Analysis of the Merging Procedure for the MSU Daily Temperature Time Series

JOHN R. CHRISTY

ESSL, GHCC, University of Alabama in Huntsville, Huntsville, Alabama

ROY W. SPENCER

NASA/Marshall Space Flight Center, Huntsville, Alabama

ELENA S. LOBL

ESSL, GHCC, University of Alabama in Huntsville, Huntsville, Alabama

(Manuscript received 10 January 1997, in final form 30 October 1997)

ABSTRACT

The merging procedure utilized to generate homogeneous time series of three deep-layer atmospheric temperature products from the nine microwave sounding units (MSUs) is described. A critically important aspect in the process is determining and removing the bias each instrument possesses relative to a common base (here being *NOAA-6*). Special attention is given to the lower-tropospheric layer and the calculation of the bias of the *NOAA-9* MSU and its rather considerable impact on the trend of the overall time series. We show that the bias is best calculated by a direct comparison between *NOAA-6* and *NOAA-9*, though there other possible methods available, and is determined to be +0.50°C. Spurious variations of individual MSUs due to orbital drift and/or cyclic variations tied to the annual cycle are also identified and eliminated. In general, intersatellite biases for the three instruments that form the backbone of the time series (MSUs on *NOAA-6*, *-10* and *-12*) are known to within 0.01°C.

After slight modifications in the treatment of the bias, drift-error, and cyclic fluctuations, the authors produced a time series in which the decadal trend is +0.03°C warmer than previously reported for the lower troposphere. Because they are of much higher precision, the midtropospheric and lower-stratospheric products are only slightly affected by alterations to procedures applied in this study.

Recent suggestions that spurious jumps were present in the lower-tropospheric time series of earlier versions of the MSU data based on SST comparisons are addressed. Using independent comparisons of different satellites, radiosondes, and night marine air temperatures, no indication is found of the presence of these ''spurious'' jumps.

1. Introduction

Microwave sounding units (MSUs) on National Oceanic and Atmospheric Administration (NOAA) polar-orbiting satellites have monitored the intensity of radiation from (primarily) atmospheric oxygen since the first was launched in December 1978. The overall stability of the instruments and the robustness of the measurements have provided a means to create relatively long time series of bulk temperatures for atmospheric layers several kilometers deep. With the time series lengthening, concern is naturally raised regarding the long-term stability of the dataset since it represents a merged product of several independent instruments.

One of the MSU products, the temperature of the lower troposphere, has been known as T_{2R} because it represents a retrieval of channel 2 view-angle temperatures (hence ''2R''). We now anticipate producing other retrievals using channel 2 data that will represent different layers besides the lower troposphere. Therefore we shall from this point forward refer to the ''channel 2 lower-tropospheric retrieval'' as T_{2LT} rather than T_{2R}.

The values of T_{2LT} require rigorous assessment because they are most widely utilized. For example, global values of MSU T_{2LT} have been used in assessments of climate variation and change because they are physically closest to the surface air layer, the temperature of which is of great concern to life on the planet (e.g., Nicholls et al. 1996). In addition, we focus here on the MSU T_{2LT} because it contains the greatest measurement uncertainty of all publicly available Spencer–Christy MSU temperature products.

Several factors come into play when an attempt is made to combine separate time series of MSU brightness temperatures from multiple satellites into a single, ho-

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

mogeneous time series. In this paper we shall report on revised procedures for processing the MSU data and, in particular, we discuss the more subtle effects that occur when selecting one particular method of merging over another. It is critical to document this procedure as there have been questions raised concerning the long-term stability of the T_{2LT} time series (e.g., Hansen et al. 1995) and such investigations have helped to spur the present study. Confidence in variations and trends of the midtroposphere (T_2) and lower stratosphere (T_4) is much higher because the signal-to-noise magnitudes are much greater than for T_{2LT}, sometimes more than an order of magnitude (Spencer and Christy 1992a; Spencer and Christy 1993; Christy and Drouilhet 1994).

Users should be aware that the manner by which the MSU data are stabilized and merged is under constant scrutiny. When a possible error is discovered, we thoroughly examine its cause and fashion a solution if possible. Fortunately, there are independent data (radiosondes) against which the MSU data may be compared and this gives us added confidence in the final results. This constant scrutiny inevitably leads to the release of revised versions of the datasets and each new version generally reveals some change in the overall long-term trend or variability. For example, when the orbital drift error in *NOAA-11* was identified and removed, the overall trend of the T_{2LT} time series was decreased by 0.03°C decade^{-1} (Christy et al. 1995). We realize that the presence of multiple versions of the MSU data can be a complication for users of the data, and our intention is to offer revisions only when clearly necessary. To this point, fortunately, the many studies that have utilized the Spencer–Christy MSU products have not been affected by the slight alterations that have been adopted. The dataset reported herein is the third version of the temperature data and since November 1997 has resided on the anonymous ftp site with filenames containing the extension ''c1,'' replacing the previous versions with filenames with extension ''b'' or ''c.''

In the following sections, we shall describe in greater detail than before what the temperatures represent (section 2); examine some of the critical aspects on how the lower-tropospheric temperature time series is constructed including filtering, perturbation annual cycles, and bias errors (sections 3, 4, and 5); and specifically examine the importance of the bias calculation of *NOAA-9* (section 6). We devote section 7 to the assessment of the reproducibility of the intersatellite biases, which are critical in constructing a reliable time series, and include comparisons with large radiosonde datasets. Section 8 will briefly report on the impacts of the new procedures on the time series of the midtroposphere and lower-stratospheric temperatures. In section 9 we shall address recent suggestions that spurious jumps appear in the MSU time series.

In concluding this study (section 10), we shall show that a slight adjustment to the overall time series is warranted and that the decadal trend of T_{2LT} is slightly warmer, by about +0.03°C, than that of the previous version b. We shall also include suggestions for future satellite missions that will enable better climate monitoring capability by reducing the types of uncertainty with which we have had to deal in the MSU time series. To streamline the following text, we shall use initials for the many repeatedly cited publications so that C will be Christy, S will be Spencer, G will be Grody, and M will be McNider. The year of publication will be concatenated to the initials (e.g., SC92a is Spencer and Christy 1992a).

2. MSU temperatures

The MSU monitors the upwelling microwave emissions from the earth at four frequencies (or channels) between 50 and 60 GHz and processes the measured intensity into digital counts. The conversion of counts to absolute brightness temperatures is basically determined through an interpolation scheme anchored on the cold side by the counts measured when viewing cold space and on the warm side by the counts recorded from the view of the onboard hot target. Known temperatures of cold space (2.7 K) and the hot target (monitored by platinum-resistance thermometers) provide the information necessary to convert digital counts into brightness temperatures. The observed digital counts measured when the MSU views the earth are then converted to brightness temperatures through this relationship (Spencer et al. 1990, hereafter SCG90).

The present MSU temperature datasets utilize observations from nine different spacecraft. Each MSU's four channels represent different broad vertical soundings of the atmosphere and, for two channels (1 and 2), detectable surface emissions. Each channel observes 11 earth views during each 26-s cross-track scan where footprints 1 and 11 are the left and right limb positions and footprint 6 the nadir (SCG90). The two calibration targets are also viewed on each scan. Because the views toward the limb become more oblique relative to the earth, the instrument senses emissions at higher altitudes than at nadir (SC92a). With four channels, each with essentially six view angles, it is possible to construct many retrieval schemes by linearly combining the view-angle observations of one or more channels to concentrate on a certain layer of atmospheric emissions (Spencer and Christy 1992b, hereafter SC92b; Goldberg and Fleming 1995).

The vertical atmospheric weighting functions of the products we produce for T_{2LT}, T_2 (channel 2, midtroposphere), and T_4 (channel 4, lower stratosphere) are displayed in Fig. 1. The methods by which the channel view angles are combined for these three products are the following:

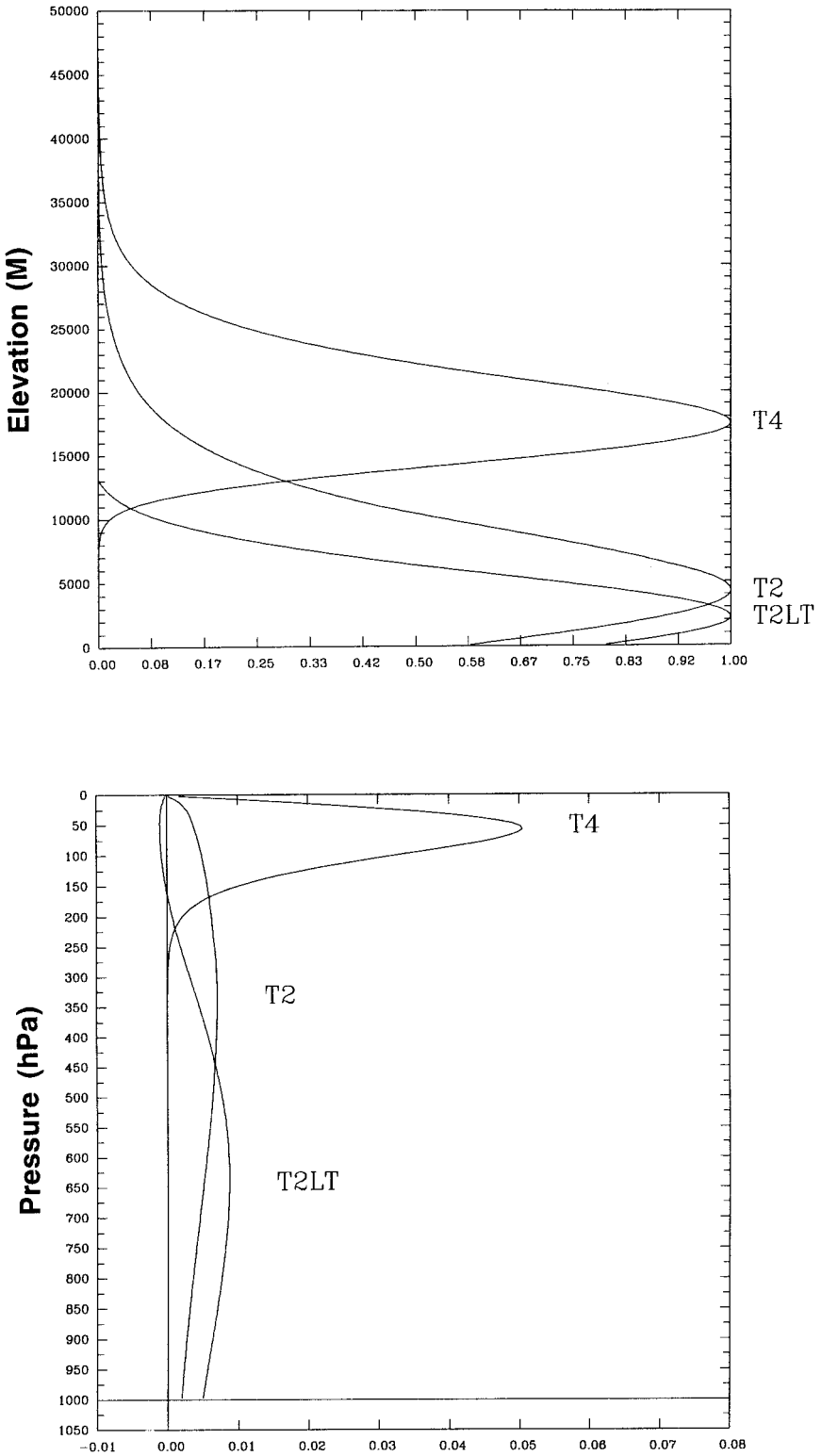


FIG. 1. MSU atmospheric weighting functions for the three MSU products described in the text: T_{2LT}, T_2, and T_4. (top) Weights presented in traditional manner as linear vs elevation for which the factor $\Delta \ln(p)$ would be applied and that are normalized to the largest weight. (bottom) Weights per 10-hPa layer presented as MSU radiometer sensitivity to 1°C perturbations for each 10-hPa layer (i.e., linear with pressure so that factor Δp would be applied per 10-hPa layer).

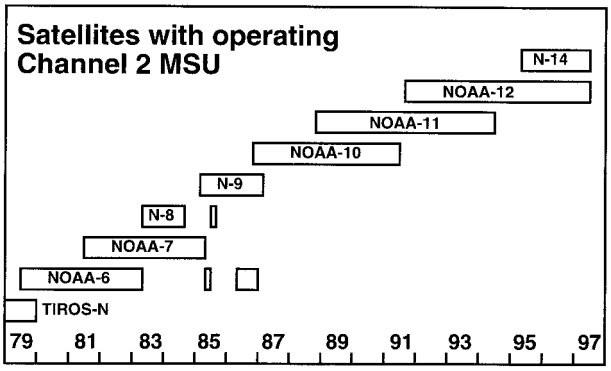


FIG. 2. Operational periods for the channel 2 receiver of the MSUs on board the nine NOAA spacecraft (NOAA- abbreviated to N- for short period satellites).

$$T_{2LT} = (T_2 3 + T_2 4 + T_2 8 + T_2 9) - 0.75\,(T_2 1 + T_2 2 + T_2 10 + T_2 11),$$

$$T_2 = 0.2\,(T_2 4 + T_2 5 + T_2 6 + T_2 7 + T_2 8), \quad \text{and}$$

$$T_4 = 0.2\,(T_4 4 + T_4 5 + T_4 6 + T_4 7 + T_4 8),$$

where $T_x Y$ is the brightness temperature of channel X at footprint position Y. Note that the resultant value of T_{2LT} is a small residual of two large numbers, and therefore the basic error contained in the individual measurements is enhanced relative to the error of a single measurement (SC92a, SC92b, CSM95). The calculation of T_2 and T_4, on the other hand, minimizes random instrument error since the observations are averaged together. In these three schemes, one value of brightness temperature is produced on each scan—that is, every 26 s.

For the daily zonal mean values used in this study, the latitudinal position of the nadir footprint determines into which 2.5° lat band the observation is assigned. These values are accumulated according to 24-h UT days and finally averaged into daily zonal means. Because the satellites do not exactly roll over the poles, the highest latitudes for which data are observed with confidence are ±82.5°. Each retrieval represents the average temperature across a roughly 1500-km swath and there are on average about 50 evenly spaced T_{2LT} retrievals (based on 400 independent footprint observations) per day per 2.5° lat band per satellite.

3. General information of overlapping data

The periods for which channel 2 data were obtained from the MSUs on various spacecraft are displayed in Fig. 2. The nominal period of performance for an individual satellite was intended to be about four years. Replacements were scheduled to be launched about every two years, alternating in their orbital positions between the A.M. [''morning'' or ''terminator''—about 0730/1930 local equator crossing time (LCT)] and P.M. (''afternoon'' about 1400/0200 LCT) orbits. For nominal operation, as an example, a P.M. satellite would experience two years of overlapping observations with an A.M. satellite already in orbit, and two years with the A.M. satellite launched to replace the earlier one. The periods of operation shown in Fig. 2 reveal several deviations from this plan, especially prior to 1988. The unusual operational sequence from 1983 to 1986, due to problems with *NOAA-8* in particular, is the subject of a later section.

There were several other problems involving the MSU instruments and/or their spacecraft. Some of the major ones are listed below.

