Using First Differences to Reduce Inhomogeneity in Radiosonde Temperature Datasets

MELISSA FREE AND JAMES K. ANGELL

NOAA/Air Resources Laboratory, Silver Spring, Maryland

IMKE DURRE

NOAA/National Climatic Data Center, Ashville, North Carolina

JOHN LANZANTE

NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

THOMAS C. PETERSON

NOAA/National Climatic Data Center, Asheville, North Carolina

DIAN J. SEIDEL

NOAA/Air Resources Laboratory, Silver Spring, Maryland

(Manuscript received 23 December 2003, in final form 24 April 2004)

ABSTRACT

The utility of a ''first difference'' method for producing temporally homogeneous large-scale mean time series is assessed. Starting with monthly averages, the method involves dropping data around the time of suspected discontinuities and then calculating differences in temperature from one year to the next, resulting in a time series of year-to-year differences for each month at each station. These first difference time series are then combined to form large-scale means, and mean temperature time series are constructed from the first difference series. When applied to radiosonde temperature data, the method introduces random errors that decrease with the number of station time series used to create the large-scale time series and increase with the number of temporal gaps in the station time series. Root-mean-square errors for annual means of datasets produced with this method using over 500 stations are estimated at no more than 0.03 K, with errors in trends less than 0.02 K decade^{-1} for 1960–97 at 500 mb. For a 50-station dataset, errors in trends in annual global means introduced by the first differencing procedure may be as large as 0.06 K decade^{-1} (for six breaks per series), which is greater than the standard error of the trend. Although the first difference method offers significant resource and labor advantages over methods that attempt to adjust the data, it introduces an error in large-scale mean time series that may be unacceptable in some cases.

1. Introduction

Use of radiosonde datasets for climate studies has been hampered by the presence of numerous inhomogeneities caused by frequent changes in instruments and practices (Gaffen 1994; Parker and Cox 1995). Adjustment methods used for surface data generally rely heavily on comparison with data from neighboring stations. These methods are of limited usefulness for radiosonde data because upper-air stations are more widely scattered than surface stations (though upper-air temperatures are more spatially coherent), and because inhomogeneities are often present at the same time throughout an entire country. The methods used so far to reduce these inhomogeneities in radiosonde data are typically labor intensive (e.g., Lanzante et al. 2003), making the resulting datasets difficult to expand in time or space. Here we examine an alternative approach using a first difference (FD) technique to combine station observations into area averages.

2. The first difference method

The FD method for combining station data (Peterson et al. 1998) was created to facilitate the use of short data segments in the analysis of surface data and is used by the National Climate Data Center (NCDC) for mon-

Corresponding author address: Dr. Melissa Free, NOAA/Air Resources Laboratory (R/ARL), 1315 East West Highway, Silver Spring, MD 20910.
E-mail: melissa.free@noaa.gov

itoring surface climate. The FD series is the time series of year-to-year changes in a variable at a station, that is,

$$D_t = x_t - x_{t-1}, \quad (1)$$

where x_t is the value of the variable at time t. For example, to get the first difference value for January temperature for 1981, we subtract the temperature for January 1980 from the temperature for January 1981. The D_t values for each time step are averaged over the stations in a region to form a regional first difference series r_t. The regional mean time series is then constructed by cumulatively summing the area-averaged first difference values from the first year to the last, setting the first year to zero:

$$R_n = \sum_{t=1}^{n} r_t. \quad (2)$$

If no gaps in the data exist, the result will be the same as conventional averaging except that the time series will be shifted up or down so that the initial value is zero. (Alternatively, the initial value could be set to equal the regional mean of the original data for that time step.) A final series cannot be reconstructed from an FD series containing gaps unless the FD series is first combined with one or more other series.

Although Peterson et al. (1998) used the method to facilitate the use of short data segments, it can also be used along with cuts in the data to reduce inhomogeneities due to instrument changes, station moves, and other changes (e.g., Tuomenvirta 2004). In a conventional time series, artificial jumps due to changes in instruments or practices can affect the entire time series after the jump (Fig. 1a). In a first difference time series, however, jumps in the original time series are reflected in only one or just a few points in time in the FD series (Fig. 1b). Thus, omitting those points should eliminate the effect of the jumps on the resulting time series. Given sufficient information about the timing of possible sources of discontinuities, we can omit segments of data that contain these potential inhomogeneities. By using the first difference method and incorporating only those segments of radiosonde data that are believed to be free of possible artificial jumps, we can, in theory, construct a series unaffected by the known changes without having to accurately determine the effects of the changes. This is especially appealing for upper-air data because different methods for adjusting radiosonde data yield contradictory estimates of the sign and magnitude of artificial jumps (Free et al. 2002) and because the vertical structure of artificial jumps in radiosonde data through the depth of the atmosphere can be complex (Lanzante et al. 2003).

