

## Temporal Homogenization of Monthly Radiosonde Temperature Data. Part II: Trends, Sensitivities, and MSU Comparison

JOHN R. LANZANTE AND STEPHEN A. KLEIN

*NOAA/Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, New Jersey*

DIAN J. SEIDEL

*NOAA/Air Resources Laboratory, Silver Spring, Maryland*

(Manuscript received 19 September 2001, in final form 24 July 2002)

### ABSTRACT

Trends in radiosonde-based temperatures and lower-tropospheric lapse rates are presented for the time periods 1959–97 and 1979–97, including their vertical, horizontal, and seasonal variations. A novel aspect is that estimates are made globally of the effects of artificial (instrumental or procedural) changes on the derived trends using data homogenization procedures introduced in a companion paper (Part I). Credibility of the data homogenization scheme is established by comparison with independent satellite temperature measurements derived from the microwave sounding unit (MSU) instruments for 1979–97. The various analyses are performed using monthly mean temperatures from a near-globally distributed network of 87 radiosonde stations.

The severity of instrument-related problems, which varies markedly by geographic region, was found, in general, to increase from the lower troposphere to the lower stratosphere, although surface data were found to be as problematic as data from the stratosphere. Except for the surface, there is a tendency for changes in instruments to artificially lower temperature readings with time, so that adjusting the data to account for this results in increased tropospheric warming and decreased stratospheric cooling. Furthermore, the adjustments tend to enhance warming in the upper troposphere more than in the lower troposphere; such sensitivity may have implications for “fingerprint” assessments of climate change. However, the most sensitive part of the vertical profile with regard to its shape was near the surface, particularly at regional scales. In particular, the lower-tropospheric lapse rate was found to be especially sensitive to adjustment as well as spatial sampling. In the lower stratosphere, instrument-related biases were found to artificially inflate latitudinal differences, leading to statistically significantly more cooling in the Tropics than elsewhere. After adjustment there were no significant differences between the latitude zones.

### 1. Introduction

Long-term variations in the horizontal and vertical temperature structure of the atmosphere play an important role in the detection and attribution of climate change. While recent efforts have resolved some outstanding issues involving temperature measurements near the surface and in the free atmosphere [National Research Council (NRC) 2000], progress has been slow in addressing the critical issue of temporal continuity of radiosonde temperature measurements (Free et al. 2002). In a companion paper (Lanzante et al. 2003), hereafter referred to as Part I, examples have been presented suggesting that in some instances historical changes in instruments and observing practices can have a large impact on the low-frequency character of tem-

perature time series through the introduction of artificial discontinuities or “changepoints.”

Motivated by these concerns, procedures have been developed (Part I) to identify artificial discontinuities (changepoints), as well as other maladies, and to reduce their influence through modification of the temperature time series. Modification consists of either adjustment to remove the artificial discontinuity or deletion of a portion of a time series if adjustment is not feasible or appropriate.

Because historical changes in instruments vary greatly by country, and often by station as well, an a priori assessment of their impacts on the global or regional field of temperature is not possible. For this purpose, our procedures have been applied to radiosonde temperatures from a select near-globally distributed network of 87 stations, with a limitation on the number of stations necessitated by the labor-intensive nature of the methodology. Temperature trends, and a few other measures, along with the sensitivities of these quantities to

---

*Corresponding author address:* Dr. John R. Lanzante, NOAA/Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, NJ 08542.  
E-mail: jrl@gfdl.noaa.gov

the manner in which data modifications were applied, are reported. The sensitivities serve to quantify the uncertainties in the trends. Both the vertical and horizontal distribution of trends are explored, as well as such issues as lower-tropospheric lapse rates, seasonality, and temporal evolution of large-scale temperatures. Finally, our results are compared with those from an independent dataset, satellite temperatures (Christy et al. 2000) from the microwave sounding unit (MSU).

Section 2 provides descriptive statistics that summarize the nature of our data modifications. While section 3 presents vertical distributions of temperature trends along with measures of sensitivity to data modification, section 4 provides similar horizontal distributions and sensitivities. Section 5 reexamines historical variations in lower-tropospheric lapse rate (Gaffen et al. 2000; Brown et al. 2000) in light of data quality uncertainties. Section 6 explores the seasonality of temperature trends. Section 7 presents the temporal evolution of temperatures over large spatial scales. In section 8 our data modifications are evaluated via comparison with MSU temperatures. A summary and conclusions are given in section 9.

## 2. Summaries of data modifications

As described in detail in Part I, the original temperature time series were modified in two fundamental ways: 1) changepoint adjustment and 2) data deletion. For the former, first each artificial discontinuity was identified and presumed to be associated with an instantaneous change in instrument or recording practices, and then an adjustment was applied to remove its effects. When adjustment was not feasible or appropriate, a portion of a temperature time series was deleted instead. To allow for assessment of the uncertainty associated with such actions, changepoint identification was performed using two levels of confidence: conservative (CON) and liberal (LIB); we deemed the former to be more confidently identified than the latter. Subsequent adjustment was performed in two ways using either a simple scheme, nonreference level adjustment, or a more complex one, reference level adjustment. A set of scenarios were defined based on various combinations of data deletion, changepoint identification, and adjustment (section 3 and Table 1 of Part I). While in this section statistics are presented for some of the scenarios, in sections 3 and 8 systematic comparisons of trend results are made using all scenarios.

Nearly two-thirds (65%) of all changepoints are of the more confident type (CON), which reflects our philosophy not to alter the data without a compelling reason. These are assigned when either station history information is indicative of instrumental or procedural change at the time of a discontinuity, or a discontinuity is large compared to the variance of the time series.

The temporal variations of data modifications are indicated by the two stair-step curves shown in Fig. 1,

while the total amount of data, prior to deletions, is given by the thick curve. Data deletions vary strongly as a function of time. Early in the record there were more gaps in the data and more problems of a nondiscontinuous nature, both of which hamper changepoint adjustment, necessitating deletion. Deletions are more frequent at both ends of the record since sufficient data must be present both before and after a changepoint to make a reasonable adjustment. The spike prior to the 1957 global observation time shift is a reflection of many time series that began just prior to this time. Part of the increase in deletions at the end of the record is due to the widespread natural steplike drop in stratospheric temperatures around 1992–93 (Part I, section 5a), which sometimes hinders adjustment. As inferred from the sharp increase in data availability, depicted by the thick curve, a substantial fraction of the available data was deleted up through the 1950s (~30%), whereas during the later two decades the deletion rate was substantially lower (~5%). Over the entire period of record the deletion rate is ~10%. By contrast, the time series of number of adjusted changepoints is more uniform, with two prominent spikes worth noting, the 1957 shift in time of observation and the Soviet instrument changes in the late 1960s.

Figure 2 shows the vertical variation of the number of data modifications expressed as a fraction of the number of months of data available in the unadjusted dataset (i.e., before any modifications are made). For the highest levels, 10 and 20 hPa, where the available data are typically quite sparse, severe data deletions were imposed because we were frequently unable to declare homogeneity with any degree of confidence; this leaves much less data available for assignment of changepoints. The deletion and changepoint rates are higher in the longer period of record (1949–97) than for the more recent period of time, corresponding to the satellite era (1979–97). This is not surprising because the technology and operating procedures of the early years were far less advanced than today and thus more prone to producing sudden changes when improvements were instituted. Except for the highest levels, the deletion rate is reasonably uniform in the vertical whereas the rate of assignment of changepoints increases somewhat from the lower troposphere upward. The latter is consistent with the theoretical work of Luers and Eskridge (1998), which suggests that the magnitude of discontinuities should increase upward, coupled with our hesitation in assigning a changepoint when the magnitude of the jump is small.

There are relative maxima near the surface for both types of data modifications in Fig. 2. Surface readings from a radiosonde sounding generally are not made with the radiosonde equipment but are instead made at the collocated surface observation station using a different type of instrumentation. The surface data used here are the values reported in the radiosonde soundings and, thus, are not necessarily identical to those found in other

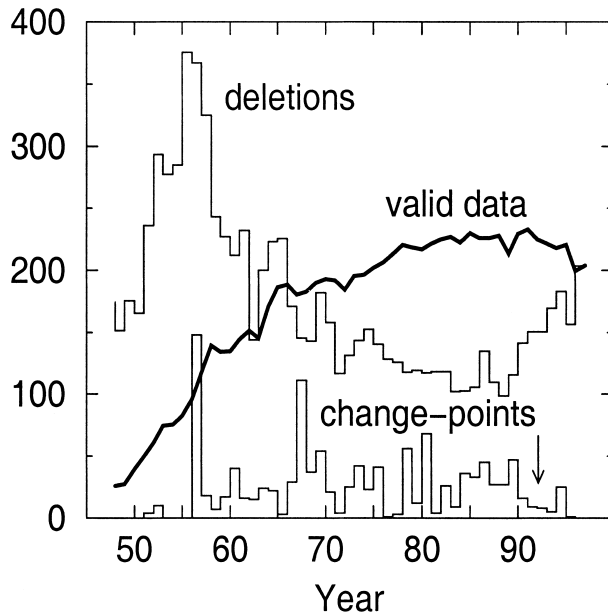


FIG. 1. Time series of annual totals of deletions (top stair-step curve), change-points (bottom stair-step curve), and number of valid data values (thick curve) summed over all stations, levels and observation times. Note that for display purposes the top curve is normalized as the number of months deleted divided by 10 while the thick curve is normalized as the number of months of valid data divided by 100. The bottom curve is based on LIBCON change-points (i.e., the union of all LIB and CON change-points) while the thick curve is based on unadjusted (UNADJ) data.

surface temperature datasets, which may involve further processing. Furthermore, data from the 1000-hPa level can be as problematic as a result of missing reports when the surface pressure is less than 1000 hPa; in such instances reported data may be fabricated by extrapolation below the surface, or if left missing would bias monthly means. Since this work relies on vertical coherence in the identification of change-points, and since boundary layer effects impair the ability to utilize this tool, it may be that the severity of data problems near the surface have been underestimated in the data modification process.

As indicated by Table 1, there is a preponderance of negative over positive adjustments. Since the adjustment is defined as the value added to the segment before the change-point, negative adjustments lead to less cooling or more warming. The underlying artificial cooling may be due to improvements in radiosonde temperature sensors or algorithms that tend to decrease the effects of solar radiation errors (Part I, sections 5b,d,f). As indicated by Table 1, negative change-points dominate in all decades except the 1970s, which are influenced by the artificial rises seen at many Soviet stations (Part I, section 5e).

Vertical profiles of the absolute value of the adjustments (right curves, Fig. 3) show that except for the near-surface levels, the adjustment magnitude generally increases upward from the lower troposphere. While this

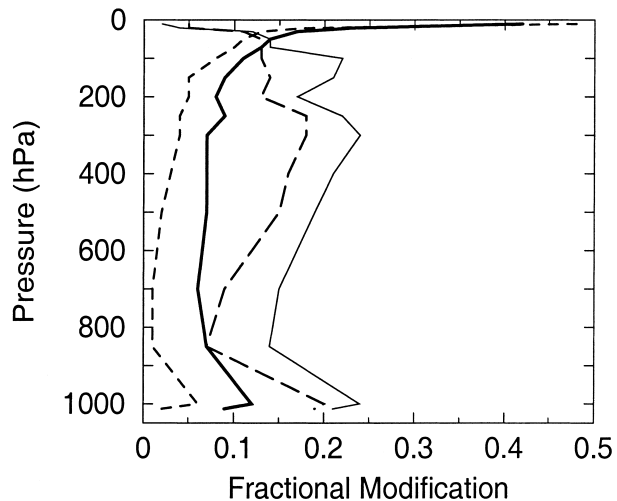


FIG. 2. Fractional deletions and change-points summed over all stations and observation times as a function of level. Deletions are indicated by the thick solid curve for 1949–97 and by the short dashed curve for 1979–97. Change-points are denoted by the thin solid curve for 1949–97 and by the long dashed curve for 1979–97. Deletions are reported as the number per month of valid data while change-points (LIBCON) are the number per decade (120 months) of valid data.

is consistent with the expectation that lower air density at higher elevations enhances instrument bias due to solar radiation, other causes are possible. Two factors may contribute to the near-surface maximum: 1) different instrumentation is often used at the surface and may partially influence derived 1000-hPa temperatures, and 2) the enhanced amplitude of the diurnal cycle near the surface may magnify the effects of instrumental bias. Adjustment values in the lower troposphere are  $\sim 0.5$  K, near the surface and upper troposphere are  $\sim 0.75$  K, and in the stratosphere  $\sim 1.0$ – $1.25$  K. While the overwhelming majority of adjustments are less than 2 K, the largest approach 5 K. Such adjustments are significant given that we find the standard deviation of monthly temperature anomalies is typically  $\sim 0.5$ – $2.5$  K.

Because of the preponderance of negative over positive adjustments (Table 1), one might expect that the median adjustment (left curves) should be approximately the mirror image of the median absolute value of the adjustment (right curves), as displayed in Fig. 3. The exceptions to this are at the surface as well as at and adjacent to the 250-hPa level; in both cases there is little bias in the *sign* of the adjustment. The former is less surprising since different instrumentation is used at the surface. For the latter, Soviet and Australian stations have a substantial contribution, although the causes are unknown.

