Sensitivity of Tropospheric and Stratospheric Temperature Trends to Radiosonde Data Quality

DIAN J. GAFFEN,* MICHAEL A. SARGENT,+ R. E. HABERMANN,# AND JOHN R. LANZANTE@

* *NOAA Air Resources Laboratory, Silver Spring, Maryland*
\+ *Department of Mathematics, University of Maryland at College Park, College Park, Maryland*
\# *NOAA National Geophysical Data Center, Boulder, Colorado*
@ *NOAA Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, New Jersey*

(Manuscript received 16 November 1998, in final form 8 July 1999)

ABSTRACT

Radiosonde data have been used, and will likely continue to be used, for the detection of temporal trends in tropospheric and lower-stratospheric temperature. However, the data are primarily operational observations, and it is not clear that they are of sufficient quality for precise monitoring of climate change. This paper explores the sensitivity of upper-air temperature trend estimates to several data quality issues.

Many radiosonde stations do not have even moderately complete records of monthly mean data for the period 1959–95. In a network of 180 stations (the combined Global Climate Observing System Baseline Upper-Air Network and the network developed by J. K. Angell), only 74 stations meet the data availability requirement of at least 85% of nonmissing months of data for tropospheric levels (850–100 hPa). Extending into the lower stratosphere (up to 30 hPa), only 22 stations have data records meeting this requirement for the same period, and the 30-hPa monthly data are generally based on fewer daily observations than at 50 hPa and below. These networks show evidence of statistically significant tropospheric warming, particularly in the Tropics, and stratospheric cooling for the period 1959–95. However, the selection of different station networks can cause network-mean trend values to differ by up to 0.1 K decade^{-1}.

The choice of radiosonde dataset used to estimate trends influences the results. Trends at individual stations and pressure levels differ in two independently produced monthly mean temperature datasets. The differences are generally less than 0.1 K decade^{-1}, but in a few cases they are larger and statistically significant at the 99% confidence level. These cases are due to periods of record when one dataset has a distinct bias with respect to the other.

The statistical method used to estimate linear trends has a small influence on the result. The nonparametric median of pairwise slopes method and the parametric least squares linear regression method tend to yield very similar, but not identical, results with differences generally less than ± 0.03 K decade^{-1} for the period 1959–95. However, in a few instances the differences in stratospheric trends for the period 1970–95 exceed 0.1 K decade^{-1}.

Instrument changes can lead to abrupt changes in the mean, or change-points, in radiosonde temperature data records, which influence trend estimates. Two approaches to removing change-points by adjusting radiosonde temperature data were attempted. One involves purely statistical examination of time series to objectively identify and remove multiple change-points. Methods of this type tend to yield similar results about the existence and timing of the largest change-points, but the magnitude of detected change-points is very sensitive to the particular scheme employed and its implementation. The overwhelming effect of adjusting time series using the purely statistical schemes is to remove the trends, probably because some of the detected change-points are not spurious signals but represent real atmospheric change.

The second approach incorporates station history information to test specific dates of instrument changes as potential change-points, and to adjust time series only if there is agreement in the test results for multiple stations. This approach involved significantly fewer adjustments to the time series, and their effect was to reduce tropospheric warming trends (or enhance tropospheric cooling) during 1959–95 and (in the case of one type of instrument change) enhance stratospheric cooling during 1970–95. The trends based on the adjusted data were often statistically significantly different from the original trends at the 99% confidence level. The intent here was not to correct or improve the existing time series, but to determine the sensitivity of trend estimates to the adjustments. Adjustment for change-points can yield very different time series depending on the scheme used and the manner in which it is implemented, and trend estimates are extremely sensitive to the adjustments. Overall, trends are more sensitive to the treatment of potential change-points than to any of the other radiosonde data quality issues explored.

[+] Current affiliation: The Cybarus Group, Inc., Takoma Park, Maryland.

Corresponding author address: Dr. Dian J. Gaffen, NOAA Air Resources Laboratory, R/ARL, 1315 East-West Highway, Silver Spring, MD 20910.
E-mail: dian.gaffen@noaa.gov

1. Introduction

Contemporary efforts to detect climate change, and to attribute changes to anthropogenic or other causes, are based principally on comparing patterns of observed changes with those predicted by models. As discussed by Santer et al. (1996) and references therein, climate change detection and attribution studies rely heavily on the vertical structure of (often zonally averaged) atmospheric temperature changes. The prominence of upper-air temperature is probably due to three factors. One factor is that long-term tropospheric warming and stratospheric cooling is one of the most confidently predicted changes associated with an enhanced atmospheric greenhouse effect. A second is that this pattern differs significantly from the patterns of temperature change associated with other changes in the radiative forcing of the climate system, such as changes in volcanic aerosol concentrations or solar output. A third is the availability of historical observations of upper-air temperature by the operational radiosonde network that span several decades and that are perceived to be of sufficient quality to yield useful information about patterns of temperature change.

Radiosonde observations are not the only source of upper-air temperature time series, but they offer some potential advantages over others for climate change detection. The microwave sounding units (MSU) on polar-orbiting satellites have provided global temperature data since 1979 with spatial coverage far exceeding that of the radiosonde network, especially in the Tropics, Southern Hemisphere, and oceanic regions. However, temperatures derived from MSU data are representative of fairly thick tropospheric and stratospheric layers (Spencer and Christy 1990). Operational radiosoundings yield temperature values with far greater vertical resolution, at minimum twelve mandatory reporting levels between 850 and 10 hPa, at some locations since the 1940's. Time series of MSU temperature data are relatively short and are potentially influenced by changes in cloud amount and tropospheric humidity (Christy 1995); the process of combining data from different satellite-borne sensors (Hurrell and Trenberth 1998); and the decay in the orbit of the satellites, especially during periods of high solar activity (Wentz and Schabel 1998), and so are not ideal for temperature monitoring (Gaffen 1998).

Temperature fields produced by re-analyses of observational data from a variety of instruments using a consistent numerical model (Kalnay et al. 1996) are another option. However, these do not eliminate, and may complicate, problems related to the temporal homogeneity of the input datasets. As shown by Santer et al. (1999) the interpretation of interannual and decadal temperature variations in reanalysis products requires in-depth understanding of the nature and treatment of the various assimilated observations. Consequently, the use of reanalyses for climate trend detection seems highly problematic.

Radiosonde data are also beset with data continuity problems, and because the network is operated by many different national meteorological services, data from different stations have different historical influences. Previous studies have identified numerous changes in radiosonde instruments and observing practices (Parker 1985; Parker and Cox 1995; Gaffen 1993, 1996; Finger et al. 1995) and have estimated the quantitative effects of some specific changes (Parker 1985; Gaffen 1994; Lanzante 1996; Zhai and Eskridge 1996). Instrument changes can lead to abrupt changes in data biases, or ‘‘change-points,’’ in radiosonde temperature time series, and examples are shown in Gaffen (1994), Parker and Cox (1995), and Lanzante (1996). Change-points are particularly noticeable in stratospheric temperature data, especially in the early years of radiosonde operations, but tropospheric data are also affected (Gaffen 1994).

Until recently, most estimates of upper-air temperature trends based on radiosonde data (e.g., Angell 1988, 1991; Oort and Liu 1993; Labitzke and van Loon 1995; Pawson and Naujokat 1997) have not taken into account these time-dependent biases, except perhaps to note their existence. To account for possible change-points, Miller et al. (1992) used a statistical regression model that included a level-shift term to adjust lower-stratospheric temperature data for a few radiosonde stations in Angell's (1988) network for the period 1970–86. Hansen et al. (1997) have noted the difficulty of assessing global and hemispheric trends in the presence of inhomogeneous data and chose to present trends for only a few stations without significant instrument changes for comparison with model predictions.

Parker et al. (1997) have attempted to adjust radiosonde temperature data for known changes in instrumentation and to estimate the effect of the adjustments on calculated trends. Their adjustments were limited to data from stations in Australia and New Zealand where instrument changes were documented during the period 1979–95. Adjustments were based on comparisons between the radiosonde data and MSU data, with the assumption that the latter time series could serve as a reference. The adjustments (of earlier data relative to 1995 data) ranged from 0 to -3.3 K and reduced the estimated zonal-mean temperature change between the periods 1987–96 and 1965–74 at about 30°S at 30 hPa from -2.5 K to about -1.25 K. These results indicate a potentially large sensitivity of estimated temperature trends to the identification and adjustment of change-points in time series.

This paper explores in greater depth the sensitivity of upper-air temperature trend estimates to radiosonde data quality, with an emphasis on the influence of change-points in radiosonde data. Our main purpose is to provide quantitative estimates of the uncertainty in upper-air temperature trends. Section 2 discusses the datasets and historical metadata used in this study. Sec-

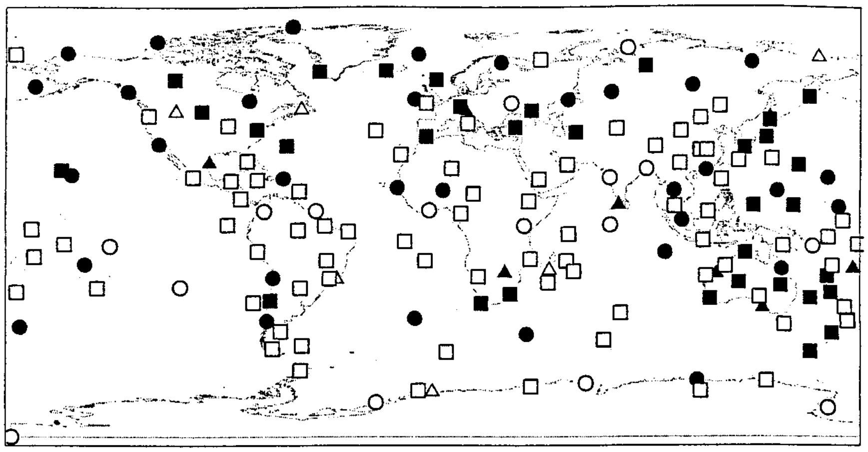

FIG. 1. Map of GCOS and Angell network stations. Circles indicate stations in both networks, squares indicate stations in the GCOS network only, and triangles indicate stations in the Angell network only. Filled symbols indicate stations used in this study; open symbols indicate stations with insufficient data.

tion 3 outlines the statistical methods employed for trend estimation and change-point detection. In section 4 we present linear regression trend estimates based on monthly CLIMAT TEMP data (Parker et al. 1997) as a ‘‘control’’ for sensitivity studies. Section 5 presents trends derived from a different radiosonde data source (Eskridge et al. 1995) to investigate the sensitivity of the trends to choice of dataset. In section 6 we examine the sensitivity of trends to the statistical method of trend estimation by comparing a parametric method (least-squares linear regression) to a nonparametric method (median of pairwise slopes). Section 7 examines the impact of two different statistical methods to remove multiple change-points in time series on trend estimates. In section 8, we attempt to remove a limited set of change-points using a single change-point detection scheme combined with station history information and again assess the impact on trend estimates. Section 9 provides a summary of the sensitivity of temperature trends to each of the factors considered and suggestions for future work.

Although our analysis covers the global domain, trend estimates are restricted to local station trends at specific pressure levels. Santer et al. (1999) recently considered the uncertainty in global and hemispheric temperature trend estimates from a variety of observational data sources and found considerable discrepancies among them. Our intent here is to elucidate the nature of some of the uncertainty in trends from radiosonde data. By examining individual stations we highlight local issues that contribute to uncertainties in trends averaged over large regions.

2. Radiosonde data and station histories

a. Radiosonde station networks defined by Angell and GCOS

To allow detailed investigation of data quality in radiosonde station records, we restrict our analysis to a subset of the complete global network. Rather than specify a new subset, we initially chose to work with two networks already specified by others: the 63-station network used by Angell (1988) to monitor stratospheric and tropospheric temperature, and the Global Climate Observing System (GCOS) Baseline Upper-Air Network (WMO 1996). The Angell network has been the basis of numerous past investigations (Angell and Korshover 1975, 1977, 1978a,b, 1983; Angell 1988, 1991), and the GCOS network is intended to serve future upper-air climate monitoring needs. By focusing on these two networks, we hope to shed light on their utility and the reliability of the results of studies of their data.

The list of stations in the GCOS network has changed slightly during the course of this study, but in 1997 it included 151 stations plus 14 ‘‘standby’’ stations, and incorporated most of Angell’s stations as shown in Fig. 1. [Information about the GCOS network can be found on the WMO home page http://www.wmo.ch/web/gcos and in WMO (1996) and Wallis (1998).] The combined Angell and GCOS networks total 180 stations. However, as discussed below, many of the GGOS stations have incomplete data records that precluded using them in our analysis.

b. Data sources

There are two sources of radiosonde data: individual soundings from stations, and monthly mean values reported by stations (CLIMAT TEMP reports). This study uses both but focuses on CLIMAT TEMP reports (supplemented by data from national meteorological services) as a primary data source because it is the most up-to-date global radiosonde dataset currently available (Parker et al. 1997). The CLIMAT TEMP data for 1959–95 were provided by the Hadley Centre for Climate Prediction and Research, U.K. Meteorological Office.

