# CREATING CLIMATE REFERENCE DATASETS

## CARDS Workshop on Adjusting Radiosonde Temperature Data for Climate Monitoring

BY MELISSA FREE, IMKE DURRE, ENRIC AGUILAR, DIAN SEIDEL, THOMAS C. PETERSON, ROBERT E. ESKRIDGE, JAMES K. LUERS, DAVID PARKER, MARGARET GORDON, JOHN LANZANTE, STEPHEN KLEIN, JOHN CHRISTY, STEVEN SCHROEDER, BRIAN SODEN, LARRY M. MCMILLIN, AND ELIZABETH WEATHERHEAD

The limited agreement between homogeneity adjustments made by different methods shows the difficulty of creating climate-quality radiosonde temperature datasets

During the past 50–60 years, radiosondes launched at least once daily at stations around the globe have provided vertical profiles of temperature and other atmospheric variables for forecast models and reanalyses. The data have also been used in studies of the long-term trends of temperature aloft and tropospheric water vapor content. Interest in the trends of temperatures measured by radiosondes has increased in recent years because the strong warming observed at the earth's surface is not reflected in the record of satellite-derived tropospheric temperatures, which begins in 1979 (NRC Panel on Reconciling Temperature Observations 2000). The radiosonde record provides more detailed vertical resolution and a longer history than the satellite record. Radiosonde data may therefore be crucial to resolving the apparent discrepancy between surface and tropospheric trends (Gaffen et al. 2000a; Brown et al. 2000). Furthermore, radiosonde observations of the vertical profile of temperature change may help distinguish among causes of climate change (Tett et al. 1996; Santer et al. 1996).

Archived time series of radiosonde measurements (like other climate series), however, are often plagued

**AFFILIATIONS:** FREE AND SEIDEL—NOAA/Air Resources Laboratory, Silver Spring, Maryland; DURRE, PETERSON, AND ESKRIDGE—NOAA/National Climatic Data Center, Asheville, North Carolina; AGUILAR—NOAA/National Climatic Data Center, Asheville, North Carolina and Universitat Rovira i Virgili, Tarragona, Spain; LUERS—University of Dayton Research Institute, Dayton, Ohio; PARKER AND GORDON—Hadley Centre, Met Office, Bracknell, United Kingdom; LANZANTE, SODEN, AND KLEIN—NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey; CHRISTY—University of Alabama, Huntsville, Alabama; SCHROEDER—Texas A&M University, College Station, Texas; MCMILLIN—NOAA/National Environmental Satellite, Data, and Information Service, Camp Springs, Maryland; WEATHERHEAD—Cooperative Institute for Research in the Environmental Sciences, University of Colorado, Boulder, Colorado

**CORRESPONDING AUTHOR:** Dr. Melissa Free, Air Resources Laboratory (R/ARL), National Oceanic and Atmospheric Administration, SSMC3, Room 3151, 1315 East–West Highway, Silver Spring, MD 20910

E-mail: melissa.free@noaa.gov

In final form 21 November 2001

by inhomogeneities that compromise the validity of trends calculated from the data. The National Climatic Data Center (NCDC) has consolidated numerous datasets of individual soundings and station history information from around the world into a single database, the Comprehensive Aerological Reference Data Set (CARDS, Eskridge et al. 1995). Even though gross errors, often caused by data transfer or keying errors, should have been removed during NCDC's complex quality control procedures, systematic time-varying biases such as instantaneous artificial jumps of up to several degrees Celsius remain in the data. Sources of these biases include radiative, convective, and conductive effects on the radiosonde instruments and housing, changes in data reduction methods, and changes in instrumentation (Gaffen 1994; Luers and Eskridge 1995, 1998).

Currently, various groups are working to identify and remove these inhomogeneities to make the data more suitable for climate studies. Dian Seidel (formerly Gaffen) and Tom Peterson convened a CARDS Workshop on Adjusting Radiosonde Temperature Data for Climate Monitoring on 11–12 October 2000, at NCDC in Asheville, North Carolina, to discuss and compare the various adjustment methods. Specifically, the workshop aimed to improve understanding of the approaches to identifying and determining the magnitude of temporal inhomogeneities, to compare and assess each method, to improve understanding of the uncertainty in adjusted time series, and to discuss radiosonde station history metadata and how they can be improved.

Representatives from seven different groups actively researching the problem and others interested in the work attended. The participants agreed in advance to focus comparison on 12 radiosonde stations chosen to present a sampling of different regions and types of adjustment problems. Each of the groups then applied its adjustment method to as many of these stations as possible before the workshop and shared the results with other participants. The workshop included the first known comparison of adjustment methods for upper-air in situ temperature records. Its primary accomplishment was to recognize major differences in the adjustments made by the different groups for many stations and the difficulties in making such adjustments with confidence. The user of radiosonde products should be aware of these problems when assessing the reliability of unadjusted or adjusted upper-air data.

