

# Homogenization of Radiosonde Temperature Time Series Using Innovation Statistics

LEOPOLD HAIMBERGER

*Department of Meteorology and Geophysics, University of Vienna, Vienna, Austria*

(Manuscript received 21 February 2006, in final form 8 August 2006)

## ABSTRACT

Radiosonde temperature records contain valuable information for climate change research from the 1940s onward. Since they are affected by numerous artificial shifts, time series homogenization efforts are required. This paper introduces a new technique that uses time series of temperature differences between the original radiosonde observations (obs) and background forecasts (bg) of an atmospheric climate data assimilation system for homogenization.

These obs – bg differences, the “innovations,” are a by-product of the data assimilation process. They have been saved during the 40-yr ECMWF Re-Analysis (ERA-40) and are now available for each assimilated radiosonde record back to 1958. It is demonstrated that inhomogeneities in the obs time series due to changes in instrumentation can be automatically detected and adjusted using daily time series of innovations at 0000 and 1200 UTC.

The innovations not only reveal problems of the radiosonde records but also of the data assimilation system. Although ERA-40 used a frozen data assimilation system, the time series of the bg contains some breaks as well, mainly due to changes in the satellite observing system. It has been necessary to adjust the global mean bg temperatures before the radiosonde homogenization.

After this step, homogeneity adjustments, which can be added to existing raw radiosonde observations, have been calculated for 1184 radiosonde records. The spatiotemporal consistency of the global radiosonde dataset is improved by these adjustments and spuriously large day–night differences are removed. After homogenization the climatologies of the time series from certain radiosonde types have been adjusted. This step reduces temporally constant biases, which are detrimental for reanalysis purposes. Therefore the adjustments applied should yield an improved radiosonde dataset that is suitable for climate analysis and particularly useful as input for future climate data assimilation efforts. The focus of this paper relies on the lower stratosphere and on the internal consistency of the homogenized radiosonde dataset. Implications for global mean upper-air temperature trends are touched upon only briefly.

---

## 1. Introduction

Since the 1940s radiosondes are an essential component of the global atmospheric observing system. They reach farther back than satellite records and they also provide relatively high vertical resolution compared to satellite instruments, such as the Microwave Sounding Unit (MSU). Therefore they are a unique source of information about the upper-air climate and they are essential for climate data assimilation efforts, such as the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) or the National Centers for Environmental Prediction–National

Center for Atmospheric Research (NCEP–NCAR) re-analysis (Uppala et al. 2005; Kistler et al. 2001). While radiosondes measure temperature, humidity, and wind, this work’s analysis is restricted to temperature. The quality of the radiosonde instrumentation has improved over time. Systematic errors of radiosonde temperature measurements comprised several kelvin in the stratosphere during the 1960s and 1970s. The main, but not the only reason for systematic errors, was radiation effects. In the late 1980s, at many sites the radiation error was still larger than 1 K at the 50-hPa level, as is indicated in Fig. 1. With today’s modern sounding equipment the radiation error is reduced to a few tenths of a kelvin (Nash et al. 2005).

Figure 2 shows time series of mean 1200 – 0000 UTC differences of composites of radiosondes located between 30°W and 40°E as well as between 120°E and 120°W at the 50-hPa level. The 50-hPa level has been

---

*Corresponding author address:* Dr. Leopold Haimberger, Department of Meteorology and Geophysics, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria.  
E-mail: leopold.haimberger@univie.ac.at

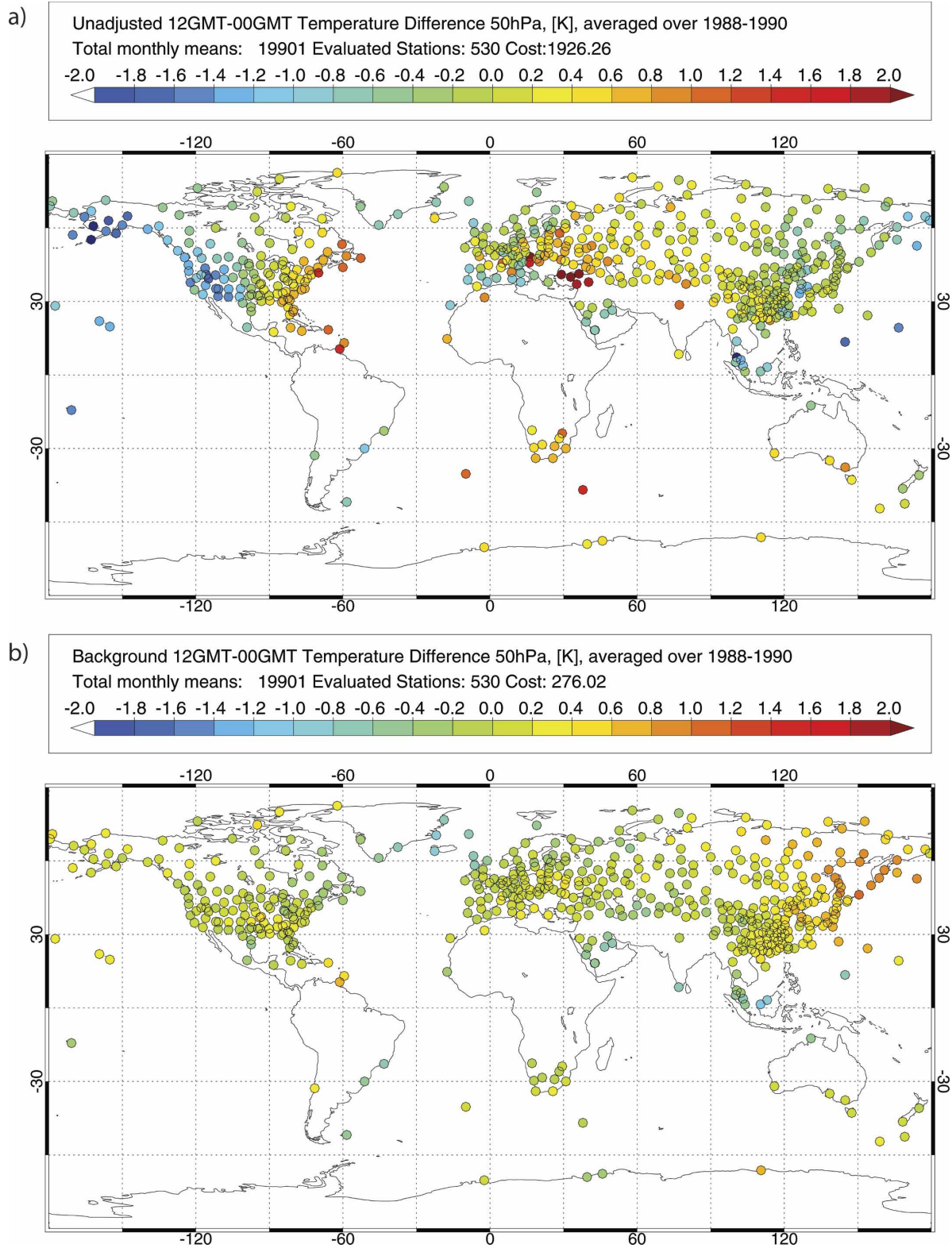


FIG. 1. The 1200 – 0000 UTC temperature differences at 50 hPa averaged over period 1988–90. Each bullet denotes a radiosonde station with more than 30 out of 36 months of data; color of bullets indicates the difference. “Cost” refers to a spatial consistency measure defined in appendix A. (a) Differences as measured by radiosondes. (b) Differences from the ERA-40 bg. These are spatially much more homogeneous.