- *TIROS-N* lost parts of orbital data throughout July 1979 and suffered loss of precision in navigation during a portion of November 1979.
- In its last year of operation, *NOAA-6* had drifted 1 h from its initial LCT after being shut down for two extended periods.
- *NOAA-7*, as was typical of afternoon orbiters, drifted from the original LCT by eventually 1.75 h.
- Channels 1 and 2 on *NOAA-9* ceased operation in early 1987 after only 102 days of simultaneous observations with *NOAA-10*.
- *NOAA-11* provided data continuously for over 6 yr, but its noticeable LCT drift (4 h) required adjustments to the data (CSM95).
- *NOAA-12's* MSU appears to have had an emitting object in the calibration field of view for cold space. As a result, the temperature of cold space, nominally 2.7 K, was being observed to be about 5 or 6 K, thus requiring special adjustment to the interpolation procedure (Mo 1994; Christy 1995, hereafter C95).
- *NOAA-13* (entire spacecraft) was lost one week after launch due to insufficient power generation from solar panels.
- *NOAA-14* experienced an early failure in the MSU's rotating mirror that was finally rectified in April 1995 by clever software changes, which carefully monitored the instrument temperature to avoid slippage in the belt that rotated the mirror assembly.

With so many operational anomalies to consider, one can understand the need to provide rigorous examination of the various adjustments that are applied to account for these operational problems.

4. Summary of merging procedure

Because several processing steps are involved in merging data from nine MSUs, it is difficult to describe in sequence every procedure without jumping from one topic to another and back again. In this section we shall describe the types of processes that are applied in the merging procedures and show some results of confidence tests. In the following section, we shall describe the exact path by which the different satellites are

merged based upon statistical tests of the processes described in this section.

The merging process begins once the daily zonal mean (2.5° bands) absolute brightness temperatures of each of the nine satellites are generated as described above and in SCG90 and CSM95. A reference annual cycle of absolute temperatures is then calculated for *NOAA-6* (A.M. satellite) and *NOAA-7* (P.M. satellite) for a period common to both: September 1981 to August 1982, a 12-month period in which both satellites observed nearly every day simultaneously. These absolute reference temperatures are then smoothed to remove fluctuations less than 15 days in length, with the smoothing wrapping around to blend the end of August 1982 to the beginning of September 1981. These two annual cycles represent only September 1981 to August 1982 and are used only as a starting reference for all anomalies.

Daily zonal-mean anomalies are then determined for each MSU time series based on the appropriate reference annual cycle. The A.M. orbiters (*NOAA-6, -8, -10,* and *-12*) utilize the reference annual cycle of *NOAA-6,* whereas the P.M. orbiters *(TIROS-N, NOAA-7, -9, -11,* and *-14)* rely on the reference annual cycle from *NOAA-7.* Because of the variation in observed brightness temperature due to diurnal effects over land surfaces, there is a noticeable difference between A.M. and P.M. MSU globally averaged temperatures (268.74 K vs 269.35 K). As such, their respective anomalies, the quantities we are concerned with here, must be determined relative to the appropriate reference annual cycle and thus eliminate the differences due to diurnal effects.

Once anomalies are determined for all time series from each MSU, there are four basic processes we apply in adjusting one satellite's anomalies to its predecessor to create a homogeneous times series in this order: (a) median filtering, (b) perturbation annual cycle removal, (c) drift-error correction (for two of the P.M. orbiters), and (d) bias removal. The last three are calculated by direct intercomparison of anomalies, latitude by latitude, during common periods of observation between the satellite of interest and its predecessor (except *TIROS-N,* which was compared with *NOAA-6,* the later satellite).

The characteristics of the these merging components are shown in Fig. 3 for the example of *NOAA-11* versus *-12* and will be referred to in the following discussion. The figure shows the two-satellite average of daily global anomalies, that is, one-half their sum (Fig. 3a, top); one-half the two-satellite daily difference (Fig. 3a, lower); the sum of the squares of the cosine-weighted latitudinal differences, no filtering (Fig. 3b); the sum of the squares of the cosine-weighted latitudinal differences after median filtering (Fig. 3c); annual cycle of the difference time series (Fig. 3d) and the latitudinal biases of the later satellite *NOAA-12* (Fig. 3e).

After all anomaly time series of each instrument are adjusted to the anchor satellite *NOAA-6* and its reference annual cycle they are merged into a single time series by averaging the anomalies on any given day when two instruments supply data. We then select more representative periods to serve as the reference for the complete anomaly time series. For T_{2LT} and T_2, we choose the 10-yr period of January 1982–December 1991 as the base. For T_4, we choose January 1984–December 1990 to avoid the massive impacts of volcanoes in 1982 and 1991.

a. Important statistical quantities

Four quantities are useful in the following discussion and are derived from the time series of the differences between two satellites. First is the standard deviation of the differences of the daily global mean time series (σ_Δ, which is twice the value of σ for Fig. 3a, lower). The second is a measure of the latitudinal agreement between the satellites and is the root of the daily sum of the squares of the cosine-weighted latitudinal differences (T_{err}). With this quantity we can judge whether there is truly latitudinal consistency of anomalies even when global values of σ_Δ are low. In equation form for a single day this is

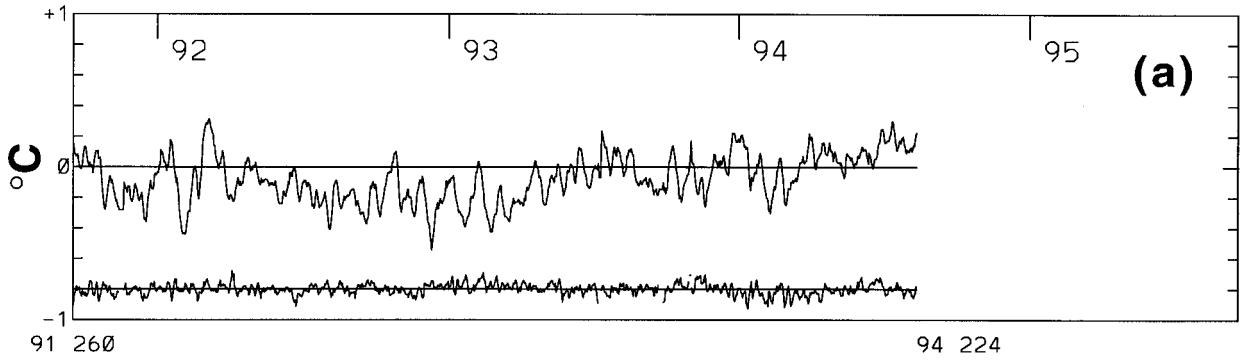
$$T_{\text{err}}^2 = \sum_{\varphi=-82.5^\circ}^{82.5^\circ} [T_1(\varphi) - T_2(\varphi)]^2 \cos(\varphi) \Bigg/ \sum_{\varphi=-82.5^\circ}^{82.5^\circ} \cos(\varphi),$$

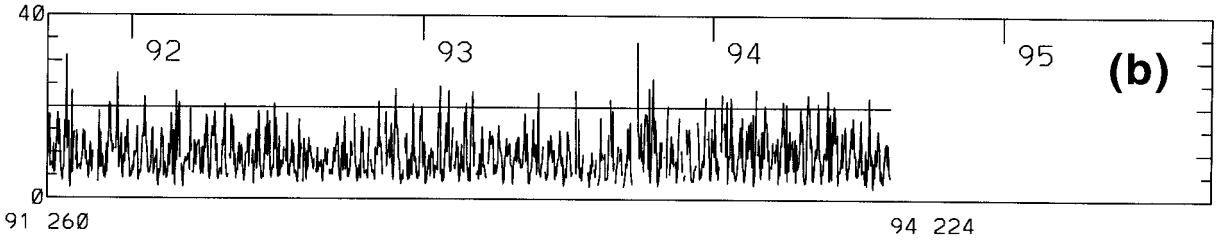
where T_1 and T_2 represent the daily zonal-mean temperature anomalies for satellites 1 and 2 (after the bias of satellite 2 is removed). Eventually, we shall select a value of T_{err} that will serve as a threshold to deselect those days of large latitudinal differences from inclusion in the processes of bias and trend determination.

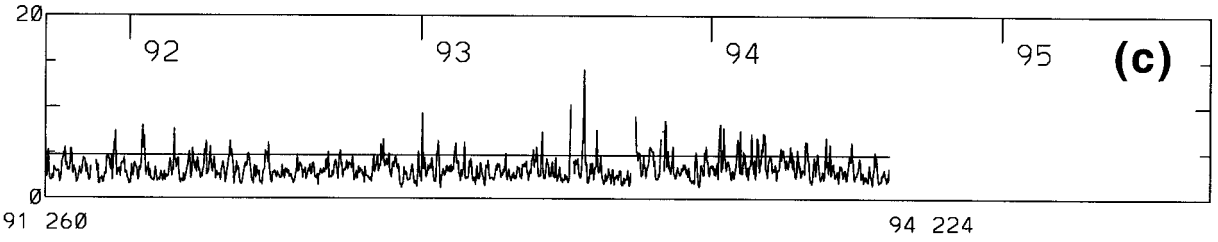
The third quantity of interest is the number of independent observations (N) contained in each of the overlap comparisons. Since a five-point median filter (described later) is applied separately and independently to each satellite's zonal anomaly time series, it is difficult to know how much ''independence'' is removed in the time series of intersatellite differences. By investigating the day-to-day replacement rate of the median filter along with the gaps that occur due to missing

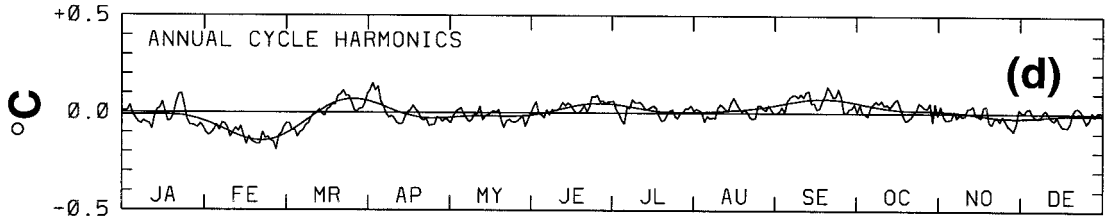
→

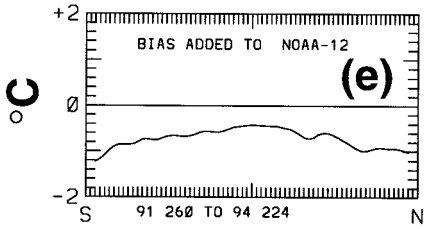
FIG. 3. Characteristics of the difference time series between overlapping satellite observations for *NOAA-11* vs *-12.* (a) Time series of global daily average anomalies (°C) from the two instruments (top) and their difference divided by 2 (lower); (b) daily time series of $T_{\text{err}}^2 \times \Sigma \cos\phi$ ($\phi = -82.5°$–$82.5°$ by 2.5°, see text) without median filtering; (c) as in (b) but with 5-day median filtering applied and threshold equivalent to $T_{\text{err}} = 0.325$°C [horizontal line; note: vertical is scale one-half of (b)]; (d) perturbation annual cycle between *NOAA-11* (already adjusted to *NOAA-6*) and *NOAA-12*; (e) zonal-mean biases, latitude-by-latitude, added to *NOAA-12* to adjust to *NOAA-6.* The numbers below the time series (a)–(c) indicate the year and Julian day of the beginning and end of the time series.











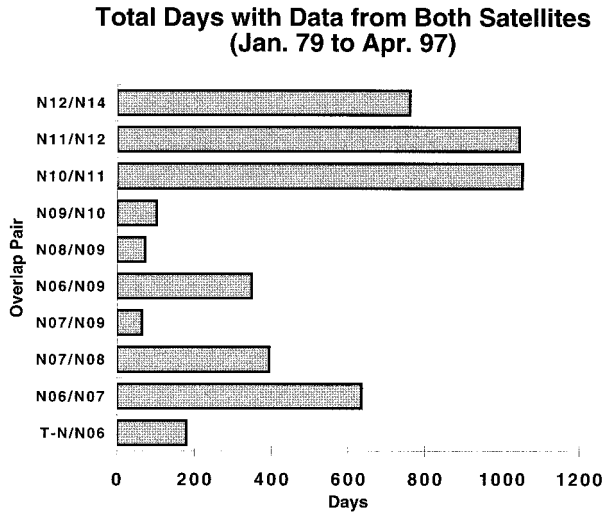


FIG. 4. Number of days for which a sufficient amount of data (19 000 footprint observations) were available from both satellites for intercomparison.

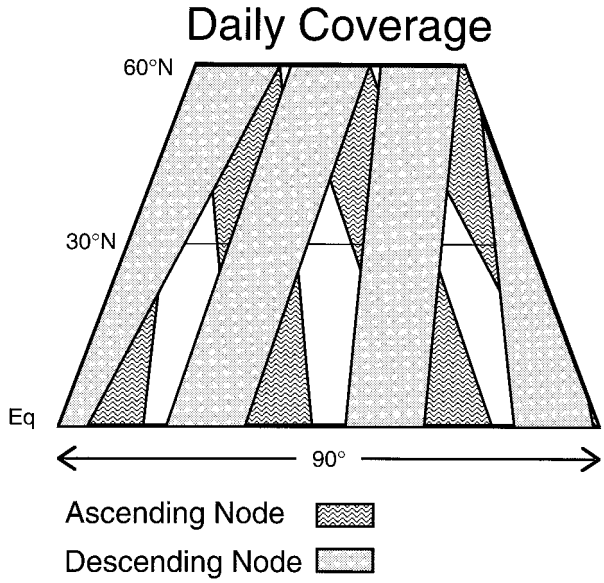


FIG. 5. Schematic ground track of a single MSU for a section of the globe 0°–60° lat with an extent of 90° long.

or deselected days, we estimate that about every second value is mostly independent. This does not refer to every second day, but rather to every second ''selected day'' that passes the T_{err} threshold criterion (since only these are used in the bias and trend determination). By this reasoning, N would be about half of the total number of selected observations in each overlapping period.

Two other estimates for N may be made from the intersatellite difference time series of global anomalies itself. First, the autocorrelations at lag indicate that $r^2 < 10\%$ at lag 2 and $<1\%$ at lag 3. This test implies that every third selected day is independent. Alternatively, we may calculate the reduction of variance due to the median filter. The ratio $\sigma^2_{\Delta\text{no-filter}}/\sigma^2_{\Delta\text{median-filtered}} = 3.07$, which indicates that the new value of N should be about one-third of the total number of original unsmoothed observations. Together, these two tests indicate that N is one-third of the total number of observations employed in the calculation of standard error ($\sigma_E = \sigma_\Delta/\sqrt{N}$), which is the final quantity of interest.

b. Characteristics of the overlapping periods

The number of days during each overlapping period for which data from both satellites was sufficient to calculate daily temperature values (19 000 footprint observations per day or 80% of possible for the required view positions) is shown in Fig. 4. Multiyear overlapping periods are preferred because the large number of observations allows for the precise determination of intersatellite biases, error trends, and perturbation annual cycles.

What is evident in Fig. 4 is that the overlap periods between *NOAA-9* and the three preceding spacecraft *NOAA-6, -7,* and *-8* are relatively short and, in each case, the periods occur during the very end of the earlier satellite's life span. The problems inherent in the calculation of the *NOAA-9* bias, given these relatively poor overlap periods, is the topic of an upcoming section.

5. Detailed description of merging procedures

a. Median temporal filtering and error-threshold value

As discussed in CSM95, once the zonal anomalies of each satellite are determined relative to the reference of *NOAA-6* or *-7,* they are filtered in time with a 5-day median filter to account for day-to-day intersatellite differences in orbital swath ground tracks. These intersatellite differences are most pronounced in the subtropical latitudes where gaps appear in a single day's coverage as schematically shown in Fig. 5. Each satellite crosses a given latitude (to $\pm 82.5°$ lat) 28 times per day during its 14 orbits. At the equator, the swaths achieve an almost even spatial distribution and therefore provide excellent spatial sampling there. Poleward of about 40° lat, the convergence of the swaths also produces excellent sampling as the swaths overlap.

