Causes of differing temperature trends in radiosonde upper air data sets

Melissa Free and Dian J. Seidel
Air Resources Laboratory, NOAA, Silver Spring, Maryland, USA

Received 30 September 2004; revised 26 January 2005; accepted 10 February 2005; published 6 April 2005.

[1] Differences between trends in different radiosonde temperature products resulting from the varying choices made by the developers of the data sets create obstacles for use of those products in climate change detection and attribution. To clarify the causes of these differences, one must examine results using a common subset of locations to minimize spatial sampling effects. When this is done for the Lanzante-Klein-Seidel (LKS) and Hadley Center (HadRT) radiosonde data sets, differences are reduced by at least one third. Differing homogeneity adjustment methods and differences in the source data are both important factors contributing to the remaining discrepancies. In contrast, subsampling the microwave sounding unit (MSU) satellite data sets according to the radiosonde coverage does not generally bring the trends in the satellite data closer to those in the radiosonde data so that adjustments and other processing differences appear to be the predominant sources of satellite-radiosonde discrepancies. Experiments in which we subsample globally complete data sets provide additional insight into the role of sampling errors. In the troposphere, spatial sampling errors are frequently comparable to the trends for 1979–1997, while in the stratosphere the errors are generally small relative to the trends. Sampling effects estimated from National Centers for Environmental Prediction reanalysis and MSU satellite data for seven actual radiosonde networks show little consistent relation between sampling error and network size. These results may have significant implications for the design of future climate monitoring networks. However, estimates of sampling effects using the reanalysis and the satellite data sets differ noticeably from each other and from effects estimated from actual radiosonde data, suggesting that these globally complete data sets may not fully reproduce actual sampling effects.

Citation: Free, M., and D. J. Seidel (2005), Causes of differing temperature trends in radiosonde upper air data sets, *J. Geophys. Res.*, *110*, D07101, doi:10.1029/2004JD005481.

1. Introduction

[2] Several radiosonde temperature data sets have been created in recent years, including the *Angell* [2003], Lanzante-Klein-Seidel (LKS) [*Lanzante et al.*, 2003a, 2003b], and Hadley Center (HadRT) [*Parker et al.*, 1997] data sets. These temperature records are important for climate change detection and attribution studies [e.g., *Santer et al.*, 1996; *Tett et al.*, 1996; *Thorne et al.*, 2002]. Unfortunately, the data sets do not give the same trends for large-scale means on multidecadal timescales. *Seidel et al.* [2004] compared trends and other signals from upper air data sets using layer mean and microwave sounding unit (MSU) satellite equivalent temperatures, showing that the data sets give relatively consistent values for the El Niño–Southern Oscillation, volcanic, and quasi-biennial oscillation signals but more widely varying results for trends. Differences between temperature trends in the tropics derived from the HadRT and LKS radiosonde data sets for a midtropospheric layer corresponding to that measured by the MSU channel 2 satellite product were 0.1 K decade^{-1} (-0.132 for HadRT versus -0.032 for LKS) for 1979–1997. Global radiosonde temperature trends differed from actual MSU satellite trends by 0.16–0.31 K decade^{-1} in the stratosphere, or 26–51% of the mean trend. Trends in HadRT2.1s and LKS adjusted data plotted by pressure level also show marked differences (Figure 1) not only in magnitude but even in sign. Moreover, HadRT shows trends that are more negative with increasing altitude while LKS shows more warming with height up to 300 mbar in the tropics. These differences may have important implications for detection of long-term climate change and attribution of its causes. They contribute to uncertainties in vertical differences in temperature trends, which are currently the subject of considerable scientific interest. How can we account for them?

[3] Possible sources of these differences include spatial sampling differences, temporal sampling differences, differences in the original data, adjustments for inhomogeneities, and other differences in processing. In this paper we assess

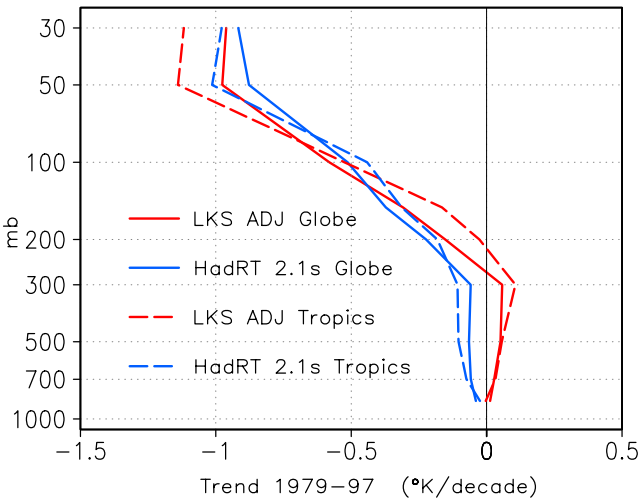

Figure 1. Least squares linear trends (K decade^{-1}) in global and tropical (30°S–30°N) mean radiosonde temperature data from HadRT2.1s and LKS adjusted data sets for 1979–1997 at the nine standard HadRT levels (850, 700, 500, 300, 200, 150, 100, 50, and 30 mbar).

the relative contribution of each to the differences between radiosonde data sets and examine the effect of sampling differences on the trend results of *Seidel et al.* [2004].

2. Data Sets and Station Networks Used

[4] We used data from two radiosonde temperature products, HadRT and LKS. The first is derived from CLIMAT TEMP reports of monthly mean temperatures calculated in the countries where the observations are made. Most radiosonde stations make observations once or twice daily, and the CLIMAT TEMP data are a mixture of all times available. The LKS data set is derived from individual radiosonde soundings in the Comprehensive Aerological Reference Data Set (CARDS) archive [*Eskridge et al.*, 1995]. In some cases both observation times were used; at other stations, only one time (typically day) was used; and at a few stations, different observation times were used during different time periods. The two data sets may thus differ in the observation times used at individual stations, as well as in the methods used for quality control and homogeneity adjustment. The HadRT2.1 data have been adjusted for changes in instruments and procedures after 1979. HadRT2.1s has been adjusted only above 200 mbar. The LKS data set uses a different method and removes inhomogeneities at all levels in the atmosphere. In addition, HadRT data sets are gridded products, while LKS consists of individual station data.

[5] For sampling error studies we also examined the station networks used in the Angell and HadAT data sets and the Global Climate Observing System (GCOS) Upper Air Network (GUAN). (The HadAT gridded data were not yet available when this work was done.)

[6] To assess sampling error, we used gridded monthly mean temperatures from two data sets with globally complete coverage: the *Christy et al.* [2003] microwave satellite data (University of Alabama at Huntsville (UAH) MSU) and the National Centers for Environmental Pre-

Table 1. Radiosonde Station Sets Used in This Work[a]

Name	Number of Stations	Reference
A 54	54	*Angell* [2003]
A 63	63	*Angell and Korshover* [1975]
LKS	87	*Lanzante et al.* [2003a]
GUAN	152	*World Meteorological Organization* [1996]
HadRT	444	*Parker et al.* [1997]
HadAT1	477	P. Thorne (personal communication, 2003)
HadAT2	676	P. Thorne (personal communication, 2003)

[a]HadAT denotes stations used as input for the future HadAT radiosonde temperature product as of the time of this work. GUAN is Global Climate Observing System (GCOS) Upper Air Network.

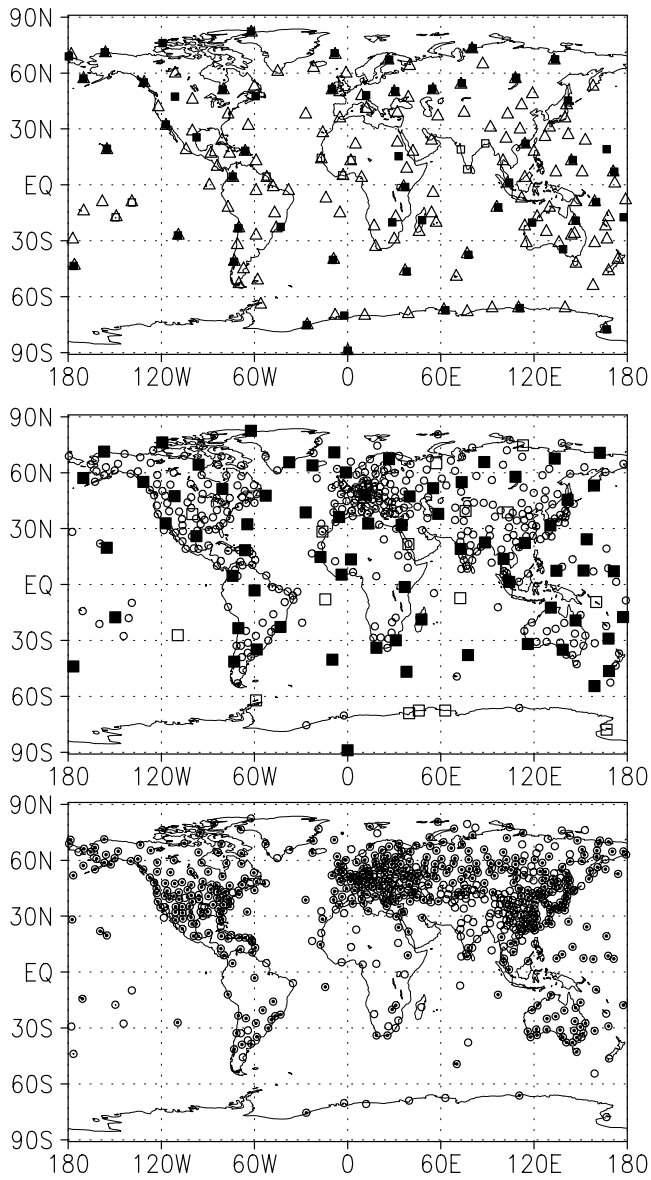

Figure 2. Locations of stations for networks listed in Table 1. (top) Angell 54 (solid squares) and 63 (open squares) and Global Climate Observing System (GCOS) Upper Air Network (open triangles). (middle) LKS (squares) and HadRT (circles). Solid squares denote 71 locations common to both LKS and HadRT. (bottom) HadAT1 (small solid circles) and HadAT2 (open circles).

