

Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC): A new data set of large-area anomaly time series

Melissa Free, Dian J. Seidel, and James K. Angell

NOAA Air Resources Laboratory, Silver Spring, Maryland, USA

John Lanzante

NOAA Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey, USA

Imke Durre and Thomas C. Peterson

NOAA National Climatic Data Center, Asheville, North Carolina, USA

Received 3 May 2005; revised 27 July 2005; accepted 1 September 2005; published 16 November 2005.

[1] A new data set containing large-scale regional mean upper air temperatures based on adjusted global radiosonde data is now available up to the present. Starting with data from 85 of the 87 stations adjusted for homogeneity by Lanzante, Klein and Seidel, we extend the data beyond 1997 where available, using a first differencing method combined with guidance from station metadata. The data set consists of temperature anomaly time series for the globe, the hemispheres, tropics (30°N–30°S) and extratropics. Data provided include annual time series for 13 pressure levels from the surface to 30 mbar and seasonal time series for three broader layers (850–300, 300–100 and 100–50 mbar). The additional years of data increase trends to more than 0.1 K/decade for the global and tropical midtroposphere for 1979–2004. Trends in the stratosphere are approximately –0.5 to –0.9 K/decade and are more negative in the tropics than for the globe. Differences between trends at the surface and in the troposphere are generally reduced in the new time series as compared to raw data and are near zero in the global mean for 1979–2004. We estimate the uncertainty in global mean trends from 1979 to 2004 introduced by the use of first difference processing after 1995 at less than 0.02–0.04 K/decade in the troposphere and up to 0.15 K/decade in the stratosphere at individual pressure levels. Our reliance on metadata, which is often incomplete or unclear, adds further, unquantified uncertainty that could be comparable to the uncertainty from the FD processing. Because the first differencing method cannot be used for individual stations, we also provide updated station time series that are unadjusted after 1997. The Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC) data set will be archived and updated at NOAA's National Climatic Data Center as part of its climate monitoring program.

Citation: Free, M., D. J. Seidel, J. K. Angell, J. Lanzante, I. Durre, and T. C. Peterson (2005), Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC): A new data set of large-area anomaly time series, *J. Geophys. Res.*, *110*, D22101, doi:10.1029/2005JD006169.

1. Introduction

[2] Changes in upper air temperatures are an important indicator for detection and attribution of human influences on climate [Santer *et al.*, 1996; Tett *et al.*, 1996; Thorne *et al.*, 2002; *International Ad Hoc Detection and Attribution Group*, 2005]. They have recently been the subject of debate because the strong warming observed at the earth's surface has not been seen in many analyses of tropospheric temperature records during the satellite measurement era [National Research Council, 2000]. Moreover, climate

models predict more warming in the troposphere than at the surface in the tropics, but this predicted amplification is not present in many upper air data sets [Santer *et al.*, 2005]. Understanding observational uncertainties is critical to resolving this disparity. Different upper air data sets give widely differing temperature trends, indicating a high level of uncertainty in observed climate changes above the surface [Seidel *et al.*, 2004]. Although radiosonde temperature records have great potential value in this debate, their use has been limited by the presence of many inhomogeneities due to changes in instruments and practices [Gaffen, 1994]. These inhomogeneities, which may result in spurious trends, are difficult to remove because of the frequent lack of neighboring homogeneous reference series, incorrect or

incomplete station history information, and the need to maintain vertical consistency in adjusted time series. Despite these problems, it is essential to create and maintain upper air temperature records from a variety of sources and to use a range of processing approaches to better understand remaining uncertainties.

[3] Several quite different approaches have been used in the past to reduce inhomogeneities in radiosonde data sets [Free *et al.*, 2002; Seidel *et al.*, 2004], including methods based on estimates of radiative heating errors for various instruments and methods using reference time series such as satellite data or neighbor stations to identify artificial changes. Two additional methods developed since Free *et al.* [2002] are based on comparisons with neighbor stations (“HadAT” [Thorne *et al.*, 2005]) or with data derived from reanalysis background fields [Haimberger, 2005]. Statistical methods to detect change points in time series are often used in homogenization procedures, but, when applied to individual radiosonde station data without the use of reference series, cannot always distinguish between natural and artificial changes in temperature [Gaffen *et al.*, 2000]. Using a different approach combining statistical tests with several other factors and using decisions by a committee of scientists, a recently created radiosonde temperature data set (LKS [Lanzante *et al.*, 2003a, 2003b]) has been carefully scrutinized for temporal homogeneity throughout its period of record and is therefore better suited for trend analysis than previously available data sets such as HadRT [Parker *et al.*, 1997] or the CARDS data set [Eskridge *et al.*, 1995]. This 87-station data set covers the period from 1948 to 1997.

[4] The Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC) project is a collaborative effort involving NOAA scientists from the Air Resources Laboratory, the Geophysical Fluid Dynamics Laboratory, and the National Climatic Data Center (NCDC). Its purpose was to create large-scale regional mean time series based on LKS, to extend those series to the present, and to allow for future updates without the LKS adjustment procedure (which is labor-intensive and requires a panel of experts) while minimizing inhomogeneities that would interfere with trend assessment. Our goal was a method that was independent of satellite or reanalysis data and that was usable in the absence of appropriate near neighbor stations. To achieve this goal, we used a new approach involving first differences (“FD”), described by Free *et al.* [2004] and in section 3.2 below. The results of this method constitute “RATPAC-A”. For comparison, we also create equivalent time series without the FD method (“RATPAC-B”).

[5] In this paper we describe the methods used to construct the RATPAC data sets from the LKS station data and to estimate the uncertainty in RATPAC-A resulting from our methods, and present some basic results. We also discuss our reasons for choosing not to expand the data set to include more stations.

2. Data Sources

[6] The new data set is based on the 87-station adjusted time series described in LKS. These were derived from data in the Comprehensive Aerological Reference Data Set

(CARDS) at NCDC [Eskridge *et al.*, 1995]. Monthly means for 87 carefully selected stations were adjusted using a multifactor expert analysis by a team of three climate scientists, without use of satellite data as references and with minimal use of neighbor station comparisons. The team visually examined time series of temperatures at multiple levels, night-day temperature differences, temperatures predicted from regression relationships, and temperatures at other nearby stations. They also considered metadata, statistical change points, the Southern Oscillation Index and the dates of major volcanic eruptions. Using these indicators, they identified artificial change points and remedied them by either adjusting the time series at each affected level or, if adjustment was not feasible, by deleting data. The adjustments were then examined for reasonableness. This procedure differs both from that used for HadAT, which relies primarily on neighbor station data, including the adjusted LKS data, to adjust station data, and from that of Haimberger [2005] which relies on information derived from reanalyses. The LKS data consist of monthly temperatures for 16 atmospheric levels from the surface to 10 mbar, from 1948 to 1997. Because of previously recognized problems with the data from India [Parker *et al.*, 1997; Free and Angell, 2002], we did not use the LKS station data for Bombay or Calcutta, leaving 85 stations (listed in Table S1 in the Supplemental Material¹). We deleted the 10 and 20 mbar levels from the RATPAC products because of the scarcity of data at those levels. LKS found the 1000 mbar data to be more erratic and less reliable than other levels in the troposphere, probably because of problems arising from days when the surface pressure is less than 1000 mbar [Lanzante *et al.*, 2003b]. (On those days, data may be reported for 1000 mbar by extrapolating below the surface. Alternatively, if no data are reported for those days, the monthly values may be biased as a result.) We therefore deleted the 1000 mbar data from RATPAC.

[7] To remedy various recently identified problems in the CARDS database, NCDC has undertaken a wholesale revision of the CARDS quality control procedures [Durre *et al.*, 2004]. In the CARDS data set, information from several sources was sometimes combined to create a single time series, and in some cases Durre *et al.* found that those sources were not sufficiently consistent to create a homogeneous record when merged. By eliminating these inconsistent data sources and other problems, the new data set may reduce or eliminate some artificial shifts present in the CARDS data. We have used the resulting data set, the Integrated Global Radiosonde Archive (IGRA), rather than the CARDS data set, to extend the station data past 1997.

