Uncertainty in Signals of Large-Scale Climate Variations in Radiosonde and Satellite Upper-Air Temperature Datasets

D. J. SEIDEL,* J. K. ANGELL,* J. CHRISTY,[+] M. FREE,* S. A. KLEIN,[#] J. R. LANZANTE,[#] C. MEARS,[@]
D. PARKER,[&] M. SCHABEL,[@] R. SPENCER,[+] A. STERIN,[**] P. THORNE,[&] AND F. WENTZ[@]

*NOAA/Air Resources Laboratory, Silver Spring, Maryland
[+]University of Alabama in Huntsville, Huntsville, Alabama
[#]NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey
[@]Remote Sensing Systems, Santa Rosa, California
[&]Met Office, Bracknell, Berkshire, United Kingdom
[**]All-Russian Research Institute of Hydrometeorological Information, Obninsk, Kaluga Region, Russia

(Manuscript received 3 May 2003, in final form 30 September 2003)

ABSTRACT

There is no single reference dataset of long-term global upper-air temperature observations, although several groups have developed datasets from radiosonde and satellite observations for climate-monitoring purposes. The existence of multiple data products allows for exploration of the uncertainty in signals of climate variations and change. This paper examines eight upper-air temperature datasets and quantifies the magnitude and uncertainty of various climate signals, including stratospheric quasi-biennial oscillation (QBO) and tropospheric ENSO signals, stratospheric warming following three major volcanic eruptions, the abrupt tropospheric warming of 1976–77, and multidecadal temperature trends. Uncertainty estimates are based both on the spread of signal estimates from the different observational datasets and on the inherent statistical uncertainties of the signal in any individual dataset.

The large spread among trend estimates suggests that using multiple datasets to characterize large-scale upper-air temperature trends gives a more complete characterization of their uncertainty than reliance on a single dataset. For other climate signals, there is value in using more than one dataset, because signal strengths vary. However, the purely statistical uncertainty of the signal in individual datasets is large enough to effectively encompass the spread among datasets. This result supports the notion of an 11th climate-monitoring principle, augmenting the 10 principles that have now been generally accepted (although not generally implemented) by the climate community. This 11th principle calls for monitoring key climate variables with multiple, independent observing systems for measuring the variable, and multiple, independent groups analyzing the data.

1. Introduction

Radiosonde and satellite observations have been used to create long-term global upper-air temperature datasets, which figure prominently in studies of large-scale climate variability and change. Different groups have addressed data quality, spatial sampling, and temporal homogeneity issues differently, and no single data product has emerged as a generally recognized reference. Indeed, because none is based on observations traceable to reference standards, and all involve complex data-processing algorithms and expert judgements regarding quality control, adjustments, etc., objectively identifying one or more "best" datasets is currently a matter of intense research, which this paper seeks to inform. Thus, a suite of datasets is available to the scientific community, which allows us to measure uncertainty in estimates of the magnitude of signals of climate variations and changes as manifest in upper-air temperature over the past 2–5 decades.

Here we compare eight upper-air temperature datasets produced by six research groups. Previous intercomparisons (Hurrell and Trenberth 1998; Santer et al. 1999, 2000; Gaffen et al. 2000; National Research Council 2000; Hurrell et al. 2000; Ramaswamy et al. 2001), including those made for the three comprehensive Intergovernmental Panel on Climate Change (IPCC) assessment reports, have examined fewer datasets and have focused on temperature trends (particularly in the lower troposphere and at the surface), sensitivity to the choice of a statistical metric, and global sampling issues. We extend these studies by incorporating new datasets and examining other aspects of climate variability, as well as trends, with a goal of better characterizing the uncertainty of observational estimates of upper-air temperature changes. Our comparisons involve large-scale

Corresponding author address: Dian J. Seidel, NOAA/Air Resources Laboratory (R/ARL), 1315 East–West Highway, Silver Spring, MD 20910.
E-mail: dian.seidel@noaa.gov

(global, hemispheric, and tropical) averages, and we do not account for differences in spatial sampling among the datasets. Therefore, our assessments should be considered conservative, because some unquantified component of the differences we report is likely due to these sampling inconsistencies (Santer et al. 1999; Hurrell et al. 2000). Nevertheless, our analyses are of practical importance because climate researchers often use these spatially incongruous datasets interchangeably.

This study attempts to ascertain whether upper-air data products are indeed interchangeable, by examining quantitatively the uncertainty in estimates of the strength of various climate signals. If the statistical uncertainty in the signal strength estimate in a single dataset (arising from variability in the time series not associated with the signal in question) is large enough to encompass the signal strength estimate (with its own associated uncertainty) from an alternate dataset, then the choice of dataset does not unduly influence the result. If, on the other hand, the difference in estimates from different datasets is larger than their individual uncertainties, the choice of dataset will more strongly influence the results, and, unless one or more datasets can be objectively identified as superior, use of multiple datasets will give a more complete picture of the overall uncertainty.

Section 2 describes the eight datasets, and section 3 presents the global and regional time series used in this intercomparison. Section 4 presents cross correlations among the datasets, and section 5 compares basic measures of the variability of the time series—their standard deviations and autocorrelation structure. Section 6 examines the magnitude and uncertainty of estimates of temperature changes associated with four particular climate signals: El Niño–Southern Oscillation (ENSO), the tropospheric temperature shift of the late 1970s, the quasi-biennial oscillation (QBO), and the stratospheric response to volcanic eruptions. Section 7 compares linear trend estimates, and their uncertainties, for three different time periods. In sections 6 and 7 we compare, for each signal, the statistical uncertainty in the estimate of that signal in an individual dataset with the uncertainty arising from differences among datasets and present statistical measures to quantify both. A concluding section discusses the implications of our results for monitoring large-scale temperature change using these data products.

2. Datasets

Our intercomparison includes eight datasets prepared by six research teams. Each has been presented in peer-reviewed journals, most have figured in IPCC assessment reports, and several will likely be updated regularly for climate monitoring and research purposes. We include three datasets based on the Microwave Sounding Units [(MSU) and the Advanced MSU (AMSU)] that have flown on National Oceanic and Atmospheric Administration (NOAA) polar-orbiting satellites since
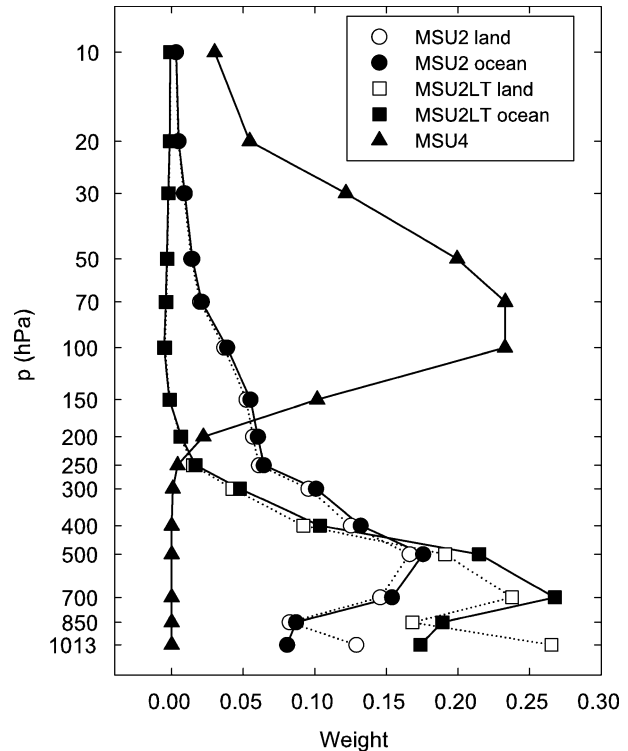

FIG. 1. Weights applied to LKS radiosonde pressure-level temperature anomaly data to simulate three MSU deep-layer mean temperature anomalies. Two separate sets of weights, for land and ocean regions, are given for the tropospheric MSU2 and MSU2LT; one set is given for the stratospheric MSU4. The level labeled 1013 hPa represents the surface.

1979 and five datasets based on subsets of the global radiosonde data archive. We employ no "blended" data products or reanalyses.

a. Satellite MSU datasets

1) UAH MSU VERSION D

John Christy, Roy Spencer, and colleagues at the University of Alabama in Huntsville (UAH) prepare this monthly, gridded, global temperature anomaly dataset. This study includes the UAH MSU data for the lower stratosphere (MSU4), troposphere (MSU2), and lower troposphere (MSU2LT). The vertical extent of these layers is shown in Fig. 1. Christy et al. (2000) describe quality control and procedures for merging data from different satellites for version D, the fourth version to be made publicly available since the first MSU temperature dataset (Spencer and Christy 1990).

2) UAH MSU VERSION 5.0

This version (5.0) is an update of the previous dataset, includes AMSU data, and incorporates a different (nonlinear rather than linear) correction for time-varying sampling of the diurnal cycle by the MSU instruments

due to drift in the local equatorial crossing time of the satellite orbits (Christy et al. 2003). This difference applies to MSU2 and MSU2LT, but not to MSU4, which is identical in versions D and 5.0. Both versions of UAH MSU data are available for the full period of our analysis.

3) RSS MSU

Frank Wentz, Carl Mears, and Matthias Schabel of Remote Sensing Systems, Inc. (RSS), have recently developed an MSU temperature dataset using different corrections and merging procedures than those used by UAH (Mears et al. 2003). This monthly, gridded, global temperature anomaly dataset covers the lower stratosphere (MSU4), tropopause region (MSU3, not examined in this paper), and troposphere (MSU2), but does not include a lower-tropospheric product comparable to the MSU2LT.

b. Radiosonde datasets

1) ANGELL-63

For several decades, Jim Angell of the NOAA/Air Resources Laboratory has presented global, hemispheric, and zonal seasonal temperature anomalies in three pressure layers (850–300, 300–100, and 100–50 hPa), based on daily sounding data from a 63-station network (Angell-63; Angell and Korshover 1975). No adjustments for data inhomogeneities are made. Details are described by Angell (1988) and references therein.

2) ANGELL-54

This new dataset is a revision of Angell-63 in which nine (out of the original 31) tropical (30°N–30°S) radiosonde stations, whose temperature trends in the 300–100-hPa layer were highly anomalous, were removed from the network (Angell-54; Angell 2003). For the two Angell datasets, we created monthly anomaly time series by linear interpolation of the seasonal anomaly data.