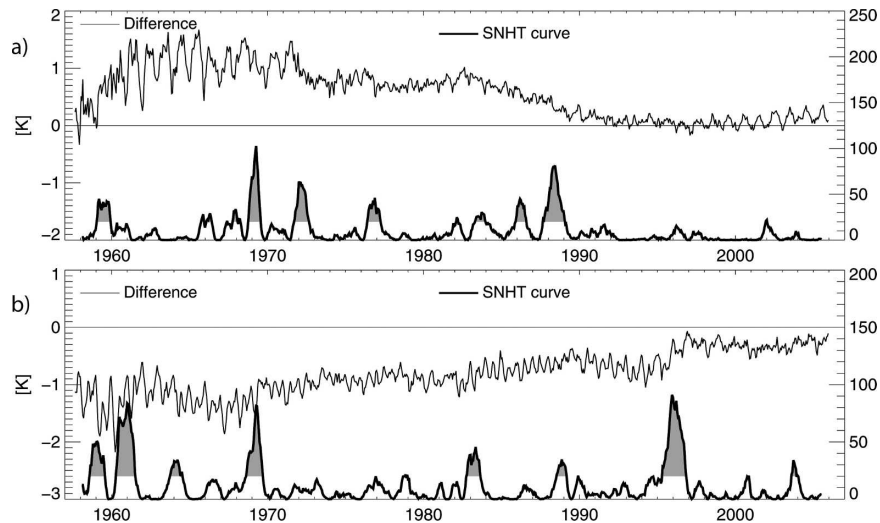


FIG. 2. Time series of composite mean 1200 – 0000 UTC differences from radiosondes (a) between  $30^{\circ}\text{W}$  and  $40^{\circ}\text{E}$  and (b) between  $120^{\circ}\text{E}$  and  $120^{\circ}\text{W}$ . Only stations and days with both 0000 and 1200 UTC data have been included in composite. Thin curve is 1200 – 0000 UTC difference; thick curve is time series of SNHT statistic as defined in Eq. (4). Peaks above 20 in SNHT test statistic are shaded. They indicate abrupt changes in the difference series. In (a) the peaks in 1969 and 1972 are caused by changes in the French/Russian radiosonde network. The peaks in 1988–89 are related to the change to Vaisala RS80 radiosondes at many sites. The main peaks in (b) are caused by changes in the Russian and Japanese radiosonde networks (1969) and the degradation of the Russian radiosonde network (circa 1996), which had relatively large day–night differences.

chosen as a compromise between data availability and susceptibility to radiation errors. Between  $30^{\circ}\text{W}$  and  $40^{\circ}\text{E}$  there is darkness at 0000 UTC and high solar elevation at 1200 UTC (except at polar regions). Between  $120^{\circ}\text{E}$  and  $120^{\circ}\text{W}$  the opposite is the case. A larger longitude range has been chosen in Fig. 2b to include more stations around the Pacific. In recent years the mean differences, which exceed 1 K in the 1960s, are gradually reduced at this altitude to practically zero. The resulting “trend” in the 1200 – 0000 UTC differences is almost certainly artificial (Sherwood et al. 2005) and must be removed before the radiosonde data can be applied for climate change research. Breaks and biases also cause problems in operational data assimilation systems as well as in climate data assimilation systems; observations with large bias but otherwise good quality tend to be rejected more frequently by a data assimilation system. If they are not rejected they cause biases in the resulting analyses unless the data assimilation system is specifically designed to deal with biased observations (Dee and Da Silva 1998).

The task of removing artificial breaks from time series is referred to as *homogenization*. Some authors attempted to use the available metadata (Gaffen 1996) and detailed knowledge of the instruments to apply

physically based corrections (Luers and Eskridge 1995; Eskridge et al. 2003; Redder et al. 2004). While this is potentially the most favorable approach, it can be applied only at selected sites since the required detailed information about equipment and launch times is often not available.

Homogeneity adjustments based on time series analysis, metadata, and expert judgment have been published by Lanzante et al. (2003a), Lanzante et al. (2003b), Free et al. (2005) (referred to as the LKS dataset), and Thorne et al. (2005a). While these datasets consider only a subset of the global radiosonde network (87 stations in the LKS dataset, 678 stations by Thorne et al. 2005a) they are important achievements and valuable tools for intercomparisons with satellite data and with climate model results.

Subsampling of the data, as performed in these analyses, is acceptable if the desire is to characterize only large-scale changes. For climate data assimilation purposes it is much harder to justify omission of at least one-third of the available radiosonde data in order to retain temporal homogeneity of the analyzed product. Since only anomaly time series have been adjusted, many records of these homogenized datasets may still have a constant bias and therefore may cause a bias in reanalyses if used as input for data assimilation. Both

aspects limit their value as direct input for future reanalyses, although the information on breakpoint timing and magnitude that they contain may prove useful.

None of the above datasets has been created by automatic procedures. Any improvement of the datasets would be rather laborious. Further there is often lack of agreement where and how large the breaks are (see Free et al. 2002). Sizeable uncertainty therefore still remains on upper-air trends and low-frequency variability (Seidel et al. 2004; Free and Seidel 2005; Thorne et al. 2005b). Sherwood et al. (2005), Randel and Wu (2006), and Santer et al. (2005) have recently raised serious doubts about the validity of temperature trends from radiosondes in the Tropics. Reduced uncertainty in observed upper-air trends has been identified as one of the most pressing needs in climate research (Karl et al. 2006).

This article describes a new homogeneity adjustment method called Radiosonde Observation Correction Using Reanalyses (RAOBCORE). It addresses some of the problems of the existing homogenized datasets. A preliminary version of RAOBCORE has been documented in Haimberger (2005). The most important characteristics of RAOBCORE are the following:

- It uses time series of innovation statistics (the difference between observation and the background forecast of the assimilating model) of a climate data assimilation system, such as the ERA-40 reanalysis (Uppala et al. 2005). The rationale for using ERA-40 innovation statistics for homogenization is discussed in section 2.
- It uses the most recent and most complete radiosonde datasets available: The Integrated Global Radiosonde Archive (IGRA; Durre et al. 2006) and ERA-40 (see section 3).
- It uses time series of individual launches, not monthly or seasonal means, for break detection and adjustment, and it analyzes daytime and nighttime launches separately. The homogeneity adjustment method is described in section 4.
- A record with no breaks may still have a large constant bias, which is problematic for reanalysis efforts. In section 7 it is outlined how RAOBCORE adjusts not only breaks of long records but also the biases of the most recent parts of the records.
- It is relatively easy to create multiple realizations of the homogenized dataset, for example, for sensitivity experiments, because the adjustment software is fully automatic.

RAOBCORE uses innovations from a climate data assimilation system as reference instead of composites of neighboring radiosondes (cf. Thorne et al. 2005a)

and makes no reference to the LKS dataset. Consequently it can be regarded as independent of these homogenization methods. The homogeneity properties of the ERA-40 background temperature time series are discussed in section 6 and in appendix B. Adjustment results for selected stations and for several sensitivity experiments are presented in sections 8 and 9.