[8] We visually compared monthly mean time series of unadjusted LKS and IGRA data for the 85 individual stations through 1997 and found them to be generally consistent above 1000 mbar. The principal apparent difference between the station data sets is that data are missing for different months in the two time series. At the surface and 1000 mbar, the unadjusted LKS data show abrupt shifts in temperature relative to IGRA at a number of stations in the earlier part of the record. Comparison of the unadjusted and adjusted LKS series at these stations suggests that the

¹Auxiliary material is available at <ftp://ftp.agu.org/apend/jd/2005JD006169>.

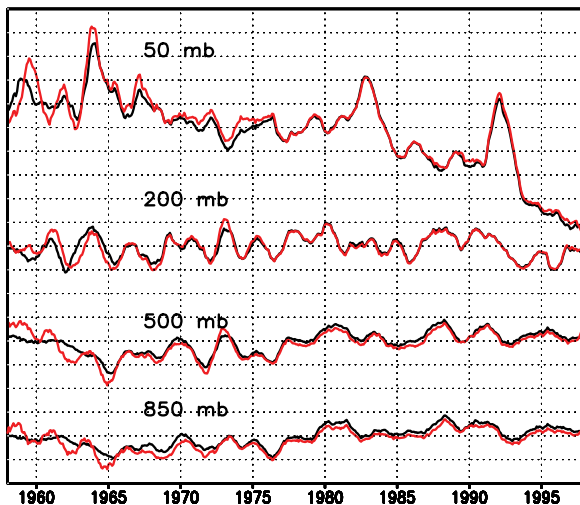


Figure 1. Twelve-month running means of monthly global mean temperature anomalies from unadjusted LKS (in black) and IGRA (in red) data sets. Horizontal grid lines are spaced 0.5 K apart. The IGRA anomalies are computed with respect to the mean for 1970–1999. The LKS anomalies are with respect to the entire period of record. In the top curve, 0.3 K has been added to the LKS anomalies to facilitate comparison.

largest of these shifts have been eliminated by the LKS adjustment process. These appear to be problems in the CARDS data set that were eliminated by the IGRA processing and also largely remedied by the LKS adjustment procedure.

[9] The global mean time series from the (CARDS-based) unadjusted LKS and IGRA for the 85 stations used in this work differ most notably before 1965 (Figure 1). Although IGRA is an improvement over CARDS, we do not have LKS adjustments based on IGRA. Because of the careful scrutiny used by the LKS team to create the adjusted LKS data, LKS is likely to be more reliable than a data set derived by applying the FD method to the IGRA data before 1995. (In 1996 and 1997, the LKS adjustments are less reliable than for earlier periods because of the small number of data points present after the adjustments. For these years, the FD result is likely to be equally valid.) We therefore use LKS instead of IGRA before 1995 to reap the substantial benefits of the LKS homogeneity adjustments. However, because of the differences between the data sets before 1965, RATPAC data from that period should be viewed with caution.

3. Updating Methods

3.1. RATPAC-B

[10] Because the first difference method used for RATPAC-A does not allow production of individual station time series, and to provide alternative large-scale mean time series for comparison with the FD time series, we also created a set of updated station time series by appending monthly mean station data from IGRA for 1997–2004 to the corresponding adjusted LKS station time series for 1958–1997 without any adjustment for inhomogeneities after 1997. We combine the 00Z and 12Z observations

where both are available, and for consistency we use only those observation times from the IGRA data that were present in the LKS adjusted station data. (Station data for the two observation times will also be available separately.) To minimize the discontinuity at 1997, we add to the IGRA monthly means a factor equal to the difference between the means of the IGRA and LKS data for 1996–1997. The effect is to shift the IGRA data so that the means of the two data sets for the last 2 years of the LKS time series are equal. If both time series are present for fewer than 9 months in those 2 years, we use instead the time period 1990–1997. At one or more levels at ~ 14 stations, LKS found a discontinuity, but deleted data after the discontinuity rather than adjusting it because adjustment was not feasible. In creating RATPAC-B we do not append the IGRA data for those levels at those stations after 1997 to avoid reintroducing known inhomogeneities. (These IGRA data are used in the creation of RATPAC-A, however, since the FD procedure is expected to deal with the discontinuity.) Four of the possible 85 stations (Preobrazheniya, Chetyrkhstolbov, Mould Bay and Ashabad) have no data after 1997, and so are not extended. Data for four other stations (Abidjan, Honiara, Bellingshausen and Molodezhnaya) stop before 2004.

[11] The result is a set of station time series for 1958 through 2004 that have been adjusted through 1997 but not afterward. Hemispheric, global, tropical and extratropical means created from these station time series are used for comparison with the FD product described in the next subsection.

3.2. RATPAC-A

3.2.1. First Differencing

[12] To update the LKS data, we used the first difference (“FD”) procedure [Peterson *et al.*, 1998]. As discussed in more detail by Free *et al.* [2004], this method allows us to reduce inhomogeneities in large-scale mean time series without making adjustments to the individual time series. In this method we take the difference in temperature between one time step and the next (the “first difference”), then compute large-scale means of the FD series, and finally reconstruct large-scale temperature series from the FD series (see Appendix A for details). By omitting portions of the station time series around the times of known changes in instruments or procedures, we attempt to eliminate the effect of inhomogeneities due to such changes. However, the method introduces a random error that increases with the number of time gaps in the data and with decreasing number of stations, so that results are limited to large-scale means. Although our method does not use neighbor stations as reference series in the usual sense, it does in effect rely on other stations in a region to supply information about temperature change at times of metadata events at an affected station, and so does not adjust individual stations independently.

[13] Here we applied the method to IGRA monthly means starting in 1996. Before 1996, the resulting time series is the mean of the adjusted LKS station data, without use of FD. Although the LKS data set runs through 1997, we substituted IGRA data for 1996 and 1997 because the short record left after the adjustments makes LKS adjustments in 1996 and 1997 less reliable than those at earlier

times. (The LKS approach requires several years of data both before and after a possible inhomogeneity to make the best adjustment.)

[14] Starting in 1995, we deleted 6 months of data from the IGRA monthly means before and after each metadata event for any station having a relevant event documented by a report from the country in which the station was located. Some events were considered relevant only for certain stations and levels. For example, a change in reporting practices affecting temperatures below -90°C was considered relevant only for stations and levels where temperatures near that cutoff were reported. We also deleted data at times in 1996 and 1997 where LKS had made adjustments or had deleted data because of homogeneity concerns. Despite recent efforts by NCDC to update the station histories, useful metadata after 1995 was available for just 38 of the 85 stations. On the basis of this metadata, cuts were made at a total of 29 stations. The series were then combined using the method described in more detail in Appendix A.

3.2.2. Spatial Averaging

[15] In an effort to obtain spatially unbiased large-scale means, we compensate for uneven longitudinal distribution of stations by creating regional means before averaging data into zonal bands. Each 30° zonal band was divided into three longitudinal regions of 120° each: 30°W to 90°E , 90°E to 150°W and 150°W to 30°W . Hemispheric (0 – 90°), tropical (30°S – 30°N) and extratropical (30 – 90°) means were calculated from these zonal means, areally weighted using the cosine of the latitude of the midpoint of the zone, and the global mean was the average of the hemispheric means. Although alternative definitions of the tropical zone, such as 20°S to 20°N , are sometimes used, we chose to use a broader definition. While the region from 30°S to 30°N will include areas beyond the tropical Hadley cell in some seasons, the alternative definition excludes some of it, so that the best choice for climate trend assessment is not clear. We therefore prefer to use the broader definition to maximize the number of stations available for the FD procedure in the tropics. To facilitate comparison with other data sets, time series for the region from 20°N to 20°S will be available in the future.