Along with these advantages, the FD method also creates some problems. First, Peterson et al. (1998) found that the method introduces a random error related to the presence of outliers at the endpoints of the series segments, which can be reduced by removing segment endpoints that exceed a given range (''endpoint outlier trimming''). Large short-term excursions in climate have large-magnitude FD values before and after the event, with the latter value in the opposite sense to the former (Figs. 1c,d). If the excursion begins at an endpoint of a segment of data, only one of the pair of large-magnitude FD values contributes to the area-averaged time series (Figs. 1e,f), distorting the end results. If large, this error could reduce the usefulness of resulting time series. Second, the method relies on external information about the data (''metadata''), such as the timing of station moves, instrument changes, etc., to identify the locations of potential discontinuities. Third, if such discontinuities occur at the same time at all stations in a large region, requiring the deletion of data at that time over the entire region, the resulting time series for that period may not be representative of the total area. Last, because it inherently involves combining several time series, the method can be used only to produce large-scale mean series and not to create homogeneous records for individual stations.

This paper addresses the effects of the first problem, the random error arising from the endpoint outlier effect. The second problem can be mitigated by better metadata. Efforts to improve the existing metadata for radiosonde observations are underway at NCDC. If future changes are better documented than past events, the FD method will be easier to use for updating climate records in the future. The estimation of errors arising from the third problem depends on the timing of individual metadata events and will not be explored further in this paper. Although the fourth problem may prevent the use of the first difference method for local climate monitoring, global and hemispheric mean time series are of considerable interest for climate studies (e.g., Folland et al. 2001).

We did several types of tests to examine the impact of the first, or random error, effect. Section 3 discusses tests done with radiosonde station data from the Lanzante–Klein–Seidel (LKS) dataset (Lanzante et al. 2003), and section 4 discusses the tests done with National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis data (Kalnay et al. 1996).

3. Tests with LKS data and adjustment points

These tests assess the ability of the method to reproduce the results of the homogeneity adjustment procedure described in LKS. In that work, the authors adjusted a radiosonde temperature dataset containing 87 stations throughout the globe using a subjective multifactor expert evaluation of the data. Our goal is to achieve a similar result using first differencing, without the labor-intensive, subjective LKS process.

We removed six consecutive months from the unadjusted LKS temperature anomaly station data im-