## 2. Interpretation of the innovation statistics of a data assimilation system

The atmospheric data assimilation process may be described as applying a filter on the multivariate time series of meteorological observations. The filtering process is optimal if the time series of differences between observations (*obs*) and the background forecast (*bg*) of the assimilating model is stationary random with zero mean (Lewis et al. 2006). The *obs* – *bg* differences are often referred to as *innovations* and the time series of the differences is called *innovation process*. In practice the filtering process is never optimal, due to systematic errors of both the observing system and the assimilating model. As a result the time series of innovations have nonzero long-term means, are not stationary, and not random. These nonzero means of the innovations are often referred to as *bias*. One can only estimate the bias of one (test) dataset with respect to a second (reference) dataset. Since the mean of the true state is not known, the real bias is also unknown. However, if there is a shift in the observation time series, for example, due to an instrument change, it should immediately result also in a shift in the innovation time series and thus of the bias. This shift can be estimated and used for homogenization purposes.

For this study mainly the innovations from the three-dimensional variational data assimilation system (3DVAR) used in ERA-40 have been used. Within the 6-h cycle the following cost function is minimized (Courtier et al. 1998):

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}). \quad (1)$$

Here  $\mathbf{x}$  is the state vector of the assimilating forecast model, and  $\mathbf{x}_b$  is the background state, which is obtained by a 6-h forecast run of the assimilating model. Then the forecast state mapped to observation space by the observation operator  $\mathbf{H}$  is compared with the available observations (the observation vector  $\mathbf{y}$ ), and  $\mathbf{B}$  and  $\mathbf{R}$  are the background error and observation error covariance matrices, respectively.

At the beginning of the minimization process,  $\mathbf{x} = \mathbf{x}_b$ ;

thus the first term vanishes. The quantity  $(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$  is called the *innovation* or *obs - bg* difference. Its accurate calculation is essential in the data assimilation process and much effort is put into the specification of the observation operator  $\mathbf{H}$ . The difference  $(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$  is available for every single observation presented to the data assimilation system and is saved in the so-called ERA-40 analysis feedback files. In the case of radiosonde temperatures,  $\mathbf{H}$  represents the transformation from the spectral space to the model grid and then simple linear interpolation from the ECMWF model grid to the observation location (ECMWF 2000).

The innovations have proven most useful for the detection and estimation of systematic observation errors (Hollingsworth et al. 1986). Monitoring of the innovations is an important task at operational forecast centers. The innovations have been used also in ERA-40 to calculate adjustments of the radiation error of radiosondes (Onogi 2000; Andrae et al. 2004). The innovations are more useful for bias estimation than differences between observations and the analyses, since the *obs - bg* differences are more directly related to systematic observation errors. If the *bg* were perfect, the *obs - bg* difference would be the *obs* error, whereas the *obs - analyses (an)* difference is already influenced in a complicated way by the erroneous observation (see, e.g., Dee 2004).

It is the working hypothesis of this paper that time series of innovations can be used to detect and remove artificial shifts in radiosonde temperature time series, that is, to *homogenize* these time series. Since the *bg* error is generally quite small, the differences between the *bg* and the corresponding observations tend to be small and therefore the series of differences between *bg* and *obs* reacts sensitively to any change in the radiosonde temperature bias.

Small random errors are only one of the necessary ingredients for successful homogenization. The *bg* forecast time series must also be temporally homogeneous and independent of the radiosonde records to be tested. These aspects are discussed in more detail in sections 6 and 8.

### 3. Input data

With the completion of the ERA-40 project (Uppala et al. 2005), a 45-yr global time series of analyses (*an*) and 6-h *bg* forecasts has become available. While the analyses are the most useful products for many research applications, the observation database assembled during the reanalysis effort is equally valuable.

The binary unified form for the representation of meteorological data (BUFR)-coded ERA-40 analysis

feedback (AF) dataset has been the main input data source for this study. The AF dataset contains all observations from 1958 to 2002 presented to the ERA-40 data assimilation system, plus quality control flags and the innovations described in section 2. No other dataset holds such long and complete time series of upper-air temperature innovations. The time series have daily resolution, and 0000 and 1200 UTC ascents were kept separate. Observations at 16 standard pressure levels (10, 20, 30, 50, 70, 100, 150, 200, 250, 300, 400, 500, 700, 850, 925, and 1000 hPa) were considered and analyzed for breakpoints.

Since the ERA-40 dataset ends in 2002, it has been decided to use operational analysis feedback data from 2001 onward. Concatenating operational and ERA-40 AF has allowed the use of radiosonde records up to December 2005. The *obs - bg* differences need to be adjusted since the *bg* of the operational version of the ECMWF data assimilation system used for 2001 and later years differs substantially from the *bg* of the ERA-40 assimilation system (see section 6). After 2001 no evidence for major temperature inhomogeneities in the *bg* temperatures due to changes in the operational ECMWF data assimilation system could be found.

The recent advent of the IGRA dataset (Durre et al. 2006) has been the second big improvement of the data basis for homogenization of the global radiosonde network. The IGRA and ERA-40/ECMWF datasets are not identical, as can be seen from Fig. 3. ERA-40 contains more data in the 1960s and 1970s, whereas IGRA contains more data in the 1990s.

The IGRA data complement ERA-40 radiosonde data at some places, especially over the United States in the 1990s. The data missing in IGRA are partly available in the older Comprehensive Aerological Reference Data Set (CARDS; Durre et al. 2006). They have not been included in IGRA because of data quality concerns. Experience from ERA-40 indicates, however, that most of these data are of sufficient quality to be assimilated. While the original IGRA data do not contain *obs - bg* differences, these can be calculated quite accurately from archived ERA-40 *bg* fields interpolated to the IGRA observation sites (Haimberger 2005).

In this paper, the union of both datasets is used. The merging procedure has been simple: if there were duplicates, preference has been given to ERA-40 data. A total of 2881 stations (most of them on land but also a few weather ships) have been identified. These have time series containing temperature on standard pressure levels and on significant levels (TEMP-Land and TEMP-Ship). The station identification procedure, which normally makes merging so complicated, has

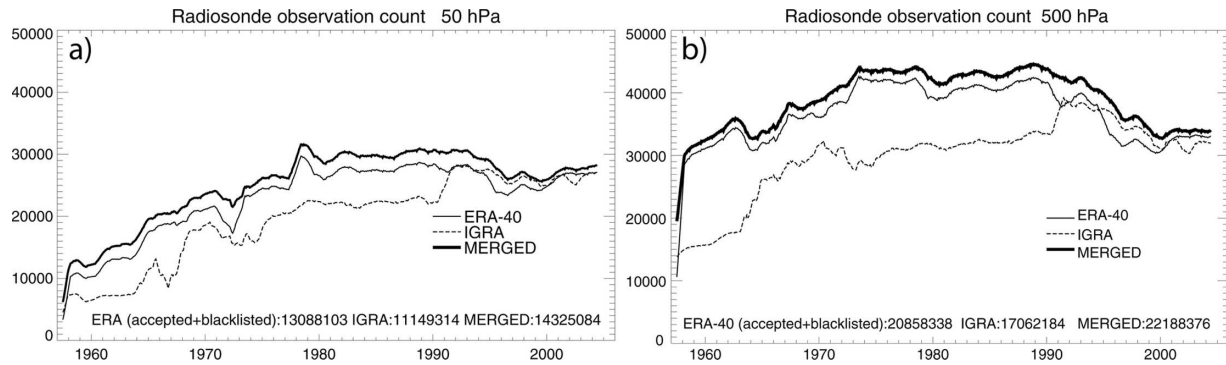


FIG. 3. Time series of monthly global radiosonde temperature observation count 1958–2004 at the (a) 10- and (b) 500-hPa level. Only TEMP-Land observations have been included. TEMP-Ship observations, which are also part of the ERA-40 archive, have been excluded in this plot but are used in RAOBCORE. Thin dashed: count from IGRA dataset. Thin solid: count in ERA-40 archive. Thick: merged ERA-40 + IGRA dataset. Numbers at the bottom are total observation counts over 47 yr.

been facilitated by the good documentation of IGRA data and the station tables available from ERA-40.