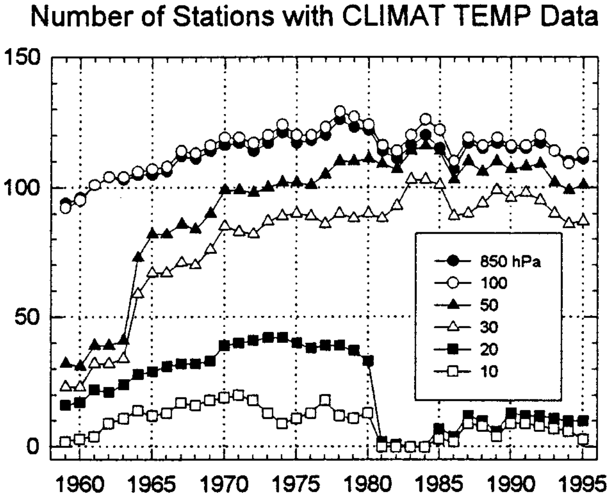

FIG. 2. Number of stations (of a possible total of 180) with available CLIMAT TEMP reports at various pressure levels during at least 1 month of each year 1959–95.

Data for each station include monthly means, and the number of days missing in the month, at the surface and at the following pressure levels: 850, 700, 500, 300, 200, 150, 100, 50, 30, 20, and 10 hPa. We use the ''raw'' station reports, not the gridded and adjusted datasets also described by Parker et al. (1997) and employed by Tett et al. (1996), Santer et al. (1999), and Santer et al. (2000). One limitation of CLIMAT TEMP monthly means is that the separation of 0000 and 1200 UTC data is not possible.

Monthly mean values computed from daily 0000 and 1200 UTC soundings by the Comprehensive Aerological Research Data Set (CARDS) Project (Eskridge et al. 1995) for the period 1959–91 are used for comparison with the CLIMAT TEMP data. The CARDS dataset is a preliminary product that is not generally available and may be affected by inclusion of some erroneous individual observations (R. Eskridge 1996, personal communication). Nevertheless, it is useful as a gauge of data availability, because it is likely the most complete radiosonde dataset ever compiled.

Unexpectedly, we found insufficient data for analysis of trends at many stations in the combined (Angell plus GCOS) network. Data records in the CLIMAT TEMP dataset spanned less than 10 yr for 32 stations of the total 180. As shown in Fig. 2, about 100–125 stations reported data at the 850–100 hPa levels (and levels between these two, not shown) during each year. The number of reports decreases rapidly above 50 hPa. In 1981 there is a sudden decline in the number of reports above 30 hPa, because some data sources were included only for the period through 1980 (D. Parker 1997, personal communication). Clearly, then, the full 180 station network would not be suitable for in-depth analysis for the period 1959–95, and smaller networks had to be developed, particularly for the stratosphere.

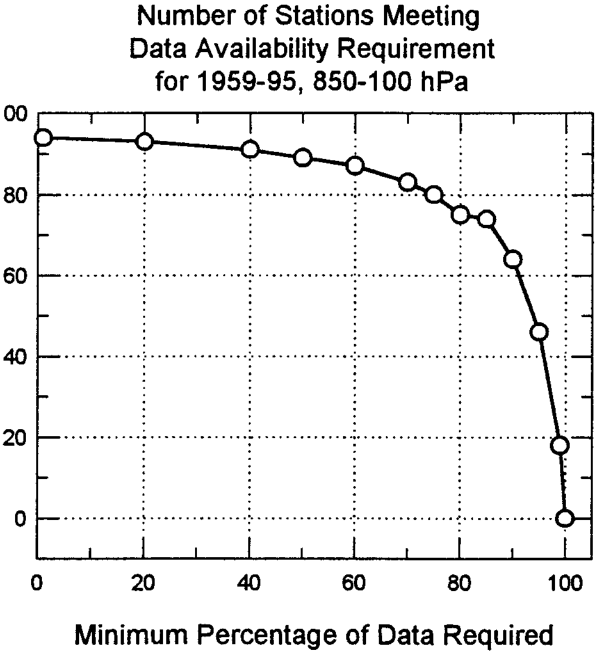

FIG. 3. Number of stations meeting minimum data requirements for the period 1959–95 and for all pressure levels between 850 and 100 hPa. The data requirement is simply nonmissing CLIMAT TEMP reports for a specified percentage of months during the period. On the basis of this analysis, the T network of 74 stations with data available for at least 85% of the months was selected.

c. Networks used in this study

To develop station networks, we examined the data availability for each station in the layer 850–100 hPa and attempted to balance the desire to include as many stations as possible with the desire to include only stations with little or no missing data. Figure 3 shows the number of stations with nonmissing monthly mean temperature data for at least a given percentage, p, of months during 1959–95 at all seven levels within the layer. By reducing p from 99% to 85%, the network size increases from 18 to 74 stations, but further relaxing of p gains only a few additional stations. This analysis yielded a nominal tropospheric (T) network of 74 stations, listed in the appendix. Starting with the combined GCOS and Angell networks of 180 stations, and applying a simple, but not especially rigorous, data availability requirement ($p = 85\%$), results in a network only 17% larger than Angell's 63 stations. Forty-one of the 74 T network stations are in Angell's network.

For the same 1959–95 period, only 22 of the 74 stations in this T network had at least 85% nonmissing monthly mean data for all levels between 850 and 30 hPa, and these define a second network, which we will call the deep (D) network. Recognizing the gradual increase in lower stratospheric data (Fig. 2), and desiring a less sparse network for stratospheric analysis, we found that by shortening the period more stations met the $p = 85\%$ criterion (Fig. 4) at the 150-, 100-, 50-,

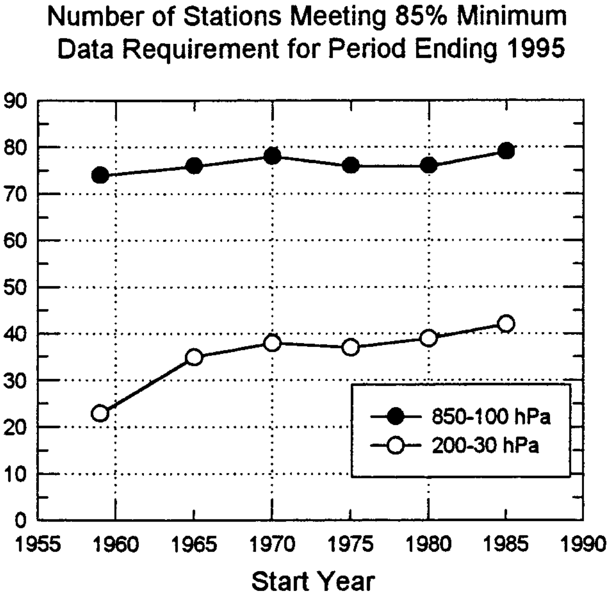

FIG. 4. Number of stations with CLIMAT TEMP data available for at least 85% of the months for various periods, all ending in 1995 but each beginning in different years. On the basis of this analysis, the S network of 38 stations was selected for the period 1970–95.

and 30-hPa levels. (Note that shortening the period does not significantly affect the number of stations meeting the criterion for the 850–100-hPa data.) A resulting third stratospheric (S) network consists of 38 stations with data for 1970–95.

To evaluate the potential influences of dataset choice on trend estimates in section 4 below, we devised a fourth network of twenty stations meeting the $p = 85\%$ criterion for the same levels as the D network, but for the shorter period 1959–91, to accommodate the shorter CARDS data record. This network is identified as C for ‘‘comparison.’’ Table 1 summarizes the basic characteristics of each network and shows how these networks, defined on the basis of data availability, favor the Northern Hemisphere (NH), particularly in the lower stratosphere, despite the fact that 45% of the GCOS stations, and 40% of Angell's stations, are in the Southern Hemisphere (SH).

d. *Representativeness of monthly means*

The $p = 85\%$ criterion applies to CLIMAT TEMP reports of monthly means and reveals nothing about the representativeness of those mean values. A 1967 report includes national practices for CLIMAT TEMP computations (U.S. Dept. of Commerce 1967) for 32 countries. Sixteen countries required at least ten observations in forming the monthly mean, and the remainder had a variety of requirements, ranging from 5 to 25 days of data. We do not know whether these practices have changed in the past three decades or what practices are used in other countries.

A CLIMAT TEMP report is supposed to include the number of missing days of data for the month. However, examination of the dataset prepared by Parker et al. (1997) for the D network defined above showed that in about 28% of the reports the number of missing observations is not known. At levels below 50 hPa about 70% of the monthly mean reports were calculated with fewer than 10 observations missing, while at 30 hPa just over 50% of the reports meet this specification. Thus it appears that the monthly means are more representative up to 50 hPa than for levels above, which may affect interpretation of the vertical profile of lower-stratospheric temperature change.

e. *Station history information*

Despite their inclusion in climate monitoring networks, the radiosonde stations in this (and other) studies have been operated essentially for meteorological forecasting purposes. Changes in instrumentation and observing practices affect the quality and temporal homogeneity of the data and therefore their value in climate studies. While both the Angell and GCOS network designs incorporated the reliability of station reporting (J. Angell and P. Julian 1997, personal communication), neither considered station history information to ensure the temporal homogeneity of the data records. Therefore, there is no a priori reason to assume that the data are homogeneous.

Detailed station history information, compiled and digitized by Gaffen (1996), is available for at least 92% of the stations in each of our four networks (Table 1). These histories include information about radiosonde models (including temperature, pressure, and humidity sensors), balloon types, ground systems, data reduction techniques (including radiation and other corrections applied to the data, methods of computing humidity variables, and the use of manual or computer techniques),

TABLE 1. Basic characteristics of radiosonde station networks used in this study. Each station in each network has at least 85% nonmissing months of data for the periods shown.

Network	No. of pressure levels	Pressure level range (hPa)	Data period	Number of stations NH	SH	Globe	Percent with station histories
Troposphere (T)	7	850–100	1959–95	48	26	74	92
Stratosphere (S)	4	150–30	1970–95	34	4	38	92
Deep (D)	9	850–30	1959–95	19	3	22	95
Comparison (C)	9	850–30	1959–91	17	3	20	95

calibration methods, windfinding equipment, etc. All of the stations in our networks for which histories are available experienced some changes in instruments and methods, and it is probably safe to assume that the few for which we have no information did as well. Although the accuracy and completeness of the histories are not perfect, these ''metadata'' can be used to estimate the dates at which one might expect changes in the error characteristics of the temperature time series, as described in section 3b.

3. Statistical methods

The main statistical features of the data explored in this paper are linear trend estimates and possible ''change-points,'' or abrupt shifts in mean value. This section describes preparation of the time series and the statistical methods used. All time series of monthly mean temperatures were first converted to monthly anomaly time series, by subtraction of long-term mean values for each month of the year from the monthly means.

For both trend estimation and change-point detection, parametric and nonparametric methods were used. As discussed by Wilks (1995) and Lanzante (1996), parametric methods are based on assumptions about the distribution of data, often that it is Gaussian. Nonparametric methods do not require such assumptions. Although monthly temperature anomalies are often assumed to have a Gaussian distribution, the potential existence of change-points introduces a strong possibility that they do not. For all but one test (explained below), there was no need to interpolate missing data values.

a. Methods of estimating trends and their confidence intervals

Least squares linear regression (LSLR) was used for most of the trend estimates in this study. It is a parametric test based on finding the equation of best fit to the paired temperature anomaly and time data by minimizing the sum of the squares of the errors in the residuals. For each trend estimate, we also compute the 99% confidence interval of the trend as a function of the trend estimate and the t-statistic, with $n - 2$ degrees of freedom (dof), where n is the number of months in the time series (Bickel and Doksum 1977). No adjustments are made for possible autocorrelation effects on dof. Because n is large (≥ 312), reducing dof would have a negligible effect on the confidence intervals.

The median of pairwise slopes (MPS) is a nonparametric estimate of trend determined by computing the slopes of lines connecting all possible pairs of points in the time series and taking the median value as the trend estimate (Lanzante 1996). The confidence interval is computed using the Kendall test statistic (Conover 1980). Note that the MPS trend estimate is not necessarily centered within the confidence interval, unlike LSLR. The main benefits of the MPS method are resistance to outliers and less sensitivity to data near the ends of the time series. Its chief disadvantage is a slightly larger sampling error in the case of normally distributed data.

b. Methods of detecting change-points

The problem of detecting change-points, either at known or unknown times, within a geophysical time series with variability on multiple timescales is a complex one. An excellent review of efforts to detect change-points for the purpose of homogenizing meteorological time series is given by Peterson et al. (1998a). They classify the methods as direct, if they rely on metadata or information about known instrument changes, or indirect, if they rely on statistical or graphical manipulations of the time series.

Peterson et al. (1998a) review methods applied to surface observations. As discussed above, very few attempts have been made with upper-air data, which pose complications different from surface data. First, radiosonde stations are much more distant from one another than surface observing sites and similar changes are more likely to be made countrywide, so techniques that employ neighboring stations to create reference time series are less useful. Second, radiosonde instrument packages are expendable, so changes in instrumentation are potentially easier to make, and more frequent, than at surface sites with permanent instrument installations. Third, radiosonde data include information at many levels in the atmosphere, and identification and adjustments of change-points should be consistent in the vertical. Fourth, the observation includes temperature, humidity, and geopotential height data, and identification and adjustments of change-points should be thermodynamically consistent (although this may not be a concern if temperature time series alone are used). Fifth, although numerous intercomparisons of various radiosonde instrument types have been made (e.g., Richner and Phillips 1982; Nash and Schmidlin 1987; Ivanov et al. 1991), no reference instrument is routinely used to ensure that data adjustments are indeed corrections. Sixth, upper-air temperature data contain natural variations that could easily be mistaken for artificial change-points. These include sudden stratospheric warmings (Scherhag 1960), abrupt stratospheric warmings associated with injection of volcanic aerosols (Angell 1996), steplike changes in the climate ''regime'' in the lower troposphere (Trenberth 1990), and tropical fluctuations associated with the quasi-biennial oscillation in the stratosphere and the El Niño–Southern Oscillation in the troposphere. Seventh, one prominent source of inhomogeneity in surface data, station location changes, has a far less obvious impact on upper-air observations than on surface observations.

