | |
| *NOAA-11* | −0.035 | −0.035 | −0.035 |
| *NOAA-12* | −0.007 | −0.007 | −0.007 |
| *NOAA-14* | −0.017 | −0.015 | −0.015 |
| $\sigma_{\Delta\text{Tdy}}$ (daily $\Delta T_{m,n}$) | 0.0320 | 0.0319 | 0.0685 |
| Global trend (K decade$^{-1}$) | +0.04 | +0.04 | +0.06 |

As indicated in (2) for any given row (i.e., specific day) of the matrix, there was one value in the appropriate bias column, one for each $T_W$ in the two columns representing the two satellites with data for the given day and one value ($\Delta T_{m,n}$) in the earth-viewed global temperature difference column (all other entries relative to nonobserving satellites in the row being zero). There were over 4000 rows representing the number of days on which data were available for two coorbiting satellites. A solution was then produced which generated the biases and $a_m$ coefficients which explained the most variance in $\Delta T_{m,n}$.

Since these are linear operations, we will apply the $T_W$ coefficients for $T_2$ to $T_{2LT}$ because $T_{2LT}$ has greater noise than $T_2$ due to the retrieval algorithm. Unlike the diurnal correction, the instrument body effect is slowly varying and is globally systematic.

Before describing the completed dataset in which the instrument body factor is incorporated, we shall address the notion of dependency of MSU temperatures on $T_W$ directly from the calibration equations. The idea that secular variations in $T_W$ are related to calibration errors appears clear from Figs. 7 and 8. We are able to estimate the magnitude of the dependency between $T_W$ and the observed antenna temperature ($T_{\text{ant}}$) in the case of *NOAA-12* because accurate calibration coefficients have been determined postlaunch (Mo 1995).

The basic NESDIS algorithm is

$$T_{\text{ant}} = T_W + (c_0 + c_1 X_e + c_2 X_e^2 - X_w) G^{-1}, \quad (3)$$

$$G = \frac{(X_W - X_c)}{(T_W - T_c)}, \quad (4)$$

where $X_e$, $X_W$, and $X_c$ are counts on the earth view, warm target, and cold space, respectively; and $T_{\text{ant}}$, $T_{\text{W}}$, and $T_c$ are the corresponding temperatures. Hence, $T_{\text{ant}} = T_{\text{lin}} + T_{\text{nonlin}}$, where $T_{\text{lin}}$ represents the simple linear calibration and

$$T_{\text{nonlin}} = [c_0 + (c_1 - 1)X_e + c_2X_e^2]G^{-1}, \quad (5)$$

where the coefficients $c_0$, $c_1$, and $c_2$ define the nonlinearity and are determined in the prelaunch calibrations. Since $X_W$ and $T_W$ vary together and are measured on board the spacecraft, a purely linear system would not have a correlation between $T_W$ and $T_{\text{ant}}$. Thus, the error component described earlier arises from the nonlinear aspect of the radiometer response.

As expressed above, $T_{\text{nonlin}}$ has no direct dependence on $T_W$, except possibly indirectly through a variation of gain with instrument temperature. However, Mo (1995, his appendix B) notes that Eq. (3) is self-consistent only if the coefficients $c_0$ and $c_1$ depend on the calibration temperatures (or counts)

$$c_0 = c_2X_WX_c, \quad (6)$$

$$c_1 = 1 - c_2(X_W + X_c). \quad (7)$$

Substituting (6) and (7) into (5) yields

$$Q = c_2(X_e - X_W)(X_e - X_c)G^{-1}, \quad (8)$$

$$\approx c_2(T_{\text{ant}} - T_W)(T_{\text{ant}} - T_c)G, \quad (9)$$

and hence

$$\partial Q/\partial T_W = -c_2(T_{\text{ant}} - T_c)G. \quad (10)$$

Thus a dependency with respect to $T_W$ may be established for $Q$, an alternate expression for $T_{\text{nonlin}}$. To consider an example, the MSU data for *NOAA-12* were calibrated by Eq. (3) [ignoring the dependencies expressed by (6) and (7)]. The calibrated antenna temperature $T_{\text{ant}}$ would contain an error $T\varepsilon_{\text{ant}} = T_{\text{nonlin}} - Q$ (note: $T_{\text{nonlin}}$ is based on digital counts and $Q$ is based on temperatures, including $T_W$) and thus

$$\frac{\partial T\varepsilon_{\text{ant}}}{\partial T_W} = \frac{-\partial Q}{\partial T_W}. \quad (11)$$

(Since the error is essentially the only part of $T_{\text{ant}}$ that is dependent upon $T_W$, one may also write $\partial T_{\text{ant}}/\partial T_W = -\partial Q/\partial T_W$.) Using the values for *NOAA-12* appearing in Mo (1995), $c_2 = 2.59 \times 10^{-5}$ and $G = 4.04$ counts per kelvin along with the mean observed values, $T_{\text{ant}} = 249.7$ K, $T_c = 2.7$ K, we calculate from (10) the estimated error factor of due to $T_W$:

$$\frac{\partial T\varepsilon_{\text{ant}}}{\partial T_W} = 2.59 \times 10^{-5}(247.0)(4.04) = 0.026. \quad (12)$$

Thus, we show that the error associated with the nonlinear component may be on the order of a few hundredths of a degree for a 1° change in $T_W$. This is similar to our empirical calculation for *NOAA-12* ($-0.025$, negative because it is applied as a correction) of the $T_W$ coefficient when calculated from the uncorrected data and indicates the magnitude of the correction, or in the present context, the magnitude of the $a_m$'s, that could be expected.

### *d. Adjustments for $T_2 \cdot D$, $T_{2LT} \cdot D$, and their completed time series*

The results of the matrix solutions are given in Table 1 for $T_2$ along with other information regarding the completely merged time series for $T_2 \cdot D$ and $T_{2LT} \cdot D$ based on the coefficients listed. The table includes the standard deviation of the nonsmoothed daily global differences between all satellite pairs ($\sigma_{\Delta\text{Tdy}}$) and the 20-yr global decadal trend.