Apart from pure measurement information and the innovation statistics, metadata information is also available in digitized form from the CARDS archive (Gaffen 1996; with updates from Aguilar 2000). The information about the radiosonde type and the radiation corrections used is useful for finding and interpreting breaks in the radiosonde temperature time series, although the information is incomplete and sometimes inaccurate.

Quite useful additional metadata information could be extracted from the ERA-40 feedback records. From 1979 onward, many stations have routinely transmitted the radiosonde type. This information is available in the ERA-40 feedback files as well [coded according to World Meteorological Organization (WMO)-BUFR table 02011; see WMO Manual on Codes, No. 306, Vol. I.2]. From this information more than 1000 well-documented and often precisely dated radiosonde type changes could be extracted.

#### 4. Break detection method

The obs – bg difference series and obs(1200 UTC) – obs(0000 UTC) difference series have been analyzed with a variant of the standard normal homogeneity test (SNHT; Alexandersson and Moberg 1997) described below. Ducre-Robetaille et al. (2003) recently compared popular homogeneity tests and the SNHT performed well in this intercomparison. However, the results of this comparison are not directly applicable for the analysis of radiosonde records. The frequency of breaks requires analysis windows of only a few (<5) years. For such short time windows it is essential to take the seasonal cycle into account and it is advisable to use daily ascents instead of monthly or annual means in

order to have full control over the sampling of the available data. It is worth noting that using anomalies does not remove the annual cycle in the case of radiosonde data since the annual cycles of, for example, 1200 – 0000 UTC differences at the same stations are rather different for different radiosonde types (see, e.g., Fig. 10).

##### a. A variant of the standard normal homogeneity test

The original SNHT (Alexandersson and Moberg 1997) calculates the series of differences between the tested series (in this study the radiosonde time series) and a reference series (the bg). Then the means of the parts of the difference series before and after a potential breakpoint  $k$  are compared. The point dividing the sample into two parts is varied but the time interval stays fixed. For each dividing point  $k$  in the sample, a test statistic can be calculated:

$$T_k = [(N - k)(\mu_{1k} - \mu)^2 + k(\mu_{2k} - \mu)^2]/\sigma, \quad (2)$$

where  $N$  is the sample size,  $\mu_{1k}$  is the mean of the subsample before  $k$ ,  $\mu_{2k}$  is the mean of the subsample after  $k$ ,  $\mu$  is the mean of the whole sample, and  $\sigma$  is the sample standard deviation. The difference series is considered inhomogeneous if the maximum value  $T^s$  of all  $k$  is above the threshold somewhere in the interval. In this case the point  $k^s$  where  $T^s$  occurs is the most likely location for the breakpoint. The difference of  $\mu_{2k} - \mu_{1k}$  calculated at the point  $k^s$  is the best estimate for the magnitude of the break.

The original SNHT has two drawbacks. First, it tends to indicate breaks near the edges of the time interval as either  $k$  or  $N - k$  becomes small (Ducre-Robetaille et al. 2003). Second, the position of a breakpoint is poorly estimated in the presence of a periodic signal. To overcome these problems the original SNHT has been

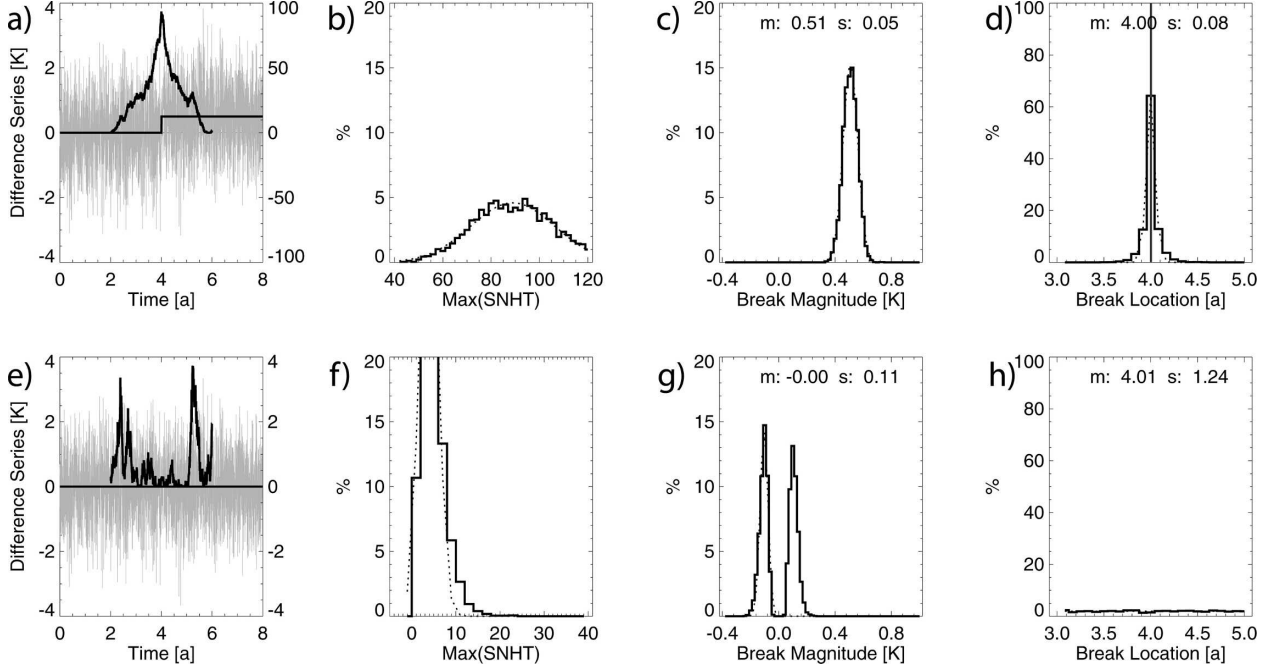


FIG. 4. Moving average SNHT applied to random series with shifts of (a)–(d) 0.5 K, (e)–(h) 0 K, and a residual variance of  $1 \text{ K}^2$ . (a), (e) One realization of a random time series with break and the corresponding SNHT test statistic  $T_k$  (right axis). A time period of 8 yr is simulated with a break right in the middle. The SNHT uses 2-yr-moving averages. (b), (f) Histograms of the  $T_k^s$ , generated with 5000 different random series. Note different x scales. (c), (g) Distribution of diagnosed size  $\Delta T$  of break. (d), (h) Distribution of diagnosed break location  $k^s$ . The smooth curves are fitted Gauss functions. Inset numbers indicate sample means (m) and standard deviations (s).

modified such that  $\mu_{1k}$  and  $\mu_{2k}$  are calculated as 2-yr moving averages:

$$T_k = \left[ \frac{N}{2} (\mu_{1k} - \mu)^2 + \frac{N}{2} (\mu_{2k} - \mu)^2 \right] / \sigma_k. \quad (3)$$

The sample sizes at point  $k$  are then fixed to  $N/2$  but the interval  $[k - N/2, k + N/2]$  now depends on  $k$ . The maximum value of the test statistic changes its meaning in the sense that it is not an absolute maximum of the  $T_k$  in a fixed interval but the local maximum of the  $T_k$  in the interval  $[k - N/2, k + N/2]$ . Therefore it is denoted  $T_k^s$ .