3) HADRT

David Parker, Margaret Gordon, and Peter Thorne of the Met Office's Hadley Centre for Climate Prediction and Research created monthly, gridded, global temperature anomalies based mainly on data from radiosonde stations providing monthly temperature (CLIMAT TEMP) reports (HadRT). Data are available at nine pressure levels between 850 and 30 hPa. Stratospheric data since 1979 have been adjusted using UAH MSU version-D data, but only at those stations where significant temperature changes were accompanied by known station history events. The version used in this paper is HadRT2.1s, which applies globally the adjustment method described by Parker et al. (1997), but only for

the stratosphere, because the tropospheric adjustments were not realistic (Thorne et al. 2002).

4) LKS

John Lanzante, Steve Klein (NOAA/Geophysical Fluid Dynamics Laboratory), and Dian Seidel (NOAA/Air Resources Laboratory) have prepared a new monthly global temperature anomaly dataset (LKS) based on data from 87 radiosonde stations by incorporating temporal homogeneity adjustments that are independent of other upper-air temperature datasets. The adjustments are based on a suite of indicators, including day–night temperature differences, the vertical structure of temperature, station history information, statistical measures of abrupt change, and indices of real climate variations. Data are available at the surface and 15 pressure levels from 1000 to 10 hPa, and are described by Lanzante et al. (2003a,b).

5) RIHMI

Alex Sterin at the All-Russian Research Institute of Hydrometeorological Information (RIHMI) has prepared monthly gridded temperature anomalies from the global radiosonde network using the Monthly Aerological Data Set (MONADS; Sterin and Eskridge 1998), which is based on the Comprehensive Aerological Reference Data Set (CARDS; Eskridge et al. 1995). Global gridded fields are based on station data, with spatial interpolation of station data to fill data-void regions (Sterin 1999). This study used data for the same three pressure layers as the two Angell datasets (850–300, 300–100, and 100–50 hPa).

3. Time series

a. Regions, layers, and time periods

To facilitate the intercomparison, we prepared monthly temperature anomaly time series from each dataset for four regions and for a subset of 23 levels and layers. The regions are defined as Globe, Northern Hemisphere (0°–90°N; NH), Southern Hemisphere (0°–90°S; SH), and Tropics (30°N–30°S). The levels and layers include the surface, 15 pressure levels, the three layers defined by Angell (1988), and four MSU layers (MSU2, -2LT, -3, and -4, see UAH MSU and RSS MSU above). In this paper we focus on three MSU layers (MSU2, -2LT, and -4) and the three pressure layers defined by Angell (1988). (The datasets are provided as an electronic supplement to this paper and can be found online at http://dx.doi.org/10.1175/3012.1.S1.)

Weighting functions were applied to zonal mean pressure-level radiosonde data from HadRT and LKS to simulate the MSU and Angell layers. To simulate the Angell layers, anomaly time series at relevant pressure levels were weighted by the logarithm of pressure. For the
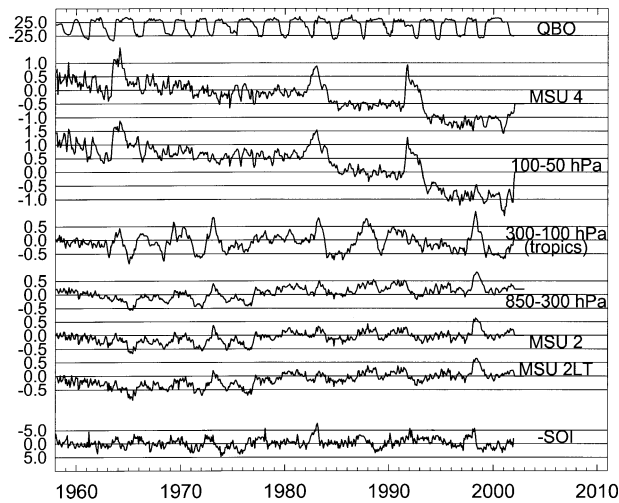
FIG. 2. The top and bottom traces are time series of the QBO (m s^{-1}) and normalized SOI (dimensionless), respectively. The QBO is represented by the 50-hPa zonal wind at Singapore. The SOI is based on Trenberth (1984). The other time series are multidataset-average monthly temperature anomalies for six different layers. All temperature time series are for the globe, except for the 300–100-hPa-layer time series, which is for the Tropics.

FIG. 3. Global temperature anomalies (K) in the 850–300-hPa layer from five radiosonde datasets. The bottom curve is the average of the five, and the other curves show deviations from this five-dataset average. LKS, RIHMI, and HadRT are monthly datasets; Angell-63 and Angell-54 data are seasonal datasets from which monthly values have been linearly interpolated.

three MSU layers, Fig. 1 shows weighting functions applied to the LKS pressure-level data. For MSU2 and MSU2LT, we used two different weighting functions—one suitable for land regions and one for ocean regions. However, because the results were nearly identical, we present only the results for land regions, because most radiosonde stations are land based. The LKS dataset has finer vertical resolution (including a surface level) than HadRT (where the lowest level is 850 hPa), so different weights (not shown) were given to individual pressure levels for HadRT data. In all cases, we required that data were available at enough levels to account for at least 75% of the weighting function, to avoid an undue influence from missing radiosonde data, particularly in the stratosphere.

Our intercomparisons are for three time periods: 1958–97 for the five radiosonde datasets, 1979–2001 for the satellite and radiosonde datasets (excluding LKS, which ends in 1997), and 1979–97 for all of the datasets.

b. Time series comparison plots

Figures 2–8 show 1958–2001 data for the six layers of interest. All are global time series, except one trace in Fig. 2 and all of Fig. 5, which show results for the tropical (rather than global) 300–100-hPa layer. Figure 2 shows the multidataset-average monthly anomaly time series (AVG) for each of the layers considered. Because of gross similarities among the datasets, the multidataset-average plots are a convenient depiction of the dominant variations in upper-air temperature, and the "stacked" time series in Fig. 2 show their vertical structure. Figure 2 also shows the QBO and Southern Os-
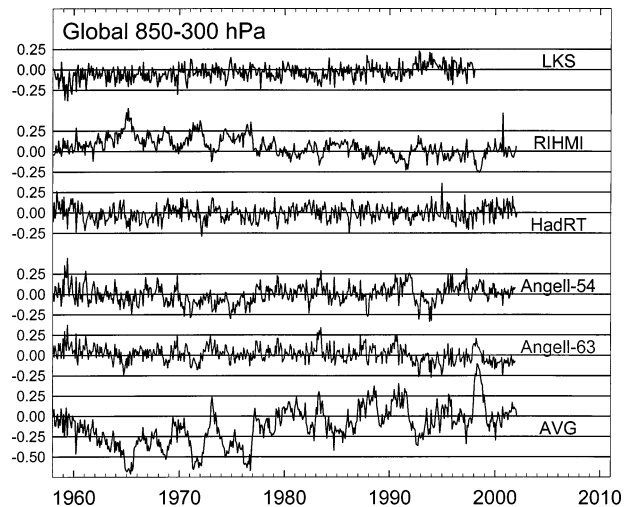
cillation index (SOI) time series. The QBO is represented by 50-hPa zonal wind variations based on radiosonde data from Singapore, and the SOI is from Trenberth (1984).

In each of Figs. 3–8, the AVG curves are the same as those in Fig. 2 and are based on all available datasets. The other curves in Figs. 3–8 show departures of individual time series from the AVG. Sections 4, 5, 6, and 7 discuss statistical aspects of the time series from which Figs. 3–8 are derived, and similar time series for the
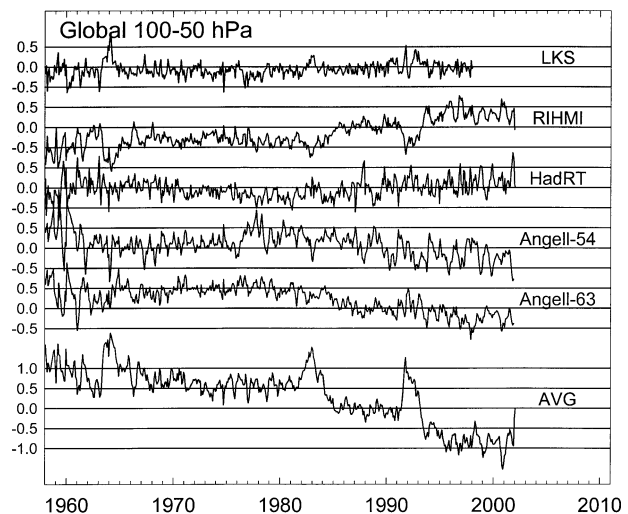

FIG. 4. Global temperature anomalies (K) in the 100–50-hPa layer from five radiosonde datasets. The bottom curve is the average of the five, and the other curves show deviations from this five-dataset average.
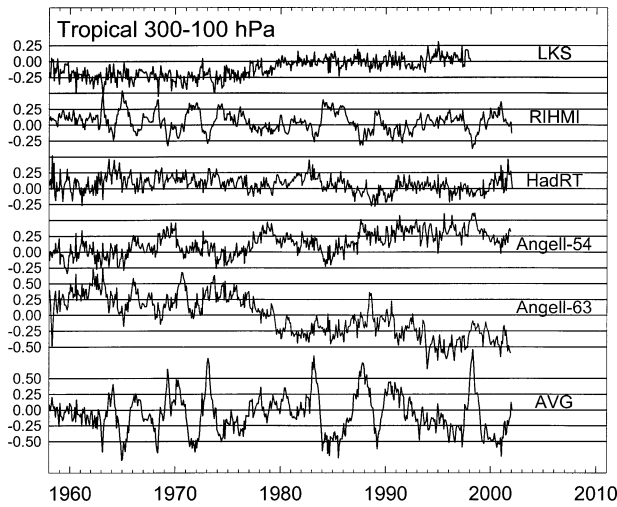
FIG. 5. Tropical temperature anomalies (K) in the 300–100-hPa layer from five radiosonde datasets. The bottom curve is the average of the five, and the other curves show deviations from this five-dataset average.
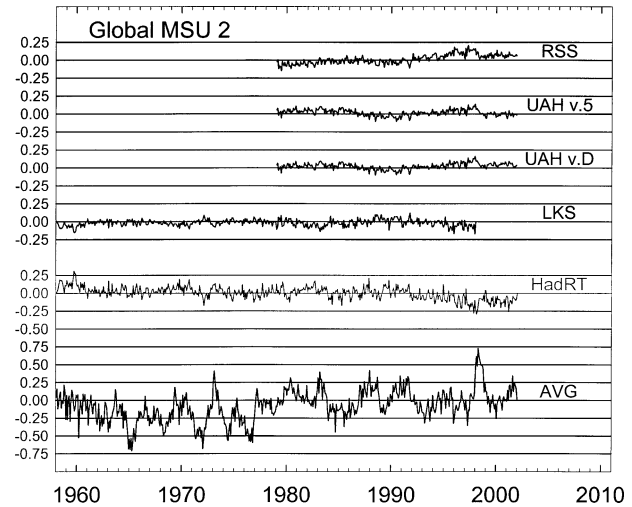

FIG. 7. Global temperature anomalies (K) in the MSU2 layer (troposphere) from two UAH MSU datasets, the RSS MSU dataset and (simulated) from two radiosonde datasets. The bottom curve is the average, and the other curves show deviations from the average.

other regions (NH, SH, and the Tropics). Here we point out a few salient features in the time series.