Figure 3 also compares adjustment for the LIBCON (solid) and NONREF (dashed) scenarios, which use the same change-points, but differ in the method of adjustment (see Part I, section 3d). The magnitude of the adjustment based on the reference level scheme is slightly less, as expected, suggesting that on average a small

TABLE 1. Total number of LIBCON changepoints by time period, stratified according to algebraic sign of the associated adjustment value. By convention, the sign of the adjustment is the same as the sign of the artificial trend induced by the changepoint. These numbers are based on sums over all stations, levels, and observation times.

Sign	Time period					
	1950–59	1960–69	1970–79	1980–89	1990–97	1950–97
+	65	83	152	127	32	459
–	122	231	118	160	79	710
Total	187	314	270	287	111	1169

fraction of the jump across each discontinuity is taken to be natural when the more complex LIBCON adjustment scheme is used. It is important to stress that the composite adjustment profiles shown in Fig. 3 are not typical of adjustment profiles for individual changepoints; often the individual profiles are more complex, being less smooth and even discontinuous in the vertical, with adjustments isolated in some discrete layer.

### 3. Vertical structure

#### a. Computation and summary of trends

Trends computed in this study have been estimated using “median of pairwise slopes” nonparametric regression. Other statistical measures used here are nonparametric as well, with the benefits articulated by Lanzante (1996, 1998). Trends are estimated for two time periods: 1979–97 and 1959–97. The former starts at the beginning of the period of record of MSU data used for satellite comparisons in section 8. The latter begins after the 1957 global, 3-h shift in observation times that had a disproportionately large effect on the data, as seen in Fig. 1; furthermore, data quantity and quality decline rapidly prior to this. The longer ~40-yr period of record is more appropriate for the study of climate change as assessed by trend analysis (Stott and Tett 1998; Santer et al. 2000), and is the focus of this paper.

One complicating factor is that the record length varies considerably by country and level; fewer data are available for developing countries, particularly in the early years, and at higher altitudes. To ensure that reported trends are reasonably representative of the nominal time period, 1959–97 or 1979–97, trends are reported for a level/station only if at least half of the months have valid data in each third of the nominal time period. Although each station has separate time series for either one or two observation times (0000, 1200, or 9900 UTC, where 9900 refers to mixed times), only one trend is reported for each station, using the average of separate 0000 and 1200 UTC trends when both are available.

Trend results are summarized using medians of all station trends, for a particular level or layer, for one of three latitude zones: Northern Hemisphere extratropics (NH; 30°–90°N), Tropics (TRPC; 30°N–30°S), and

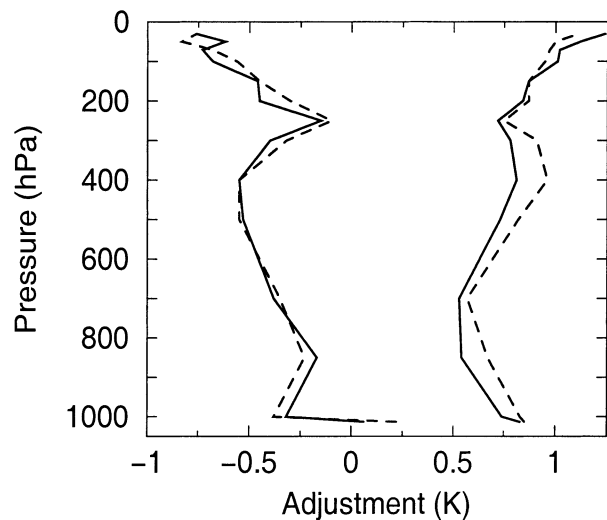


FIG. 3. Median adjustments (left curves) and median absolute value of adjustments (right curves) over all stations and observation times as a function of level. Solid curves are based on the LIBCON scenario and dashed curves correspond to the NONREF scenario (see Part I, Table 1 for details of scenarios).

Southern Hemisphere extratropics (SH; 30°–90°S). Using this scheme, the number of stations varies by time period, level, and scenario due to data deletions (Table 2). Although the more problematic surface level has slightly fewer stations, the numbers of stations are fairly stable throughout the troposphere. Data modification through deletion has only a slight effect on the number of stations available for trend calculations. However, sparsity of data and data deletions hamper analysis at the highest levels. Using the longer period of record, the drop-off is not too severe for the NH and TRPC; however, lack of stations is a problem for the SH even in the more recent period. Any conclusions drawn from SH aggregates should be viewed tentatively.

Some results are also summarized using layers (hPa) which consist of the aggregate of several levels: 50–100 (lower stratosphere), 150–250 (near tropopause, except upper troposphere in the Tropics), 300–500 (upper troposphere), 700–850 (lower troposphere), and the surface. Results are also reported for the 10–30-hPa layer, but these are less reliable due to lack of data. Also, the 1000-hPa level is usually disregarded due to a lesser quantity of data as well as concerns about the fabrication of data in at least some cases, as discussed in section 2.

#### b. Trend sensitivity and bias for each data modification scenario

Our first consideration is the sensitivity of station temperature trends to data modification, using the scenarios defined in Table 1 of Part I. Here Table 3 indicates the percentage of station trend values that are influenced by a particular detail of the methodology. These results are summarized by vertical layer and separately for the



TABLE 2. Number of stations by latitude zone for a particular scenario (UNADJ or DEL), time period (1959–97 or 1979–97), and selected levels (20, 50, 100, 200, 300, 850 hPa, or the surface). Each triplet of numbers corresponds to the NH (30°–90°N), TRPC (30°N–30°S), and SH (30°–90°S). These values are the number of stations for which sufficient data exist to report a trend over the nominal time period (see text for definition of “sufficient data”). The last row (upper limit) represents the number of stations that would be available if no data were missing.

Level (hPa)	1959–97		1979–97	
	UNADJ	DEL	UNADJ	DEL
20	9/8/0	7/7/0	21/9/0	16/9/0
50	21/13/5	17/13/4	32/21/10	26/19/9
100	37/22/7	30/21/5	34/28/15	29/25/14
200	37/24/9	33/23/8	34/30/16	31/28/16
300	37/25/9	36/24/8	35/32/17	32/29/17
850	37/24/8	37/23/7	35/29/16	34/29/16
Surface	30/22/8	29/18/7	33/29/15	32/29/15
Upper limit	37/33/17	37/33/17	37/33/17	37/33/17

two time periods used for trend estimation, 1959–97 and 1979–97. Each column in Table 3 can be used to assess the influence of a particular aspect of the data modification process. For example, “U–D,” “D–C,” “C–L,” and “L–N” can be used to infer the effects of data deletions, CON changepoints, LIB changepoints, and nonreference level adjustment, respectively. The last column, U–L, is a measure of the combined influence of data deletions and the adjustment of CON and LIB changepoints using the reference level scheme; it represents the effects of our preferred LIBCON approach. The pair of values in each cell indicate the percentage of trend estimates changed by more than a slight amount, on the left, or by a statistically significant amount, on the right. The fact that the 1959–97 values almost always exceed those from 1979–97 indicates that more stations were affected during the earlier time period when the measurement technology and procedures were less sophisticated. Trends are not particularly sensitive to the inclusion of the less confident changepoints (C–L) or to the scheme used to perform adjustment

(L–N). Data deletions (U–D) and adjustment of CON changepoints (D–C) have a greater effect. For 1959–97 the LIBCON scheme (U–L) has some impact on about half of all stations, and a significant impact on about a quarter, while for 1979–97 some impact occurs for about a third and significant impact for about a sixth. Consistent with the vertical profiles given in Figs. 2 and 3, Table 3 also shows that for the more serious modifications (i.e., data deletions and CON changepoints, as well as the cumulative effects in LIBCON) the surface data quality has more in common with the upper rather than the lower troposphere.

A large-scale perspective to the issue of sensitivity is given in Table 4, which displays the relative change in global temperature trend as a function of the manner of data modification. The row/column structure is similar to that of Table 3 except that additional columns ( $T_L$ ) report the LIBCON global trends. Generally speaking, as in Table 3, Table 4 shows that data deletions (U–D) and conservative changepoints (D–C) tend to have more influence than the other factors. Since Tables 3 and 4 present complementary information (the former examines local and the latter global sensitivities) a correspondence between the results is not guaranteed. Table 4 estimates the net global bias in trend while Table 3 tallies how many station trends are affected, without regard to sign of adjustment. It is worth noting that the values for 1979–97 will tend to be larger than those for 1959–97 because regression estimates have larger sampling variability for shorter periods (Santer et al. 2000); this is reflected by fewer significant values on the right half of Table 4 in spite of some larger magnitudes.

The vertical structure of the global trends ( $T_L$ ) indicates tropospheric warming, significant only for the longer period, and even stronger stratospheric cooling, significant during both time periods. Lower-stratospheric trends are roughly twice as great during the more recent period because, as illustrated in section 7, there is little cooling prior to about 1980. By contrast the tropospheric warming is much stronger for 1959–97.

TABLE 3. Percentage of stations whose trends differ between two scenarios. The paired scenario differences (column headings) are as follows: U–D (UNADJ – DEL), D–C (DEL – CON), C–L (CON – LIBCON), L–N (LIBCON – NONREF) and U–L (UNADJ – LIBCON). The scenarios are defined in Part I, Table 1. Each cell consists of a pair of values that represent the percentage of all station/level values, for a particular layer and time period, for which trends differ slightly (left value) or significantly (right value) between the two scenarios being compared. The degree of difference is defined using the Spearman correlation coefficient (Lanzante 1996) associated with the trend relationship. A slight difference occurs when the correlation differs between the two scenarios by at least 0.01. Testing for a significant difference was accomplished by applying the Fisher  $z$  transformation to the correlation coefficients and then using a  $z$  test (Zar 1996) based on the standard error defined by the effective sample sizes, which were computed from the lag-1 autocorrelation of the time series (Laurmann and Gates 1977).

Layer (hPa)	1959–97					1979–97				
	U–D	D–C	C–L	L–N	U–L	U–D	D–C	C–L	L–N	U–L
10–30	32/17	28/10	3/0	10/5	51/25	21/18	16/6	2/0	8/4	34/22
50–100	29/17	29/13	13/4	11/1	55/30	21/17	19/7	4/1	7/0	37/23
150–250	30/11	27/14	19/5	11/2	56/27	12/9	17/3	8/1	9/1	34/13
300–500	25/6	31/13	22/6	24/4	55/25	12/8	13/3	10/2	11/0	33/12
700–850	21/4	23/6	14/2	15/2	48/13	7/3	8/1	7/3	5/0	21/5
Surface	26/13	31/21	12/2	17/3	49/33	8/2	17/9	16/5	15/2	31/15

TABLE 4. Difference between median trends ( $K \text{ decade}^{-1} \times 100$ ) for two scenarios for a particular layer and time period. For a given scenario, the median trend is computed from the pool consisting of trends from all stations, for all levels within the given layer. The paired scenario differences (column headings) are as defined in Table 3. The last column, " $T_L$ ," is not a difference, rather the LIBCON trend. Values significant at the 5% level are in italics with 1% significance indicated by bold italics. The robust rank-order test (Lanzante 1996) has been used to assess whether the medians for each pair of scenarios are significantly different. Significance of the trends ( $T_L$ ) have been determined by applying the binomial test (Siegel and Castellan 1988) to the pool of values. To try to account for the fact that the station/level trend values in the pool are not all independent, due to horizontal and vertical correlation, a conservative approach is employed such that the 1% level is required for claims of significance; values passing at the 5% level are considered borderline or suggestive.

Layer (hPa)	1959–97						1979–97					
	U–D	D–C	C–L	L–N	U–L	$T_L$	U–D	D–C	C–L	L–N	U–L	$T_L$
10–30	–1	–12	–1	0	<i>–14</i>	<i>–41</i>	6	0	0	–1	6	<i>–117</i>
50–100	2	–5	–1	–1	<i>–5</i>	<i>–41</i>	–8	–3	–3	–4	<i>–15</i>	<i>–77</i>
150–250	–2	–3	0	0	–5	0	–1	–3	–1	0	–4	<i>–21</i>
300–500	–2	<i>–1</i>	1	1	–2	<b>13</b>	–1	–2	0	2	–3	4
700–850	0	–1	1	1	0	<b>14</b>	–2	0	0	1	–2	2
Surface	1	4	–1	2	5	<b>8</b>	–4	5	–1	3	0	12

The column labeled U–L is our best estimates of bias in global temperature trends due to changes in instruments and measurement practices and represents the first estimates of their kind. These can now be factored in to the ongoing effort to reconcile trends derived from various temperature datasets (Santer et al. 1999). Except for the surface, whose measurements are based on non-radiosonde equipment, and the highest layer during 1979–97, which has far less data than the other levels (Table 2), the biases are all negative and tend to have greater magnitude at higher elevation, in accord with Fig. 3. Thus, according to our assessments, the unadjusted data overestimate the stratospheric cooling by  $\sim 10\%$ , which is enough to yield a statistically different trend for both time periods. Table 4 also suggests an underestimate of tropospheric warming; while this is not significant for 1979–97, the estimated bias is comparable to the trend, so that virtually all of the warming is due to adjustment.

Table 4 also shows that during the satellite era the surface warmed more than the troposphere while for the longer time period the opposite was true. The adjustments do not greatly alter the relative warming between the surface and free troposphere during 1979–97, but for the longer period the surface adjustments are much more influential. For 1959–97 the unadjusted data suggest nearly equal warming, but after adjustment the surface warms noticeably less than aloft. While we are less confident in our treatment of the surface (due to the complexities of the boundary layer and the use of non-radiosonde equipment) this finding does raise some concern as to the ability to assess variations in the tropospheric lapse rate. A more detailed examination of this issue is given in section 5.