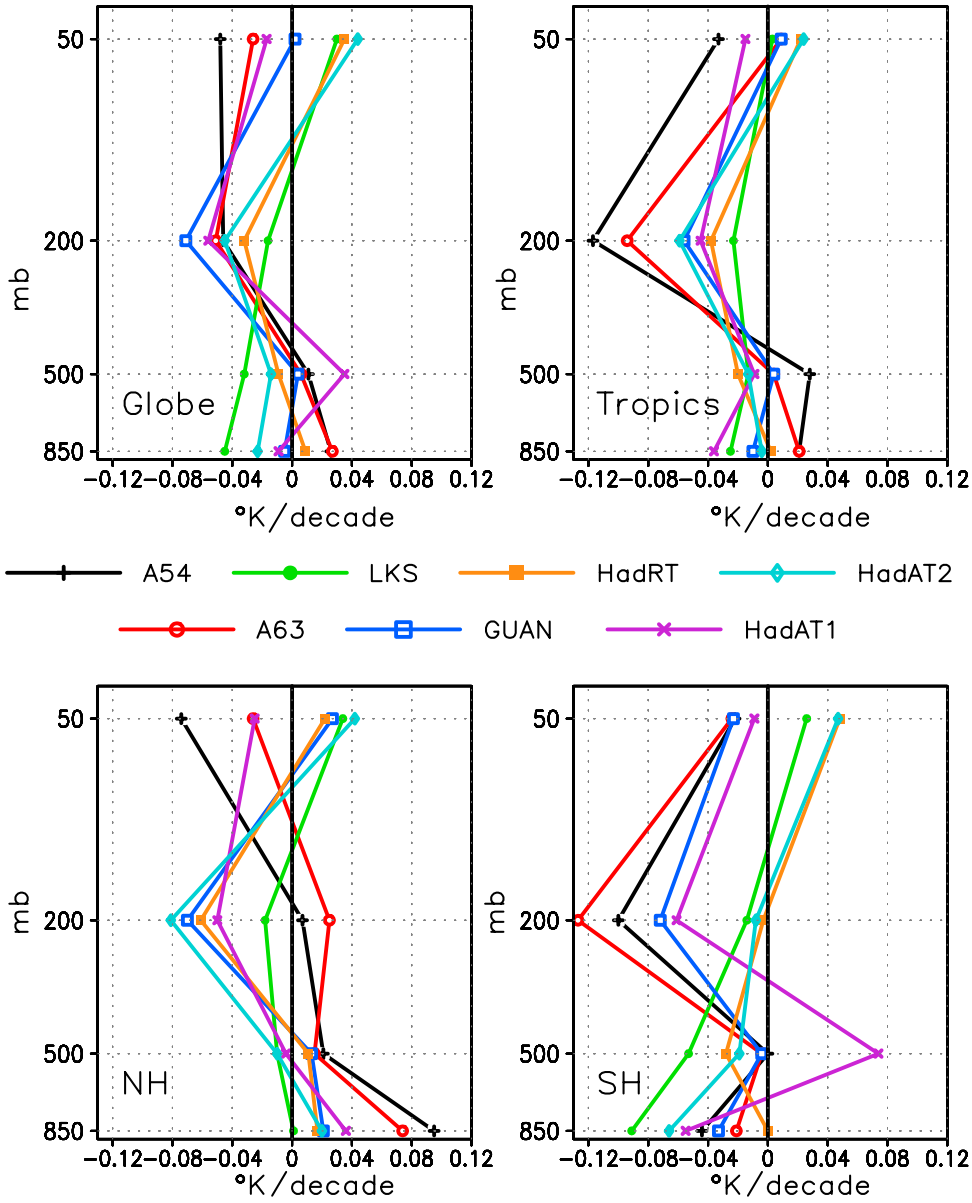

Figure 3. Spatial sampling error in trends as estimated from National Centers for Environmental Prediction (NCEP) reanalysis: trends (K decade^{-1}) in reanalysis data subsampled according to the locations of stations in the networks in Table 1 minus trends in the full gridded data set for 1979–1997 in four regions.

diction (NCEP)/National Center for Atmospheric Research reanalysis [*Kalnay et al.*, 1996]. In section 4.1 we also show results from the *Mears et al.* [2003] satellite data set (Remote Sensing Systems, Inc. (RSS)), derived from the same satellite observations as Christy et al. but using different processing. While radiosonde data sets are derived from point measurements at specific locations and discrete pressure levels, the satellite observations represent emissions over a large horizontal and vertical area. Sonde data sets include at least nine atmospheric levels, while the satellite data available for the last 23 years are reported for only two or three broad vertical layers. For comparison purposes we translate sonde data into layer means approximately equivalent to the MSU layers using a globally uniform weighting function as in *Seidel et al.* [2004], applied to the temperatures at pressure levels available in the radiosonde data sets. The results are referred to herein as MSU equivalent layer means. Because the satellite data have much less vertical and horizontal resolution, they would not be expected to show the same spatial and temporal variability as the radiosonde observations.

[7] The reanalysis data set is derived from in situ and satellite observations combined using a numerical weather forecasting model. As with the satellite data, gridded data in the reanalysis represent the mean of conditions within an area spanning several degrees latitude and longitude, rather than the single point represented by radiosonde data. The use of several data sources and the influence of the model might be expected to reduce

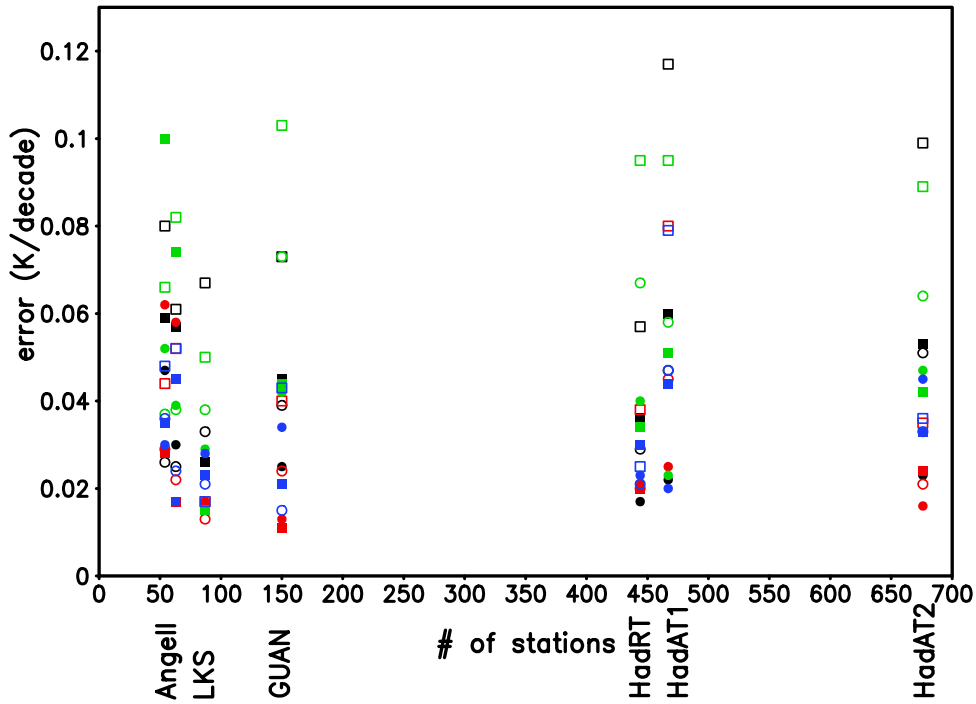

Figure 4. Means of spatial sampling errors (K decade^{-1}) in trends for 25 twenty-year segments between 1958 and 1977 and 1982 and 2001 for time series subsampled from NCEP reanalysis to simulate seven actual radiosonde networks. Means for global (open circles), Northern Hemisphere (solid circles), Southern Hemisphere (open squares), and tropical (solid squares) mean time series are plotted as a function of the number of locations in the subset. Black symbols are errors at 850, red symbols are errors at 500, green symbols are errors at 200, and blue symbols are errors at 50 mbar.