1) TROPOSPHERIC TIME SERIES

The global tropospheric (MSU2LT, MSU2, and 850–300 hPa) temperature anomaly datasets show, in the AVG curves (Fig. 2), long-term warming of the troposphere that is not monotonic but has considerable variability on monthly, interannual, and interdecadal time scales. The time-lagged influence of the Southern Oscillation (Fig. 2) is apparent, as is a relatively abrupt warming of several tenths of a degree in 1976–77. The range of anomalies is slightly larger than 1 K in all three global tropospheric time series.

Departures of individual datasets from the AVG are

not simply random noise but have steplike changes, long-term trends, and interannual signals. For example, the global 850–300-hPa LKS data (Fig. 3) have a stronger upward trend than the average, and the RIHMI data have a weaker trend. For the MSU2 and MSU2LT layers (Figs. 6 and 7), the two versions of UAH data are very similar. Differences between the radiosonde and satellite data products for the MSU2 layer are particularly noticeable in the 1990s, when both radiosonde datasets (HadRT and LKS) are cooler than AVG, while the RSS
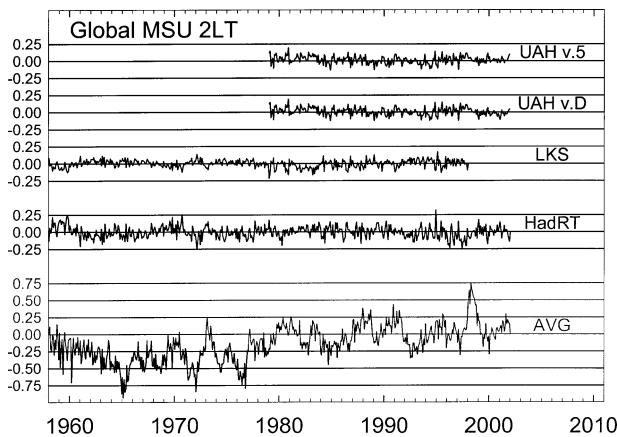

FIG. 6. Global temperature anomalies (K) in the MSU2LT layer (lower troposphere) from two UAH MSU datasets and (simulated) from two radiosonde datasets. The bottom curve is the average, and the other curves show deviations from the average.
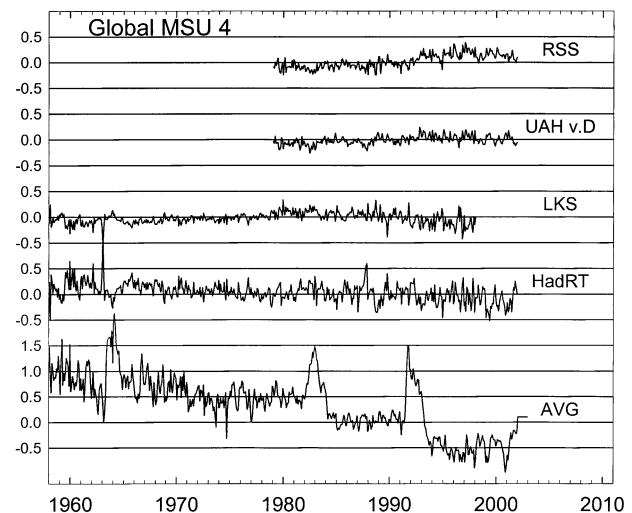

FIG. 8. Global temperature anomalies (K) in the MSU4 layer (lower stratosphere) from the UAH MSU and the RSS MSU dataset and (simulated) from two radiosonde datasets. UAH MSU versions D and 5.0 are identical for this layer and so only version D is shown. The bottom curve is the average of the four, and the other curves show deviations from the average.

TABLE 1. Correlations among global radiosonde temperature anomaly time series for 1958–97 in three layers. Values in the lower-left triangles of each matrix are for detrended monthly anomalies, and those in the upper right are for annual anomalies.

850–300 hPa	Angell-63	Angell-54	HadRT	LKS	RIHMI
Angell-63		0.97	0.96	0.92	0.96
Angell-54	0.94		0.95	0.92	0.94
HadRT	0.81	0.76		0.95	0.97
LKS	0.75	0.70	0.86		0.93
RIHMI	0.81	0.77	0.87	0.82	

300–100 hPa	Angell-63	Angell-54	HadRT	LKS	RIHMI
Angell-63		0.89	0.80	0.49	0.86
Angell-54	0.88		0.83	0.63	0.85
HadRT	0.72	0.66		0.79	0.95
LKS	0.77	0.72	0.84		0.78
RIHMI	0.74	0.66	0.88	0.83	

100–50 hPa	Angell-63	Angell-54	HadRT	LKS	RIHMI
Angell-63		0.96	0.92	0.94	0.94
Angell-54	0.86		0.89	0.91	0.92
HadRT	0.62	0.50		0.98	0.98
LKS	0.75	0.64	0.83		0.96
RIHMI	0.70	0.59	0.86	0.80	

TABLE 2. Correlations among global satellite and radiosonde temperature anomaly time series for 1979–97 in three layers. Values in the lower-left triangles of each matrix are for detrended monthly anomalies, and those in the upper right are for annual anomalies.

MSU4	UAH (versions D and 5)		RSS	HadRT	LKS
UAH			0.99	0.98	0.98
RSS	0.98			0.97	0.96
HadRT	0.89		0.89		0.99
LKS	0.92		0.90	0.86	

MSU2	UAH version D	UAH version 5.0	RSS	HadRT	LKS
UAH version D		1.00	0.91	0.83	0.86
UAH version 5.0	1.00		0.88	0.85	0.86
RSS	0.98	0.98		0.66	0.78
HADRT	0.82	0.82	0.83		0.93
LKS	0.84	0.84	0.85	0.86	

MSU2LT	UAH version D	UAH version 5.0		HadRT	LKS
UAH version D		1.00		0.92	0.91
UAH version 5.0	1.00			0.92	0.90
HadRT	0.79	0.78			0.87
LKS	0.79	0.79		0.78	

and UAH version D satellite datasets are warmer than AVG.

2) STRATOSPHERIC TIME SERIES

The two AVG global stratospheric time series (Fig. 2) show cooling over the 1958–2001 period, punctuated by strong transient warming episodes in the early 1960s, early 1980s, and early 1990s, associated with volcanic eruptions. These global time series also reveal a hint of a QBO signal, which is more prominent in their tropical counterparts (not shown).

The stratospheric anomaly time series appear to have a much higher signal-to-noise ratio than in the troposphere (Fig. 2). The range of the anomalies in the AVG stratospheric curves is ~3 K, compared with ~1 K for the troposphere. Similarly, differences from the AVG for the stratosphere (Figs. 4 and 8) are also greater than for the troposphere (Figs. 3, 6, and 7).

Compared with the radiosonde AVG, the LKS, RIHMI, and HadRT datasets indicate less cooling, while the Angell-63 dataset shows more cooling (Fig. 4), and the LKS dataset appears to have a stronger volcanic warming signal than AVG. Compared with the MSU4 layer AVG (Fig. 8), the RSS and UAH (version D, shown, and version 5.0, which is not shown but is identical), indicate less cooling, while the equivalent MSU4 temperature in HadRT and LKS shows more cooling.

3) TROPICAL 300–100-HPA-LAYER TIME SERIES

As we show in sections to follow, the 300–100-hPa layer exposes, in several respects, the most significant

differences among the radiosonde datasets, and these are particularly apparent in the tropical region, shown in Fig. 5. The variations in the AVG curve are large (with a range of almost 2 K), and the deviations of individual datasets from the average show more structure than for other layers. The LKS difference (from the AVG) time series has an upward shift near the middle of the time series, whereas the HadRT difference has a downward shift in the mid-1980s. The RIHMI difference time series is anticorrelated with the average, perhaps owing to the weakening of anomalies by interpolation, which also allows for information from extratropical regions to influence the tropical data. The Angell-63 dataset shows significantly more cooling than the AVG, while the Angell-54 data (with nine fewer tropical stations) shows substantially more warming. (In fact, the two Angell time series for the tropical 300–100-hPa layer have a correlation of only 0.50, much lower than their correlations for other regions and layers.)

4. Cross correlations among datasets

Correlations among pairs of datasets, shown in Tables 1, 2, and 3, can give us a quantitative sense of their overall agreement, but without identifying specific common or disparate signals, which we address in sections 5–8. Table 1 deals with global radiosonde datasets for 1958–97, Table 2 addresses global MSU datasets and radiosonde-simulated MSU for 1979–97, and Table 3 compares the NH and SH correlations for the MSU2 layer. For each table, we show correlations computed

TABLE 3. Correlations among hemispheric satellite and radiosonde temperature anomaly time series for 1979–97 for the MSU2 layer. Values in the lower-left triangles of each matrix are for detrended monthly anomalies, and those in the upper right are for annual anomalies.