The 61–121-day averaging periods used to calculate the coefficients provided a robust calculation of the instrument body effect, giving values whose standard deviation of $T_W$ coefficient differences ($\sigma_{\Delta a}$) among the averaging periods of 0.0033, or about 10% of the typical $T_W$ coefficient. The $T_W$ coefficient for *NOAA-9* is the largest, which is consistent with the fact its NESDIS-estimated nonlinear term was the smallest, and the instrument was noticed to be particularly sensitive until its local oscillator failed after only two years of operation. On the other hand, the $T_W$ coefficient for *NOAA-6* is virtually zero, indicating the original NESDIS calibration coefficients are evidently quite accurate.

To test the reproducibility of the time series as determined by the $T_W$ coefficients, we generated four new realizations in which each $T_W$ coefficient was adjusted by 0.003, or about 1 $\sigma_{\Delta a}$. The four tests were as follows: 1) all seven $T_W$ coefficients adjusted downward by 1 $\sigma_{\Delta a}$; 2) all adjusted upward; 3) alternating adjustments positive, negative, positive, etc.; and 4) alternating adjustments, negative, positive, negative, etc. The resulting trends ranged from $+0.033$ to $+0.043$ K decade$^{-1}$, or $+0.038 \pm 0.005$ K decade$^{-1}$, thus the possible error in the calculation of the coefficients has a rather small impact on the resulting global trends given the estimated error range of the global trends of $\pm 0.06$ K decade$^{-1}$ (described later).

In viewing the results of the four tests above we noticed that there was some reduction in $\sigma_{\Delta\text{Tdy}}$ when certain coefficients were adjusted. We take this opportunity a posteriori to adjust some of the $T_W$ coefficients to produce a slightly lower value of $\sigma_{\Delta\text{Tdy}}$ but without affecting the overall trend produced from the original $T_W$ coefficients. The adjustments to *NOAA-7*, *-8*, *-9*, and *-14* in this process are less than 1 $\sigma_{\Delta a}$.

### *e. Further description of merging procedure*

Note that for the time series discussed above, the values of $T_2$ used to produce $\Delta T_{m,n}$ for the $T_W$ coefficient calculations are already adjusted for orbit decay and diurnal drift. This is followed by the removal of the dependency on the instrument body temperature (as

TABLE 2. Error characteristics of the three temperature products for earlier version C and the present version D (Jan 1979–Dec 1998). Here, $\sigma_{error}$ (K) is $\sigma_{\Delta Tdy}/\sqrt{2}$ where $\sigma_{\Delta Tdy}$ is the standard deviation of the daily differences between two co-orbiting satellites. Also, S/E is the ratio of the variances for the daily average global anomalies of the two co-orbiting satellites vs their average difference. Trends are 1979–98 (in K decade$^{-1}$).

| | $\sigma_{error}$ C | $\sigma_{error}$ D | S/E C | S/E D | Trend C | Trend D |
|---|---|---|---|---|---|---|
| $T_{2LT}$ | 0.050 | 0.049 | 30 | 30 | +0.03 | +0.06 ± 0.06 |
| $T_2$ | 0.027 | 0.023 | 90 | 118 | +0.08 | +0.04 ± 0.06 |
| $T_4$ | 0.032 | 0.029 | 201 | 266 | −0.50 | −0.49 ± 0.10 |

measured by $T_W$) using the coefficient $a_m$. At this point, we have time series of daily, zonal anomalies for each of the nine satellites from which has been removed the spurious effects of orbit decay, diurnal drift and instrument body response. These anomalies are then merged by the calculation and removal of bias as described in Christy et al. (1998) for version C. Unlike version C, however, no intersatellite difference trends are explicitly removed in version D as we have already dealt with those based on the three effects above. Thus we have attempted to keep the minimization of error as unbiased as possible, not selecting one satellite over another in terms of preferential accuracy. For days on which two satellites report data (over 80%), the average of their anomalies is supplied for the completed time series.A comparison of statistical results is given in Table 2 for versions C and D. Note that the improvement in $T_2 \cdot D$ over $T_2 \cdot C$ indicates the $T_W$ information is quite useful in explaining the intersatellite variations. The lack of improvement between $T_{2LT} \cdot C$ and $T_{2LT} \cdot D$ is due to the procedure in version C by which one satellite was more or less forced to agree with its coorbiting counterpart, thus artificially creating smaller errors. The lower stratospheric channel results ($T_4 \cdot D$) for which the same procedures were applied are also given in Table 2 and as with $T_2 \cdot D$ show improvement over version C.

## 3. Discussion

### *a. Confidence estimates for annual anomalies and trends*

How accurate are the annual anomalies and trends of version D? We will show below that the 95% confidence interval (CI) for annual anomalies $T_2 \cdot D$ and $T_{2LT} \cdot D$ is about ±0.10 K and that the CI of the trend is ±0.06 K decade$^{-1}$. It is important to understand we are only describing one type of error in these ranges: measurement error. This is the error that answers the question, how well do we know the trend for the specific period observed by this system? We are not answering the different question, how well does the trend of this 20-yr period represent trends of all 20-yr periods? The latter question deals with ideas of statistical sampling and the representativeness of a relatively short period of observations (20 yr) and is illustrated below (see also Santer et al. 1999).

The interannual variations in these time series can be quite substantial as is indicated by the warmest temperatures of the 20-yr period occurring in 1998 being over 0.3 K warmer than any previous year due to the strong warm ENSO event. The trend (K decade$^{-1}$) for $T_{2LT} \cdot D$ through 1997 was −0.01, however, this became +0.06 with the addition of only one more year, 1998. So, even if the anomalies had been known perfectly, the trend of 1979–97 would have been a poor predictor for the trend for 1979–98.

The nature of the variability of the climate system requires extreme caution (i.e., large error bars) when using the trend as a tool for climate prediction (Santer et al. 1999). We state again here that our usage of the trend is as a metric to identify the effect of slight changes in adjustment procedures, and that our usage of the error bars is only to characterize measurement error. The experience of the single warm year of 1998 indicates changes of several hundredths per decade up or down in the trend should not be viewed as especially newsworthy given the fact the time series is only two decades long and affected by large climatic events of fairly short duration (Christy and McNider 1994).