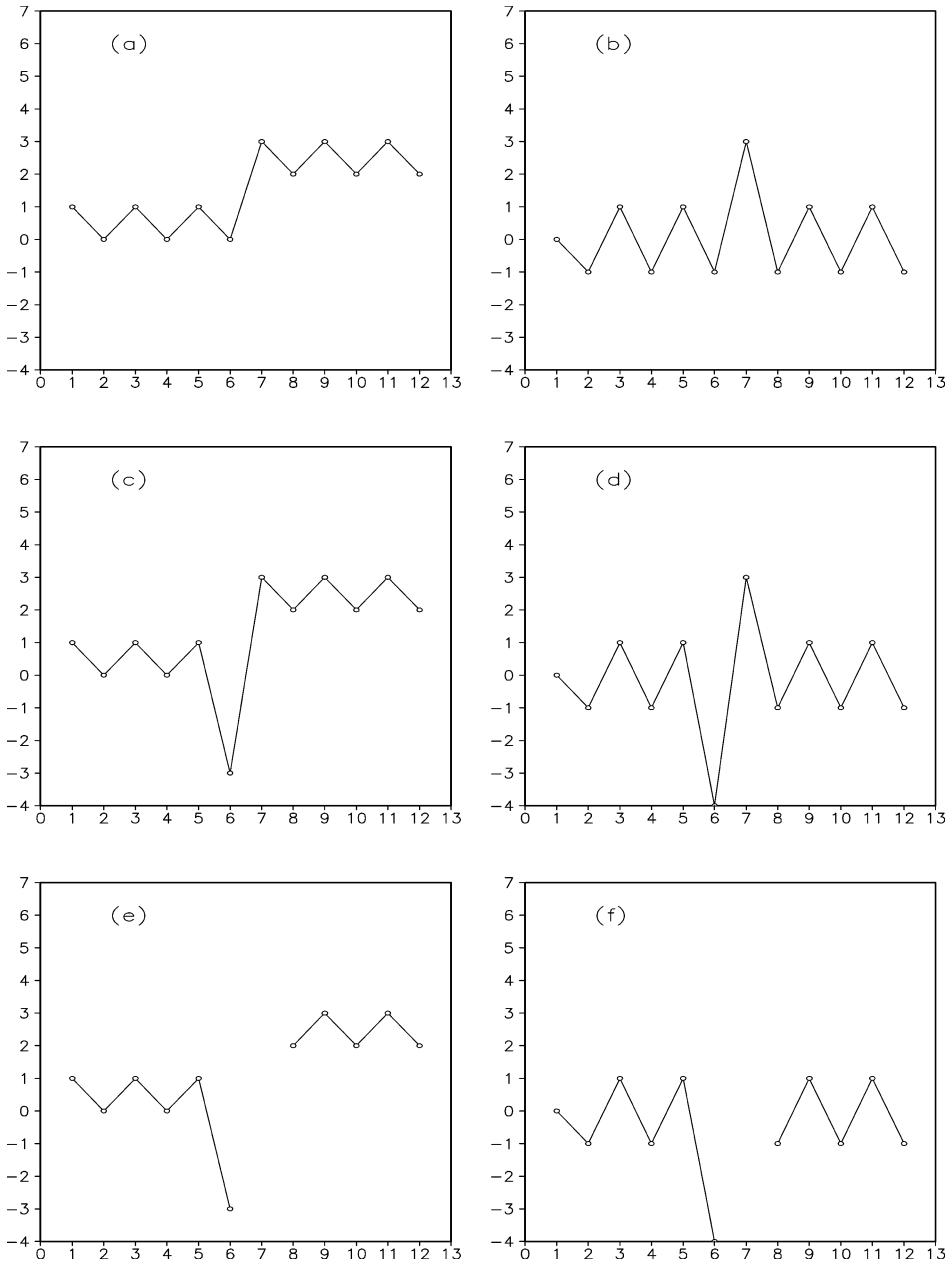

FIG. 1. (a) Example of time series with step change in level at $t = 7$. (b) FD series derived from series in (a). (c) Time series as in (a), but with an ''outlier'' before the step change. (d) FD series derived from series in (c). (e) Time series in (c) with the data point at the beginning of the step change removed. (f) FD series derived from series in (e). The final time series is constructed by cumulatively summing the elements in the FD series. [For (b) and (d), the final series is the same as in (a) and (c), respectively, but shifted so that the first point is zero. For (f), a final series can be constructed only if the series is first combined with another series to eliminate the gap.]

mediately before and after the dates of the LKS adjustments. The resulting series were combined using first differences and the results were compared to the means of the 87-station LKS-adjusted time series. The number of adjustment dates for individual stations ranges from zero to seven, with an average of between one and two adjustments per time series. The series are differenced by month, for example, January 1959 minus January 1958, and the monthly results are then averaged to produce annual means. This was repeated with and without endpoint outlier trimming, in which anomaly data immediately before or after a gap were deleted if they were more than two standard deviations from the mean. A similar procedure was applied to missing data at the

TABLE 1. Standard deviations (K) of the global time series for 1948–97 obtained from unadjusted LKS data, the adjusted LKS data, the unadjusted LKS data combined using the FD method, and the unadjusted LKS data using the FD method with endpoint outlier trimming, as described in the text.

	500 mb	50 mb
UNADJ	0.222	0.866
LKS ADJ	0.213	0.839
FD	0.254	1.033
FD trim	0.237	0.737

start of series that began after 1948 and at the end of series that ended before 1997. For purposes of illustration, we focus particularly on 500 mb in the midtroposphere and 50 mb in the stratosphere. (Analysis of data at other levels yielded comparable results.)

The results presented in Table 1 show that the standard deviations of the FD-produced time series without endpoint trimming (shown in Figs. 2a,c) are 19%–22% larger than that of the adjusted LKS series and are smaller with endpoint trimming than without it. Plots of the differences between these series and the original data (Figs. 2b,d) indicate the effect of using FD in comparison with the effect of the LKS adjustments on the time series. The effect of FD is larger and more variable with time than the effect of the LKS adjustments, and