Summing up Tables 3 and 4, the largest impacts are produced by data deletions and the inclusion of CON changepoints; the distinctions between the confidence level of the changepoints and between the two adjustment schemes are of secondary importance. On this basis, as well as comparisons with satellite data (section 8), further analyses are limited mostly to a comparison

between unadjusted (UNADJ) and LIBCON. The latter is our preferred method of adjustment because it includes LIB changepoints, which denote features we consider artificial, and uses reference level adjustment, which we believe is more likely to adjust the data in a vertically consistent manner.

### c. Trends by latitude zone

Temperature trends computed for each of the three latitude zones as a function of pressure level are displayed in Fig. 4. A number of statistical tests have been performed on these values but in the interest of brevity only key findings are summarized in reference to features in Fig. 4. The significance of the latitude zone trends have been assessed using the binomial test (see caption for Table 4). Significance of differences between latitude zone trends, UNADJ versus LIBCON, and one zone versus another have been assessed using the robust rank-order test (see caption for Table 4). The fraction of stations whose trends are locally significant have been determined using a  $z$  test (see caption for Table 3). Following the rationale given in the caption for Table 4, the 1% level is used for claims of statistical significance while the 5% is used for borderline significance. All significance tests have been applied to vertical layers as defined in section 3a.

Some degree of caution is needed in interpreting the physical significance of the trend profiles in Fig. 4. Since the number of stations used to define the trend for a particular latitude zone varies by pressure level and scenario, the trend profiles may have some additional component of sampling variability. However (see Table 2), this is not a great concern in the troposphere where the numbers of stations vary only slightly. The use of the median (over all stations) to summarize the trends in Fig. 4 further protects against sampling variations. However, the stratosphere is more problematic, due to the considerable decrease in available stations with altitude. This concern is somewhat offset by the more zonally symmetric nature of the stratospheric climate. Never-

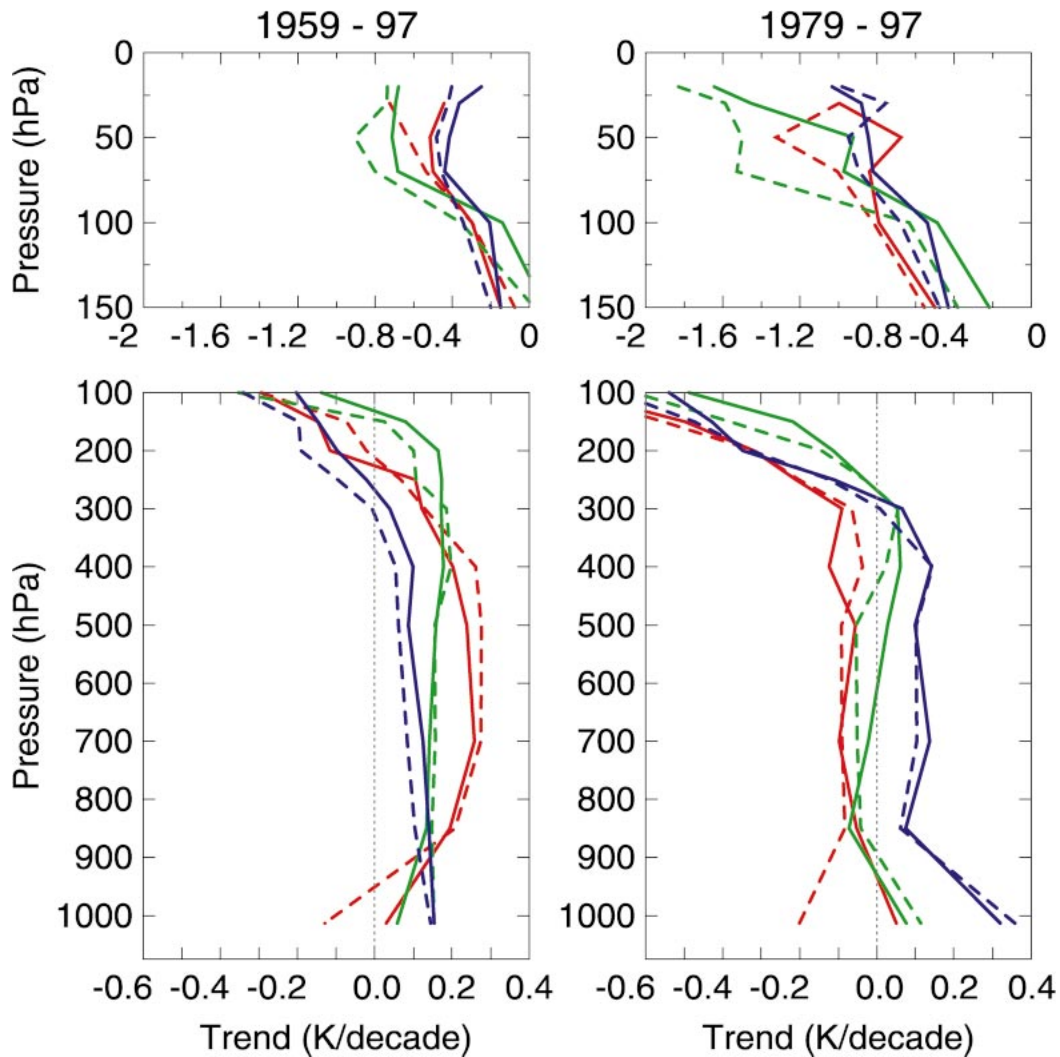


FIG. 4. Temperature trend ( $\text{K decade}^{-1}$ ) as a function of pressure level (hPa) (left) for 1959–97 and (right) for 1979–97. Because of the different ranges, trend profiles have been plotted separately for (top) stratospheric and (bottom) tropospheric levels, with different increments on the abscissa. Trend is median over all stations in a particular latitude zone: NH ( $30^{\circ}$ – $90^{\circ}$ N, blue), TRPC ( $30^{\circ}$ N– $30^{\circ}$ S, green), and SH ( $30^{\circ}$ – $90^{\circ}$ S, red). Each zone has curves based on unadjusted (UNADJ, dashed) and adjusted (LIBCON, solid) data.

theless, the trend estimates should be viewed with less confidence higher in the stratosphere.

The trend profiles shown in Fig. 4 are consistent with the widely accepted notion of tropospheric warming and much stronger stratospheric cooling. Globally the stratospheric cooling is significant in both time periods while the tropospheric warming is only significant for the longer one. In accord with the tendency for instrument changes to lead to artificial cooling with time, the adjusted trend profiles (solid) are typically slightly more positive than the unadjusted ones (dashed). In the lower stratosphere, UNADJ cooling is greater than LIBCON, especially for the TRPC during 1979–97. Unadjusted TRPC trends are statistically significantly less than adjusted ones during 1979–97 and borderline for 1959–97; global stratospheric trends (not shown) are signifi-

cantly less for both time periods. Prior to adjustment, the TRPC zone has a significantly more negative trend than the NH during both time periods, and the SH during 1979–97 (1959–97 is borderline significant); after adjustment the TRPC zone is not statistically different from the other zones. This finding has important implications for the validation of GCMs being used to study the effects of anthropogenically induced changes in greenhouse gas and ozone concentrations since, according to our estimates, the apparent latitudinal differences in the stratosphere are not real. While we believe we have removed a major part of the stratospheric latitudinal bias, later (section 8) we suggest that the TRPC may still be showing too much cooling, and the NH too much warming.

While the magnitude of the estimated trend bias is

larger in the stratosphere, the troposphere has a comparable relative bias, typically  $\sim 10\%$ , except higher during 1979–97 in zones in which the trend approaches zero. However, trend relationships are more complex in the troposphere in terms of features that cannot be explained by artificial effects: 1) differences in trends among latitude zones and 2) relationships between latitude zones that are quite different for the two time periods of analysis. Of the three zones, the NH has the least warming during 1959–97 and the most during 1979–97, while the opposite is true for the SH; warming for the TRPC is intermediate during both times. Adjustment alters this ordering in only one instance, resulting in a statistically significant difference between TRPC and SH in the upper troposphere for 1979–97, which was not the case prior to modification.

Although data adjustment results in some qualitative differences in tropospheric trends, most of the changes do not attain statistical significance. While for 1979–97 none of the changes are significant, for 1959–97 the increased warming is significant for the NH upper troposphere and is borderline significant for the TRPC upper troposphere. For global trends (not shown) data modification results in a borderline significant additional warming during 1959–97 of the upper troposphere and near-tropopause regions. Particularly for the TRPC, adjustment shifts the level of maximum trend upward. These changes may have some relevance to the attribution of global warming. In this regard some idealized data-only calculations (not presented) have been performed aimed at assessing the potential impact of adjustment on fingerprint attribution analysis; observed trend patterns have been used in lieu of GCM fingerprints. In summary, these analyses suggest that for 1959–97 our adjustments may reduce the signal strength of stratospheric cooling by  $\sim 10\%$ , but increase the tropospheric warming signal by as much as  $\sim 20\%$ – $40\%$ . Confirmation of these results awaits the use of GCM-derived fingerprints.

In light of recent controversy (NRC 2000), the behavior of trends at the surface and their relationship to those in the troposphere merits discussion. For the satellite era, enhanced warming of the surface relative to the troposphere is robust to adjustment for the NH and TRPC, however, a similar relationship is found in the SH only after adjustment. The surface in the SH shows considerable sensitivity to adjustment such that the sense of the relationship is reversed. For 1959–97, only the NH is insensitive to adjustment, whereas adjustment warms the surface considerably for the SH and cools it for the TRPC. Note that in Fig. 4 UNADJ TRPC and LIBCON NH are nearly coincident at the surface. Although none of the latitude zone surface trends differ significantly between LIBCON and UNADJ, for 1959–97 a high percentage of the individual stations in the SH (50%) and TRPC (41%) have significantly different trends. By comparison, typical percentages in the lower stratosphere are only  $\sim 20\%$ – $30\%$ . It would seem that

surface data are a major source of uncertainty, particularly for the earlier periods of record. A related, more detailed analysis of sensitivities in terms of the lapse rate is given in section 5.

#### 4. Horizontal structure

##### *a. Tropospheric trends*

Although maps of the horizontal distribution of trends have been prepared for each standard level, for each of the two time periods, 1959–97 and 1979–97, owing to the considerable vertical coherence the essential features can be presented using far fewer maps. While the character differs considerably between the two time periods, there is considerable similarity among levels within the main bodies of the troposphere and stratosphere for a given time period. In this section trends characteristic of only the main body of the troposphere are shown. The presentation of stratospheric trends is deferred until section 8 where radiosonde trends are plotted along with those from the MSU. Stratospheric trends involving data prior to the satellite era are not shown because there is little trend during the earlier years as demonstrated in section 7.

The 400-hPa trends shown in the form of vector maps in Fig. 5 have patterns similar to those throughout the free troposphere, consistent with the general notion of small vertical gradients of trend (Fig. 4). Use of this level somewhat enhances the distinction between unadjusted and adjusted trends by virtue of the general increase in magnitude of adjustment with height (Fig. 3). For the longer time period (Fig. 5a) the most obvious disparities between unadjusted and adjusted trends occur over Asia and Africa; in a number of cases trends differ substantially. The tendency, noted earlier, for adjustment to make the trends more positive can be seen, particularly for Soviet stations. By contrast, some areas, especially the Americas, are hardly affected by adjustment. After adjustment the pattern of trends has reduced complexity with the majority of stations reporting positive values.

The satellite-era trends shown in Fig. 5b indicate a much more complex pattern of change, even after adjustment. As was true for the longer time period, the greatest sensitivity to adjustment is also seen over Africa and Asia. Some of the localized features are revisited in section 8 in conjunction with the comparison with MSU data. However, assuming that our treatment of data problems via adjustment is reasonable, then instrumental changes cannot be used to explain the complex structure of this trend map. This additional complexity may simply reflect the greater effects of sampling variability in a shorter sample.

##### *b. Data quality*

The sensitivity of trends and other measures to our data modifications can be used as an indicator of un-



certainty due to historical changes in instruments and measurement practices. To the extent that our adjustments represent enhancement of the data, as shown in section 8, sensitivity to adjustment can also be interpreted as a measure of data quality, where lower sensitivity indicates higher quality. Through the examination of many time series and trend maps it has been found that there exist strong regional contrasts in sensitivity. To summarize these findings, sensitivity statistics have been computed for 10 different regions and are shown in Table 5. The assignment of stations to regions is indicated by the colors in Fig. 6; an effort was made to group like stations in terms of country of control and overall data quality. The values presented in Table 5 are data quality expressed as a relative rank, among the regions, based on data from all levels. One quality measure is based on the sensitivity of regional trends, computed separately for each layer, to LIBCON modification. Other measures were computed based on the fraction of months for which data were deleted (in the DEL scenario) or the number of LIBCON change-points assigned, expressed as a fraction of available months of data. Each measure was computed separately for the 1959–97 and 1979–97 time periods.