However, the ground track of the satellites in the subtropical regions produces a pattern that is not spatially complete. In the unshaded gaps of Fig. 5, which are maximized at about 24° lat, no observations are possible on a given day. This entire pattern precesses eastward about 700 km (at equator) each day so that within a 3–4-day period all gaps are filled. However, any two satellites in orbit do not have exactly the same precession rate so their sampling patterns fluctuate in and out of phase with a period of about 10–12 days. This is illustrated in Fig. 3b (no median filter) where T^2_{err} reveals substantial 10–12 day oscillations in which T^2_{err} is low

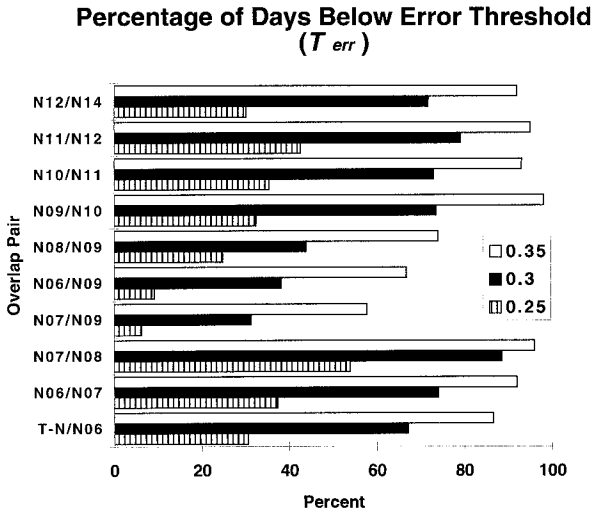


FIG. 6. Percentage of days from Fig. 4 on which the agreement between two satellite temperatures fell below average zonal-mean thresholds: T_{err} = 0.25, 0.30, and 0.35°C.

when the ground track patterns are in phase (matching) and large when out of phase.

When the two satellites are out of phase, one satellite views the subtropical locations that mostly represent the data gaps of the other. The existence of longitudinal atmospheric temperature variations, therefore, is enough to cause increases in T_{err}. In the subtropics, however, there are also substantial longitudinal variations in surface topography. For land below 500 m, T_{2LT} emissions from the surface account for about 15%–20% of the total signal and for oceanic surfaces, about 10%. At higher elevations, the surface ''shines through'' more and more because the oxygen overburden, from which the atmospheric emissions originate, becomes less of the observational signal.

As long as the surface is sampled in correct proportion for all ocean–land–mountain surface types in each latitude band, the surface emission effect, being very systematic, may be eliminated in the anomaly dataset for constant LCT. This happens for the near-equator and extratropical latitudes. However, for subtropical latitudes, one cannot assume that on a daily basis a consistent proportion of ocean–land–mountains will be observed by a given satellite. For example, the width of the Andes Mountains is approximately the width of the gaps shown in Fig. 5. Thus, the extent to which the Andes are sampled on a given day would impact the zonal-mean temperature and the anomalies produced therefrom.

In Fig. 6 we show the percentage of days for each overlapping period in which three thresholds of T_{err} (0.25°, 0.30°, and 0.35°C) are used as constraints and after the merging procedure is complete for path C1 (discussed later). The best agreement between satellites occurs for *NOAA-6* versus *-7, NOAA-9* versus *-10, NOAA-10* versus *-11, NOAA-11* versus *-12,* and *NOAA-12* versus *-14.* What is somewhat remarkable in this

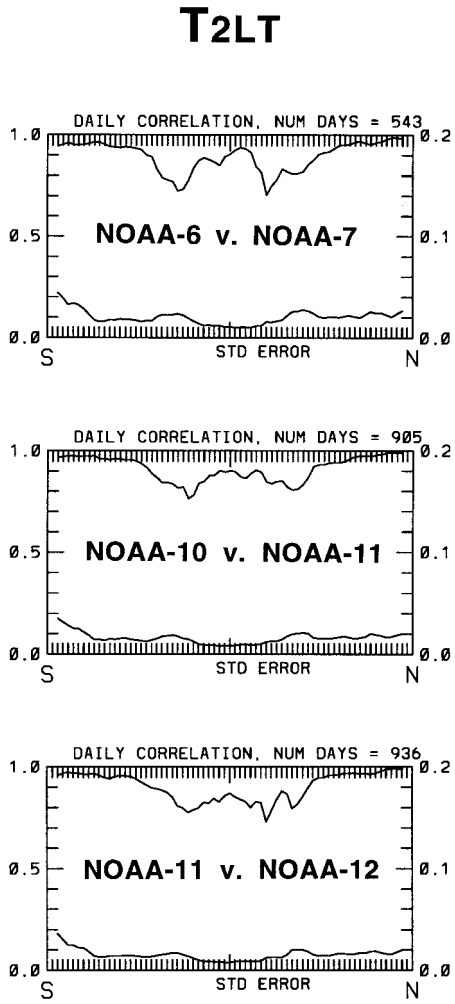


FIG. 7. Correlation of daily latitudinal anomalies during overlapping comparisons of *NOAA-6* vs *-7, NOAA-10* vs *-11,* and *NOAA-11* vs *-12* (top curve of each, scale on left) for T_{2LT}. Latitudinal values of σ_E from the differences between the pairs of satellites (lower curve of each, scale on right in °C). The horizontal scale is from 87.5°S to 87.5°N lat.

group is that the *NOAA-9* versus *-10* overlap contains only 102 possible days on which statistics may be determined, and yet the agreement is excellent, being as good or better than those for which multiyear overlaps are available.[1]

The largest values of T^2_{err} in Fig. 3b are due to anomalies in the subtropical latitudes and are confirmed in

[1] Though channel 2 failed on *NOAA-9* after only 102 days of overlap with *NOAA-10,* channel 4 (lower stratosphere) remained in operation for the nominal 2-yr overlap period. The error magnitude of the *NOAA-9* vs *-10* overlap in channel 4 was the lowest of all overlap comparisons. It is often the case that the error magnitudes are systematic for all channels of an MSU. This, then, is further indication that though we have only a brief channel 2 overlap for *NOAA-9* vs *-10,* the bias has been accurately determined.

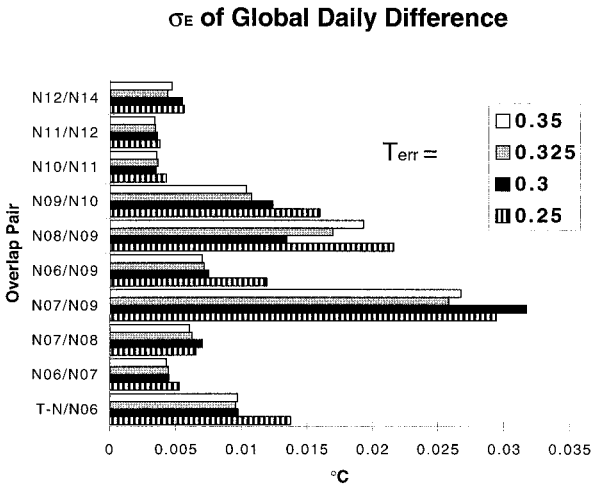


FIG. 8. Global-mean values (°C) of σ_E for four threshold values of T_{err}.

the latitudinal correlation and $\sigma_{E\ latitude}$ diagrams produced from the daily, latitude-by-latitude intersatellite differences (Fig. 7). The correlations are lowest in the subtropical zones, lower even than the equatorial latitudes where correlations would be expected to be lowest due to small signal magnitude (Hurrell and Trenberth 1996). We demonstrated in CSM95 that the noise due to this sampling artifact was largely eliminated by the application of a 5-day median filter to the daily zonal anomalies (Fig. 3c). Note that there still remains a pattern of fluctuating T_{err}^2, but it is much reduced.

Because the long-term stability of the MSU time series is a critical issue, we have decided to apply a constraint on those data that are used to determine the intersatellite biases and trends between two satellites. The value of T_{err} is to be based on the smallest σ_E achievable from the difference time series or, in other words, that we select the subset of days that keeps N large and σ_Δ small. In Fig. 8 we show σ_E for four thresholds of T_{err}. Except for *NOAA-10* versus *-11*, σ_E is lowest or second lowest when $T_{err} = 0.325$ for all pairs. We select 0.325 then as the threshold (represented by the horizontal line in Fig. 3c) and note that 80% of the possible daily comparisons are accepted for bias and trend determination (see Fig. 6). In the previous dataset version b, all data were utilized for bias and trend calculation (described below), no matter how large T_{err} became.

b. Perturbation annual cycle

The perturbation (or error) annual cycle (PAC) was addressed in CSM95 though its origin was not explained. In CSM95 we reported that there were fluctuations in phase with the annual cycle in the difference time series between some of the satellite pairs. In particular, the peculiar shape of the PAC in *NOAA-12* versus *-11* (Fig. 3d), which had repeated itself for 3 yr, prompted us to examine other instrument parameters to discover the source of the fluctuations. The PACs arise from the differences that occur between each satellite and its reference instrument (*NOAA-6* for A.M. and *NOAA-7* for P.M.). Because these were systematically repeating fluctuations, their magnitudes could be determined and removed (in CSM95 to annual harmonic five) from the newer satellite, thus improving σ_E by not reducing N by a significant amount.

Several health parameters are monitored on each MSU including the temperature of its various components. Because the radiometers are externally calibrated, fluctuations in instrument temperature (theoretically) are not intended to affect the calculation of the earth-viewed temperature. However, we find that the secular temperature fluctuations of both *NOAA-12* hot targets, ranging up to 7 K and repeating each year, are highly correlated (0.96) with the PAC we had discovered in *NOAA-12* (Fig. 3d).

This same annual variation appears in thermistors in other MSU components and in the data for channels 3 and 4 (which have separate calibration targets and circuitry). Thus, we assume that the problem is pervasive to this particular MSU. Further study reveals detectable annual cycle features in *NOAA-11* and *-14* also. Therefore, it seems there is a systematic dependency between brightness temperature anomalies of the MSU and the temperature fluctuations of its mechanical components (Mo 1994). This dependency is slight (less than 2% of the secular range of the hot target variations of *NOAA-12*) but does affect the merged data.

These PACs apparently relate to the seasonal variations in the exposure of the instruments to direct or reflected solar radiation. Each spacecraft carries a slightly different suite of instruments with different instrument configurations and thus experiences different shadowing effects during the course of a year, which in turn affects the temperature of the instrument itself. In future space observation missions it would be beneficial before launch to determine the impact of variational solar effects on each instrument after full spacecraft integration is complete.

Because these cycles are systematic features, they are easily quantified and removed. Our present investigation indicates that the character of the hot-target fluctuations are better represented by annual harmonics one to eight, instead of one to five as applied in CSM95. Increasing the number of harmonics to eight reduces σ_Δ of the intersatellite comparisons by around 10% over that of harmonics one through five. For the determination of the PAC, we found no difference between inclusion of all days on which data were available from the two operating satellites, or only those days that met the T_{err} threshold. Therefore, the calculation of the magnitude of annual cycle harmonics one to eight of the perturbations will utilize all available data. These PACs were removed from *NOAA-8, -11, -12,* and *-14.*

To test the reproducibility of the PACs we divide the overlap data into two independent halves and calculate

TABLE 1. Comparison of the standard deviation of the daily global differences (σ_Δ, °C) for three satellites calculated under four examples related to the removal of PACs. The first experiment has no adjustment for PACs. The second and third experiments use only half of the overlap data to calculate the PAC, and the last column utilizes all overlap data as in T_{2LT} (version c1). The values of σ_Δ for *TIROS-N, NOAA-7, NOAA-9,* and *-10,* which did not have PACs removed, are shown for comparison purposes.

	No adjustment	First half	Second half	All ($T_{2LT.c}$)
	σ_Δ	σ_Δ	σ_Δ	σ_Δ
NOAA-11 (vs *-10*)	0.082	0.064	0.065	0.063
NOAA-12 (vs *-11*)	0.098	0.063	0.075	0.061
NOAA-14 (vs *-12*)	0.117	0.064	0.065	0.064
TIROS-N (vs *-6*)	—	—	—	0.067
NOAA-7 (vs *-6*)	—	—	—	0.060
NOAA-9 (vs *-6*)	—	—	—	0.056
NOAA-10 (vs *-9*)	—	—	—	0.060

the PACs separately from each half. In Table 1, the first column shows σ_Δ for which no PACs are removed. The sudden increase in noise in *NOAA-11* and the following satellites (represented by σ_Δ, cf. the other satellites) is clearly a sign that something new developed.

The overlap periods of *NOAA-11* versus *-10* and *NOAA-12* versus *-11* are over 1000 days in length and *NOAA-14* versus *-12* about 800 days (to date). In the second and third columns of Table 1 we see that σ_Δ is reduced considerably after each PAC, based on only one-half of the overlap data, is removed. The reduction is substantial and approaches the magnitude of the last column for which the entire overlap is utilized.[2]

We must note here that the PAC of *NOAA-8* (vs *-7*) has also been calculated and removed. However, only one year of overlapping observations were present for this procedure, so there is no means to test the reproducibility of this particular PAC. We are certain that nonrepeating fluctuations between *NOAA-7* and *-8* have been artificially removed in this process as the PAC overfits the difference time series. We justify the removal of this particular PAC because there were attitude control problems with *NOAA-8* before it was shut down and these probably contributed to the variance in the difference time series. In addition, *NOAA-8* is not a factor in the construction of the backbone of the time series and serves merely to enrich the day-to-day variations (already observed by *NOAA-7*) during its brief lifetime. Therefore, calculations in later exercises will present *NOAA-8*'s σ_Δ and σ_E as being artificially low and should be viewed with caution.

In general, we expect the lowest magnitude of σ_Δ to be about 0.060°C as determined by comparisons between satellites with no significantly anomalous operations (e.g., Table 1, last column). It is possible to reduce σ_Δ below 0.060°C by introducing artificial overfitting procedures not based on physical effects. All of the procedures described in this paper are based on physical factors determined from independent information (e.g., source of PACs in instrument body temperatures). The results in Table 1 provide high confidence in the reproducibility of the non-*NOAA-8* PACs and indicate further reductions in σ_Δ are not warranted at this time. The existence of PACs also serves to highlight the need for future satellite systems to require overlapping observations of at least one year with satellites already in orbit.

c. Drift error

A spurious variation in the intended temperature time series can occur as a spacecraft drifts away from its original LCT, observing the earth at earlier or later points in the local earth position's diurnal cycle. Our calculation of the drift error of a P.M. satellite consists of finding the slope of the regression line (for each latitude) determined from the time series of difference anomalies relative to the concurrently operating A.M. satellite (C95). The variation of greatest interest appears in *NOAA-11* as the spacecraft drifted to later LCTs.[3]

NOAA-10 drifted by no more than 15 min from its average LCT (1925/0725) in five years of operation. *NOAA-11,* in contrast, drifted from 1335/0135 to 1405/0205 (30 min) between operational start (October 1988) and April 1990. By the end of *NOAA-10's* operation (August 1991) *NOAA-11's* LCT was an additional 35 min later. From September 1988 to July 1994, *NOAA-11's* LCT continued to drift to 1700/0500. The fact that *NOAA-11* began operation at an LCT just after the usual maximum of local surface skin temperature and then drifted relatively rapidly through 3+ h of the diurnal cycle requires that special attention be given this satellite since it forms the bridge between *NOAA-10* and *-12,* two of the backbone satellites of the T_{2LT} time series.

In selecting the starting date for the trend removal of *NOAA-11* versus *-10,* we considered several factors: (a) the statistical significance of the trend over the portion of the time series for which the correction was to be applied, (b) the effect of the trend removal on the entire *NOAA-11* versus *-10* difference trend, and (c) the reduction of σ_Δ in the entire overlap.

For reference, a time series of length 500 (1000) days with $N = 167$ ($N = 333$) and $\sigma_\Delta = 0.070$°C will possess

[2] When using halves, the PAC will overfit the variations of the respective halves including variations that may be interannual. This is especially true for the second half of the *NOAA-12* vs *-11* overlap when the volcanic effects of Mt. Pinatubo caused a large temperature perturbation.

