

## Temporal Homogenization of Monthly Radiosonde Temperature Data. Part I: Methodology

JOHN R. LANZANTE AND STEPHEN A. KLEIN

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

DIAN J. SEIDEL

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

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

### ABSTRACT

Historical changes in instrumentation and recording practices have severely compromised the temporal homogeneity of radiosonde data, a crucial issue for the determination of long-term trends. Methods developed to deal with these homogeneity problems have been applied to a near-globally distributed network of 87 stations using monthly temperature data at mandatory pressure levels, covering the period 1948–97. The homogenization process begins with the identification of artificial discontinuities through visual examination of graphical and textual materials, including temperature time series, transformations of the temperature data, and independent indicators of climate variability, as well as ancillary information such as station history metadata. To ameliorate each problem encountered, a modification was applied in the form of data adjustment or data deletion. A companion paper (Part II) reports on various analyses, particularly trend related, based on the modified data resulting from the method presented here.

Application of the procedures to the 87-station network revealed a number of systematic problems. The effects of the 1957 global 3-h shift of standard observation times (from 0300/1500 to 0000/1200 UTC) are seen at many stations, especially near the surface and in the stratosphere. Temperatures from Australian and former Soviet stations have been plagued by numerous serious problems throughout their history. Some stations, especially Soviet ones up until ~1970, show a tendency for episodic drops in temperature that produce spurious downward trends. Stations from Africa and neighboring regions are found to be the most problematic; in some cases even the character of the interannual variability is unreliable. It is also found that temporal variations in observation time can lead to inhomogeneities as serious as the worst instrument-related problems.

### 1. Introduction

Change in the vertical profile of atmospheric temperature is an important diagnostic for climate change detection and attribution (Santer et al. 1996; Tett et al. 1996; Allen and Tett 1999; Hill et al. 2001). Results from general circulation model (GCM) climate change simulations [Hansen et al. 1997; Bengtsson et al. 1999; the National Research Council (NRC) 2000; Santer et al. 2000; Ramaswamy et al. 2001] suggest that the vertical structure of the temperature response, from the surface up to the stratosphere, depends critically on the particular forcings that are included (e.g., increases in well-mixed green-house gases, stratospheric ozone loss, stratospheric water vapor increases, volcanic aerosols, and solar radiation changes). Unfortunately, highly reliable estimates of long-term global temperature trends

at different altitudes are not possible because existing observational datasets do not meet the standards for long-term monitoring of the climate system, articulated by Karl et al. (1995) and promulgated by the National Research Council (NRC 1999). Those tenets of climate observing systems set forth system design and maintenance principles, operating procedures, and data and metadata analysis and archival policies that would vastly improve the long-term continuity and quality of climate datasets. Because they were initiated primarily to support weather forecasting rather than climate monitoring, existing upper-air temperature observing systems, whether based on satellite, lidar, rocketsonde, or radiosonde observations, fall far short of these goals, introducing considerable uncertainty in trend estimation.

The degree of consistency in trends computed from different temperature datasets yields insights as to the overall uncertainty of the estimates (e.g., Santer et al. 1999; Hurrell et al. 2000; Ramaswamy et al. 2001). However, adequate explanation for the discrepancies that have been found is lacking at this time. Radiosonde

---

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

data offer a potential means of reconciling some of these differences (e.g., Brown et al. 2000; Gaffen et al. 2000a), particularly because of their superiority, as compared to other products, in the combination of length of record and vertical resolution.

Substantial effort has been devoted to developing improved global temperature datasets from surface observations (Hansen et al. 1999; Jones et al. 1999; Peterson and Vose 1997; Vinnikov et al. 1990; Jones et al. 2001; Folland et al. 2001); rocketsondes (Keckhut et al. 1999); and the microwave sounding unit (MSU) on the National Oceanic and Atmospheric Administration (NOAA) polar-orbiting satellites (Christy et al. 2000). These efforts focus on adjusting, or homogenizing, the data to remove both gradual and abrupt artificial temperature changes that might result from station moves, instrument and procedural changes, and urbanization (in the case of surface observations), or changing orbital configuration, instrument drift, and differences between instruments on different platforms in the case of satellites.

Comparable efforts to create more temporally homogeneous radiosonde temperature datasets have only been attempted in the last few years: previously, radiosonde temperature data were used without adjustment to estimate trends (e.g., Angell and Korshover 1975; Oort and Liu 1993). Several homogenization methods were presented at an October 2000 workshop and are described and intercompared in a meeting report (Free et al. 2002). All but one method are still in the developmental stage and have not yet been evaluated or used to create time series for trend analysis. The single exception is the United Kingdom Meteorological Office (UKMO) method (Parker et al. 1997), which is based on comparison of monthly mean radiosonde and MSU temperature data in conjunction with station history metadata. Therefore, it is limited to the period beginning in 1979, the first year of MSU data, and to stations for which such metadata are available. Furthermore, the radiosonde data adjustments are potentially affected by any remaining temporal inhomogeneities in MSU data and are limited by the much coarser MSU vertical resolution. Most importantly, the resulting radiosonde time series are no longer independent of MSU.

The diversity of approaches currently under investigation for homogenizing radiosonde data (Free et al. 2002) is evidence of the more complex issues that must be addressed for radiosonde data than for satellite or surface data. For satellite data only one or two instruments observe the globe at a time and are replaced, often with overlap, with new versions of the same instrument. Surface observations are made with permanently installed instruments. By contrast, radiosondes are expendable instruments, thus the global data archive consists of tens of millions of soundings, each made with a different instrument; expendability facilitates relatively easy and frequent instrument changes. The radiosonde network is operated at a national level, and instrument types and observing practices vary from

country to country, as well as within national networks. Some of this diversity is documented in station history metadata (Gaffen 1993, 1996), but these records are neither complete nor fully reliable. Furthermore, radiosonde data are temperature profiles that must be homogenized with vertical structure, so it is inappropriate to use the same adjustment at all levels.

A myriad of approaches, both objective and subjective, have been used to deal with inhomogeneous climate data (Peterson et al. 1998). Our previous attempts to develop objective schemes to homogenize radiosonde data (Gaffen et al. 2000b) did not yield useful time series but did suggest that completely objective methods are not well suited to this particular problem. The statistical methods employed to identify abrupt shifts in mean temperature could not distinguish between real and artificial change points (i.e., discontinuities), and resulted in adjustments that removed practically all of the original trends. When these methods were used in combination with station history information, the number of change points that could be adjusted fell dramatically, leaving obviously artificial change points unadjusted. Based on this experience, and our desire to develop a homogenized radiosonde dataset that is independent of satellite data, we have developed a procedure that applies critical reasoning, with a subjective element, to identify artificial change points using a more diverse and largely independent set of objective tools.

This paper presents our new radiosonde temperature homogenization procedure. Section 2 describes the data as well as the broad statistical approach utilized. Section 3 outlines the entire procedure used in attempting to render the data more temporally homogeneous. Section 4 describes specific tools utilized in identifying the features responsible for inhomogeneity. Section 5 consists of case studies, from a selection of the station data employed here, exemplifying our procedure as well as some outstanding problems found throughout our network of stations. A summary and concluding remarks are given in section 6. A companion paper (Lanzante et al. 2003), hereafter referred to as Part II, presents and evaluates the results of applying the homogenization procedure to data from our near-globally distributed network of 87 stations.

## 2. Data and statistical considerations

The radiosonde temperature data used in this study are from the Comprehensive Aerological Reference Data Set (CARDS) Project (Eskridge et al. 1995) and were obtained in the form of station soundings for the period 1948–97. As indicated in appendix A, not all stations have usable data for the full time period. We eliminated values flagged by CARDS as suspect or erroneous, using CARDS-provided replacement or corrected values when available. Monthly means were computed from the soundings, with the requirement of at least 16 valid values per month. Given sufficient num-

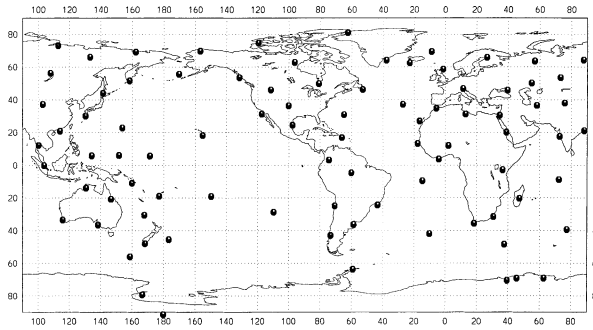


FIG. 1. Network of 87 radiosonde stations (filled circles).

bers of observations, separate 0000 and 1200 UTC monthly means were computed. For a small number of stations, where 0000 or 1200 UTC data were insufficient, means were computed after pooling data from all available observation times and are referred to as 9900 UTC means. This choice was based on the desire to include some remote areas where adherence to 0000 and 1200 UTC observation times creates voids in spatial coverage where stations exist.

A systematic global change in observation time from 0300/1500 to 0000/1200 UTC occurred around 1957, although the exact timing varies among countries/stations from 1957 to early 1958. Although we refer to time series as 0000/1200 UTC, this change is implicit in our time series. In some cases this observation time shift introduces an inhomogeneity that is dealt with in the same fashion as an instrumental change.

This work utilizes data from 16 mandatory levels: the surface, and the 1000-, 850-, 700-, 500-, 400-, 300-, 250-, 200-, 150-, 100-, 70-, 50-, 30-, 20-, and 10-hPa levels; note that standard practices dictate that surface values be measured using surface instrumentation, at a nearby instrument shelter, rather than radiosonde equipment. Because our approach is tedious and requires typically 5–10 person-hours per station we selected a minimal set of stations that would reasonably sample the globe for as long a time span as possible. An 87-station subset (appendix A) of the Global Climate Observing System (GCOS) Baseline Upper-Air Network (WMO 1994) is employed, which includes 48 stations from Angell's (1988) 63-station network.