Despite this compendium of good excuses to abandon

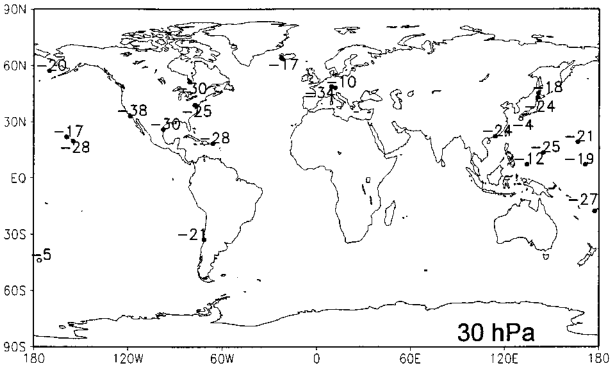

FIG. 5. Least squares linear regression trends (×100 K decade⁻¹) for the D network for 1959–95 based on CLIMAT TEMP data at 30 hPa. Filled circles indicate trends that are statistically significant at the 99% confidence level.

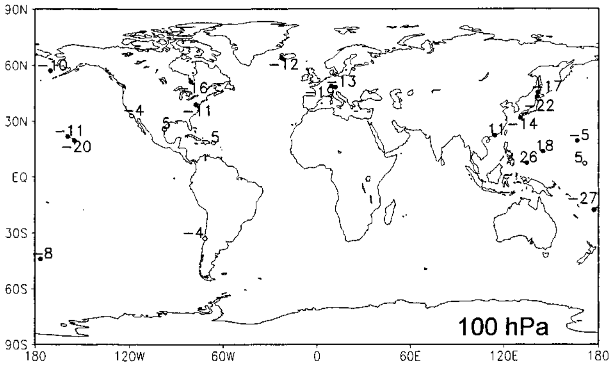

FIG. 6. Same as Fig. 5 but for 100 hPa.

all hope of detecting and adjusting for change-points, we present two indirect methods (using statistical approaches alone), and one hybrid method (using statistical methods combined with station histories) for adjusting radiosonde temperature time series. We stress that our intent is not to correct or improve the existing time series with these adjustments, but to determine the sensitivity of trend estimates to the adjustments and the resulting uncertainty in the trends.

The two indirect methods are objective, statistical tests for the detection of multiple change-points in time series. One, developed by Habermann (1987) for application to seismological time series, is a parametric method involving iterative testing of the difference in segment means to partition the times series into a sequence of shorter segments. Each segment is the longest one such that every point within it shows no change at a prescribed, but adjustable, confidence level. Segments cannot be shorter than an adjustable, arbitrary minimum length. For this method, there must be no missing data in the series. Therefore, for this test only, we linearly interpolated the anomaly time series to fill data gaps.

The second method, developed by Lanzante (1996), is a nonparametric method in which the test statistic is based on a cumulative sum of ranks. As the test is applied iteratively, the time series is adjusted, and the test is reapplied to the adjusted series. Before identifying a change-point, the time series is tested for possible trends to avoid mistaking a trend for a change-point. Like the Habermann method, this scheme has adjustable parameters to control the significance of detected change-point.

The hybrid approach involves a much simpler statistical method combined with station history metadata. The station history information presented by Gaffen (1996) categorizes historical information as one of two types of ‘‘events,’’ termed dynamic or static. A static event is information of the type ‘‘radiosonde instrument type *R* was in use at station *S* in year *Y*.’’ A dynamic event is a change in observing method, such as ‘‘*S* changed from R_1 to R_2 in *Y*.’’ Dynamic event information was used to identify the dates of potential artificial change-points in the data. These dates were then tested using the nonparametric Wilcoxon signed-rank test (Wilks 1995). The test was made using data segments of 1, 3, and 5 yr preceding and following the potential change-point (but ignoring data within 6 months of the point, to allow for potential inaccuracies in the station histories). The most consistent results (from station to station and at different levels for a given station) were achieved using the 5-yr data window, and the results reported below are based on that window size.

In addition to our efforts, at least two other groups are currently attempting to make adjustments for change-points in radiosonde data. Following the methods outlined by Parker et al. (1997), the Hadley Centre of the U.K. Meteorological Office is using the MSU satellite temperature data as a reference time series to make adjustments for the period 1979 to present. Using the detailed radiation correction models described by Luers and Eskridge (1998) and a statistical change-point detection scheme (Eskridge et al. 1995), the National Climatic Data Center of the National Oceanic and Atmospheric Administration (NOAA) is developing methods to adjust monthly mean temperature data for a core set of stations described by Wallis (1998). Both of these efforts will rely on approximately the same station history information we employ. Comparison of the results of these two methods and the approach presented here is planned.

4. Tropospheric and stratospheric trends during 1959–95 based on CLIMAT TEMP data

a. 1959–95 trends in the ‘‘deep’’ network

Figures 5, 6, 7, and 8a present decadal LSLR trends at 30, 100, 300, and 700 hPa for the D network for 1959–95 using the CLIMAT TEMP data. The overall results are qualitatively consistent with those of Parker et al. (1997), Hansen et al. (1997), and Santer et al. (1996, 1999), and show cooling of the lower strato-

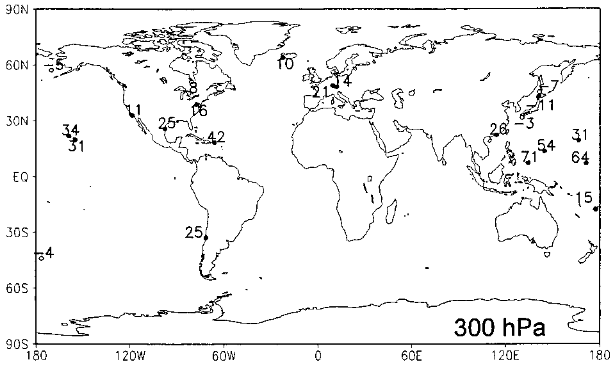

Fig. 7. Same as Fig. 5 but for 300 hPa.

sphere and warming of the troposphere, both of order tenths of a degree per decade. Figure 9 summarizes the trend results for all nine mandatory pressure levels in the D network in box plot form, showing the spread among the trend values from station to station.

At 30 hPa, in the lower stratosphere, all stations have negative trends and 91% are statistically significant at the 99% confidence level (Fig. 5). They range from -0.04 ± 0.09 K decade^{-1} at Kagoshima (WMO No. 47827) to -0.38 ± 0.08 K per decade at San Diego (72293). The 99% confidence intervals for the trends at these two stations do not overlap. Even at very closely spaced stations, the differences can be large. For example, the 30-hPa trend at Stuttgart (10739, a GCOS network station) is -0.34 ± 0.12 K decade^{-1}, while at Munich (10868, one of Angell's stations) it is -0.10 ± 0.13 K decade^{-1}.

The 100 hPa level (Fig. 6) is near the tropical tropopause (Frederick and Douglass 1983) but within the stratosphere at higher latitudes. The decadal trends range from -0.27 ± 0.08 K decade^{-1} at Nandi (91680) to $+0.26 \pm 0.08$ K decade^{-1} at Koror (91408). Thus two tropical stations have almost equal but opposite trends, and their confidence intervals show the trends to be significantly different from zero at the 99% level.

By 300 hPa (Fig. 7) all but five stations have positive trends, and only one negative trend (at Sapporo, 47412) is statistically significant. The magnitude of the warming tends to increase with increasing pressure from the upper to the midtroposphere (Figs. 8 and 9).

At 700 hPa (Fig. 8a), 85% of the stations in the D

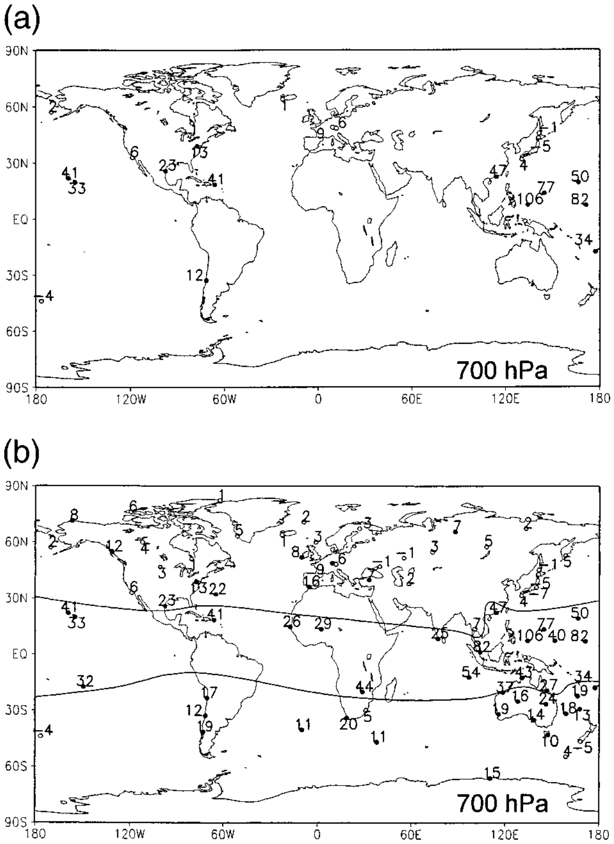

Fig. 8. (a) Same as Fig. 5 but for 700 hPa; (b) same as (a) but for the T network. Isolines indicate regions of trends above and below 0.25 K decade^{-1}.

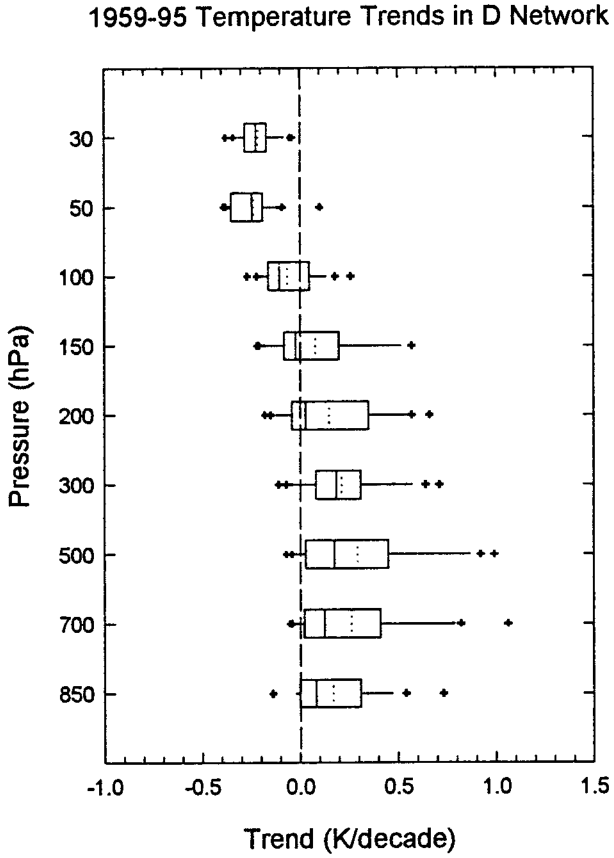

Fig. 9. Box plot of least-squares linear regression trends for 1959–95 based on CLIMAT TEMP data for the 22 stations in the D network as a function of pressure. Each box shows the 25th, 50th, and 75th percentile values; the whiskers show the 10th and 90th percentile values; and outliers are shown as + signs. The dotted lines indicate mean values.

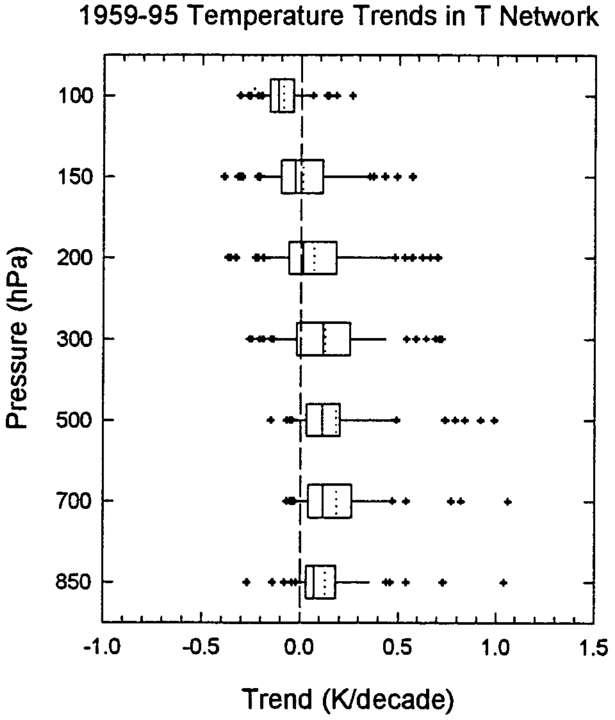

FIG. 10. Same as Fig. 9 but for the 74 stations in the T network.

network have warming trends, more than half of which are statistically significant. In the larger T network, 88% of the stations have positive trends for the same data period (Fig. 8b). The most pronounced warming is in the Tropics, where the trends exceed 0.25 K decade^{-1}; in the western tropical Pacific, they are closer to 1.0 K decade^{-1}. As we discuss below, these linear trend estimates are not particularly good estimates of the true temperature change in the time series, which is more steplike than smoothly linear.