The adjustment of radiosonde data generally requires three steps: identification of artificial discontinuities in the data, estimation of the size of these discontinuities, and application of adjustments. Some of the groups at the workshop use all three steps, while others do only the first step. Of the seven techniques discussed at the workshop, five operate on temperature measurements alone. The Texas A&M (TAMU) method adjusts both temperature and humidity, while the method used by Brian Soden [Geophysical Fluid Dynamics Laboratory (GFDL)-Humidity] focuses on humidity. Most of these methods are still under development.

**NCDC METHOD.** As part of the CARDS project, Robert Eskridge and James Luers have developed heat transfer models that adjust temperatures for errors associated mainly with the radiative properties and response times of the radiosondes. They have completed models for VIZ, Vaisala RS-80, and several other widely used radiosondes (Luers and Eskridge 1995, 1998). Several of these models have been evaluated using data from the World Meteorological Organization (WMO) International Intercomparison Experiments (Nash and Schmidlin 1987) in which different types of radiosondes from different countries were flown on the same balloon. The NCDC method begins by removing any prior radiation adjustments. Next, the appropriate heat transfer models are used to adjust the temperature at every level of every sounding. Finally, a plot of the 1200–0000 UTC temperature differences that has been smoothed with a Kolmogorov–Zurbenko filter (Zurbenko et al. 1996) is analyzed for any remaining discontinuities from instrument changes.

Since the workshop, Imke Durre and coworkers have compared unadjusted and adjusted CARDS data with microwave sounding unit (MSU) temperature data (Durre et al. 2002). Based on this work, NCDC has decided not to include these adjustments in the CARDS data archive.

**GFDL METHOD.** John Lanzante and Steve Klein (both at NOAA/GFDL) and Seidel [National Oceanic and Atmospheric Administration/Air Resources Laboratory (NOAA/ARL)] have developed a method to distinguish between artificial and natural abrupt changes in time series of monthly mean radiosonde temperatures. They have demonstrated that purely statistical methods (e.g., Lanzante 1996) can identify large abrupt changes, and that station history information does not always conclusively identify the cause of these changes (Gaffen et al. 2000b). Therefore, in addition to statistics and station metadata, they use indicators such as (in approximate order of importance) 0000 minus 1200 UTC temperature differences, temperatures at nearby levels at the same station, predicted temperatures based on statistical regression of the observed temperatures and winds, the time of observation history, the Southern Oscillation index, dates of major volcanic eruptions, and temperature data from relatively nearby stations.

Using these tools, the investigators individually examine the data and recommend adjustments for abrupt changes and deletions of other poor data. The team then meets to compare recommendations and reach a consensus that becomes part of the metadata. The process is very time consuming and so is not suitable for a large network.

The GFDL/ARL group makes adjustments for a given pressure level using time series of temperature at nearby "reference" levels at the same station. The reference levels either need no adjustment or have already been adjusted. If no reference data are available, an automated adjustment is made based on the difference in the means of data before and after the abrupt change. All adjusted time series are inspected to ensure that they are reasonable. The team has applied this GFDL method to a global network of 87 radiosonde stations.

**TAMU METHOD.** A method designed by Steve Schroeder of Texas A&M University attempts to reconstruct missing metadata and remove artificial biases from humidity and temperature data by looking for signals in the data corresponding to documented changes in instrumentation. Time series of monthly values of 170 (mostly tropospheric) quantities including the basic meteorological measurements as well as the number of levels present, maximum and minimum values reported, presence of unrealistic values, and other indicators are prepared for over 1200 stations, including ships. Data are derived from National Center for Atmospheric Research (NCAR) Dataset 353.4, a data source independent of the CARDS dataset. Stations with reliable histories in Gaffen (1996) are searched for consistent signals of each instrument type. Since similar signals are found at stations with little or no metadata, the changes in instrumentation at those stations can be inferred.

Adjustment algorithms are prepared for each instrument type based on all stations with sufficient data around the time of instrument transition. The average of certain Vaisala and VIZ radiosondes serves as a standard. The dewpoint adjustment transforms the statistical distribution of the dewpoint depressions of the instrument of interest to match the distribution of the standard. In future work, temperature adjustments will precede dewpoint adjustments.