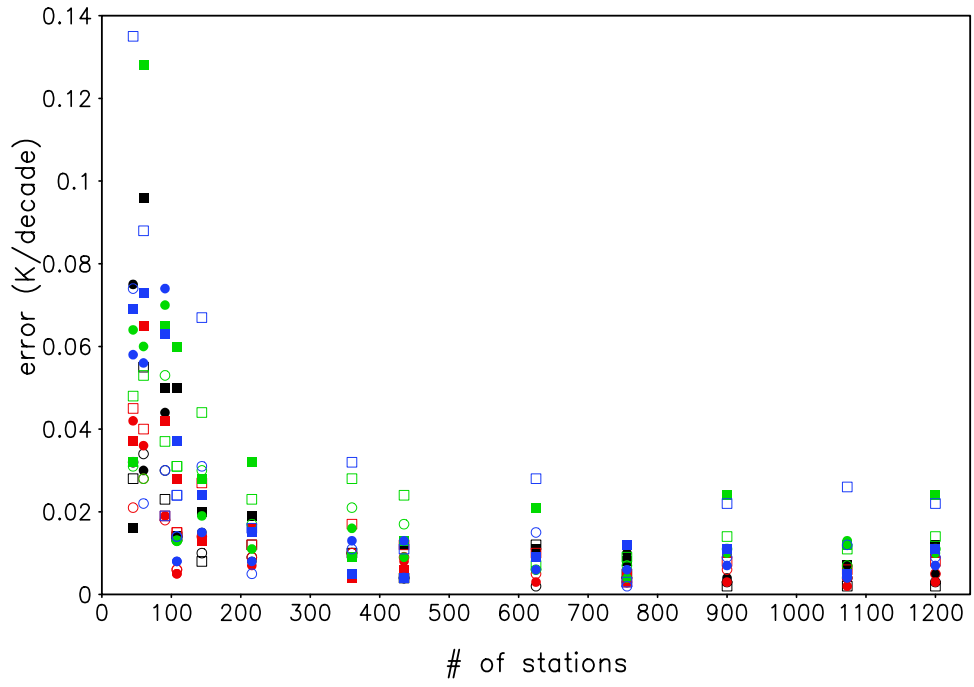

Figure 5. As in Figure 4 but for hypothetical networks with approximately evenly spaced locations.

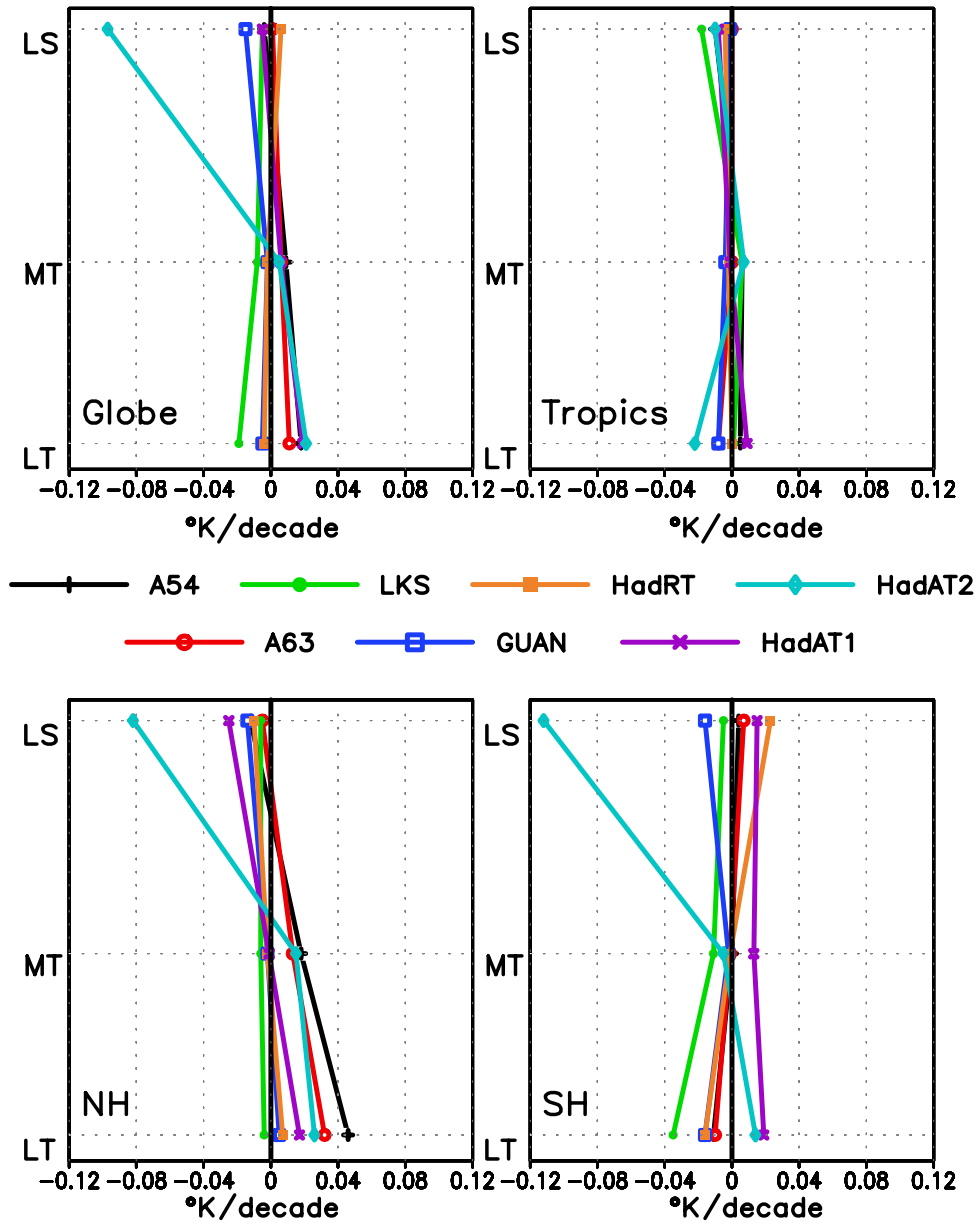

Figure 6. As in Figure 3 but using trends from microwave sounding unit (MSU) temperature data for 1979–1997 and MSU atmospheric layers.

noise in the observations so that, like the satellite data, the reanalysis temperatures are not expected to contain all of the spatial and temporal variability of the radiosonde data.

3. Sampling Error

3.1. Previous Work

[8] Several earlier studies have subsampled globally complete data sets according to radiosonde network coverage to assess sampling error. Using Geophysical Fluid Dynamics Laboratory model output, *Oort* [1978] estimated that RMS errors in monthly mean global tropospheric upper air temperatures due to spatial sampling gaps at individual pressure levels were 0.5–1.0 K for an 855-station network. *Oort* [1978] concluded that the network was generally adequate for determining large-scale circulation statistics and trends in the Northern Hemisphere (NH) but not in the Southern Hemisphere (SH). *Trenberth and Olson* [1991] used European Centre for Medium-Range Weather Forecasts (ECMWF) operational analyses subsampled according to the Angell 63-station network [*Angell and Korshover*, 1975]. They found that RMS differences between complete and subsampled seasonal mean layer mean series in the troposphere were 0.07–0.11 K for global means, 0.10–0.20 for the NH, and 0.08–0.20 for the SH.

[9] In more recent work, *Santer et al.* [1999] compared the effects of coverage differences on trends for the globe, NH, and SH for MSU equivalent layers using the coverage of the HadRT data set applied to NCEP and ECMWF reanalysis (ERA) [*Gibson et al.*, 1997] and MSU satellite data. For global mean trends from 1979 to 1993 the differ-

Table 2. Total Sampling Error (in Time and Space) in LKS and HadRT Networks Estimated From Reanalysis[a]

	1979–1997					1960–1997				
Millibars	Globe	NH	SH	30°S–30°N	20°S–20°N	Globe	NH	SH	30°S–30°N	20°S–20°N
					LKS					
50	0.057	0.017	0.097	0.050	0.073	−0.033	0.036	−0.101	0.013	−0.002
200	−0.022	−0.018	−0.026	−0.041	−0.045	−0.061	−0.023	−0.101	−0.022	−0.021
500	−0.028	−0.011	−0.046	−0.021	−0.010	−0.025	0.013	−0.064	0.009	0.020
850	−0.049	−0.002	−0.098	−0.017	0.018	−0.083	−0.003	−0.163	−0.077	−0.064
					HadRT					
50	0.067	0.045	0.089	0.018	−0.007	−0.023	0.031	−0.076	0.008	−0.007
200	−0.041	−0.080	−0.002	−0.077	−0.042	−0.079	−0.043	−0.115	−0.059	−0.023
500	−0.010	−0.005	−0.014	−0.038	−0.034	−0.008	0.026	−0.044	0.012	0.009
850	0.002	0.018	−0.015	−0.000	0.033	−0.027	0.008	−0.063	−0.043	0.006

[a]Trends in data subsampled in space and time according to the LKS and HadRT radiosonde data minus trends in full reanalysis data set (K decade^{-1}) are given. NH is Northern Hemisphere; SH is Southern Hemisphere.

ences ranged from near zero for the midtroposphere using the ERA temperatures to 0.131 K decade^{-1} (out of a trend of ~0.3) for the stratosphere. *Rosen et al.* [2003] subsampled NCEP reanalysis data to evaluate various radiosonde station networks in the United States and North America and found that differences in 500-mbar trends among the networks were no larger than 0.03 K decade^{-1} (~10–50% of the trend). In contrast, *Agudelo and Curry* [2004] found differences of up to 0.08 K decade^{-1} between tropospheric trends in full global data sets and the same data sets subsampled according to the LKS network using a nonparametric trend estimator. They concluded that the LKS network overestimated the temperature trend because of the shortage of observations over the oceans.