As shown in Fig. 9, there is considerable spread among the trend estimates at each level. The full range (maximum minus minimum) of the trends is at least twice, but up to 10 times, the magnitude of the median value. The distributions are clearly not normal; the mean and median values differ by up to 0.15 K decade^{-1} (at 700 hPa). This suggests that the median or mean should not be taken as representative of the overall network trend.

b. 1959–95 trends in the ''tropospheric'' network

Trend estimates for the same period for the larger (74 station) T network, shown in Figs. 8b and 10, have somewhat larger full ranges than the 22 station D network. However, the interquartile ranges are smaller in the T network, so that expanding the network results in better agreement among the majority of stations but introduces more outliers. At most levels, the median trend values from the two networks are in excellent agreement, but at 300 and 500 hPa they differ by more than 0.10 K decade^{-1}. Thus global mean (or median) trends, computed as the mean of station trend values, can be quite sensitive to the station network used. Recall that the same data availability criterion ($p = 85\%$) was applied to both the D and T networks, but at more levels for the D network (Table 1).

c. Comparison of stratospheric trends for two time periods

To examine the sensitivity of stratospheric trend estimates to length of record and network size, we computed LSLR trends for the 38-station S network for 1970–95. First we compare the decadal trends for 1970–95 with those for 1959–95 at the 22 stations common to the S and D networks. (The distributions of the D network trends for 1959–95 are shown in Fig. 9, and the 30-hPa trends are shown in Fig. 5.) The distributions of trends for 1970–95 (not shown) have mean and median values that are almost identical to those for 1959–95 at all four levels (150, 100, 50, and 30 hPa), and the 5th to 95th percentile ranges are also in excellent agreement. This result is somewhat surprising; we expected to see greater cooling for the shorter record. It is generally believed that most of the lower-stratospheric cooling in the past four decades has occurred since the mid-1980s in association with stratospheric ozone loss, although there are uncertainties associated with all the datasets used in establishing stratospheric trends (Chanin et al. 1999). The similarity of the trends for the two periods suggests some combination of the following circumstances: the recent cooling dominates both sets of time series so that the 11-yr difference has little influence on the trends; the stratosphere was indeed cooling between 1959 and 1969 at a rate comparable to that following 1969; the early data contain some artificial abrupt cooling signals, as suggested by Gaffen (1994). Because our network is extremely sparse in the Southern Hemisphere, and includes no stations in the southern high latitudes, the main region of recent ozone loss and related cooling, these effects are difficult to separate.

d. 1970–95 stratospheric trends in two networks

With the increased stratospheric data available after 1970, we compared the 1970–95 trends for 22 stations with those for the complete, 38-station S network. The distributions of the trends for the 38-station network, shown in Fig. 11, is in close agreement with those for the 22-station subnetwork at 150 hPa. At higher levels, the 38-station network shows median cooling trends that are smaller in magnitude (by about 0.03 to 0.05 K decade^{-1}) than the median cooling of the smaller network. The distribution of trends in the larger network at 30 and 50 hPa is more skewed; the mean trends exceed the medians by about 0.05 K decade^{-1} because of a few large positive trends. As we saw for the tropospheric

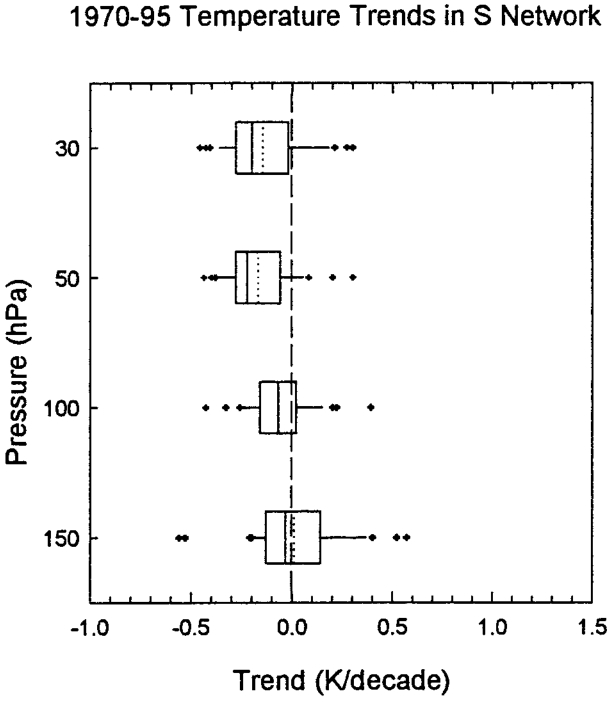

FIG. 11. Same as Fig. 9 but for the 38 stations in the S network for the period 1970–95.

networks (section 4b), network size has a marked impact on estimates of global mean trends.

5. Sensitivity of trends to choice of dataset

Because there is no standard upper-air dataset, investigations of upper-air temperature trends have been based on somewhat different datasets, with different station networks, periods of record, and statistical approaches. In this section we isolate the effect of choice of data source by comparing trends for a fixed network, over a fixed period, using a single trend estimation method. Using the twenty stations in the C network (appendix), we computed LSLR trends at the nine pressure levels between 850 and 30 mb for 1959–91 using both the CLIMAT TEMP and CARDS monthly temperature anomalies. (The trends from the CLIMAT TEMP data for this period were generally very close to the trends presented in the previous section for the longer period ending in 1995.)

Trends computed from the CLIMAT TEMP data were generally different from those computed with the CARDS data. In most cases they differed by less than ±0.1 K decade^{-1}; as shown in Fig. 12, at each level, at least 14 (or 70%) of the 20 station trends agreed to this tolerance. At 30 hPa, six stations had trend differences of at least 0.1 K decade^{-1}. In this (admittedly rather small) network, the CARDS record shows more stratospheric (30 and 50 hPa) cooling and less midtropospheric (300, 500, and 700 hPa) warming than the CLIMAT TEMP record.

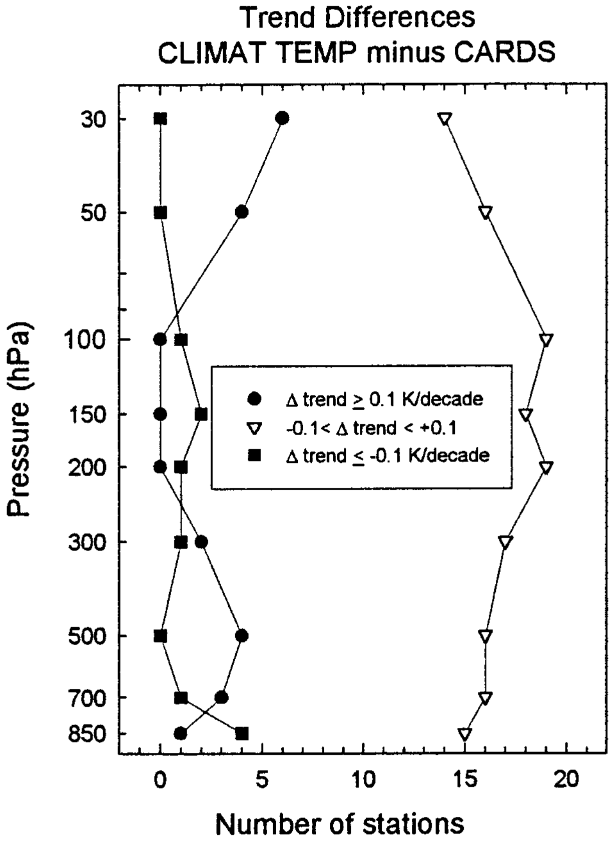

FIG. 12. Categories of trend differences for the twenty-station C network for 1959–91. Differences are related to choice of dataset and are shown as CLIMAT TEMP trends minus CARDS trends.

In some cases, large trend differences are related to change-points in the differences between the two temperature time series. For example, Fig. 13 shows the 30-hPa monthly mean temperature data from Kagoshima (47827) from the CLIMAT TEMP and CARDS datasets and the differences between the two time series. The linear regression trends for the anomaly time series created from these data are 0.00 ± 0.11 for the CLIMAT TEMP data and -0.25 ± 0.09 K decade^{-1} for the CARDS data, and the two confidence intervals have no overlap. The CARDS dataset has temperatures consistently higher (by up to 2 K) than the CLIMAT TEMP dataset through 1981. During 1982–91, the monthly differences seem more random but still relatively large, several tenths of a degree or more (Fig. 13 bottom). The CARDS data are based on both 0000 and 1200 UTC data throughout the period, but the CLIMAT TEMP data are for a variety of observing times before 1970, but only for 1200 UTC (nighttime) data thereafter. This difference may explain the differences for the period 1970–81, but does not explain the very high values before 1970.

Although the influence of dataset choice is largest at

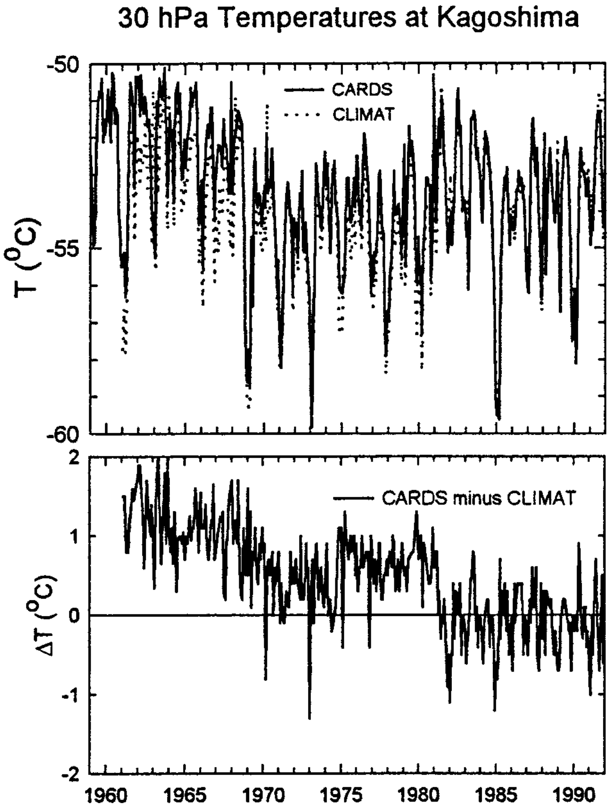

FIG. 13. Time series of 30-hPa temperatures from the CARDS and CLIMAT TEMP datasets at Kagoshima (47827, top) and differences (bottom).

30 hPa at Kagoshima (and this difference is among the largest found at any station or level in our C network), the effect is not negligible at other levels. Figure 14 shows linear regression trends, and their 99% confidence intervals, for Kagoshima based on both CLIMAT TEMP and CARDS data for the C network period 1959–91. At six of the nine levels, the two trends differ by more than 0.1 K decade^{-1}, but only at 30 and 50 hPa do the confidence intervals for the two trends not overlap.

Figure 14 also shows linear regression trends for the longer D network period 1959–95. Shortening these records by about 12% (4 yr) changes the trends by less than 0.05 K decade^{-1}, and both trends fall within each other's 99% confidence intervals.

A second example is Majuro; Fig. 15 shows trends based on both datasets. The trends are largest in the troposphere, where both datasets indicate warming approaching 1.0 K decade^{-1}, but they differ by as much as 0.2 K decade^{-1}. Because of the large confidence intervals, these differences are not statistically significant. At this station, the effect of shortening the CLIMAT TEMP dataset is to change the trend by up to about 0.2 K decade^{-1}, but the differences are not significant at the 99% level.

Examination of the time series of differences between the Majuro time series from the two datasets (not shown) revealed abrupt changes, as was the case at Kagoshima. From 1966 through 1972, the CLIMAT TEMP temperatures are up to 2K higher than the CARDS values. One possible explanation can be found in the different observation times used in computing the monthly means. For the CLIMAT TEMP data, 0000 UTC observations were used for the periods January 1959–February 1962 and August 1963–December 1991, and 1200 UTC data were used in the remaining months. However, the CARDS data have little 0000 UTC data before 1962, and very few 1200 UTC observations for 1964–65, and for March 1973–October 1989.

In summary, we find that trends computed from the CLIMAT TEMP and CARDS dataset are not identical, but that, for the period 1959–91, they generally differ by less than 0.1 K decade^{-1}. Nevertheless, there are cases in which the differences are larger, because of periods of record when one dataset has a distinct bias with respect to the other, often related to time of observation. This highlights one advantage of daily sounding data (e.g., CARDS) over CLIMAT TEMP monthly

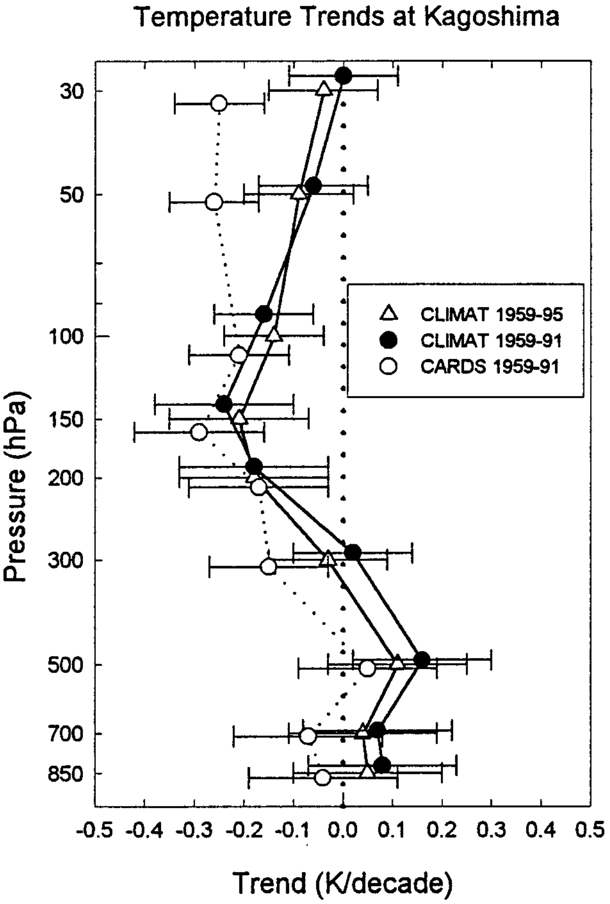

FIG. 14. Linear regression temperature trends at Kagoshima (47827) based on the CLIMAT TEMP data for two periods (the D network period 1959–95 and the C network period 1959–91) and based on the CARDS data for 1959–91.