This way the means are more reliably estimated and the annual wave is averaged out exactly if  $N/2$  corresponds to an integer multiple of 365 if daily data are used and none are missing. To make the test more robust against combined effects of an annual cycle and missing data, data before and after the midpoint  $k$  of the analyzed time interval have been placed into 12 bins, one for each month. If the data count of, for example, the January values in the earlier 2-yr interval, was less than the count of January values in the later 2-yr interval, the excessive January values in the later interval were removed (starting with the values with the

largest time distance from the middle of the investigated interval). This version of the SNHT, which ensures equal sampling of the annual cycle in the intervals before and after a potential breakpoint, is referred to as *equal sampling SNHT*.

Figure 4 shows examples of the behavior of the moving average SNHT test statistic and related quantities with simulated data. The synthetic time series has 2920 data points, corresponding to 8 yr if daily data are considered. The interval for the moving averages before and after a suspected breakpoint is 2 yr, the same as used for the real data below.

The grey curve in Fig. 4a depicts one of 5000 realization of a  $N(0, 1)$  normal distributed random series with a break of size  $0.5\sigma$  in the middle of the interval. The black  $T_k^s$  curve in Fig. 4a shows a distinct maximum at the right location. Figure 4b shows the distribution of  $T_k^s$  gained from the 5000 realizations. Values are practically always above 50. Figure 4c shows the distribution of the break estimates  $\mu_{1k^s} - \mu_{2k^s}$  at the locations  $k_s$  where  $T_k$  is largest. It is normally distributed around 0.5 K with a standard deviation of 0.05 K. Figure 4d shows the distribution of the detected break locations  $k^s$ . The location of a break with size 0.5 K can be detected with a standard deviation of 0.08 yr (1 month).

The significance levels for rejecting the null hypothesis can be gained from applying the SNHT to a sample of 5000 random series with no break. Figure 4e shows a realization of such a series. The  $T_k$  vary erratically and  $T_k^s$  is much smaller. The significance levels of the moving average SNHT can be estimated from the distribution of the  $T_k^s$  in Fig. 4f. For the 95% level it is 9.6, and for the 99% level it is 12.5. Figure 4g shows again the distribution of the difference between the moving averages at the point  $k^s$ . Figure 4h shows the distribution of the break location  $k^s$  for the case of no breaks. It is close to uniform.

From this result, it becomes clear that breaks with a size of  $0.5\sigma$  are practically always detected since in this case more than 99% of the  $T_k^s$  reach values above 20. Figure 4g indicates the problems to be expected if a break is falsely detected: the estimates break size and therefore the false correction would be of magnitude  $0.25\sigma$ .

The significance levels derived above are valid only for purely random processes (apart from an annual cycle) and for complete time series (1460 days). If data are missing, the  $T_k^s$  value necessary to reach the 95% significance level decreases, no matter whether the data loss occurs as one big gap or intermittently. This can be verified with Monte Carlo-type experiments similar to those presented above (not shown) and has been documented by Alexandersson and Moberg 1997 as well. If there is an annual cycle in the obs – bg, the significance threshold also decreases, since it generates additional variance. In these cases smaller thresholds for the maximum SNHT should be used.

Only if the time series are generated by stochastic (autoregressive) processes, the significance levels become higher since the degrees of freedom are reduced (von Storch and Zwiers 1999). A check of the obs – bg difference time series (with annual cycle removed) showed weak autocorrelation for lags of up to 4 days at some remote areas and practically zero autocorrelation for areas with good data coverage and for obs(1200 UTC) – obs(0000 UTC) time series. Data show generally much less autocorrelation than a first-order autoregressive model with coefficient  $\alpha = 0.3$ , and even if such a model would apply, the 95% threshold is still below 20. An exception are remote island stations before the satellite era where the autocorrelation is similar to a first-order autoregressive model with  $\alpha = 0.5$ , for which a Monte Carlo experiment similar to those in Fig. 4 yields a 95% significance threshold of 25. Additional sensitivity experiments with the equal sampling SNHT are documented in Haimberger (2005). It performs well in general as long as there is only one break within the

tested interval. If there are more breaks only the largest break is detected.

To be safe, thresholds larger than  $T_k^s > 12.5$  suggested from the above analysis have been used for break detection. For obs – bg time series  $T_k^s$  values above 50 are considered significant, if no metadata are available. For obs(1200 UTC) – obs(0000 UTC) time series, which show no autocorrelation and cannot be influenced by bg errors, values above 20 are considered significant if there are no metadata available. An SNHT value above 50 is almost always reached by breaks with size  $0.5\sigma$  (see Fig. 4b). SNHT values above 20 are reached by about half of the series with breaks of size  $0.25\sigma$  but are still not attained by homogeneous time series.

The equal sampling SNHT is now applied to the following time series:

- Log pressure-weighted 1200 – 0000 UTC temperature difference in the stratosphere: depending on data availability analysis starts with the 20–30-hPa layer. Then all the means from two pressure levels down to the 200–300-hPa layer are calculated. Each layer mean time series is analyzed with the equal sampling SNHT. For each point in time the maximum SNHT value from the analyzed pressure layers is kept. The resulting time series of maximum SNHT values is then examined for significant values. Note that the 1200 – 0000 UTC time series is independent of the ERA-40 bg. A break detected in this time series has high credibility since it can be safely attributed to the obs series.
- Log pressure-weighted layer mean obs – bg temperature difference at 0000 and 1200 UTC in the stratosphere. These time series, if available, are analyzed in the same manner as the 1200 – 0000 UTC temperatures.
- The 300–850-hPa log pressure-weighted layer mean obs – bg temperature difference at 0000 and 1200 UTC. These time series are sensitive to station relocations and are more complete than the stratospheric time series. Temperature means from this relatively thick layer have quite small variance, and therefore relatively subtle breaks having the same sign throughout this layer can be detected.

#### b. Decision algorithm for break detection

The probability for a break depends not only on the SNHT test statistics but also on the occurrence of events, such as instrument changes. These information sources may be combined with the Bayesian rule (De-Groot 1986). Let  $A_1$  be the event that a break with a given size occurs within a time interval of  $\pm 2$  yr around