[10] Similar approaches have been used for surface data [e.g., *Hansen and Lebedeff*, 1987; *Karl et al.*, 1994]. An alternative approach to spatial sampling error is to estimate spatial degrees of freedom from correlation decay lengths [e.g., *Jones et al.*, 1997]. That approach is deferred for future investigation.

[11] Here we expand on previous sampling error work by using a longer time period and additional station networks, and we examine effects at individual pressure levels as well as for layer means. We compare results based on the NCEP reanalysis with those from the UAH MSU satellite record [*Christy et al.*, 2003] and from actual radiosonde data to determine sensitivity of the estimates to choice of data set. We also directly address the causes of previously identified differences in trends and compare trends in different upper air data sets using similar spatial and temporal sampling for each.

3.2. Spatial Sampling Error Estimated From Globally Complete Data Sets

[12] The usefulness of error estimates obtained by subsampling a global data set depends on the realism of the spatial and temporal variability in the globally complete data set. As discussed in section 2, the globally complete data sets report temperatures at different spatial and temporal resolutions than the radiosonde data sets. We calculated the spatial standard deviations of trends in annual mean data in the MSU and NCEP data sets using locations corresponding to 444 stations used in the HadRT data set. As might be expected, the standard deviations of the globally complete data sets are one half to two thirds of those for actual radiosonde data (which range from 0.37 K decade^{-1} at 500 mbar to 0.70 K decade^{-1} at 50 mbar). Given this difference, the error estimates from our subsampling method are likely to be smaller than actual sampling

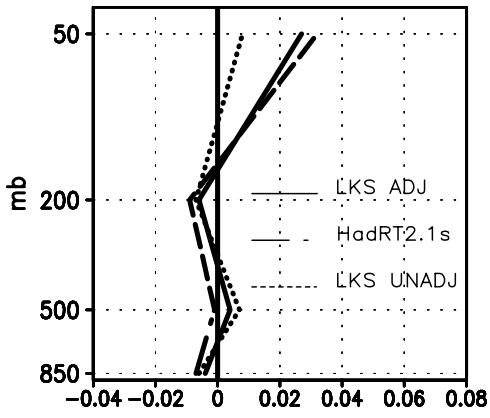

Figure 7. Effect of temporal sampling error in LKS and HadRT networks on global mean trends for 1979–1997 in reanalysis data: trend in reanalysis data subsampled in space and time according to the LKS and HadRT data sets minus trends from reanalysis data subsampled in space only (K decade^{-1}).

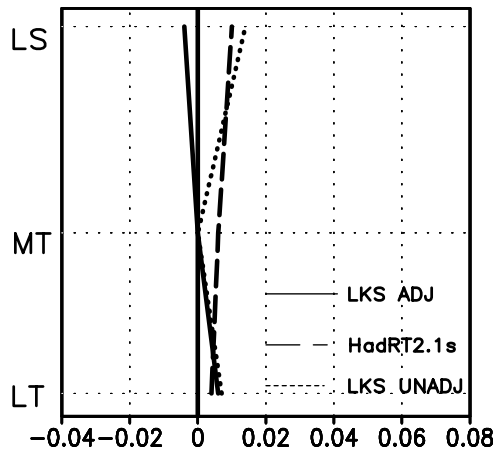

Figure 8. As in Figure 7 but using MSU rather than reanalysis data.

Table 3a. Effect of Reducing Spatial Sampling From 444 to 71 Locations as Estimated From the NCEP Reanalysis and HadRT Data Sets[a]

	Globe		NH		SH		30°S–30°N		20°S–20°N	
Level	Reanalysis	Actual	Reanalysis	Actual	Reanalysis	Actual	Reanalysis	Actual	Reanalysis	Actual
50	−0.058	−0.059	−0.031	−0.020	−0.084	−0.098	−0.018	−0.128	−0.003	−0.168
200	0.058	−0.045	0.044	−0.030	0.071	−0.059	0.025	−0.036	−0.016	−0.094
500	−0.002	−0.037	0.006	−0.024	−0.011	−0.049	0.015	−0.073	0.025	−0.134
850	−0.065	−0.008	0.009	−0.019	−0.140	0.003	−0.040	0.018	−0.044	−0.018

[a]Trends (K decade^{-1}, 1979–1997) in large-scale mean reanalysis temperatures from 444 minus those from 71 locations (columns labeled "Reanalysis") and trends in the HadRT radiosonde data for the 444 stations minus those for 71 locations (columns labeled "Actual") are given. NCEP is National Centers for Environmental Prediction.

errors. In addition, spatial patterns of temperature changes are not quite the same in the MSU and reanalysis data sets [*Agudelo and Curry*, 2004], so the error estimates are likely to differ between the two. As discussed in section 3.4, error estimates from this method also differ noticeably from those derived from actual radiosonde data.

[13] We assess large-scale spatial sampling error by subsampling NCEP reanalysis and MSU temperature data according to several radiosonde station subsets listed in Table 1. The locations of stations included in each network are shown in Figure 2. Each of these networks has been used in the past or proposed for use in the future for climate monitoring. For each network we subsampled the NCEP reanalysis data by selecting grid boxes containing the stations in the network. We then created global, hemispheric, and tropical (30°S–30°N) mean time series by averaging the selected grid boxes into 10° zonal means and combining the zonal means with weighting equal to the cosine of the latitude at the middle of each zone. We calculated linear least squares regression trends for each such mean time series and for the mean time series from the full data set, and the differences between the subsampled and full data set trends, for four atmospheric pressure levels chosen to represent the lower troposphere (850 mbar), midtroposphere (500 mbar), upper troposphere-tropopause region (200 mbar), and stratosphere (50 mbar). Although nonparametric trend estimation methods are more robust than linear regression, they are seldom used in discussions of upper air temperatures, and other studies have found little difference between least squares trends and the nonparametric median of pairwise slopes [*Gaffen et al.*, 2000; *Huth and Pokorna*, 2004].

[14] The differences between full and subsampled global mean trends for 1979–1997 (Figure 3) range from less than 0.002 K decade^{-1} for the GUAN at 50 mbar to 0.071 for the same network at 200 (or 24% of the trend from the complete reanalysis data set). At 850, 500, and 50 mbar, most differences are less than 0.05 K decade^{-1}. (For comparison, standard error of the trends in the reanalysis is ∼0.07 K decade^{-1} for 850 and 500 mbar, ∼0.2 at 200 mbar, and ∼0.5 at 50 mbar.) There are no clear overall best or worst networks and no apparent relationships between size of network and size of sampling error. The size of the errors is reasonably consistent with *Santer et al.* [1999], and the finding of relatively little difference in trends from smaller versus larger networks is consistent with *Rosen et al.* [2003] in the NH.

[15] We also examined trend differences for 1960–1997 and found generally similar overall results. Because these two time periods might not be representative of sampling errors for other periods, we calculated trends for 25 segments of 20 years each, starting with 1958–1977 and ending with 1982–2001, using the same subsampled and full reanalysis time series. When errors are plotted as a function of network size, there is again no apparent relationship between network size and spatial sampling error (Figure 4). Although errors for the LKS network are smaller overall than for the Angell networks, large outliers from the GUAN, HadRT, and HadAT results suggest no reliable improvement with increasing network size. (We also calculated errors using the median of pairwise slopes instead of linear least squares trends and found that although many details were different, our overall conclusions were unaffected.)

[16] To explore the reasons for this surprising result, we constructed 13 hypothetical networks with approximately evenly spaced locations by taking every *n*th grid box in the NCEP reanalysis data set, with *n* chosen to give network sizes from 48 to 1200. This method gives a greater density of sampling at higher latitudes, but this geographic imbalance is eliminated by our use of zonal averaging. The spatial sampling trend errors in these

Table 3b. Same as Table 3a but Using UAH MSU Temperatures and MSU Equivalent Temperatures Calculated From HadRT Radiosonde Data[a]

	Globe		NH		SH		30°S–30°N		20°S–20°N	
Level	MSU	Actual	MSU	Actual	MSU	Actual	MSU	Actual	MSU	Actual
LS	−0.046	−0.023	−0.045	0.006	−0.046	−0.053	−0.002	−0.036	0.003	−0.077
MT	−0.004	−0.031	0.002	−0.033	−0.011	−0.028	−0.005	−0.046	−0.003	−0.099
LT	−0.002	−0.028	0.010	−0.042	−0.015	−0.014	−0.003	−0.042	0.009	−0.127

[a]UAH is University of Alabama at Huntsville; MSU is microwave sounding unit; LS is lower stratosphere; MT is middle troposphere; LT is lower troposphere.