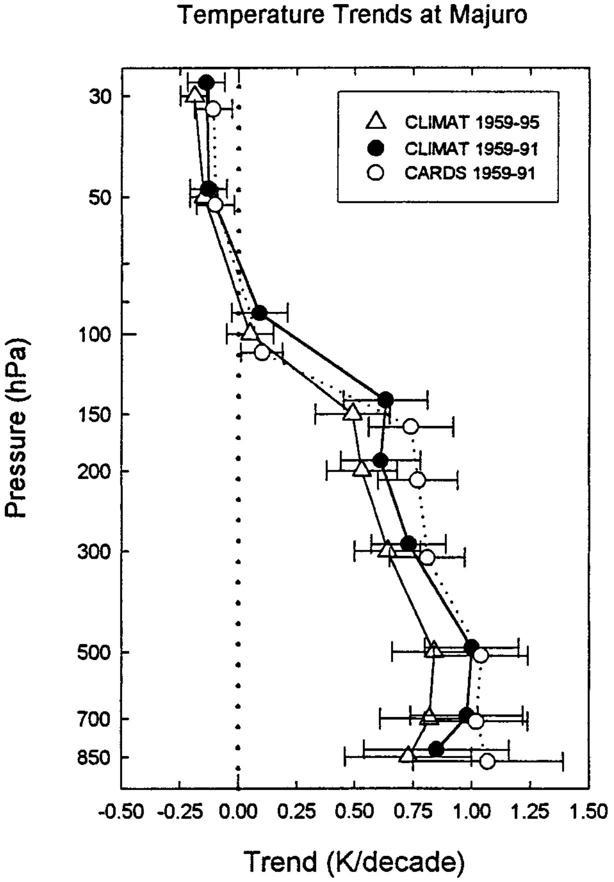

FIG. 15. Same as Fig. 14 but for Majuro (41376).

mean reports; the changing mix of observations at different times of day in the latter cannot be rectified.

6. Sensitivity of trends to trend estimation method

As has been demonstrated by several investigators (Parker 1985; Gaffen 1994; Finger et al. 1995; Lanzante 1996), and as illustrated in Fig. 13, temperature time series from radiosonde data can exhibit change-points that will clearly influence LSLR trends. Because the MPS method is less sensitive to change-points than LSLR, it may yield more realistic trends. To examine influence of trend estimation method on computed trends, we computed both LSLR and MPS trends, and their confidence intervals, for all the time series in the T and S networks using the CLIMAT TEMP data. In the majority of cases, the difference is smaller than ±0.03 K decade^{-1}, which is well within the 99% confidence intervals of each trend estimate.

The largest difference in 1959–95 trends in the T network is 0.10 K decade^{-1} at 700 hPa at Hong Kong (45004). The LSLR trend is 0.47 ± 0.17 and the MPS trend is 0.37 (with confidence interval 0.02–0.52) K decade^{-1}, so the estimates are not significantly different at the 99% confidence level. The time series of monthly anomalies, shown in Fig. 16 (top), shows a steplike warming in about 1974. As discussed in section 3, the MPS method (which yields the lower trend estimate) is less sensitive to this feature, which may be related to the December 1974 instrument changes (from the Vaisala RS 13 or RS 15 radiosonde to the Vaisala RS18), or to a shift in climate state, as discussed by Trenberth (1990). A similar example is at 850 hPa at Singapore (48698), where the LSLR trend 1.04 ± 0.16 and the MPS trend is 0.08 K decade^{-1} smaller. Here the mid-1970s warming is even more pronounced (Fig. 16, bottom), but again it is convoluted with instrument changes at the station, four of which are documented during the 1970s.

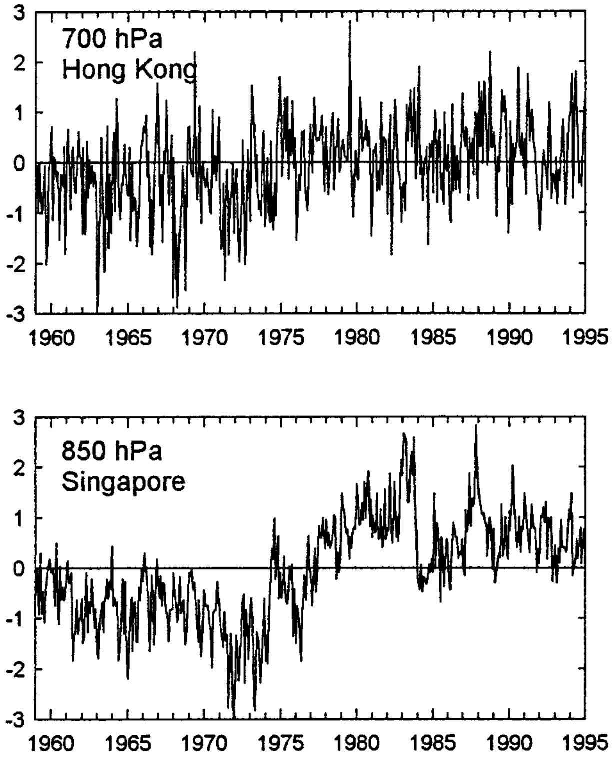

FIG. 16. Time series of monthly temperature anomalies (K) at (top) 700 hPa at Hong Kong (45004), and (bottom) 850 hPa at Singapore (48698). The steplike warmings in the mid-1970s may be related, in part, to radiosonde instrument changes.

The largest difference in 1970–95 trends in the S network is 0.16 K decade^{-1} at 30 hPa at Lerwick (03005). There the LSLR estimate is −0.23 ± 0.15 and the MPS estimate is −0.07 (with confidence interval −0.17, +0.02) K decade^{-1}. The LSLR method yields cooling at a rate more than three times that of the MPS, demonstrating the sensitivity of LSLR to the high anomalies at the start of the time series (shown in Fig. 17).

In summary, the MPS and LSLR decadal trend estimates tend to yield very similar, but not identical results, with differences generally less than ±0.03 K decade^{-1}. In a few instances, the differences are larger, but not so large as to be significant at the 99% confi-

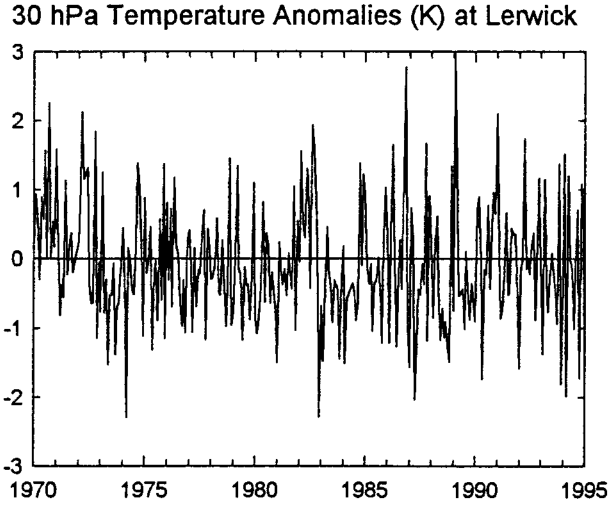

FIG. 17. Time series of monthly 30-hPa temperature anomalies at Lerwick (03005).

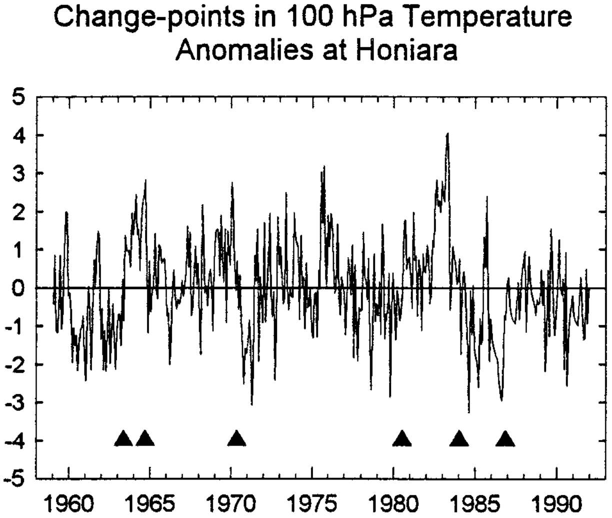

FIG. 18. Time series of monthly 100-hPa temperature anomalies (K) at Honiara (91517) and dates of change-points detected by the Lanzante scheme. Table 2 gives numerical values of the size of each detected discontinuity.

dence level. Larger differences are found for the 1970–95 than for the longer period 1959–95. This result is consistent with those of Santer et al. (2000), who consider trends in global mean temperatures from various surface and upper-air datasets. Comparing LSLR with least absolute deviation linear regression trends, they find differences generally less than 0.05 K decade^{-1} for 1959–96, and larger differences for shorter periods. Peterson et al. (1998b), on the other hand, found differences in global surface temperature trends, between LSLR and a resistant trend estimator, of up to ~0.1 K decade^{-1} for the much longer period 1880–90. Thus, like Peterson et al. (1998b) and Santer et al. (2000), we caution that trends can be sensitive to trend estimation method (particularly when the time series shows highly nonlinear change) and recommend that investigators specify the method used when reporting trends.

7. Sensitivity of trends to removal of change-points using multiple change-point statistical methods only

The prospect of using objective, statistical methods to detect, measure, and make corrections for artificial change-points in time series is an attractive one. However, as enumerated in section 3c, many complications make this a less-than-straightforward goal. As an example, consider a single time series of 100-hPa temperature anomalies at Honiara (91517) and one change-point detection scheme—that of Lanzante (1996). The scheme identifies segments of the time series that are not interrupted by change-points, as well as the location and magnitude of the change-points separating those homogeneous segments. An adjustable parameter α is the confidence level associated with the change-point detection.

Figure 18 shows the time series and Table 2 gives the sizes of all detected change-points for α equal to 0.05, 0.01, and 0.001. For $\alpha = 0.05$, six change-points were detected, and four of those correspond with dates of known changes in radiosonde type. With the more stringent $\alpha = 0.01$, only four change-points were detected, and for $\alpha = 0.001$, only two were detected. Thus the number of detected change-points decreases as the confidence level of their detection increases. This result typifies our impression of the performance of both the Lanzante (1996) and Habermann (1987) schemes, and is probably also applicable to other objective statistical methods for multiple change-point detection. That is, the methods generally succeed in detecting the primary (largest, visually obvious) discontinuities and they agree quite well in identifying their timing. However, they often disagree about the existence of secondary (smaller, less obvious) change-points.

More important, however, the methods tend to dis-

TABLE 2. Results of application of the Lanzante scheme for multiple change-point detection to time series of 100-hPa temperature anomalies at Honiara (91517, Fig. 18). The number of change-points detected and their magnitude depends on the significance level α. The dates of documented changes in radiosonde instrument type are shown for comparison with the independently detected change-point dates.

Date of detected change-point	Date of radiosonde type change	Detected discontinuity (K)		
		$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
May 1963	Nov 1963	2.33	1.34	1.09
Sep 1964	None	−1.23	Not detected	Not detected
May 1970	None	−0.45	−0.69	Not detected
Jul 1980	Aug 1979	1.14	1.14	Not detected
Jan 1984	Aug 1984	−2.29	−1.64	−0.95
Nov 1986	Dec 1986	0.98	Not detected	Not detected

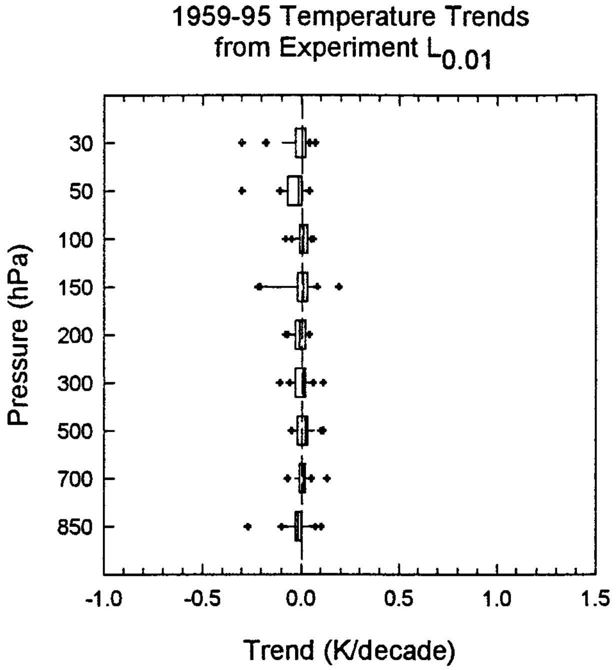

FIG. 19. Results of removing change-points as specified by experiment $L_{0.01}$. Trends are presented as box plots as in Fig. 9.