The various calculations performed utilize nonparametric statistical methods (Lanzante 1996, 1998) that provide alternatives to common operations such as computing means, standard deviations, correlation coefficients, etc. Nonparametric techniques are particularly advantageous in the analysis of "messy" data because they greatly diminish the impact of outliers without having to explicitly identify the offending values, and since they make no assumptions regarding the underlying statistical distribution (e.g., Gaussian). Most noteworthy for this paper is our use of the biweight mean instead of the traditional arithmetic mean as well as the scheme

for detection of multiple statistical change points in a time series, previously employed by Gaffen et al. (2000b).

### 3. Data homogenization procedure

#### a. Overview

The procedure consists of two parts: 1) identification of artificial change points and other maladies in temperature time series, and 2) modification of the time series in an attempt to remove a major portion of the artificial effects. Furthermore, the first part, which is accomplished through the examination of a variety of graphical and textual information, consists of two steps. First, two of us (Lanzante and Klein) examined the materials as individuals to form preliminary opinions. Next, we met and discussed each case until we were able to come to agreement as to the actions needed. Our third member (Seidel) was involved in the group discussions for a subset of the stations, and served as a tiebreaker as needed.

An attempt has been made to apply, in a consistent manner, a set of objective rules or operating principles that have been developed a priori as well as a posteriori. For example, one a priori principle is to consider only the largest change points because Gaffen et al. (2000b) showed that trends depend crucially on the largest change points whose impacts overwhelm smaller discontinuities. This is also motivated by pragmatism, acknowledging that radiosonde data adjustment is in its infancy (Free et al. 2002), since weaker discontinuities are less easily distinguished from natural variability. An a posteriori principle is that pronounced vertical inconsistencies are often indicators of artificial changes. An example is when time series at nearby levels, which normally covary strongly, differ markedly regarding the presence or absence of a discontinuity. More principles will be illustrated through explanation of the scheme and by way of example.

The result of our new approach is much higher confidence, relative to our prior attempts, in identification of artificial discontinuities. In Part II, a comparison with independent satellite data from the MSU demonstrates an overall increase in consistency between the two datasets as a result of our homogenization. Beyond any improvements made through modification, quality has been added by documenting data limitations and strengths in records that may be of value to prospective users. Furthermore, in examining sensitivities to the procedures used to alter the data, some measure of confidence can be attached to trend calculations reported in Part II.

#### b. The nature of data modification decisions

Assignment of artificial change points is specific to station, level, and observation hour. While our original

intent was to assign changepoints that would apply to all levels for a station-observation time, it became readily apparent that this would be an unrealistic imposition. The effects of artificial discontinuities can be isolated in the vertical (even limited to a single level) or discontinuous in the vertical (e.g., a cluster of levels with substantial discontinuities, with adjacent levels having only a trivial effect). While this is different from our expectations based on Luers and Eskridge (1998), their study is more theoretical in nature than is ours.

Although our original intent was merely to adjust for the effects of artificial steplike changes, it became obvious that some maladies could not be handled in such a fashion. As a result, deletion of selected portions of individual time series was added as one of the decisions made. As shown in Part II, overall, the impact of data deletions is substantial and of comparable magnitude to adjustment of artificial changepoints. In general, there are three situations warranting data deletion. One situation is excessive uncertainty regarding data quality: long gaps in time series that preclude assessment of temporal continuity, or periods of erratic data characterized by unrealistically large month-to-month variance. Another justification for deletion is the inability to make a desired adjustment due to problems in the proximity of a changepoint: an insufficient amount of data prior to or after the changepoint, or the presence of a natural changepoint (due to a volcano or other causes) in which case our methods do not always produce satisfactory results. Finally, some artificial features such as drifts or low-frequency meanders are not well characterized by changepoints because they represent gradual rather than steplike changes. Assignment of data deletions, like that of changepoints, is specific to station, level, and observation hour.

#### *c. Classification scheme and documentation of data modification decisions*

Once an artificial changepoint has been assigned to a specific location within a time series, a categorical measure of confidence is attached. Those changepoints identified with a higher degree of confidence are designated as conservative (CON), and those for which we have less confidence are designated as liberal (LIB). In the case of CON, the changepoint is associated with either of the following: 1) a station history metadata event (i.e., some documented change in instruments or practices), or 2) a change of such large magnitude that in our judgement it is beyond the realm of natural variability. If the changepoint does not meet one of these two criteria its designation defaults to LIB. Some leeway is allowed in interpreting the dates for condition 1 since our assignment of a changepoint date is inexact and since the station history dates can be approximate or uncertain (Gaffen 1996). Generally a year or so is allowed but this depends on our confidence in the me-

tadata, which can be shaped by quality indicators included with the metadata as well as our experiences with other stations from the same country. It is worth noting that application of condition 2 is not limited to the raw time series at a single level; derived time series also examined include smoothed series (low-pass filtered), difference series (0000/1200 UTC difference), time series at other levels at the same station (to judge the nature of the vertical structure), and, in a few instances, time series from other stations in the region.

All data modifications are documented in station-specific text files in a machine-readable format. Each file includes the level-specific time periods of data deletions and dates of changepoints, along with a commentary explaining our rationale. Systematic and detailed documentation of all of the data modification decisions has two important benefits: (i) creation of derived metadata that can be used by other researchers and (ii) consistency and rationality to the procedure, since written justification of all of our actions is required. These decision files are available by request from the corresponding author.

#### *d. Adjustment procedures and scenarios*

Ideally the amount of adjustment required would be the difference between the mean values of the data segments before and after a changepoint. In reality, uncertainty arises because there is no guarantee that this difference is due solely to artificial effects. For example, consider the impact of an instrument change at a station in the tropical Pacific that occurs at the time of a phase change of the El Niño–Southern Oscillation (ENSO). Depending on the signs and magnitudes of the natural and artificial signals, adjustment using the difference of segment means could erroneously remove the natural signal or fail to remove the artificial component. One way to try to overcome this problem is to make the segments long enough so that the shorter-term natural signal averages out. Of course, there is no way to ensure this; furthermore, the length of the segments is not always easily controlled. When a time series has multiple changepoints, the segments used for adjustment cannot extend past the nearest neighboring changepoint on each side; also, segment length may be constrained by the beginning or end of the usable data record. Finally, it is worth noting that the above concerns apply to all natural signals, including very low frequency signals due to external forcings or anthropogenic causes. However, since these are typically of considerably smaller amplitude over the segments than the higher-frequency signals, such as ENSO, the latter are a greater source of uncertainty.

In order to deal with the uncertainties of adjustment, two fundamentally different approaches have been used to enable sensitivity testing. In addition to the simple adjustment procedure described above, a more complex

procedure was used, inspired by detailed inspection of station time series for the most confidently identified changepoints. It was frequently observed that while certain levels were very strongly influenced by a particular artificial discontinuity, other nearby levels at the same station appeared to be unaffected. Furthermore, the interannual variations of these nearby levels were otherwise well correlated with the affected levels. We reasoned that in the absence of the artificial effects, the shape of the affected levels would resemble that of the unaffected nearby levels. Altering the affected levels so that their low-frequency behavior most clearly matches the nearby unaffected levels yields the potential to retain the natural component of a jump across a changepoint; the simple method does not have this ability. In this procedure we not only include time series from other levels, but from the other observation time, 0000 or 1200 UTC, if present. Thus, for the complex method, the daytime sounding data can be adjusted using nighttime soundings, which in general are less affected by instrumental changes.

The simple approach or “nonreference level” scheme computes the adjustment value as the difference of the means of the two segments adjacent to a changepoint. The complex approach or “reference level” scheme uses one or more levels that are well correlated with the affected level as a reference series and proceeds iteratively, at each step adjusting an affected level until it resembles as closely as possible its reference levels. Reference level adjustment is preferred because it has the ability to retain the natural vertical structure. The interested reader is referred to appendix B for general information as well as more details on the reference level adjustment scheme.

It should be noted that our adjustment schemes make *relative* adjustments in that they seek to eliminate a discontinuity between two adjacent segments. Adjustments, which additionally seek to adjust the mean of the resulting time series to some standard, for example, to some common instrument type, are far beyond our present capabilities. As a result of this limitation we operate on and produce time series in the form of monthly *anomalies*. While this is a handicap for some applications, for others, such as trend estimation, it is inconsequential.

Besides deriving an improved dataset, one of the broader goals of this study is to examine the sensitivities of results to the procedures. To this end, five data scenarios were created differing in the degree and manner of data alteration. The scenarios are distinguished by the level of confidence in changepoint identification and the method of data alteration. The first four scenarios (Table 1) represent a progressive increase in data modification. For UNADJ no data modifications are made, for DEL only data deletions apply, and for CON (LIBCON) conservative (both liberal and conservative) changepoints are adjusted using the reference level scheme. The NONREF scenario is the same as

TABLE 1. Definitions of data scenarios (columns). Rows define modifications that may be made to the data, with an “X” indicating applicability to a particular scenario.

Data modification	Scenario				
	UN-ADJ	DEL	CON	LIB-CON	NON-REF
Data deletions		X	X	X	X
Conservative changepoints			X	X	X
Liberal changepoints				X	X
Reference level adjustment			X	X	
Nonreference level adjustment					X

LIBCON except that simple nonreference level adjustment is used instead.