**MET OFFICE (UKMO) METHOD.** David Parker and Margaret Gordon of the Met Office presented a method that adjusts monthly radiosonde temperature data from 1979 onward with satellite-based MSU temperature retrievals as a reference. They apply an update of the technique reported by Parker et al. (1997). Radiosonde temperatures at nine standard levels between 850 and 30 hPa are taken mainly from "CLIMAT TEMP" monthly mean messages and are subjected to quality control (Parker and Cox 1995). Observing times vary and are often a combination of 0000 UTC and 1200 UTC, complicating comparisons with methods using these hours separately. MSU data are available only since 1979 (Christy et al. 2000).

Radiosonde temperature anomalies are vertically averaged with weights approximating the relevant MSU profile. Monthly collocated MSU anomalies are then subtracted from the radiosonde anomalies. Dates of known radiosonde instrument and computer changes are taken from Gaffen (1996). Working with the troposphere and stratosphere separately, average differences (radiosonde minus MSU) before ($\Delta 1$) and after ($\Delta 2$) the most recent known change, but excluding data before any prior change, are compared first. If $\Delta 2 - \Delta 1$ differs significantly from zero at the 95% confidence level according to a *t* test, a seasonally invariant bias adjustment equal to $\Delta 2 - \Delta 1$ is allocated to the earlier radiosonde data for the layer as far back as the previous instrument change or, if no such change occurred, January 1979. All instrument changes in a time series are treated the same way, working back to 1979. Here $\Delta 2$ is calculated from the entire record following the instrument change, incorporating already allocated bias adjustments. The adjustments are then apportioned to individual levels based on average bias adjustments estimated at each level in the Tropics and extratropics for different classes of instrument or computer change.

**UNIVERSITY OF ALABAMA AT HUNTSVILLE (UAH) METHOD.** Developed by John Christy at UAH, this method compares monthly anomalies of radiosonde-simulated MSU brightness temperatures ($\mathrm{RaobT}_b$) and actual satellite MSU brightness temperatures ($\mathrm{T}_b$) to identify artificial discontinuities. In contrast to the Met Office method, which involves a similar comparison, the UAH approach uses radiosonde data from CARDS and considers 0000 UTC and 1200 UTC observations separately whenever possible. Furthermore, the UAH method does not require metadata and identifies data breakpoints for a bulk section of the atmosphere rather than for specific levels.

First a radiosonde–MSU difference time series is formed by subtracting $T_b$ from Raob$T_b$. Next, differences of consecutive 30-month running averages of this time series are calculated. Comparison between Raob$T_b$ and $T_b$ at 30 United States controlled stations that operated with the VIZ radiosonde from 1979 to 1998 shows that changes of about 0.3 K in the Tropics to 0.6 K in high latitudes at individual stations are significant. Adjustments are based on the 30-month differences associated with the breakpoints.

**NESDIS METHOD.** Larry McMillin and collaborators at NOAA/National Environmental Satellite, Data, and Information Service (NESDIS) have been comparing operational retrievals from NOAA high-resolution infrared sounder (HIRS) and MSU satellite instruments to radiosonde measurements. Temperature profiles of individual soundings are adjusted to resemble profiles that would be measured by the widely used Vaisala RS-80 radiosonde.

A function (or set of coefficients) describes the difference between the temperature profiles of each radiosonde model and those of an RS-80. Given a particular sounding taken by an RS-80, all available observations from the radiosonde of interest are searched to find a sounding that closely resembles the RS-80 sounding. Since the two soundings selected are generally not taken at the same location or time, any difference between the two is assumed to stem from a combination of instrument biases and atmospheric conditions. The difference in atmospheric conditions should also be reflected in collocated satellite measurements, so the difference between satellite measurements is subtracted from the observed difference in the pair of soundings to obtain the "true" difference between the two radiosonde types for that particular pair. Coefficients are based on numerous such pairs from a range of atmospheric conditions.

Adjustments for operational radiosondes have been generated with radiosonde data used in models run by the NOAA/Environmental Modeling Center. Unfortunately, radiation corrections already have been applied to these data, and coefficients derived from them are not appropriate for climate studies. A new system that collects two radiosonde measurements (uncorrected and corrected) for each satellite observation started in spring 2001. Once a sufficient sample has been collected with the new system, new coefficients will be generated.

**GFDL-HUMIDITY METHOD.** To highlight artificial discontinuities, Soden of GFDL computes monthly mean differences between satellite infrared radiance measurements from the Television Infrared Observational Satellite (TIROS) Operational Vertical Sounder (TOVS) and radiances calculated from CARDS radiosonde observations. The radiosonde–satellite differences are compared with documented instrument changes and composited according to radiosonde type and satellite model. These composites are evaluated for consistency of the radiosonde adjustments among stations and/or countries. Future efforts will involve application of various change point detection algorithms to supplement existing metadata.