It is the purpose of this section to estimate the actual confidence limits of the annual anomalies and trends in terms of measurement precision. (In this section we shall focus mostly on $T_{2LT} \cdot D$ because the potential errors are greater than for $T_2 \cdot D$.) There is no simple method by which such an estimate may be derived considering the fact several adjustments in creating the dataset are applied, each of which has some associated error. However, because there are several adjustments, the accumulated effect of their errors will tend to cancel to the extent they are random. The only true estimate of error is possible from accurate, independent data of the same quantity. Though their accuracy is even now being assessed, we shall use radiosonde comparisons for this purpose. Before exploring that test, we shall estimate the confidence limits from internal considerations.

#### 1) ESTIMATES FROM INTERNAL ERROR CHARACTERIZATIONS

As mentioned, it is difficult to fully describe all of the sources of error in satellite datasets such as $T_{2LT} \cdot D$ and $T_2 \cdot D$. For example, the instrument noise of a single Tb measurement has been estimated at ±0.25 K. However, with 15 000 to 22 000 Tb observations per day per satellite, this effect becomes minuscule when large-scale averages are performed. We are interested in the effects which impact longer period quantities such as the annual anomalies and trends.

The orbit decay calculation (for $T_{2LT} \cdot D$ only) is a straightforward geometric problem. We have applied the altitude loss effect using the altitude data from NOAA/NESDIS to standard atmospheric profiles of tropical,

midlatitude, and subarctic regions (winter and summer for the latter two). Ultimately, the magnitude of the effect depends on the lapse rate of the troposphere, which has not been shown to change over the past 20 yr by even 2% (though it is not, apparently, perfectly constant). Thus small decadal changes in the lapse rate are of little consequence to the quantity calculated here, which is dominated by a cooling of 50–100 K over the depth of the troposphere.

We have calculated the global mean orbit decay effect as $-0.10$ K decade$^{-1}$, which is based on assignments to latitude bands of the three standard atmospheric profiles. There is a possibility that the latitudinal assignments of the standard profiles could be a source of error. We tested various latitudinal assignments for the profiles including the complete substitution in the polar regions with the midlatitude profiles. The variations of profile application were no more than $\pm 0.01$ K decade$^{-1}$ on the overall trend of the time series.

The magnitudes of the diurnal adjustments are determined by an empirical calculation based on several million individual observations and is greatest for $T_{2LT}$ D. The cross-scan differences on which the effect is based were very consistent for each of the instruments. For example, the temperature drift of 1 h eastward movement for afternoon satellites in the tropical latitudes (the region with greatest impact on the global temperature) was found to be 0.15 K with 95% Cl of $\pm 0.031$ K based on the five afternoon orbiters over 16 yr. The accumulated effect of the diurnal adjustments on the trend of $T_{2LT}$ D was 0.03 K decade$^{-1}$. As can be seen, the accumulated adjustments would require significant errors of the order of 33% to translate into an error of $\pm 0.01$ K decade$^{-1}$.

The $T_w$ coefficient values were tested by generating the $T_2 \cdot D$ time series from combinations of errors of order 1 $\sigma_{\Delta a}$ (0.0033). The trends in $T_2 \cdot D$ from these tests show only minor impact on the trend ($\pm 0.005$ K decade$^{-1}$) in the unlikely event that all coefficients were in error by 1 $\sigma_{\Delta a}$ of the same sign. However, there is other information in these test runs that is useful. The standard deviation of the annual global temperature anomaly differences ($\sigma_{\Delta ann}$) among the tests is 0.0054 K for $T_2 \cdot D$. However, $\sigma_{\Delta ann}$ ranges from 0.0130 K (1994) to 0.0004 K (1991) when checking the reproducibility of individual years. Assuming that 0.0130 K represents a realistic $\sigma_{\Delta ann}$ for all years (or 95% Cl of 0.025 K for annual anomalies) in the time series, we estimate that the 95% Cl for the trend is $\pm 0.014$ K decade$^{-1}$. Scaling these values to $T_{2LT}$ D we arrive at estimates for $\sigma_{\Delta ann}$ of 0.029 K (95% Cl of 0.057 K) and 95% Cl trend error of $\pm 0.058$ K decade$^{-1}$.[5]

Note that in the above estimates we have taken $\sigma_{\Delta ann}$ from the worst-case year of the 20 years to represent all years, a somewhat conservative approach. It should also be noted that errors contained in the orbit decay and diurnal effect may be compensated by the calculation of $T_w$ coefficients and thus reduced. However, we still do not know if some type of systematic error might be hidden in the instruments which would then be undetectable in intersatellite comparisons. The most desirable test for determining measurement error bounds is to compare against independent measurements of the same atmospheric layer monitored by the MSUs and is discussed in the following section.

### 2) Precision estimates from collocated radiosondes

We may estimate the precision of the annual anomalies and global trend of $T_{2LT}$ D in two ways. First, we calculate the monthly anomalies of simulated $T_{2LT}$ from 97 U.S.-controlled radiosonde stations in the western Northern Hemisphere for 1979–94.[6] This set of sondes was chosen because of their consistency in instrumentation and observational methodology for the time period examined (Luers and Eskridge 1998). Anomalies of $T_{2LT}$ D from the 2.5° gridded monthly maps of MSU temperatures are then collocated with the radiosonde sites with anomalies averaged to a 97-station composite. Statistics were then calculated on the composited anomalies (except individual trend comparisons.) The composite signal-to-noise ratio for monthly anomalies is 13.2 (variance of differences = 0.0185 K$^2$, variance of anomalies = 0.2440 K$^2$) and for annual anomalies is 11.0 (variance of differences = 0.0037 K$^2$, variance of anomalies = 0.0407 K$^2$).

The results appear in Fig. 9 in which the annual (monthly) correlation of anomalies is 0.97 (0.96) and the trends agree within 0.005 K decade$^{-1}$. (The trend of the composite anomalies for these 97 grids is $+0.16$ K decade$^{-1}$, indicating the geographic distribution of these stations is not representative of the globe whose trend was $-0.01$ K decade$^{-1}$ during this period.) It is important to note that during the construction of the grid point datasets, which contain over 10 000 grids, there is no knowledge of the anomalies and trends of these particular 97 locations; that is, there is no way that we might manipulate the construction procedure to preferentially induce agreement between the MSU and these radiosondes. This test is a means to select 97 ''random''

[5] The scale factor for standard errors of $T_{2LT}$ D vs $T_2 \cdot D$ anomalies (from daily to annual) consistently appears as between 2.0 and 2.4 (see Tables 1 and 2). However, the error estimation for the trend of $T_{2LT}$ D here is performed using the 95% Cl applied to the individual annual anomalies.