#### 4. Tools for changepoint identification

All data decisions are based on examination of numerous materials (12 graphical and 5 textual products for each station-observation time) with the intent of separating true from artificial signals. The use of multiple, independent tools is a crucial factor that often bolsters confidence considerably because weaknesses or uncertainties in one indicator can be overridden by another indicator. The graphs display temperature as well as derived time series, in both raw and low-pass-filtered form. They also include natural indicators such as the Southern Oscillation index (SOI) and the times of major volcanic eruptions, station history metadata events, and changepoints derived from a purely statistical time series analysis method. Typical plots have stacked time series from ~10 levels, usually with multiple series per level. The text files include various inventories (counts as a function of time), metadata, and derived statistics.

The tools are introduced in sections 4a–e below, in order of importance as we perceive it over the entire station network, although in any particular instance importance may vary considerably. After this, the tools and operating principles for data decisions are illustrated through examples. To conserve space and enhance clarity, this discussion focuses on the major tools and display is limited to severely edited versions of the graphs we have used.

##### a. Diurnal (0000/1200 UTC) differences

A major source of the difference in measuring capabilities between two different radiosonde temperature sensors is due to the influence of solar radiation, since inadequate shielding or ventilation can lead to spuriously high readings (Zhai and Eskridge 1996); this is particularly true in the stratosphere due to the lower density of air. A useful indicator in this regard is the time series of the difference between 0000 and 1200 UTC data. Differencing largely eliminates the real climate signal, which is common to both, leaving mostly the time-varying relative bias. Ideally the difference se-

ries should be white noise punctuated by discontinuities at times of instrument change. Although reality can sometimes be more complex (e.g., drifts or low-frequency meanders) the idealization is true frequently enough to make this by far our most powerful tool. One of our operating principles is that any irregularity that rises well above the natural noise in the difference series is a virtual guarantee of a problem. Because neither 0000 nor 1200 UTC data are known to be “correct,” irregularities in the difference series do not indicate whether one or both are “at fault.” Frequently other tools can be used to attribute the problem to one or both observation times. In most cases either just the daytime time series is corrupted, or it is corrupted more, in accordance with expectations.

For polar stations or those near 90°E and W, the diurnal difference has little value due to the limited difference in intensity of solar radiation between 0000 and 1200 UTC; both observations are “daytime” during summer and “nighttime” during winter. The complications of seasonal variations are not directly addressed but are reduced by examining a low-pass-filtered version of the difference series in addition to the unfiltered version. The possibility of natural variations in day–night differences is not of great concern because natural variations would be either of short duration, associated with an event such as volcanic eruption, or driftlike if associated with climate change. A trend or slow fluctuation would have little impact on changepoint identification.

#### b. Vertical structure/coherence

While the vertical structure of natural phenomena is constrained by physical laws, artificial variations are virtually unconstrained. Visually, characteristic natural vertical structures are very striking: low-frequency variations are very coherent throughout the free troposphere and lower stratosphere, with a rapid disconnect in approaching and crossing the tropopause. Other features such as the character of ENSO or stratospheric (quasi-biennial oscillation) QBO-related variations, the signature of volcanos, the ~1976–77 climate regime shift, rapid drops in stratospheric temperature during the last two decades, etc., are phenomena seen at numerous stations and believed to be real. Features that do not follow these known patterns are viewed with suspicion.

#### c. Station history metadata

Station history metadata provides information on radiosonde manufacturer, model, sensor type, station relocations, ground and computer equipment, data reduction algorithms, procedures, etc. The metadata employed (Gaffen 1996) were derived from a number of different sources. Metadata events are of two types: dynamic, indicating a *change* of some sort occurred at a particular time; and static, indicating that a particular

instrument or procedure *was in use* at a particular time. Static events are less useful because it is only possible to infer that a change took place at some indeterminate time between events.

It is important to keep in mind that instruments and practices can vary widely not only among different countries, but sometimes among stations within a country. Also, the reliability and completeness of the metadata can vary greatly. In some instances, information from different sources can be contradictory or ambiguous; dates and instrument characteristics can be incorrect or vague. Furthermore, not every instrumental or procedural change results in an artificial change of any practical importance. However, despite these shortcomings, metadata can be a very powerful tool, particularly when one or more other indicators suggest a change at the same time.

#### d. Statistical changepoints

A statistical procedure to objectively identify multiple changepoints in a time series has been used (Lanzante 1996, 1998). For each station and level, results are displayed graphically in the form of step function curves, that is, line segments that join changepoints. Statistical changepoint identification is very powerful because it can identify discontinuities in noisy time series. However, there is a certain error rate and the procedure does not distinguish between artificial and natural changepoints. Natural phenomena such as ENSO phase transitions, the climate regime shift around 1976–77, the stratospheric response to volcanic eruptions, etc., are often characterized by approximately discontinuous temperature change. Statistical changepoints are most useful in conjunction with other indicators that help pinpoint the date. They are also useful when examining the vertical profile of time series at a station; when changepoints line up in the vertical for a number of consecutive levels it signals that closer examination is warranted.

#### e. Other indicators

A number of minor tools can occasionally have moderate to considerable value:

- 1) Predicted temperature series based on regression of temperature on winds and SOI. Since winds are measured independently of temperature they can potentially confirm or contradict temperature discontinuities as being natural. Although occasionally useful, the strength of the statistical relationship is generally too weak to instill great confidence.
- 2) The SOI time series.
- 3) Dates of major volcanic eruptions.
- 4) Time series of estimated surface station elevation derived hydrostatically from surface and low-level radiosonde parameters (Collins and Gandin 1990).

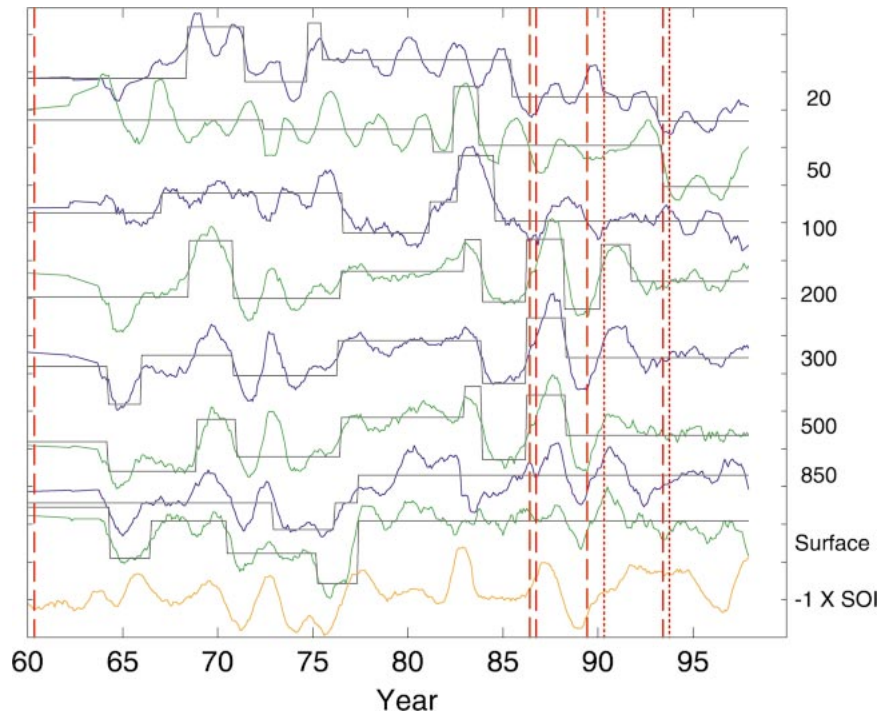


FIG. 2. Smoothed time series of 0000 UTC temperature at Majuro (91376) for every other available level from the stratosphere (20 hPa) to the surface; smoothing is based on a 15-point running median. The tick interval on the ordinate is one nondimensional unit. For clarity, temperature time series curves have been standardized to unit variance (i.e., are nondimensional) and alternate between blue and green. Black step function curves connect statistical changepoints. The orange curve depicts the smoothed inverted SOI time series. Dynamic (static) station history metadata events are denoted by dotted (dashed) red vertical lines.

Occasionally, comparison of the reported versus derived elevations points strongly toward an undocumented station move, which suggests possible undocumented instrument changes.

- 5) Time series of temperature from stations in different countries in the same region. These are compared, in an attempt to ascertain whether a particular feature is natural; unfortunately, the typical distance between stations within our network limits the between-station correlation and, thus, the applicability of this tool.
- 6) A listing of the number of observations per month by hour (0000–2300 UTC) as a function of time. These are vital in a few instances, particularly for the 9900 UTC stations, for associating temperature discontinuities with systematic changes in time of observation.
- 7) Counts of numbers of observations per month as a function of time and by level; these aid in finding sampling biases or less reliable time periods.

## 5. Case studies

### a. Majuro

The first example (Fig. 2) is Majuro, a station for which we did not assign any changepoints, and serves

to illustrate the graphical setup along with some of the basic tools. Because the ranges of values of temperature time series vary by level, the curves have been standardized to unit variance to make for compact display. For further compactness, only every other level is presented, with colors alternating between blue and green; the graphs used in practice contain the full vertical resolution.

At first glance, there appear to be a number of possible inhomogeneities, but the convergence of evidence suggested that they are all manifestations of natural climate variability. Since this station lies in the deep tropical western Pacific, tropospheric temperature variations correlate negatively with the SOI, which accordingly has been plotted in inverse form. Particularly above the surface, temperature variations associated with major SOI fluctuations are coherent through a deep layer in the troposphere, until damping in the vicinity of the tropopause ( $\sim 100$  hPa). The well-known climate regime shift  $\sim 1976$ – $77$  (Trenberth and Hurrell 1994) is evident in both the tropospheric temperatures and the SOI. Vertically coherent variations in the stratosphere are quite different, dominated by the QBO as well as a pronounced downward trend during the last  $\sim 15$  yr. This example illustrates the danger of relying on a purely statistical method of changepoint identification (black step function curves). Many of the ENSO-related

events, the  $\sim 1976$ – $77$  regime shift, and a few QBO-related events are identified synchronously at multiple levels by the statistical changepoint identification method.