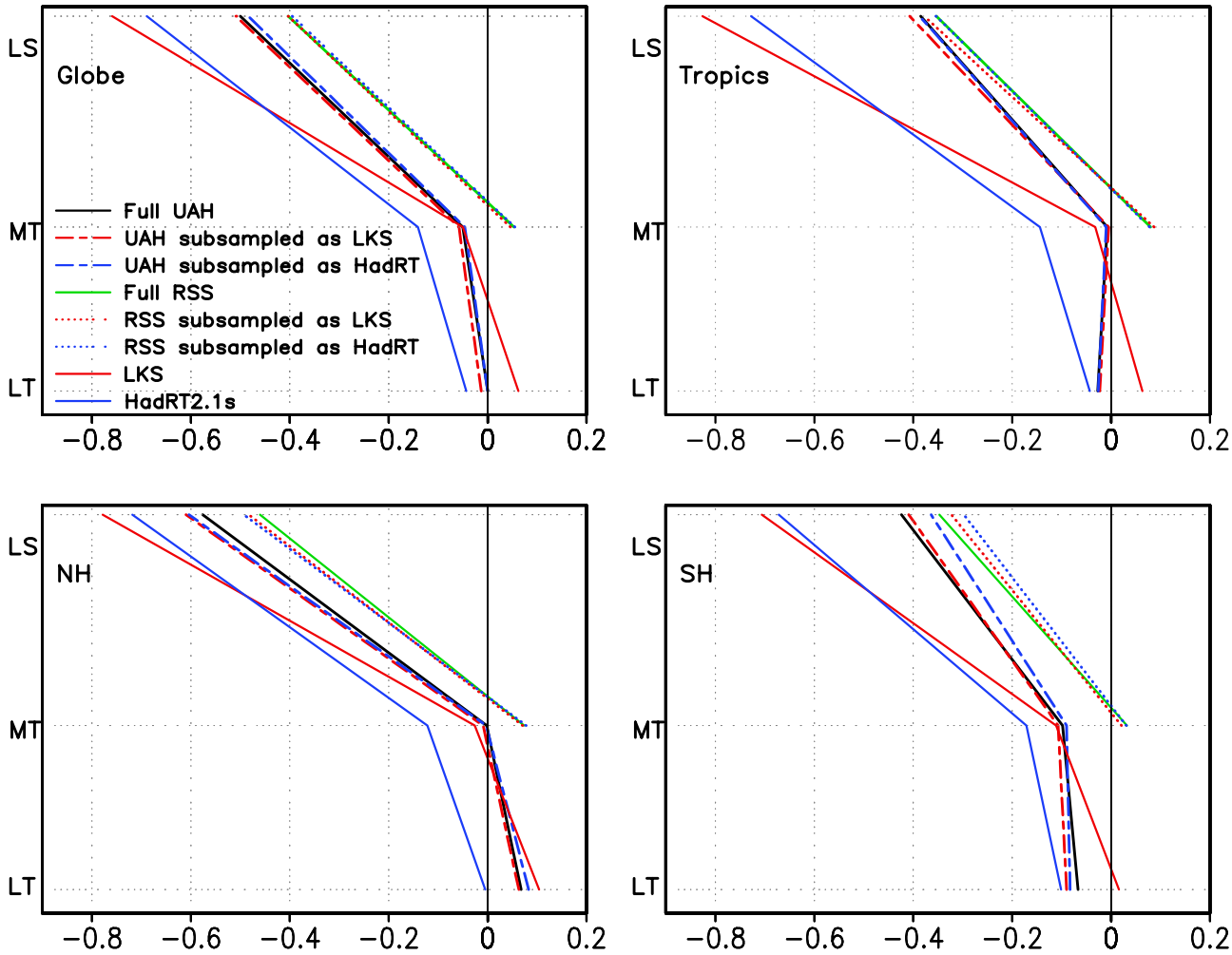

Figure 9. Trends (K decade^{-1}) for 1979–1997 in MSU temperature data from *Christy et al.* [2003] ("Full UAH") and *Mears et al.* [2003] ("Full RSS") and from the LKS and HadRT radiosonde data sets. The two MSU data sets have been subsampled as for LKS and HadRT in both time and space. (The Remote Sensing Systems, Inc., (RSS) data set exists only for lower stratosphere (LS) and middle troposphere (MT).) In the global plot the green RSS line is obscured by the red dotted line denoting RSS subsampled as for LKS.

networks (Figure 5) decline as network size increases from 48 to around 400 locations and are then roughly similar for larger networks, showing a clear relationship to size only for the smaller networks. The error for networks larger than 360 stations is less than 0.02 K decade^{-1}, in contrast to the much larger errors for the HadRT and HadAT networks. We speculate that the advantage of a larger number of stations is overcome in the radiosonde networks by errors caused by the unrepresentative distribution of the additional stations. These results suggest that relatively small (e.g., under 100 stations) but carefully designed radiosonde networks may be almost as good as larger ones for monitoring long-term temperature trends. We also tested a limited number of hypothetical uneven networks created by "flipping" the HadAT network east to west and/or north to south or shifting it a fixed number of degrees. In most cases these alternative uneven networks showed smaller sampling error than the actual HadAT network but more than the evenly spaced networks. These experiments suggest that the concentration of stations in North America, Europe, and China may be particularly unrepresentative, perhaps because it oversamples continental areas as opposed to oceans. These findings may be useful for network design. Further investigation of these and related network design issues is left for future work.

[17] *Santer et al.* [1999] showed that coverage effects can be different when tested using reanalysis versus satellite data. Since the reanalysis may not accurately reproduce the regional distribution of temperature variability, we also subsampled the MSU lower troposphere (LT), middle troposphere (MT), and lower stratosphere (LS) data [*Christy et al.*, 2003] in the same manner as for the reanalysis (Figure 6). Overall, the error estimates from MSU are smaller than those from reanalysis, and the results from the MSU and reanalysis experiments differ in many details. As with the reanalysis, however, the differences between the trends in the full and subsampled time series are in most cases relatively small, no networks are consistently superior to the others, and no consistent

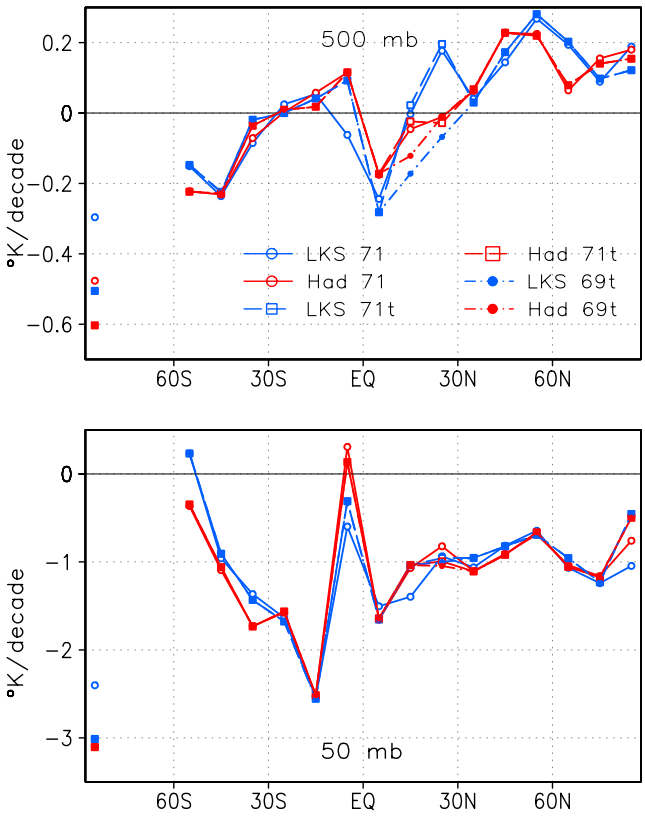

Figure 10. Trends for 1979–1997 in 10° zonal mean temperature anomalies from LKS and HadRT 71- and 69-station subsets. Results labeled "t" are subsampled in time as well as in space.

relationship between size of network and sampling error is apparent.

3.3. Temporal Sampling Error From Globally Complete Data Sets

[18] The temporal sampling differences resulting from missing monthly means can be simulated by dropping out the missing months from a complete data set such as reanalysis or satellite data. Information on missing months was readily available only for the HadRT and LKS radiosonde data sets. Furthermore, we did not have detailed information on the within-month sampling for the CLIMAT TEMP data, so we did not examine temporal sampling errors at submonthly scales. Temporal sampling differences due to use of different days within the month and different observation times within days are part of the difference between the CLIMAT TEMP and CARDS data sources (see section 4.1) used in the HadRT and LKS radiosonde data sets, respectively.