[3] The drifting is intentionally introduced in the injection trajectory for the P.M. satellites to prevent the spacecraft from accidentally drifting ''backward'' to local solar noon, a time at which some measurements may be impaired by sunlight directly behind the spacecraft.

a slope significantly different from zero if it exceeds ± 0.05 (± 0.02) °C yr^{-1}. We evaluate the relative significance of the difference trend beginning at various points in the *NOAA-11* versus *-10* overlap counting back from the last common day of observation in 25-day increments. Essentially all time periods from 300 days to 1000 days produce significant difference trends, but those of greatest significance are between the last 400 and 700 days of common observations. This equates to drift errors beginning in *NOAA-11* after 15–40 min from original LCT. The greatest reduction in σ_Δ and the least magnitude of the residual difference trend occurs for a drift period of 475 days (28 May 1990–15 September 1991, or about 30 min of LCT drift) so this period is selected to represent the drift error of *NOAA-11* versus *-10*. After this trend is removed from *NOAA-11*, we calculate the latitudinal difference trends between *NOAA-11* and *-12*. The difference trends for the 475-day portion of *NOAA-11* versus *-10* and all of *NOAA-11* versus *-12* (1045 days) are $+0.066$ and $+0.059$°C yr^{-1} respectively.[4] Since all operational aspects of the instruments were nominal, we assume that these significant trends are due to *NOAA-11's* LCT drift.

We approximate the drift error of *NOAA-11* by calculating and removing the two trends or line segments (latitude by latitude) in the difference trends of *NOAA-11* versus *-10* and *NOAA-11* versus *-12*. By calculating two line segments, we allow for a crude approximation of any second-order effect associated with the nonsinusoidal component of the diurnal cycle from a drift of 1405/0205 LCT to 1700/0500 LCT. (If the T_{2LT} diurnal cycle were a first-order sinusoid at every latitude, the effect of the forward drift of both ascending and descending nodes would exactly cancel, resulting in no net temperature change.)

In general, latitudes with a large proportion of ocean or that are tropical tend to show relatively strong, systematic warming throughout the forward drift, whereas latitudes with considerable coverage of dry land reveal less warming. The P.M. drift affects both ascending (daytime, variations of which dominate the diurnal cycle effect) and descending (nighttime) nodes, and we are looking only at the net drift effect in the average of both nodes.

The latitudinal variations in trend adjustments between the first and second segments are interesting. For example, in the latitudinal band 17.5°–32.5°S the *NOAA-11* relative trend for the first segment (1405/0205–1440/0240) is $+0.11$°C yr^{-1} and in the second segment (1440/0240–1700/0500) is $+0.00$°C yr^{-1}. This is consistent with the observation that the surface skin temperatures tend to peak shortly after solar noon, especially in dry areas, whereas the atmospheric temperatures (above land and ocean) tend to continue warming throughout the afternoon. For example, early summer lower-tropospheric temperatures at several European stations reached maximum values on average at 1746 LT (1638 LT) for the 850 hPa (700 hPa) pressure levels (Richner and Phillips 1984).

The net effect for this subtropical band, which includes much ocean surface but also dry regions of Australia, southern Africa, and South America, suggests that atmospheric heating dominated the first portion of the drift, whereas the waning atmospheric heating (after 1440) and continued cooling of the skin summed to zero in the second portion. In contrast, the trend for the 10.0°–32.5°N band, which includes the larger deserts from the Sahara through Tibet, is only about $+0.03$°C yr^{-1} in both periods. This suggests that the skin cooling over the greater desert areas almost cancels the afternoon atmospheric warming.

As stated, the 3-h *NOAA-11* drift error of T_{2LT} as defined by two line segments is a rather crude approximation of the diurnal cycle. Improvements to this approximation may become a task to address in future versions of the dataset. We are reluctant to pursue higher-order approximations to the diurnal effects at this point since the reduction of σ_Δ already achieved by this crude method matches the level found in the overlap statistics of the other nontrending satellites (Table 1). At this time we feel we would risk overfitting the data and removing real interannual variations (though reducing σ_Δ) by the higher-order estimates.

NOAA-7 experienced a drift from its original LCT of 1.75 h by the last day of operation. Because it was in a P.M. orbit, we must assume there is a spurious temperature change present; however, we have no stable A.M. satellite against which to compare during most of the drift. We have therefore taken the drift error trend from *NOAA-11* versus *-10* as a starting point for drift error removal. The drifting speed of *NOAA-7* was about 75% the rate of *NOAA-11's* drift (16 months vs 21 months for the same drift in LCT). Thus we reduce the latitudinal trends of the *NOAA-11* versus *-10* by 75% and then remove them from *NOAA-7* beginning in May 1983. Though this seems ad hoc, we base the removal procedure on the fact that *NOAA-7* indeed drifted through the same diurnal times as did *NOAA-11*, which we know experienced a drift error in temperature. Fortunately, the backbone of the time series does not require information from *NOAA-7*, whose contribution to the time series is information from April 1983 to January 1985. Thus, the effect of the trend removal reduces the temperature for 1984 by about 0.05°C but does not interact with the calculations for the following years.

d. Bias

Though calibrated precisely on earth before launch, the MSU instruments show systematic intersatellite biases once in the environment of space. Part of the bias

[4] If averaged, the effect would be $+0.061$°C yr^{-1}, which is twice the magnitude reported earlier (e.g., CSM95) due to a factor of 2 missing on a printout.

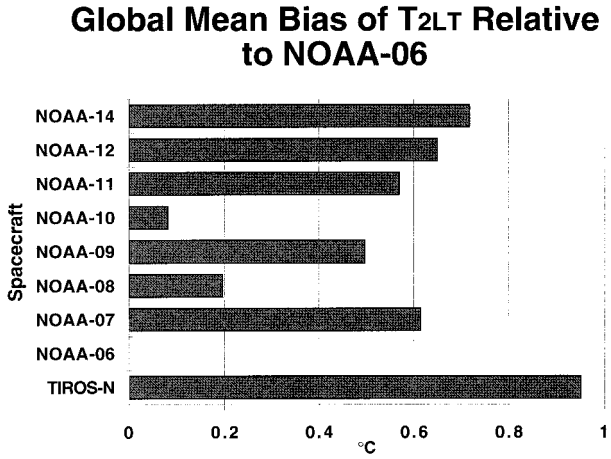


FIG. 9. Global mean (°C) of latitudinal biases of T_{2LT} relative to *NOAA-6*'s MSU.

is due to the point in the diurnal cycle that each satellite observes. We calculate the biases by first determining the average difference in the zonal-mean temperature anomalies between two simultaneously operating satellites for all days passing the T_{err} threshold. By accumulating these differences from one satellite to the next, we are able to calculate the bias for any satellite relative to the anchor satellite *NOAA-6* (e.g., Fig. 3e).

The global mean biases relative to *NOAA-6* for path C1 (described later) are shown in Fig. 9. All P.M. orbiter temperatures are warmer than the A.M. MSUs due to the maximum emission observed in the sunlit afternoon (~1430 LCT). Note, however, in Fig. 9 the bias was unusually large for *NOAA-12* (an A.M. satellite), being over 1°C at higher latitudes (Fig. 3e), probably due to the emitting object in the ''cold space'' view (Mo 1994). The confidence we have in the magnitude of the intersatellite biases will be addressed following the determination of the exact set of satellites to be utilized in the merging process.

6. The special problem of determining the bias for *NOAA-9*

The MSU time series described in SCG90 was produced by removing the biases of each MSU relative to *NOAA-6.* In that and subsequent studies, the *NOAA-9* bias was calculated by direct comparison with *NOAA-6.* SCG90 demonstrated that the MSU on *NOAA-6* was stable from 1979 to 1986 by comparing radiosonde-simulated MSU T_2 and *NOAA-6* observations during an early 13-month period (November 1979–November 1980) and the last 13 months of *NOAA-6's* operation (November 1985–November 1986) even though it drifted ''backward'' in LCT. The concern here is whether the net effect between the cooling observed as the descending node drifted from 0730 to 0630 LCT and the warming as the ascending node moved from 1930 to 1830 actually cancel each other.

The net effect of any drift is maximized over land where the diurnal impact of the surface is greater than is observed over the oceans. In SCG90, 45 radiosonde stations were selected in the eastern United States whose twice-daily regular observation times were very close to those of the *NOAA-6* passes. Since the radiosondes maintained a constant observing time, any net temperature drift of the satellite would be detected in the interannual differences. The ''interannual'' temperature difference (1985/6 minus 1979/80) was +0.013°C greater in the radiosondes than the MSU. Since this was only one standard deviation of the measurement error for the combined eastern U.S. stations, we rejected the hypothesis that there was a spurious trend in *NOAA-6* for its period of record. This result implied that whatever cooling appeared in the descending node was essentially balanced by the warming of the ascending node for the A.M. orbit drift.

As we examine the entire time series from the present-day vantage point, we see that how one calibrates *NOAA-9* to *NOAA-6* involves decisions that can substantially change the trend of the data. No other aspect of constructing the overall time series has such a pronounced impact on the magnitude and even sign of the long-term trend. For example, we may create separate global time series of T_{2LT} (from the same raw satellite data) whose trends range from −0.08 to +0.04°C decade^{-1} by essentially adjusting the procedures by which the bias of *NOAA-9* is determined and nothing else. These large ranges in the 18+ yr trend can be created only from error-prone data observed during brief overlap periods that will be shown are inappropriate to employ for the merging process. In any case, a careful examination of the calculation of *NOAA-9* bias is necessary here.

To provide background for the following discussion, Fig. 10 displays the daily global anomalies of T_{2LT} for 1982–86 as they are calculated from the individual MSUs and adjusted for bias to *NOAA-6.* That there are multiple, direct methods or paths available to calculate the bias of *NOAA-9* is due ultimately to the instability of the *NOAA-8* satellite. As indicated in Fig. 2, *NOAA-8* was launched in 1983 to replace *NOAA-6* but developed problems and was shut down in 1984 when the primary oscillator behaved sporadically. *NOAA-6* was reactivated at that time.[5] In mid-1985, *NOAA-8* was recovered, reactivated, and data for about 5 months were recorded. However, the satellite had lost several thermal channels during its hiatus due to detector heating during tumbling, so it was again shut down in favor of *NOAA-*

[5] Due to an unreported change in the identifier number for the *NOAA-6* spacecraft when it was reactivated in 1984, *NOAA-6* data were not extracted in the data rescue process when the last remaining magnetic tape medium on which the data were stored was converted to a newer medium. As far as we are able to determine, the data for *NOAA-6* from August 1984 to April 1985 are extinct.

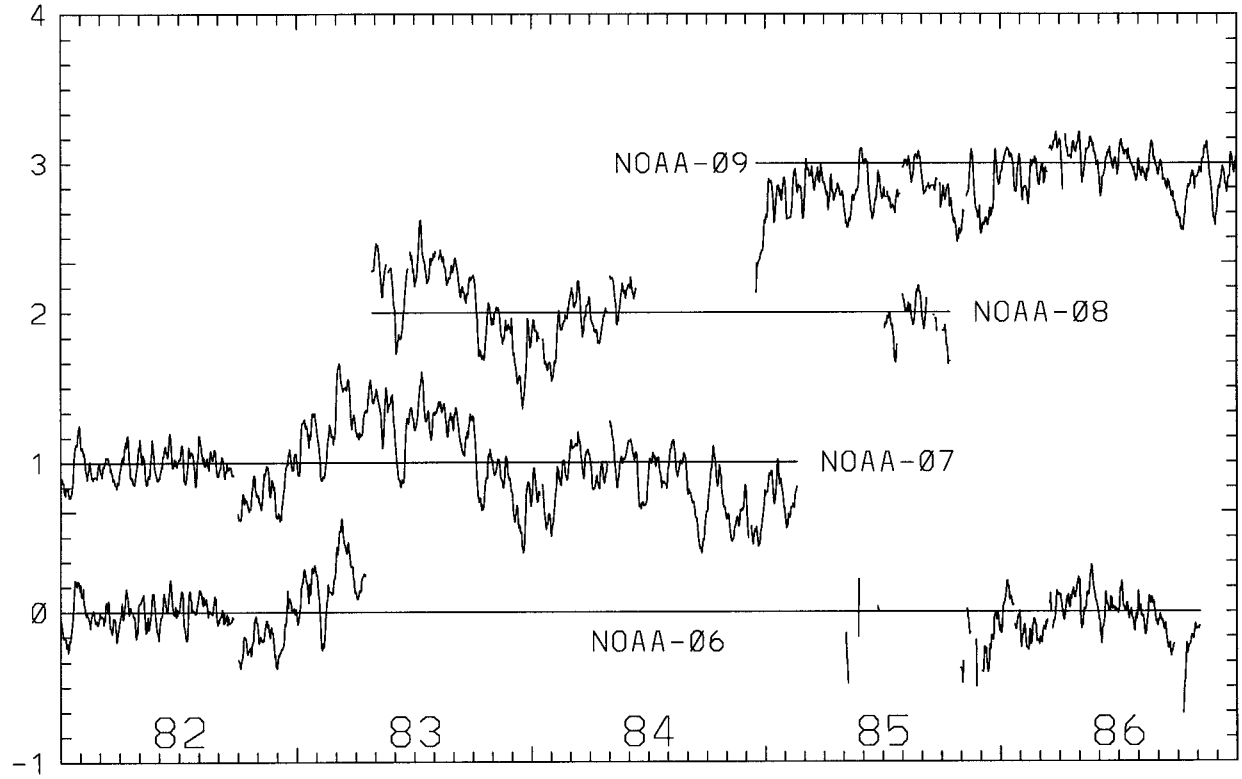


FIG. 10. Daily anomalies (°C) of global-mean T_{2LT} for the period 1982–86 as provided by the indicated MSUs. Each time series has been displaced by 1°C for clarity.

6. For the third time, *NOAA-6* was activated and served as the A.M. orbiter until *NOAA-10* was put into service in December 1986 (Kidwell 1995).

Had *NOAA-8* performed nominally, there would have been a 2-yr period of overlapping observations first with *NOAA-7* and then with *NOAA-9*. With such intended overlaps, a precise value of the *NOAA-9* bias could have been determined through *NOAA-8* and *NOAA-7*. However, with the repeated activation and deactivation of *NOAA-6* and *NOAA-8*, a considerable amount of uncertainty is introduced into the procedure to calculate *NOAA-9*'s bias.

Our goal is to produce a time series in which data from the MSU instruments alone are used. Subsequent comparisons with radiosondes, therefore, will be completely independent validation exercises. As such, we must determine the bias of *NOAA-9* relative to *NOAA-6* from one or more of the satellites that were in operation with *NOAA-9*. From Fig. 10, we observe three different paths, call them A, B, and C, through which the *NOAA-9* bias may be directly calculated:

path A: *NOAA-6* to *NOAA-7* to *NOAA-9*,
path B: *NOAA-6* to *NOAA-7* to *NOAA-8* to *NOAA-9*,
path C: *NOAA-6* to *NOAA-9*.

At first glance, path C would seem most appropriate because this is a direct route and will not incur added uncertainty associated with the calculation of the intervening satellites' biases. However, we know that *NOAA-6* was out of operation for relatively long periods and was at the end of its life when the *NOAA-6* versus *-9* overlap occurred. Satellite stability and drift are therefore issues for *NOAA-6* though we have shown that the drift apparently had no significant effect on the observed temperature (SCG90).