In Table 5 the regions have been ordered from best to worst according to the average rank of the six statistics. With only a few exceptions, there is a great deal of consistency among the different measures and time periods. Noteworthy exceptions are the improvement with time in data quality over southern Asia and the degradation in the tropical Pacific. The latter is influenced by the transition from VIZ sondes to Vaisala sondes for some stations during the 1990s (Stendel et al. 2000). The highest quality data is found in the Americas and Antarctica while the lowest quality is found in Africa and the former Soviet Union. Some of the regionality in data quality is evident in Fig. 5, as well as other maps shown later.

One might expect that a trend-based measure, which emphasizes the longest timescales, would exhibit more sensitivity to data discontinuities than would measures that are more strongly influenced by interannual variability. However, for the most severely degraded data even interannual timescales are markedly affected (see, e.g., Fig. 5 of Part I). To this end, the correlation coefficient between UNADJ and LIBCON time series has been computed separately by level, averaged over all levels for a given station, and then squared (Fig. 6). For the three regions of highest data quality as indicated by Table 5, the squared correlation measure is near 100%. In the tropical Pacific and Australian regions the values are ~85%–90%. The Indian stations have values ~75% and their tropospheric warming trends during the satellite era (Fig. 5b) are larger than anywhere else; this issue is revisited later. For the poorest quality stations, in western equatorial Africa, the squared correlation is only ~50%, indicating quite severe problems. However, since the correlations plotted are averaged over all lev-

els, the values for the worst levels are even smaller. Finally, as indicated by the open circles in Fig. 6, lack of data during the presatellite era is a serious problem for the Tropics and especially the Southern Hemisphere, further exacerbating the poorer spatial coverage in those regions.

## 5. Lower-tropospheric lapse rate

Motivated by observations showing that while considerable warming has occurred at the surface during the last two decades, much less, if any has occurred in the free troposphere, recent studies have examined historical variations in lower-tropospheric temperature lapse rate quantities in the Tropics (Gaffen et al. 2000; Brown et al. 2000). Both studies confirmed the greater surface warming and found other low-frequency variations. This subject is reexamined here to determine the effects of radiosonde record inhomogeneities on the earlier conclusions. Also, analyses examine regions outside of the Tropics, horizontal variations in lapse rate trends, and the nature of the temporal behavior. As an approximation to the lower-tropospheric lapse rate we use the difference in temperature, surface minus 700 hPa, following Brown et al. (2000). Our motivation is that in order to calculate the true lapse rate we would need to utilize geopotential height data, for which we have not made continuity adjustments. Although this approximation adds some uncertainty, it seems reasonable given the following: (i) the qualitative agreement in results between Brown et al. (2000), utilizing the approximation, and Gaffen et al. (2000), utilizing actual lapse rate; and (ii) calculations by Gaffen et al. (2000), which suggest that trends in the approximate quantity are dominated by trends in the actual lapse rate rather than changes in layer thickness.

To examine the variation of lapse rate trends by latitude zone such statistics have been computed in several different ways, three of which are given in Table 6. The first method uses the difference in median trends at each of the two levels, equivalent to taking the difference between the values displayed in Fig. 4. There is no requirement that the station locations used are the same at both levels; these may differ due to missing data. This procedure is analogous to past comparisons of in situ surface data with that of the free atmosphere derived from satellite data. The second approach is more reasonable in that trends are computed from station time series of the monthly lapse rate, insuring that both surface and 700 hPa have valid temperature values in the same month; the reported value is then the median of all station trends within a zone. The third approach is the same as the second except that the station network is reduced by using only those for which enough data is available to compute trends for both 1959–97 and 1979–97.

The lapse rate trends in Table 6 show a considerable range for some of the subsets. For the SH there is ex-

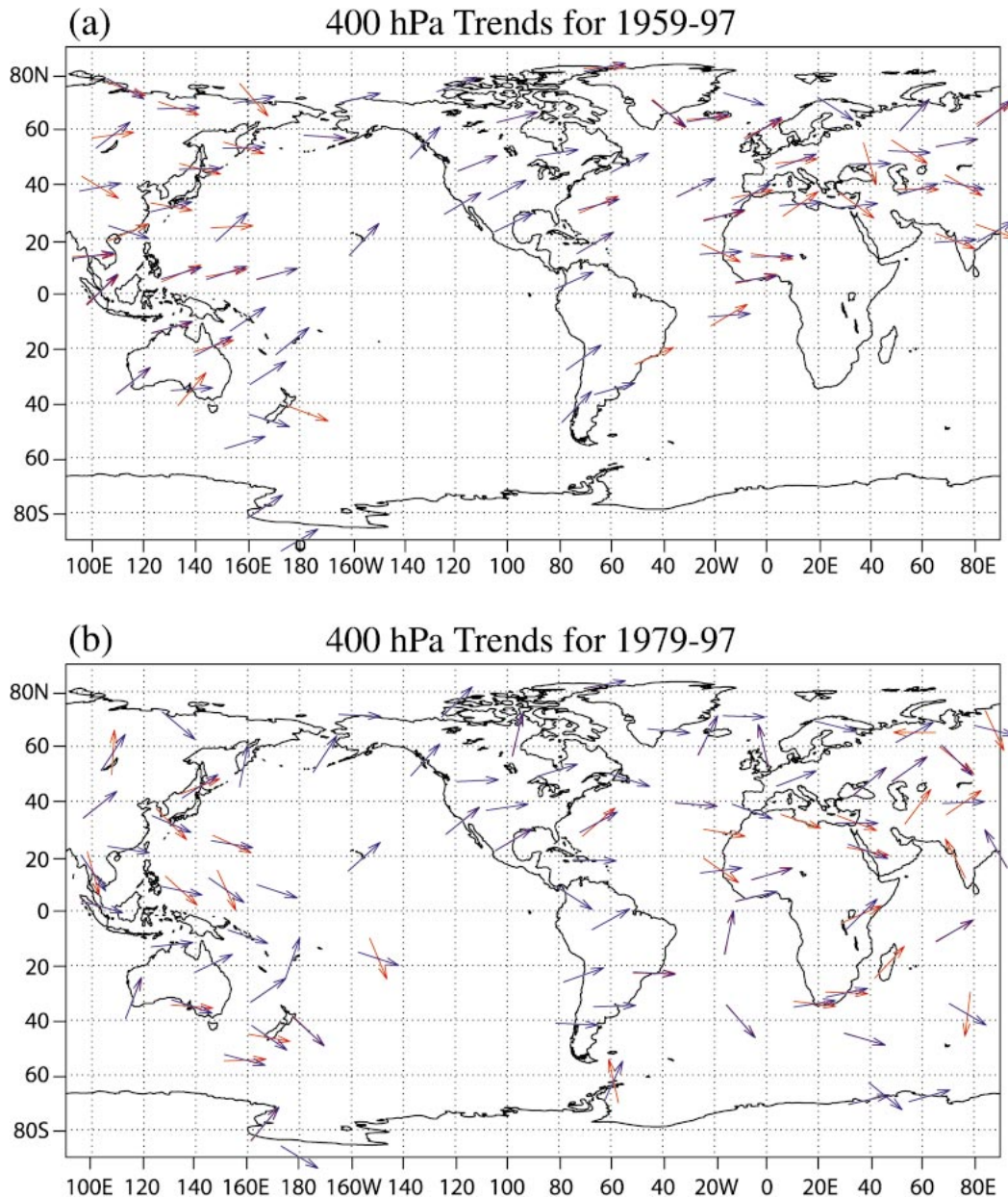


FIG. 5. Temperature trends at 400 hPa for (a) 1959–97 and (b) 1979–97, where red vectors correspond to UNADJ and blue to LIBCON data. With a vector pointing toward the right as a reference of a zero trend, counterclockwise rotation to an upward orientation denotes a positive trend while clockwise rotation to a downward orientation denotes a negative trend. The magnitude of the trend is proportional to the angle of rotation such that a vector pointing directly up or down indicates a trend magnitude of  $0.8 \text{ K decade}^{-1}$ . The presence of only a blue vector indicates that UNADJ and LIBCON trends are identical. The presence of only a red vector indicates that data deletion resulted in too little data for the computation of a LIBCON trend.

treme sensitivity, which stems in part from the small number of stations (less than 10 for 1959–97; Table 2). However, for both the SH and TRPC the uncertainty stems in largest part from the sensitivity of surface temperatures to adjustment (not shown) and is consistent with larger adjustment magnitudes there indicated by the vertical profiles given in Fig. 3. For comparison, tropical lapse rate trends calculated by Gaffen et al.

(2000) are  $\sim 0.08$  for 1979–97 and  $\sim 0.03$  for 1960–97, while Brown et al. (2000) calculated  $\sim 0.20$  for 1979–98. In conjunction with the values in Table 6, these prior estimates underscore the considerable degree of uncertainty. Nevertheless, for the satellite era, given the qualitative similarity in lapse rate trend between Brown et al. (2000) and Gaffen et al. (2000), and the consistency of relationships involving other data products, such as

TABLE 5. Data quality (i.e., sensitivity to data modification) by region expressed as a rank relative to other regions (1–10, where 1 = highest quality/least sensitivity). Ranks to the left of the slash are based on the period 1959–97 and to the right are for 1979–97. Assignment of stations to regions is as shown in Fig. 6, with number of stations per region indicated by *N*. For “Trend,” the quantity of interest is the absolute value of the *difference* in the median trends between the UNADJ and LIBCON scenarios, with the medians taken from a collection of trends consisting of all levels in a vertical layer (layers as per Table 3) and all stations in the region. Ranks computed separately for each layer were averaged vertically and then these averages were ranked to produce the reported value. The use of ranks for each layer protects against a disproportionate influence of one or more layers. For DEL, the fraction of months that were deleted under the DEL scenario is the quantity of interest. For LIBCON, the number of changepoints expressed as a fraction of available months of data under the LIBCON scenario is the quantity of interest. These quantities were summed over all levels and stations in a region and then ranked. The column “Avg” consists of the average of the six ranks shown to its left.

Region	<i>N</i>	Trend	DEL	LIBCON	Avg
South American	8	2/1	3/2	2/2	2.0
Antarctica	6	4/2	1/1	3/1	2.0
North America	13	1/3	2/6	1/3	2.7
Tropical Pacific	7	3/9	5/5	5/5	5.3
Europe	8	6/5	7/7	4/4	5.5
South Asia	10	9/4	4/3	9/6	5.8
Australia	8	5/6	6/4	7/9	6.2
South Africa	8	7/8	10/8	6/7	7.7
Former Soviet Union	11	8/7	9/10	8/8	8.3
North Africa	8	10/10	8/9	10/10	9.5

TABLE 6. Median lapse rate (surface – 700 hPa) trends ( $K \text{ decade}^{-1} \times 100$ ) by latitude zone (NH, TRPC, or SH) and scenario (UNADJ or LIBCON) for (left) 1959–97 and (right) 1979–97. Each entry is based on the median of the trends computed for each station in the zone. Trends have been estimated using three different procedures: 1) Separate latitude zone trends were computed for the surface and for 700 hPa, and then these were differenced. 2) The trend at each station was computed from a monthly time series of lapse rate and then latitudinal medians were taken of these. 3) The procedure is the same as for 2) except that fewer stations were used; only stations for which sufficient data were available for the calculation of trends during both the 1959–97 and 1979–97 time periods were used.

Zone	Scenario	1959–97			1979–97		
		1	2	3	1	2	3
NH	UNADJ	6	6	6	25	17	15
NH	LIBCON	3	0	0	18	16	16
TRPC	UNADJ	0	-6	-8	17	17	10
TRPC	LIBCON	-8	-5	-7	10	0	0
SH	UNADJ	-41	1	-2	-11	-4	-23
SH	LIBCON	-23	4	-6	15	14	5
GLOBAL	UNADJ	1	2	0	13	16	11
GLOBAL	LIBCON	-5	0	-4	10	14	7

sea surface and nighttime marine air temperatures (Christy et al. 2001), the reality of some component of trend in vertical temperature differences cannot be discounted.