[6] Station representation: 6 in central and western Pacific; 11 in Mexico, Caribbean, and Bermuda Islands; 20 in Canada, Alaska, and Iceland and the remainder in the conterminous United States. Comparisons beyond 1994 should be made with caution as the U.S. sonde type began a shift from VIZ to Vaisala in 1995, the latter being systematically warmer than the former (Robert Eskridge 1999, personal communication).

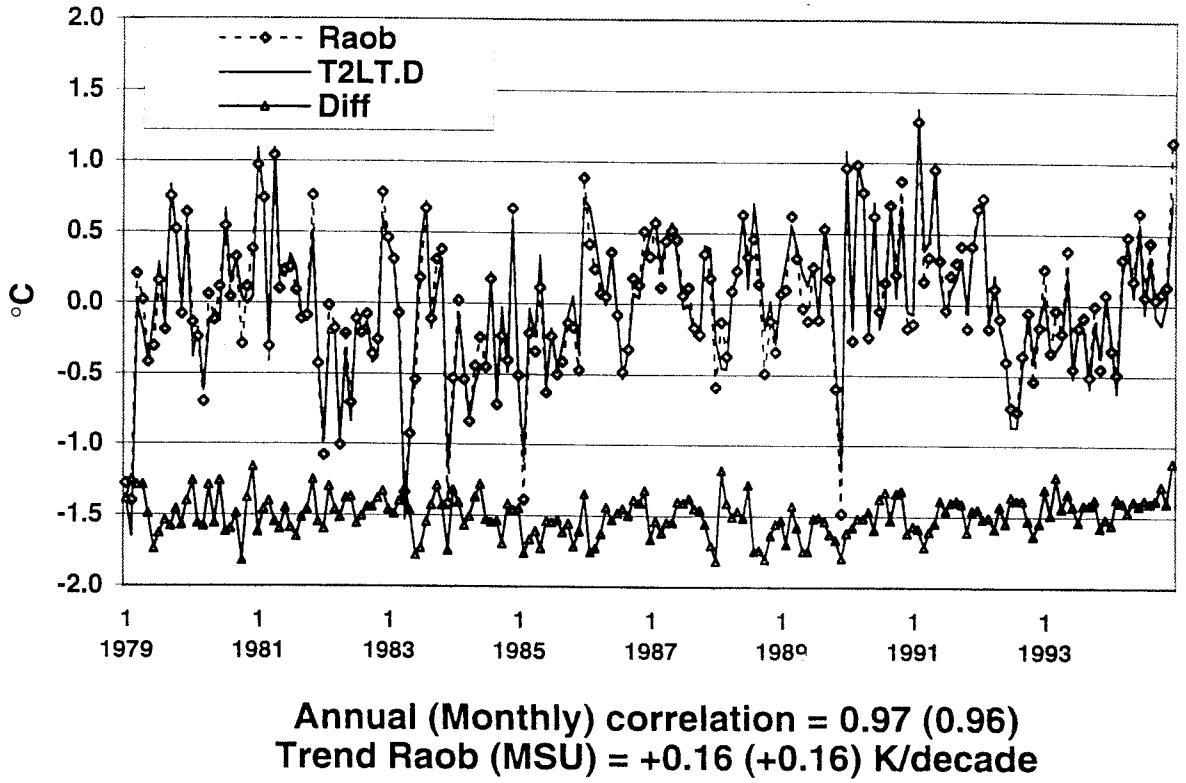


FIG. 9. Independent comparison between simulated values of $T_{2LT}$ from 97 U.S.-controlled radiosonde stations (1979–94) using radiative transfer theory and gridpoint values of MSU $T_{2LT}$_D. Station representation: 6 in central and western Pacific; 11 in Mexico, Caribbean, and Bermuda Islands; 20 in Canada, Alaska, and Iceland; and the remainder in the conterminous United States. The difference time series is offset by −1.5 K for clarity.

grids (or subsets as indicated below) from the MSU datasets for comparison. With a sample this large, the statistics will be quite useful in estimating errors.

To show that the excellent agreement between the MSU and the average of these 97 radiosonde stations is not a coincidence, we select three geographically distinct subregions and perform the same statistical comparison. The first set contains 20 stations[7] north of 50°N, the second utilizes 12 stations[8] in the northcentral United States, and the third comparison employs 24 stations[9] south of 30°N latitude (but including Bermuda). The results are given in Table 3.

The selection of the high latitude and north-central U.S. sites in one sense is a means to test the impact that surface emissivity variations might have on the MSU anomalies since about 20% of the $T_{2LT}$ signal originates from the surface over land (about 10% over ocean). Specifically, wet conditions tend to reduce emissivity and thus cause a cooling in the satellite-observed Tb. In the high latitudes this might occur as variations in the thawing and freezing of the tundra landscapes as well as variations in snow cover. In the north-central United States, variations in wetness due to flooding or droughts, snow cover, icing of the Great Lakes, etc., might also be a factor.

The radiosonde-computed Tb utilizes a constant value of emissivity, while the MSU observes the effects of the true emissivity. The results for both independent sets of analyses in Table 3 indicate that emissivity variations must be very small since the actual differences in the observation systems are small. Indeed, the standard deviation of monthly anomalies for the radiosonde-simulated MSU was 1.579 K, while that of the observed MSU was 1.427 K in the north-central United States. If surface emissivity variations were a factor, one would expect the observed MSU to have the greater standard deviation. The comparisons of the three subregions list-

[7] In Alaska: Barrow, Kotzebue, Nome, Bethel, McGrath, Fairbanks, Anchorage, St. Paul Island, Cold Bay, King Salmon, Kodiak, Yakutat, Annette. In Canada: Goose, Moosonee, Big Trout Lake, Cambridge, Fort Smith, and Inuvik. In Iceland: Keflavik.

[8] Salem, IL; Dodge City, KS; Topeka, KS; Peoria, IL; Omaha, NB; Green Bay, WI; Huron, SD; St. Cloud, MN; Sault Ste. Marie, MI; International Falls, MN; and Bismarck, ND.