3.2.3. Endpoint Outlier Trimming

[16] To reduce the random errors introduced by the FD procedure, we used an endpoint outlier trimming procedure. As described by Peterson *et al.* [1998] and Free *et al.* [2004], this procedure removes data exceeding a prescribed multiple of the standard deviation of the original time series if the data fall at the end of a data segment (immediately before or after a gap). If a larger multiple is used as a cutoff, fewer data points are removed than with a smaller multiple. Results from the FD procedure are sensitive to the choice of this multiple, or trim factor (see section 5.1 below). Tests with reanalysis data indicate that the range of trends in randomly cut series combined with FD so as to simulate the construction of the RATPAC products is less for smaller trim factors, i.e., when more data points are trimmed. To supplement these tests we combined the LKS input data through 1997 using FD, after cutting the data according to the LKS adjustment dates. We compared the resulting trends to the trends from the actual LKS adjusted data for the hemispheres and the tropics and found the best match was

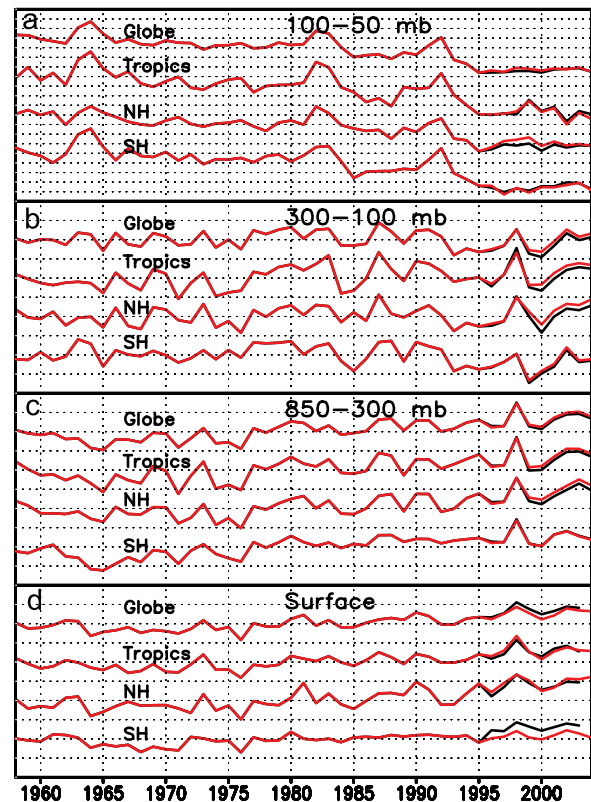


Figure 2. Time series of annual mean temperature anomalies from RATPAC-A (in red) and B (in black) for (a) 100–50 mbar, (b) 300–100 mbar, (c) 850–300 mbar, and (d) the surface. Intervals between horizontal grid lines are 0.5 K.

with trim factors 1.0–1.2. We therefore chose to use a factor equal to one standard deviation from the mean for endpoint outlier trimming.

[17] In another effort to reduce random errors, before making cuts at the times of metadata events, we filled in missing data in gaps of less than 4 months using linear interpolation between months of data. We used this interpolation only at stations that were to be cut because of metadata events. The mean number of months of data added by interpolation to these 38 stations was ~ 6 per station. The difference between trends for 1979–2004 with and without this interpolation is <0.02 K/decade below 50 mbar in all 6 regions. At 30 mbar the differences are up to 0.10 K/decade. The trends using interpolation are more negative than those created without interpolation.

4. Results

[18] Figure 2 shows the time series of annual mean temperature anomalies in RATPAC-A and B for three atmospheric layers for the globe, tropics and hemispheres. The layer means are constructed using volume weighting (weighting by the log of pressure) and resemble the layers used by Angell and Korshover [1975]. For the 100–50 mbar layer (Figure 2a), the period after 1995 shows an apparent leveling off of the stratospheric cooling in the global mean, with continued cooling in the tropics and a suggestion of

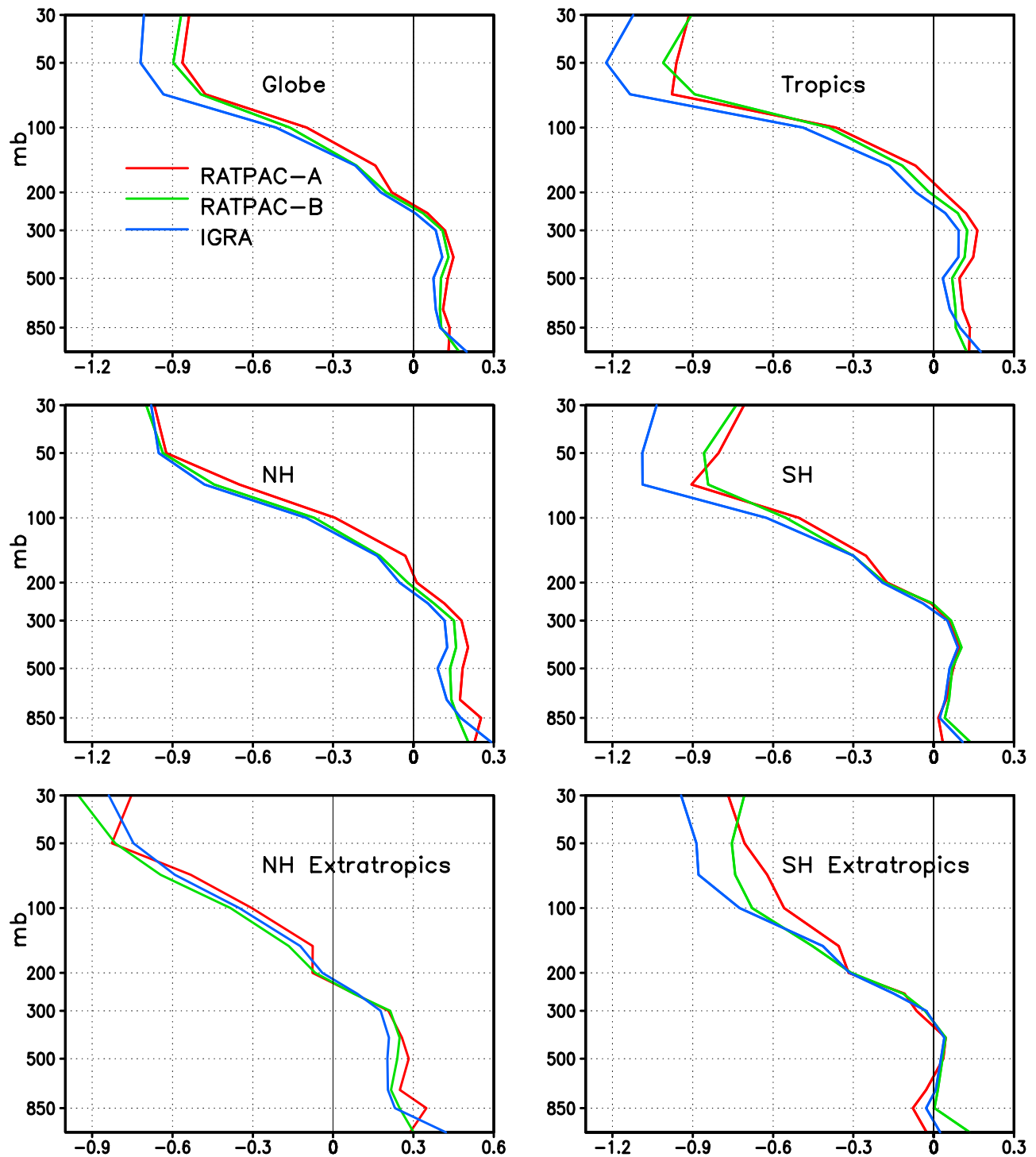


Figure 3. Least squares linear trends in annual mean temperatures for 1979–2004 for the globe, tropics, hemispheres and extratropics from IGRA, RATPAC-A, and RATPAC-B, in K/decade.

warming beginning in the SH. The 850–300 mbar layer (Figure 2c) shows the expected large warming in 1998 coincident with the large El Niño–Southern Oscillation event and a more sustained warming for 2001 and later. At 300–100 mbar (Figure 2b), temperatures since 1997 have been lower than in the 1980s in the SH, but similar or warmer in the tropics and NH. Although the FD procedure

tends to increase temporal variability, the series before 1995 appear similar in character to those after, and this is true also for the time series at individual pressure levels (not shown).