Consider too, information from Fig. 8 shows that the *NOAA-6* versus *-7* bias is robust and known to high precision. Therefore, little risk is taken by utilizing information from this overlap, so path A or B may serve as the best route. In the following discussion we shall consider the merits of each path to calculate the bias of *NOAA-9* relative to *NOAA-6*.

a. Path A

Information in Figs. 6 and 8 and supplemented in Table 2 indicate that path A, particularly the *NOAA-7* versus *-9* overlap period, does not provide enough data for a confident assessment of the bias. Not only is this the shortest overlapping period available, but σ_E is clearly the largest of all pairs (Fig. 8). Part of the problem is found in Fig. 10, in which the first week of observations of *NOAA-9* are substandard. In any case, we calculate the bias of *NOAA-9* relative to *NOAA-6* as

TABLE 2. Characteristics of different paths used to calculate the bias of the *NOAA-9* MSU for T_{2LT} (T_{err} = 0.325°C). Decadal trend is for the period January 1979–April 1997.

Path	Sat. pair	Total days	Accepted T_{err} = 0.325	Percent	σ_Δ °C	*NOAA-9* bias °C	Trend °C decade^{-1}
A	06/07	636	543	85	0.060		
	07/09	64	34	53	0.086	+0.368	+0.045
A1	06/07*	636	543	85	0.060		
	07*/09	64	34	53	0.087	+0.458	−0.016
B	07/07	636	543	85	0.060		
	07/08	395	371	94	0.066		
	08/09	73	52	71	0.071	+0.410	+0.014
B1	06/07	636	543	85	0.060		
	07/08*	395	371	94	0.063		
	08*/09	73	51	70	0.071	+0.495	−0.049
B2	06/07*	636	543	85	0.060		
	07*/08	395	366	93	0.069		
	08/09	73	51	70	0.070	+0.428	+0.006
B3	06/07*	636	543	85	0.060		
	07*/08*	395	366	93	0.065		
	08*/09	73	51	70	0.071	+0.546	−0.081
C	06/09	349	184	53	0.056	+0.497	−0.052
C1	06/09, 7*	349	184	53	0.056	+0.497	−0.046
C2	06/09, 7*8*	349	184	53	0.056	+0.497	−0.051

* Relative trend removed (see text).

about +0.37°C through this path. In constructing the entire time series therefore, the latitudinal biases, whose global mean is 0.37°C, are subtracted from *NOAA-9* in path A. The resulting 18+ yr trend in the data becomes +0.045°C decade^{-1}, a value much warmer than that of version b (−0.07°C decade^{-1}) and of global tropospheric radiosonde datasets for the same period (as shown later in Table 5).

The *NOAA-7* spacecraft, being a P.M. orbiter, is subject to the intended ''forward'' orbit drifting pattern (LCT becomes later and later) as mentioned earlier. By the time of the brief *NOAA-7* versus *-9* overlap, *NOAA-7* had drifted to 0415/1615 local observation time from 0230/1430, or 1.75 h. Evidence presented above from the time series of *NOAA-11* indicates that such a drift will induce a spurious warming trend as the satellite samples the earth at later points in the local diurnal cycle.

The results of the merging process in which *NOAA-7* has been adjusted for an assumed spurious warming trend are identified as path A1. The main effect of this trend removal will be maximized in path A1 since the *NOAA-7* versus *-9* overlap occurs at the very end of *NOAA-7's* life. Thus, the bias determined for *NOAA-9* is now quite different, now being +0.458°C, and the overall 18+ yr trend becomes −0.016°C decade^{-1}. This is a clear example of how important the decisions one makes are in merging satellite data as they may have considerable impact on the resulting global trends. Because of the paucity of data with which to calculate the bias, the 1.75-h *NOAA-7* LCT drift, and the large σ_E, we conclude path A is not sufficiently robust to be utilized.

b. Path B

Path B takes the route of determining the bias of *NOAA-7* versus *-6*, then of *NOAA-8* versus *-7*, and finally *NOAA-9* versus *-8*. We therefore require the accumulation of biases of three overlapping pairs, one of which is quite brief (*NOAA-9* versus *-8*, about 50 useable days). This last bias calculation is the weak link in path B as it occurs at the end of *NOAA-8's* life after it had experienced operational problems and had been allowed to tumble for a year. Exactly what impact these problems may have had on the MSU is unknown, but we do see that σ_E is second worst for *NOAA-8* versus *-9*, even though the removal of *NOAA-8's* PAC probably artificially reduced σ_E (Fig. 8).

The global mean of the zonal biases for *NOAA-9* as calculated in path B is +0.410°C. This bias value is larger than calculated in path A, but less than path A1 generating an overall trend of +0.014°C decade^{-1}. The robustness of the overlap statistics (Table 2) reveal slightly better values than in path A.

As mentioned, we do not know how the MSU on *NOAA-8* was affected by the prolonged shutdown and subsequent tumbling. As an exercise, we calculate the trend in the difference between *NOAA-8* and *NOAA-7* over their 412-day overlap (path B1). For such short periods, slight difference trends will naturally occur between any two MSUs. This trend (a slight relative warm-

ing of *NOAA-8* versus -*7*) is then assumed to have been contaminating the *NOAA-8* data throughout its life and so is extrapolated to the period when *NOAA-8* was reactivated in 1985. (Such a procedure abounds with uncertainties but is shown here for illustrative purposes.) The resulting bias for *NOAA-9*, determined from direct comparison with ''detrended'' *NOAA-8*, is +0.495 essentially agreeing with the value of +0.49°C used in version b and producing an overall trend of −0.049°C decade^{-1}.

We had indicated that *NOAA-7* contains a slight spurious warming trend in the last year or so of operation. We adjust for this trend in *NOAA-7* and produce what we label path B2 in Table 2, while leaving *NOAA-8* untouched. The *NOAA-9* bias and overall trend are +0.428°C and +0.006°C decade^{-1}, respectively.

One further exercise may be performed, that of removing the late trend in *NOAA-7*, then the relative trend in *NOAA-8* versus the ''detrended'' *NOAA-7*. We call this route path B3. The resulting global mean bias for *NOAA-9* is +0.546°C and the overall trend is −0.081°C decade^{-1}. Immediately, one notices that the biases differ substantially (over 0.1°C) between the values in paths B, B1, B2, and B3, indicating path B and its variants are not very robust. In addition, the trend of the global time series by path B3 is considerably cooler than found in the nominal time series and is quite a bit cooler than observed in global radiosonde measurements. We conclude that path B is too uncertain for bias calculation, due to the brevity in the *NOAA-8* versus -*9* overlap, the relatively large value of σ_E, the reliance on three intervening biases, and the uncertain effect of *NOAA-8*'s hiatus.[6]

c. Path C

From a practical point of view, path C offers the most reasonable route for directly calculating the bias of *NOAA-9*. Not only do we avoid the possible error introduced through the use of intervening bias uncertainties, but this overlap contains the largest amount of data in which any satellite simultaneously observed the earth with *NOAA-9*. We see also that $\sigma_\Delta = 0.056$°C is lower than any of the other paths and that σ_E is less than half its nearest competitor (path B, Fig. 8). In addition to path C, we show the results computed with detrended *NOAA-7* (path C1) and detrended *NOAA-7* and *NOAA-8* (path C2). All path Cs here are very close in statistics and overall trends because the *NOAA-9* bias is identical in each variant. We conclude that path C1 is the most appropriate as the evidence for *NOAA-7* drift is substantial (CSM95) and we cannot show that the trend difference in *NOAA-8* is significantly different from zero (path C2).

TABLE 3. The interbiennial differences (1986/87 minus 1981/82) for six global temperature datasets are as follows.[a]

T_{2LT} (Path C1):	+0.06°C
Angell 850–300-hPa raobs[b]	+0.05
Oort and Liu 850–300-hPa raobs[c]	+0.07
U.K. Meteorological Office raob-simulated T_{2LT}[d]	+0.07
IPCC land and ocean surface	+0.07
NASA/GISS land/ocean index[e]	+0.05

[a] Annual anomalies of each dataset were normalized to the variance of the T_{2LT} dataset before the subtraction.
[b] Angell (1988) and updates.
[c] Oort and Liu (1993).
[d] Parker et al. (1997).
[e] Hansen et al. (1996).

The key question is, do we have confidence that the drift in *NOAA-6* (which during the *NOAA-9* overlap averaged 45–50 min from original LCT) has not affected the temperature to a significant degree as has been found for two P.M. orbiters? We discussed earlier our reasons for concluding that there is insufficient evidence to apply a drift correction to the end of the *NOAA-6* MSU time series. Here we also note that the drift of *NOAA-6* from 0730/1930 to 0630/1830 occurs at points in the diurnal cycle when less net temperature variations occur (Richner and Phillips 1984). Thus, until new information becomes available that indicates the presence of a net effect on the *NOAA-6* drift, we will assume it has an insignificant impact on the time series.

There are other datasets that will help in evaluating the decision to assume there is no significant drift in the temperature measured by *NOAA-6* between its first tour of duty and its last. From the global mean record of T_{2LT}, we select the years 1981–82 as years representing the first period of *NOAA-6* and 1986–87 as representing the end effect of the period over which *NOAA-6* had the greatest impact on the trend of the time series—that is, the relative anomalies for 1981–82 and 1986–87 are anchored by the *NOAA-6* record in path C1. We then take the difference between the two 2-yr averages and compare with those of other surface temperature and radiosonde datasets (Table 3). In each case, the variation from the value of T_{2LT} is very slight in comparison with the others. None of these radiosonde or surface comparisons is sufficient in and of themselves to prove a lack of drift in *NOAA-6* [e.g., surface values are not the same physical quantity as T_{2LT} (Hurrell and Trenberth 1996)]. Together, however, they strongly suggest that we have made a reasonable decision that path C1 is the best among several less than robust routes to calculate the bias in *NOAA-9*.

The bias of *NOAA-9* is estimated to be very close to 0.50°C (Table 4). With the path C1 choice thus made,

[6] We note there is a fourth path that may be used to calculate the bias of *NOAA-9*. We may assume that *NOAA-9's* MSU has the same bias as *NOAA-7* (+0.615 °C). However, subtracting a global mean bias of +0.615 °C from *NOAA-9* would result in a global trend even cooler than shown in path B3 and would be significantly cooler than radiosonde-simulated MSU temperatures.

TABLE 4. Column values are 1) the number of days used to produce the intersatellite biases in version c1, 2) the biases themselves for version c1, 3) estimated standard deviation (σ, size = $n - 1$) of the biases or trend (°C, °C decade^{-1}) produced from seven subsets (three thirds and four quarters) of the overlapping periods, and 4) as in column 3 but for six subsets each using half of the overlap data. In these 13 tests, the entire time series is recreated using no more than 50% of the overlap data to determine the intersatellite biases.

No. tests	# Days	$T_{2LT.c}$ bias	σ, thirds, fourths 7	σ, halves 6
TIROS-N	147	0.950	0.006	0.023
NOAA-7	543	0.615	0.003	0.011
NOAA-8	366	0.187	0.002	0.020
NOAA-9	184	0.497	0.002	0.006
NOAA-10	93	0.081	0.002	0.007
NOAA-11	905	0.573	0.005	0.007
NOAA-12	936	0.641	0.010	0.008
NOAA-14	642	0.703	0.007	0.007
Dec. trend		−0.046	0.007	0.006

it is worthy to note that the *NOAA-9* bias and overall trend produced in path C1 fall very near the center of the distribution of values calculated in a variety of ways listed in Table 2. If we use our previously estimated accuracy on the trend of T_{2LT} of ±0.03°C decade^{-1}, we note that paths A1, B1, C, and C2 are consistent with the trend of path C1.

We emphasize here that the extreme range of trend and bias results in Table 2 should not be misinterpreted. The values from paths A and B represent calculations based upon overlapping data periods with characteristics that are inferior to those utilized in path C. The sample sizes in A and B are clearly smaller and standard errors much greater. Thus the variation in trends in Table 2 is not applicable to the estimate of possible trend errors in path C1.

Utilizing path C1, then, the ''backbone'' of the MSU dataset in terms of long-term stability is formed by the three A.M. orbiters: *NOAA-6, -10,* and *-12*. The bias from *NOAA-6* to *NOAA-10* is calculated through *NOAA-9* and from *NOAA-10* to *NOAA-12* through *NOAA-11*.

7. Bias reproducibility

Now that the specific merging sequence is selected, we shall examine the reproducibility of the global mean biases shown in Fig. 9. In the following we shall provide results from methods that employ statistical techniques and empirical testing.

a. Reproducibility from statistical information

The global standard error of the intersatellite bias is found by calculating σ_E as defined earlier (Fig. 8) with N being one-third of the number of simultaneous observations, less 16 for those overlaps that have had PACs removed. Calculated in this way, the global values of σ_E range from 0.004°C for *NOAA-11* versus *-12* to 0.011°C for *NOAA-9* versus *-10* for $T_{err} = 0.325$. However, we may also estimate the reproducibility of the biases by reconstructing the time series with only portions of the overlapping data and comparing with these results.

b. Reproducibility from evenly sampled subsets of overlapping data

In this first set of tests we generate subsets of the overlapping periods that contain either one-third or one-fourth of the total overlapping data. This is accomplished by selecting for the first three subsets every third day and for the second four subsets every fourth day. For example, in the first subset containing one-third of the data we select days starting from January 1979 dates 1, 4, 7, . . . , and for the second third, January 2, 5, 8, . . . , etc. Analogous procedures were applied to the subset of fourths.

The individual members of each of these subsets are independent of the other members within the same subset. The subsets themselves, however, are quasi independent, as there will be some common members—for example, every third or fourth value will be shared between any ''third'' and ''fourth'' subset. These tests are designed to compare the robustness of the bias calculation using seven much smaller (though not quite independent) subsets of the original data and to monitor how bias errors might propagate to later satellites. The procedures for removing the PACs and trends are unchanged from T_{2LT}, and applied before the subset biases are calculated. In this way we could monitor the singular effect of subsampled data. The results (Table 4, second to last column) show that when using sample sizes even as small as one-quarter of the original, the critical biases (*NOAA-9, -10, -11,* and *-12*) are reproduced very well. We conclude then that by using all of the data in the overlap periods, we should know the biases with an error of less than 0.01°C.

c. Reproducibility from segments of overlapping data

We may further test the robustness of the bias calculation by dividing the overlapping data into three pairs of disjoint subsets using long segments of the data. In the first pair, we bisect each overlap period and reconstruct the complete time series with one of the pair utilizing only the first half of the overlap data, then the second of the pair using only the second half of overlap data. This would mimic the effect produced if every satellite either had failed earlier than was the actual case or was launched later than in reality. For the next pair of tests, we select the first and last quarters of the overlap data for the first of the pair, then the second and third quarters for the second of the pair. For the final pair of tests we choose alternating 25-day segments such that in the first test we choose days 1–25, 51–75, etc., and in the second, 26–50, 76–100, etc. Thus, in total, we

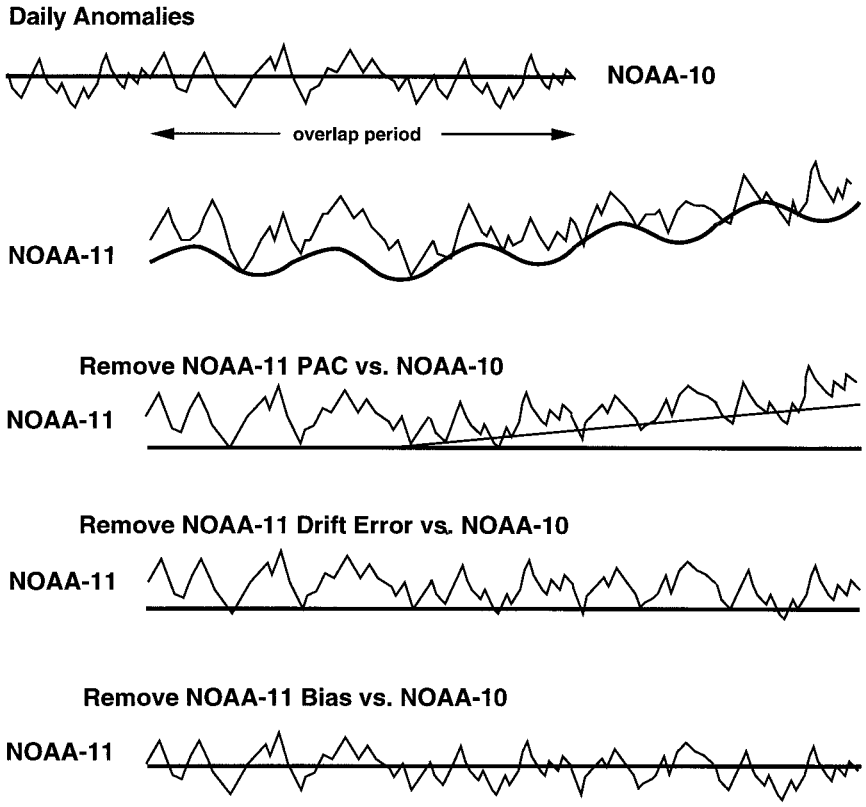


FIG. 11. Schematic diagram of the processes applied to *NOAA-11* to adjust the anomalies prior to merging. The features have been exaggerated for clarity.

have six additional reconstructions each based on only 50% of the overlap data.