[9] Key West, FL; Lake Charles, LA; Brownsville, TX; Del Rio, TX. In Caribbean: San Juan, PR; St. Martin, Christ Church, Piarco, Curacao, and Roberts, Cayman Island; St. George, Bermuda. In Pacific: Midway Is., Truk Is., Ponape, Majuro, Lihue, HI and Hilo, HI. In Mexico: Chihuahua, Guaymas, Monterrey, Mazatlan, Guadalajara, Merida, Manzanillo, and Vera Cruz.

TABLE 3. Comparison of composited anomalies of groups of U.S.-controlled radiosonde stations for 1979–94.

| Stations | Radiosonde trend (K decade$^{-1}$) | MSU $T_{2LT}$ trend (K decade$^{-1}$) | Correlation Ann, Mo | Std. dev. Ann. diff. (K) |
|---|---|---|---|---|
| 97 W. N. Hemisphere | +0.16 | +0.16 | 0.97, 0.96 | 0.061 |
| 20 High latitude | −0.01 | +0.01 | 0.99, 0.99 | 0.048 |
| 12 North-central (United States) | +0.21 | +0.23 | 0.98, 0.98 | 0.110 |
| 24 Tropical/subtropical | −0.03 | −0.05 | 0.94, 0.89 | 0.049 |

ed in Table 3 provide substantial evidence that the MSU $T_{2LT}$ data, as corrected here, have excellent interannual and long-term stability.

The standard deviation of the individual, collocated, site-by-site (i.e., not composited) difference trends of the 97 sites is 0.114 K decade$^{-1}$, while the standard deviation of the radiosonde trends is 0.277 K decade$^{-1}$, giving a signal-to-noise ratio of 5.8 for trends. For the following calculation, we shall assume that all of the variance of the difference trends between the radiosondes and the MSU is due to errors in the MSU data. In doing so, we assume the radiosondes are perfect and we neglect the fact the MSU values are 2.5° gridded values, while the radiosondes monitor a narrow column of air through which the balloon ascends. In addition, we assume no impact from the time of observation differences; that is, the anomalies of monthly average of 0000 UTC and 1200 UTC radiosonde observations will be virtually identical with anomalies of the monthly averages produced by satellites which measure at approximately 1930/0730 and 1400/0200 UTC.

Let us now assume there are 25 spatial degrees of freedom (DOF) in the 2.5° × 2.5° global geographical distribution of $T_{2LT}$.D from which global mean anomalies and trends are determined. This is a reasonable value as Hurrell and Trenberth (1997) estimate 8 DOFs for the one-third of the global area from 20°S to 20°N. In addition, we may assume that the aggregate of the 97 sites in the western Northern Hemisphere represents about one-sixth of the total global DOF. The 95% Cl for monthly global anomalies here would become ±0.119 K [$=1.96 \times (0.136/\sqrt{5})$]. Similarly for annual anomalies the 95% Cl is ±0.053 K for $T_{2LT}$.D, a value very similar to that estimated from internal factors in the previous section.

Continuing in the same manner with the trend consideration, we estimate a 95% Cl for the global trend of $T_{2LT}$.D to be ±0.046 K decade$^{-1}$ [$= 1.96 \times (0.114/\sqrt{24})$]. (The full number of DOFs are used here because the standard deviation of trend differences was calculated on a site-by-site basis.) Since this test utilizes the completed version of the MSU data versus independent data, this is, in effect, an estimate of the net error of the combined errors which arise (and perhaps cancel to some extent) from all of the processes involved in the dataset construction.

### 3) PRECISION ESTIMATES FROM ''GLOBAL'' RADIOSONDE DATASETS

The test above is confined to those locations, which have relatively high quality time series of radiosonde data. It is possible that the MSU data may happen to be more (or less) precise in these sites and their subgroupings than for the globe as a whole, even though the MSU dataset has systematic procedures applied everywhere. A more appropriate test would be a comparison with a truly global distribution of radiosondes. Unfortunately, large areas of the globe are not sampled (hence global is in quotes), though this is not such an extreme a problem as for other quantities (i.e., surface temperatures) because there are fewer DOFs in the deep-layer atmospheric measurements due to the strong spatial coherency of the quantity (Hurrell and Trenberth 1996; Wallis 1998). However, the lack of geographic coverage is indeed a shortcoming of these datasets and is compounded by serious discontinuities in several radiosonde time series (Gaffen 1994; Christy 1995; Parker et al. 1997; Luers and Eskridge 1998). The annual anomalies of the datasets described below are displayed in Fig. 10.

For ''global'' averages, Angell (1988) utilizes a relatively small number of stations (63), but whose data are mostly continuous. The quantity reported is the 850–300-hPa thickness temperature derived from the geopotential heights of the two levels (Trenberth and Olson 1991; Christy 1995). The data are binned into seven latitudinal bands and averaged into zonal means from which ''global'' values are determined. This two-step averaging procedure is a technique designed to provide more realistic geographic representation of sparsely observed regions. The standard deviation of annual anomalies ($\sigma_{ANGann}$) is 0.218 K, and thus contains more variance than $T_{2LT}$.D ($\sigma_{MSUann} = 0.166$ K). The standard deviation of the annual anomaly differences between Angell and MSU $T_{2LT}$ ($\sigma_{\Delta Angell}$) is 0.087 K. Angell's 1979–98 trend is +0.02 K decade$^{-1}$ and the $T_{2LT}$ versus Angell correlation is 0.93.[10]

A series of radiosonde-based datasets of annual global

[10] Angell has found that a few of the tropical stations appear to have spurious cooling so that the global trend, without these sites, becomes about 0.02 K decade$^{-1}$ warmer (J. Angell 1997, personal communication).