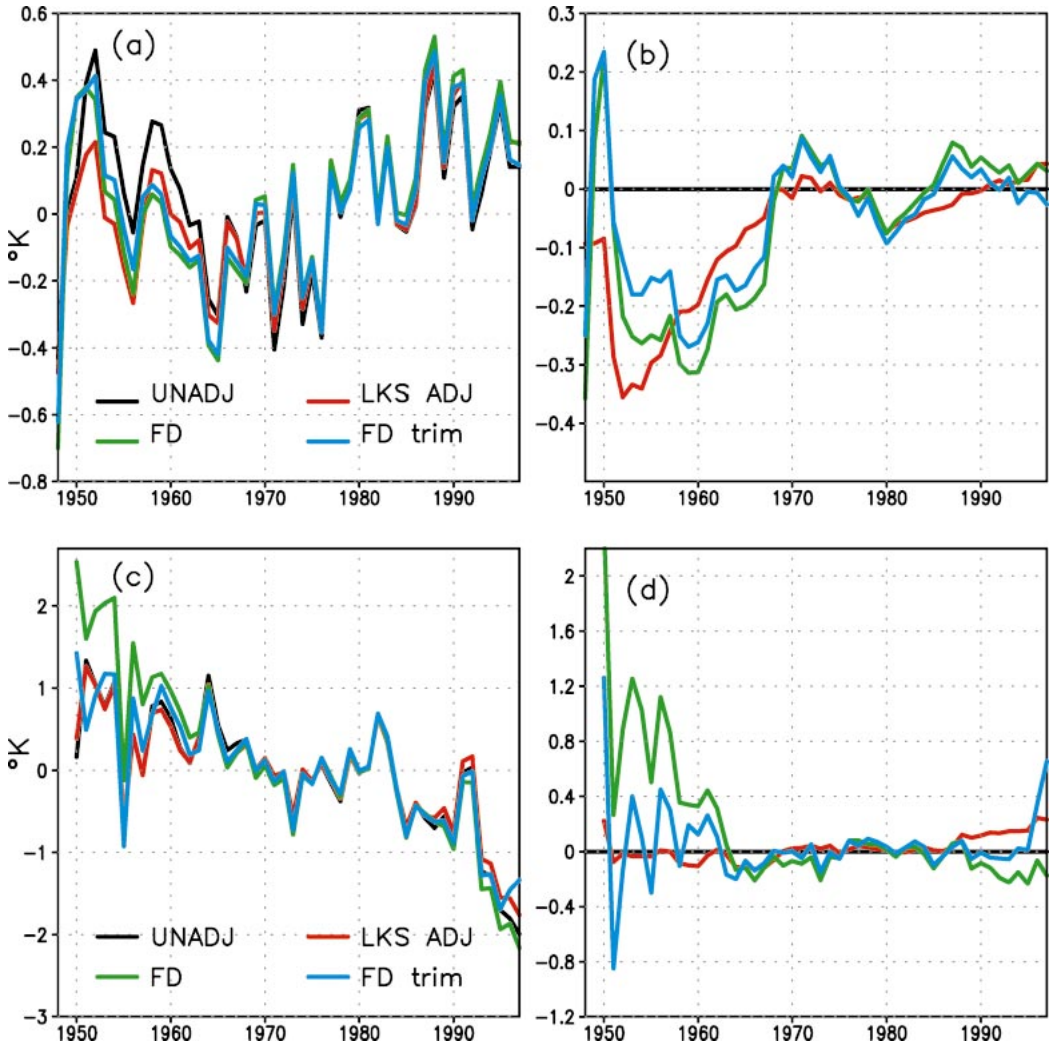

FIG. 2. (a) Annual mean temperature anomalies resulting from the FD procedure using unadjusted LKS radiosonde station data as described in the text, with (FD trim) and without (FD) endpoint outlier trimming, along with the mean of LKS adjusted data (LKS ADJ) and LKS unadjusted (UNADJ) data, at 500 mb. The unadjusted LKS data have undergone deletions of questionable data (DEL version of LKS). Anomalies are calculated with respect to 1960–90. (b) Adjusted and FD time series in (a) minus the mean of the unadjusted data, showing the effect of the FD procedure and the LKS adjustments on annual global mean temperature at 500 mb. (c) As in (a), but for 50 mb. Note difference in scale from (a). (d) As in (b), but for 50 mb. Note difference in scale from (b).

TABLE 2. Least squares linear trends (K decade^{-1}) in global annual mean temperatures for 1959–97 from unadjusted LKS data combined using the FD method and the FD method with endpoint outlier trimming, as described in the text; from adjusted LKS data; and from the unadjusted LKS data. The confidence intervals shown are twice the standard error of the LKS trends. Confidence intervals for the other trends are similar.

Level (hPa)	FD	FD trim	LKS ADJ	UNADJ	Confidence interval
Surface	0.04	0.07	0.09	0.10	0.08
1000	0.12	0.12	0.10	0.05	0.09
850	0.13	0.11	0.11	0.11	0.08
700	0.16	0.12	0.12	0.10	0.08
500	0.15	0.13	0.12	0.09	0.08
400	0.16	0.13	0.11	0.07	0.08
300	0.14	0.11	0.08	0.03	0.06
250	0.12	0.11	0.04	0.01	0.06
200	0.00	−0.01	−0.02	−0.05	0.07
150	−0.06	−0.06	−0.08	−0.13	0.09
100	−0.39	−0.31	−0.22	−0.30	0.16
70	−0.71	−0.54	−0.41	−0.47	0.31
50	−0.56	−0.46	−0.42	−0.49	0.33
30	−0.52	−0.46	−0.33	−0.37	0.28
20	−0.34	−0.54	−0.26	−0.26	0.25

the trimmed FD is generally closer to the LKS adjustments than is the untrimmed FD. The higher variability of the FD series is most noticeable at the beginning and, in some cases, near the end of the time period, when the total number of series included is smaller and more variable.

Least squares linear regression trends are generally larger in absolute value for the FD than those in the LKS-adjusted series, as shown in Table 2. The absolute value of the difference between the trimmed FD and the LKS trends is 0.00–0.07 K decade^{-1} below 150 mb for 1959–97 but increases to more than 0.1 K decade^{-1} at some levels above 100 mb. The difference is in most cases less than the standard error of the trend, so that the uncertainty introduced by FD is generally less than that created by the interannual variability of the series. This difference seems to increase with altitude above 850 mb but is larger at the surface than at 850–500 mb and is generally larger when endpoint outlier trimming is not used. The divergence between the FD and LKS results is smallest at the global scale and increases for smaller regions such as the hemispheres and Tropics (not shown).