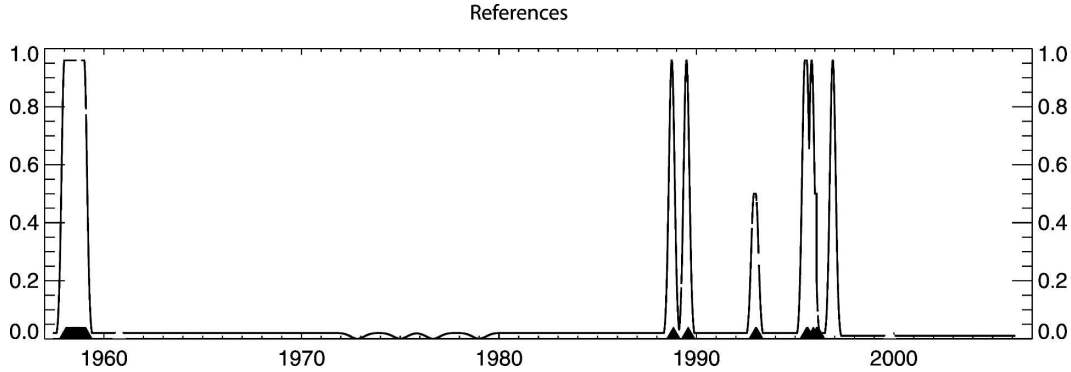


FIG. 5. Prior “probabilities” for station Bethel (70219, Alaska, 60.78°N, 161.80°W) derived from CARDS/BUFR metadata. Triangles/trapezoids denote metadata events from CARDS. Trapezoid in 1958 indicates that only the year of the event is known from CARDS. A constant high prior break probability is assigned throughout this year. Peaks without triangles/trapezoids are triggered by radiosonde-type changes derived from ERA-40 feedback metadata. Gaussians are used to crudely represent metadata uncertainty. Prior probabilities in 1973, 1975, 1976, and 1979 are lowered to reduce the chance of false detections due to major changes in the satellite observing system.

a point in time and  $A_0$  be the event that no break occurs. Let  $B$  be the event that  $T_k^s$  reaches a value above a certain threshold  $x$ . We now want to know the probability of a break given that the event  $B$  occurs. Bayes’ theorem states that this probability can be calculated as

$$\Pr(A_1/B) = \frac{\Pr(A_1)\Pr(B/A_1)}{\sum_{j=0}^1 \Pr(A_j)\Pr(B/A_j)}. \quad (4)$$

The challenge is now to specify the probabilities on the rhs of this equation. The probabilities  $\Pr(B/A_1)$  and  $\Pr(B/A_0)$  can be found from the histograms shown in the second column of Fig. 4. It is the area under the histograms between  $x$  and  $\infty$ . Metadata information can be included by specifying prior probabilities  $\Pr(A_1)$ . If, for example, the date of a radiosonde change is known, one can apply a higher prior probability  $\Pr(A_1)$  to this particular date. In the examples presented in this paper, prior probabilities of 0.96, 0.5, and 0.02 have been chosen for radiosonde changes, radiation correction changes, and no documented changes, respectively. This choice may appear extreme given the limited reliability of metadata but is necessary to make the detection system sensitive enough for metadata. With this choice, SNHT values of 31, 43, and 50 (14, 18, 20 for 1200 – 0000 UTC difference series) are necessary to reach scores above 0.5 when put into Bayes’ formula. A documented break with size  $0.3\sigma$  reaches similar scores as an undocumented break with size  $0.5\sigma$ . A score above 0.5 is necessary to trigger the breakpoint adjustment algorithm. This happens with breaks of size  $0.5\sigma$  without metadata.

Since the metadata information may be imprecise,

the prior probabilities have been modeled as Gaussians with standard deviation of 60 days. If only the year of a change has been reported the same high prior probability is specified throughout the year in order to assure that high weight is given to the metadata and to let the SNHT detect the break date. If Global Telecommunications System (GTS) metadata have been available, the prior probability has been lowered to 0.01 between events since the sequence of radiosonde-type reports explicitly states that there was no radiosonde-type change.

Figure 5 shows the breakpoint analysis for the radiosonde station Bethel (70219, Alaska). Prior probabilities are assigned according to the CARDS/BUFR metadata. Figures. 6a,b show time series of the test statistic  $T_k$  of the SNHT at Bethel for the stratospheric obs(1200 UTC) – obs(0000 UTC) and for the tropospheric obs – bg time series at 0000 UTC. The test statistics exceed the significance levels (20 for Fig. 6a, 50 for Fig. 6b) several times at this station (see also Fig. 10 for the corresponding difference time series). One can also see that the peaks are very sharp for the large breaks (1989, 1995), so that the break location is well determined in these cases, consistent with Fig. 4.

Since there are five time series of probability scores that may contain conflicting information (e.g., the location of the diagnosed breakpoint may differ slightly between the time series) one has to choose the location. The obs(1200 UTC) – obs(0000 UTC) score, shown in Fig. 6c is given the highest priority, since it is independent of possible inhomogeneities in the bg and therefore always attributable to the radiosondes. This is consistent with previous findings (Lanzante et al. 2003a;

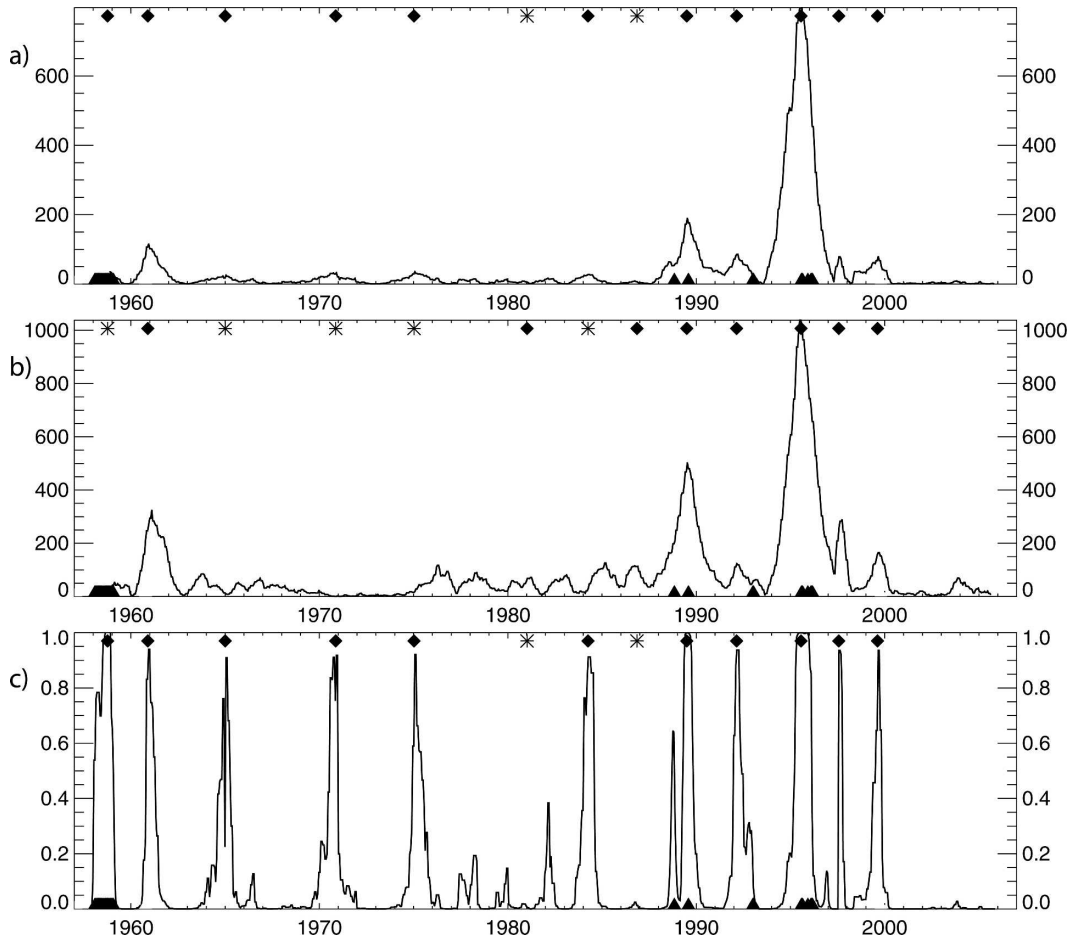


FIG. 6. SNHT results and posterior probabilities for Bethel. (a), (b) Equal sampling SNHT test parameter for stratospheric obs(1200 UTC) – obs(0000 UTC) and for tropospheric obs – bg (0000 UTC) time series. Values above 20 (50) are considered significant. Squares denote breakpoints where SNHT test value is above significance threshold. Stars denote points where SNHT or break probability is below threshold in the respective series but is above the threshold in any other of the five analyzed time series. All breaks are detectable in obs(1200 UTC) – obs(0000 UTC) series (see Fig. 10) except in 1981 and 1987, which are detected in 0000 UTC obs – bg difference series. (c) “Probability” score for a break in the obs(1200 UTC) – obs(0000 UTC) time series, given the prior “probability” from (a) and the test statistic (b).