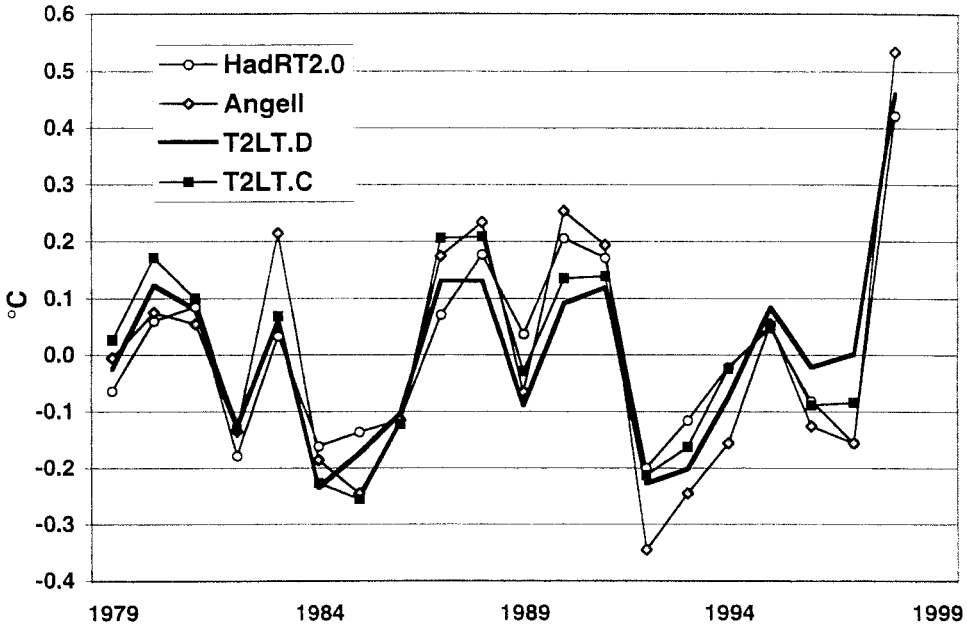


FIG. 10. Time series of annual anomalies of globally estimated temperature of the troposphere (see text). Note the increase in temperature in 1984–85 from version C to version D of $T_{2LT}$.

anomalies is available for which the atmospheric weighting profile of $T_{2LT}$ has been applied to create simulated MSU temperatures. This dataset is generated at the Hadley Centre of the United Kingdom Meteorology Office and is based on data described in Parker et al. (1997) for which about 400 stations are employed. The Hadley Centre produces a series of versions of the analyses of radiosonde data and we use here the dataset known as HadRT2.0. After preliminary checks for hydrostatic and horizontal consistency are made, the station data for HadRT2.0 are binned into 5° lat × 10° long grids. No infilling of data void grids and no corrections relative to the MSU are applied.

We find for HadRT2.0 that $\sigma_{\text{HADann}}$ of the simulated $T_{2LT}$ is 0.159 K, $\sigma_{\Delta\text{HAD}} = 0.071$ K, the trend is +0.05 K decade$^{-1}$ and the correlation versus MSU is 0.91. Assuming HadRT2.0 better represents the global anomaly differences than Angell, we estimate a 95% Cl measurement error range for the MSU trend as ±0.058 K decade$^{-1}$, given the assumption that the HadRT2.0 dataset is the standard and all of the error lies in the MSU measurements. Assuming part of the difference error lies with HadRT2.0, this confidence interval would accordingly be reduced.

We may estimate the 95% Cl for the measurement error of the trend of $T_{2LT}$ D, based on the various analyses above, as ±0.06 K decade$^{-1}$. The results of (i) internal analyses, (ii) the comparison with collocated radiosondes, and (iii) the comparisons with ''global'' radiosonde datasets all agree at this level of precision of measurement error. The results for assessing the measurement error range for individual annual global anomalies is less certain. In the internal analysis, we found the 95% Cl was ±0.058 K for annual global anomalies in $T_{2LT}$ D, and then using collocated radiosondes we estimated the value to be ±0.053 K. In comparison with Angell and HadRT2.0 ''global'' datasets we calculate a standard deviation of annual differences versus MSU of 0.087 K and 0.071 K, respectively.

Given the sparseness of the ''global'' radiosonde data sets, the slight difference in the quantity observed (in Angell's case), and their inclusion of some stations with temperature shifts due to instrumentation and procedural changes, we have less confidence that these values are as useful in documenting errors in the MSU anomalies. We note, for example, that the standard deviation of the differences between Angell and HadRT2.0 is 0.089 K, a value greater than that calculated when comparing each individually with the independent MSU data. Given the confidence estimates from the internal variations and the comparisons with systematic collocated radiosondes it is very likely that the 95% Cl for annual $T_{2LT}$ anomalies is less than 0.10 K.

We apply the results above to $T_2 \cdot D$ and also estimate the 95% Cl for least squares trend of global annual anomalies as ±0.06 K decade$^{-1}$. Though the internal consistency of $T_4$ anomalies (i.e., intersatellite signal to noise level; Table 2) is exceptionally high, we are unable to estimate the same level of confidence with radiosonde comparisons. This is due to the fact errors in radiosonde soundings tend to increase with elevation where instrumentation changes can have exceptionally large impacts (e.g., about 3 K for the lower stratosphere in the Australian region when the instrument was changed from Phillips to Vaisala, Parker et al. 1997.) Thus, for trend

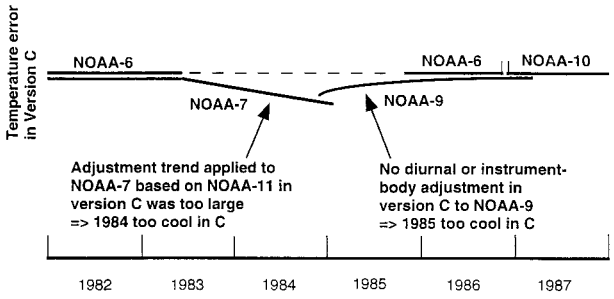


FIG. 11. Schematic diagram of the errors in version C due to *NOAA-7* and *-9* (see text). In version D, the last portion of *NOAA-7* and the first portion of *NOAA-9* are warmer.

precision will shall only estimate a 95% C.l. of $\pm 0.10$ K decade$^{-1}$ for $T_4$.

### b. Comments on version C

Many comparison studies between $T_{2LT}.C$ and radiosondes or global reanalyses indicated that this dataset appeared to contain reliable trends (e.g., Christy 1995; Christy et al. 1997; Parker et al. 1997; Stendel and Bengtsson 1997; Christy et al. 1998; Pielke et al. 1998). Considering the absence of the 0.10 K decade$^{-1}$ orbit decay correction in C, how could this agreement occur? In addition, a comparison of trends between versions D and C reveal differences that are fairly small overall, even after the orbit decay correction is included (a net increase in trend for $T_{2LT}$ of only 0.03 K decade$^{-1}$). How is this explained?