The standard deviations (σs) of the six additional tests are given in the last column of Table 4. The variation of the biases in this second set of tests is greater than that of the first set of tests in which the overlap data were sampled evenly at every third or fourth entry. This is due to the presence of short periods of relatively poor data in one or both satellites, which are then included or excluded entirely from each reconstruction in the one set of the pair. For example, the pair of tests for which the overlap period of *TIROS-N* versus *NOAA-6* is bisected was the major source of the relatively large σ in the last column of Table 4 for these two. The first run of that pair, which includes the poor *TIROS-N* data in July 1979, contained 68 acceptable entries, whereas the second contained 79, hence there were differences in the quality of the halves. Though such variations occurred within halves of the data, the overall character of the time series suffered little since the key biases (*NOAA-9* through *-12*) are well-known.

A ''poor man's'' way to check the impact of these bias tests is to calculate the overall trend of each of the 13 test cases. The trends are all within ± 0.012°C decade^{-1} of the original trend. This tight range is due to the highly reproducible biases of the specific satellites involved in the critical overlapping periods that provide the backbone of the time series. The relatively high range of biases for *TIROS-N*, *NOAA-7*, and *NOAA-8* are of little consequence for the overall structure of the time series.

Since the calculation of the intersatellite biases utilizes the complete set of data in the overlap periods (i.e., N is twice as large as in the tests above), we would infer from Table 4 that the standard error is roughly 0.707 ($\sqrt{2}/2$) multiplied by the values in the last column. As such, we again conclude that the critical biases (*NOAA-9* through *-12*) are known to within about 0.01°C. This result agrees with the values calculated directly from σ_E earlier.

We now have described the necessary steps to perform the complete merging procedure beginning with the median-filtered anomalies of each satellite relative to the reference annual cycle of *NOAA-6* or *NOAA-7*. We first calculate and remove the bias of *TIROS-N* and *NOAA-9* relative to *NOAA-6*. Next, we calculate and remove the bias of *NOAA-10* relative to adjusted *NOAA-9*. We then calculate and remove, in order, the PAC, the drift error, and the bias from *NOAA-11* relative to the adjusted *NOAA-10*. Since this overlap involves all pro-

cedures described, we present a schematic of the procedure in Fig. 11.[7]

With the drift error of *NOAA-11* quantified, we then use these latitudinal values to remove the drift error from *NOAA-7* for the period described earlier. We calculate and remove the bias from *NOAA-7* relative to *NOAA-6*. For *NOAA-8* we calculate and remove the PAC and bias relative to adjusted *NOAA-7*. Recall that *NOAA-7* and *-8* are not utilized for the computations involving the three backbone satellites, so their adjustments may be performed at any time once the *NOAA-11* drift error is determined.

We then calculate and remove the PAC from *NOAA-12* relative to adjusted *NOAA-11*. The drift error of *NOAA-11* is calculated and removed relative to *NOAA-12* for their common overlap period. Then, the bias of *NOAA-12* is calculated and removed relative to readjusted *NOAA-11* data. Finally, the PAC and bias of *NOAA-14* relative to adjusted *NOAA-12* is calculated and removed. The entire procedure beginning with median-filtered anomalies is sketched out in Fig. 12.

It is important to note that we have dealt exclusively with the daily zonal-mean anomalies of the MSU datasets in this paper. A gridpoint dataset that relies on the monthly zonal anomalies from the daily datasets discussed here is also produced. However, the gridpoint datasets incorporate several simple spatial interpolation schemes to fill in the data gaps between the swaths so that local or gridpoint errors are enhanced. The datasets we describe here are produced before and completely separately from the gridpoint datasets, the procedures for which will be described in a future publication.

d. Radiosonde check

Given the existence of global radiosonde-based tropospheric datasets, we are able to compare the overall trends since 1979 with those of T_{2LT} (Table 5). Though the radiosonde datasets are constructed with different stations and methods and contain a unique set of problems (C95), all of the trends reported here are within ± 0.03°C decade^{-1} of T_{2LT}. These results provide independent information to support the previously estimated precision of the T_{2LT} trend for the globe of ± 0.03°C decade^{-1} though a value of ± 0.05°C decade^{-1} is recommended here.[8] Additionally, these comparisons support the choice of path C1 as the best route for satellite intercalibration as certain versions of paths A and B would have produced much larger differences between the satellites and the radiosondes.

8. T_2 and T_4

The several measures of error magnitudes are much smaller for T_2 and T_4 MSU datasets. In Fig. 13 we show the correlation and σ_E of latitudinal differences between anomalies for the three overlapping satellite pairs shown in Fig. 7 for T_2 and T_4. Correlations are nearly all above 0.95, except in the subtropical latitudes of T_2 (see section 5a above), and for T_4 are often above 0.995. This astounding agreement lends high confidence to the calculation of the quantities required in the merging procedure for these two products. With much lower errors, we shall require for the calculation of intersatellite biases and the *NOAA-11* trend only those days on which T_{err} is less than 0.20°C for T_2 and 0.25°C for T_4 (recall for T_{2LT} we set 0.325°C as the threshold leaving 80% of data for use). This threshold retains 96% of T_2 comparisons and 97% of T_4. The application of a threshold in producing the intersatellite biases and *NOAA-11* trend, the adjustment of the drift error in *NOAA-7* and the use of harmonics one through eight for determining the PACs are all adopted for T_2 and T_4.

The error characteristics of T_{2LT}, T_2, and T_4 are given in Table 6. We point out that T_2 and T_4 are almost unchanged in trend. Note that $\sigma_{error} = \sigma_\Delta/\sqrt{2}$ except that here, σ_Δ is calculated using every possible day on which two satellites observed the earth (i.e., no constraint of T_{err} is applied). Since global values of these days are calculated as the mean of the two available satellites, we estimate the daily error as σ_{error}. This value applies only to those days for which two satellites were in operation, encompassing about 80% of the time series. The entire time series of the three MSU temperature products are displayed in Figs. 14a–c.

9. Comments on studies of previous versions of T_{2LT}

Recently, two similar papers (Hurrell and Trenberth 1997, 1998, both hereafter HT) suggested that the earlier versions of T_{2LT} data (a but mostly b) were prone to errors such as spurious jumps of up to 0.25°C, which inflicted on the time series global trend errors of about 0.1°C decade^{-1}. Considering the levels of confidence reported here for version c1, and that the levels are only slightly improved from version b, these are remarkably significant claims if they are shown to be real. Though most aspects discussed about T_{2LT} in HT may be challenged, we shall here briefly focus on what we believe to be the three most significant claims regarding the satellite time series. [A very limited statement of these issues appears in Christy et al. (1997).]

[7] Note, however, that the post-*NOAA-10* data of *NOAA-11* will not be completely adjusted until comparison with *NOAA-12* (Fig. 12).

[8] Several Australian stations had slight adjustments made to their time series based on MSU comparisons, so that the Southern Hemispheric comparison vs Parker et al. is ''mostly'' but not completely independent.

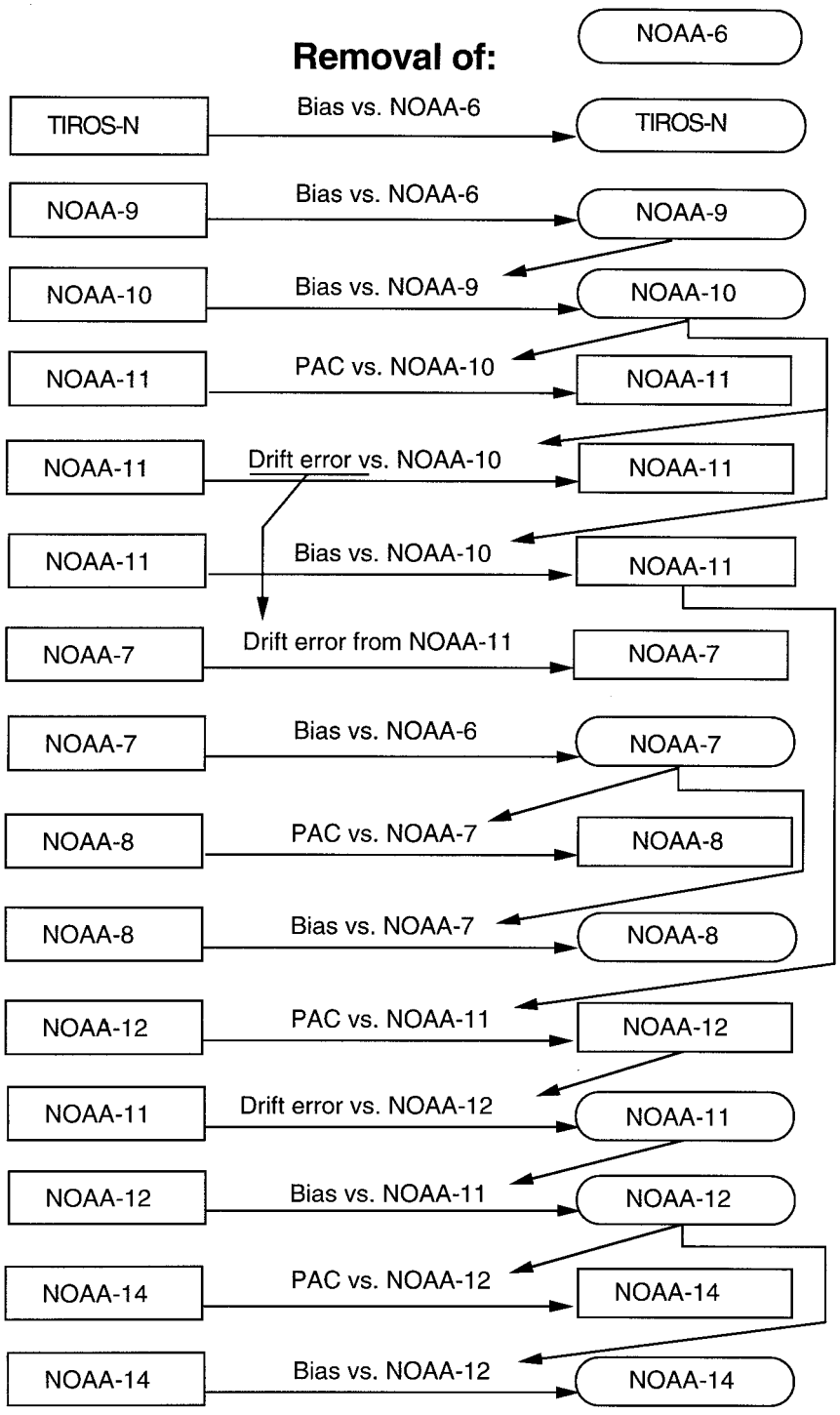


FIG. 12. Schematic diagram of the steps performed to adjust the anomalies of each instrument to the level of *NOAA-6*. Borders represent the dataset of daily, zonal-mean anomalies for each MSU on the various spacecraft. Those framed with an oval border indicate the process is complete for that instrument. Icons framed with a rectangle require further steps.

TABLE 5. Comparisons of trends since 1979 for T_{2LT} vs various radiosonde-based tropospheric datasets, which, except for the 850–300-hPa-layer temperature, are weighted to match the MSU T_{2LT} weighting function (Fig. 1).

	No. stations	Trend °C decade^{-1}	T_{2LT} trend for same region °C decade^{-1}	Difference (radiosonde minus MSU)	Years
Global (850–300 hPa)[a]	63	−0.06	−0.04	−0.02	1979–96
No. Hemisphere[b]	250+	+0.01	+0.02	−0.01	1979–96
So. Hemisphere[b]	50+	−0.11	−0.08	−0.03	1979–96
W. No. Hemisphere[c]	97	+0.16	+0.14	+0.02	1979–94

[a] Angell (1988) and updates.
[b] Parker et al. (1997) using simulated T_{2LT} from maps of objectively analyzed radiosondes.
[c] Stations are in an area roughly bounded by Truk, South Pacific, to Pt. Barrow, Alaska, to Keflavik, Iceland, to Trinidad. This is a comparison of radiosonde-simulated T_{2LT} with colocated T_{2LT} from the gridpoint dataset, which depends on the zonal dataset described here.

a. Confidence in the magnitude of the intersatellite biases

The accuracy of the intersatellite biases is important because errors in these offsets would produce an inhomogeneous time series. To estimate the precision of these biases for the 20°S–20°N lat band, HT utilized error statistics from a gridpoint dataset that does not have the benefit of daily median filtering and requires simple spatial interpolation that increases gridpoint noise. Therefore it does not provide the necessary information to judge the accuracy of the bias for a *zonal*

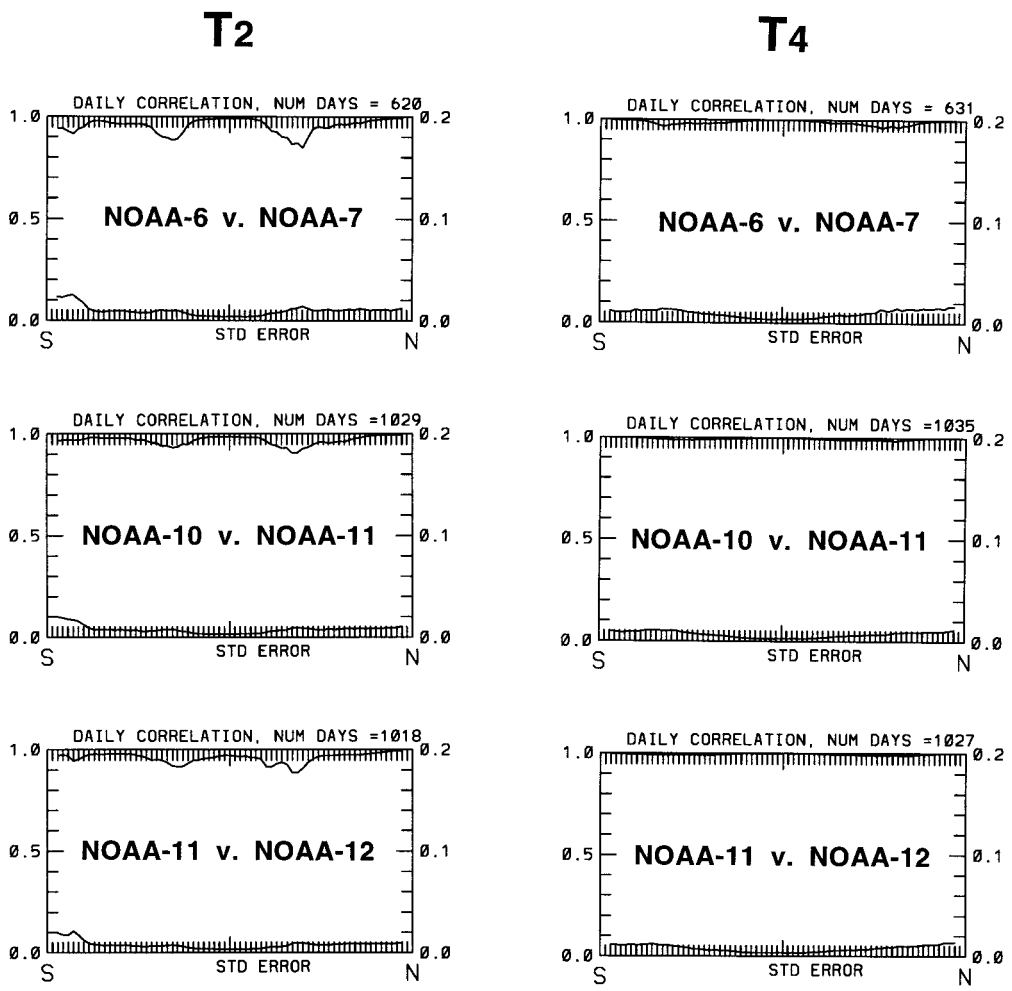


FIG. 13. As in Fig. 7 for T_2 (left) and T_4 (right).