Sherwood et al. 2005) that most changes seem to be accompanied with changes in the obs(1200 UTC) – obs(0000 UTC). The peaks whose score is above 0.5 and that are the highest peaks within a  $\pm 2$  yr interval are selected. Then the two probability time series from the stratospheric obs – bg time series are analyzed (the 0000 UTC series is shown in Fig. 6b). The locations with the highest peaks are chosen as breakpoints unless a breakpoint has already been detected within a  $\pm 2$  yr interval in the obs(1200 UTC) – obs(0000 UTC) series. Finally, the probabilities from the tropospheric obs – bg time series are analyzed. Other priority choices for the breaks from obs – bg series are possible. One could give the tropospheric obs – bg time series priority to

stratospheric obs – bg series or one could combine the probability scores again with the Bayesian rule. Little overall sensitivity has been found in respect of this choice.

The diamonds and stars in Fig. 6c indicate the location of the finally chosen breakpoints. The diamonds indicate breaks detected already in the obs(1200 UTC) – obs(0000 UTC) time series, the stars are breaks detected in at least one of the obs – bg time series.

There are many possible ways to further improve this break detection algorithm. Perhaps the most important path for improvement is to use different time intervals (not only  $\pm 2$  yr) for the SNHT since a running mean with fixed length may be incapable of detecting all breaks.

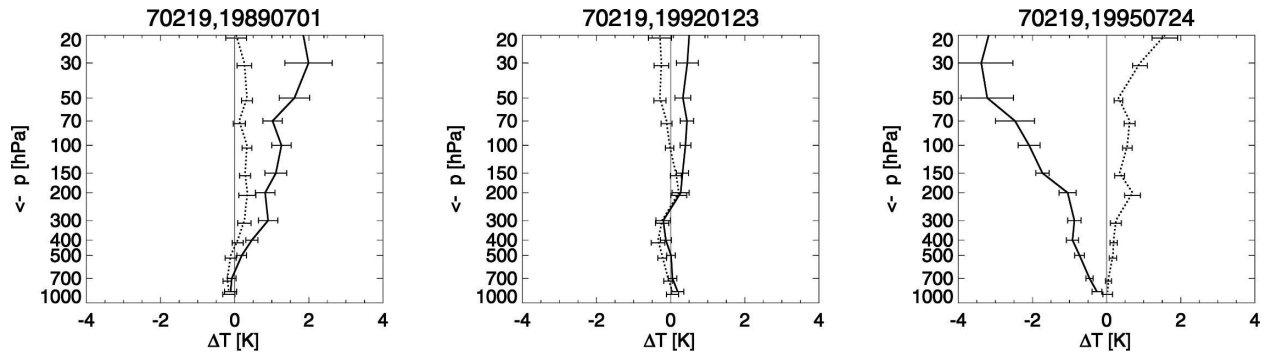


FIG. 7. Break profiles diagnosed for three breaks at station Bethel. Solid curves are break amplitudes at 0000 UTC, and dashed curves are amplitudes at 1200 UTC. Dates in headers indicate diagnosed date of breaks. Break profiles are applied to all data preceding the breakpoint. Error bars are 90% percentiles of obs – bg difference before/after the break.

### 5. Estimation of the break profiles

After the break detection, the mean obs – bg differences before and after the break to be adjusted are calculated at each pressure level, again ensuring that the annual cycle is sampled equally before and after the break. The difference of the means at every pressure level yields the estimated profile of the breaks (the solid and dotted curves in Fig. 7). While this profile inevitably contains small-scale vertical noise it is not smoothed vertically, as in sensitivity experiments the smoothing did not improve the spatiotemporal consistency of adjusted trends or 1200 – 0000 UTC differences.

The time interval used for estimating the break size is varied between 1 and 8 yr. The default of 8 yr is necessary for stable statistics in the presence of frequent data gaps. It is reduced only if there is another breakpoint before the breakpoint considered within 8 yr. The interval after the breakpoint is always 8 yr (if the time series is long enough), since the time series after the breakpoint has already been homogenized. The minimum interval of 1 yr is also the minimum interval between breaks allowed by the break detection algorithm.

After estimating the break at each level, its significance is tested using Student's *t* test. Only profiles where at least two pressure levels contain breaks that are significant at the 95% level are used for the adjustment.

Since the bg may be affected by changes in the satellite observing system, a second criterion to test the significance of a break profile has been developed. It compares obs – bg time series of the tested radiosonde, with a composite of obs – bg time series from radiosondes surrounding the tested site. This composite was originally intended only for adjustment of the climatology of a tested time series (see section 7) but is quite useful for checking the break estimates gained from the

obs – bg information at the tested site as well. Only if the break profile from difference series between obs – bg time series of the tested radiosonde and composite obs – bg time series of the surrounding radiosondes contains significant breaks at two or more pressure levels as well, is the break profile adjusted. The second criterion is more robust against jumps in the bg since the obs – bg series of the neighboring radiosondes are likely affected by very similar bg jumps, as these have a very large-scale structure. The neighbor composite criterion used here has similarities to the methods used by Thorne et al. (2005a). They use, however, composites of obs-anomalies, not composites of bg – obs differences for break detection and adjustments. Using the above criteria, only about 70% of the detected breaks are actually adjusted.

If a break profile has been considered significant, adjustments have been applied at all levels, even if the adjustment amounts have been below the significance threshold at some pressure levels. Adjustments at very high levels may be less accurate due to lack of data but the spatiotemporal consistency could be improved by the adjustments even at the 10-hPa level. No data deletions have been performed.