**STRENGTHS AND WEAKNESSES OF THE METHODS.** Each approach has unique strengths and weaknesses. Methods of identifying level shifts in observations inherently run two risks: false identification of an actual shift as an artificial shift, and failure to identify an artificial shift. Betsy Weatherhead of the Cooperative Institute for Research in Environmental Sciences (CIRES) pointed out at the workshop that adjusting data likely affects both the magnitude of the trend derived and the confidence interval associated with that trend. Previously, Weatherhead et al. (1998) had shown that even one level shift in a dataset can increase the period of record needed to detect a given trend by as much as 50%. Adjustments made before a statistical analysis is complete may tend to diminish the derived trend (see also Gaffen et al. 2000b)

Techniques exist for deriving trends at the same time as the level shifts are adjusted. In theory such techniques optimize estimates of the magnitude of the level shift and the trend. On the other hand, the resulting adjustments may depend on assumptions about the trend, thereby introducing additional uncertainty about adjustments. None of the techniques presented at the workshop uses this optimization.

One might expect that methods that apply adjustments only to documented instrument or computer changes might create less uncertainty than methods that derive the time of intervention. Station history information is, however, frequently incomplete or inaccurate. Therefore, some of the methods use independent reference data (e.g., satellite measurements) or statistical techniques to identify undocumented changes and artificial discontinuities.

The NCDC method is the only approach presented at the workshop that uses physical models calibrated to the manufacturers' specifications for the individual thermistors. A disadvantage of this approach is that these models require precise launch times and rise rates of radiosondes, atmospheric humidity, and aerosol content, as well as cloud amount and height. Furthermore, the NCDC method fails to adjust for discontinuities not due to radiation or lag errors. The other six methods, based on signals in the data rather than physical models, may account for a larger variety of errors, but may also run a higher risk of removing natural variability.

Methods based on comparisons between radiosonde and satellite data have the advantage of using an independent reference but do not compensate for instrument changes before 1979. Furthermore, these methods generally assume that the satellite data are homogeneous and that the effects of such natural events as volcanic eruptions are similar in both satellite and radiosonde data. Finally, adjusted data are no longer independent of the satellite record; neither the adjustments nor any conclusions based on them can be verified with the satellite data, and the adjusted data cannot be used to reconcile differences between surface and satellite trends.

The spatial and temporal resolution varies among the techniques. While many methods work with monthly anomalies and/or temperatures of deep atmospheric layers, the NCDC, TAMU, and NESDIS methods adjust the entire vertical profile of each sounding. Most approaches examine 0000 UTC and 1200 UTC observations separately but the Met Office and GFDL-Humidity methods mix the two observation times, which may introduce inhomogeneities when observing schedules change.

**METADATA.** Most methods to produce a homogeneous upper-air dataset rely heavily on accurate and up-to-date information about the data. In the early 1990s, a major effort was made to collect and digitize a comprehensive world metadata bank (Gaffen 1993, 1996). Under WMO auspices, Seidel surveyed member countries and several original sources, greatly improving our knowledge of station histories. Nevertheless, dramatic changes in the upper-air observing network have occurred recently, particularly the concentration of radiosonde production in two main sources, Vaisala and Sippican (which acquired VIZ's radiosonde business), and further gains in market share by Vaisala. In addition, technological advances, such as GPS and computer upgrades, have improved ground equipment. The disintegration of the Soviet Union and economic difficulties in developing countries have also affected the network.

For these reasons, Enric Aguilar, visiting at NCDC, recently updated the metadata. The update relies on WMO publications and information provided directly by national meteorological offices. (The WMO is planning a new survey of the world's radiosonde network history soon.) In addition, information extracted from the headers of the soundings for all the stations in the CARDS database provides valuable information about the radiosonde network.

Whenever possible, the updated metadataset includes the exact times of changes in the network, to the hour. Unfortunately, in many cases, the records are approximate—to the month or year—or uncertain. For this reason, the changes identified by homogeneity analyses of upper-air data are being added to the metadata.

Seidel's database and some of the modifications recently made by Aguilar are available online at the CARDS Web site (lwf.ncdc.noaa.gov/oa/climate/cards/cards_homepage.html).

**COMPARISON OF RESULTS.** In addition to discussing methods, the workshop included comparisons of results at 12 stations. The stations were chosen with emphasis on countries with large station networks and good available metadata. Two stations from Australia were selected because of the known discrepancy between MSU and radiosonde data there (Parker et al. 1997), and stations in India and Africa were included to compare the performance of the techniques given incomplete or unreliable data or metadata. The NESDIS method was not included in the comparison because results were not available at the time of the workshop. We present here a summary of the results along with additional details for Darwin, Australia, as an example.