For further insight, the map of lapse rate trends for 1979–97 (Fig. 7a) based on the monthly lapse rate as determined using the second method is examined. At some locations, particularly in the SH and TRPC the sensitivity to data modification is very large. Further-

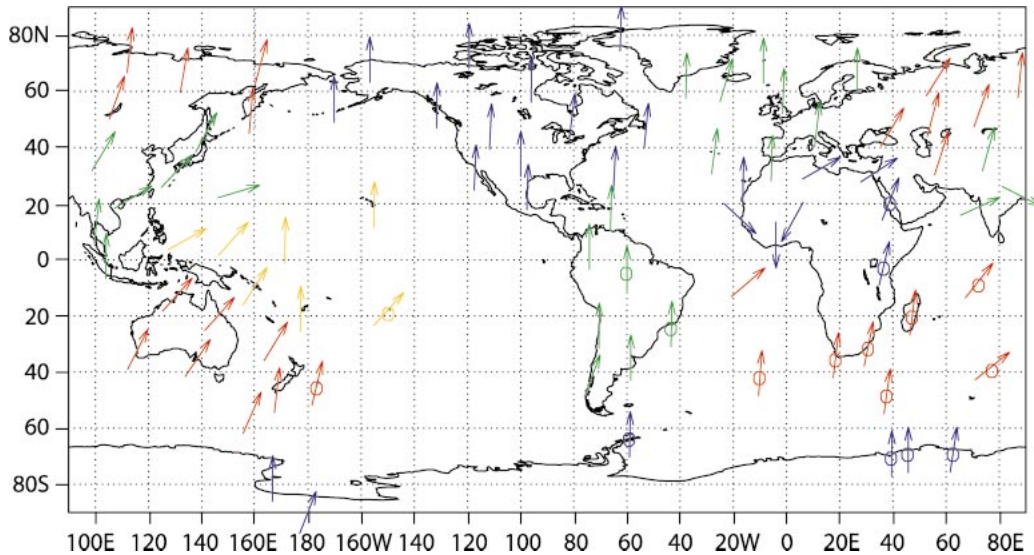


FIG. 6. Square of correlation coefficient, expressed as a percentage, between UNADJ and LIBCON time series averaged over all levels. Spearman correlations (Lanzante 1996) computed separately for each level were averaged, using the Fischer *z* transformation (Zar 1996), and then squared, to form a single value for each station. Plotting is such that a vector pointing up indicates a value of 100% with a vector pointing down corresponding to 50%; vectors rotate clockwise such that magnitude is proportional to the angle of rotation. Data are based on the time period 1959–97 with the criterion that for a given level there must be valid data available for at least half of the months, separately for each third of the time period; otherwise the time period 1979–97 is used instead and is indicated by an open circle. The colors serve only to delineate the different regions as indicated in Table 5.



more, the spatial pattern of trends is complex, so it is easy to imagine that the average over some region might be sensitive to the spatial sampling. In the extratropics of North America and Eurasia there appear to be wave train-like structures and in the Arctic there are some strong positive trends. To a considerable extent the pattern of lapse rate trend and surface trend (not shown) resemble one another, and in particular it can be seen that the sensitivity to adjustment stems largely from surface sensitivity; however, the reader should keep in mind that we have less confidence in surface than upper-air adjustments. The lapse rate trend map for 1959–97 (not shown) has a much more spatially uniform pattern with a very weak hint of the wave trains noted but without the strong positive trends in the Arctic; again strongest sensitivities to adjustment are in the Southern Hemisphere and the Tropics.

We conclude this section with the examination of lapse rate time series computed by latitude zone (Fig. 7b). As indicated in the figure caption, stations are limited to those with sufficient data over the 1959–97 time period; results are similar without this restriction. The curve for the TRPC is qualitatively similar to that of Gaffen et al. (2000) and to a somewhat lesser extent that of Brown et al. (2000). The most striking feature is the downward discontinuity ~1976–77. Variability related to ENSO is also prominent; lag correlations suggest that the SOI leads the TRPC lower-tropospheric lapse rate series by ~6 months, consistent with that notion that El Niño warms the tropical troposphere making the lapse rate more negative (stable). However, the ENSO signature appears to be considerably weaker after the 1976–77 transition. It is interesting that the aspect that has received the most attention is the weak upward trend over the 1980s and 1990s, which, according to Table 6, is not robust to adjustment and is dwarfed in magnitude by the 1976–77 discontinuity. For the NH there is prominent variability on the timescale of ~10 years. When viewed from this context the upward trend for 1979–97 captured by Table 6 is less impressive, largely an artifact of the time period chosen. For the SH the veracity of the considerable drop in late 1960s must be tempered by the large sensitivity to adjustment. In summary, the complexity of both the spatial and temporal variability of the lower-tropospheric lapse rate, as well as its sensitivity to instrument-related changes and the method of estimation, argues for caution at this time in ascribing physical significance to any apparent secular changes, and for further study of this matter.

## 6. Seasonality of trends

The seasonality of trends may have some relevance to the detection and attribution of climate change in conjunction with the strong seasonality of ozone depletion in the stratosphere as well as with the response of the troposphere to increases in greenhouse gases (Stott et al. 2001). A simplified analysis of the seasonality of

temperature trends is presented using standard 3-month seasons and the latitude zones defined previously for the two time periods. The analysis is confined to three vertical regions: the surface, 300–500 hPa, and 50–100 hPa, the latter two being representative of the main body of the troposphere and the lower stratosphere, respectively. The Kruskal–Wallis one-way analysis of variance test based on ranks (Siegel and Castellan 1988) was used to determine whether there are any differences in median trend among the four seasons. Median trends for each season were computed from the collection of values consisting of trends at all stations for all levels in a particular vertical region, latitude zone, and time period.

The credibility of any claims of seasonal differences will be bolstered by both a high level of statistical significance as well as robustness to adjustment. For the period 1959–97, as well as for the surface for 1979–97, the test results fail to meet these criteria. However, for 1979–97 the tests suggest seasonality in trend both in the lower stratosphere and upper troposphere, as shown in Table 7. For the NH stratosphere there is less cooling during September–October–November (SON; hereafter all 3-month seasons abbreviated similarly) than the other seasons. There is more sensitivity to adjustment in the TRPC, and the most robust results indicate more cooling during JJA and less during MAM. Some sensitivity is also seen in the SH, with greatest cooling during DJF and less during JJA and MAM. Seasonality in global trends may not have much meaning due to the considerable differences between latitude zone.

For the upper troposphere, the most striking feature is exemplified by the NH, which shows greater warming during SON at a borderline level of significance. Although they lack significance and robustness, the TRPC, with greater warming, and the SH, with less cooling, share this aspect of seasonality during SON so that the global result is both highly significant and robust. Globally too, the most cooling is during MAM, with contributions from both the TRPC and SH, but not the NH. For the enhanced SON warming in the NH, both the spatial structure and temporal behavior deserve comment. The trends in the troposphere (~850–400 hPa) have a horizontal structure reminiscent of the North Atlantic Oscillation (NAO), with enhanced positive trends over and near Greenland and much weaker positive, or negative trends outside of this region in the extratropics of the Northern Hemisphere. Another manifestation of this free-tropospheric behavior is a corresponding NAO-like SON seasonality in lapse rate trends, although this seasonality is only marginally significant for the NH during 1979–97.

To examine the time history of this phenomenon, a simple index has been constructed by averaging the temperature anomalies for three of our stations in the Greenland–Iceland region, each with a different country of control. The time series for this index are shown in Fig.



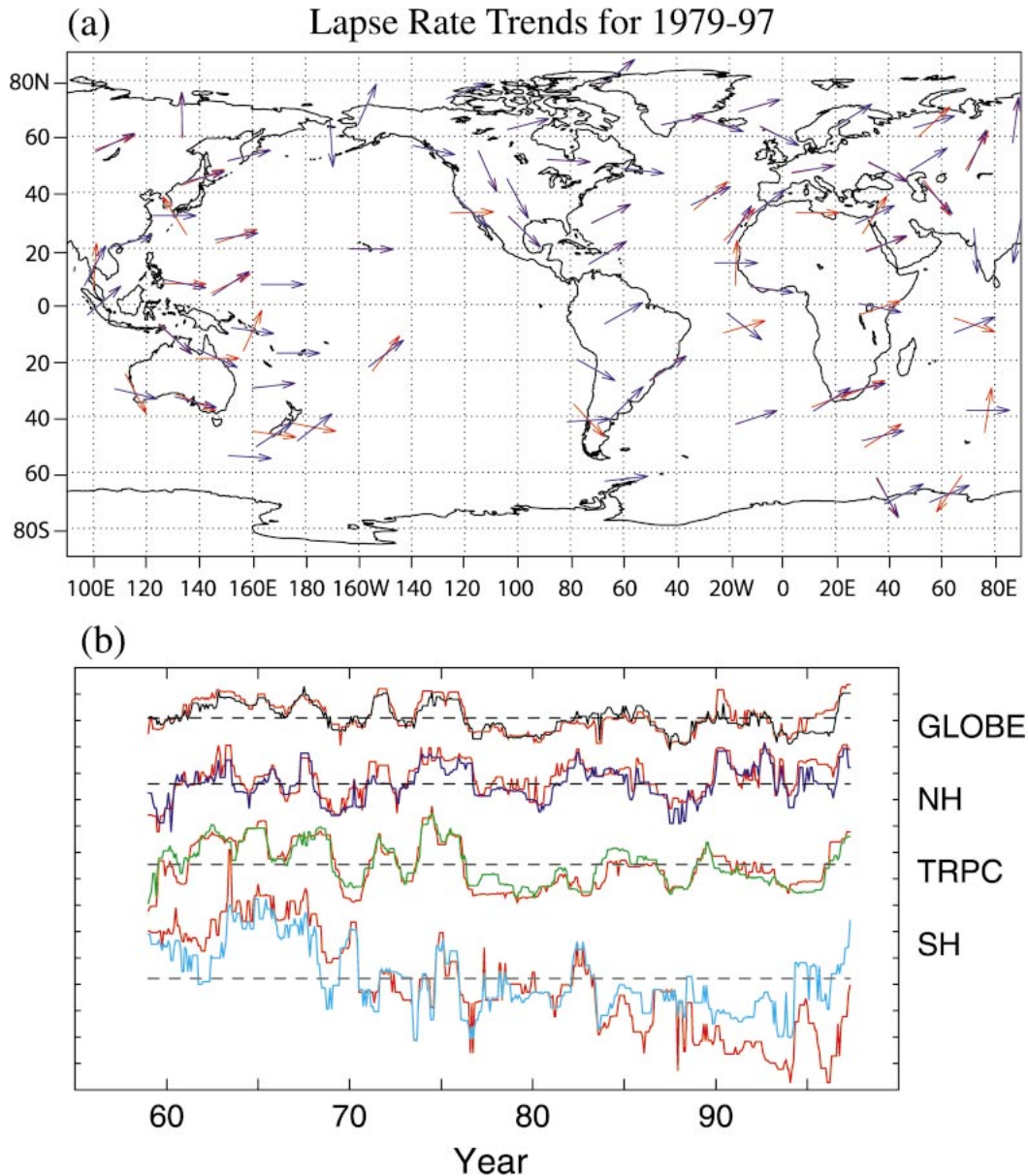


FIG. 7. Trends in lapse rate (surface minus 700 hPa) (a) for 1979–97, as in Fig. 5 except that a vector pointing directly up or down indicates a trend of  $1.0 \text{ K decade}^{-1}$ . (b) Time series of lapse rate by latitude zone based on LIBCON data for the globe (black), NH (blue), TRPC (green), and SH (cyan). Companion red curves are based on UNADJ data. Each monthly value is the median of the values at all of the stations in a latitude zone. Dashed horizontal lines are LIBCON means provided for visual reference. The station set has been restricted to those for which sufficient data were available to calculate a trend for 1959–97, based on the requirement that at least 50% of the months had valid data for each third of the time period. The time series have been smoothed using a 15-point running median. The tick interval on the ordinate is  $0.2 \text{ K}$ .

8, separately for each season. For MAM and JJA the amplitude of interdecadal variability is much less than for DJF and SON. While the latter two have comparable amplitude there does not appear to be any clear correspondence between them. It can be seen that although the strong SON temperature trends appear to have been a fortuitous result of the choice of the 1979–97 time period, they nevertheless are indicative of low-frequen-

cy variability. The similar timing of the enhanced NAO-like warming signature in the troposphere with the reduced NH stratospheric cooling, both occurring during SON is curious as is the apparent global-scale signature of the enhanced tropospheric warming during SON. Note with regard to the former that there have been suggestions of a connection between the Arctic Oscillation and stratospheric dynamics (Shindell et al. 1999),

TABLE 7. Median temperature trends ( $\text{K decade}^{-1} \times 100$ ) by latitude zone (NH, TRPC, or SH), scenario (UNADJ or LIBCON) and season (DJF, MAM, JJA, or SON) for two layers (50–100 or 300–500 hPa) for 1979–97. Each entry is the median of all trends over all stations in the zone and over all levels in the layer. The Kruskal–Wallis one-way analysis of variance (Siegel and Castellan 1988) was used to determine the statistical significance (SIG) in testing whether there are any differences in median trends among the four seasons; significance is rounded to the nearest 1%. As for Table 4, the 1% level is considered significant while the 5% level is borderline significant.

Zone	Scenario	50–100 hPa					300–500 hPa				
		DJF	MAM	JJA	SON	SIG (%)	DJF	MAM	JJA	SON	SIG (%)
NH	UNADJ	–86	–90	–87	–58	2	5	11	0	25	3
NH	LIBCON	–89	–81	–84	–53	0	5	14	2	19	5
TRPC	UNADJ	–114	–88	–124	–128	0	2	–11	4	3	16
TRPC	LIBCON	–81	–50	–103	–85	0	9	–5	2	8	3
SH	UNADJ	–122	–77	–46	–114	0	–15	–12	–7	0	26
SH	LIBCON	–115	–63	–40	–82	0	–7	–19	–8	–1	8
GLOBAL	UNADJ	–103	–88	–94	–89	7	0	–4	1	8	0
GLOBAL	LIBCON	–91	–64	–84	–71	1	3	–5	1	10	0

and with regard to the latter that there may be an association between tropical forcing and the NAO (Hoerling et al. 2001).