Visual inspection of the 500-mb time series (Fig. 2a) suggests that the LKS and the FD series are more similar to each other than to the original time series, particularly in the earlier time period. At this atmospheric level, the trimmed FD series is not much different from the LKS after the mid-1960s. At 50 mb (Fig. 2c), in contrast, the LKS-adjusted series seems to be closer to the original than to the FD versions. This is consistent with the pattern seen in the trends in Table 2, where the FD trends resemble the LKS trends more closely in the troposphere than in the stratosphere.

In the troposphere, both FD and the LKS adjustment methods give generally increased warming over time compared to the unadjusted data. In the stratosphere, however, the effect of the FD method at 100 and 70 mb is to make the trends more negative, while the LKS adjustments make them less negative. Comparisons with satellite data from the Microwave Sounding Unit suggest that unadjusted radiosonde trends may have too much cooling in the stratosphere (Shine et al. 2003; Seidel et al. 2003). This in turn suggests that the LKS result is likely to be more realistic than the FD result in the stratosphere, perhaps because of the random errors introduced by the FD method.

4. Tests using reanalysis data

One shortcoming of the tests in the previous section is the lack of ''ground truth.'' Since we cannot be certain that the LKS adjustments are correct, we do not know how much of the difference between the LKS-adjusted data and the results of FD is due to errors produced by the FD method. Using NCEP–NCAR reanalysis data gives us a temporally complete dataset to compare with the results of FD procedures. It also allows us to combine larger numbers of time series to examine the dependence of FD effects on dataset size.

To clarify the effects of FD on data containing gaps, we calculated trends in the mean of 87 time series taken from the gridded NCEP–NCAR reanalysis data at points corresponding to the locations of the 87 LKS stations with and without using the FD method. In both cases, the time series were masked to duplicate the time coverage of the actual LKS radiosonde data by deleting months missing in the LKS data. (Due to geographic sampling error, the 87-station network will not exactly reproduce the true global mean from the full reanalysis grid. The size of this sampling error is the subject of current research, but is beyond the scope of this paper.) Table 3 shows the difference between the trends for the mean time series with temporal masking and the means

TABLE 3. Differences in global mean trends (subsampled and FD series trends minus trend from complete 87 time series) from NCEP–NCAR reanalysis data sampled as for LKS 87 stations (K decade^{-1}). The trend for the complete 87 time series (87 STNS) is shown at the top of the table for comparison. "Masked" means that months missing in the actual LKS data are deleted from the reanalysis data. For the FD time series, "FD" means that no endpoint outlier trimming was used, "trim 1" means that endpoints exceeding one standard deviation from the mean were deleted, and "trim 2" means that endpoints exceeding two standard deviations from the mean were deleted before the FD procedure was implemented. See text for further explanation.

Pressure level	850 mb		500 mb		200 mb		50 mb	
Time period	1960–97	1979–97	1960–97	1979–97	1960–97	1979–97	1960–97	1979–97
Trend								
87 STNS, unmasked	0.194	−0.022	0.134	−0.072	0.259	−0.323	−0.452	−0.882
Differences in trends								
87 STNS, masked	−0.038	0.010	−0.028	0.011	−0.031	−0.007	0.016	0.044
FD	−0.026	−0.050	−0.014	0.020	0.054	0.018	0.115	0.172
FD trim 1	−0.068	−0.036	−0.041	0.015	−0.002	0.007	0.161	0.264
FD trim 2	−0.061	−0.002	−0.041	0.012	−0.020	−0.053	0.177	0.272

from the complete, unmasked 87-station time series for the time periods 1960–97 and 1979–97 and four pressure levels. The test was performed with no endpoint outlier trimming and repeated with trimming at one and two standard deviations. For all three trim scenarios, in six of the eight cases, the FD result is farther from the "true" trend than is the non-FD result. Averaging the eight cases together shows a mean absolute error of ~0.02 K decade^{-1} for the non-FD trends and ~0.06 to ~0.08 for the FD methods. Here the error in trends for the FD method is at least 3 times the error without it.

Using NCEP–NCAR reanalysis data, we also performed many randomized FD procedures to assess the potential uncertainty in trends resulting from the random FD error. Starting with gridded reanalysis temperature anomalies for 500 mb, we selected subsets containing various numbers of grid boxes and then deleted 12 consecutive months of data from the individual time series at randomly chosen times. The subsets were selected so as to sample all areas of the globe approximately equally. The deletions are intended to simulate the effect of deleting data around the times of documented changes in instruments or procedures. The resulting series were combined using first differences, and the "global" series were compared to the mean of the same time series without the cuts.