[19] Using the MSU or reanalysis grid boxes that contained the locations of the radiosonde stations in the LKS and HadRT networks, we deleted data from these locations for months when the corresponding radiosonde data were not present. We used 50-mbar radiosonde data to mask the LS (MSU4) data, 500-mbar data for MT (2T), and 850-mbar data for LT. For LKS we masked using both unadjusted and adjusted data. (The unadjusted LKS data set has fewer missing months than the adjusted data set because some months were deleted when data were sparse or erratic or could not be adjusted satisfactorily.) The masked data were then averaged as described in section 3.2, and the trends were compared to the means of the full MSU or reanalysis data set (Table 2).

[20] The difference between trends in data subsampled in time only and trends in data sampled in both space and time gives an estimate of the effect of temporal sampling error in the LKS and HadRT networks (Figures 7 and 8). The errors for 1979–1997 trends are mostly less than 0.02 K decade^{-1} except in the stratosphere and, for reanalysis, at 200 mbar. In some cases, particularly in the stratosphere, temporal sampling errors are larger than spatial sampling errors. Because of the deletions made in the adjusted LKS data sets we expect time sampling errors to be smaller for the unadjusted than the adjusted LKS sampling, and this is generally true, with a few noticeable exceptions in the reanalysis results.

[21] Total sampling errors (differences between full grid trend and trend from data subsampled in both time and space) from reanalysis tests (Table 2) range from less than 0.001 K decade^{-1} for HadRT at 850 mbar in the tropics to 0.098 for LKS at 850 mbar in the SH for 1979–1997. In the SH, LKS errors are noticeably larger than those for HadRT, but in the global mean the two networks have similar overall performance. For MSU (not shown), in contrast, the largest error is for HadRT in the SH stratosphere. MSU total sampling error estimates elsewhere are much smaller than those from reanalysis and show little difference between HadRT and LKS.

3.4. Estimates of Sampling Error Using Radiosonde Data

[22] How well does the sampling error in reanalysis or MSU data represent actual sampling errors? We subsampled actual HadRT radiosonde temperature data in space and time using 71 locations common to both the HadRT and LKS networks (Figure 2). (For HadRT we used data from grid boxes corresponding to the LKS station locations. For 15 of these boxes, more than one station was used to produce the HadRT product, so the comparison does not use data from exactly the same radiosonde stations in those locations. This will not affect the analysis in this section, but it contributes to the differences between LKS and HadRT data discussed in section 4.2.) The difference between trends from the full 444 HadRT stations and trends for the 71–grid box subset shows the effect of differing sampling in the radiosonde data. In Table 3a we compare this effect to the effect of the same sampling experiment applied to the reanalysis data by subsampling the reanalysis in space and time using the same 444-station and 71-station networks and subtracting the trend for 71 stations from the trend for 444 stations. While the overall magnitude of sampling effects is similar in the radiosonde and reanalysis experiments, the differences at individual levels and regions show little relation. For example, at 850 mbar in the SH, where the reanalysis shows a difference of 0.140 K decade^{-1} between trends in the 444- and 71-station networks, the actual radiosonde data show a difference of only −0.003. Using the nonparametric median of pairwise slopes method to estimate trends, we find a similar lack of agreement between reanalysis and radiosonde results.

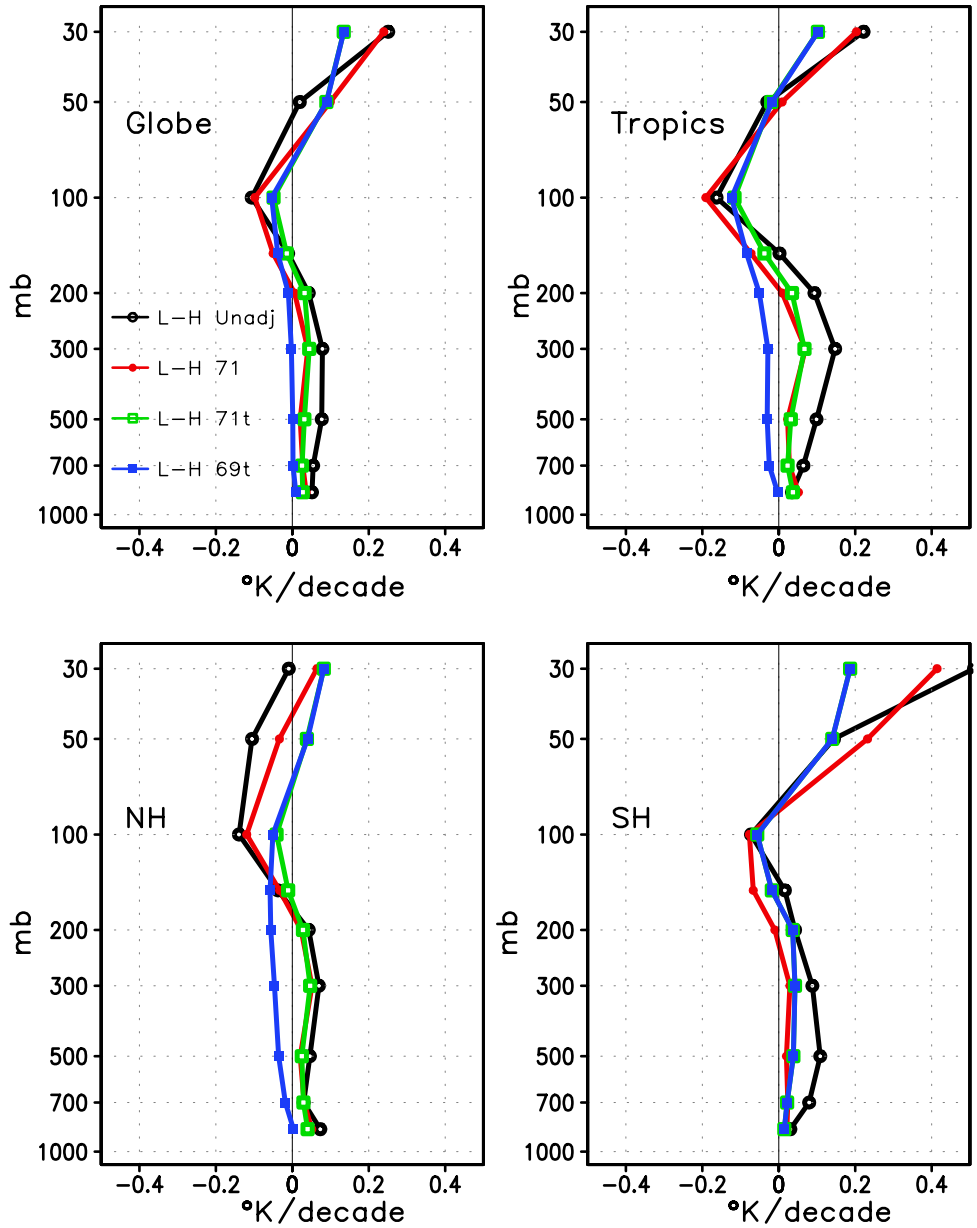

Figure 11. Trends in unadjusted LKS minus trends in (unadjusted) HadRT 2.0 for 1979–1997 for full networks (L-H Unadj), a common 71-station subset (L-H 71), the common 71-station subset with the same time coverage (L-H 71t), and a 69-station subset without Indian stations (L-H 69t).

[23] We did the same test using the UAH MSU data (Table 3b). As with the reanalysis experiment the actual sampling effect and that estimated from UAH MSU data often differ in size and show different vertical and regional patterns.

[24] Thus the sampling error estimates from reanalysis and MSU data are not a foolproof guide to the effect of network selection in actual radiosonde data. Some differences between estimated and actual errors are expected given the differences between the quantities measured by radiosondes and those measured by satellites and the reanalysis (see section 2). Nevertheless, the results suggest the possibility that the large-scale geographic patterns of trends in the reanalysis and MSU may not be sufficiently similar to those in radiosonde upper air data to permit confident assessment of sampling errors. Alternatively, the effects of small-scale sampling error and random instrumental error present in the radiosonde data sets but not in the globally complete data sets may be so large as to overwhelm effects of large-scale sampling errors in the radiosonde data.