[19] Figure 3 shows the vertical profiles of least squares linear trends in the IGRA and the two RATPAC time series for 1979–2004. RATPAC-B generally shows more warming than the IGRA data above the surface except in the NH

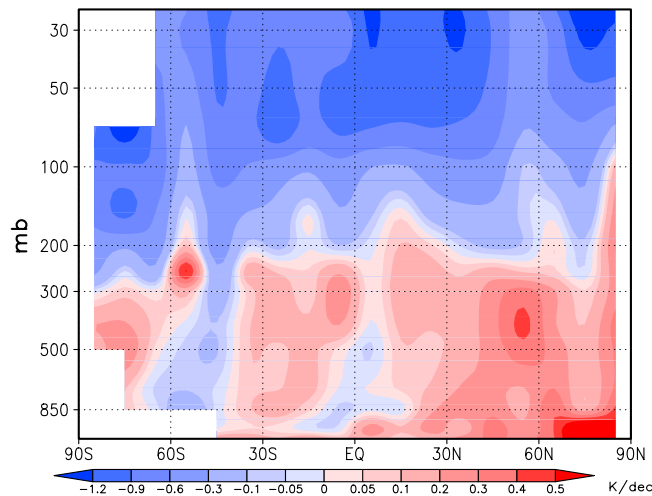


Figure 4. Least squares linear trends in 10° latitudinal means of RATPAC-B for 1979–2004.

extratropics above 200 mbar. RATPAC-A shows more lower-tropospheric warming than B in the NH and tropics but less in the SH. Because of the small number of stations in the SH extratropics, trends in that area must be interpreted with caution. Both the RATPAC series show less cooling in the SH and tropical stratosphere than IGRA. All three data sets show similar stratospheric trends in the NH at 50 mbar and above. In all three, trends in the tropical stratosphere are larger than in the extratropics, as found by *Thompson and Solomon* [2005]. The validity of these large tropical stratospheric trends has, however, recently been questioned (see section 5.4 below). For the longer time period 1960–2004 (not shown), trends for the A and B products are very similar to each other, but IGRA trends remain more negative in all regions above 400 mbar and in the troposphere in the NH. For both time periods, the more positive trends in the RATPAC data sets compared to IGRA are primarily the result of the LKS adjustments before 1995, and the similarity between trends for the two versions of RATPAC arises because the two are identical before 1996.

[20] Because the FD procedure is not suitable for combining the small number of time series present in individual 10° zones, we use RATPAC-B to illustrate details of the zonal mean profile of trends. The latitude-height plot of RATPAC-B trends for 1979–2004 (Figure 4) shows cooling at all levels above 200 mbar and warming in most of the troposphere below 200 mbar, with tropospheric cooling bands at $0\text{--}10^\circ\text{N}$ and $40\text{--}60^\circ\text{S}$. Because the number of stations in some 10° latitude bands is small, some of these latitudinal details may represent noise rather than true climatic signals, and these trends may include effects of inhomogeneities after 1997.

[21] Trends for layer mean series (Table 1 and Table S2 in Supplemental Material) are generally similar for RATPAC-A and B, with B showing slightly less positive trends than A except in the SH. The similarity in trends is expected since the time periods examined (1979–2004 and 1960–2004) are dominated by the years before 1996, in which the two series are identical. The A series show warming of ~ 0.13 K/decade in the global mean troposphere and slightly less in the tropical troposphere for 1979–2004. Trends in the strato-

sphere for the same period are approximately -0.5 to -0.9 K/decade. The addition of the years 1998–2004 increases the warming trends in the troposphere and reduces the cooling trends in the stratosphere as compared to the trends for 1979–1997.

[22] Trends at the surface minus those in the troposphere for 1979–2004 (Table 2) are less in RATPAC-B than in IGRA for all regions except the SH and SH extratropics, and are less in RATPAC-A than in RATPAC-B for all regions. For the globe, RATPAC-A shows no difference between surface and tropospheric trends, and in the SH, RATPAC-A shows the troposphere warming slightly more than the surface. Since most climate models show greater warming in the troposphere than at the surface in the tropics for this time period [*Santer et al.*, 2000, 2005; *Gaffen et al.*, 2000], the LKS adjustments in RATPAC-B and the first difference procedure used for RATPAC-A appear to bring trend profiles into better agreement with those in models. However, in the tropics, surface trends are still slightly greater than tropospheric trends, and, as noted above, areas of tropospheric cooling exist in the deep tropics in RATPAC-B. For 1960–2004, tropospheric trends exceed those at the surface in all three data sets.

[23] Seasonal trends (Table 3) suggest more cooling in the tropical stratosphere in SON and DJF than in boreal spring or summer. The difference between stratospheric temperatures in boreal spring and those in winter or fall shows a least squares linear trend that is statistically significant at the 95% level. Stratospheric trends in the tropics are similar to those in the extratropics in MAM and JJA, but the tropical trends are significantly more negative (at the 95% level) than SH extratropical trends in SON and NH extratropical trends in DJF. Other layers also show apparent differences between trends for different seasons (see Table S3 in Supplemental Material), but examination of monthly trends (not shown) shows no coherent seasonal pattern, so those differences may not be physically meaningful.

5. Sources of Uncertainty

5.1. First Difference Method

[24] To estimate the error introduced into the extended time series by the first difference procedure, we did Monte Carlo tests using NCEP/NCAR reanalysis data [*Kistler et al.*, 2001] to simulate the radiosonde station data. Using grid points collocated with the RATPAC stations, we selected artificial break dates after 1996 at random, deleted the data for 12 months surrounding those dates and then combined the reanalysis time series using the FD method described above. The number of break dates used at a given station was equal to the number actually found in the metadata for that station. The process was repeated 10,000 times with different randomly selected dates for the cuts. For each iteration, we calculated trends in the resulting series for 1960–2004 (not shown) and 1979–2004. The input time series were masked to match the missing months in the actual IGRA data sets for the corresponding stations. All tests used endpoint outlier trimming at 1.0 standard deviation and the spatial averaging and interpolation procedures described in section 3.2 above.

[25] We then took the difference between the 5th and 95th percentiles of the trends from the 10,000 iterations as a

Table 1. Least-Squares Linear Trends in Annual Mean Temperatures for 1979–2004 and 1960–2004 From RATPAC-A and B^a

	NH	SH	GLOBE	TROPICS	NH EX	SH EX
<i>1979–2004</i>						
850–300 mbar						
RATPAC-B	0.15	0.07	0.11	0.09	0.24	0.03
RATPAC-A	0.19	0.07	0.13	0.12	0.26	0.01
IGRA	0.12	0.06	0.09	0.07	0.20	0.02
CI	(0.15)	(0.08)	(0.11)	(0.11)	(0.18)	(0.09)
300–100 mbar						
RATPAC-B	–0.07	–0.22	–0.14	–0.06	–0.09	–0.35
RATPAC-A	–0.01	–0.19	–0.10	–0.02	–0.05	–0.30
IGRA	–0.09	–0.24	–0.16	–0.12	–0.07	–0.35
CI	(0.19)	(0.22)	(0.19)	(0.18)	(0.19)	(0.27)
100–50 mbar						
RATPAC-B	–0.71	–0.80	–0.75	–0.83	–0.63	–0.73
RATPAC-A	–0.63	–0.82	–0.73	–0.87	–0.54	–0.62
IGRA	–0.74	–1.08	–0.88	–1.04	–0.57	–0.85
CI	(0.51)	(0.81)	(0.67)	(0.76)	(0.37)	(0.52)
<i>1960–2004</i>						
850–300 mbar						
RATPAC-B	0.15	0.15	0.15	0.16	0.17	0.13
RATPAC-A	0.17	0.15	0.16	0.17	0.18	0.12
IGRA	0.12	0.16	0.15	0.16	0.12	0.14
CI	(0.09)	(0.09)	(0.08)	(0.09)	(0.10)	(0.11)
300–100 mbar						
RATPAC-B	0.02	–0.07	–0.02	0.08	–0.06	–0.20
RATPAC-A	0.05	–0.06	–0.01	0.09	–0.05	–0.19
IGRA	–0.03	–0.15	–0.09	–0.02	–0.06	–0.27
CI	(0.05)	(0.08)	(0.06)	(0.07)	(0.07)	(0.13)
100–50 mbar						
RATPAC-B	–0.40	–0.49	–0.45	–0.51	–0.32	–0.45
RATPAC-A	–0.37	–0.50	–0.44	–0.53	–0.28	–0.41
IGRA	–0.49	–0.70	–0.60	–0.71	–0.37	–0.59
CI	(0.23)	(0.36)	(0.32)	(0.34)	(0.13)	(0.25)