We repeated these experiments 1000 times with different dates for the cuts to get an idea of the range of possible results. This was repeated for 1–6 cuts and 35–1000 grid boxes to test the sensitivity of the FD results to those factors. We also did similar tests using 87 grid boxes located near the LKS stations, masking the reanalysis data to reproduce the data gaps in the actual station data, to assess the effects of missing data on the FD results. For most of these, we tried the first differencing with and without endpoint outlier trimming (removing endpoints exceeding one standard deviation). We used the difference between FD results and the mean of the complete series as a measure of the error introduced by the first-differencing procedure. We also compared the trends in the FD results cut at different random times.

Figure 3 shows an example of the difference between mean time series created by FD and the mean of the original time series for 20 randomized iterations. The example used reanalysis data from 500 mb for 87 arbitrarily selected points and made two random cuts in each gridpoint time series, deleting 6 months of data before and after each cut. The difference between the mean of the full series and the FD results generally grows with time, consistent with the "random walk" nature of the errors. Also shown in Fig. 3 is the difference between the original time series and the randomly cut series combined without using FD. It is clear from comparison that most of the difference in the FD results is derived from the FD procedure rather than from the random cuts themselves.

When measured by the rms differences between orig-

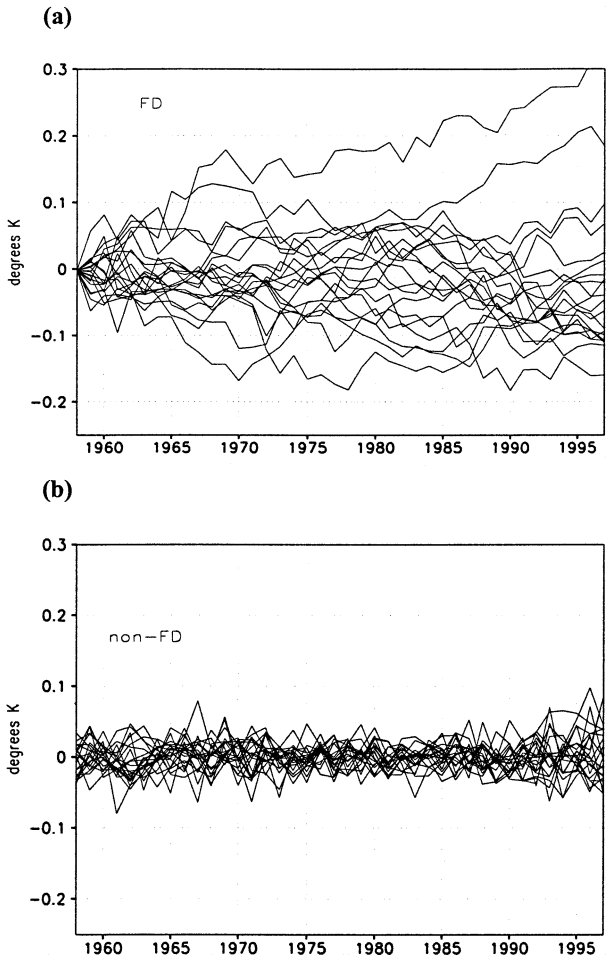

FIG. 3. (a) Error in global mean 500-mb temperature time series produced by the FD method, measured by the difference between those series and the arithmetic mean of the complete time series. Each of the 20 time series shown was produced using temperature series taken from the same 87 locations in the NCEP–NCAR reanalysis gridded data. Each grid box series was cut at two randomly selected times before the 87 series were combined using FD without endpoint outlier trimming. (b) As in (a), but for error caused by the random cuts when the series are combined without using FD.