NH	UAH version D	UAH version 5.0	RSS	HadRT	LKS
UAH version D		1.00	0.96	0.81	0.92
UAH version 5.0	1.00		0.95	0.81	0.91
RSS	0.96	0.96		0.66	0.85
HadRT	0.87	0.87	0.85		0.92
LKS	0.85	0.85	0.84	0.86	

SH	UAH version D	UAH version 5.0	RSS	HadRT	LKS
UAH version D		0.99	0.85	0.78	0.78
UAH version 5.0	1.00		0.79	0.80	0.78
RSS	0.96	0.96		0.64	0.67
HadRT	0.66	0.66	0.69		0.85
LKS	0.76	0.76	0.76	0.72	

two ways. Values in the lower-left triangle of the matrix are based on detrended monthly anomaly time series, and so measure the degree of association of short-term (monthly to interannual) variations. Values in the upper-right triangle are based on annual anomaly time series (without detrending), and so are more sensitive to longer-term (multiyear to multidecade) variability.

Over the 40-yr period 1958–97, the detrended monthly anomaly time series from radiosonde datasets have correlations ranging from ~0.6 to ~0.9 (Table 1, lower-left triangles of each matrix), so that the time series share only about 30%–80% common variance on short time scales. The strongest correlations are between the two Angell datasets. The correlations with the other datasets are lower for Angell-54 than for Angell-63, which may be due to the even sparser sampling in Angell-54. It also may suggest that some of the problems associated with the nine stations removed from Angell-63 to create Angell-54 may remain in the LKS, HadRT, and RIHMI products. On longer time scales (Table 1, upper-right triangles), the correlations in the 850–300- and 100–50-hPa layers are substantially higher, always exceeding 0.89, but the 300–100-hPa-layer correlations are lower. The lower correlations at longer time scales in the 300–100-hPa layer may be due to the dominant effects of data adjustments (or lack thereof) over relatively weak trends in this layer.

For the satellite period 1979–97, Table 2 shows that the UAH version D, UAH version 5.0, and RSS versions of MSU4 and MSU2 are very highly correlated (0.98–1.00) on short time scales. On long time scales, for the MSU2 layer, correlations between UAH and RSS are somewhat lower (0.88–0.91), suggesting that these data products agree very well in terms of monthly to interannual scale temperature changes and differ mainly in terms of trend (Mears et al. 2003). The MSU-layer-

average time series simulated from radiosonde data show lower correlations with the actual MSU data than the correlations between MSU products, perhaps due to differences in spatial sampling. The generally lower correlations for the detrended monthly anomalies than for the annual anomaly time series suggests that this spatial sampling problem in the radiosonde data affects the representation of shorter time-scale variations more than longer time scales.

Table 2 also indicates that correlations with MSU data are similar for LKS and HadRT data. This is somewhat surprising because the HadRT stratospheric data from some stations are adjusted to the UAH MSU version D data, whereas the LKS data are independently adjusted. In addition, LKS data are adjusted in the troposphere, but HadRT data are not. That the two sets of adjustments yield similar, and high, correlations with MSU on the global scale is encouraging.

However, the correlations among radiosonde datasets for the MSU2 layer are systematically higher for the NH than the SH, as shown in Table 3, which is likely due to the poorer sampling of the SH than the NH by the radiosonde network. Correlations among the two UAH and the RSS datasets do not show this hemispheric difference for short-time-scale variations. The correlations between LKS and HadRT, for short time scales, are 0.86 and 0.70 in the NH and SH, respectively, and, for long time-scale variations, are 0.92 (NH) and 0.85 (SH).

In summary, these cross-correlation tables show that correlations are lower among radiosonde datasets than among MSU satellite datasets, correlations for radiosonde datasets are generally higher for longer-time-scale variations than for shorter time scales, satellite datasets, on the other hand, are better correlated at short than long time scales (probably because the datasets differ mainly in their treatment of transitions between satellite platforms), and correlations involving radiosonde datasets are poorest in the 300–100-hPa layer for long time scales, and in the SH, where spatial sampling is poor and differs among the radiosonde datasets.

5. Variability and autocorrelation within individual temperature time series

a. Standard deviations of time series

Standard deviations are a basic measure of variability of time series and provide a context within which to examine the more climate-relevant measures of temperature variability discussed below. Standard deviation (and, in the next section, autocorrelation) results are based on the global monthly anomaly time series for 1979–97, the period covered by all the datasets. The time series were each detrended before the standard deviations and autocorrelations were computed, so that differences in trends, which enhance both statistics, would not influence the intercomparison.

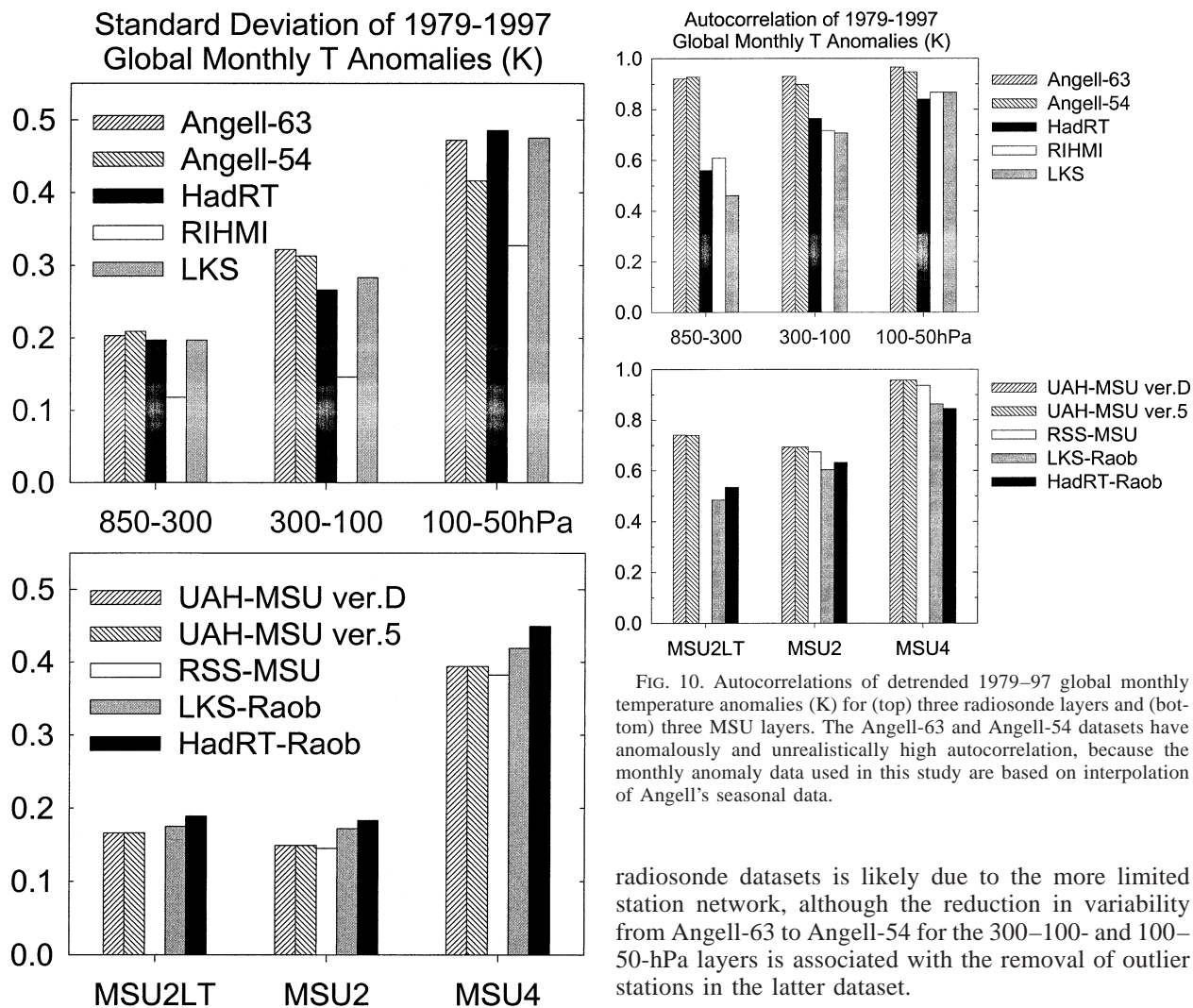## Standard Deviation of 1979-1997 Global Monthly T Anomalies (K)

FIG. 9. Standard deviations (K) of detrended 1979–97 global monthly temperature anomalies (K) for (top) three radiosonde layers and (bottom) three MSU layers.

Autocorrelation of 1979-1997 Global Monthly T Anomalies (K)
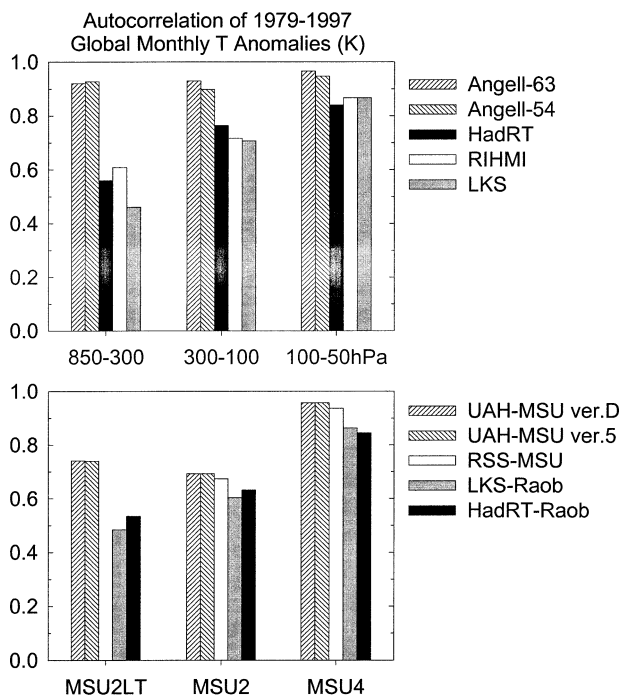

FIG. 10. Autocorrelations of detrended 1979–97 global monthly temperature anomalies (K) for (top) three radiosonde layers and (bottom) three MSU layers. The Angell-63 and Angell-54 datasets have anomalously and unrealistically high autocorrelation, because the monthly anomaly data used in this study are based on interpolation of Angell's seasonal data.