4. Comparison of Trends in Radiosonde and MSU Data Sets Using Similar Sampling

4.1. MSU Versus LKS and HadRT

[25] How do the subsampled MSU data compare to the actual radiosonde data? We calculated MSU equivalent global, hemispheric, and tropical temperature series from HadRT2.1s gridded temperature data and LKS adjusted data

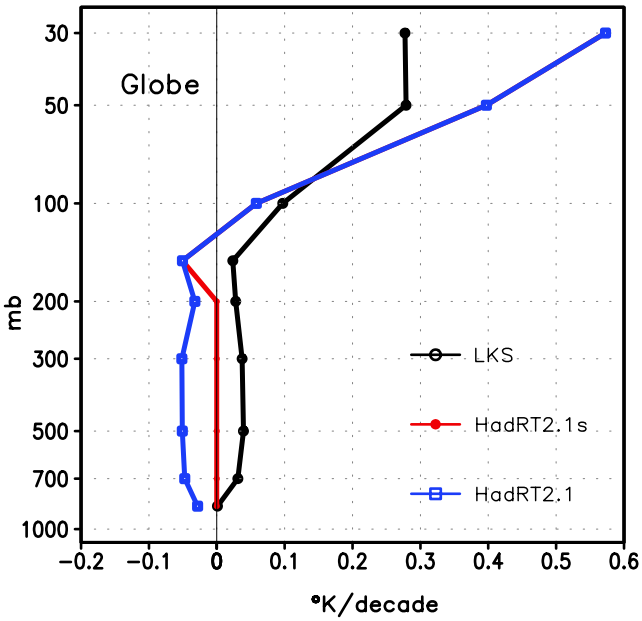

Figure 12. Effect of homogeneity adjustments in LKS, HadRT 2.1, and HadRT2.1s time series on global mean trends, measured by the trend in the adjusted series minus the trend in the unadjusted series for 1979–1997. The traces for HadRT2.1s and HadRT2.1 are identical in the stratosphere, where both have the same adjustments, but differ in the troposphere, where HadRT2.1s has no adjustments.

using the static weighting function method as described by *Seidel et al.* [2004]. In that paper, only latitude zones with at least three grid boxes of data were used to construct large-scale means for HadRT. Here we use all grid boxes with data. UAH and RSS MSU data were subsampled in space and time as described in section 3.3.

[26] Figure 9 shows that trends in the subsampled MSU series are not appreciably closer to the actual radiosonde trends than are those for the complete MSU. In some cases, such as the lower stratosphere in the SH, the trend in MSU subsampled as for HadRT shows noticeably less cooling than the complete MSU, but the actual HadRT trend is more negative than the full MSU trend by more than 0.2 K.

[27] From these results it appears that inadequate spatial coverage and missing months of data are not major causes of differences between MSU and radiosonde trends. More likely sources of discrepancies are time-varying biases in one or more data sets, differences in processing and adjustments, and the inherent difference between the point measurement of a radiosonde and the horizontally and vertically averaged measurement of the satellite instrument.

4.2. HadRT Versus LKS

[28] *Gaffen et al.* [2000], examining a 20-station set, showed that differences between trends in CLIMAT TEMP and CARDS radiosonde data from individual stations were generally less than 0.1 K decade^{-1} but occasionally as much as 0.2 K decade^{-1} for 1959–1991. Here we extend that work to a larger set of stations and examine the large-scale effects of those differences.

[29] We identified 71 locations common to the LKS and HadRT networks (Figure 2). (As discussed above, 16 of these locations included more than one station in the relevant HadRT grid box. For these locations the comparison is not an exact station-to-station match.) To test the influence of differing temporal sampling, we deleted from each unadjusted data set those months that were missing in the corresponding station or grid box in the other data set. Comparing least squares linear trends at 500 mbar for 1979–1997 for each of these locations individually, we find that 10 of the 71 locations, all in the Northern Hemisphere, have trend differences greater than 0.2 K decade^{-1} even with uniform temporal sampling. Seven of these ten locations contain exactly the same stations in both data sets. Trends at an additional 11 locations differed by 0.1–0.2 K decade^{-1}. The difference of 1.07 K decade^{-1} for the Calcutta area is the largest, followed by 0.57 at Bombay and 0.39 at Tripoli. Because of the particularly large differences for the Indian locations and because Indian sonde data may be of particularly poor quality, we used a 69-station subset excluding Calcutta and Bombay in addition to the 71-station set in the large-scale mean comparisons below.

[30] When these 71- and 69-station data sets are combined into 10° zonal means for 500 mbar (Figure 10), their trends show little effect from temporal masking except at the South Pole and at 0°–10°S (where the zonal means include only one or two stations). Removing the Indian locations reduces the difference between the LKS and HadRT trends at 20°–30°N by more than 0.15 K decade^{-1}. Even after the exclusion of locations in India and the temporal masking, trends at 0°–10°N and 60°–70°N still differ by more than 0.1 K decade^{-1} in the troposphere and by as much as 0.4–0.6 K decade^{-1} at some latitudes in the stratosphere.

[31] Of the 59 identical stations present in both HadRT and LKS, 7 have trends that differ by more than 0.2 K decade^{-1}, and 8 more differ by more than 0.1 K decade^{-1}. Using only these 59 locations, we still find differences of 0.1 K decade^{-1} or more in zonal means at 10°–20°N, 20°–30°N, 60°–70°N, and 80°–90°S (not shown).

[32] To assess the effect of these differences on larger-scale means, we subtracted the trend in the HadRT mean data set from the trend in the LKS mean for the globe, tropics, and hemispheres using the complete data sets and the 71-station and 69-station subsets. The subsets were masked as described above to use the same months in each data set. The global mean trends for 1979–1997 for the full unadjusted data sets (87 stations for LKS and 444 for HadRT) differ by up to 0.1 K decade^{-1} below 30 mbar (Figure 11). This difference is almost eliminated in the troposphere when only the common station set (without India) is used. However, this agreement is the result of compensating differences of ~0.05 K decade^{-1} in the two hemispheres. Results are similar for trends calculated using the median of pairwise slopes.

[33] Where the common locations include more than one HadRT station, differences may be caused primarily by the subgrid-scale spatial variability, reflected in the differences in trends for nearby stations. The differences between HadRT and LKS data for the same stations may arise from differences in within-month time sampling, including different days in the month and different times within a day, from different procedures used to calculate monthly means, or from differing errors in transmission of the two sets of

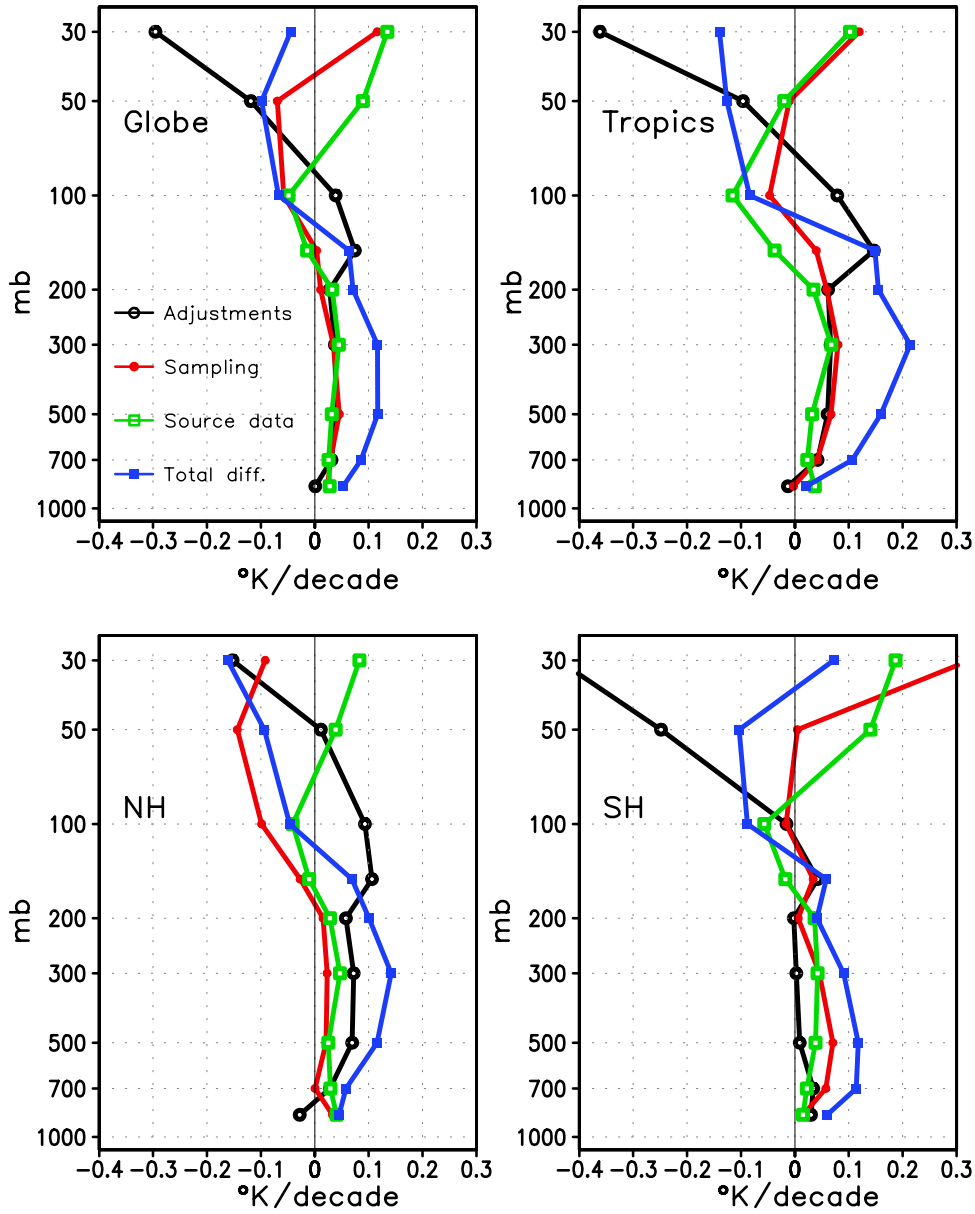

Figure 13. Effects of adjustments, sampling differences, and differences in source data on the difference between trends for 1979–1997 in LKS and HadRT2.1s time series. See text for details.

data. In most cases, LKS data are likely to be more reliable because their monthly means are based on a uniform averaging and quality control procedure, but several problems have been identified in the CARDS database (on which LKS is based) since the LKS data set was created (I. Durre et al., Overview of the integrated global radiosonde archive, submitted to *Journal of Climate*, 2005), which could have affected the LKS data set, particularly in early years.