There were two main factors that had a more-or-less compensating influence against the orbit decay effect missing in version C, which lead to C's apparent accuracy. First, in version C we assumed there was a spurious temperature trend in *NOAA-7* due to LECT drift. The trend subsequently removed was based on the trend determined from *NOAA-11* in its comparison with *NOAA-10* and *-12*. It was not apparent at the time that the *NOAA-11* trend was a combination of both the diurnal effect and the instrument body effect. In *NOAA-11* the latter effect was twice as large (3.5%, Table 1) than determined for *NOAA-7* (1.8%), thus the trend removed in C for *NOAA-7* was too large. The result was an anomaly for the period around 1984, dependent on *NOAA-7,* which was too cool (Fig. 11).

Second, since we did not adjust for the interannual component of the instrument body effect, we did not remove any warming trend from *NOAA-9* based on its $T_W$, nor did we remove any warming due to diurnal considerations, thus creating an anomaly for 1985 which was also too cool (Fig. 11). (*NOAA-9* was calibrated into the time series during its last year when its overlap with reactivated *NOAA-6* and *NOAA-10* occurred.) The ''corrections'' applied to *NOAA-7* and the lack thereof to *NOAA-9* in version C created a spuriously cool period in 1984–85, which was justifiably noted as being suspicious in comparisons with surface data in Jones et al. (1997). However, because the anomalies subsequent to 1985 were dependent on *NOAA-6* and *NOAA-10,* the comparison in Christy et al. (1998) revealed excellent agreement with other temperature datasets for the 1986–87 anomalies versus those in 1981–82 when a check was performed on the stability of *NOAA-6.* However, no comparison for 1984–85 was included at the time.

The depressed temperature values in 1984–85 caused a spurious positive tilt to the overall trend in $T_{2LT}.C$. Thus, a large portion of the reduction in trend due to the lack of orbit decay corrections in $T_{2LT}.C$ was compensated by this inadvertent depression of the values for 1984–85. When the proper corrections are applied for *NOAA-7* and *NOAA-9,* the temperatures in 1984–85 are increased, and the overall trend of $T_{2LT}.D$ declines in value simply due to these corrections (Fig. 10).

### c. Vertical coherence of trends

As we have seen, the value of the trend for $T_2 \cdot D$ is less than that of $T_2 \cdot C$, while that of $T_{2LT}.D$ is more positive than $T_{2LT}.C$. This occurs because $T_2 \cdot C$ also contained the spuriously depressed temperatures of 1984–85 (as did $T_{2LT}.C$), which are now corrected in $T_2 \cdot D$, thus decreasing its overall trend. However, there is essentially no impact from the orbit decay correction so the net effect on the trend of $T_2 \cdot D$ of the corrections in version D is to reduce the trend (by among other things warming up $1984 \cdot 85$) relative to version C.

We now see that there is a systematic decrease of trends as the weighting functions increase in altitude with $T_{2LT}.D$ more positive than $T_2 \cdot D$, which in turn is more positive than $T_4$ (not shown). Version C of these datasets presented the curious result that $T_2$ maintained a more positive trend than either $T_{2LT}$ or $T_4$, and this was the source of criticism of one aspect of version C (Hurrell and Trenberth 1998; Wentz and Schabel 1998). With version D of the dataset, the inconsistency in vertical coherence of trends found in C is removed.

However, one must note that for specific periods, $T_2$ may indeed show greater warming than $T_{2LT}$ due to the differential effects of volcanic eruptions or ENSO events. For example, when Mt. Pinatubo erupted in June 1991, the tropical stratosphere warmed suddenly by 2.0 K while the tropical troposphere cooled by 0.6 K. Since $T_2$ includes emissions from the stratosphere, the volcanic impact warmed $T_2$ relative to $T_{2LT}$. Thus for particular periods in which volcanic eruptions are included near the end of the time series it is possible to observe trends for which $T_2$ is more positive than $T_{2LT}$. For example, we show the monthly global anomalies of $T_2 \cdot D$ and $T_{2LT}.D$ in Fig. 12. It is evident that $T_2$ reveals warming relative to $T_{2LT}.D$ following 1982 and 1991, years in which major volcanic eruptions occurred. Also shown is the relative warming of $T_2$ in the 1998 ENSO event where upper tropospheric temperatures were char-