^aTrend is in K/decade. The confidence interval (“CI”) is twice the standard error of the trend in RATPAC-A. Standard errors were computed with adjustment for autocorrelation of the time series. “NH EX” is Northern Hemisphere extratropics, 30°N–90°N, and “SH EX” is Southern Hemisphere extratropics, 30°S–90°S.

measure of the uncertainty introduced by the procedure. Since the FD procedure is used in only approximately one third of the time period 1979–2004, this uncertainty metric is less than that which would occur if FD were used for the entire period. We chose this measure, rather than a metric focused on the period when FD was used, because it most closely resembles the quantity of interest to most users of the data set. As the RATPAC-A data set is extended further in the future using FD, the trend uncertainty can be expected to increase.

[26] The uncertainties for annual mean trends at individual pressure levels in the hemispheres and tropics fall between 0.01 K/decade for the SH near the surface and 0.25 K/decade for the NH at 30 mbar (Figure 5 and Table S4 in Supplemental Material) and are usually a few hundredths

of a degree in the troposphere, but over 0.1 K/decade in the stratosphere. The ratio of uncertainty to trend is usually much less than one, but exceeds one for a few cases, generally where the trends themselves are very small. The uncertainties are typically less than the standard error of the trends, except for a few cases near the surface. Largest uncertainties from the FD procedure occur for the NH extratropics in the troposphere, because of the large number of metadata events for which cuts are made in this area, and for the tropics in the stratosphere.

[27] Similar tests for surface-troposphere trend differences indicate uncertainties of 0.05 K/decade for the globe and 0.08 for the tropics for 1979–2004. Uncertainties for trend differences at adjacent levels in the troposphere are smaller. For example, differences between trends at 700 and

Table 2. Trend in Surface Temperature Minus Trend in 850–300 mbar Layer Mean Temperature for 1979–2004 and 1960–2004^a

	NH	SH	GLOBE	TROPICS	NH Extra	SH Extra
<i>1979–2004</i>						
IGRA	0.17	0.05	0.11	0.11	0.22	0.00
RATPAC-B	0.06	0.07	0.06	0.03	0.06	0.11
RATPAC-A	0.04	–0.04	0.00	0.01	0.03	–0.04
<i>1960–2004</i>						
IGRA	0.09	–0.06	0.00	–0.02	0.13	–0.06
RATPAC-B	0.01	–0.05	–0.02	–0.05	0.03	–0.04
RATPAC-A	0.00	–0.09	–0.05	–0.05	0.01	–0.10

^aTrend is in K/decade.

Table 3. Least Squares Linear Trends for 1979–2004 in Seasonal Mean Time Series From RATPAC-A for Stratospheric Layer Means (100–50 mbar)^a

	DJF	MAM	JJA	SON
NH	−0.60 (0.47)	−0.58 (0.33)	−0.62 (0.54)	−0.72 (0.48)
SH	−1.12 (0.79)	−0.65 (0.46)	−0.69 (0.48)	−0.84 (0.86)
GLOBE	−0.86 (0.65)	−0.61 (0.42)	−0.65 (0.54)	−0.78 (0.70)
TROPICS	−1.08 (0.75)	−0.58 (0.44)	−0.71 (0.67)	−1.14 (0.77)
NH extratropics	−0.24 (0.38)	−0.64 (0.33)	−0.62 (0.47)	−0.60 (0.34)
SH extratropics	−1.04 (0.66)	−0.65 (0.43)	−0.58 (0.48)	−0.24 (0.65)

^aTrends are in K/decade. Numbers in parentheses are two times the standard error of the trends.

500 mbar had estimated uncertainties of 0.02–0.04 K/decade. The data set therefore appears suitable for analysis of changes in lapse rates within the troposphere.

[28] On the basis of similar tests with reanalysis data, trends for seasonal layer means show uncertainties generally larger than those for annual pressure level data, reaching 0.18–0.51 in the stratosphere in boreal fall (Table S5 in Supplemental Material). As with the annual pressure level data, trend uncertainties in other seasons are generally largest for NH and NH extratropics where the most metadata cuts are made. In the troposphere, seasonal trend uncertainties are no more than 0.07 K/decade for regions other than the NH extratropics and NH. For trends in individual months, uncertainties estimated by this method (not shown) are larger, up to 0.20 for global 850–300 mbar layer means. Because of these high uncertainties, we do not include monthly data in the RATPAC-A data set.

[29] In general, uncertainties are larger for the NH than the SH, presumably because of the greater number of metadata events there. In the SH, lack of metadata limits the number of cuts and therefore the estimated uncertainty from the FD procedure, but the lack of metadata itself creates uncertainty of an unknown size. Because of this lack of metadata, as well as the limited number of stations in some parts of the SH, overall uncertainty could well be greater in the SH than in the NH.

[30] Another possible source of uncertainty in the FD method is the sensitivity to methodological details such as the endpoint outlier trimming procedure described in section 3.2.3 above. Different trimming choices can change the resulting trends by up to 0.06 K/decade in the troposphere and 0.07 in the stratosphere (Figure 6), with the largest effect in the NH. In most cases these differences are small in comparison with the trends, so that choice of trim parameter does not appear to be a major source of uncertainty.

[31] Results from the FD procedure will also be sensitive to the timing of cuts made for metadata events. Because the metadata is out of date, incomplete or unclear for the majority of the stations in the LKS network, many changes have undoubtedly occurred that are not accounted for in our procedure. At least half of the discontinuities found by the LKS team and the HadAT team were not supported by available metadata. This suggests that metadata problems could significantly reduce the ability of our procedure to reduce inhomogeneities. The effect of this error has not been quantified, but could exceed the other uncertainties described in this subsection.

5.2. Limited Spatial Coverage

[32] Previous work has shown the potential importance of spatial sampling issues in results from other radiosonde data

sets [e.g., Trenberth and Olsen, 1991; Santer et al., 1999]. Because of similar questions about the adequacy of spatial coverage in the 85-station RATPAC network, we considered expanding the network to include additional carefully selected stations. Our original plan was to use the first difference method applied to data from stations with good station history metadata, few metadata events, and relatively complete records. We hoped that the first difference method would allow incorporation of new stations without the arduous work that had already been done for the original LKS stations. However, we found that in most of the apparent voids in the network, such as the Northern Pacific ocean, no stations with long records and good metadata exist, so opportunities to improve the gross spatial coverage are limited. Thirty-five potential new stations were identified on the basis of data and metadata archives. Most were in the Northern Hemisphere, particularly in North America

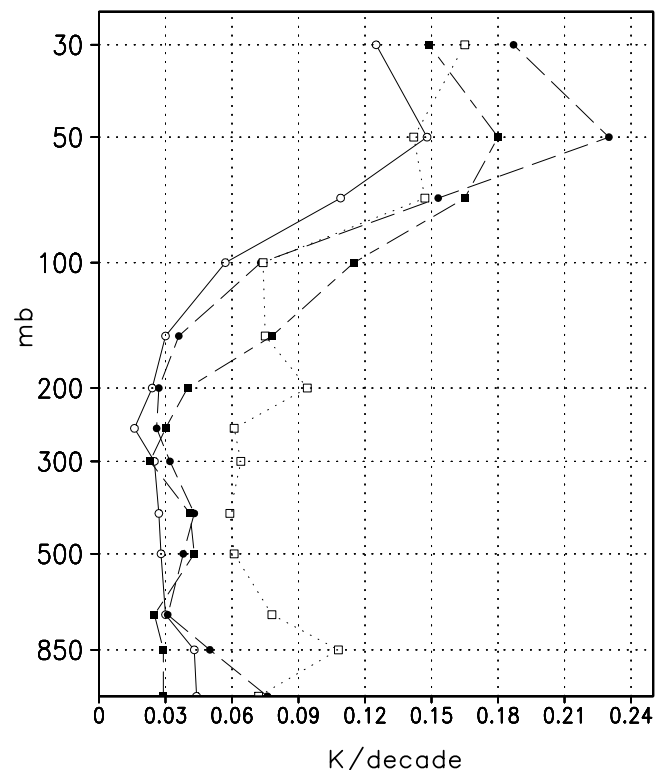


Figure 5. Estimated uncertainty in trends in annual mean temperature for 1979–2004 related to the use of the FD method for the globe (open circles), tropics (solid circles), NH extratropics (open squares) and SH extratropics (solid squares).