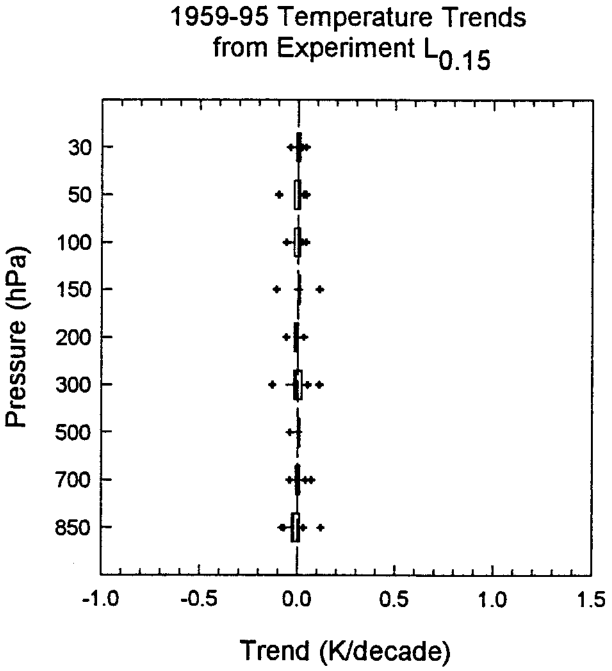

FIG. 20. Results of removing change-points as specified by experiment $L_{0.15}$.

agree about the magnitude of the detected change-points. The Honiara example (Fig. 18, Table 2) is typical in that the discontinuity (which is computed based on the median values of the surrounding segments, and thus their length) depends strongly on α (which controls the number of change-points and thus the lengths of the segments). The 1963 upward jump in temperature is estimated to be 2.33, 1.34, or 1.09 K, for the three α values used. This diversity of jump sizes is problematic if one wishes to adjust the time series. Furthermore, the abrupt warming follows the March 1963 eruption of Mt. Agung, and thus is probably not altogether spurious. In light of these complications, we reiterate that the adjusted time series, and resulting trends estimates, presented below are meant primarily to determine the sensitivity of trends to adjustments, and are not necessarily improvements over the unadjusted data.

Using the least squares linear regression trends for the D network presented in section 4 (see Fig. 9) as controls, we discuss here the results of four experiments to measure the sensitivity of trends to the removal of change-points. In each case, change-points are identified by statistical methods alone, and the time series are adjusted to remove each change-point. The change-point methods are applied independently to the data at each pressure level; nothing constrains the results at a given level to agree with those at other levels. The experiments are as follows:

$L_{0.01}$—change-points are detected using the Lanzante scheme with $\alpha = 0.01$,
$L_{0.15}$—change-points are detected using the Lanzante scheme with $\alpha = 0.15$,
$H_{0.01}$—change-points are detected using the Habermann scheme with $\alpha = 0.01$,
$H_{0.15}$—change-points are detected using the Habermann scheme with $\alpha = 0.15$.

Figures 19 and 20 show the distribution of trends for experiments $L_{0.01}$ and $L_{0.15}$. The overwhelming effect of removing change-points in the data is to remove the trends (see Fig. 9 for comparison), and the trends are most effectively removed when change-point detection and adjustment is at a lower confidence level ($L_{0.15}$, Fig. 20). The results of $H_{0.01}$ and $H_{0.15}$ (not shown) lead to the same conclusion, which is subject to various interpretations. If one believes the statistical methods have detected all artificial change-points and only artificial change-points, the lack of trend in the adjusted time series calls into question the claims of stratospheric cooling and tropospheric warming that have been generally accepted in the literature (e.g., Santer et al. 1996).

However, independent lines of evidence (surface temperature records, satellite data, etc.) support the finding of global temperature trends. Our examination of the detected change-points, along with station history information, indicates that, although the statistical methods frequently detect change-points at times of known instrument changes, they also often detect change-point at times when there is no known change. Furthermore, there is nothing built into the purely statistical methods to distinguish between real and artificial changes. The detected change-points can often be associated with what we believe to be real atmospheric temperature change, related to volcanic eruptions, ENSO events, or climate regime shifts. Recall that the Lanzante (1996) scheme specifically tests to ensure that time series segments that exhibit a trend are not mistakenly attributed a change-point within the segment. Therefore, it is not the case that change-point removal would automatically eliminate trends.

For these reasons, we interpret the results of these experiments as evidence that atmospheric trends are generally not completely linear, but rather result, at least in part, from steplike changes in temperature. Some of these may be artificial but others are likely real (see, e.g., Pawson et al. 1998). Therefore, change-point detection schemes that do not integrate other information (such as station history) will remove both artificial and real temperature changes and the resulting trends are both negligibly small and negligibly meaningful.

8. Sensitivity of trends to removal of change-points using a single-change-point statistical method combined with station histories

From the previous analysis, it is clear that improved adjustments of time series must involve effectively removing artificial change-points without removing true atmospheric variability. This can only be accomplished using additional information. Two sorts of information are potentially useful. One is station history information, which can help identify times of potential artificial change-points and the other is information that would allow one to identify true temperature variability. Here we focus on the use of station history information to isolate artificial change-points. In section 9 we propose an alternate approach involving removal of natural variations to isolate artificial ones.

We apply the Wilcoxon signed-rank test for single change-points at the dates of changes in radiosonde models as suggested by the station histories, for the S and T networks (to maximize the number of stations tested). For each instance of a particular historical event (a change from one particular radiosonde type to another) that occurred at two or more stations (but not necessarily at the same time), we tally the results of the change-point tests at each pressure level. If at least half of the station tests at a given pressure level indicate the existence of a change-point at the 99% confidence level, we adjust the time series of those stations, and at those levels, for which the test was statistically significant. The adjustment is made from the date of the instrument change all the way back to the start of the time series. The size of the adjustment varies from station to station and is the difference in the median monthly temperature anomaly between the 5-yr periods preceding and following the instrument change. However, we do not consider the data within 6 months of the documented instrument change, to allow for uncertainty in the instrument change date.

This approach has the advantage that it incorporates station history information and is based on objective statistical methods. As in the Habermann (1987) and Lanzante (1996) methods, it does not rely on reference time series or on information about instrument measurement characteristics, which might not always be available. By requiring some concurrence among stations before adjusting the time series at one station, we reduce (but do not eliminate) the possibility of mistaking real change associated with interannual climate variations for artificial change associated with instrument changes. Compared with section 7, in which the multiple change-point detection schemes allowed for fairly liberal adjustment of many change-points, our application of this single change-point method is much more conservative and allows adjustment only when relatively strict criteria are met.

There are several deficiencies with this approach. First, there is no constraint to ensure the vertical or horizontal consistency of the adjustments. Second, the estimated magnitude of the change-point may be enhanced or reduced by real changes that are close in time to a given instrument change. Third, known inhomogeneities [e.g., those associated with changes between VIZ and Vaisala radiosondes, McMillin et al. (1988)] are not treated if the change-point tests are not statistically significant. Fourth, obvious change-points are not adjusted if there is insufficient station history information or there are no other instances of the particular instrument change in question to justify the change.

These last two constraints explain why, despite a station history rich with instrument change information, our method yielded adjustments for only five types of instrument changes. Those adjustments, and their effects on LSLR trends, are shown in Tables 3–7. We discuss each set of adjustments in turn.

The first set of adjustments is associated with the circa 1969 change from Mesural 1940B radiosondes to Mesural 1943B radiosondes (Table 3). The former carried bimetal temperature sensors, and the latter carried thermistors. This change was documented to have occurred at two stations, Tahiti (91938) and Noumea (91592), although it is likely that it also took place at other French-operated sites (e.g., in Africa) for which we have little or no station history. The change-point test yielded significant results at the levels shown in Table 3. Note that at 150 hPa, only Tahiti showed a significant change. The detected discontinuities are all positive, indicating an artificial upward jump, so the LSLR trends after adjustment all show more cooling (or less warming) than the trends based on the original data. The discontinuities vary with height and between the two stations. Not all pressure levels were adjusted. Figure 21 shows the adjustments made at Tahiti and their effects on the trends. The original data indicated warming at 850, 700, and 500 hPa, and cooling at higher

TABLE 3. Adjustments, and their influence on 1959–95 trends and confidence intervals, associated with the 1969 change from Mesural 1940B radiosondes to Mesural 1943B radiosondes.

Pressure (hPa)	Station	Adjustment (K)	Trend (K per decade)	
			Original	After adjustment
150	Tahiti (91938)	0.41	−0.31 ± 0.04	−0.44 ± 0.03
200	Noumea (91592)	0.56	−0.37 ± 0.08	−0.54 ± 0.08
	Tahiti	0.99	−0.22 ± 0.06	−0.54 ± 0.06
300	Noumea	0.79	−0.21 ± 0.10	−0.45 ± 0.09
	Tahiti	1.28	−0.09 ± 0.08	−0.51 ± 0.07
500	Noumea	1.39	0.04 ± 0.14	−0.39 ± 0.13
	Tahiti	1.53	0.11 ± 0.11	−0.40 ± 0.10
700	Noumea	0.92	0.19 ± 0.14	−0.10 ± 0.13
	Tahiti	1.14	0.32 ± 0.12	−0.06 ± 0.12
850	Noumea	0.93	−0.04 ± 0.13	−0.33 ± 0.12
	Tahiti	1.03	0.35 ± 0.15	0.01 ± 0.15

levels. The trends after adjustment show no warming that is significant at the 99% level, but cooling at each level from 700-hPa upward. The adjustment enhances the cooling in the upper troposphere; it is more than doubled at 200 hPa and five times larger at 300 hpa (where the original trend was only −0.09 K decade^{-1}).

The second set of adjustments is associated with the change from Meisei RSII-56 radiosondes, carrying bimetal temperature sensors, to the Meisei RSII-80 radiosondes, carrying bead thermistors. This change took place circa 1981 at the five Japanese stations in our S network; however, the change-point test yielded significant results only at the three stations listed in Table 4. The Kagoshima (47827) time series (Fig. 13) shows the upward jump in temperatures associated with this instrument change. The detected discontinuities at 50 and 30 hPa are fairly consistent from station to station (Table 4), all indicating an upward jump in monthly temperature anomaly between 0.48 and 0.76 K. Adjusting the time series for those jumps changes the 1970–95 trends from warming to cooling at all three stations and at both levels, and in five of the six cases (two levels at each of three stations) there is no overlap in the 99% confidence intervals of the original trends and the trends after adjustment (Table 4).

Table 5 shows a third set of adjustments for the change from RKZ-2 radiosondes, carrying rod thermistors, to MARS radiosondes, with the same sensors but treated with antiradiation coating. This change occurred at only two of the nine stations in the former U.S.S.R. included in our T network. In the U.S.S.R. frequent instrument changes were made, but not always in the same sequence at each station (Gaffen 1993), making it difficult to attempt adjustments in this large network. The effect of these adjustments on computed trends is enhanced cooling at 100 and 200 hPa. According to Luers and Eskridge (1998), there is little error in the nighttime observations from either the RKZ-2 or MARS radiosondes; however, both have large (±1.0 K) daytime errors, and corrections were made for the MARS data.

The final two sets of adjustments are for stations in

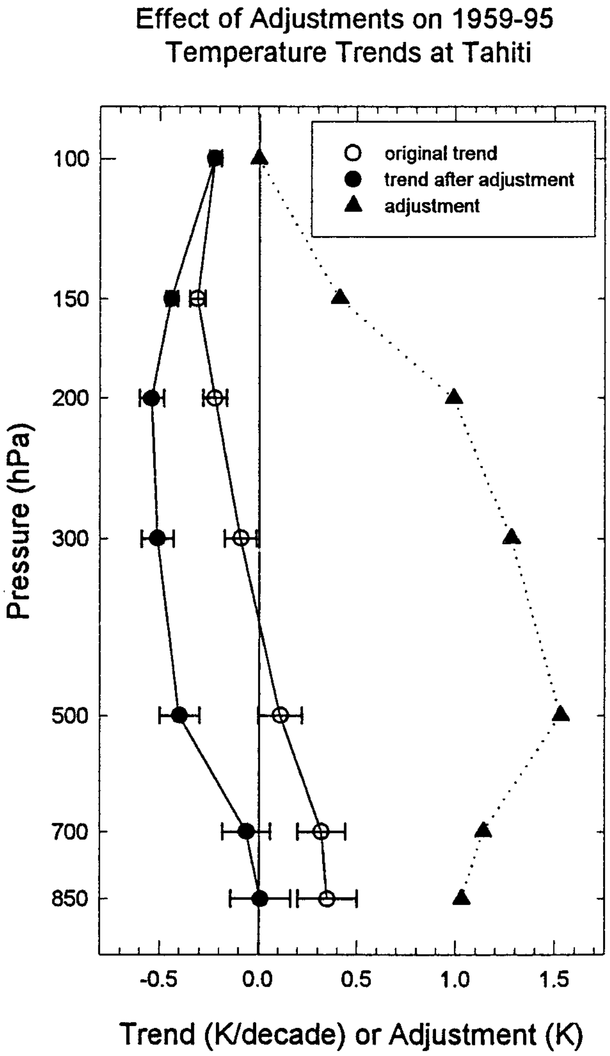

FIG. 21. Original trend estimates and revised estimates after adjustments made to Tahiti temperature records. The magnitudes of the adjustments are also shown.