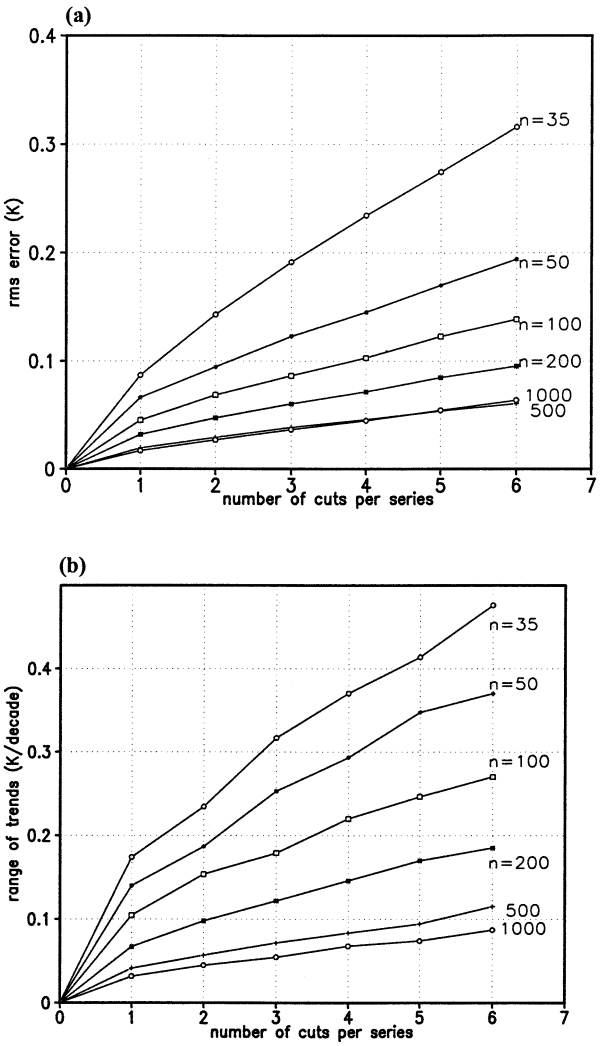

FIG. 4. (a) The rms difference between arithmetic mean of original time series for Jun and the mean produced by FD using endpoint outlier trimming at one standard deviation, as a function of the number of cuts made in the series and the number of series used (*n*). (b) Range of trends in global mean temperatures for Jun, 1958–97, calculated using the FD method, as a function of the number of cuts made in each time series and the number of time series used. The range is computed as the 95th percentile minus the 5th percentile of the trends generated in 1000 randomized iterations. See text for details.

inal and FD results or by the range of trends in results from many iterations, the error is larger for smaller numbers of time series and for a larger number of cuts in the series. In individual cases, however, the results are not necessarily predictable due to the random nature of the errors. Figure 4a shows the rms difference between the time series produced by FD and the arithmetic mean of the input data as a function of the number of cuts made in the data and the number of time series that are combined, for the month of June at 500 mb. (June was chosen arbitrarily as an example. Results for June are not systematically larger or smaller than results for other months.) With two cuts made in each series, the rms difference varies from ~0.14 K for 35 series to ~0.025 K for 500 series.

Figure 4b shows the dependence of the range of trends on the number of cuts and number of time series. The range in trend is measured by the difference between the 95th and 5th percentiles of trends in 1000 iterations using different randomly timed cuts. For two cuts per series, the range varies from ~0.24 K decade^{-1} for 35 series to less than 0.05 K decade^{-1} for 1000 series. The relatively large random walk effect for individual months is greatly reduced for annual means, as shown in Fig. 5, where the rms errors for the same case discussed above range from only ~0.05 to ~0.01 K, with

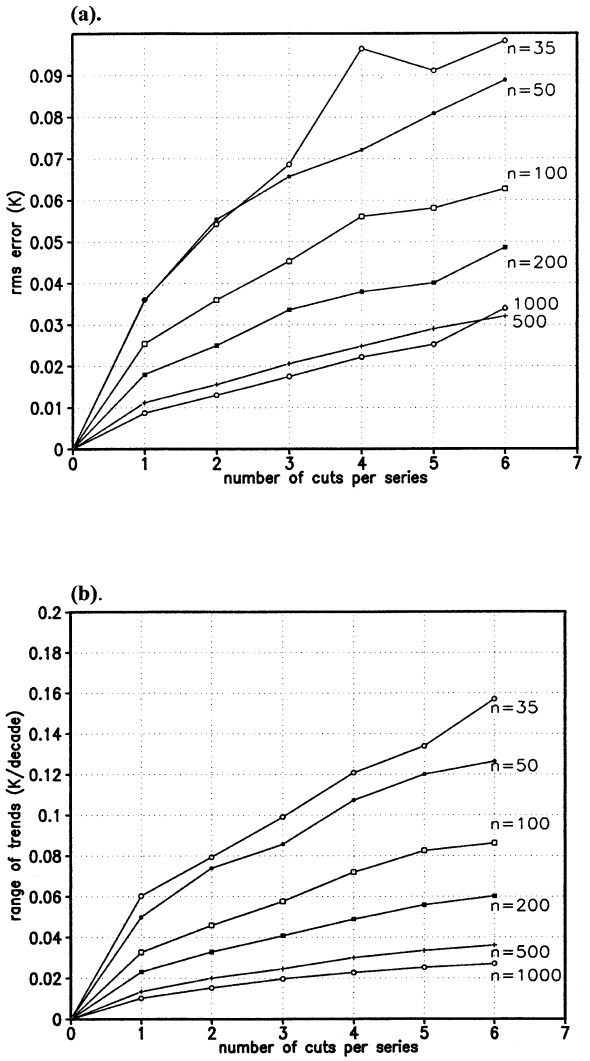

FIG. 5. (a) As in Fig. 4a, but using annual mean temperatures. (b) As in Fig. 4b, but using annual mean temperature.

the range of the trends varying only from 0.08 to 0.02 K decade^{-1}.

The calculations shown in Figs. 4 and 5 use endpoint outlier trimming at one standard deviation. Trend ranges without trimming (not shown) are between 50% and 300% larger; rms differences are only slightly larger. Using annual 500-mb data, the range of trends for 1000 iterations in a scenario approximating the number of cuts and missing data in the LKS 87-station dataset is $\pm$0.05 K decade^{-1}, compared to a trend of 0.13 with a standard error of 0.05 in the reanalysis data for 1958–97. The effects of the FD process are slightly larger near the surface and in the stratosphere than at 500 mb, and the range of trends for shorter time periods is generally larger than for longer periods. For 1979–97, for example, the range of trends can be almost twice the range for 1958–97.