The groups did not all start with data from the same source. Some groups work with daily data, while others use monthly means. These differences are an additional possible source of discrepancies in results.

Among the six groups included in the comparison, the average number of changes detected per decade examined is largest for UAH and smallest for GFDL, and the number of changes is substantially less than the number of metadata events for most methods and stations. We would expect considerable agreement on dates since all but UAH were working with similar metadata, yet agreement among all groups is the exception rather than the rule (Fig. 1). Out of 21 changes identified by NCDC at 10 stations after January 1979, in only four cases do all groups that examined the station find changes within one year. The NCDC, GFDL, and Met Office groups agree on only six change points (out of eight stations). Results from the two groups (GFDL–Humidity and TAMU) that work primarily with humidity show better agreement than the temperature-based groups.

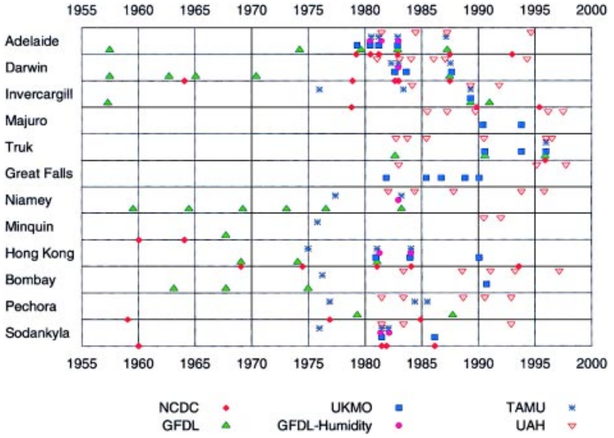

**Fig. 1. Temperature change points found at any level by the six participating groups for the 12 stations considered. Symbols on or above the grid line for each station show change points for that station. Countries for each station are shown in Table 1b.**

Table 1 shows that although several discontinuities are identified by most groups, most cases have agreement of only about 50%. There is even less agreement on the signs and pressure levels of the adjustments. Only once (for a change point in Truk in 1995) do NCDC, GFDL, and the Met Office agree on sign and time of adjustment in both the troposphere and stratosphere.

Closer examination of the results for Darwin (which were the most complete) illustrates the limited agreement between methods. Figure 2 shows that some change points there (e.g., 1987 at 50 hPa) are reasonably clear, while others are not obvious without statistical analysis. Figure 3 compares adjustments at each level at Darwin. For NCDC, we calculated effective adjustments from the mean difference between adjusted and unadjusted time series. The adjustments agree fairly well for some times and levels, but often have differing sizes or even signs. The vertical profiles of the changes made by different groups are noticeably different. The large apparent discontinuity detected by GFDL at the beginning of the record is not adjusted by the Met Office, UAH, or GFDL–Humidity because it is before 1979, or by NCDC, because the 1962 shift is not associated with a documented change in instruments or practices. The 1957 shift is associated with the change in observation time from 0300 to 0000 UTC. Other differences may be due to the fact that the data used by GFDL and CARDS are for observations at 0000 UTC only, whereas the Met Office data combines both 0000 UTC and 1200 UTC.

The differences between the trends produced by the Met Office, GFDL, and NCDC are as big as the trends themselves at many levels (see Fig. 4). In Table 2, the difference in trend between unadjusted and adjusted series at 50 hPa is between 35% and 80% of the original trend. The effect in the lower troposphere is smaller, but for NCDC, the difference is still three times the original trend at 850 hPa.