5. Comparison of Sampling, Homogeneity Adjustment, and Source Data Effects on Trends in LKS and HadRT Data Sets

[34] Section 4.2 considers the differences in sampling and source data between LKS and HadRT. The other principal difference is the adjustments made by the two groups to remove temporal inhomogeneities in the data. Here we show the effect of those adjustments on trends in large-scale mean time series and compare the effects of all three factors.

[35] Comparing the unadjusted and adjusted versions of radiosonde data sets provides a measure of the impact of homogeneity adjustments. *Free et al.* [2002] showed major differences between homogeneity adjustments by different groups at 12 stations and their effects on local trends but did not address the effect of these differences on large-scale mean trends. *Lanzante et al.* [2003b] compared unadjusted and adjusted trends for 87 stations by pressure level. In that work the authors calculated trends by station and then combined them to get global and other large-scale trends. Here we created global, hemispheric, and tropical mean

time series from the station data and compared trends in those series. The effects of the LKS and HadRT adjustments (using HadRT 2.1) are opposite in sign in the troposphere (LKS adjustments increase the warming, and HadRT adjustments increase the cooling trend in their respective data sets), while both reduce the cooling trend in the stratosphere in the global (Figure 12), hemispheric, and tropical means (not shown). Results using the median of pairwise slopes are similar.

[36] In Figure 13 we compare the effects of adjustments, input differences, and sampling differences between LKS and HadRT2.1s. To estimate the effect of adjustments on the LKS-HadRT differences, we first compute the effect of adjustments ($\Delta\text{Trend}_{\text{adj}}$) for each data set by subtracting trends in the unadjusted from the trends in the adjusted series (2.1s for HadRT) for each data set separately. The effect on LKS-HadRT differences is then $\Delta\text{Trend}_{\text{adj}}(\text{LKS}) - \Delta\text{Trend}_{\text{adj}}(\text{HadRT})$. We compare differences between subsampled and full series in the same way. The effect of differences in source data is shown by the trend in the 71-station unadjusted LKS data minus the trend in the 71-station unadjusted HadRT subset.

[37] The effect of adjustments in the troposphere in the global mean (1979–1997) is roughly similar in size to the effects of sampling differences and differences in input data and is of the same sign. In the tropics the pattern is similar, but the effects are larger, with almost 0.2 K decade^{-1} total difference in trend at 300 mbar. If the tropics are limited to 20°S–20°N (not shown), the total difference between data sets is larger, and sampling effects predominate, while source data effects become minimal. Using trends from the median of pairwise slopes, again, sampling effects are the largest factor, and source data differences are less important. In the stratosphere, large effects of opposite sign result in smaller net differences between trends in the two data sets. The relative roles of the three factors differ in the SH and NH and when we use a smaller subset of locations with identical stations.

[38] The specific contributions of sampling, source data, and adjustments will also differ with time period, geographic area, and data set, but it is reasonable to expect all three factors to be important for other data sets as well. The roughly similar contributions of differences in source data, adjustments, and station networks have implications for efforts to create improved upper air temperature data sets from radiosonde data. To narrow uncertainties significantly, these efforts will need to address all three issues.

6. Conclusions

[39] If MSU and reanalysis depictions of long-term spatial variability are adequate, the effects of spatial sampling differences on trends are usually less than 0.05 K decade^{-1} but can occasionally be as large as 0.12 K decade^{-1}. In the lower troposphere, where trends are small, spatial sampling effects are often larger than the trends themselves and are comparable to the standard error of the trends. In the upper troposphere and stratosphere, errors are usually no more than 25% of the trends and are much smaller than the standard error. The effect of missing months of data is typically ~0.02 but may be as large as 0.08 K decade^{-1}. However, estimates of coverage effects based on reanalysis and MSU data differ noticeably from those seen in actual radiosonde data as well as from each other. Effects in actual data are generally larger than those in the globally complete data sets. It is thus unclear to what extent these complete global data sets can reliably reproduce the behavior of actual radiosonde data.

[40] In this work, although spatial sampling errors decrease with network size in hypothetical evenly spaced networks, the larger actual radiosonde networks, with less regular station distributions, do not give consistently smaller spatial sampling error than the smaller networks. Thus these experiments provide little basis for preferring one network over another and suggest that a carefully chosen network with fewer than 100 stations may be almost as good as much larger networks for monitoring long-term trends. Other benefits of a larger network, such as reduction of the effects of random instrumental errors and errors due to inadequate subgrid or submonthly sampling, are beyond the scope of this work.

[41] Differences in spatial coverage and in months for which data are available do not explain the differences between trends in MSU and radiosonde data sets. In some cases, subsampling increases the differences between these trends (as shown by *Santer et al.* [1999]). We do not address differences in within-month sampling.

[42] Using 71 locations common to the LKS and HadRT data sets improves the agreement between trends in the two data sets for 1979–1997, but LKS trends are still 0.03–0.05 K decade^{-1} more positive in the troposphere than HadRT even for unadjusted data. Results are very sensitive to elimination of the stations in India. Trends for the same individual locations in the two data sets before homogeneity adjustments can differ by more than 0.2 K decade^{-1}.

[43] Homogeneity adjustments, sampling differences, and differences in input data make roughly comparable contributions to total differences between trends in the LKS and HadRT temperature data set for 1979–1997. It follows that to narrow uncertainties in radiosonde temperatures, we must consider all three sources of disagreement. New data sets using improved data sources and new adjustment methods are currently under construction and may show smaller differences in trends than the HadRT and LKS data sets discussed in this paper.

[44] **Acknowledgments.** We thank Peter Thorne, Mark McCarthy, and David Parker of the U. K. Met Office for providing the Hadley Center data and station lists. John Lanzante, Jim Angell, and an anonymous reviewer provided helpful comments on the manuscript. This work was supported in part by the Climate Change Data and Detection element of NOAA's Office of Global Programs.

References

Agudelo, P. A., and J. Curry (2004), Analysis of spatial distribution in tropospheric temperature trends, *Geophys. Res. Lett.*, *31*, L22207, doi:10.1029/2004GL020818.

Angell, J. K. (2003), Effect of exclusion of anomalous tropical stations on temperature trends from a 63-station radiosonde network, and comparison with other analyses, *J. Clim.*, *16*, 2288–2295.

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 MSU/AMSU bulk atmospheric temperatures, *J. Atmos. Oceanic Technol.*, *20*, 613–629.

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., et al. (2002), Creating climate reference datasets: CARDS workshop on adjusting radiosonde temperature data for climate monitoring, *Bull. Am. Meteorol. Soc.*, *83*, 891–899.

Gaffen, D. J., M. Sargent, R. E. Habermann, and J. R. Lanzante (2000), Sensitivity of tropospheric and stratospheric temperature trends to radiosonde data quality, *J. Clim.*, *13*, 1776–1796.

Gibson, J. K., P. Kallberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serrano (1997), ERA description, *ECMWF Reanal. Proj. Rep. Ser.*, *1*, 66 pp., Eur. Cent. For Medium-Range Weather Forecasts, Reading, U. K.

Hansen, J., and S. Lebedeff (1987), Global trends of measured surface air temperature, *J. Geophys. Res.*, *92*, 13,345–13,372.

Huth, R., and L. Pokorna (2004), Parametric versus non-parametric estimates of climatic trends, *Theor. Appl. Climatol.*, *77*, 107–112.

Jones, P. D., T. J. Osborn, and K. R. Briffa (1997), Estimating sampling errors in large-scale temperature averages, *J. Clim.*, *10*, 2548–2568.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*, 437–471.

Karl, T., R. Knight, and J. Christy (1994), Global and hemispheric temperature trends: Uncertainties related to inadequate spatial sampling, *J. Clim.*, *7*, 1144–1163.

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.

Oort, A. (1978), Adequacy of the rawinsonde network for global circulation studies tested through numerical model output, *Mon. Weather Rev.*, *106*, 174–195.

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.

Rosen, R., J. M. Henderson, and D. Salstein (2003), Sensitivity of continental-scale climate trend estimates to the distribution of radiosondes over North America, *J. Atmos. Oceanic Technol.*, *20*, 262–268, doi:10.1175/1520-0426.

Santer, B., et al. (1996), A search for human influences on the thermal structure of the atmosphere, *Nature*, *382*, 39–46.

Santer, B., et al. (1999), Uncertainties in observationally based estimates of temperature change in the free atmosphere, *J. Geophys. Res.*, *104*, 6305–6333.

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.

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.

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.

Trenberth, K., and J. Olson (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.

World Meteorological Organization (1996), GCOS plans progress, *World Clim. News*, *9*, 9–10.

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