The considerable negative trend of stratospheric temperatures, relative to the interannual variability, and the associated abrupt declines in the latter part of the record, are rather typical over our entire network. Prominent downward stratospheric temperature trends, commencing during the 1980s are found at almost every station. Furthermore, a substantial part of the trend can be explained by one or more discontinuous declines, in accord with Pawson et al. (1998), who found such features in both radiosonde and MSU temperature records. We find that the vast majority of stations exhibit a drop in  $\sim 1992$ – $93$ , and other somewhat less dramatic declines are seen during the 1980s, particularly in the Tropics and Southern Hemisphere. We note that drops occur generally  $\sim 2$  yr after major volcanic eruptions (El Chichon in 1982; Pinatubo in 1991).

Although sudden drops in stratospheric temperature during the last two decades are quite widespread, careful examination of the materials for other stations has identified a few cases where such drops are likely artificial. In some of these cases changepoint adjustment can be made, while in others the close proximity of natural and artificial discontinuities has prompted us to delete the end of the time series due to an inability to make a suitable adjustment. In the case of Majuro, there was no compelling evidence favoring an artificial cause. The synchronicity of the  $\sim 1992$  event with many similar events worldwide, the occurrence of the drop prior to the metadata event, the irrelevance of the metadata event that involves changes related to humidity measurements, and the natural-looking vertical structure (i.e., vertically coherent, but confined to the stratosphere) all played a role in our decision. The danger of relying blindly on metadata is also evident in this example given the timing of metadata events near the 1992–93 stratospheric temperature drop as well as the ENSO-related tropospheric temperature rise  $\sim 1990$ .

#### *b. Rostov*

The 11 stations from the former Soviet Union account for nearly a third of all of our stations in the Northern Hemisphere extratropics. Unfortunately, they are beset with a number of serious systematic problems that may impact derived large-scale statistics in this study, and by inference other studies utilizing radiosonde temperatures. In examining time series over our entire network we have noticed a general tendency, demonstrated more quantitatively in Part II, for artificial steplike declines in temperatures during the 1950s–60s. This is most prominent for stations in the former Soviet Union, as well as China, which used Soviet instrumentation during the early years, but occurs at some other stations as well.

One such example is Rostov (Fig. 3a), which has very

large declines in temperature, increasing in magnitude from the surface upward. Not only are these abrupt declines suspiciously large, but they have the same sign and relative magnitudes in both the upper troposphere and stratosphere (not shown); even the magnitude of the jump changes dramatically between some adjacent levels. Two distinct temperature transitions (early 1960s and late 1960s) appear to affect about half of the Soviet stations, although the timing and correspondence with station history events varies. Soviet metadata appears to be especially plagued by internal inconsistencies and ambiguities, as well as a general lack of correspondence with almost certain artificial effects. Although the possibility of a larger-scale signal related to the Arctic Oscillation was considered, this explanation was rejected due to lack of related features in appropriate stations from other countries.

For Rostov we have assigned two times for artificial discontinuities (late 1960 and early 1970) that, as shown by the dynamic metadata events in Fig. 3a, correspond reasonably well with the temperature drops in the upper troposphere. Additional complications associated with the global change in observation time prompted us to delete the data prior to 1957 as well. While we have no way of ascertaining the reason for the systematic temperature drops, we speculate that they are associated with rapid improvements in the early years associated with sonde design related to solar shielding and instrument ventilation. Such improvements would tend to reduce artificial solar warming.

The stratospheric 0000–1200 UTC difference series at Rostov (Fig. 3b) and other Soviet stations are often characterized by low-frequency meanders during the first couple of decades. At Rostov some of these are associated with notches corresponding to the observation time change and instrument transitions noted above. A particularly troubling feature of the stratospheric difference series is the upward drift ( $\sim 1$ – $1.5$  K) from the late 1970s to the mid-1990s, which is interrupted by a downward jump in 1986 associated with a major sonde change. About half of the Soviet stations are affected, in a geographic pattern suggestive of solar radiation effects. The drift is seen most clearly at far western and eastern stations, locations at which 0000 and 1200 UTC better approximate the extremes of day versus night, as well as lower latitudes, where seasonal variations in solar radiation are less extreme. We reject the possibility of natural causes since we have found no such effect at any other locales. Examination of separate 0000 and 1200 UTC time series (not shown) led us to conclude that the problem is largely or entirely associated with the daytime soundings, which lack the accelerated stratospheric cooling seen worldwide at other stations; therefore, we opted to delete the daytime stratospheric series.

#### *c. Kagoshima and Omsk*

One characteristic that distinguishes artificial from natural temperature discontinuities is a lack of vertical



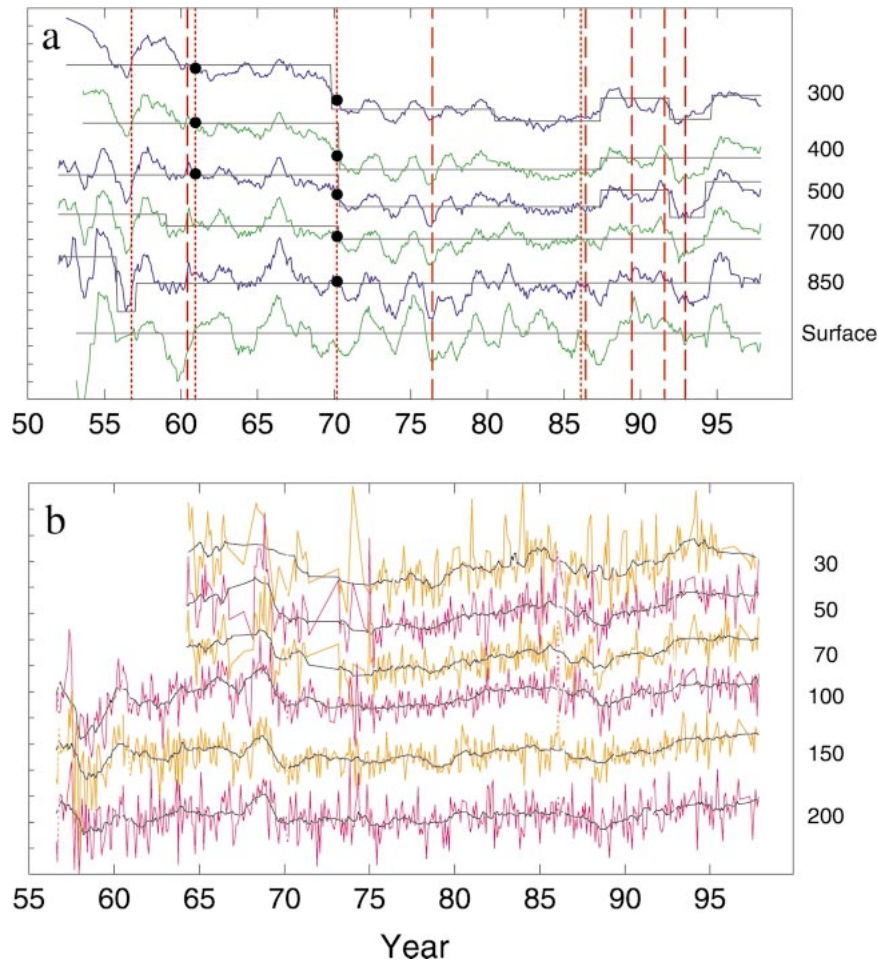


FIG. 3. (a) Smoothed time series of 1200 UTC temperature (K) at Rostov (34731) from 300 hPa to the surface; smoothing is based on a 15-point running median. The tick interval on the ordinate is 1 K. For clarity, temperature time series alternate between blue and green. Black step function curves connect statistical change points. Dynamic (static) station history metadata events are denoted by dotted (dashed) red vertical lines. Black dots denote assigned change points relevant to the discussion. (b) Diurnal temperature (K) difference (1200 – 0000 UTC) time series at Rostov from 30 to 200 hPa. For clarity, difference time series alternate between orange and magenta, with smoothed difference series in black.

coherence between levels that are otherwise highly coherent. Such behavior can be quite striking when only a single level acts out of character, such as for the two examples given here. Caution is advised when making inferences based on vertical coherence near the surface because of boundary layer effects, which can vary considerably from station to station. However, in some instances, such as at Kagoshima, Japan (Fig. 4a), the deviant behavior leaves little doubt of its artificial nature. Temporal variations in temperature at the 700- and 850-hPa levels are very similar, and reasonably similar to that at the surface except for the steplike surface changes near the ends of the record. The bottom of Fig. 4a shows the time series of surface elevation (i.e., the baseline) as estimated from the radiosonde data itself. Changes in the baseline may indicate either real changes in station elevation that may occur when a station relocates, or

changes in vertical structure of the data. The small amplitude annual cycle in the baseline is of no practical consequence; it arises due to the annual cycle in near-surface lapse rate. The discontinuity in 1957 can be explained by the global 3-h shift in observation time. Temperatures based on daytime soundings (Fig. 4a) drop as the observation time shifts from 1200 to 0900 LST, whereas nighttime temperatures (not shown) exhibit no appreciable change. The discontinuity in 1993, which affects both observation times, has no metadata explanation but is obviously artificial. Radiosonde surface observations are not actually measured using the sonde sensor, rather they are taken from the collocated surface observation station (FCM-H3 1997). Thus, changes in instruments and practices used at the surface may be independent of those aloft.