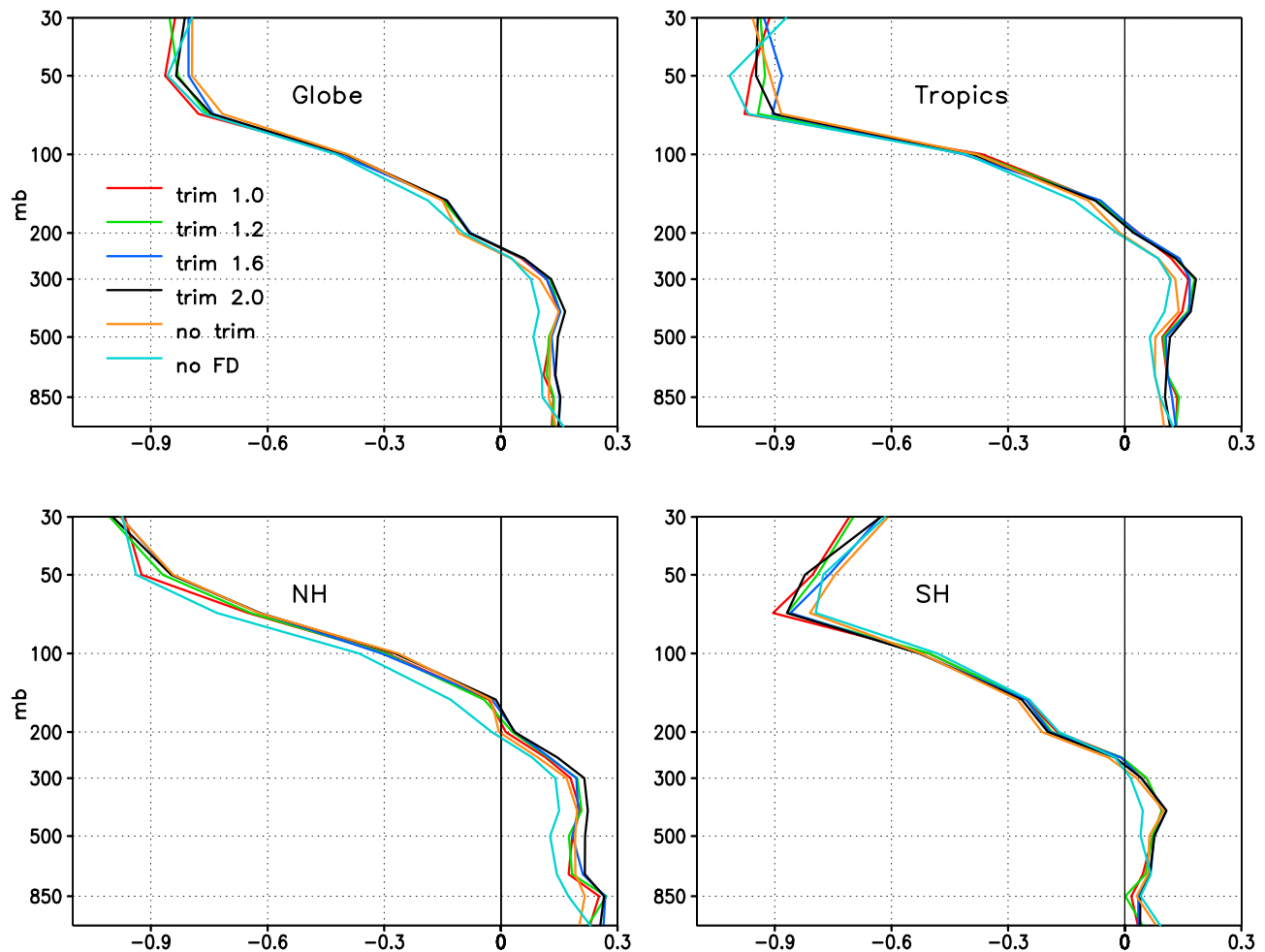


Figure 6. Least squares linear trends in annual mean temperatures for 1979–2004 from RATPAC-A using trim factors of 1.0, 1.2, 1.6 and 2.0 times the standard deviation, along with trends for series combined with FD but no trimming (“no trim”) and for series combined without using FD (“no FD”), in K/decade. The “no FD” series contain LKS adjusted data through 1995 and IGRA data afterward. Note that this differs slightly from RATPAC-B which uses LKS adjusted data through 1997.

and China. Although the new stations were in areas already covered relatively well, we hoped the additional stations could still improve the sampling error of the large-scale mean time series.

[33] We tested the effect of the proposed expansion on sampling error using the NCEP reanalysis data. For this test, the “global mean” or “regional mean” data sets created by subsampling the NCEP reanalysis data according to (1) the LKS network locations and (2) the expanded network locations (LKS plus new) were compared to (3) the global mean of the NCEP data for all points. Table 4 shows sampling error and changes in sampling error for global mean trends at four pressure levels. We measure the sampling error by the absolute value of the difference between the trend in the full global mean and the trend in the subsampled data set. The “improvement” from the expansion is the LKS sampling error minus the extended set sampling error.

[34] For global means, the extended set includes 35 additional stations. The sampling error in the trends ranges from less than 0.01 to 0.06 K/decade for 1979–1997. The

changes in sampling error with the addition of extension stations are no more than 0.04 K/decade, and are slightly negative in four of the eight cases (implying a degradation of the result with the expanded network in those cases, although the change is statistically insignificant). Most of these sampling errors and their differences are much smaller than the ~ 0.05 standard error of the trends. These trend comparisons do not show a consistent or significant improvement from adding the additional stations to the LKS data set. Results from the NH and tropics confirmed this conclusion. This is consistent with the findings of *Free and Seidel* [2005] showing little or no improvement in large-scale sampling error for upper air networks of greater than 100 unevenly spaced stations in comparison to the LKS 87-station network.

[35] For comparison, we used the reanalysis data to estimate the error that could be introduced by the first difference method if we used it to extend the data set by 35 new stations, in a test like those described in section 5.1 above. We used two randomly timed cuts for each of the new series that had metadata events in real life. The results

Table 4. Spatial Sampling Error in LKS and Extended Networks, as Measured by the Difference Between the Trend in the Complete Global Mean Reanalysis Data Set and the Trend in the Smaller Network, Along With the Change in That Error From Expansion of the Network From 85 LKS Stations to the 120-Station Extended Network (“Improvement”)^a

	1960–1997			1979–1997		
	LKS Error ^b	Extended Error ^b	Improvement ^c	LKS Error ^b	Extended Error ^b	Improvement ^c
50 mbar	0.020	0.022	−0.002	0.030	0.029	0.001
200 mbar	0.044	0.060	−0.016	0.016	0.053	−0.037
500 mbar	0.003	0.000	0.003	0.032	0.014	0.018
850 mbar	0.040	0.042	−0.002	0.045	0.036	0.009

^aError is in K/decade.

^bAbsolute value of subsampled trend minus full global trend.