TABLE 6. Error characteristics of the three MSU products for the earlier version b and the present version c1 (January 1979–April 1997). Here σ_{error} is $\sigma_{\Delta}/\sqrt{2}$ between two simultaneous satellite readings for all days, whether meeting the T_{err} criteria or not (°C). Also, S/E is the ratio of the variances for the daily average global anomalies of the two simultaneous satellite readings vs the average difference for all days in which a difference could be determined (i.e., again, no constraint from T_{err}). Trends are °C decade^{-1}.

	σ_{error}	σ_{error}	S/E	S/E	Trend	Trend	T_{err}
Product	b	c1	b	c1	b	c1	c1
T2LT	0.055	0.050	25	30	−0.076	−0.046	0.325
T2	0.030	0.027	75	90	−0.002	0.003	0.200
T4	0.031	0.031	237	242	−0.482	−0.485	0.250

band since the zonal anomalies of the gridpoint dataset come directly from the zonal datasets discussed here. Thus, we are able to calculate directly (not inferentially) the standard errors of the biases as was done for the globe earlier.

Actual $\sigma_{\text{E Tropics}}$ values from version c1 range from 0.005 to 0.011°C, with the main source of the variations being the number of independent observations (N) as $\sigma_{\Delta\ \text{Tropics}}$ are all similar in magnitude. The estimate of HT for a single month (0.05°C) was exaggerated due to the source of their calculation (gridpoint data) and was incomplete since they did not incorporate the temporal extent of the periods of overlapping observations.

b. NOAA-7 *as the cause of a possible spurious jump in late 1981*

Hurrell and Trenberth (1997, 1998) also suggest that two spurious jumps appear in the tropical T_{2LT} record, one in late 1981 (0.25°C) and one in late 1991 (0.10°C). Their suggestion is based on a comparison of T_{2LT} against simulated tropospheric temperatures estimated from sea water (i.e., surface) temperatures (SSTs). Their simulated values are produced from a linear regression equation of tropically averaged SSTs versus T_{2LT} over a calibration period of January 1982–December 1991, hence both alleged jumps occur at the edges of their carefully selected calibration period. Their view is that the addition of *NOAA-7* into the database is the source of the largest jump in 1981 and *NOAA-12* is responsible for the 1991 jump.

Since the alleged jump in 1981 is very large, we shall examine this event closely. In Fig. 15, we show the actual monthly tropical anomalies of *NOAA-6* and *NOAA-7* for 1980–83 relative to their reference annual cycles. *The only dependency in these two anomaly time series is the common time period chosen for the reference annual cycle; no temperature values are communicated between them.* Each time series is calculated independently and are the actual anomalies used in T_{2LT}. It is clear that with or without *NOAA-7,* the anomalies and their variation for 1980 to April 1983 are not influenced by *NOAA-7* to a magnitude approaching 0.25°C as suggested by HT.[9]

Perhaps it is possible that both *NOAA-6* and *NOAA-7* independently experienced a spurious 0.25°C jump at nearly the same time in late 1981. We examine several tropical radiosonde profiles during this period to check this possibility. We choose only the 2 yr prior to the alleged jump (December 1979–November 1981) and the 2 yr after (December 1981–November 1983) to largely (or completely) eliminate any radiosonde instrumental changes that may confound the signal we seek. The values we compare are calculated as collocated sonde-simulated T_{2LT} versus observed T_{2LT}. Table 7 displays the 2-yr differences on either side of the break for the radiosonde T_{2LT} minus MSU T_{2LT}. The next to last column is the difference of the differences and shows no value approaching +0.25°C, which, if HT were correct, would be evident.

Both MSU instruments on *NOAA-6* and *NOAA-7* behaved exceptionally well during this period. *NOAA-6* (the anchor) remained within ±5 min of LCT and no anomalous variations in the redundantly monitored target temperatures of either instrument occurred. Given the information in Fig. 15 and Table 7, we conclude that if HT's SST regression formula is accurate, then the tropically averaged vertical temperature profile changed by a couple of tenths of a degree in late 1981.

It is important to note that in the final time series of T_{2LT}, we average the values from each satellite, such as those in Fig. 15, and that we have accounted for diurnal effects because each is referenced to its own point in the diurnal cycle. In addition, the MSU datasets are not, as HT imply, constructed as time series requiring transitions from one satellite to the next as is the procedure of OLR datasets because these lack long-period overlapping data (Walliser and Zhou 1997). All MSU time series are created by true merging of satellite data utilizing the hundreds of thousands of independent observations contained in the overlapping periods. Over 80% of the T_{2LT} time series (over 96% since 1988) contains data from two simultaneously orbiting spacecraft.

c. Additional independent evidence for T_{2LT} anomalies

In Fig. 16 we show the annual tropically averaged (20°S–20°N) differences between anomalies of SSTs and night marine air temperatures (NMATs; Parker and Folland 1991; Parker et al. 1995). The solid line represents differences for all SST and NMAT observations (including SSTs from daytime observations when temperature variations due to depth are greatest) and the dashed line for collocated observations only (i.e., ships and buoys reporting coincident NMATs and SSTs).

[9] A similar plot was produced for the 1991 period in which it is evident that *NOAA-12's* data did not introduce a spurious jump in the tropical average temperature.

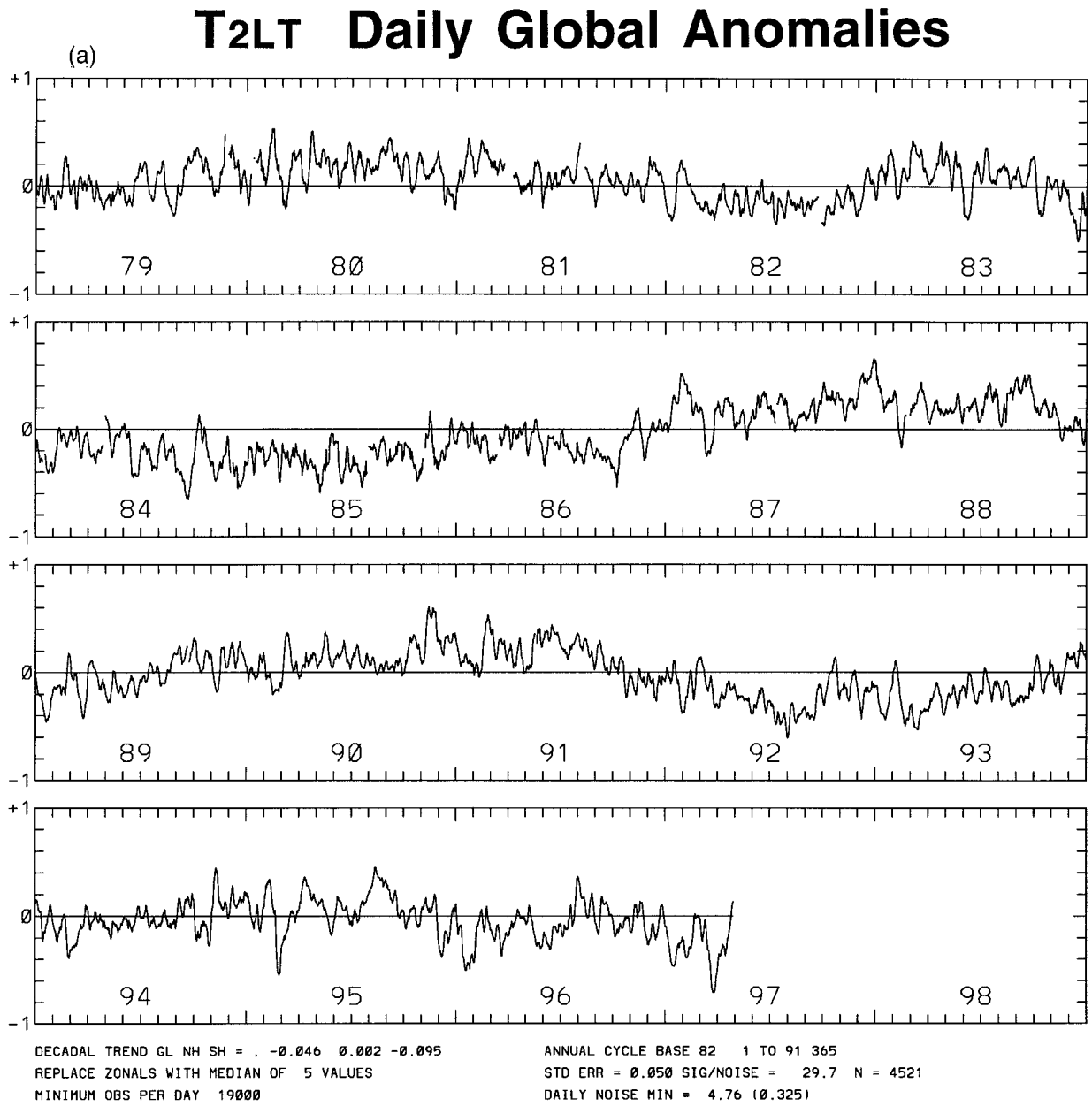


FIG. 14. Time series (January 1979–April 1997) of daily global anomalies of (a) T_{2LT}, (b) T_2, and (c) T_4. The base period is 1982–91 for T_{2LT} and T_2 and 1984–90 for T_4.

These are provisional values as the dataset is undergoing constant updating and quality-control processing (D. Parker 1997, personal communication). It is clear in these comparisons, however, that since 1979, and especially since 1991, there are different changes in SSTs and NMATs, the trends of the differences being -0.07 (solid) and -0.09 (dashed) °C decade^{-1} for 1979–96.

Recognizing the uncertainties in both SSTs and NMATs (see Parker et al. 1995), the variations in Fig. 16 imply that the vertical stability of the low-level tropically averaged ocean–atmosphere system may have fluctuated since 1979. The variations after 1991 are pronounced and may be due in part to impacts of the Mt. Pinatubo volcano. Indeed, low-level ocean–atmosphere variations have been reported from independent measurements of the Tropical Atmosphere–Ocean buoy array. In the equatorial Pacific, SSTs were consistently anomalously warmer than atmospheric temperatures (ATs) for 5 yr (1990–94) and were over 1°C warmer during the 2 yr beginning in mid-1991 (165°E–160°W; Weisberg and Wang 1997).

Upon further analysis of the more proven, noncollocated dataset (Fig. 16, solid line), we discover that the tropical Pacific and Indian Oceans are the source regions

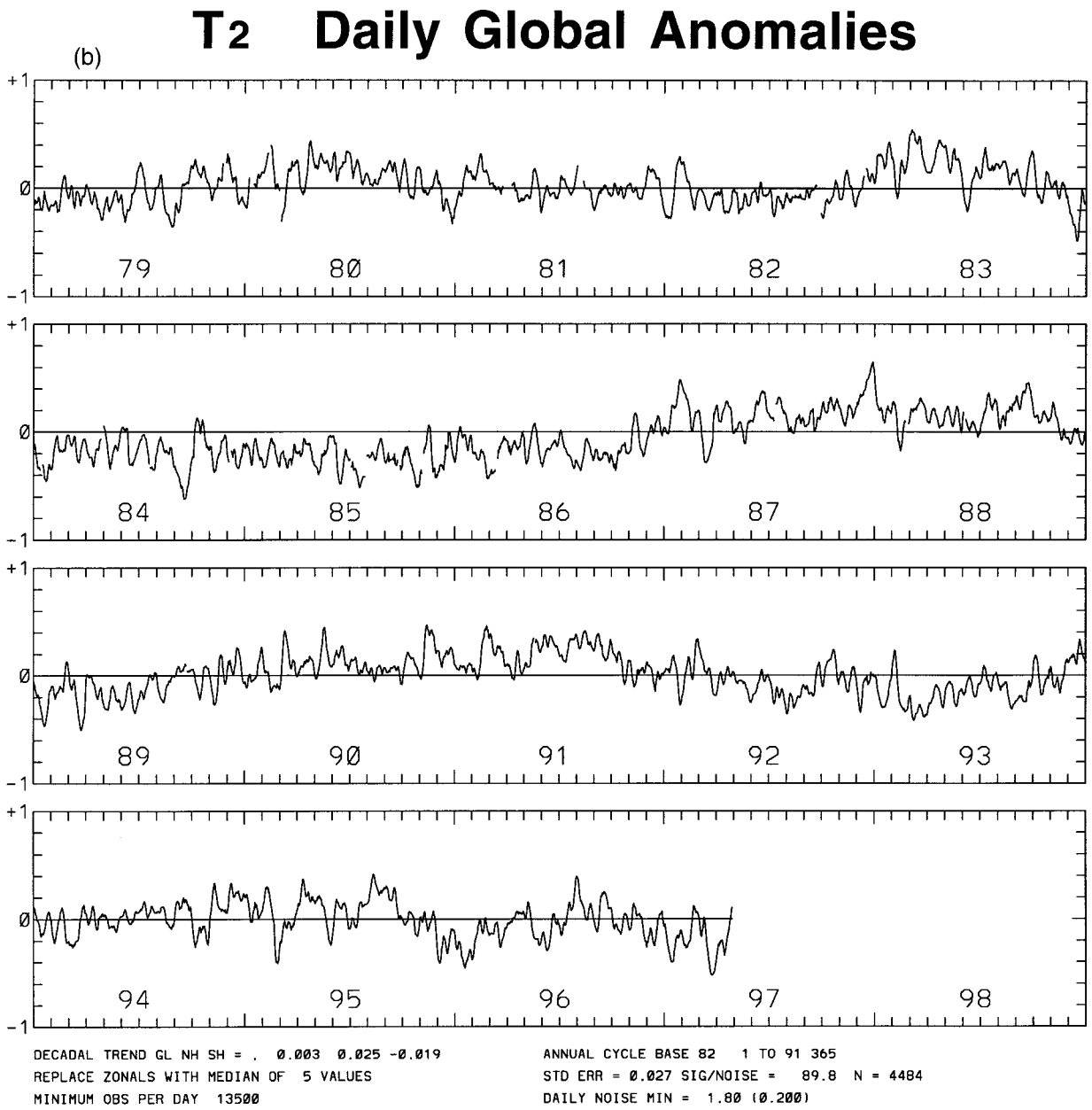


FIG. 14. (*Continued*)

for the two apparent shifts (1981–82 and 1991–92).[10] The magnitude of each of the shifts in these regions is about 0.09°C (average of 1979–81 minus 1982–84 and average of 1987–91 minus 1992–96) with NMATs becoming cooler relative to the SSTs in both cases. Because of partially compensating variations between 1984 and 1987, the total relative difference between the average of 1979–81 and 1992–96 is 0.14°C. (These difference variations between SSTs and NMATs are absent in the tropical Atlantic.) Thus relatively long-term difference variations between SSTs and ATs, on the order of at least 0.1°C in the Indian–Pacific region, are supported by the various measurements shown above.