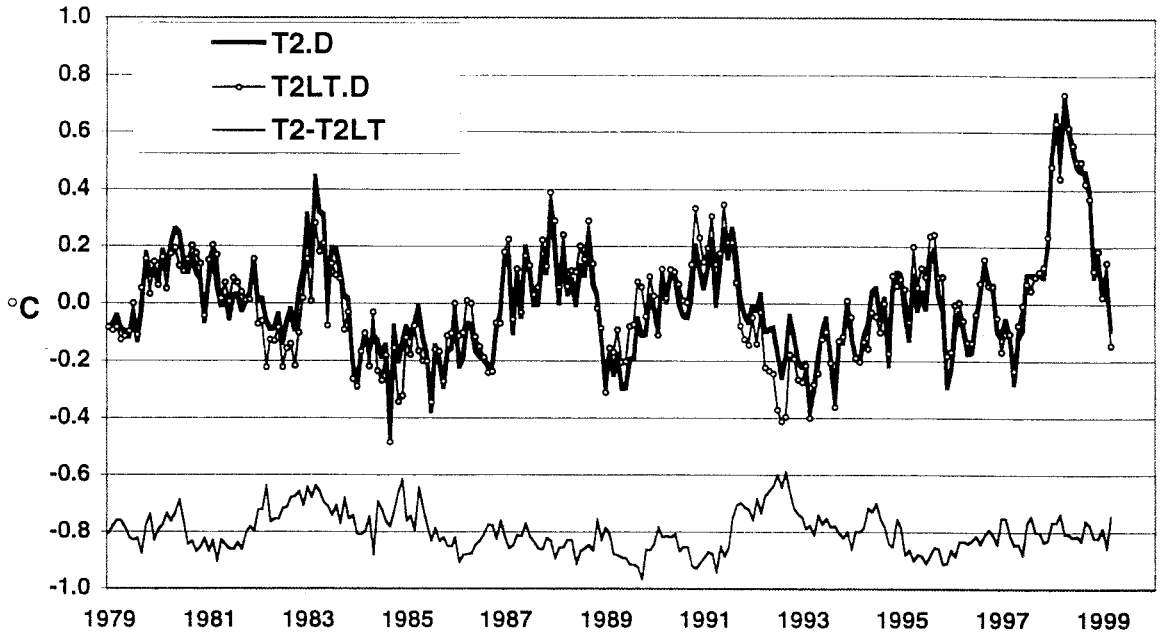


FIG. 12. Monthly global mean anomalies of $T_2 \cdot D$ and $T_{2LT} \cdot D$ with their differences. Though the difference trend ($T_2 \cdot D$ minus $T_{2LT} \cdot D$) is negative, note the warming in $T_2$ relative to $T_{2LT}$ following volcanic episodes in 1982 (El Chichon) and 1991 (Mt. Pinatubo). The difference time series is offset by $-0.8$ K for clarity.

acterized by warmer anomalies than observed in the lower troposphere as seen by $T_{2LT}$.

## 4. Concluding remarks

When we began the construction of a time series of global temperatures from sequentially launched satellites in 1989, we could not have known the many factors that would ultimately come to our attention which in some way affect the time series. Some studies, motivated by the lack of agreement between trends of surface records and those of the MSU troposphere, offered several suggestions for the differences, which we have addressed elsewhere (Spencer et al. 1996; Christy et al. 1997, 1998). More recently, the discovery by Wentz and Schabel of one significant and necessary adjustment due to satellite orbit decay, the finding of Mo (1995) regarding NESDIS calibration errors, and our discoveries related to diurnal effects and instrument heating all lead to the present effort.

The important issue in this paper is that we have shown the instrument body temperature, as indicated by the warm target temperatures, is a key piece of information that explains considerable variance of the differences between coorbiting satellites. Still, even with all the adjustments applied here, the instruments do not produce perfect agreement, but we are able at least to quantify those differences. In addition, we are able to state that the decadal trend for 1979–98 for $T_2 \cdot D$ ($T_{2LT} \cdot D$) is $+0.04$ ($+0.06$) K decade$^{-1}$ $\pm$ 0.06, but that these quantities have little predictive value because of the shortness of the time series. (In particular, the error range here is limited strictly to measurement error only and does not include uncertainties due to sampling of a relatively short time series.) We recognize other factors unknown to us at the present may yet be discovered which will aid in further improvements of the products described here, thus releases beyond version D should be infrequently expected in which the error range of the trend may be reduced.[11]

*Acknowledgments.* The research presented herein was supported by NOAA/NASA Enhanced Datasets for Analysis and Applications (NA97AANAG0247), with a portion of the lead author's support provided by DOE/NIGEC through the Southeast Regional Center at the University of Alabama, Tuscaloosa (Research Agreement No. 920287-ALA by DOE Cooperative Agreement No. DE-FC03-90ER61010 through the University of California). Financial support does not constitute an endorsement by DOE of the views expressed in this article. We gratefully acknowledge the following who supplied global, annual anomalies from their radiosonde datasets: J. Angell (NOAA); S. Brown and D. Parker (Hadley Centre, UKMO). In addition, the reviewers of this manuscript were extremely helpful in suggesting improve-

[11] For example, it would be desirable to determine new nonlinear calibration coefficients for each MSU based on laboratory and on-orbit checks as demonstrated by Mo (1995). Our preliminary investigation indicates that for many of the MSUs, the existence of the original laboratory data may be inaccessible so that we will continue to rely on the empirical adjustments based on $T_w$ as described here.

ments, particularly one reviewer who developed the justification of the warm target coefficients outlined in Eqs. (3)–(12).

## REFERENCES


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

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., 1995: Temperature above the surface layer. *Climatic Change,* **31,** 455–474.

——, and R. T. McNider, 1994: Satellite greenhouse warming. *Nature,* **367,** 325.

——, 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 W. D. Braswell, 1997: How accurate are satellite ''thermometers''? *Nature,* **389,** 342–343.

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

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

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

——, and ——, 1997: Spurious trends in satellite MSU temperatures from merging different satellite records. *Nature,* **386,** 164–167.

——, and ——, 1998: Difficulties in obtaining reliable temperature trends: Reconciling the surface and satellite microwave sounding unit records. *J. Climate,* **11,** 945–967.

Jones, P. D., T. J. Osborn, T. M. L. Wigley, P. M. Kelly, and B. D. Santer, 1997: Comparisons between the microwave sounding unit temperature record and the surface temperature record from 1979 to 1996: Real differences or potential discontinuities. *J. Geophys. Res.,* **102,** 30 135–30 145.

Luers, J. K., and R. E. Eskridge, 1998: Use of radiosonde temperature data in climate studies. *J. Climate,* **11,** 1002–1019.

Mo, T., 1995: A study of the Microwave Sounding Unit on the *NOAA-12* satellite. *IEEE Trans. Geosci. Remote Sens.,* **33,** 1141–1152.

Nicholls, N., G. V. Gruza, J. Jouzel, T. R. Karl, L. A. Ogallo, and D. E. Parker, 1996: Observed climate variability and change. *Climate Change 1995,* J. T. Houghton et al., Eds., Cambridge University Press, 572 pp.

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.

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,** 6305–6338.

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

——, ——, and ——, 1996: Analysis of 'Examination of ''Global Atmospheric Temperature Monitoring with Satellite Microwave Measurements'''. *Climatic Change,* **33,** 477–389.

Stendel, M., and L. Bengtsson, 1997: Toward monitoring the tropospheric temperature by means of a general circulation model. *J. Geophys. Res.,* **102,** 29 779–29 788.

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 Publishers, 249–260.

——, J. R. Christy, and J. W. Hurrell, 1992: Monitoring global monthly mean surface temperatures. *J. Climate,* **5,** 1405–1423.

Wallis, T. W. R., 1998: A subset to core stations from the Comprehensive Aerological Reference Dataset. *J. Climate,* **11,** 272–282.

Wentz, F. J., and M. Schabel, 1998: Effects of satellite orbital decay on MSU lower tropospheric temperature trends. *Nature,* **394,** 361–364.