### 7. Temporal evolution on large spatial scales

To characterize the time evolution of temperature on large spatial scales, monthly median time series have been calculated for different latitude zones and levels. Such time series for the 400- and 70-hPa levels, which are typical of the troposphere and lower stratosphere, respectively, are shown in Fig. 9. Because of the large horizontal scales covered, the effects of adjustment are generally subtle. The most prominent exception is regarding the NH during the first decade or two when the UNADJ tropospheric data were noticeably warmer, another manifestation of the artificial cooling that was found especially at the Soviet stations (see Part I). From a global perspective the behavior in the troposphere

could be characterized as approximately linear warming from the 1960s to 1990s. However, examination of time series by latitude zone reveals a more complex evolution. For the NH, there was little warming prior to the 1980s, whereas in the SH, the warming was concentrated from the 1960s up to about 1980. In the TRPC, most if not all of the warming occurred in conjunction with the well-known abrupt climate regime shift in the mid-1970s. Variations on ENSO timescales are very prominent in the TRPC and are also conspicuous globally. For the stratosphere, only curves for the global domain are shown since the latitudinal differences are much smaller than in the troposphere. Warming associated with the major volcanoes (Agung in 1963; El Chichon in 1982; Pinatubo in 1991) is prominent and a signature of the quasibiennial oscillation (QBO) can also be seen. It is also evident that little, if any, of the long-term stratospheric cooling occurred prior to ~1980. Thereafter, considerable cooling occurred, and was concentrated during the periods of a couple of years after the last two volcanoes.

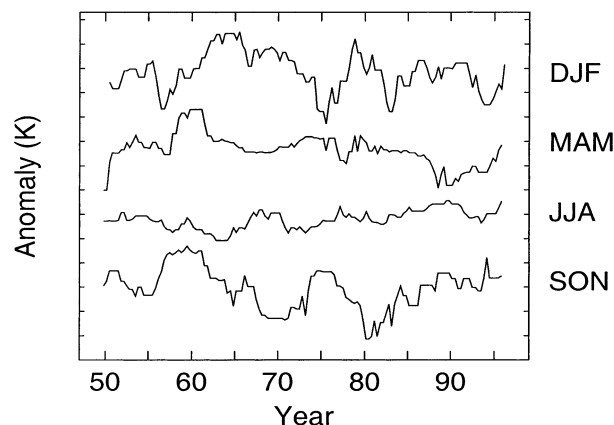


FIG. 8. Time series of 700-hPa temperature anomaly (K) by season (DJF, MAM, JJA, and SON) computed as an average of three stations in the Greenland–Iceland region (Jan Mayen, Keflavik, and Angmagssalik). These time series have been smoothed using a 12-point running median. The tick interval of the ordinate is 0.1 K and the curves have been offset an arbitrary amount for display purposes. Data are UNADJ, but LIBCON time series (not shown) are nearly identical.

### 8. MSU comparisons

To evaluate the credibility of our data modifications, a comparison is made with independent measures of atmospheric temperatures derived from the microwave sounding unit (MSU), version d, which is described by Christy et al. (2000). To facilitate the comparison, static weighting functions (Fig. 10), kindly supplied by John Christy, are employed to convert radiosonde temperatures to values equivalent to those from MSU channels  $2_{LT}$  (lower troposphere), 2 (upper troposphere), and 4 (lower stratosphere); tropospheric functions differ between land and ocean due to different surface emissivities. The use of static weighting functions, as opposed to the more complex approach of radiative transfer modeling requires some assumptions and compromises. For example, stations must be assigned to land or ocean. It is also assumed that there are no trends in emissivity, which might arise due to changes in sea ice, snow, or soil moisture; channel  $2_{LT}$  will be most sensitive to var-

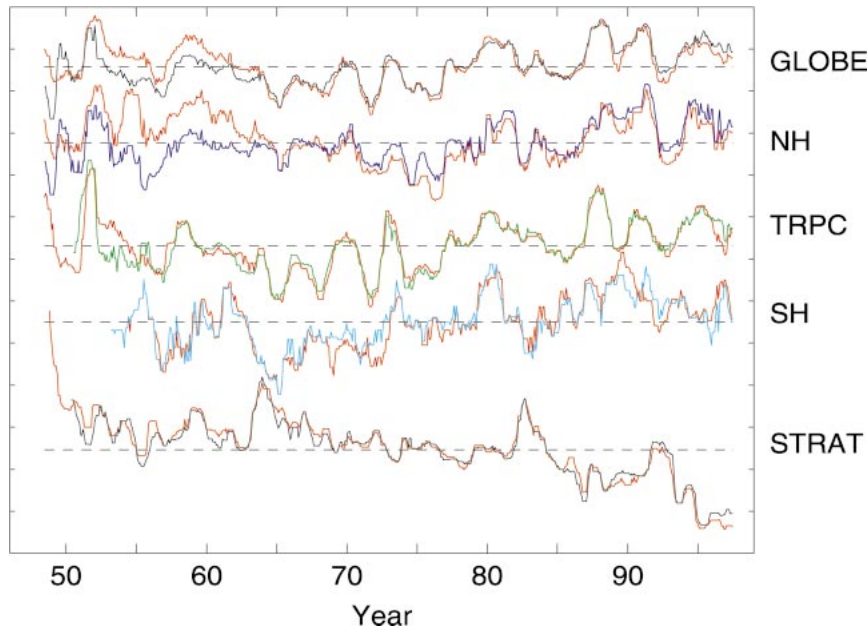


FIG. 9. Time series of temperature derived as the median of all monthly station values in a latitude zone. Top four pairs of curves, representative of the troposphere, are for the 400-hPa level while the bottom pair (STRAT), representative of the global stratosphere, are for the 70-hPa level. Red curves are based on UNADJ data while black (GLOBE), blue (NH), green (TRPC), cyan (SH), and black (STRAT) are based on LIBCON data. The time series have been smoothed using a 12-point running median. Dashed horizontal lines are LIBCON means provided for visual reference. The tick interval on the ordinate corresponds to 0.5 K for tropospheric and 1 K for stratospheric curves. The station set has been restricted to those for which sufficient data were available to calculate a trend for 1959–97, based on the requirement that at least 50% of the months had valid data for each third of the time period. The qualitative characteristics of the time series are not sensitive to the level chosen; for example, related time series computed as weighted vertical averages, for either the troposphere or stratosphere, using as weights the value of the trend for 1959–97, yield similar time series.

iations in emissivity (Shah and Rind 1995). Also, because some stations have little or no data at some levels, particularly the highest ones, a given monthly MSU equivalent value was required to be based on enough levels to account for at least 75% of the total weighting function, otherwise it is considered missing. Furthermore, while the MSU temperature products have undergone a series of homogeneity adjustments in an attempt to account for factors such as changes in satellites and satellite drift, it was implicitly assumed that the latest version of data is largely temporally homogeneous. In so much as a number of important assumptions have been made, comparisons concentrate on whether our radiosonde data modifications result in a closer match with MSU data. Future work will involve more detailed evaluation involving several different satellite and radiosonde products.

Statistics based on the squared correlation ( $r^2$ ) between radiosonde and MSU temperatures are given in Table 8. The columns denote changes between data modification scenarios. As was seen in analogous statistics presented in Table 4, the effects of data deletion (U–D) and adjustment of the more confident change-points (D–C) dominate over the type of adjustment

scheme (L–N) and the confidence level of the change-points (C–L). Over the various treatments, and particularly for our preferred LIBCON approach (U–L), our modifications overwhelmingly result in data that are more highly correlated with MSU. The last two rows in Table 8 give the median changes in a nonparametric statistic akin to the root-mean-square (rms), and indicate that even in the small minority of cases when data modification results in a lower correlation, the magnitude of the degradation is typically much less than the magnitude of the enhancement that occurs in the overwhelming majority of cases.

A related comparison based on other types of metrics is given in Table 9 and further supports the conclusions drawn from Table 8. It is worth noting that the absolute agreement is greatest for channel 2. Poorer agreement for  $2_{LT}$  is not surprising due to concerns regarding surface emissivity. Channel 4 has the largest improvement via data modification but also the largest absolute discrepancy, both of which may be attributable to the effects of the few highest levels, which have large weighting (Fig. 10). While the improvement may be associated with the fact that these levels have the largest adjustment magnitudes (Fig. 3), the absolute discrepancy may be

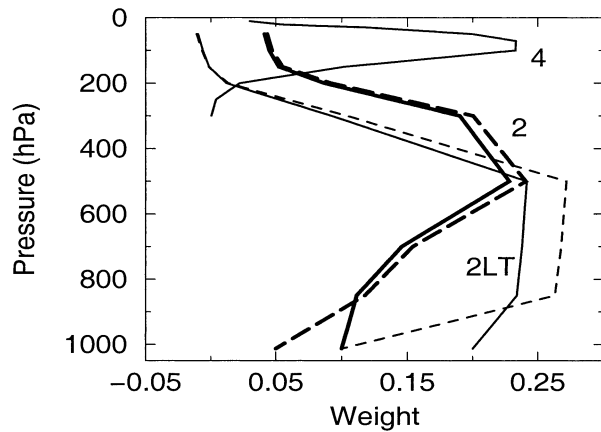


FIG. 10. Weighting function used to compute radiosonde layer-average temperatures equivalent to MSU channels  $2_{LT}$  (bottom, light curves), 2 (middle, dark curves), and 4 (top curve). For channels  $2_{LT}$  and 2 solid curves correspond to land and dashed to ocean.

related to the fact that these levels have the most missing data.

The horizontal distribution of trends for channels 2 and 4 are given in Fig. 11 in the vector format used earlier, red for UNADJ and blue for LIBCON, with the addition of green for MSU. The channel-2 trend pattern (Fig. 11a) bears considerable similarity to that at 400 hPa (Fig. 5b) except for greater cooling in polar regions where the lower tropopause allows for more stratospheric influence on channel 2. Unlike the troposphere, the stratosphere (Fig. 11b) has large trends of the same sign (negative) almost everywhere. Some of the largest discrepancies between UNADJ and either LIBCON or MSU trends correspond to data problems found by Parker et al. (1997) for Australian stations and Stendel et al. (2000) for stations in the western tropical Pacific. However, one curious feature is the station southwest of New Zealand (Macquarie Island), where *both* radiosonde and MSU exhibit a near-zero trend. As seen in Fig. 12, at this station long-term cooling is interrupted by a dramatic warming in the early 1990s. For a discussion the

TABLE 9. Statistics of comparison between MSU and radiosonde temperatures by channel ( $2_{LT}$ , 2, and 4). Comparisons are made by the scenarios defined in Part I, Table 1 (UNADJ, DEL, CON, LIBCON, and NONREF). Each value is the median over the statistics computed separately by station. The statistics are as follows: 1) the square of the Spearman correlation ( $\times 1000$ ) between MSU and radiosonde temperatures ( $r^2$ ), 2) the MAD between MSU and radiosonde temperatures ( $K \times 1000$ ), and 3) the absolute value of the difference between MSU and radiosonde trends ( $K \text{ decade}^{-1} \times 1000$ ).

Statistic	UNADJ	DEL	CON	LIBCON	NONREF
Channel $2_{LT}$					
$r^2$	843	845	845	844	843
MAD	370	350	360	358	355
$ \Delta\text{Trend} $	165	165	128	107	128
Channel 2					
$r^2$	866	876	877	877	877
MAD	232	226	218	216	215
$ \Delta\text{Trend} $	163	150	120	120	119
Channel 4					
$r^2$	848	858	867	868	868
MAD	465	440	420	406	400
$ \Delta\text{Trend} $	342	259	260	246	224

reader is referred to Compagnucci et al. (2001), who first discovered this feature in the MSU data.

As shown in Fig. 11, adjustments made at individual stations generally push the radiosonde data toward the MSU. This is illustrated in the top half of Table 10, which gives the median of the absolute value of the trend difference (between radiosonde and MSU) as a function of latitude zone and channel. With one minor exception, all entries indicate a closer match with MSU after adjustment. While the changes in the NH are minor those in the TRPC are considerable. Nevertheless, a closer examination of Fig. 11 reveals that before and to a lesser extent after adjustment there is a tendency for radiosonde trends to have a negative bias *relative* to the MSU. This is quantified by the bias measure given in the bottom of Table 10, which indicates that although channel  $2_{LT}$  is somewhat different, for channels 2 and 4 adjustment reduces negative bias; bias reduction is

TABLE 8. Statistics on the impact of various scenario changes on a comparison between MSU and radiosonde temperatures. The paired scenario differences (column headings) are as follows: U-D (UNADJ - DEL), D-C (DEL - CON), C-L (CON - LIBCON), L-N (LIBCON - NONREF), U-L (UNADJ - LIBCON). The scenarios are defined in Part I, Table 1. In each table cell, the numbers ordered from left to right correspond to MSU channels  $2_{LT}$ , 2, and 4, respectively. The  $r^2$  has been computed as the square of the Spearman correlation coefficient between MSU and radiosonde temperatures. The first row gives the number of stations for which  $r^2$  increases going from the first to second scenario, while the second row gives the corresponding numbers for which  $r^2$  decreases. The third row is similar except for an increase of greater than 0.01; the fourth row corresponds to a decrease of greater than 0.01. The fifth row gives the median (Med) over all stations of the change in median absolute deviation (MAD) between MSU and radiosonde temperatures ( $K \times 1000$ ), going from the first to second scenario, for stations for which the  $r^2$  with MSU increases; the sixth row is similar except for stations for which the  $r^2$  decreases.