^cLKS sampling error minus extended sampling error.

indicated potential errors of up to 0.02 K/decade from the procedure.

[36] We also examined the root-mean-square differences between the subsampled and full annual time series. Like the trends, these showed no consistent improvement in the error for the extended data set as compared to the LKS network.

[37] These estimates suggest that the error in global and hemispheric temperature would likely be no better in the expanded network than in the LKS network. This finding is consistent with results from *Trenberth and Olsen* [1991] suggesting that the *Angell and Korshover* [1975] network of only 63 stations is reasonably adequate to define global and hemispheric upper air temperature variations. On the basis of our subsampling results, the potential for additional error from first differencing, and the fact that the candidate extension stations are in regions that are already reasonably well sampled by the LKS network, we concluded that the RATPAC data set would probably be more accurate without the additional stations.

5.3. Variations in Spatial Coverage

[38] The actual number of stations whose data is used varies from month to month because of gaps in the original data, the cuts made at the dates of metadata events, and the effects of endpoint outlier trimming. Figure 7 shows the number of stations contributing to the global, NH and SH means for each month from 1996 to 2004, for the original data and the RATPAC time series. The total declines over time because some stations in the LKS network have stopped reporting (see section 3.1 above). The metadata events for many stations are clustered around 1998 and 1999, contributing to a drop in the station count for those years. The IGRA input data also have gaps for many stations during this time period. The effect of endpoint outlier trimming on data availability is small for most months.

5.4. Residual Inhomogeneities in LKS Data

[39] As recognized by *Lanzante et al.* [2003a, 2003b], the LKS adjustment process undoubtedly missed some significant inhomogeneities. LKS showed that the adjustments generally reduced the differences between the radiosonde trends and trends from the *Christy et al.* [2003] satellite data for individual stations, but did not eliminate them. Since this work, two alternate satellite temperature data sets [*Mears et al.*, 2003; *Vinnikov and Grody*, 2003] have been published that suggest even larger differences between sonde and satellite trends. Recent work by other authors comparing

sonde data to satellite data [*Randel and Wu*, 2005] and examining day-night differences in sonde data [*Sherwood et al.*, 2005] suggests that the trend effects of remaining inhomogeneities in LKS and other adjusted radiosonde data

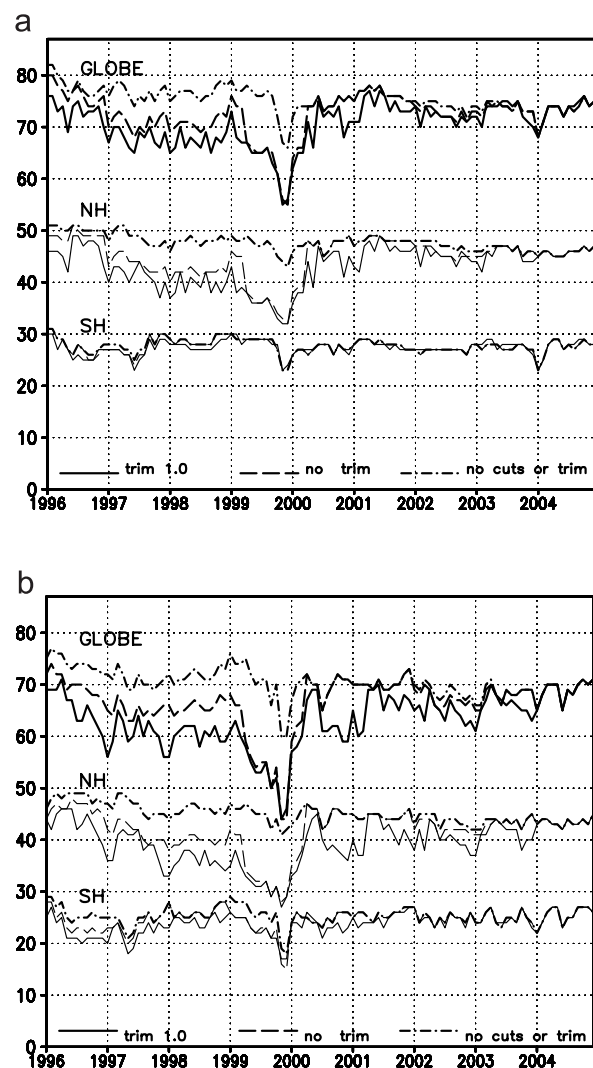


Figure 7. (a) Number of stations contributing to the global, NH and SH means at 500 mbar for each month from 1996 to 2004, for the original data without cutting or trimming (“no cuts or trim”), the original series processed using FD without endpoint outlier trimming (“no trim”) and the RATPAC-A time series with trim factor 1.0 (“trim 1.0”). (b) As in Figure 7a but for 50 mbar.

sets could be as large as the effects of the LKS adjustments. Thus shortcomings in the adjusted LKS data set are a significant possible source of additional uncertainty in the RATPAC products. However, RATPAC trends are generally similar [Santer *et al.*, 2005] to those in the HadAT radiosonde data set, which was constructed with a different adjustment approach. We believe that the current RATPAC product, despite its limitations, is as reliable as any available at this time.

6. Conclusion

[40] We have constructed large-scale mean time series of upper air temperatures suitable for climate change analysis, extending from 1958 through 2004. The RATPAC-A data set is based on the LKS homogeneity-adjusted data through 1995, extended using a first difference (FD) procedure that reduces potential inhomogeneities inferred from changes present in station metadata. Because this procedure can produce only large-scale means, we also extended the individual station data without adjustments after 1997 and, for comparison to RATPAC-A, created large-scale means without the FD procedure (RATPAC-B). For analysis of interannual and longer-term changes in global, hemispheric and tropical means, we recommend RATPAC-A. For individual station data, monthly data, or regional means on smaller scales, we recommend use of RATPAC-B, with careful attention paid to the potential of inhomogeneities impacting analysis after 1997. We also advise caution in use of either RATPAC product prior to 1965.

[41] Comparison of RATPAC with other existing and future data sets will give additional insight into climate changes above the surface and their uncertainties. Global mean trends for 1979–2004 from RATPAC-A (created using FD) in the troposphere are ~ 0.13 K/decade, and slightly less in the tropical troposphere. Trends in the stratosphere are approximately -0.5 to -0.9 K/decade for the same period and are more negative in the tropics than elsewhere. Trends in RATPAC-B (created without FD) are generally slightly less positive than those in RATPAC-A, except in the SH. For the NH, globe and tropics, the adjustment procedure used by LKS before 1995 and the FD procedure used after 1995 reduce the surface-troposphere trend differences found in the unadjusted IGRA data for the satellite period. For the globe and SH extratropics, the tropospheric trends in RATPAC-A are greater than or equal to those at the surface.

[42] Subsampling experiments using reanalysis data indicate that if reasonable requirements for data completeness and metadata quality are imposed, adding qualified stations to the LKS network using FD is unlikely to improve the error characteristics of the resulting large-scale mean time series. Uncertainty introduced by the FD method as estimated by Monte Carlo experiments is 0.02 – 0.04 K/decade in the troposphere and up to 0.15 K/decade in the stratosphere for the global annual mean trend for 1979–2004. Uncertainties for smaller regions are generally larger. Estimates of uncertainty arising from the FD procedure indicate that trends from RATPAC-A for layer means are more reliable than those for individual levels, and those for annual mean time series are more reliable than seasonal or

monthly values [see also Free *et al.*, 2004]. Sensitivity to choice of trim parameter seems to be a less important source of uncertainty as compared to the random error from the FD process. On the basis of these factors, we provide annual time series for individual pressure levels and seasonal time series for layer means, but no monthly time series and no seasonal data for pressure levels for the RATPAC-A data set.

[43] Results are also sensitive to timing of metadata cuts and our metadata are incomplete and in some cases ambiguous. The future usefulness of our method will depend critically on continuing efforts to expand, clarify and update the radiosonde metadata.