Artificial problems associated with isolated levels are

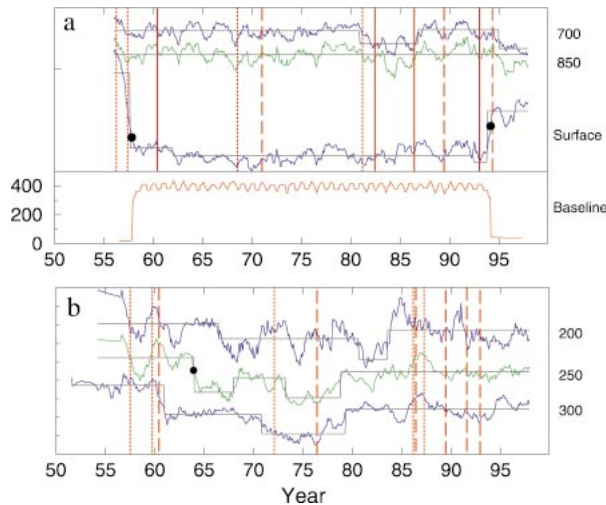


FIG. 4. (a) The top is the same as Fig. 3a except for 0000 UTC temperature at Kagoshima (47827) from 700 hPa to the surface; the bottom is the estimated surface elevation (m). (b) Same as Fig. 3a except for 0000 UTC temperature at Omsk (28698) from 200 to 300 hPa.

not limited to the surface. As an example, several of the Soviet stations (Omsk, Rostov, and Orenburg, Russia) have isolated 250-hPa discontinuities occurring at nearly the same time. In the case of Omsk (Fig. 4b) temperatures in the upper troposphere (250 and 300 hPa) are very well correlated from the early 1970s onward, while at 200 hPa, temperatures are different, instead characteristic of variations at the other stratospheric levels (not shown). The point of interest here is the downward drop of  $\sim 2$  K in 1964 that is limited to the 250-hPa level and is not explained by station history metadata. We occasionally find similar upper-tropospheric jumps limited to one or two levels at non-Soviet stations as well, particularly Australian, and can only speculate as to the cause. Correction factors are sometimes applied at certain specific levels in converting the signals recorded by the sensor to a temperature. It may be that the levels to which corrections are applied change over time, possibly in response to feedback provided by operational weather analysts/forecasters or due to further laboratory study. Finally, it is simply noted that a number of inhomogeneities irrelevant to this section, as they affect multiple levels, occur as follows: 200 (1957), 250 (1957, 1979), and 300 hPa (1960, 1968, 1979).

#### d. Niamey

Niamey, Niger, serves to demonstrate the degradation of temporal continuity resulting from changes in observation time, since more than 10 different mixes of observation times were used over the period of record, because insufficient data were available at either 0000 or 1200 UTC. A very similar history of mixes was found at the other two French colonial stations in our network (Abidjan, Ivory Coast, and Dakar, Senegal). The consequences are quite severe as shown in Fig. 5a, which

displays temperature time series for selected stratospheric and tropospheric levels. Numerous artificial discontinuities were found, with those in boldface associated with a change in the mix of observation times: 1959, 1964, 1969, 1971, 1973, 1976, and 1983. The station history metadata are incomplete, particularly in the first half of the record, and are not very useful. However, the coherence of some of these events between the troposphere and stratosphere, for which natural variations are usually uncorrelated, raises confidence in declaring them artificial. As was the case for Rostov and other Soviet stations, there is a tendency for systematic declines in temperature with time during the first few decades. Unfortunately, the problems at Niamey are typical of those found in the African sector, which compounds the lack of spatial coverage.

The effects of adjustment can be seen in Fig. 5b, which consists of both unadjusted (red) and adjusted (blue) temperature series for selected tropospheric levels along with trend lines at 200 and 850 hPa. Adjustment eliminates most of the strong downward trend in the upper troposphere as well as the warming in the lower troposphere. During the first half of the record, adjustment substantially reduces the artificially large interannual variability to a magnitude found in the latter half. Given the number of changepoints and their large magnitudes it is legitimate to question whether, in cases such as this, the true variability can be recovered by any means.

#### e. Pechora

While the examples thus far have focused on *relatively* straightforward decisions, there are cases in which we faced dilemmas. Pechora, Russia, exemplifies problems affecting several of our Soviet stations (Turuhansk, Preobrazheniya, Omsk, and Verkhoyansk, Russia). The tropospheric time series (Fig. 6) show a time period (1979–87) during which temperature is elevated, the magnitude of which grows with height from the lower to upper troposphere. The large magnitude in the upper troposphere ( $\sim 3$ –4 K) with little signal in the mid- to lower troposphere suggests this feature is almost certainly artificial. Also, due to the extreme nature of the problem in the upper troposphere, the typically weak temperature–wind regression (not shown) is a useful tool and points toward artificial causes. Based on a lack of related features at appropriate non-Soviet stations a connection to the Arctic Oscillation was rejected. However, there are several counterarguments including the following: 1) a lack of metadata support, which has major sonde changes in 1976 and 1984, 2) absence of any deviant behavior in the diurnal difference time series, and 3) the fact that like natural phenomena, this feature grows with height in the troposphere but vanishes upon reaching the tropopause.

While it is not unreasonable to judge this event as artificial at Pechora, there is less comfort in doing so at the other four stations, where the magnitude of the effect

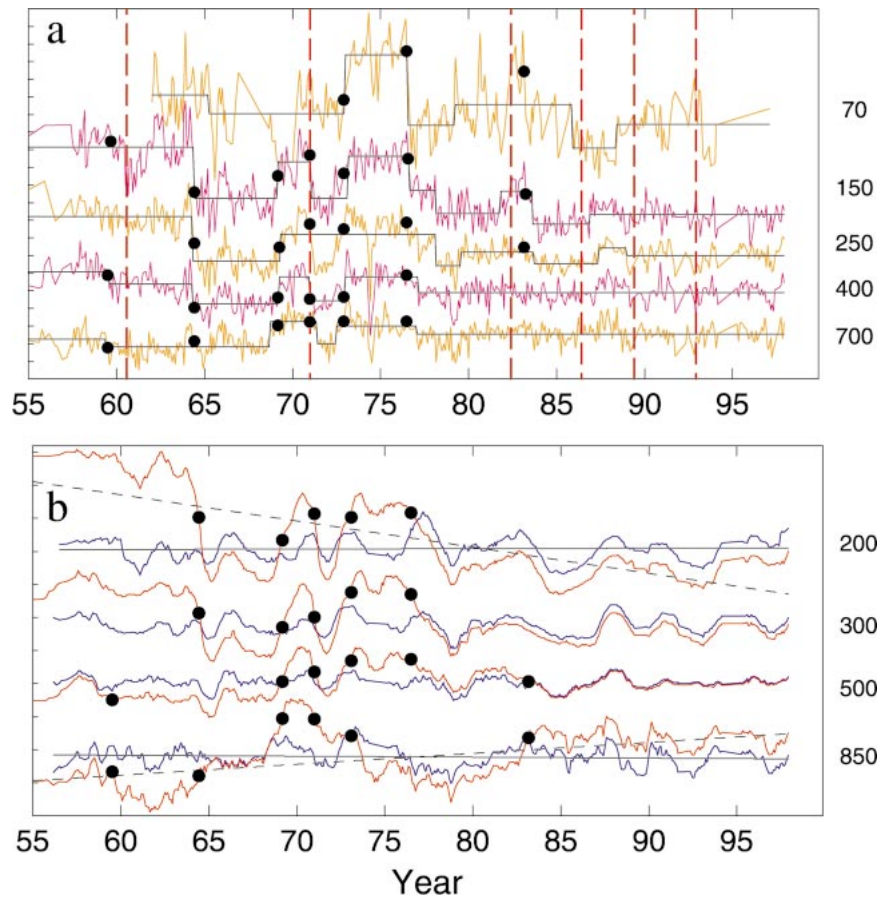


FIG. 5. (a) Same as Fig. 3a except for unsmoothed 9900 UTC temperature at Niamey (61052) for selected stratospheric and tropospheric levels, using alternate orange and magenta curves for clarity. Black dots denote assigned change points relevant to the discussion. (b) Smoothed time series of 9900 UTC temperature at Niamey (61052) for selected tropospheric levels; smoothing is based on a 15-point running median. The tick interval on the ordinate is 1 K. The red (blue) curves are for the unadjusted (adjusted) data. Trend lines at 200 and 850 hPa are based on the unadjusted (dashed) and adjusted (solid) time series. Black dots indicate change points for which adjustments were made.

is comparable to the natural variability. Furthermore, the downward jump found at Pechora in 1987 is absent from one of the other stations and occurs three years later at another. Indications of major sonde changes in the metadata that do not correspond with discontinuities reinforces the notion of serious problems with the Soviet station history information. Nevertheless, we cannot ignore the similarity in timing and appearance of this feature and thus have opted, in this rare instance, to factor neighboring stations strongly into our decision-making process. Accordingly, we have designated this feature as artificial in all of the affected stations except at Verkhoyansk, where it is too weak to allow reasonable adjustment via the methods we employ. Some reassurance of the validity of our decisions can be derived from comparisons with MSU temperatures reported in Part II. As to the cause of the problems, again we can only speculate. It may be that the metadata dates are wrong and that for the sonde used from  $\sim 1979$  to 1987 an

arbitrary decision was made to apply data correction factors only to stratospheric levels.