Figure 9 shows standard deviations of the detrended radiosonde time series (top panel) and of the satellite and radiosonde-simulated satellite datasets (bottom panel). Standard deviations in the stratosphere (100–50-hPa and MSU4 layers) are typically ~0.4 K, about twice as large as in the troposphere (850–300-hPa and MSU2LT and MSU2 layers). The 300–100-hPa layer has standard deviations of about 0.3 K.

The variability of most of the datasets is comparable for a given layer, with the exception of RIHMI, which has markedly smaller standard deviations than the other radiosonde datasets. This may be due both to the inclusion of many more stations in RIHMI than in the other radiosonde datasets and to the spatial interpolation of zonal mean values to areas of missing data in the RIHMI dataset. The slightly greater variability of the Angell tropospheric datasets compared with the other radiosonde datasets is likely due to the more limited station network, although the reduction in variability from Angell-63 to Angell-54 for the 300–100- and 100–50-hPa layers is associated with the removal of outlier stations in the latter dataset.

b. Autocorrelations and degrees of freedom

Lag-one autocorrelations[1] (Fig. 10) of the detrended global monthly anomaly time series generally exceed 0.8 in the stratosphere (100–50-hPa and MSU4 layers). Autocorrelations are smaller in the troposphere, but there is more variability among the datasets. Most radiosonde-based autocorrelations in the 850–300- and 300–100-hPa layers are approximately 0.5 and 0.7 in the two layers, respectively (Fig. 10, top panel). However, the two Angell datasets have anomalously and unrealistically high autocorrelation (exceeding 0.9), because the monthly anomaly data used in this study are based on interpolation of Angell's seasonal data. A similar pattern of higher autocorrelation in the stratosphere than in the troposphere is also evident in the MSU layers (Fig. 10, bottom panel). The radiosonde-simulated MSU

---

[1] Lag-one autocorrelation is the correlation between a given time series and the same time series shifted by one time unit, in this case 1 month.
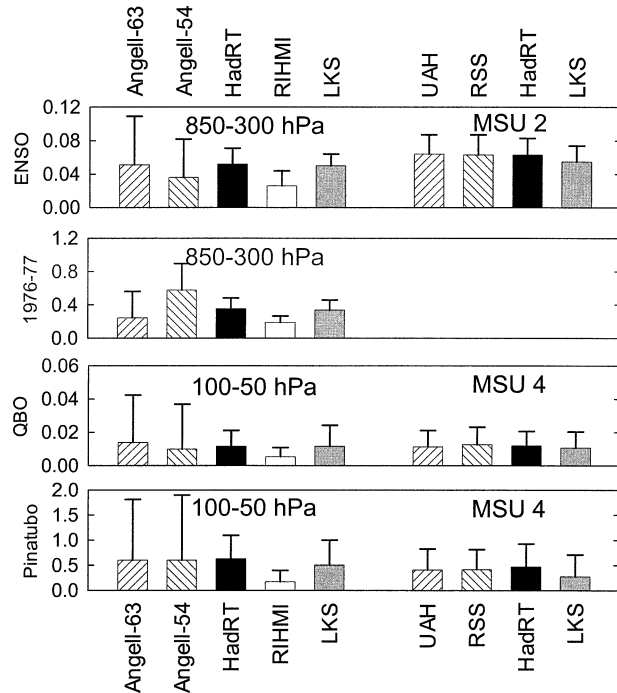
FIG. 11. Estimated magnitudes of four different climate signals from each applicable dataset. (top) The negative of the 5-month-lag regression coefficient (K), and its standard error, between tropical tropospheric temperature anomalies (K) and the Southern Oscillation index (nondimensional). (second panel) The magnitude of the tropical 850–300-hPa warming (K) during 1976–77, and its standard error. (third panel) The negative of the 3-month-lag regression coefficient (K s m^{-1}) between tropical stratospheric temperature anomalies and the QBO index (m s^{-1}), and the error bars are one standard error of the estimated regression coefficient. (bottom) The magnitude of the tropical stratospheric warming (K) following the Mount Pinatubo eruption, computed after removal of the QBO signal, and one standard error estimates.

layer time series have somewhat lower autocorrelation than the actual MSU time series, perhaps in part because the poorer spatial sampling of the radiosonde networks degrades the temporal autocorrelation of large-scale average temperature anomalies.

These high lag-one autocorrelations are associated with significant autocorrelations at longer time lags, which reduce the effective number of degrees of freedom n' in the time series. In these 228-month (1979–97) detrended global monthly temperature anomaly time series, n' can be estimated as $n(1 - r)/(1 + r)$, where r is the lag-one autocorrelation (Laurmann and Gates 1977). For r values of 0.9, 0.7, and 0.5, we obtain n' of 12, 40, and 76, respectively. The significant effect of this large reduction in n' is to enlarge estimates of uncertainty in climate signal strengths, as demonstrated by Santer et al. (2000) and as seen below.

6. Estimates of the magnitude of large-scale climate variations

In this section we explore how each dataset reveals four particular types of temperature variability: ENSO

TABLE 4. The response of global and tropical tropospheric temperature to the ENSO. Units are Kelvins per unit change in the dimensionless (normalized) SOI. Values in the table are the coefficients of linear regression between the detrended temperature anomaly time series and the SOI (Trenberth 1984), with a 5-month lag. Results are shown for (top) the tropical 850–300-hPa layer and (bottom) the MSU2 layer, and standard errors are given in parentheses.

850–300 hPa	Globe	Tropics
Angell-63	−0.020 (0.043)	−0.051 (0.058)
Angell-54	−0.009 (0.049)	−0.036 (0.046)
HadRT	−0.012 (0.014)	−0.052 (0.019)
LKS	−0.013 (0.012)	−0.050 (0.019)
RIHMI	−0.005 (0.009)	−0.026 (0.014)
MSU2	Globe	Tropics
UAH	−0.025 (0.013)	−0.064 (0.023)
RSS	−0.024 (0.012)	−0.063 (0.024)
HadRT	−0.024 (0.015)	−0.063 (0.020)
LKS	−0.019 (0.013)	−0.055 (0.019)

variability, steplike decadal change, the QBO, and climate response to episodic volcanoes. Quantitative estimates of the magnitude of each signal are given, together with estimates of their uncertainty.

a. Global and tropical response to El Niño–Southern Oscillation

A simple estimate of the strength of the ENSO signal is the coefficient of linear regression between tropospheric temperature (in the 850–300-hPa and MSU2 layers) and SOI [updated time series based on Trenberth (1984)] time series, with a 5-month lag (Angell 2000; Free and Angell 2002), based on detrended 1979–97 data. The standard error (adjusted for temporal autocorrelation effects) is a measure of the uncertainty in the regression coefficient. Santer et al. (2001) give a more comprehensive analysis, examining sensitivity to the treatment of volcanic effects and to the assumed lag time.

As seen in Fig. 11 (top panel), all the tropical datasets show a negative correlation with the SOI, and Table 4 shows larger values for the Tropics than globally. The regression coefficients can be interpreted as the temperature change (K) per unit change in the SOI. For the 850–300-hPa layer in the Tropics, the median value (from Table 4) of the signal is −0.050 K. Given an SOI range of about 10 units (Fig. 2), tropical tropospheric temperature changes of up to ~0.5 K can be associated with ENSO variations. (Note that the negative regression coefficients have been multiplied by −1 for presentation in Fig. 11.)

The signal is stronger in MSU2 than 850–300 hPa, as seen by direct comparison of the two layers for the LKS and HadRT results and by the overall patterns. The signal strength for RIHMI is the weakest, consistent with the low standard deviations of the RIHMI time series. The MSU datasets show better agreement than do the radiosonde datasets, which is likely due to the

TABLE 5. For signals of various climate variations in upper-air temperature, for selected regions [globe (G), Tropics (T), Northern Hemisphere (NH), and Southern Hemisphere (SH)] and for various radiosonde pressure layers and MSU layers, the table shows the median value of the signal strength, the MSE, the PSD, and the radio $R = |2 \times \text{MSE/PSD}|$. For each signal, the statistics are based on estimates from four or five different datasets, depending on the time period and layer. Trends are expressed in K decade^{-1}; the volcanic signals and the 1976–77 shift are expressed in K; and the QBO and ENSO signals are regression coefficients, as discussed in the text, with units of K s m^{-1} and K, respectively.

Signal	Region	Layer	Median	MSE	PSD	R
ENSO	T	850–300	−0.050	0.019	0.008	5.0
ENSO	T	MSU2	−0.064	0.023	0.001	40.6
ENSO	G	850–300	−0.013	0.014	0.005	5.5
ENSO	G	MSU2	−0.025	0.013	0.002	16.3
1976–77 shift	T	850–300	0.34	0.13	0.10	2.7
1976–77 shift	G	850–300	0.30	0.09	0.07	2.6
QBO	T	100–50	−0.0115	0.0127	0.0015	16.8
QBO	T	300–100	−0.0051	0.0101	0.0010	20.1
QBO	T	MSU4	−0.0116	0.0098	0.0009	21.6
QBO	G	MSU4	−0.0003	0.0077	0.0022	7.1
Mount Agung	T	100–50	0.76	0.68	0.29	4.7
El Chichón	T	100–50	0.50	0.51	0.13	8.1
El Chichón	T	MSU4	0.34	0.71	0.03	41.9
Mount Pinatubo	T	100–50	0.58	0.50	0.18	5.4
Mount Pinatubo	T	MSU4	0.41	0.74	0.07	20.0
Mount Pinatubo	G	100–50	0.37	0.47	0.16	5.8
Mount Pinatubo	G	MSU4	0.33	0.61	0.12	9.9
1958–97 trend	G	100–50	−0.39	0.06	0.10	1.2
1958–97 trend	T	100–50	−0.41	0.09	0.10	1.7
1958–97 trend	NH	100–50	−0.37	0.05	0.02	3.9
1958–97 trend	SH	100–50	−0.46	0.07	0.32	0.4
1958–97 trend	G	300–100	−0.08	0.03	0.10	0.6
1958–97 trend	T	300–100	−0.02	0.04	0.11	0.7
1958–97 trend	G	850–300	0.10	0.02	0.03	1.6
1958–97 trend	T	850–300	0.13	0.03	0.03	1.9
1958–97 trend	NH	850–300	0.08	0.03	0.03	1.9
1958–97 trend	SH	850–300	0.13	0.02	0.04	1.0
1979–2001 trend	G	MSU4	−0.59	0.20	0.18	2.2
1979–2001 trend	T	MSU4	−0.49	0.16	0.26	1.2
1979–2001 trend	NH	MSU4	−0.60	0.13	0.19	1.3
1979–2001 trend	SH	MSU4	−0.56	0.14	0.15	1.8
1979–2001 trend	G	MSU2	0.01	0.04	0.06	1.5
1979–2001 trend	T	MSU2	0.03	0.09	0.08	2.1
1979–2001 trend	NH	MSU2	0.08	0.05	0.09	1.1
1979–2001 trend	SH	MSU2	−0.06	0.04	0.02	3.2

different spatial sampling of the radiosonde datasets, while the MSU datasets have almost identical sampling. (The RSS dataset includes no data poleward of 83.75° latitude, whereas UAH interpolates over the poles to fill this data void.) The signal strength for the global time series is about a factor of 3 smaller than for the Tropics (Table 4).