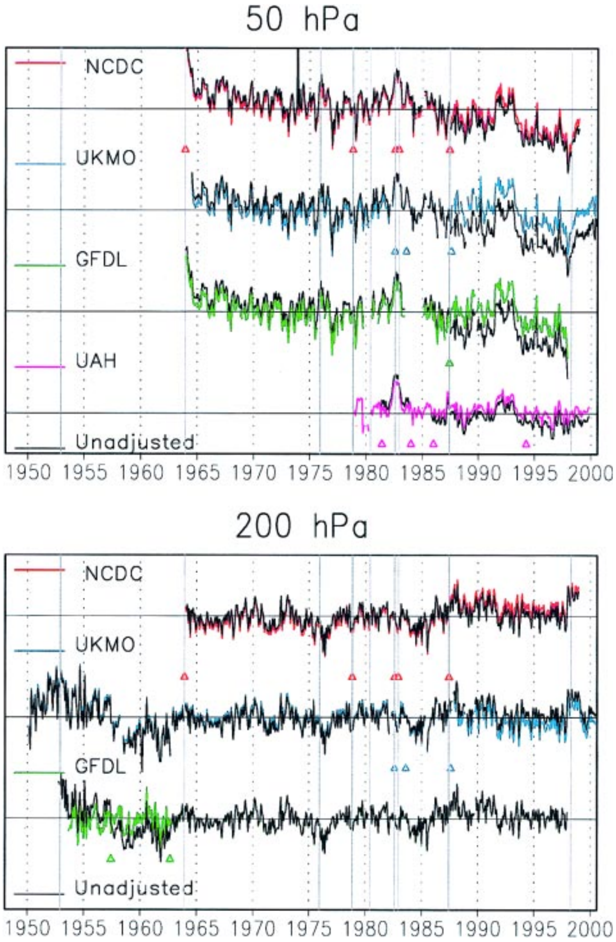

**Fig. 2. (top) Unadjusted and adjusted monthly temperature anomalies at Darwin for 50 hPa, for NCDC (0000 UTC), Met Office (UKMO) (0000 UTC and 1200 UTC combined), and GFDL (0000 UTC), and MSU4 equivalent temperatures calculated from Darwin radiosonde data by UAH. The Met Office series were adjusted only after 1979. Vertical lines indicate times of metadata events. Triangles indicate times of change points found by NCDC, GFDL, Met Office, and UAH groups. Horizontal grid lines representing zero levels for each series are separated by 10 K. (bottom) As in (top), but for 200 hPa and without UAH results. Horizontal grid lines are separated by 5 K.**

These results for Darwin may not be representative of all stations, but they suggest that adjustments for inhomogeneities may affect trends significantly. This finding is consistent with those of Santer et al. (1999) and Gaffen et al. (2000b) as well as with preliminary results presented at the workshop by Lanzante.

**Table 1a. Comparison of change points identified by the different methods, showing percentages of change points that occurred within 6 months of each other for each pair of groups. Where one or both groups found change points, the percent agreement is the number of common change points divided by the total number of change points found by the two groups. Stations where neither of the two groups found any change points are considered to be in 100% agreement. Each entry in the last column is the average of the percentages shown in that row. Comparisons with the Met Office and UAH are limited to 1979–97. The stations may not constitute a representative sample of all radiosonde data, and not all groups produced results for all stations.**

| | NCDC | Met Office | GFDL | GFDL-Humidity | TAMU | UAH | Average |
|---|---|---|---|---|---|---|---|
| **NCDC** | X | 56 | 55 | 53 | 70 | 21 | 50.9 |
| **Met Office** | 56 | X | 39 | 58 | 45 | 23 | 44.0 |
| **GFDL** | 55 | 39 | X | 44 | 49 | 12 | 39.9 |
| **GFDL-Humidity** | 53 | 58 | 44 | X | 82 | 25 | 52.6 |
| **TAMU** | 57 | 45 | 49 | 82 | X | 13 | 51.6 |
| **UAH** | 21 | 23 | 12 | 25 | 13 | X | 18.5 |

**Table 1b. Number of change points identified by each group for 1979–97. An X denotes stations for which a group did not provide data, and an 0 denotes stations where a group examined the record but found no break points from 1979 through 1997.**

| | NCDC | Met Office | GFDL | GFDL-Humidity | TAMU | UAH |
|---|---|---|---|---|---|---|
| **Adelaide, Australia** | 6 | 4 | 3 | 3 | 4 | 4 |
| **Darwin, Australia** | 3 | 3 | 1 | 1 | 3 | 6 |
| **Invercargill, New Zealand** | 2 | 1 | 2 | X | 2 | 3 |
| **Majuro, Marshall Islands** | 0 | 2 | 0 | 0 | 0 | 5 |
| **Truk, Micronesia** | 1 | 3 | 3 | X | 1 | 6 |
| **Great Falls, MT, United States** | 0 | 5 | 0 | X | 0 | 3 |
| **Niamey, Niger** | X | X | 1 | 1 | 1 | 5 |
| **Minquin, China** | 0 | 0 | 0 | X | 0 | 2 |
| **Hong Kong, China** | 3 | 3 | 1 | 2 | 2 | X |
| **Bombay, India** | X | 1 | 0 | X | 0 | 5 |
| **Pechora, Russia** | 1 | X | 2 | X | 2 | 5 |
| **Sodankyla, Finland** | 3 | 2 | X | 2 | 2 | 3 |

**CONCLUSIONS.** The workshop clearly showed the importance of ongoing efforts to insure a climate-quality data record for upper-air temperature and to improve access to station history records. A recent National Research Council report (NRC Panel on Reconciling Temperature Observations 2000) addresses these issues with recommendations on how we can avoid the adjustment problem in the future. The report points out that changes in instruments and observing methods should take into account the need