#### f. Adelaide and Perth

The final examples presented are intended to further illustrate some of the difficulties faced and compromises required, as well as to display some of the more widespread problems of Australian stations. All of our six Australian stations exhibit artificial temperature changes during the late 1980s, primarily in the form of stratospheric cooling, probably associated with the transition from Phillips to Vaisala sondes. This artificial cooling was discovered by Parker et al. (1997) using a comparison with MSU temperatures. Although the evidence makes for confident identification, the exact nature of the problem and the needed remedies are less clear cut, as illustrated by Fig. 7a, which displays both the 50-hPa 0000 UTC temperature (blue) as well as the diurnal

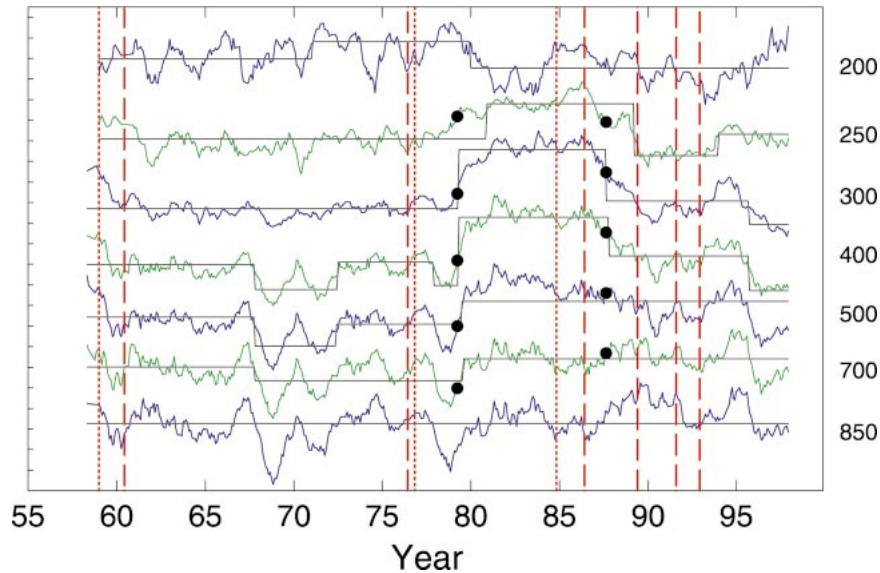


FIG. 6. Same as Fig. 3a except for 0000 UTC temperature at Pechora (23418) from 200 to 850 hPa.

difference series (green) at Adelaide, Australia. During the 1980s the stratospheric temperature declined substantially ( $\sim 1$  K) in an irregular, multiple steplike fashion. However, the diurnal difference series exhibits much of the same behavior. Over this time period the metadata indicates eight significant changes, although some of the dates are uncertain. Since we feel neither

our identification nor adjustment methods are well suited for shortly spaced changepoints, we have compromised, as is sometimes the case, and have placed a single changepoint in 1987, corresponding to the sharpest downward step in both the temperature and difference series, accepting the fact that we cannot ameliorate the behavior during the time of rapid artificial changes. As was the case for Soviet stations, where appropriate we pool information across sites controlled by Australia.

There are several natural features in Fig. 7a worthy of comment as well. There is stratospheric warming associated with major volcanic eruptions (Agung in 1963, Pinatubo in 1991, and possibly El Chichon in 1982), and knowledge of these events enables us to avoid erroneous changepoint assignment. Note how the effects of Agung at the beginning of the record give the false impression of early stratospheric cooling. The step-like drop around 1992–93, noted earlier at Majuro (Fig. 2), which is found in the stratosphere for most stations in our network is another natural feature that we retain.

A substantial fraction of our Australian stations have serious problems near the ground (surface and 1000 hPa). An example is Perth whose 850 hPa and surface temperature series are shown in blue (green) for 0000 (1200) UTC in Fig. 7b. At 850 hPa and nearby levels above, daytime and nighttime temperature series are quite similar. However, at the surface and 1000 mb (not shown) the series abruptly separate in the early and latter parts of the record in both the temperature and the diurnal difference series (not shown). Accordingly, we have assigned changepoints in 1973 and 1984. Some of our other Australian stations have near-surface problems as well, some more complex than this, and with a tendency toward artificial cooling. As a result we have less confidence in near-surface temperature trends in this region.

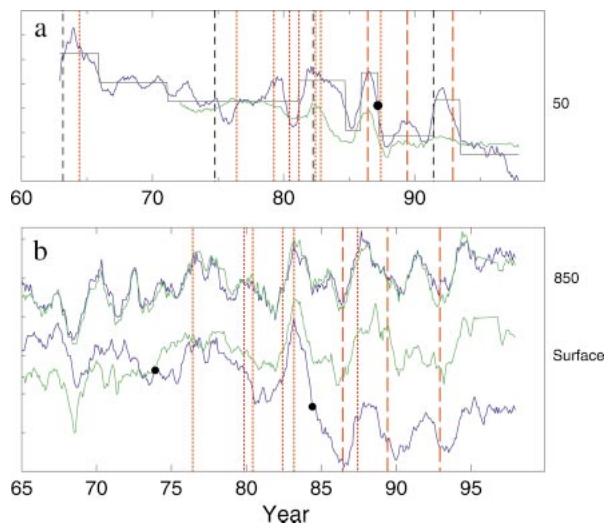


FIG. 7. (a) Blue (green) curve is smoothed time series of 0000 UTC temperature (0000 minus 1200 UTC temperature difference) at Adelaide (94672) at 50 hPa; smoothing is based on a 15-point running median. The tick interval on the ordinate is 1 K. Black step function curve connects statistical changepoints. Dynamic (static) station history metadata events are denoted by dotted (dashed) red vertical lines. Dates of major volcanic eruptions are indicated by dashed black vertical lines. Black dot denotes assigned changepoint relevant to the discussion. (b) Blue (green) curves are smoothed time series of 0000 (1200) UTC temperature at Perth (94610) at 850 hPa and the surface. Smoothing, tick interval, red lines, and black dots are same as in (a).

## 6. Summary and discussion

The problem of temporal inhomogeneity, induced by changes in instrumentation and practices, is a serious concern when attempting to estimate long-term trends of atmospheric temperatures derived from radiosonde observations. The difficulty of this problem results from the fact that the time history of instruments, which is not always known, is unique to a specific country, and sometimes to particular stations within a country. To address this problem a two-step procedure has been developed, involving identification of artificial discontinuities (change points) and other maladies, followed by change point adjustment or deletion of unusable data. Identification of data problems involves a subjective element acting through critical decision-making based on a variety of graphical and textual materials that display the data in its original and transformed states, along with auxiliary information regarding data characteristics, as well as independent indicators of climate variability. The procedures developed have been applied to monthly radiosonde temperatures extending back more than four decades for a near-globally distributed network of 87 stations. Detailed examination of these data indicates that a number of tools are particularly useful in identifying artificial data problems: 1) the time series of the difference between separate 0000 and 1200 UTC monthly means of temperature, 2) the time-varying vertical structure/coherence properties, 3) station history metadata, and 4) statistical identification of discontinuities. For future reference, a detailed record of the maladies found has been created, with information by station and vertical level. This information is available by request from the corresponding author.

The goals of this work are not limited to production of an improved (i.e., more temporally homogeneous) temperature dataset, but also include better understanding of the nature and scope of the problem. It has been found that problems are not only widespread spatially, affecting data from many countries, but that they affect the entire period of record. Instantaneous artificial rises or falls of temperature of  $\sim 0.5$  K are not uncommon, with some instances up to several K. A number of systematic problems have been identified. For example, the global 3-h shift in observation times that occurred in 1957 affects temperature at many stations, particularly near the surface and in the stratosphere. Up to the late 1960s there is a tendency, particularly at former Soviet stations, for large artificial drops in temperature to occur in the upper troposphere and stratosphere, leading to spurious downward trends. Former Soviet and Australian stations, which dominate large regions, were found to be especially problematic, having numerous artificial discontinuities. An artificial drift of  $\sim 1$ – $1.5$  K affecting daytime stratospheric temperatures at some Soviet stations occurs from the late 1970s to the mid-1990s. Spurious drops  $\sim 1$ – $2$  K are found in the stratosphere of Australian stations in the late 1980s and for some western tropical Pacific stations during the 1990s. However, data from Africa and adjacent areas were found to be

the most dubious, due to lack of spatial coverage and severe problems with temporal continuity; not only are derived trends in doubt but prior to about 1980 even the nature of interannual variability can be questioned. Other phenomena, judged to be natural because of their widespread nature and realistic vertical structure include a sudden rise in tropospheric temperatures  $\sim 1976$ – $77$  in the Tropics (Trenberth and Hurrell 1994), stratospheric warming and upper-tropospheric cooling associated with volcanic eruptions, and step-like drops in stratospheric temperatures (Pawson et al. 1998), almost everywhere  $\sim 1992$ – $93$ , as well as during the 1980s, particularly in the Southern Hemisphere and Tropics.

Monthly means derived from mixed observation times, rather than from soundings near one of the standard times (i.e., 0000 or 1200 UTC), can be quite problematic. A change in the mix of times can introduce a spurious discontinuity as big as the largest instrumentally induced artificial changes, because the portion of the diurnal cycle that is sampled has changed. The potential effect is greatest in the low latitudes where solar/diurnal heating cycles have largest amplitude. This serves as a cautionary note on the use of CLIMAT TEMP monthly mean data, which is the basis for the radiosonde products produced by the Hadley Centre (Parker et al. 1997), and which at some stations appear to include mixed observation times (Gaffen et al. 2000b).

Application of the new procedures presented herein yields much higher confidence, relative to our prior attempts (Gaffen et al. 2000b), in identification of artificial discontinuities. In our companion paper (Part II), a comparison with independent MSU satellite data demonstrates an overall increase in consistency between these two datasets as a result of our homogenization. Beyond any improvements made through modification, quality has been added by documenting data limitations and strengths in records that may be of value to prospective users. Further results reported in Part II include estimates of trends of temperature and lower-tropospheric lapse rate for different regions, levels, and time periods, along with uncertainties based on the sensitivities of the trends to the data adjustment; the general lack of sensitivity to the details of our homogenization procedures adds some additional measure of confidence to the results reported.