It is clear from Fig. 11 (top) that the spread of the ENSO signal strength among the datasets is generally smaller than the purely statistical uncertainties of individual estimates of signal strength. (Note, however, that standard errors in the plot are about 2–10 times larger than they would be if the reduction in n' due to autocorrelation, discussed above, had not been taken into account.) This suggests that the datasets are in good agreement regarding this particular signal, and that the estimate from almost any of the datasets, along with the associated uncertainty, will fairly represent the result one would obtain from the others.

To illustrate this point more quantitatively, Table 5 presents an analysis of uncertainties in estimating the ENSO signal, and the other large-scale climate variations addressed in the following sections of this paper. The table compares the uncertainty associated with individual estimates of signal strength to the uncertainty associated with the spread among estimates from different datasets using two basic statistics. The first is the median value of the standard error (MSE) of the signal strength, which represents the typical uncertainty associated with a single dataset. The second is a measure of the spread among the estimates of the signal strength from all available datasets. The standard deviation is a commonly used parametric statistic characterizing the spread. We employ instead the pseudo-standard deviation (PSD), which is simply the interquartile range (IQR) divided by 1.349. The latter factor makes the IQR, which encompasses only 50% of the spread, comparable to twice the standard error. Both the MSE and PSD are nonparametric statistics; therefore, outliers will not distort our results (Lanzante 1996).

TABLE 6. Increase in the 850–300-hPa temperature anomaly (K) associated with the 1976–77 climate shift. The increase is calculated as the difference between the 5-yr periods 1971–75 and 1978–82, and the standard error of each estimate is given in parentheses.

	Angell-63	Angell-54	HadRT	LKS	RIHMI
Globe	0.30 (0.24)	0.42 (0.23)	0.34 (0.09)	0.28 (0.07)	0.17 (0.05)
NH	0.27 (0.23)	0.50 (0.24)	0.35 (0.10)	0.28 (0.08)	0.21 (0.06)
SH	0.33 (0.27)	0.35 (0.22)	0.33 (0.08)	0.29 (0.07)	0.12 (0.04)
Tropics	0.25 (0.32)	0.58 (0.32)	0.35 (0.13)	0.34 (0.12)	0.19 (0.08)

To compare the two sources of uncertainty, we examine the ratio $R = 2 \times$ MSE/PSD, where the numerator reflects the purely statistical uncertainty of the individual estimates and the denominator reflects the uncertainty associated with the multiplicity of datasets. For $R \gg 1$, the uncertainty in signal strength estimates from individual datasets tends to encompass the spread in signal strength estimates from different datasets, suggesting that it is not vital to examine multiple datasets to fully capture the uncertainty. For climate signals with $R \sim 1$ or $R < 1$, the spread associated with different datasets is a significant factor in the overall uncertainty, underscoring the importance of using multiple datasets. Table 5 also shows the median values of the signal strength, for comparison with the uncertainties.

The first four rows of Table 5 show statistics for the ENSO signal strength, based on the data in Table 4. For both the Tropics and Globe regions, and for both the MSU2 and 850–300-hPa layers, MSE is substantially larger than PSD, confirming the visual impression given by Fig. 11, that the uncertainty in ENSO signal strength from individual datasets exceeds the spread among estimates from different datasets. Thus, we obtain $R > 1$ for the ENSO signal, which implies that the signal strength estimate based on any individual dataset, along with the associated uncertainty, will encompass the uncertainty associated with the spread among datasets. For some applications at least (e.g., assessing the strength of the ENSO signal in a climate model simulation), it may be sufficient to rely on a single dataset.

b. Tropospheric warming during 1976–77

Many investigators have noted the decadal-scale step-like warming of the troposphere that occurred in 1976–77 (e.g., Trenberth 1990; Graham 1994). In Table 6 we show the 850–300-hPa-layer warming, in all four regions, between the two periods 1971–75 and 1978–82, for all the radiosonde datasets. (This climate shift occurred before the start of the MSU observations.) The uncertainty in the warming signal strength is estimated based on the variances of the temperature anomalies in the two 5-yr periods, again taking into account the reduction in n' due to autocorrelation. Results for the Tropics are shown in Fig. 11 (second panel).

All five datasets show warming of a few tenths of a degree in all four regions. However, the magnitude and pattern of warming vary considerably. For all four regions, the Angell-54 dataset shows the greatest warm-

ing, while RIHMI shows the smallest, with differences of up to a factor of 3. Three datasets show the strongest signal in the Tropics, but RIHMI and Angell-63 have their largest shifts in the NH and SH, respectively (Table 6).

Table 5 provides uncertainty statistics for this climate signal. The median signal strength estimate for the Tropics and Globe regions are 0.34 and 0.30 K, respectively, with associated MSE values of 0.13 and 0.09. The PSD values are slightly smaller than the MSE values, yielding R values of ~3. Thus, like the ENSO signal, the 1976–77 warming signal can be reasonably well estimated from individual datasets, because the uncertainty in individual signal strength encompasses the spread in signal strength from the available suite of datasets.

c. Stratospheric manifestation of the quasi-biennial oscillation

To compare the stratospheric warming following three major volcanic eruptions, each occurring at different phases of the QBO, we first remove the QBO signal from the temperature time series. We estimate the strength of the QBO signal using the linear regression coefficient between stratospheric temperature anomaly time series and the QBO index (50-hPa zonal winds at Singapore), using data from June 1984 through May 1991, when there were no major volcanic eruptions. The standard error of the regression measures its uncertainty. The regression coefficients are largest in the tropical stratosphere, where they are maximum at a 3-month lag (temperature-lagging QBO). However, note that our tropical region (30°N–30°S) spans both the equatorial region, in which QBO winds lag the temperature, and the more poleward regions, in which the phase is reversed (Randel et al. 1999; Baldwin et al. 2001). Therefore, our QBO signal strength estimates are not comparable with those from other studies in which station data, or data from narrower zonal bands, are used. Figure 11 (third panel) shows the negative of the regression coefficients (K s m^{-1}) and their standard errors for the tropical 100–50-hPa- and MSU4-layer data.

The signal strength ranges from -0.0053 (for RIH-MI) to -0.0139 (for Angell-63) K s m^{-1}, with typical standard errors of comparable magnitude (except for Angell-54 and Angell-63). The median regression coefficient for the tropical 100–50-hPa layer is -0.0115 K s m^{-1}, so that changes in QBO winds exceeding 50 m s^{-1} between easterly and westerly phases explain

TABLE 7. Tropical stratospheric temperature increase (K) following three major volcanic eruptions (Mount Agung in Mar 1963, El Chichón in Apr 1982, and Mount Pinatubo in Jun 1991) in the 100–50-hPa layer, and (for the post-1979 eruptions) for the MSU4 layer. The standard errors of each estimate are shown in parentheses.

100–50 hPa	Mount Agung	El Chichón	Mount Pinatubo
Angell-63	0.95 (1.26)	0.50 (1.05)	0.60 (1.21)
Angell-54	0.76 (0.99)	0.77 (1.18)	0.58 (1.30)
HadRt	0.70 (0.38)	0.50 (0.48)	0.63 (0.48)
LKS	1.01 (0.68)	0.55 (0.51)	0.50 (0.50)
RIHMI	0.24 (0.12)	0.21 (0.27)	0.18 (0.23)

MSU4		El Chichón	Mount Pinatubo
UAH		0.40 (0.36)	0.41 (0.42)
RSS		0.34 (0.37)	0.41 (0.40)
HadRT		0.35 (0.47)	0.47 (0.45)
LKS		0.32 (0.48)	0.28 (0.43)

~0.5 K changes in stratospheric temperature. The signal strength in the tropical stratosphere is typically about 4 times larger than in the NH stratosphere, and comparable to the signal in the tropical 300–100-hPa layer. For other layers and regions, the QBO signal is not consistent in magnitude or sign among the datasets.

In Table 5, MSE values tend to exceed the median value of QBO signal strengths, partly because the signal is relatively weak in these large zonal bands, and partly because high autocorrelations inflate the standard error estimates. However, the PSD values are relatively small, so R values mostly exceed 10. Therefore, as with the ENSO and climate shift signals examined above, for the QBO signal, estimates from individual datasets encompass the spread among datasets.

d. Stratospheric warming following volcanic eruptions

The warming of the stratosphere following the eruptions of Mount Agung in Bali, in March 1963, El Chichón in Mexico, in April 1982, and Mount Pinatubo in the Philippines, in June 1991 is the most prominent signal in stratospheric temperatures (Fig. 2). Santer et al. (2001) have shown that quantifying the temperature response to volcanic eruptions can be complex due to coincident occurrences of El Niño events. Here our main interest is in comparing estimates from different datasets, not accurately measuring the climatic response to volcanic aerosol forcing, so we adopt a relatively straightforward approach.

Following Free and Angell (2002), we measure the magnitude of the warming in the tropical stratosphere as the difference in average temperature anomaly in the 24-months following minus the 24-months preceding each eruption, after first removing the QBO signal determined in the previous subsection. The uncertainty in signal strength is estimated based on the variances of the temperature anomalies in the two 24-month periods. Table 7 shows the results for all three eruptions for the 100–50-hPa layer, and for the two more recent eruptions

for the MSU4 layer. Figure 11 (bottom panel) shows the Mount Pinatubo results only.