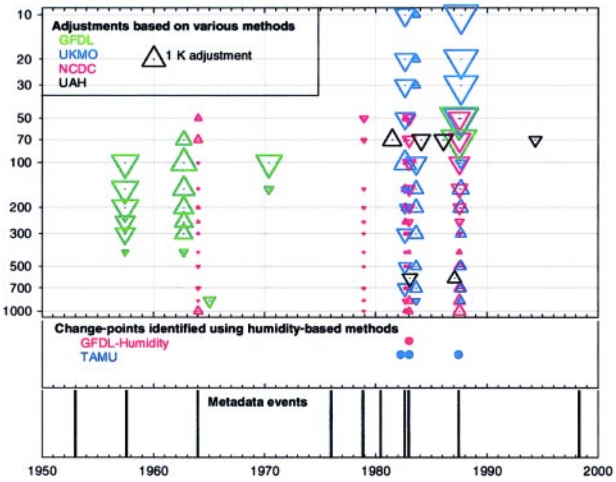

**Fig. 3. Change points found at each atmospheric level for Darwin, Australia. Triangles with apex up indicate that the data segment preceding the change point required positive adjustment. The area of each triangle is proportional to the size of the adjustment, and the area of the black triangle at the upper left indicates a 1 K adjustment. UAH identified single change points and amounts for the stratosphere and troposphere as a whole plotted at 70 and 600 hPa, respectively. TAMU and GFDL-Humidity, whose methods focus on humidity adjustments, did not identify levels, directions, or amounts of adjustments for temperature series. The black lines in the bottom panel show the dates of metadata events.**

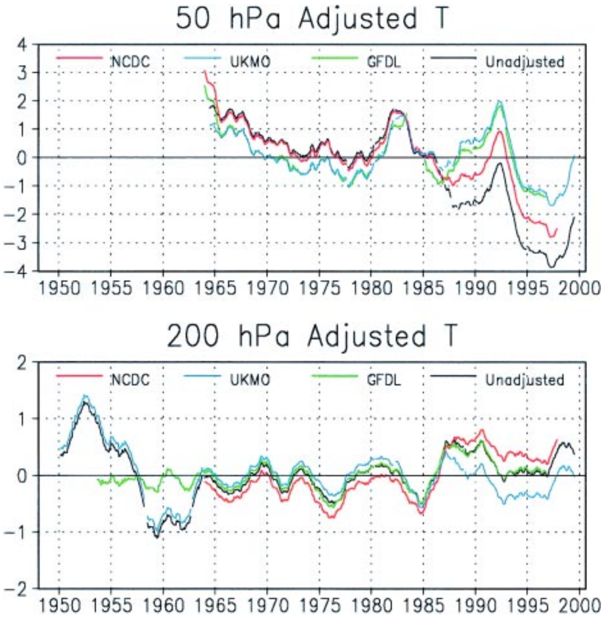

**Fig. 4. (top) Smoothed adjusted monthly temperature anomalies at Darwin, at 50 hPa, for NCDC (0000 UTC), Met Office (UKMO) (0000 UTC and 1200 UTC combined), and GFDL (0000 UTC), with adjusted anomalies from the Met Office. Anomalies have been smoothed with a 25-month running mean. The Met Office series is adjusted only after 1979. (bottom) Same as in (top), but for 200 hPa.**

for continuity. Radiosonde changes should be infrequent and should involve simultaneous launches of old and new instruments at representative sites for a full year to allow future adjustments. The panel also recommended that metadata should be updated and enhanced for the full global network, not just for a selected network, and that records, including raw soundings, should be readily available to facilitate adjustment of past data. We fully support these recommendations and add that use of a reference radiosonde, such as that developed at NCAR by Dave Carlson and coworkers, would also help calibrate operational radiosondes.

If the adjustment results compared at the workshop had shown good agreement, that

**Table 2. Linear temperature trends from adjusted and unadjusted time series for 1979–97 and their differences (adjusted minus unadjusted), in K decade$^{-1}$, for GFDL, NCDC, Met Office, and UAH at Darwin, Australia. UAH trends are for MSU4 equivalent temperatures calculated from Darwin radiosonde data. Uncertainty is at the 95% confidence level. Uncertainty in the differences is calculated as twice the square root of the sum of the squares of the standard errors of the individual time series.**