[44] The data set will be updated monthly at NCDC and included in NOAA's State of the Climate reporting and other operational efforts to monitor variations and trends in the global climate. These reports are provided online at <http://www.ncdc.noaa.gov/oa/climate/research/monitoring.html>, and the data will be available at <http://www.ncdc.noaa.gov/oa/climate/ratpac>.

Appendix A: The First Difference Method

[45] The method used to combine the station time series into the RATPAC-A global, hemispheric, tropical and extratropical means, described in more detail by Free *et al.* [2004], involves the following steps: (1) Identify dates of potential inhomogeneities from metadata. (2) (Optional) Interpolate to fill gaps in the input station data. (3) From each station time series, drop data for 6 months before and after each potential inhomogeneity. (4) (Optional) Remove data for months immediately before and after the cuts made in step 3 if the temperature for that month exceeds a specified multiple of the standard deviation ("endpoint outlier trimming"). (5) Create FD time series from these series by calculating differences in temperature from one year to the next for each of the 12 calendar months. (6) Combine the station FD series into large-scale mean FD series by averaging the station FD values. (7) Construct temperature anomaly series from the large-scale mean FD series by cumulatively summing the FD data forward in time.

[46] Within each geographical region used for the spatial averaging (see section 3.2.2 above), we combined the IGRA time series that had no suspected inhomogeneities using biweight means [Lanzante, 1996] and took the FD of the resulting mean time series. This reduces the random error effect that would be produced by applying FD to each series individually because fewer missing months of data will be present in the means than in the individual station time series. The station time series with suspect metadata events were first-differenced individually and those FD series were combined with the FD of the mean of the uncut time series (weighted by the number of uncut stations) to give large-scale mean FD series. The result was appended to the LKS data by cumulatively summing forward from 1995 to 2004.

[47] **Acknowledgments.** This work was funded by NOAA's Office of Global Programs. We thank the anonymous reviewers for helpful comments.

References

- Angell, J. K., and J. Korshover (1975), Estimate of the global change in tropospheric temperature between 1958 and 1973, *Mon. Weather Rev.*, *103*, 1007–1012.
- Christy, J. R., R. W. Spencer, W. B. Norris, W. D. Braswell, and D. E. Parker (2003), Error estimates of version 5.0 of MSY/AMSU bulk atmospheric temperatures, *J. Atmos. Oceanic Technol.*, *20*, 613–629.
- Durre, I., R. Vose, and D. Wertz (2004), Overview of the Integrated Global Radiosonde Archive, *J. Clim.*, 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. Am. Meteorol. Soc.*, *76*, 1759–1775.
- Free, M., and J. K. Angell (2002), Effect of volcanoes on the vertical temperature profile in radiosonde data, *J. Geophys. Res.*, *107*(D10), 4101, doi:10.1029/2001JD001128.
- Free, M., and D. J. Seidel (2005), Causes of differing temperature trends in radiosonde upper air datasets, *J. Geophys. Res.*, *110*, D07101, doi:10.1029/2004JD005481.
- Free, M., et al. (2002), Creating climate reference datasets: CARDS workshop on adjusting radiosonde temperature data for climate monitoring, *Bull. Am. Meteorol. Soc.*, *83*, 891–899.
- Free, M., J. K. Angell, I. Durre, J. Lanzante, T. C. Peterson, and D. J. Seidel (2004), Using first differences to reduce inhomogeneity in radiosonde temperature datasets, *J. Clim.*, *21*, 4171–4179.
- Gaffen, D. J. (1994), Temporal inhomogeneities in radiosonde temperature records, *J. Geophys. Res.*, *99*, 3667–3676.
- Gaffen, D. J., B. D. Santer, J. S. Boyle, J. R. Christy, N. E. Graham, and R. J. Ross (2000), Multi-decadal changes in the vertical temperature structure of the tropical troposphere, *Science*, *287*, 1239–1241.
- Haimberger, L. (2005), Homogenization of radiosonde temperature time series using ERA-40 analysis feedback information, *ERA-40 Proj. Rep. Ser.*, *22*, 73 pp., Eur. Cent. for Med.-Range Weather Forecasts, Reading, U. K.
- International Ad Hoc Detection and Attribution Group (2005), Detecting and attributing external influences on the climate system: A review of recent advances, *J. Clim.*, *18*, 1291–1314.
- Kistler, R., et al. (2001), The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation, *Bull. Am. Meteorol. Soc.*, *82*, 247–267.
- Lanzante, J. R. (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.
- Lanzante, J. R., S. A. Klein, and D. J. Seidel (2003a), Temporal homogenization of monthly radiosonde temperature data. Part I: Methodology, *J. Clim.*, *16*, 224–240.
- Lanzante, J. R., S. A. Klein, and D. J. Seidel (2003b), Temporal homogenization of monthly radiosonde temperature data. Part II: Trends, sensitivities, and MSU comparison, *J. Clim.*, *16*, 241–262.
- Mears, C. A., M. C. Schabel, and F. J. Wentz (2003), A reanalysis of the MSU channel 2 tropospheric temperature record, *J. Clim.*, *16*, 3650–3664.
- National Research Council (2000), *Reconciling Observations of Global Temperature Change*, 85 pp., Natl. Acad. Press, Washington, D. C.
- Parker, D. E., 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.
- Peterson, T. C., T. R. Karl, P. F. Jamason, R. Knight, and D. R. Easterling (1998), First difference method: Maximizing station density for the calculation of long-term global temperature change, *J. Geophys. Res.*, *103*, 25,967–25,974.
- Randel, W. J., and F. Wu (2005), Biases in stratospheric temperature trends derived from historical radiosonde data, *J. Clim.*, in press.
- Santer, B. D., et al. (1996), A search for human influences on the thermal structure of the atmosphere, *Nature*, *382*, 39–46.
- Santer, B. D., et al. (1999), Uncertainties in observationally based estimates of temperature change in the free atmosphere, *J. Geophys. Res.*, *104*, 6305–6333.
- Santer, B. D., et al. (2000), Interpreting differential temperature trends at the surface and in the lower troposphere, *Science*, *287*, 1227–1232.
- Santer, B. D., et al. (2005), Amplification of surface temperature trends and variability in the tropical atmosphere, *Science Online*, doi:10.1126/science.1114867.
- Seidel, D. J., et al. (2004), Uncertainty in signals of large-scale climate variations in radiosonde and satellite upper-air temperature datasets, *J. Clim.*, *17*, 2225–2240.
- Sherwood, S. C., J. Lanzante, and C. Meyer (2005), Radiosonde daytime biases and late 20th century warming, *Science Online*, doi:10.1126/science.1115640.
- 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.
- Thompson, D. W. J., and S. Solomon (2005), Recent stratospheric climate trends: Global structure and tropospheric linkages, *J. Clim.*, in press.
- Thorne, P. W., P. D. Jones, T. J. Osborn, T. Davies, S. F. B. Tett, D. E. Parker, P. Stott, G. Jones, and M. Allen (2002), Assessing the robustness of zonal mean climate change detection, *Geophys. Res. Lett.*, *29*(19), 1920, doi:10.1029/2002GL015717.
- Thorne, P. W., D. E. Parker, S. F. B. Tett, P. D. Jones, M. McCarthy, H. Coleman, P. Brohan, and J. R. Knight (2005), Revisiting radiosonde upper-air temperatures from 1958 to 2002, *J. Geophys. Res.*, *110*, D18105, doi:10.1029/2004JD005753.
- Trenberth, K. E., and J. G. Olsen (1991), Representativeness of a 63-station network for depicting climate changes, in *Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations*, edited by M. Schlesinger, pp. 249–260, Elsevier, New York.
- Vinnikov, K. Y., and N. C. Grody (2003), Global warming trend of mean tropospheric temperature observed by satellites, *Science*, *302*, 269–272.

J. K. Angell, M. Free, and D. J. Seidel, NOAA Air Resources Laboratory, Silver Spring, MD 20910, USA. (melissa.free@noaa.gov)

I. Durre and T. C. Peterson, NOAA National Climatic Data Center, Asheville, NC 28801-5001, USA.

J. Lanzante, NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ 08540-6649, USA.