Except in the Angell-54 dataset (and consistent with the 100-hPa findings of Free and Angell 2002), the response to Mount Agung was the largest of the three, with the other radiosonde datasets showing warming of 0.24–1.01 K in the 100–50-hPa layer. The radiosonde datasets show less warming in the simulated MSU4 layer than in 100–50 hPa. This may be due to the broad MSU4 weighting function (Fig. 1), which includes part of the tropical upper troposphere, where the volcanic signal is weaker. The warming of the global stratosphere (not shown) is typically about 30%–50% smaller than that of the tropical stratosphere in these datasets. All the radiosonde datasets show a stronger global response to the Mount Agung eruption than to the eruption of El Chichón or Mount Pinatubo, with median warming signals of 0.53, 0.17, and 0.37 K, respectively. As was the case with the other three climate signals, the volcanic signal is weakest in the RIHMI dataset.

Table 5 shows R values ranging from 4 to 42 for the volcanic signals, due to the relatively large values of MSE, which are comparable in magnitude to the median signal strength. This result may seem surprising, given the prominence of the volcanic warming signal in the stratospheric temperature time series (Figs. 2, 4, and 8). The reason is the inflation of the standard error estimates due to the very high lag-one autocorrelation of these time series (Fig. 10), which reduces the effective degrees of freedom in the 24-month time series segments used to estimate the volcanic signals. As with the other climate signals examined above, the large R values for the volcanic signal indicate that the uncertainty associated with individual datasets is larger than that associated with the spread among datasets.

7. Linear trends

Linear temperature trend estimates, based on ordinary least squares regression, and their ± 1 standard error confidence intervals are shown in Figs. 12, 13, and 14 for three different time periods and for different layers. Each figure shows trends for the Globe, Northern and Southern Hemispheres, and Tropics regions. Table 5 gives median trend values and uncertainty statistics. All trend estimates are expressed in units of K decade^{-1}.

a. Radiosonde trends for 1958–97

The radiosonde-layer trends for 1958–97 (Fig. 12) show warming in the 850–300-hPa layer (bottom panel) and cooling in the 100–50-hPa layer (top panel). The magnitudes of these trends vary markedly, particularly in the stratosphere, where the strongest trends (from Angell-63) are a factor of 2–4 larger than the weakest (from RIMHI). In a few cases, trends for a given region and layer do not overlap within the ± 1 standard error confidence intervals, but they generally do overlap with-
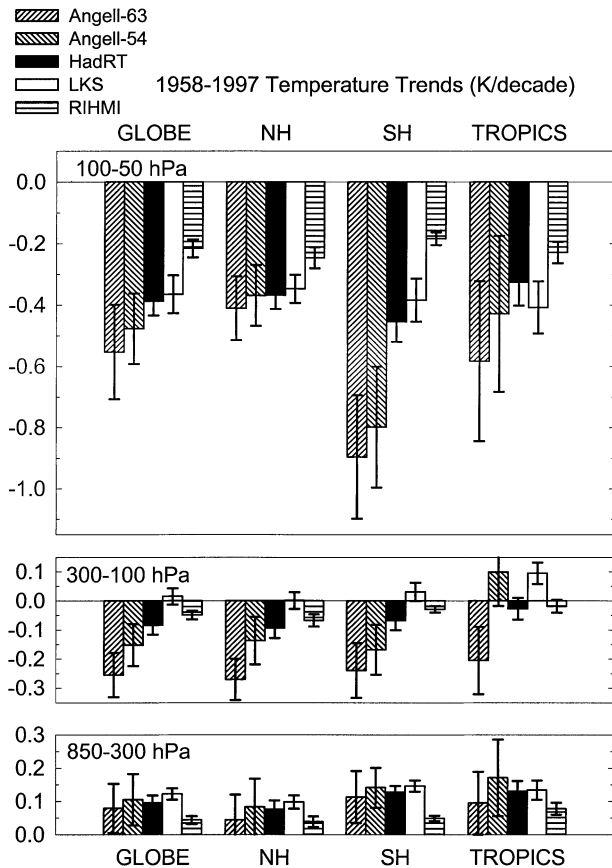
FIG. 12. Trends (K decade^{-1}) in global temperature for 1958–97 for three layers (top) 100–50, (middle) 300–100, (bottom) 850–300 hPa, in four regions, from radiosonde datasets. The confidence intervals shown are the ± 1 standard error uncertainty estimates.

FIG. 13. Trends (K decade^{-1}) in global temperature for 1979–97 for three layers (top) MSU4, (middle) MSU2, (bottom) MSU2LT, in four regions, from satellite and radiosonde datasets.

in ± 2 standard errors. In general, the RIHMI data have lower trends than the other radiosonde datasets and are clear outliers in the 850–300- and 100–50-hPa layers, but not in the 300–100-hPa layer. The LKS and HadRT data are in good agreement in the 100–50- and 850–300-hPa layers, but not in the 300–100-hPa layer, where the trends are of opposite sign. All the datasets indicate greater warming in the SH than the NH in the 850–300-hPa layer, and most suggest greater warming in the tropical belt than in either hemisphere.

In the 100–50-hPa layer, most of the datasets indicate greater cooling of the SH than the NH. The stratospheric trends are more consistent in the NH than in the SH, where they vary by about a factor of 4, probably reflecting the relatively better sampling of the NH by each radiosonde dataset. In the Tropics, where Angell removed nine stations from Angell-63 to create Angell-54, trends from Angell-54 are in better agreement with LKS and HadRT. The LKS data show fairly uniform cooling (0.35–0.40 K decade^{-1}) in each of the four regions, which is not the case in other datasets.

In the 300–100-hPa layer (Fig. 12, middle panel), there is little agreement among the datasets, particularly
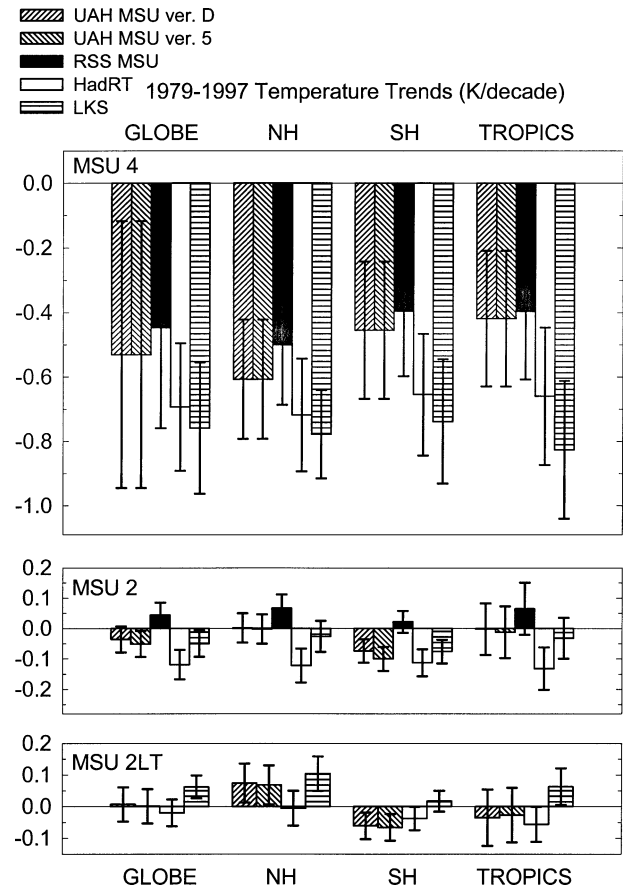
in the Tropics (see Fig. 5), where even the sign of the trend seems highly uncertain. The cooling in this layer is greatest for Angell-63, and is substantially reduced in Angell-54. (In the Tropics a 0.21 K decade^{-1} cooling in Angell-63 is transformed to a 0.10 K decade^{-1} warming in Angell-54.) Similarly, LKS, the one other dataset for which data adjustments affect tropospheric data below 150 hPa, also has positive trends in the Tropics, unlike HadRT and RIHMI, which have slight negative trends in the tropical 300–100-hPa layer.

b. Satellite and radiosonde trends since 1979

Figures 13 and 14 shows satellite-layer trends for 1979–97 and 1979–2001, respectively. In the stratosphere (MSU4, top panels), strong cooling, exceeding 0.3 K decade^{-1}, is found in all available datasets for both time periods. (Recall that the two UAH versions are identical for MSU4.) Stratospheric cooling in UAH is slightly larger than in RSS, but both satellite datasets show substantially less cooling than the HadRT (and for 1979–97 LKS) equivalent MSU4 temperatures. Furthermore, lower-stratospheric (100–50 hPa) trends for
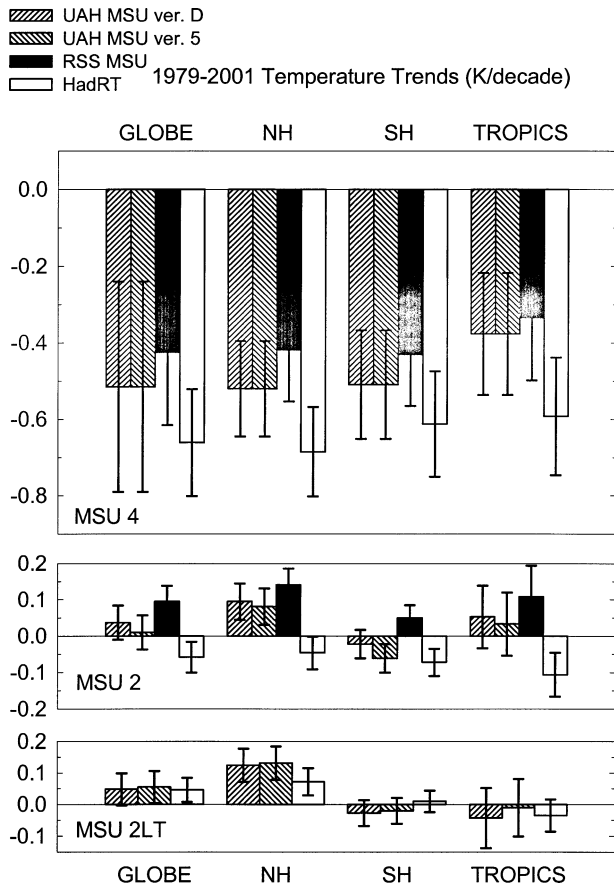
FIG. 14. Trends (K decade^{-1}) in global temperature for 1979–2001 for three vertical layers (top) MSU4, (middle) MSU2, (bottom) MSU2LT, in four regions, from satellite and radiosonde datasets.