*Acknowledgments.* The radiosonde data were kindly supplied by Mike Changery and Amy Holbrooks of the National Climatic Data Center under the auspices of the CARDS project. The NOAA Office of Global Programs, Climate Change Data and Detection program provided partial support for this project. We acknowledge the encouragement given by Jerry Mahlman and Bram Oort for this project and the related work that preceded it. We thank Tom Knutson, Brian Soden, Kevin Trenberth, John Christy, and Jim Angell for comments on an earlier version of this manuscript. The three anonymous reviewers provided very thorough and thoughtful comments that improved this manuscript.

## APPENDIX A

## Radiosonde Stations Used in this Study

For observation time, 00 or 12 indicates 0000 or 1200 UTC only, 99 indicates all available observation hours combined, and TD indicates both 0000 and 1200 UTC (i.e., twice daily). “Start” is the first year of the earliest 5-yr period having valid data at 500 hPa for at least 50% of its months; similarly, “End” is the last year of the latest 5-yr period. Start and End range from 1948 to 1997.

WMO No.	Station	Location	Lat	Lon	Obs time	Start	End
01001	Jan Mayen	Greenland Sea	70.93	-8.67	TD	1948	1997
02836	Sodankyla	Finland	67.37	26.65	TD	1948	1997
03005	Lerwick	United Kingdom	60.13	-1.18	TD	1957	1997
04018	Keflavik	Iceland	63.97	-22.60	TD	1948	1997
04360	Angmagssalik	Greenland	65.60	-37.63	TD	1957	1997
08495	North Front	Gibraltar	36.25	-5.55	TD	1957	1997
08508	Lajes Janta Rita	Azores	38.73	-27.07	TD	1948	1997
10868	Munchen	Germany	48.25	11.58	TD	1957	1997
21504	Preobrazheniya	Russia	74.67	112.93	TD	1957	1996
21965	Chetyrekholstolbov	Russia	70.63	162.40	TD	1952	1995
23418	Pechora	Russia	65.12	57.10	TD	1957	1997
23472	Turuhansk	Russia	65.78	87.95	TD	1956	1997
24266	Verkhoyansk	Russia	67.55	133.38	TD	1953	1997
28698	Omsk	Russia	54.93	73.40	TD	1950	1997
30230	Kirensk	Russia	57.77	108.12	TD	1951	1997
32540	Petropavlovsk	Russia	53.08	158.55	TD	1949	1997
34731	Rostov-na-Donu	Russia	47.25	39.82	TD	1949	1997
35121	Orenburg	Russia	51.68	55.10	TD	1949	1997
38880	Ashabad	Turkmenistan	37.97	58.33	TD	1950	1995
40179	Bet Dagan	Israel	32.00	34.82	TD	1955	1997
41024	Jeddah	Saudi Arabia	21.67	39.15	TD	1965	1997
42809	Calcutta	India	22.65	88.45	TD	1954	1997
43003	Bombay	India	19.08	72.85	TD	1955	1997
45004	Hong Kong	Hong Kong	22.32	114.17	TD	1955	1997
47401	Wakkanai	Japan	45.42	141.68	TD	1963	1997
47827	Kagoshima	Japan	31.63	130.60	TD	1954	1997
47991	Minamitorishima	North Pacific Ocean	24.30	153.97	TD	1962	1997
48455	Bangkok	Thailand	13.73	100.50	00	1953	1997
48698	Singapore	Singapore	1.37	103.98	00	1957	1997
51709	Kashi	China	39.47	75.98	TD	1956	1997
52681	Minqin	China	38.72	103.10	TD	1956	1997
60020	Tenerife	Canary Islands	28.47	-16.25	TD	1978	1997
61052	Niamey	Niger	13.48	2.17	99	1955	1997
61641	Dakar	Senegal	14.73	-17.50	12	1960	1997
61902	Ascension Island	Tropical Atlantic Ocean	-7.97	-14.40	12	1980	1997
61967	Diego Garcia	Tropical Indian Ocean	-7.35	72.48	00	1972	1997
61996	Martin de Vivies	South Indian Ocean	-37.80	77.53	12	1975	1997
62010	Tripoli	Libya	32.68	13.17	TD	1948	1996
63741	Nairobi	Kenya	-1.30	36.75	TD	1974	1997
65578	Abidjan	Ivory Coast	5.25	-3.93	99	1958	1997
67083	Antananarivo	Madagascar	-18.80	47.48	00	1980	1993
68588	Durban	South Africa	-29.97	30.95	TD	1970	1997
68816	Capetown	South Africa	-33.96	18.60	TD	1969	1997
68906	Gough Island	South Atlantic Ocean	-40.35	-9.88	TD	1970	1997
68994	Marion Island	South Indian Ocean	-46.88	37.87	TD	1970	1997
70026	Point Barrow	Alaska	71.30	-156.78	TD	1948	1997
70308	Saint Paul Island	Aleutian Islands	57.15	-170.22	TD	1948	1997
70398	Annette Island	Alaska	55.03	-131.57	TD	1948	1997
71072	Mould Bay	Canada	76.23	-119.33	TD	1948	1997
71082	Alert	Canada	82.50	-62.33	TD	1949	1997
71801	St. Johns	Canada	47.67	-52.75	TD	1951	1997
71836	Moosonee	Canada	51.27	-80.65	TD	1954	1997
71926	Baker Lake	Canada	64.30	-96.00	TD	1954	1997
72250	Brownsville	United States	25.92	-97.42	TD	1948	1997
72293	San Diego	United States	32.85	-117.12	TD	1948	1997
72451	Dodge City	United States	37.77	-99.97	TD	1948	1997
72775	Great Falls	United States	47.48	-111.35	TD	1948	1997
78016	Bermuda	Bermuda	32.37	-64.68	TD	1948	1997
78526	San Juan	Puerto Rico	18.43	-66.00	TD	1948	1997
80222	Bogota	Columbia	4.70	-74.15	12	1959	1997
82332	Manaus	Brazil	-3.15	-59.98	12	1967	1997

## APPENDIX A

(Continued)

WMO No.	Station	Location	Lat	Long	Obs time	Start	End
83746	Rio de Janeiro	Brazil	-22.82	-43.25	12	1963	1997
85442	Antofagasta	Chile	-23.42	-70.47	12	1956	1997
85469	Easter Island	South Pacific Ocean	-27.17	-109.43	99	1969	1993
85799	Puerto Montt	Chile	-41.43	-73.10	12	1956	1997
87576	Buenos Aires	Argentina	-34.82	-58.53	TD	1956	1997
89009	Amundsen	Antarctica	-90.00	180.00	00	1961	1992
89050	Bellingshausen	Antarctica	-62.20	-58.93	00	1977	1997
89532	Syowa	Antarctica	-69.00	39.58	TD	1977	1997
89542	Molodezhnaya	Antarctica	-67.67	45.85	TD	1975	1997
89564	Mawson	Antarctica	-67.60	62.88	99	1972	1997
89664	McMurdo	Antarctica	-77.85	166.67	00	1984	1997
91285	Hilo	Hawaii	19.72	-155.07	TD	1949	1997
91334	Truk	Caroline Islands	7.47	151.85	00	1962	1997
91376	Majuro	Marshall Islands	7.08	171.38	00	1962	1997
91408	Koror	Caroline Islands	7.33	134.48	00	1962	1997
91517	Honiara	Solomon Islands	-9.42	160.05	00	1957	1997
91680	Nandi	Fiji	-17.45	177.27	00	1956	1997
91938	Papeete	Tahiti	-17.55	-149.62	00	1965	1997
93844	Invercargill	New Zealand	-46.40	168.33	TD	1966	1997
93986	Chatham Island	New Zealand	-43.95	-176.57	00	1969	1997
94120	Darwin	Australia	-12.43	130.87	00	1951	1997
94294	Townsville	Australia	-19.25	146.77	00	1951	1997
94610	Perth	Australia	-31.92	115.97	TD	1951	1997
94672	Adelaide	Australia	-34.95	138.53	TD	1953	1997
94996	Norfolk Island	South Pacific Ocean	-29.03	167.93	00	1951	1997
94998	Macquarie Island	South Pacific Ocean	-54.50	158.95	TD	1951	1997

## APPENDIX B

## Details of Changepoint Adjustment

## a. Reference level adjustment

The level requiring adjustment is termed the “adjustment level,” and levels used to adjust it the “reference levels,” all of which are from the same station, but not necessarily the same observation time. The procedure begins by determining, for a particular adjustment level, which other levels may serve as a reference series. This is done by correlating the anomaly time series at the adjustment level with those at all other levels using only homogeneous segments, that is, segments whose endpoints are our previously determined changepoints. A minimum correlation of  $\sqrt{0.5}$  (i.e., at least half of the variance of either series could be predicted linearly from the other) has proven reasonable to select the candidate levels; this requirement effectively prevents stratospheric and tropospheric levels from being selected as reference levels for one another. For any adjustment/reference level pair, the adjustment offset (i.e., the additive adjustment factor) is determined by moving the segments of the adjustment level adjacent to the changepoint up or down in small increments, until the correlation coefficient between adjustment and reference levels is maximized. If multiple reference levels (with correlations exceeding the minimum value) are available, a separate adjustment offset is computed for

each, and then a weighted average is computed, using as weights the square of the correlation between adjustment level and reference level.