	U-D	D-C	C-L	L-N	U-L
$N_{\text{stns}}, \Delta r^2 > 0$	44/55/52	31/36/27	18/22/17	24/26/18	53/65/57
$N_{\text{stns}}, \Delta r^2 < 0$	38/26/27	25/20/14	23/19/9	29/28/16	31/18/22
$N_{\text{stns}}, \Delta r^2 > 0.01$	9/18/21	7/16/15	6/7/4	2/5/4	19/36/32
$N_{\text{stns}}, \Delta r^2 < 0.01$	7/2/8	2/3/1	2/2/0	1/4/2	7/5/9
Med $\Delta\text{MAD}$ ( $\Delta r^2 > 0.01$ )	-37/-30/-95	-25/-20/-50	-12/-25/-19	-25/-10/-26	-25/-37/-86
Med $\Delta\text{MAD}$ ( $\Delta r^2 < 0.01$ )	10/-5/-12	20/20/-14	8/6/-	20/10/5	12/-23/-1



especially prominent in the tropical stratosphere. A noteworthy exception is the NH stratosphere that has positive relative bias before and to a somewhat greater extent after adjustment. This NH bias appears most prominently at Soviet stations and is probably largely a manifestation of the spurious upward stratospheric drift noted in Part I (section 5b and Fig. 3b). That this bias is found at a majority of Soviet stations *after* data modification stems from the fact that 1) at some locations longitude–latitude are such that the 0000–1200 UTC differences do not represent day versus night extremes of solar radiation, thus preventing us from identifying the drift; and 2) our remedy of deleting the daytime soundings appears inadequate in so much as the drift seems to occur at night as well, although with a reduced magnitude.

The predominantly negative relative bias in the unadjusted data appears, from a cursory examination of station time series in conjunction with station history metadata (Gaffen 1996), to be due in large part to the tendency for transition to the Vaisala RS80 sonde in numerous countries during the last 15 years. A closer examination of Fig. 11 suggests that the presence of this bias after adjustment is due to a combination of underadjustment relative to MSU as well as lack of adjustment in some cases. While a detailed examination is beyond the scope of this paper, some of the more outstanding discrepancies between adjusted data and MSU have been reexamined; except for Soviet stations, these are primarily in the Tropics and subtropics. While the remaining stratospheric discrepancies are almost exclusively excessive cooling, it is noted that excessive tropospheric warming occurs at Indian stations, excessive tropospheric cooling at South African stations, and other more localized problems elsewhere. Factors that may have led to these apparent omissions on our part include incomplete and ambiguous metadata, gaps and sparseness in the temperature time series, and especially the gradual or erratic nature of some of the time-varying biases. Our approach is less well suited to handle the latter since the artificial signal is less distinct from natural variability. It is speculated that the gradual introduction of new instruments or practices, or frequent shifts among several ones may have contributed considerably to our inability to properly adjust the data.

Time series are given in Fig. 12 which illustrate some instances in which our adjustments improve or degrade the correspondence between the radiosonde and MSU trends. For Pechora, Russia, in the troposphere (section 5e and Fig. 6 of Part I) and Adelaide, Australia, in the stratosphere (section 5f and Fig. 7a of Part I) adjustment has dramatically decreased the discrepancy with MSU. For channel  $2_{LT}$  at McMurdo, Antarctica, where we found no inhomogeneities, the trend discrepancy is the result of one of the largest positive radiosonde trends and one of the largest negative MSU trends. Because of its location near the ice margin, emissivity variations not captured by the static weighting function may play

a role. On the other hand, the large drop in MSU temperature  $\sim$ 1985 corresponds to a time of minimal overlap during satellite transitions (Christy et al. 2000). Finally, the tropospheric trends at Calcutta, India (Figs. 5b and 11a), as well as Bombay, India (not shown), are inconsistent with neighboring stations and among the strongest positive values anywhere. The time series at individual levels that contribute most to the channel-2 tropospheric average (Fig. 12) do not suggest an abrupt artificial change; we suspect that undocumented changes were instituted in a more gradual fashion, mimicking natural interannual variability.

Given the regional variations in sensitivity to adjustment illustrated earlier (Table 5), it is of interest to see how well this sensitivity corresponds to the UNADJ radiosonde–MSU discrepancy. For this purpose Table 11 has been constructed, giving relative rankings among regions of the trend discrepancy by channel, and averaged over the channels. Since the ordering of the rows in Table 11 is based on the 1979–97 trends rankings from Table 5, a perfect correspondence would be indicated by average rankings of 1–10 for rows 1–10 in Table 11. It can be seen that there is a reasonable correspondence, with all but two ranks differing by 2 or less. One exception, North Africa, is an artificial result due to the drop out of all but two stations in the stratosphere, which are by selection higher quality, due to lack of upper-level data. Since the calculations for Table 5 were done by level it was possible to determine that this region has the most sensitive stratospheric data. However, there is no easy explanation for the unexpectedly large discrepancies over South America. In addition, note the large channel- $2_{LT}$  discrepancy in Antarctica that is possibly due to emissivity variations not captured by the static weighting functions or perhaps problems with the MSU such as discussed in regard to McMurdo.

In conclusion, much of the unadjusted radiosonde–MSU discrepancy appears to arise from radiosonde instrumental or operational changes. The postadjustment discrepancy is probably a combination of inadequacies in our data modification procedures, use of a static weighting function to compute radiosonde temperatures commensurate with the MSU, and remaining inhomogeneities in the MSU record. With regard to the latter, Hurrell et al. (2000) note the considerable sensitivity of MSU adjustment through a comparison of versions c and d. At this time we are unable to quantify the relative contributions of these factors to the remaining discrepancy. Of most importance, however, is the fact that our modifications make the radiosonde data more like MSU during 1979–97, which yields more confidence in the use of the adjusted data for the presatellite era. In addition, the correspondence between MSU–radiosonde discrepancies and radiosonde sensitivity to adjustment (UNADJ versus LIBCON) suggests that the latter may serve as a proxy measure of data quality in the presatellite era as well.

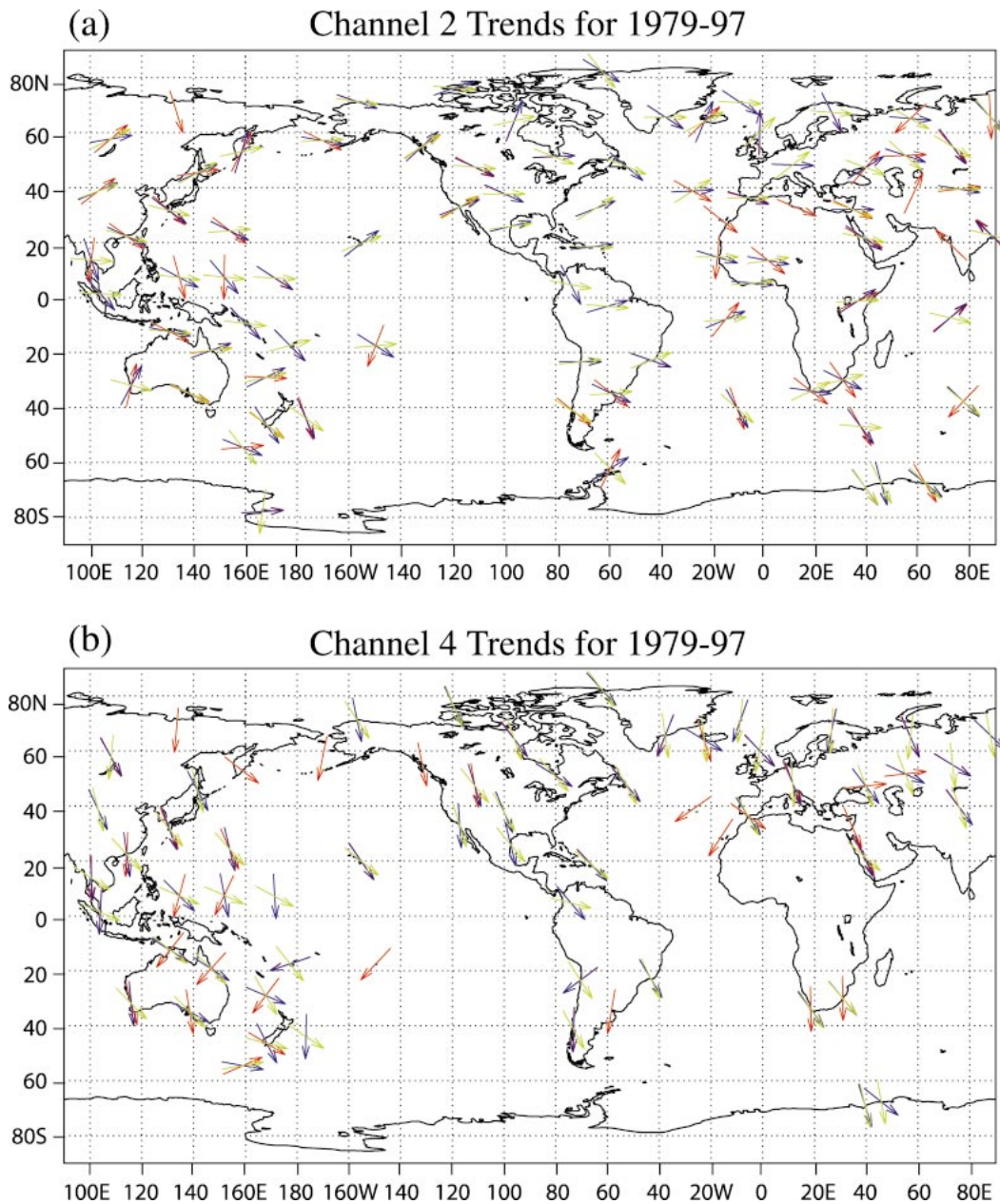


FIG. 11. Temperature trends over 1979–97 for (a) channel 2 and (b) channel 4, where red vectors correspond to UNADJ and blue to LIBCON radiosonde data; green are for MSU, with temporal sampling as for LIBCON. Vector orientation is as in Fig. 5 except that a vector pointing directly up or down indicates a trend of  $0.5 \text{ K decade}^{-1}$  for (a) and  $1.0 \text{ K decade}^{-1}$  for (b). The presence of only a red vector indicates that data deletion resulted in too little data for the computation of a LIBCON trend; since MSU trends use the same temporal sampling as LIBCON, these are absent as well.

## 9. Summary and discussion

Long-term trends of radiosonde temperatures have been examined using data from a near-globally distributed network of 87 stations. Of particular interest was the impact of temporal inhomogeneities, resulting from historical changes in instruments and measurement practices, on these time series. An assessment of the impact has been made based on a methodology introduced in

a companion paper (Part I). This methodology constitutes procedures to identify artificial inhomogeneities, especially discontinuities, and then remove their impact through either adjustment of the data or removal of a portion of the record. By comparing results based on several different strategies to accomplish this goal, some robustness to the details of the method was demonstrated. A comparison with an independent set of satellite-

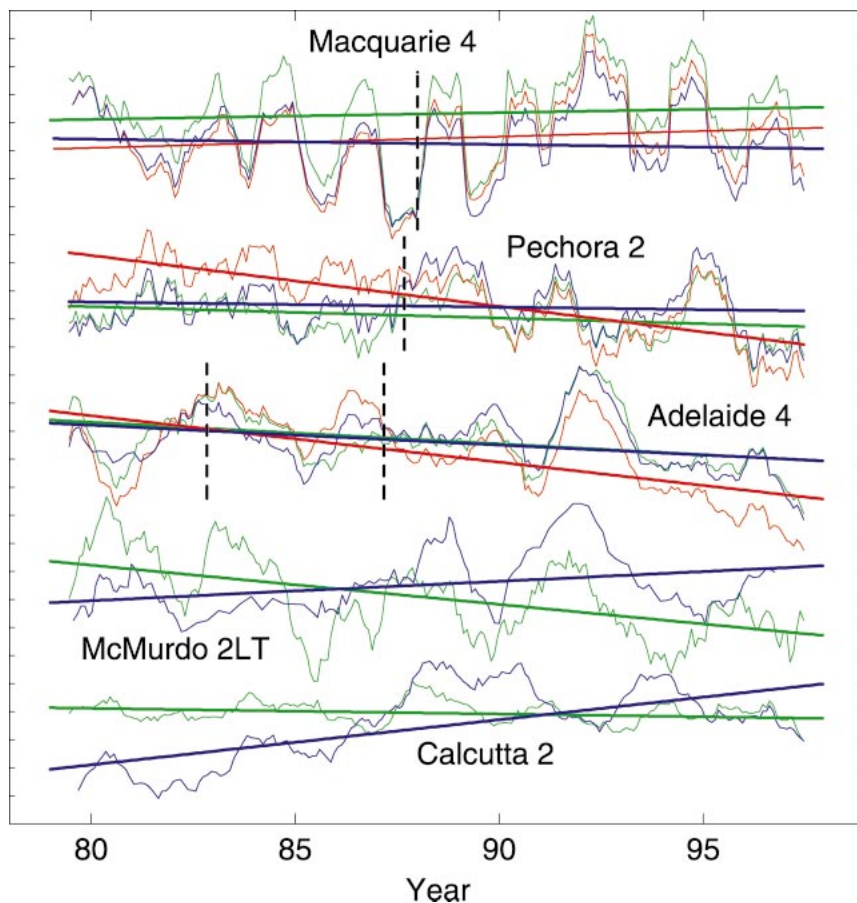


FIG. 12. Time series of temperature anomalies for selected stations and channels (as indicated). Red curves are based on UNADJ and blue on LIBCON radiosonde data; green is for MSU. When UNADJ and LIBCON are coincident, only a blue curve is plotted. Trend lines are given corresponding to each curve using the same color conventions. Dashed vertical lines indicate times of change points for which adjustments were made. Curves are smoothed, based on a 12-point running mean. Tick spacing on the ordinate is 0.5 K.