The difference between tropical SSTs and ATs over this 18-yr period might also be in part a manifestation of the rapid increase in SST reports from buoys in the otherwise poorly monitored tropical Pacific. Buoy-depth temperatures (~0.5 m) in the western tropical Pacific can be over

[10] The subregions of greatest differences between NMATs and SSTs in the Indian and Pacific Oceans are the ''Flohn'' (10°S–15°N, 65°E–180°) and ''Niño 3.4'' (5°S–5°N, 170°–120°W) for which trends of the differences averaged −0.11°C decade^{-1}.

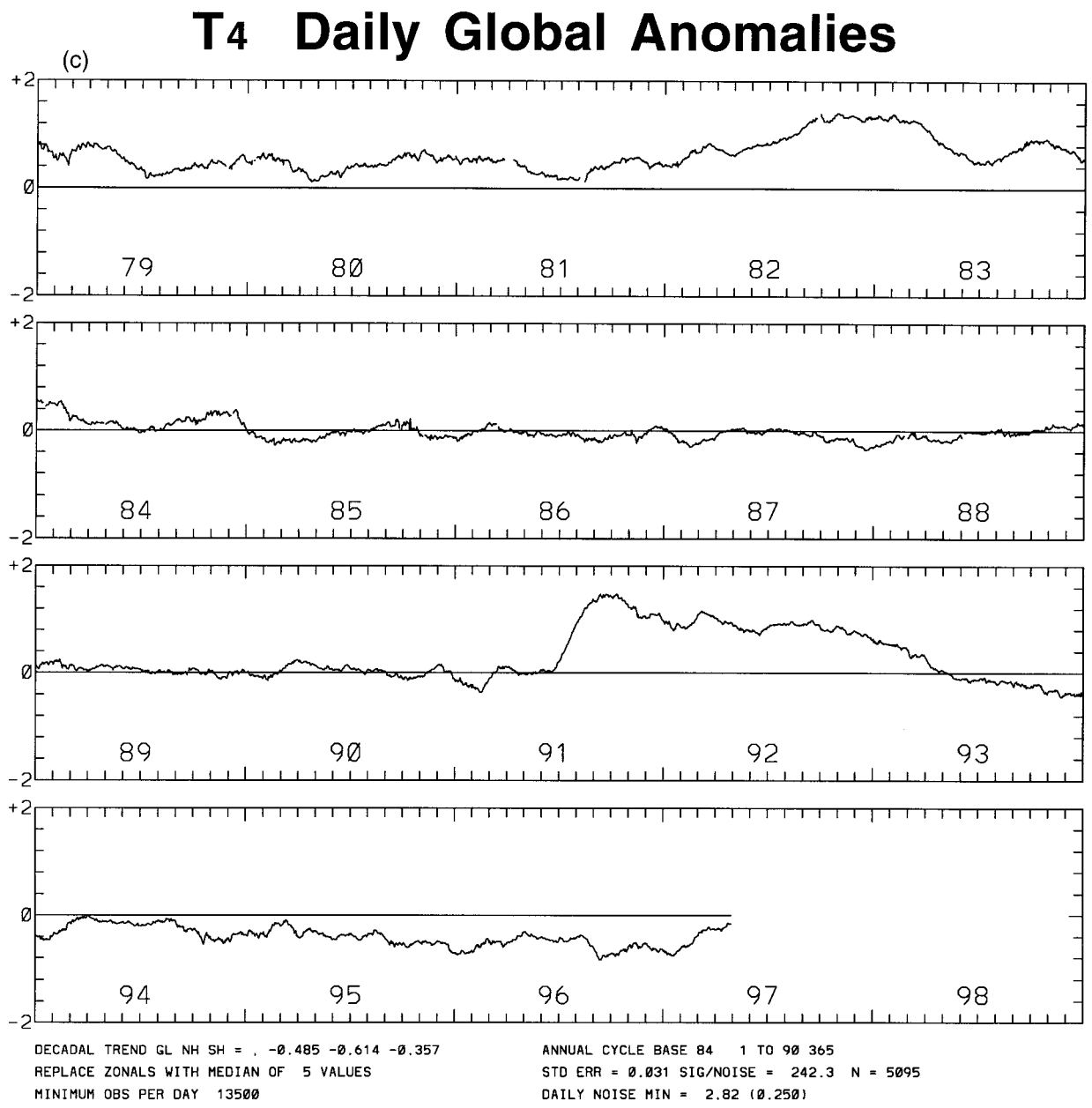


FIG. 14. (*Continued*)

0.5°C warmer on average than traditional ship-depth reports (~5 m) in calm sea conditions, which prevail in the region. Even under rarer, windy conditions, the temperatures at 0.5 m are on average 0.1°–0.2°C warmer than at 5 m (Webster et al. 1996). It is possible, therefore, that a portion of the relative positive trend in the SSTs is due to artificial warming of the SSTs related to the changing mix of observational platforms in poorly sampled tropical regions. This is only speculation at this point and is currently under investigation.

In any case, we have three independent measures of atmospheric temperature [T_{2LT}, radiosondes (see also Parker et al. 1997), and NMATs] that indicate that long-term variations in the relationship of SSTs versus ATs on the order of 0.2°C do occur. Linearly regressing tropical T_{2LT} based on tropical SSTs over a specific calibration period therefore will likely generate errors of that order for periods outside of the calibration period. This is apparently what has happened in the procedure of HT so that their reconstruction of tropospheric temperatures from SSTs is inaccurate. Furthermore, the evidence from independent satellite observations confirms that the specified spurious jumps in the tropical T_{2LT} data did not occur.

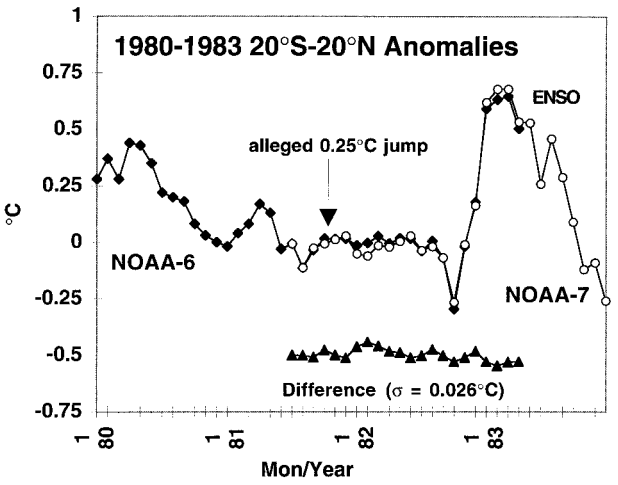


FIG. 15. Independent anomalies of 20°S–20°N produced from *NOAA-6* and *NOAA-7*. Anomalies in both time series are departures from the reference annual cycle based on September 1981–August 1982.

10. Conclusions

We have shown that the long-term trend of the lengthening record of the MSU products is highly dependent on the manner in which biases are calculated from one instrument to the next, especially for the lower troposphere, T_{2LT}. By examining the possibilities for producing the most crucial bias relative to *NOAA-6*, that of *NOAA-9*, we demonstrate that a direct method, path C1, is preferred as shown by the statistical comparisons. Independent surface and radiosonde data support path C1 as not introducing a spurious trend in the overall time series.

Three changes have been instituted in the merging process as a result of this study. First, we now use only certain days, those on which the best agreement between two satellites occurs, for the calculation of the intersatellite biases and the *NOAA-11* drift error. Second, we now remove an assumed spurious temperature drift error in the *NOAA-7* MSU as was observed in *NOAA-11*. Third, we have increased the resolution of the PACs from harmonics one through five to one through eight, which are removed from some satellites. The combination of these changes causes the 18+ year trend of T_{2LT} to be warmer by $+0.03$°C decade^{-1} (-0.076 to -0.046°C decade^{-1} for January 1979–April 1997). We estimate the precision of the overall trend as ± 0.05°C decade^{-1}, which is consistent with independent tropospheric measurements from multistation radiosonde datasets. Similar changes in merging procedure applied to T_2 and T_4 have little effect. For future versions of the MSU datasets we will account for the largely compensating effects in T_{2LT} of satellite orbit decay (artificial cooling, F. Wentz, personal communication) and interannual variations in instrument-body temperature (artificial warming).

We examined recent claims that large and spurious jumps are present in the tropical T_{2LT} time series. We found no such jumps by comparison with independent satellite and traditional atmospheric measurements. Our conclusion is that estimates of tropical atmospheric temperature variability derived solely from SSTs are not accurate to a level less than 0.2°C.

We have presented here the method used in generating the Spencer–Christy MSU temperature datasets. These datasets should not be confused with the many types of satellite data employed in the construction of global meteorological analyses utilized for climate studies. In general, global analyses include sophisticated and complicated satellite retrievals of atmospheric quantities based on radiances observed by multiple sensors in the visible, infrared, and microwave spectra. These retrievals of temperature and humidity (among other quantities) are primarily applied in operational settings and therefore are not subject to careful determination of instrumental biases (which we show are on the order of tenths of a degree) as has been done with the products discussed here. These retrievals have considerable impact on the temperature and humidity in the analyses over oceanic regions, which have no local radiosonde input.

TABLE 7. Direct, collocated comparisons of radiosonde-simulated T_{2LT} (°C) and MSU T_{2LT} in the tropical latitude belts identified. Under columns 1980–81 and 1982–83 are the differences (radiosonde minus MSU) of the 2-yr averages of their temperatures. Diff. is the difference of these differences (1982–83 minus 1980–81) and shows no value approaches +0.25°C. (Note: −0.00 indicates the value is between −0.005 and 0.000 °C.) The last column is the standard deviation of the individual 2-yr differences over nine 2-yr periods.

	# Stations	1980–81	1982–83	Diff.	$\sigma_{\Delta\ 2\text{-yr}}$
10°S–10°N	7	+0.01	−0.03	−0.04	0.068
15°S–15°N	14	+0.00	−0.03	−0.03	0.058
20°S–20°N	30	−0.01	+0.00	+0.01	0.049
25°S–25°N	38	−0.01	+0.00	+0.00	0.041
30°S–30°N	49	−0.00	+0.00	+0.01	0.038

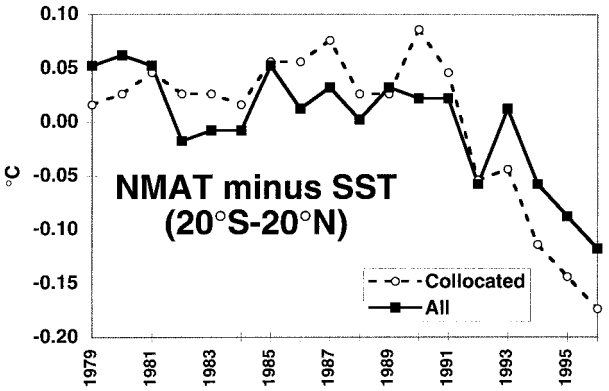


FIG. 16. Tropically averaged, annual anomaly differences between NMATs and SSTs (20°S–20°N) from U.K. Meteorological Office [MOHSST6, Parker et al. (1995) with updates]. Solid line represents differences based on all observations and dashed line differences of coincident (collocated) observations only.

Preliminary comparisons show excellent agreement between lower-tropospheric temperatures produced from reanalyses projects and the MSU product T_{2LT} over North America, Eurasia, and Australia (National Centers for Environmental Prediction: Basist and Chelliah 1997; European Centre for Medium-Range Weather Forecasts: Stendel and Bengtsson 1997). Ironically, the reanalyses and T_{2LT} are most *independent* over these regions because the analysis methods do not typically rely on satellite retrievals wherever radiosondes are available. However, over oceans, where satellite retrievals (which include some MSU information different from the products described here) are the dominate source of input, trend differences between the reanalyses and the T_{2LT} are noticeable. This disagreement introduces the possibility that complicated satellite retrievals are probably not stable enough to determine sensitive quantities such as the decadal temperature variations and trends for many regions. The reason, in part, is likely related to the inability of analysis schemes and/or retrieval schemes to detect and adjust for the small biases that we know exist between instruments on separate spacecraft.

Other issues are raised in this study regarding the operation of future satellite systems, which may be relied on to produce climate-quality datasets. First, we have shown that satellites must have overlapping observations with instruments already in space for a period of at least 1 yr. Second, it is highly desirable to document the effects of solar illumination through an annual cycle on each instrument after it has been integrated into the spacecraft system and before launch. Finally, polar-orbiting satellites are far more useful for climate studies if their LCTs are maintained to within about ± 15 mins so that the diurnal cycle will not influence the desired climate signal.

Acknowledgments. The research reported herein was funded by the NASA Pathfinder Program NAGW-4749 and NOAA/NASA Joint Program for Enhanced Datasets for Analysis and Applications. We thank David Parker for the time series of the U.K. Meteorological Office SSTs and NMATs and for several suggestions for the manuscript. In addition we thank two anonymous referees who provided rigorous, thorough, and helpful reviews of earlier versions of the manuscript.

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. *Climate Change*, **31,** 455–474.

——, and S. J. Drouilhet, 1994: Variability in daily, zonal mean lower-stratospheric temperatures. *J. Climate,* **7,** 106–120.

——, 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.

Goldberg, M. D., and H. E. Fleming, 1995: An algorithm to generate deep-layer temperatures from microwave satellite observations for the purpose of monitoring climate change. *J. Climate,* **8,** 993–1004.

Hansen, J., H. Wilson, M. Sato, R. Ruedy, K. Shah, and E. Hansen, 1995: Satellite and surface temperature data at odds? *Climate Change,* **30,** 103–117.

——, R. Ruedy, and M. Sato, 1996: Global surface air temperature in 1995: Return to pre-Pinatubo level. *Geophys. Res. Lett.,* **23,** 1665–1668.

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.

Kidwell, K. B., 1995: NOAA polar orbiter data users guide. National Oceanic and Atmospheric Administration, 410 pp. [Available from NOAA, Federal Office Bldg. #3, Room G-233, Washington, DC 20233.]

Mo, T., 1994: A study of the microwave sounding unit on the *NOAA-12* satellite. *Proc. Int. Geoscience and Remote Sensing Symp.,* Pasadena, CA, California Institute of Technology, 1535–1537.

Nicholls, N. 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, 133–192.

Oort, A. H., and H. Liu, 1993: Upper-air temperature trends over the globe, 1958–1989. *J. Climate,* **6,** 292–307.

Parker, D. E., and C. K. Folland, 1991: Worldwide surface temperature trends since the mid-19th century. *Greenhouse-Gas-Induced Climate Change: A Critical Appraisal of Simulations and Observations,* M. E. Schlesinger, Ed., Elsevier Science, 173–193.

——, ——, and M. Jackson, 1995: Marine surface temperature: Observed variations and data requirements. *Climate Change,* **31,** 429–470.

——, 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.

Richner, H., and P. D. Phillips, 1984: A comparison of temperatures from mountaintops and the free-atmosphere—Their diurnal variation and mean difference. *Mon. Wea. Rev.,* **112,** 1328–1340.

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

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

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

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

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

Walliser, D. E., and W. Zhou, 1997: Removing satellite equatorial crossing time biases from OLR and HRC datasets. *J. Climate,* **10,** 2125–2146.

Webster, P. J., C. A. Clayson, and J. A. Curry, 1996: Clouds, radiation, and the diurnal cycle of sea surface temperatures in the tropical western Pacific. *J. Climate,* **9,** 1712–1730.

Weisberg, R. H., and C. Wang, 1997: Slow variability in the equatorial west-central Pacific in relation to ENSO. *J. Climate,* **10,** 1998–2017.