TABLE 4. Adjustments, and their influence on 1970–95 trends and confidence intervals, associated with the 1981 change from Meisei RSII-56 to Meisei RSII-80 radiosondes.

Pressure (hPa)	Station	Adjustment (K)	Trend (K per decade)	
			Original	After adjustment
30	Tateno	0.48	0.18 ± 0.18	−0.10 ± 0.18
	Kagoshima	0.70	0.30 ± 0.15	−0.09 ± 0.14
	Minamitoroshima	0.60	0.27 ± 0.13	−0.06 ± 0.13
50	Tateno (47646)	0.62	0.06 ± 0.15	−0.29 ± 0.14
	Kagoshima (47827)	0.67	0.20 ± 0.14	−0.18 ± 0.13
	Minamitoroshima (47991)	0.76	0.30 ± 0.14	−0.12 ± 0.13

the Australian network. Because so many Australian stations were included in the GCOS and Angell networks, and because of fairly consistent instrument changes within the network, our method of identifying change-points yielded a relatively large number of adjustments. These circumstances should not be misinterpreted as indication that the Australian data are more prone to discontinuities than the data from other nations.

The circa 1979 change from Astor Mark I radiosondes to Philips Mark II radiosondes, both carrying rod thermistors, is associated with 700 (or 500) hPa upward jumps in temperature anomalies between 0.2 and 1.2 K at six (or seven) stations. However, 12 stations in our T network experienced this instrument change, and close to half showed no significant discontinuity. Furthermore, in several instances, although the change-point test yielded significant results, visual examination of the time series raises doubts that these are artificial change-points. For example, Fig. 22 (bottom) shows the 500-hPa anomaly time series at Townsville (94294). While warming does seem to follow the 1979 instrument change, it is not long-lived. Note, however, that other instrument changes occurred in 1976, 1983, and 1987, but our test criteria were not met for the 1976 or 1983 changes. As shown in Table 6 and Fig. 23, an adjustment of 0.65 reduces the 1959–95 trend from 0.30 to 0.04 K decade^{-1}. At most stations, 500-hPa warming, significant at the 99% confidence level, is eliminated by the adjustment (Fig. 23). Five Australian stations experienced the same instrument change, but no adjustments were made because significant change-points were not detected. Warming at these stations is intermediate to the original and adjusted trends at the other seven stations (Fig. 23).

The final adjustment, associated with the circa 1987 change from Philips Mark III radiosondes, with rod thermistors, to Vaisala RS80 instruments, with capacitive bead temperature sensors, is for 200-hPa temperatures at five of the ten stations experiencing the change (Table 7). The data for Townsville (94294, Fig. 22, top) indicate a rise in temperature around the time of the instrument change, although high values are not sustained to the end of the record. As discussed by Talbot (1972), the thermistor had a cold bias at night. The effect of removing this approximately 1 K temperature rise near the end of the time series is to substantially reduce the trends (Table 7). At the 500-hPa level, the time series also appears to have a steplike change (Fig. 22, bottom), but our test results at this and other stations did not warrant adjustment.

For each of the five adjustments (Tables 3–7) the trends after adjustment tend to show somewhat better agreement among the affected stations than the original trends. Another general result is that the adjustments reduce the confidence interval, probably because of the reduced variance in the adjusted time series.

Parker et al. (1997) adjusted radiosonde data from Australia and New Zealand for discontinuities at four dates of documented instrument changes using the MSU data as a reference for the period 1979–95. Their largest adjustments were in the stratosphere, and the effect of the adjustments on zonal trends was to reduce both stratospheric cooling and tropospheric warming between 15° and 45°S. Our adjustments in this regions are only for two instrument changes and only at tropospheric levels. They also reduce tropospheric warming, but because no Australian stations were accepted in our S network, no adjustments were made at stratospheric levels. Again, we make no claim that one or another method brings the data closer to reality.

TABLE 5. Adjustments, and their influence on 1959–95 trends and confidence intervals, associated with the 1969 change from RKZ2 to MARS radiosondes.

Pressure (hPa)	Station	Adjustment (K)	Trend (K per decade)	
			Original	After adjustment
100	Rostov-na-Donu (34731)	0.43	−0.10 ± 0.07	−0.25 ± 0.07
	Orenburg (35121)	0.51	−0.12 ± 0.07	−0.28 ± 0.06
200	Rostov-na-Donu	0.37	0.01 ± 0.07	−0.11 ± 0.07

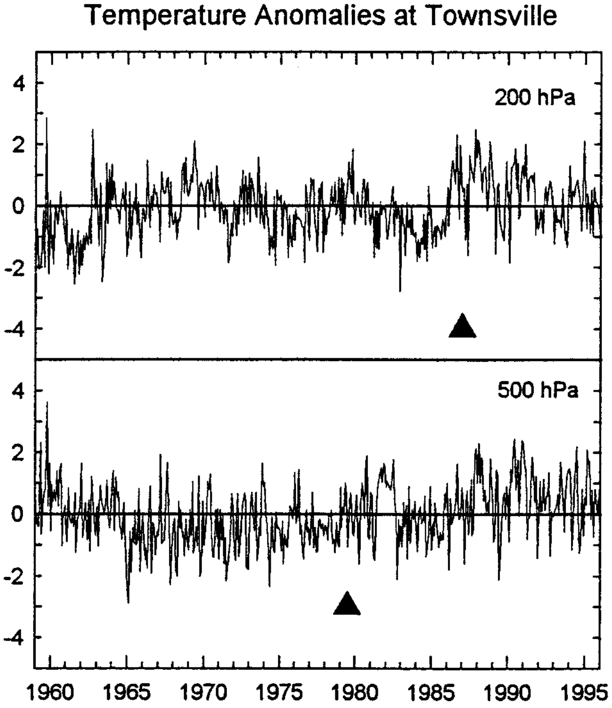

Fig. 22. Temperature anomaly time series for Townsville (94294) at 200 and 500 hPa. Dates of two instrument changes for which adjustments were made are shown, although other instrument changes occurred during this period.

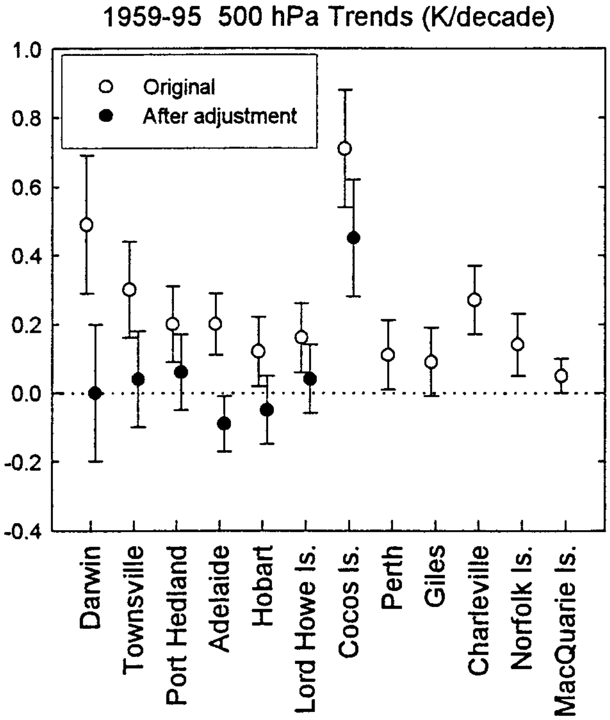

Fig. 23. Temperature trends and confidence intervals at 500 hPa for 1959–95 using unadjusted data at 12 Australian radiosonde stations. At 7 of the 12, the data were adjusted for the circa 1979 radiosonde change, and the resulting trends are also shown. See also Table 7.

9. Summary

In this paper we have examined the sensitivity of temperature trend estimates based on radiosonde data to several issues relating broadly to the quality of the data. Our main findings can be summarized as follows.

a. Data availability and representativeness

For the analysis of trends, long station data records with minimal missing data are most desirable. The records for the 180 stations in the combined Global Climate Observing System Baseline Upper-Air Network and Angell networks, however, do not generally meet this standard, and the problem is particularly acute for the lower stratosphere.

Trend estimates at a given pressure level vary considerably from station to station. The range among stations can be two to ten times the magnitude of the median value, and the trend estimates are not normally

Table 6. Adjustments, and their influence on 1959–95 trends and confidence intervals, associated with the 1979 change from Astor Mark I to Philips Mark II radiosondes.

Pressure (hPa)	Station	Adjustment (K)	Trend (K per decade) Original	Trend (K per decade) After adjustment
500	Darwin (94120)	1.23	0.49 ± 0.20	0.00 ± 0.20
	Townsville (94294)	0.65	0.30 ± 0.14	0.04 ± 0.14
	Port Hedland (94312)	0.40	0.20 ± 0.11	0.06 ± 0.11
	Adelaide (94672)	0.36	0.20 ± 0.09	−0.09 ± 0.08
	Hobart (94975)	0.73	0.12 ± 0.10	−0.05 ± 0.10
	Lord Howe Island (94995)	0.45	0.16 ± 0.10	0.04 ± 0.10
	Cocos Island (96996)	0.64	0.71 ± 0.17	0.45 ± 0.17
700	Giles (94461)	0.68	0.16 ± 0.12	−0.11 ± 0.12
	Adelaide	0.64	0.14 ± 0.09	−0.11 ± 0.09
	Lord Howe Island	0.45	0.18 ± 0.11	0.00 ± 0.11
	Cocos Island	0.95	0.54 ± 0.16	0.16 ± 0.16
	Port Hedland	0.34	0.37 ± 0.13	0.23 ± 0.13
	Hobart	0.21	0.10 ± 0.09	0.01 ± 0.09

TABLE 7. Adjustments, and their influence on 1959–95 trends and confidence intervals, at 200 hPa, associated with the 1987 change from Philips Mark II to Vaisala RS80 radiosondes.

Station	Adjustment (K)	Trend (K per decade)	
		Original	After adjustment
Darwin (94120)	1.25	0.62 ± 0.18	0.26 ± 0.18
Townsville (94294)	1.23	0.27 ± 0.16	−0.06 ± 0.17
Port Hedland (94312)	0.77	0.47 ± 0.12	0.26 ± 0.12
Perth (94610)	0.66	0.09 ± 0.08	−0.10 ± 0.08
Cocos Island (96996)	1.33	0.70 ± 0.13	0.34 ± 0.12

distributed. Therefore, the median or mean trend should not be taken as representative of trends throughout the network, especially because changing network size can change the mean trend value by up to 0.1 K decade^{-1}.

b. Sensitivity to choice of dataset

We compared 1959–91 LSLR trends for twenty stations at levels between 850 and 30 hPa using two datasets: CLIMAT TEMP monthly mean temperatures reported by stations CARDS monthly means computed from archived daily soundings. Temperature trends in these two datasets are not identical, but they generally differ by less than 0.1 K decade^{-1}. In a few cases, involving periods of record when one dataset has a distinct bias with respect to the other, the differences are larger, and the trends are statistically significantly different at the 99% confidence level.

c. Sensitivity of trends to method of estimating trends and their confidence intervals

The distribution of radiosonde temperature data may not always meet the assumptions underlying the least-squares linear regression trend estimation method, in which case the nonparametric median of pairwise slopes method may be more appropriate. The two methods tend to yield very similar but not identical results for 1959–95, with differences generally less than ±0.03 but occasionally up to 0.10 K decade^{-1} or more.

d. Sensitivity of trends to the existence of change-points

Instrument changes (and other nonclimatic signals) can lead to abrupt changes in the mean, or change-points, in radiosonde temperature data records. We explored two approaches to adjusting change-points in radiosonde temperature data.

The first approach involved two statistical schemes for multiple change-point detection and adjustment of all detected change-points (Habermann 1987; Lanzante 1996). Both schemes succeeded in detecting the largest, visually obvious, change-points and generally agreed very well on their timing, but differed in their detection of smaller change-points. More important, the magnitude of the detected change is very sensitive to the particular scheme employed and its implementation. The overwhelming effect of adjusting time series using these schemes is to remove the trends, probably because some of the detected change-points are not spurious signals but represent real atmospheric change.

The second approach incorporated station history information (Gaffen 1996) to help identify dates of potential change-points. These were tested using a single change-point detection scheme and adjustments were made only if change-points were detected in time series for a given pressure level for at least half of the stations experiencing a given instrument change. This approach yielded adjustments, for five different instrument changes at various stations, which reduced tropospheric warming trends (or enhanced tropospheric cooling) during 1959–95 and (in one case) enhanced stratospheric cooling during 1970–95. The trend differences due to the adjustment were often statistically significant at the 99% confidence level.

The two different approaches used in this study (statistics only vs statistics plus station histories) represent opposite extremes. The former is very liberal in its adjustment strategy because it cannot distinguish between artificial and natural variability. The latter is very conservative in that it adjusts only artificial changes, which are identified with a high degree of confidence. From this analysis we conclude that adjustment for change-points can yield very different time series depending on the scheme used to make adjustment and the manner in which it is implemented. Accordingly, trend estimates are extremely sensitive to the adjustments. These two extremes serve to illustrate the uncertainties in the methodology and highlight the crucial issues.