NH has warmed over both periods, but the results for the other three regions are mixed. For MSU2LT, the two radiosonde-based datasets are the most different, with LKS showing warming in all four regions and HadRT showing cooling (Fig. 13). For MSU2LT, there is a striking contrast between SH cooling and NH warming for most datasets.

As shown in Table 5, only a few temperature trend estimates have $R > 2$. These include the NH 100–50-hPa cooling during 1958–97, the global MSU4 cooling during 1979–2001, and the tropical warming and SH cooling in MSU2 during 1979–2001. For all of the other trend estimates, for both the radiosonde and satellite data periods, $0.4 < R < 2$. Among the lowest R values shown are those for 1958–97 trends in the global and tropical 300–100-hPa layers, confirming that the spread among datasets for that layer is quite large. For these small values of R, the precision of estimates of signal strengths from individual datasets is high compared with the spread among estimates from the suite of datasets, which suggests that using more than one dataset would give a better sense of overall uncertainty. In this respect, the temperature trend signal is qualitatively different from the other climate signals examined above.

One implication of this result is that climate change detection and attribution studies, in which model simulations are compared with observations of temperature change, would be more robust if they included more than one observational estimate. Although the AVG time series presented here is not necessarily the ideal choice (because it included older and improved versions of some datasets), accounting somehow for the spread among datasets seems prudent.

1979–97 (not shown) from both Angell-63 and Angell-54 (but not from RIHMI) are also larger than the MSU4 trends, suggesting that radiosonde datasets, in general, show more stratospheric cooling than do satellite datasets. The satellite data show less cooling in the Tropics than the Globe whereas the radiosonde datasets show stronger cooling in the Tropics than the global average. The regions of greatest disparity in trend estimates are the Tropics and SH, where the stratosphere is particularly poorly sampled by radiosondes.

In the troposphere (MSU2, middle panel of Figs. 13 and 14) the HadRT data show cooling in all four regions for both time periods, whereas the RSS dataset indicates warming in all four regions. The UAH trends are generally smaller than RSS and show cooling in the SH for both periods. Other UAH cooling trends for 1979–97 (Fig. 13, middle) become warming trends for 1997–2001 (Fig. 14, middle), which includes the warming associated with the 1997/98 El Niño. Version 5.0 of the UAH data has smaller positive trends (or a more negative trend in the SH) than the earlier version D.

In the lower troposphere (MSU2LT, bottom panels of Figs. 13 and 14), the datasets appear to agree that the

8. Conclusions

We have examined and quantified the magnitude and uncertainty of signals of climate variations and change in eight upper-air temperature datasets based on radiosonde and satellite observations. Estimates of the ENSO and QBO signals, stratospheric warming following three major volcanic eruptions, the abrupt tropospheric warming of 1976–77, and temperature trends have been presented, along with quantitative uncertainty estimates based on the individual time series and on the spread among estimates from different time series. For most of the climate signals examined here, the statistical uncertainty in estimates of signal strength from individual datasets is large enough to encompass the uncertainty due to spread among estimates from different datasets. However, for temperature trends, the spread among estimates is large, and supports the notion of using multiple datasets to best characterize overall uncertainty. Thus, this research suggests that, for estimating global, hemispheric, and tropical upper-air temperature trends, both in the troposphere and stratosphere, it is important that the scientific community maintain and analyze sev-

eral climate monitoring datasets to best measure overall uncertainty.

This result is consistent with proposals for an 11th climate-monitoring principle, that complements the 10 principles proposed by Karl et al. (1995) and recommended by the National Research Council (1999), and calls for redundant methods of monitoring key climate variables. Multiple and independent observing systems should provide measurements, and multiple independent research groups should analyze the observations and provide climate-monitoring data products. As this paper demonstrates, such redundancy exists for upper-air temperature, and enhances our understanding of uncertainty in estimates of climate variations.

Future work should explore the effects of spatial and temporal sampling differences among the datasets and should address new versions of some of the datasets shown here. The RIHMI dataset is currently being revised based, in part, on the results of this study. The LKS dataset is being extended from 1997 to present to form the core of a new NOAA Radiosonde Atmospheric Temperature Products for Assessing Climate (RAT-PAC). HadRT is also being upgraded using spatial consistency checks and additional data. Analysis of MSU data by UAH and RSS is ongoing.

The global and regional temperature anomaly time series used in this study are available as an electronic supplement to this paper (see http://dx.doi.org/10.1175/3012.1.S1).

REFERENCES

Angell, J. K., 1988: Variations and trends in tropospheric and strato-spheric global temperatures, 1958–87. *J. Climate,* **1,** 1296–1313.

——, 2000: Tropospheric temperature variations adjusted for El Niño, 1958–1998. *J. Geophys. Res.,* **105,** 11 841–11 849.

——, 2003: Effect of exclusion of anomalous tropical stations on temperature trends from a 63-station radiosonde network, and comparison with other analyses. *J. Climate,* **16,** 2288–2295.

——, and J. Korshover, 1975: Estimate of the global change in tropospheric temperature between 1958 and 1973. *Mon. Wea. Rev.,* **103,** 1007–1012.

Baldwin, M. P., and Coauthors, 2001: The quasi-biennial oscillation. *Rev. Geophys.,* **39,** 179–229.

Christy, J. R., R. W. Spencer, and W. D. Braswell, 2000: MSU tro-

pospheric temperatures: Dataset construction and radiosonde comparisons. *J. Atmos. Oceanic Technol.,* **17,** 1153–1170.

——, ——, 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. Chemykh, Z. Panmao, A. C. Polansky, and S. R. Doty, 1995: A Comprehensive Aerological Reference Data Set (CARDS): Rough and systematic errors. *Bull. Amer. Meteor. Soc.,* **76,** 1759–1775.

Free, M., and J. K. Angell, 2002: Effect of volcanoes on the vertical temperature profile in radiosonde data. *J. Geophys. Res.,* **107,** 4101, doi:10.1029/2001JD001128.

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

Graham, N. E., 1994: Decadal-scale climate variability in the tropical and North Pacific during the 1970s and 1980s: Observations and model results. *Climate Dyn.,* **10,** 135–162.

Hurrell, J. W., and K. E. Trenberth, 1998: Difficulties in obtaining reliable temperature trends: Reconciling the surface and satellite microwave sounding unit records. *J. Climate,* **11,** 945–967.

——, S. J. Brown, K. E. Trenberth, and J. R. Christy, 2000: Comparison of tropospheric temperatures from radiosondes and satellites: 1979–98. *Bull. Amer. Meteor. Soc.,* **81,** 2165–2178.

Karl, T. R., and Coauthors, 1995: Critical issues for long-term climate monitoring. *Climatic Change,* **31,** 185–221.

Lanzante, J. R., 1996: Resistant, robust and nonparametric techniques for analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.,* **16,** 1197–1226.

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

——, ——, and ——, 2003b: Temporal homogenization of monthly radiosonde temperature data. Part II: Trends, sensitivities, and MSU comparison. *J. Climate,* **16,** 241–262.

Laurmann, J., and L. Gates, 1977: Statistical considerations in the evaluation of climatic experiments with atmospheric general circulation models. *J. Atmos. Sci.,* **34,** 1187–1199.

Mears, C. A., M. C. Schabel, and F. J. Wentz, 2003: A reanalysis of the MSU channel 2 tropospheric temperature record. *J. Climate,* **16,** 3650–3664.

National Research Council, 1999: *Adequacy of Climate Observing Systems.* National Academy Press, 51 pp.

——, 2000: *Reconciling Observations of Global Temperature Change.* National Academy Press, 85 pp.

Parker, D. E., M. Gordon, D. P. N. Cullum, D. M. H. Sexton, C. K. Folland, and N. Rayner, 1997: A new global gridded radiosonde temperature data base and recent temperature trends. *Geophys. Res. Lett.,* **24,** 1499–1502.

Ramaswamy, V., and Coauthors, 2001: Stratospheric temperature trends: Observations and model simulations. *Rev. Geophys.,* **39,** 71–122.

Randel, W. J., F. Wu, R. Swinbank, J. Nash, and A. O'Neill, 1999: Global QBO circulation derived from UKMO stratospheric analyses. *J. Atmos. Sci.,* **56,** 457–474.

Santer, B. D., J. J. Hnilo, T. M. L. Wigley, J. S. Boyle, C. Doutriaux, M. Fiorino, D. E. Parker, and K. E. Taylor, 1999: Uncertainties in observationally based estimates of temperature change in the free atmosphere. *J. Geophys. Res.,* **104,** 6305–6333.

——, T. M. L. Wigley, J. S. Boyle, D. J. Gaffen, J. J. Hnilo, D. Nychka, D. E. Parker, and K. E. Taylor, 2000: Statistical significance of trend differences in layer-average temperature time series. *J. Geophys. Res.,* **105,** 7337–7356.

——, and Coauthors, 2001: Accounting for the effects of volcanoes and ENSO in comparisons of modeled and observed temperature trends. *J. Geophys. Res.,* **106,** 28 033–28 059.

Spencer, R. W., and J. R. Christy, 1990: Precise monitoring of global temperature trends from satellites. *Science,* **247,** 1558–1562.

Sterin, A. M., 1999: An analysis of linear trends in the free atmo-

sphere temperature series for 1958–1997 (in Russian). *Meteor. Gidrol.,* **5,** 52–68.

——, and R. E. Eskridge, 1998: Monthly aerological data set: Some features and comparison of upper-air temperature data to the NCAR/NCEP reanalysis monthly data. *Proc. 22nd Annual Climate Diagnostics and Prediction Workshop,* Berkeley, CA, NOAA, 210–213.

Thorne, P. W., and Coauthors, 2002: Assessing the robustness of zonal mean climate change detection studies. *Geophys. Res. Lett.,* **29,** 1920, doi:10.1029/2002GL015717.

Trenberth, K. E., 1984: Signal versus noise in the Southern Oscillation. *Mon. Wea. Rev.,* **112,** 326–332.

——, 1990: Recent observed interdecadal climate changes in the Northern Hemisphere. *Bull. Amer. Meteor. Soc.,* **71,** 988–993.