| | | | | |
|---|---|---|---|---|
| **Unadjusted** | | | | |
| | **GFDL** | **NCDC** | **Met Office** | **UAH** |
| **50 hPa** | −2.58 ± 1.30 | −2.44 ± 1.24 | −2.41 ± 1.12 | −0.89 ± 0.74 |
| **850 hPa** | −0.12 ± 0.24 | −0.07 ± 0.20 | −0.07 ± 0.20 | |
| **Adjusted** | | | | |
| | **GFDL** | **NCDC** | **Met Office** | **UAH** |
| **50 hPa** | −0.62 ± 0.96 | −1.60 ± 1.04 | −0.47 ± 0.86 | −0.24 ± 0.48 |
| **850 hPa** | −0.10 ± 0.22 | −0.27 ± 0.22 | −0.09 ± 0.20 | |
| **Difference** | | | | |
| | **GFDL** | **NCDC** | **Met Office** | **UAH** |
| **50 hPa** | 1.96 ± 1.61 | 0.83 ± 1.62 | 1.94 ± 1.41 | 0.65 ± 0.88 |
| **850 hPa** | 0.02 ± 0.32 | −0.20 ± 0.30 | −0.02 ± 0.28 | |

agreement would have helped to validate the methods. Instead, the results of the workshop suggest significant uncertainties in adjustments at individual stations. It is not readily apparent at this time what approach to homogeneity adjustments is best. The limited agreement between adjustment methods suggests that we must compare the effects of adjustments on the trends both at additional individual stations and for large-scale spatial averages. Any significant differences in large-scale average trends must be understood and reconciled before we can use radiosonde temperature records with full confidence in climate change assessment. The workshop will, we hope, initiate an ongoing assessment of adjustment methods, including further comparisons of radiosonde homogeneity adjustments, to improve our understanding of past variability and trends in upper-air temperature.

**ACKNOWLEDGMENTS.** NOAA's Office of Global Programs, Climate Change Data and Detection Element, funded the workshop. We thank Bill Murray and Chris Miller of that office for their support. Rick Rosen provided helpful comments. The Public Met Service Research and Development Program and the Department of the Environment, Transport, and Regions supported the Met Office authors under Contract FECD/7/12/37. This work was performed while Melissa Free and Imke Durre held National Research Council associateships at NOAA ARL and NCDC, respectively.

## REFERENCES

Brown, S. J., D. E. Parker, C. K. Folland, and I. Macadam, 2000: Decadal variability in the lower-tropospheric lapse rate. *Geophys. Res. Lett.,* **27,** 997–1000.

Christy, J. R., R. W. Spencer, and W. D. Braswell, 2000: MSU tropospheric temperatures: Dataset construction and radiosonde comparisons. *J. Atmos. Oceanic Technol.,* **17,** 1153–1170.

Durre, I., T. C. Peterson, and R. S. Vose, 2002: Evaluation of the effect of the Luers–Eskridge radiation adjustments on radiosonde temperature homogeneity. *J. Climate,* in press.

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

Gaffen, D. J., 1993: Historical changes in radiosonde instruments and practices. WMO/TD-541, WMO Instruments and Observing Methods Rep. 50, 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, 38 pp.

——, B. D. Santer, J. S. Boyle, J. R. Christy, N. E. Graham, and R. J. Ross, 2000a: Multi-decadal changes in the vertical temperature structure of the tropical troposphere. *Science,* **287,** 1239–1241.

——, M. Sargent, R. E. Habermann, and J. R. Lanzante, 2000b: Sensitivity of tropospheric and stratospheric temperature trends to radiosonde data quality. *J. Climate,* **13,** 1776–1796.

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.

Luers, J. K., and R. E. Eskridge, 1995: Temperature corrections for the VIZ and Vaisala radiosondes. *J. Appl. Meteor.,* **34,** 1241–1252.

—— and ——, 1998: Use of radiosonde temperature data in climate studies. *J. Climate,* **11,** 1002–1019.

Nash, J., and F. Schmidlin, 1987: WMO international radiosonde comparison (UK, 1984/USA, 1985). WMO Tech. Doc. 30/TD-195, 103 pp.

NRC Panel on Reconciling Temperature Observations, 2000: *Improving Atmospheric Temperature Monitoring Capabilities.* National Academy Press, 18 pp.

Parker, D. E., 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 gridded radiosonde temperature data base and recent temperature trends. *Geophys. Res. Lett.,* **24,** 1499–1502.

Santer, B. D., and Coauthors, 1996: A search for human influences on the thermal structure of the atmosphere. *Nature,* **382,** 39–46.

——, 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.

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.

Weatherhead, E. C. and Coauthors, 1998: Factors affecting the detection of trends: Statistical considerations and applications to environmental data. *J. Geophys. Res.,* **103,** 17 149–17 161.

Zurbenko, I., P. S. Porter, S. T. Rao, J. J. Ku, R. Gui, and R. E. Eskridge, 1996: Detecting discontinuities in time series of upper-air data: Demonstration of an adaptive filter technique. *J. Climate,* **9,** 3548–3560.