TABLE 10. Median temperature trend differences (radiosonde minus MSU), taken over all stations in a given latitude zone (NH, TRPC, SH, and GLOBAL) by scenario (UNADJ and LIBCON) and channel ( $2_{LT}$ , 2, and 4). The top half of the table is based on the absolute value of the trend difference while the bottom is based on the signed trend difference. The units are  $\text{K decade}^{-1} \times 100$ .

Zone	Channel $2_{LT}$		Channel 2		Channel 4	
	UNADJ	LIBCON	UNADJ	LIBCON	UNADJ	LIBCON
$ \Delta\text{Trend} $						
NH	15	12	13	12	22	20
TRPC	17	10	22	12	65	33
SH	21	16	22	13	35	37
GLOBAL	17	11	16	12	34	25
$\Delta\text{Trend}$						
NH	-5	-6	-6	-5	9	14
TRPC	3	7	-11	-9	-65	-30
SH	-4	0	-6	-1	-33	-21
GLOBAL	-2	0	-8	-5	-18	-5



TABLE 11. Data quality, assessed via comparison of UNADJ radiosonde temperature with MSU for 1979–97, by region and channel ( $2_{LT}$ , 2, and 4), expressed as a rank relative to other regions (1–10, where 1 = highest quality). Assignment of stations to regions is as shown in Fig. 6, with number of stations per region indicated by  $N$ . For the three channels, the quantity that was ranked is the median, over all stations in the region, of the absolute values of the temperature trend differences between radiosonde and MSU. The weighted averages of the ranks for the three channels were ranked to produce the overall average rank given in the rightmost column. In accordance with the redundancy between channels  $2_{LT}$  and 2 (see Fig. 10), these tropospheric channels were each given half the weighting of channel 4. Note that the ordering of the rows corresponds to the 1979–97 trend ranks from Table 5.

Region	$N$	Channel			Avg
		$2_{LT}$	2	4	
South America	8	6	4	7	7
Antarctica	6	9	1	4	3
North America	13	3	3	1	1
South Asia	10	2	6	2	2
Europe	8	5	7	5	6
Australia	8	1	2	9	5
Former Soviet Union	11	8	8	8	9
South Africa	8	10	9	6	8
Tropical Pacific	7	4	10	10	10
North Africa	8	7	5	3	4

derived temperatures (MSU) placed the data modification methodology in a favorable light. Aggregating over all stations, the MSU data were found to be in better agreement with the modified rather than the original radiosonde data. However, confidence in the data modification procedure is lower at individual stations than in aggregate. Furthermore, unresolved discrepancies remain, especially in the Tropics and in the former Soviet Union. Motivated by the fact that the MSU-derived temperatures used in this study are not necessarily an absolute standard, future work will make use of alternative products that differ in their treatment of time-varying satellite biases. A preliminary version of one such product suggests some sensitivity of trends to satellite data homogenization methods (F. Wentz 2002, personal communication).

Overall, the magnitude of data adjustments increases from the lower troposphere up into the stratosphere. However, surface temperatures, perhaps since they are measured using nonradiosonde equipment, stand out as more problematic, with adjustments more comparable to those for the stratosphere. In the free atmosphere, historical changes in radiosonde instruments introduce a systematic artificial negative trend to temperature time series. We speculate that this is due to improvements over time, which have reduced the solar heating of the instruments. The implications are a reduction in estimated stratospheric cooling and an increase in tropospheric warming of typically  $\sim 10\%$ . It was also found that the severity of homogeneity problems varies considerably by region. Superior data quality was found in North America, while much lower quality was found in Africa and the former Soviet Union. While adjustment

was found to have modest effects on the global scale, locally it can have large effects, reversing the sign of trends and/or significantly altering the magnitude. If not taken into account, these local effects may impact “fingerprint” studies that seek a particular signature of climate change. During the satellite era, stratospheric cooling was found to be excessively strong for Australian and western tropical Pacific stations and excessively weak for Soviet stations. Artificial tropospheric cooling was particularly pronounced for Soviet stations during the 1950s and 1960s. In the worst cases even the interannual variability can be compromised, for example, in equatorial Africa prior to  $\sim 1980$  and for the former Soviet Union  $\sim 1965$ .

The vertical structures of trends, examined separately by latitude zone, were found to display some, but not an overwhelming sensitivity to data adjustment aimed at removing artificial effects. Particularly in the Tropics, adjustment was found to preferentially enhance the upper-tropospheric warming, moving the trend profile upward. Qualitatively, this would seem to move the radiosonde record into closer agreement with GCM estimates of an anthropogenic response (see Figs. 9.8 and 12.8 of Houghton et al. 2001); quantitative confirmation is left for future work. While the unadjusted data suggest a statistically significantly greater stratospheric cooling in the Tropics as compared to the extratropics, after adjustment there was no significant difference.

Apart from any artificial effects, the long-term behavior of tropospheric temperatures shows marked differences between latitude zones. Over the longer 1959–97 period, the extratropics of the Southern Hemisphere (SH) were found to warm more than the extratropics of the Northern Hemisphere (NH), whereas during the satellite era (1979–97) the roles were reversed and the SH actually cooled slightly; warming in the Tropics is intermediate during both time periods. These trend differences are attributable to a very different temporal evolution by latitude zone. While the SH warming occurs primarily during the 1960s to 1970s, the NH warming occurs primarily after 1980; the bulk of the tropical warming seems to occur in association with the previously documented climate regime shift  $\sim 1976$ –77 (Trenberth and Hurrell 1994). However, any conclusions involving the SH must be regarded as tentative due to the paucity of stations and limited areal coverage. The most prominent feature of stratosphere temperatures is the pronounced cooling that occurred almost exclusively after  $\sim 1980$ . Shorter timescales are dominated by warming associated with three major volcanic eruptions as well as the quasibiennial oscillation.

Regarding sensitivity to adjustment, the surface presents particular problems. It is unclear as to whether this is due to the different instrumentation used or to inadequacies of our data modification procedures due to shallow boundary layer effects. Uncertainties in behavior of surface temperature translate into large uncertainties in lower-tropospheric lapse rate. Consider-



able latitudinal differences in both the spatial structure and temporal evolution of lapse rate have been found. During the satellite era, in the NH the pattern of trends shows a complex wave train-like structure in the mid-latitudes as well as some large positive trends associated with surface warming in the Arctic. Particularly in the Tropics and SH the local sensitivity to data adjustment is sometimes very large. The NH lapse rate has prominent decadal timescale variations. In the Tropics, lapse rate variations lag those of the Southern Oscillation by ~6 months, such that El Niño is associated with more static stability and La Niña with less; in addition, there is an abrupt increase in static stability in the Tropics associated with the climate regime shift ~1976–77. The amplitude of these features in the Tropics dwarfs the previously studied upward trend in lapse rate. The complexity of the spatial and temporal variations in lapse rate, the range of values reported for the trends in tropical lapse rate found in prior studies, and the sensitivity found here based on different treatments of the data is cause for concern. Since these issues have not been resolved here, caution is urged in ascribing physical significance to any apparent changes in lapse rate; further study is advised.

In conclusion, it has been found that time-varying instrumental biases are not large enough to alter the basic pattern of stratospheric cooling and tropospheric warming as viewed from a global perspective. However, these biases may alter some of the details of the vertical, horizontal, and temporal structure, with possible implications for detection and attribution of climate change. These findings motivate future work to compare the output from climate models with our observed data; the use of both unadjusted and adjusted data may help bracket uncertainties in the degree of correspondence. Such work might involve tropospheric and/or stratospheric temperatures as well as derived quantities such as the lower-tropospheric lapse rate. At least two studies of this type are under way, one by the lead author of this paper, utilizing Geophysical Fluid Dynamics Laboratory (GFDL) GCMs, and another led by Peter Thorne, involving GCMs from the Hadley Centre. Through such cross comparisons a better understanding of the operation of the climate system as well as the strengths and deficiencies of observed data and complex climate models may be gained.

*Acknowledgments.* The radiosonde data were kindly supplied by Mike Changery and Amy Holbrooks of the National Climatic Data Center under the auspices of the CARDS project. The satellite (MSU) data and vertical weighting functions were kindly supplied by John Christy. The NOAA Office of Global Programs, Climate Change Data and Detection program provided partial support for this project. We acknowledge the encouragement given by Jerry Mahlman and Bram Oort for this project and the related work that preceded it. We thank Tom Knutson, Brian Soden, Kevin Trenberth,

John Christy, and Jim Angell for comments on an earlier version of this manuscript. The three anonymous reviewers provided very thorough and thoughtful comments which improved this manuscript.

#### REFERENCES

- Brown, S., D. Parker, C. Folland, and I. Macadam, 2000: Decadal variability in the lower-tropospheric lapse rate. *Geophys. Res. Lett.*, **27**, 997–1000.
- Christy, J., R. Spencer, and W. Braswell, 2000: MSU tropospheric temperatures: Dataset construction and radiosonde comparisons. *J. Atmos. Oceanic Technol.*, **17**, 1153–1170.
- , D. Parker, S. Brown, I. Macadam, M. Stendel, and W. Norris, 2001: Differential trends in tropical sea surface and atmospheric temperature since 1979. *Geophys. Res. Lett.*, **28**, 183–186.
- Compagnucci, R., M. Salles, and P. Canziani, 2001: The spatial and temporal behavior of the lower stratospheric temperature over the Southern Hemisphere: The MSU view. Part I: Data, methodology, and temporal behavior. *Int. J. Climatol.*, **21**, 419–437.
- Free, M., and Coauthors, 2002: Creating climate reference datasets: CARDS workshop on adjusting radiosonde temperature data for climate monitoring. *Bull. Amer. Meteor. Soc.*, **83**, 891–899.
- Gaffen, D., 1996: A digitized metadata set of global upper-air station histories. NOAA Tech. Memo. ERL ARL-211, 38 pp.
- , B. Santer, J. Boyle, J. Christy, N. Graham, and R. Ross, 2000: Multi-decadal changes in the vertical temperature structure of the tropical troposphere. *Science*, **287**, 1239–1241.
- Hoerling, M., J. Hurrell, and T. Xu, 2001: Tropical origins for recent North Atlantic climate change. *Science*, **292**, 90–92.
- Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson, Eds., 2001: *Climate Change 2001: The Scientific Basis*. Cambridge University Press, 881 pp.
- Hurrell, J., S. Brown, K. Trenberth, and J. Christy, 2000: Comparison of the tropospheric temperatures from radiosondes and satellites: 1979–98. *Bull. Amer. Meteor. Soc.*, **81**, 2165–2177.
- Lanzante, J., 1996: Resistant, robust and nonparametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.*, **16**, 1197–1226.
- , 1998: Correction to “Resistant, robust and nonparametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data.” *Int. J. Climatol.*, **18**, 235.
- , S. Klein, and D. Seidel, 2003: Temporal homogenization of monthly radiosonde temperature data. Part I: Methodology. *J. Climate*, **16**, 224–240.
- Laurmann, J., and L. Gates, 1977: Statistical considerations in the evaluation of climatic experiments with atmospheric general circulation models. *J. Atmos. Sci.*, **34**, 1187–1199.
- Luers, J., and R. Eskridge, 1998: Use of radiosonde temperature data in climate studies. *J. Climate*, **11**, 1002–1019.
- NRC, 2000: *Reconciling Observations of Global Temperature Change*. Panel on Reconciling Temperature Observations, National Academy Press, 85 pp.
- Parker, D., M. Gordon, D. Cullum, D. Sexton, C. Folland, and N. Rayner, 1997: A new global gridded radiosonde temperature data base and recent temperature trends. *Geophys. Res. Lett.*, **24**, 1499–1502.
- Santer, B., J. Hnilo, T. Wigley, J. Boyle, C. Doutriaux, M. Fiorino, D. Parker, and K. Taylor, 1999: Uncertainties in observationally based estimates of temperature change in the free atmosphere. *J. Geophys. Res.*, **104** (D6), 6305–6333.
- , T. Wigley, J. Boyle, D. Gaffen, J. Hnilo, D. Nychka, D. Parker, and K. Taylor, 2000: Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. *J. Geophys. Res.*, **105** (D6), 7337–7356.
- Shah, K., and D. Rind, 1995: Use of microwave brightness temperatures with a general circulation model. *J. Geophys. Res.*, **100** (D7), 13 841–13 874.

- Shindell, D., R. Miller, G. Schmidt, and L. Pandolfo, 1999: Simulation of recent northern winter climate trends by greenhouse-gas forcing. *Nature*, **399**, 452–455.
- Siegel, S., and N. Castellan, 1988: *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, 399 pp.
- Stendel, M., J. Christy, and L. Bengtsson, 2000: Assessing levels of uncertainty in recent temperature time series. *Climate Dyn.*, **16**, 587–601.
- Stott, P., and S. Tett, 1998: Scale-dependent detection of climate change. *J. Climate*, **11**, 3282–3294.
- , —, G. Jones, M. Allen, W. Ingram, and J. Mitchell, 2001: Attribution of twentieth century temperature change to natural and anthropogenic causes. *Climate Dyn.*, **17**, 1–21.
- Trenberth, K., and J. Hurrell, 1994: Decadal atmosphere–ocean variations in the Pacific. *Climate Dyn.*, **9**, 303–319.
- Zar, J., 1996: *Biostatistical Analysis*. Prentice-Hall, 662 pp.