Reference level adjustment is implemented using three independent steps: 1) adjustment using reference levels that have not been adjusted themselves, 2) adjustment using previously adjusted reference levels, and 3) nonreference level adjustment. The process arbitrarily begins at the changepoint nearest the end of the time series. After adjusting all possible levels using step 1, proceed to step 2 and allow as candidate reference levels those levels that have been adjusted in step 1; multiple reference levels are down-weighted inversely by the amount of their earlier offsets so that levels that have been previously adjusted the most are weighted the least. Iteration proceeds so that once a level is adjusted it may serve immediately as a reference level. After all possible levels have been exhausted using step 2, adjust any remaining levels using the simple nonreference level scheme. Then move backward in time to the next changepoint and perform steps 1–3 again; this continues until all changepoints have been adjusted for all levels at the given station. Note that the process involves simultaneous use of 0000 and 1200 UTC time series if available, so that adjustment may use reference levels from either or both times.

## b. General considerations

While both schemes for changepoint adjustment are first implemented in an objective fashion, later visual

examination of all adjusted time series leads to the option of further refinement. The initial adjustment is accepted as long as visual inspection suggests that the major part of what we have judged by way of change-point assignment to be the artificial signal is removed. However, on occasion the adjustment process (primarily the reference level scheme because of its complexity) produces a clearly unacceptable result. Failure is usually attributable to the presence of some prominent complicating natural feature (e.g., a volcanic eruption) and/or the interaction of multiple reference levels. There are several potential remedies. First is the insertion of a “natural change-point,” which simply reduces the length of a segment used to determine the offset so as to exclude the complicating feature, but is not itself adjusted. Another option is to reclassify surrounding levels, by either adding or removing their changepoints; the presence or absence of these levels, which themselves only marginally require adjustment, can change the result since they serve as reference levels. The most severe option is to replace the changepoint by a data deletion. The objective is to always use the least intrusive action to achieve a reasonable result.

## REFERENCES

- Allen, M., and S. Tett, 1999: Checking for model consistency in optimal fingerprinting. *Climate Dyn.*, **15**, 419–434.
- Angell, J., 1988: Variations and trends in tropospheric and stratospheric global temperatures. *J. Climate*, **1**, 1296–1313.
- , and J. Korshover, 1975: Estimate of the global change in tropospheric temperature between 1958 and 1973. *Mon. Wea. Rev.*, **103**, 1007–1012.
- Bengtsson, L., E. Roeckner, and M. Stendel, 1999: Why is the global warming proceeding much slower than expected? *J. Geophys. Res.*, **104**, 3865–3876.
- Brown, S., D. Parker, C. Folland, and I. Macadam, 2000: Decadal variability in the lower-tropospheric lapse rate. *Geophys. Res. Lett.*, **27**, 997–1000.
- Christy, J., R. Spencer, and W. Braswell, 2000: MSU tropospheric temperatures: Dataset construction and radiosonde comparisons. *J. Atmos. Oceanic Technol.*, **17**, 1153–1170.
- Collins, W., and L. Gandin, 1990: Comprehensive hydrostatic quality control at the National Meteorological Center. *Mon. Wea. Rev.*, **118**, 2752–2767.
- Eskridge, R., A. Alduchov, I. Chernykh, Z. Panmao, A. Polansky, and S. Doty, 1995: A Comprehensive Aerological Research Data Set (CARDS): Rough and systematic errors. *Bull. Amer. Meteor. Soc.*, **76**, 1759–1775.
- FCM-H3, cited 1997: Federal meteorological handbook no. 3: Rawinsonde and pibal observations. [Available online at <http://www.ofcm.gov/fmh3/text/default.htm>.]
- Folland, C., and Coauthors, 2001: Global temperature change and its uncertainties since 1861. *Geophys. Res. Lett.*, **28**, 2621–2624.
- Free, M., and Coauthors, 2002: Creating climate reference datasets: CARDS workshop on adjusting radiosonde temperature data for climate monitoring. *Bull. Amer. Meteor. Soc.*, **83**, 891–899.
- Gaffen, D., 1993: Historical changes in radiosonde instruments and practices. WMO Tech. Doc. 541, Instruments and Observing Methods Rep. 50, World Meteorological Organization, Geneva, Switzerland, 123 pp.
- , 1996: A digitized metadata set of global upper-air station histories. NOAA Tech. Memo. ERL ARL-211, 38 pp.
- , B. Santer, J. Boyle, J. Christy, N. Graham, and R. Ross, 2000a: Multi-decadal changes in the vertical temperature structure of the tropical troposphere. *Science*, **287**, 1239–1241.
- , M. Sargent, R. Habermann, and J. Lanzante, 2000b: Sensitivity of tropospheric and stratospheric temperature trends to radiosonde data quality. *J. Climate*, **13**, 1776–1796.
- Hansen, J., and Coauthors, 1997: Forcings and chaos in interannual to decadal climate change. *J. Geophys. Res.*, **102**, 25 679–25 720.
- , R. Reudy, J. Glasco, and M. Sato, 1999: GISS analysis of surface temperature change. *J. Geophys. Res.*, **104**, 30 997–31 022.
- Hill, D., M. Allen, and P. Stott, 2001: Allowing for solar forcing in the detection of human influence on atmospheric vertical temperature structures. *Geophys. Res. Lett.*, **28**, 1555–1558.
- Hurrell, J., S. Brown, K. Trenberth, and J. Christy, 2000: Comparison of tropospheric temperatures from radiosondes and satellites: 1979–98. *Bull. Amer. Meteor. Soc.*, **81**, 2165–2177.
- Jones, P., M. New, D. Parker, S. Martin, and I. Rigor, 1999: Surface air temperature and its changes over the past 150 years. *Rev. Geophys.*, **37**, 173–199.
- , T. Osborn, K. Briffa, C. Folland, E. Horton, L. Alexander, D. Parker, and N. Rayner, 2001: Adjusting for sampling density in grid box land and ocean surface temperature time series. *J. Geophys. Res.*, **106**, 3371–3380.
- Karl, T., and Coauthors, 1995: Critical issues for long-term climate monitoring. *Climate Change*, **31**, 185–221.
- Keckhut, P., F. Schmidlin, A. Hauchecorne, and M. Chanin, 1999: Stratospheric and mesospheric cooling trend estimates from U.S. rocketsondes at low latitude stations (8°S–34°N), taking into account instrumental changes and natural variability. *J. Atmos. Terr. Phys.*, **61**, 447–459.
- Lanzante, J., 1996: Resistant, robust and nonparametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.*, **16**, 1197–1226.
- , 1998: Correction to “Resistant, robust and nonparametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data.” *Int. J. Climatol.*, **18**, 235.
- , S. Klein, and D. Seidel, 2003: Temporal homogenization of radiosonde temperature data. Part II: Trends, sensitivities, and MSU comparison. *J. Climate*, **16**, 241–262.
- Luers, J., and R. Eskridge, 1998: Use of radiosonde temperature data in climate studies. *J. Climate*, **11**, 1002–1019.
- NRC, 1999: *Adequacy of Climate Observing Systems*. NRC Panel on Climate Observing Systems Status, National Academy Press, 51 pp.
- , 2000: *Reconciling Observations of Global Temperature Change*. NRC Panel on Reconciling Temperature Observations, National Academy Press, 85 pp.
- Oort, A., and H. Liu, 1993: Upper-air temperature trends over the globe, 1958–1989. *J. Climate*, **6**, 292–307.
- Parker, D., M. Gordon, D. Cullum, D. Sexton, C. Folland, and N. Rayner, 1997: A new global gridded radiosonde temperature data base and recent temperature trends. *Geophys. Res. Lett.*, **24**, 1499–1502.
- Pawson, S., K. Labitzke, and S. Leder, 1998: Stepwise changes in stratospheric temperature. *Geophys. Res. Lett.*, **25**, 2157–2160.
- Peterson, T., and R. Vose, 1997: An overview of the Global Historical Climatology Network temperature data base. *Bull. Amer. Meteor. Soc.*, **78**, 2837–2849.
- , and Coauthors, 1998: Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. Climatol.*, **18**, 1493–1517.
- Ramaswamy, V., and Coauthors, 2001: Stratospheric temperature trends: Observations and model simulations. *Rev. Geophys.*, **39**, 71–122.
- Santer, B., and Coauthors, 1996: A search for human influences on the thermal structure of the atmosphere. *Nature*, **382**, 39–46.
- , J. Hnilo, T. Wigley, J. Boyle, C. Doutriaux, M. Fiorino, D.



- Parker, and K. Taylor, 1999: Uncertainties in observationally based estimates of temperature change in the free atmosphere. *J. Geophys. Res.*, **104**, 6305–6333.
- , and Coauthors, 2000: Interpreting differential temperature trends at the surface and in the lower troposphere. *Science*, **287**, 1227–1232.
- Tett, S., J. Mitchell, D. Parker, and M. Allen, 1996: Human influence on the atmospheric vertical temperature structure: Detection and observations. *Science*, **274**, 1170–1173.
- Trenberth, K., and J. Hurrell, 1994: Decadal atmosphere–ocean variations in the Pacific. *Climate Dyn.*, **9**, 303–319.
- Vinnikov, K., P. Groisman, and K. Lugina, 1990: The empirical data on modern global climate changes (temperature and precipitation). *J. Climate*, **3**, 662–677.
- WMO, 1994: Report of the GCOS Atmospheric Observation Panel, first session, Hamburg, Germany. WMO Tech. Doc. 640, WMO, Geneva, Switzerland, 16 pp. [Available online at <http://www.wmo.ch/web/gcos/publications.htm>.]
- Zhai, P., and R. Eskridge, 1996: Analysis of inhomogeneities in radiosonde temperature and humidity time series. *J. Climate*, **9**, 884–894.