Using the lessons gained herein, work is in progress to attempt to make adjustments with more confidence. A multifaceted approach to distinguish between abrupt natural and artificial changes incorporates the statistical change-point detection and station history information used in the present study, but adds other types of metadata, which can be derived from the data (e.g., day–night differences, time of observation changes, and station elevation). It also utilizes vertical coherence of features in the temperature series as well as other (non-temperature) measures of natural variability. We anticipate that by combining a number of independent indicators, we will have much more confidence in the identification and adjusment of artificial change points, and, therefore, in the trends in the resulting time series.

Acknowledgments. We thank David Parker, Margaret Gordon, and David Cullum (UKMO) for providing the CLIMAT TEMP data; Bert Eskridge and Helen Frederick (NOAA/NCDC) for providing the CARDS data; Eric Slud (University of Maryland) for statistical advice; Jim Angell (NOAA/ARL) for many helpful discussions; David Parker and Simon Brown (UKMO), Tom Knutson and Gabriel Lau (NOAA/GFDL), and two anonymous reviewers for constructive comments on the manuscript; and Becky Ross (NOAA/ARL) for creating the map figures. The NOAA Office of Global Programs, Climate Change Data and Detection program provide partial support for this work.

APPENDIX

Radiosonde Stations and Networks Used in this Study

WMO number	Lat	Long	Station	Networks
01001	70.95	−8.67	Jan Mayen	T S
02836	67.37	26.65	Sodankyla	T
03005	60.13	−1.18	Lerwick	T S
03953	51.93	−10.25	Valentia	T S
04018	63.95	−22.62	Keflavik	T S D C
04270	63.43	−51.18	Marrak Point	T
08495	36.15	−5.35	North Front	T S
10739	48.83	9.20	Stuttgart	T S D C
10868	48.25	11.58	Munchen	T S D C
17130	39.95	32.88	Ankara	T S
23472	65.78	87.95	Turuhansk	T
24266	67.52	133.38	Verkhoyansk	T S
28698	54.93	73.40	Omsk	T S
30230	57.77	108.12	Kirensk	T
32540	53.08	158.55	Petropavlovsk	T S
34731	47.08	39.73	Rostov-na-Donu	T
35121	51.75	55.12	Orenburg	T S
38880	37.97	58.33	Ashabad	T S
43371	8.48	76.95	Trivandrum	T
45004	22.32	114.17	Hong Kong	T S D C
47401	45.42	141.68	Wakkanai	T S D C
47412	43.05	141.33	Sapporo	T S D C
47646	36.05	140.13	Tateno	T S
47827	31.63	130.58	Kagoshima	T S D C
47991	24.30	153.97	Minamitorishima	S
48455	13.73	100.50	Bangkok	T
48698	1.37	103.98	Singapore	T S
61052	13.48	2.17	Niamey	T
61641	14.73	−17.50	Dakar	T
67964	−20.15	28.62	Bulawayo	T
68588	−29.96	30.95	Durban	T
68816	−33.96	18.60	Capetown	T
68906	−40.35	−9.88	Gough Island	T
68994	−46.88	37.87	Marion Island	T
70026	71.30	−156.78	Barrow	T
70308	57.13	−170.26	St. Paul Island	T S D C
70398	55.03	−131.56	Annette	T S
71072	76.23	−119.33	Mould Bay	T
71082	82.50	−62.33	Alert	T
71836	51.27	−80.65	Moosonee	T S D C
71934	60.03	−111.95	Ft. Smith	T S
72250	25.92	−97.47	Brownsville	T S D C
72293	33.02	−118.58	San Diego	T S D
72403	38.98	−77.47	Sterling	T S D
72764	46.77	−100.75	Bismarck	T S
78016	32.27	−64.85	Bermuda	T
78526	18.45	−66.10	San Juan	T S D C
85442	−23.41	−70.47	Antofagasta	T
85543	−32.78	−71.52	Quintero	T S D C
85799	−41.43	−73.12	Puerto Montt	T
89611	−66.26	110.57	Casey	T
91165	21.98	−159.35	Lihue	T S D C
91217	13.55	144.83	Guam	T D C
91245	19.30	166.62	Wake Island	T D C
91285	19.73	−155.06	Hilo	T S D C
91334	7.45	151.83	Truk	T S
91376	7.10	171.40	Majuro	T S D C
91408	7.33	134.48	Koror	T S D C
91517	−9.43	199.95	Honiara	
91592	−22.26	166.45	Noumea	T S
91680	−17.75	177.45	Nandi	T S D C
91938	−17.55	−149.61	Papeete	T
93844	−46.41	168.33	Invercargill	T
93986	−43.95	−176.56	Chatham Island	T S D C
94120	−12.43	130.87	Darwin	T
94294	−19.25	146.77	Townsville	T
94312	−20.38	118.62	Port Hedland	T
94461	−25.03	128.30	Giles	T
94510	−26.41	146.27	Charleville	T
94610	−31.93	115.97	Perth	T
94672	−34.95	138.53	Adelaide	T
94975	−42.83	147.50	Hobart	T
94995	−31.53	159.07	Lord Howe Island	T
94996	−29.05	167.93	Norfolk Island	T
94998	−54.50	158.95	Macquarie	T
96996	−12.18	96.83	Cocos Island	T

REFERENCES

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

——, 1991: Changes in tropospheric and stratospheric global temperatures, 1958–1988. *Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations,* M. E. Schlesinger, Ed., Elsevier, 231–247.

——, 1996: Stratospheric temperatures after volcanic eruptions. *The Mount Pinatubo Eruption: Effects on the Atmosphere and Climate,* G. Fiocco, D. Fua, and G. Visconti, Eds., NATO ASI Series I: Global Environmental Change, Vol. 42, Springer, 83–93.

——, and J. Korshover, 1975: Estimate of the global change in tropospheric temperature between 1958 and 1973. *Mon. Wea. Rev.,* **103,** 1007–1012.

——, and ——, 1977: Estimate of global change in temperature, surface to 100 mb, between 1958 and 1975. *Mon. Wea. Rev.,* **105,** 375–385.

——, and ——, 1978a: Global temperature variation, surface-100 mb: An update into 1977. *Mon. Wea. Rev.,* **106,** 755–770.

——, and ——, 1978b: Estimate of global temperature variations in the 100-30 mb layer between 1958 and 1977. *Mon. Wea. Rev.,* **106,** 1423–1432.

——, and ——, 1983: Global temperature variations in the troposphere and stratosphere, 1958–1982. *Mon. Wea. Rev.,* **111,** 901–921.

Bickel, P. J., and K. A. Doksum, 1977: *Mathematical Statistics.* Holden-Day, 492 pp.

Chanin, M.-L., and Coauthors, 1999: Trends in stratospheric temperatures. Scientific Assessment of Stratospheric Ozone Change: 1998. WMO Global Ozone Research and Monitoring Project Report 44, World Meteorological Organization, 5.1–5.59 pp.

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

Conover, W. J., 1980: *Practical Nonparametric Statistics.* Wiley, 493 pp.

Eskridge, R. E., O. A. Alduchov, I. V. Chernykh, Z. Panmao, A. C. Polansky, and S. R. Doty, 1995: A Comprehensive Aerological Research Data Set (CARDS): Rough and systematic errors. *Bull. Amer. Meteor. Soc.,* **76,** 1759–1775.

Finger, F. G., R. M. Nagatani, M. E. Gelman, C. S. Long, and A. J. Miller, 1995: Consistency between variations of ozone and temperature in the stratosphere. *Geophys. Res. Lett.,* **22,** 3477–3480.

Frederick, J. E., and A. R. Douglass, 1983: Atmospheric temperatures near the tropical tropopause: Temporal variations, zonal asymmetry and implications for stratospheric water vapor. *Mon. Wea. Rev.,* **111,** 1397–1403.

Gaffen, D. J., 1993: Historical changes in radiosonde instruments and practices. WMO/TD-No. 541, Instruments and Observing Methods Report 50, World Meteorological Organization, 123 pp.

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

——, 1996: A digitized metadata set of global upper-air station histories. NOAA Tech. Memo. ERL-ARL 211, Silver Spring, MD, 38 pp. [Available from NOAA Air Resources Laboratory, 1315 East West Highway, Silver Spring, MD 20910.]

——, 1998: Falling satellites, rising temperatures? *Nature,* **394,** 615–616.

Habermann, R. E., 1987: Man-made changes in seismicity rates. *Bull. Seis. Soc. Amer.,* **77,** 141–159.

Hansen, J. M., and Coauthors, 1997: Forcings and chaos in interannual to decadal climate change. *J. Geophys. Res.,* **102,** 25 679–25 720.

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

Ivanov, A., A. Kats, S. Kurnosenko, N. Nash, and N. Zaitseva, 1991: WMO International Radiosonde Intercomparison Phase III (Dzhambul, USSR 1989) final report. WMO/TD-No. 451, World Meteorological Organization, Instruments and Observing Methods Report 40, 135 pp.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Labitzke, K., and H. van Loon, 1995: A note on the distribution of trends below 10 hPa: The extratropical Northern Hemisphere. *J. Meteor. Soc. Japan,* **73,** 883–889.

Lanzante, J. R., 1996: Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.,* **16,** 1197–1226.

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

McMillin, L. M., M. E. Gelman, A. Sanyal, and M. Sylva, 1988: A method for use of satellite retrievals as a transfer standard to determine systematic radiosonde errors. *Mon. Wea. Rev.,* **116,** 1091–1102.

Miller, A. J., R. M. Nagatani, G. C Tiao, X. F. Niu, G. C. Reinsel, D. Wuebbles, and K. Grant, 1992: Comparisons of observed ozone and temperature trends in the lower stratosphere. *Geophys. Res. Lett.,* **19,** 929–932.

Nash, J., and F. J. Schmidlin, 1987: WMO International Radiosonde Intercomparison (U.K., 1984, U.S.A., 1985) final report. WMO/TD-No. 195, World Meteorological Organization, Instruments and Observing Methods Report 30, 103 pp.

Oort, A. H., and H. Liu, 1993: Upper-air temperature trends over the globe, 1956–1989. *J. Climate,* **6,** 292–307.

Parker, D. E., 1985: On the detection of temperature changes induced by increasing atmospheric carbon dioxide. *Quart. J. Roy. Meteor. Soc.,* **111,** 587–601.

——, and D. I. Cox, 1995: Towards a consistent global climatological rawinsonde data-base. *Int. J. Climatol.* **15,** 473–496.

——, M. Gordon, D. P. N. Cullum, D. M. H. Sexton, C. K. Folland, and N. Rayner, 1997: A new global gridded radiosonde temperature data base and recent temperature trends. *Geophys. Res. Lett.,* **24,** 1499–1502.

Pawson, S., and B. Naujokat, 1997: Trends in daily wintertime temperatures in the northern stratosphere. *Geophys. Res. Lett.,* **24,** 575–578.

——, K. Labitzke, and S. Leder, 1998: Stepwise changes in stratospheric temperature. *Geophys. Res. Lett.,* **25,** 2157–2160.

Peterson, T. C., and Coauthors, 1998a: Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. Climatol.,* **18,** 1493–1517.

——, T. R. Karl, P. F. Jamason, R. Knight, and D. R. Easterling, 1998b: First difference method: Maximizing station density for the calculation of long-term global temperature change. *J. Geophys. Res.,* **103,** 25 967–25 974.

Richner, H., and P. D. Phillips, 1982: The radiosonde intercomparison SONDEX Spring 1981, Payerne. *Pure Appl. Geophys.,* **120,** 852–1198.

Santer, B. D., T. M. L. Wigley, T. P. Barnett, and E. Anyamba, 1996: Detection of climate change and attribution of causes. *Climate Change 1995: The Science of Climate Change,* J. T. Houghton et al., Eds., Cambridge University Press, 407–443.

——, J. J. Hnilo, T. M. L. Wigley, J. S. Boyle, C. Doutriaux, M. Fiorino, D. E. Parker, and K. E. Taylor, 1999: Uncertainties in observationally based estimates of temperature change in the free atmosphere. *J. Geophys. Res.,* **104,** 6305–6333.

——, and Coauthors, 2000: Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. *J. Geophys. Res.,* in press.

Scherhag, R., 1960: Die explosionsartigen Stratosphärenenerwärmungen des Spätwinters 1951/1952. *Ber. Dtsch. Wetterdienstes,* **6,** 51–63.

Spencer, R. W., and J. R. Christy, 1990: Precise monitoring of global temperature trends from satellites. *Science,* **247,** 1558–1562.

Talbot, J. E., 1972: Radiation influences on a white-coated thermistor temperature sensor in a radiosonde. *Austr. Meteor. Mag.,* **20,** 22–33.

Tett, F. B., J. F. B. Mitchell, D. E. Parker, and M. R. Allen, 1996: Human influence on the atmospheric vertical temperature structure: Detection and observations. *Science,* **274,** 1170–1173.

Trenberth, K., 1990: Recent observed interdecadal climate changes in the Northern Hemisphere. *Bull. Amer. Meteor. Soc.,* **71,** 988–993.

U.S. Dept. of Commerce, 1967: Monthly climatic data for the world—1967 supplement: Computation of climate temperature values. USCOMM-ESSA-Asheville, NC, 11/14/67-2000, 14 pp. [Available from NOAA National Climatic Data Center, 151 Patton Avenue, Asheville, NC 28801.]

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

Wentz, F. J., and M. Schabel, 1998: Effects of orbital decay on satellite-derived lower-tropospheric temperature trends. *Nature,* **394,** 661–664.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.

WMO, 1996: GCOS plans progress. *World Climate News,* **9,** 9–10.

Zhai, P., and R. E. Eskridge, 1996: Analyses of inhomogeneities in radiosonde temperature and humidity time series. *J. Climate,* **9,** 676–705.