The differences between the FD series and the mean of the complete series primarily represent random errors introduced by the FD process. These errors are relatively minor when combining thousands of time series, as was done for surface temperature data by Peterson et al. (1998), but become more troublesome when the number of series is small. For an 87-station dataset with three cuts per series, the rms error at 500 mb is estimated at 0.051 K and the error in trends is estimated at $\pm$0.034 K decade^{-1} for 1958–97. Since gaps in the original data will produce the same effects as cuts introduced deliberately, the number of data gaps should be minimized (as, e.g., through judicious use of interpolation).

5. Conclusions

We have examined the utility of using a first difference method to create large-scale radiosonde temperature time series free from data inhomogeneities. Our tests of the method indicate that FD offers significant resource and labor advantages over the LKS method, but introduces an error in large-scale mean time series that may not be acceptable in some cases. When we apply FD with endpoint outlier trimming to unadjusted LKS radiosonde temperature data, the trends in the resulting global mean time series for 1959–97 differ from the trends in the global mean of the adjusted LKS data by 0.00–0.07 K decade^{-1} in the troposphere and more than 0.1 K decade^{-1} in the stratosphere. Tests using NCEP–NCAR reanalysis data with simulated data deletions show that FD introduces a ''random walk'' error that is clearly greater than the error created by the data gaps without FD. This error increases approximately linearly with the number of data gaps and decreases nonlinearly with the number of time series combined. Using NCEP–NCAR data from locations corresponding to the 87 LKS sonde stations and masking the data to match the missing months in the actual LKS data, trends in the FD global series show an error 2–3 times the error of trends in the non-FD global mean when compared to temporally complete data.

In experiments with monthly mean 500-mb reanalysis temperature data, random walk effects introduced rms errors of over 0.1 K and changed trends by more than 0.1 K decade^{-1} for individual months. The use of annual means reduces these errors substantially. For datasets of 500–1000 stations, annual mean rms errors were no more than 0.03 and errors in trends were under 0.02, less than the standard error of the trends and small in comparison with the interdataset differences described for upper-air datasets by Seidel et al. (2003).

Whether the errors introduced by the FD process are tolerable in a particular case will depend on the required accuracy and the offsetting benefits to be derived from its use. The primary benefit considered here is the reduction in the error caused by temporal inhomogeneities in the original series. The extent of the

error created by inhomogeneities is unfortunately not known. The adjustments made in LKS (excluding the effects of data deletions) have effects on global mean trends in the range of 0.01–0.10 K decade^{-1} (depending on pressure level and time period), but the true effects of inhomogeneities could be larger (or smaller). A definitive conclusion about the net effect of the proposed FD procedure on total error is therefore not feasible.

First differencing provides a potential solution to the problem of homogeneity adjustment in radiosonde data. However, the approach must be used with caution when the number of stations is small or the data contain a large number of discontinuities or missing data. Our tests with reanalysis data suggest that to limit the rms error in annual global means to no more than ~0.05 K at 500 mb, a dataset containing 87 stations should have an average of no more than three temporal gaps per station for the period 1958–97.

Acknowledgments. This work was partially funded by the Climate Change Data and Detection element of NOAA's Office of Global Programs. We thank Ellen Cooter, Bruce Hicks, and anonymous reviewers for their helpful suggestions.

REFERENCES

Folland, C. K., and Coauthors, 2001: Observed climate variability and change. *Climate Change 2001: The Scientific Basis,* J. T. Houghton et al., Eds., Cambridge University Press, 99–182.

Free, M., and Coauthors, 2002: Creating climate reference datasets: CARDS workshop on adjusting radiosonde temperature data for climate monitoring. *Bull. Amer. Meteor. Soc.,* **83,** 891–899.

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

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

Lanzante, J. R., S. A. Klein, and D. J. Seidel, 2003: Temporal homogenization of monthly radiosonde temperature data. Part I: Methodology. *J. Climate,* **16,** 224–240.

Parker, D. E., and D. I. Cox, 1995: Toward a consistent global climatological rawinsonde database. *Int. J. Climatol.,* **15,** 473–496.

Peterson, T., T. Karl, P. Jamason, R. Knight, and D. 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.

Seidel, D. J., and Coauthors, 2004: Uncertainty in signals of large-scale climate variations in radiosonde and satellite upper-air temperature datasets. *J. Climate,* **17,** 2225–2240.

Shine, K., and Coauthors, 2003: A comparison of model-simulated trends in stratospheric temperatures. *Quart. J. Roy. Meteor. Soc.,* **129,** 1565–1588.

Tuomenvirta, H., 2004: Reliable estimation of climatic variations in Finland. Finnish Meteorological Institute Contribution No. 43, Finnish Meteorological Institute, Helsinki, Finland, 41 pp.