### 6. Global mean differences between ERA-40 bg and radiosonde observations

It is essential to be aware of any inhomogeneities of the ERA-40 bg since these reduce the applicability of the ERA-40 bg as a reference. Inhomogeneities in the bg time series may be introduced by changes in the ERA-40 observation coverage, in the observation biases correction, and in the overall observation quality. Apart from radiosondes the satellite data are mainly affected by changing biases. The satellite data contributed to the realistic representation of many strato-

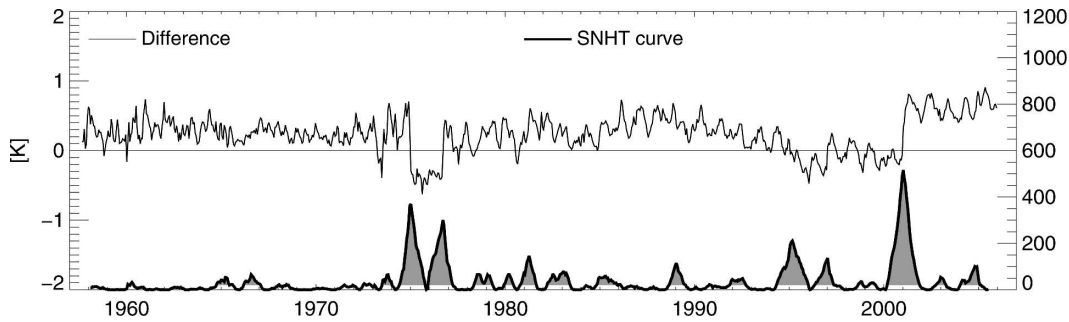


FIG. 8. Global mean 0000 UTC obs – bg difference averaged over 130 radiosonde stations of the Global Climate Observing System (GCOS) Upper Air Network (GUAN; Daan 2002) at the 50-hPa level. Shifts in 1975 and 1976 caused by erroneous *NOAA-4* bias correction. Shift in 1989 caused by change of assimilation streams in ERA; shift in 1995 and 1997 related to problems with *NOAA-11/14*; and shift in 2001 is caused by switch from ERA-40 bg to ECMWF operational bg.

spheric features in ERA-40 analyses (Randel et al. 2004) and to a much better overall quality of the ERA-40 analyses and forecasts. However, the quality control and the adjustment of the changing biases of satellite radiances as described by Hernandez et al. (2004), Li et al. (2006), and Harris and Kelly (2001) is challenging. Suboptimal radiance bias correction can easily compromise the homogeneity of the global mean ERA-40 bg forecasts (Uppala et al. 2006).

As can be seen from the global mean obs – bg difference series in Fig. 8, some changes in the ERA-40 satellite observing system did indeed lead to breaks in the global mean bg forecasts. The most prominent breaks evident in Fig. 8 occurred in January 1975, September 1976, and April 1986 and are related to problems with the National Oceanic and Atmospheric Administration (NOAA) satellites *NOAA-4* and *NOAA-9*. Jumps in 1995–97 coincide with the end of *NOAA-11*, and the start/end of *NOAA-14* (see also Christy and Norris 2006). At high altitudes the effects of insufficient bias correction of radiances from the stratospheric sounding unit (SSU), particularly in the early 1980s, are noticeable (see Haimberger 2005; Uppala et al. 2006). Trenberth and Smith (2006) have recently diagnosed a spurious break in ERA-40 temperature analyses related to the assimilation of MSU-3 radiances at the end of the *NOAA-9* period. These problems are the likely reason for the rather weak stratospheric cooling and rather strong upper-tropospheric heating trends of the ERA-40 bg compared to available radiosonde and satellite datasets (see Karl et al. 2006, their Fig. 3.4) in the ERA-40 analysis. This peculiar vertical pattern is also evident in the global mean trends of the unadjusted bg shown in Fig. 9. In the Tropics the upper-tropospheric heating is even more pronounced.

However, pervasive radiosonde temperature biases especially in the Tropics (Sherwood et al. 2005; Randel

and Wu 2006; Christy and Norris 2006) have most likely contributed to inhomogeneities in the global mean bg – obs series as well. From the 1990s onward, there is also good correspondence between trends derived from MSU radiances (Santer et al. 2004) and ERA-40, which supports the bg. For this period the ERA-40 analysis also looks more homogeneous in the study by Trenberth and Smith (2006). Therefore, with the present knowledge it is difficult to tell to what extent the radiosonde temperatures or the bg temperatures are responsible for the breaks and in particular for the trends in the global mean obs – bg departures of Fig. 8. Despite these uncertainties, the adjustment of the global mean bg has been tried such that the chance for spurious break detection due to breaks in the bg is reduced.

This is most straightforward for the large break in January 2001, which is certainly caused by the switch from ERA-40 to the operational ECMWF data assimilation system. It is a clear indication of the temperature uncertainties that still exist at stratospheric levels. The break can be adjusted relatively accurately at each radiosonde station from an overlap year (2001) between operational and ERA-40 innovations.

The remaining shifts in the global mean bg – obs series are much harder to attribute to either the bg or the obs. In the bg adjustment procedure applied here, it is assumed that much of the trend of the global mean obs – bg difference stems from time-varying biases in the global mean bg. Therefore the global mean obs – bg difference shown in Fig. 8 has been subtracted from the bg time series at individual stations. The bg adjustment has not been applied completely uniformly but has been varied with radiosonde observation density and with latitude (see appendix B for details). Other choices for the global mean bg adjustment are definitely possible. One could, for example, try to homog-

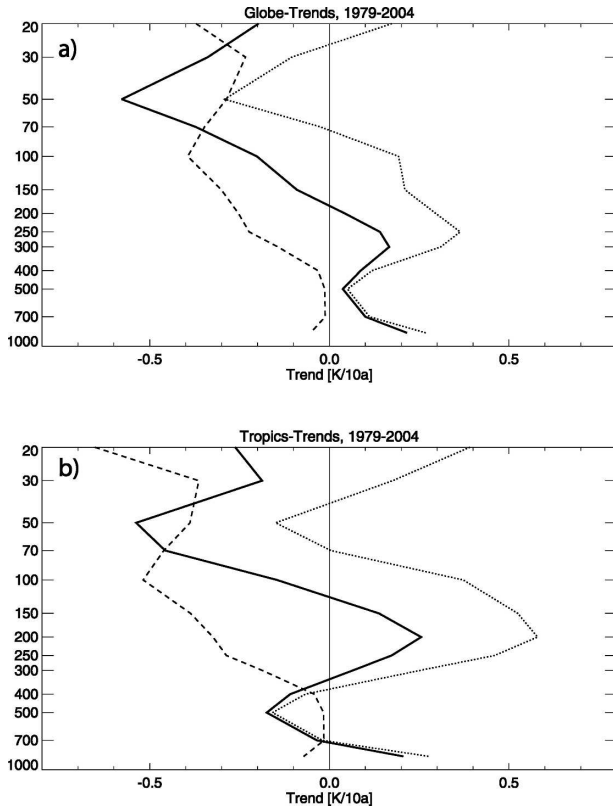


FIG. 9. Vertical profiles of global mean trends 1979–2004. Dotted line is trend of unmodified background forecast (bg) temperatures, solid line is modified bg trend, and dashed line is trend of the bg modification. (a) Global mean trends; (b) tropical mean ( $20^{\circ}\text{S}$ – $20^{\circ}\text{N}$ ) trends. The shift in 2001 due to the switch from ERA-40 to ECMWF operational bg temperatures has already been removed from the unmodified bg temperature trends. Trends have been averaged from station time series within  $10^{\circ} \times 10^{\circ}$  grid boxes and then averaged over the globe to reduce the effect of the nonuniform radiosonde station distribution.

enize the global mean bg in a similar manner as the individual radiosonde records. Further investigation of the properties of the bg is needed to design an optimal bg adjustment.

At individual stations this bg adjustment is subtle compared to the typical sizes of breaks in radiosonde records. In the global mean, however, the bg adjustment procedure leads to stronger cooling/weaker heating trends (Fig. 9), especially in the lower stratosphere/tropical upper troposphere. In the lower troposphere the adjustment has less impact since the bg – obs differences are smaller and show little trend below 300 hPa. It is interesting to note that both the unadjusted radiosondes and the bg are practically neutral [ $-0.01$  and  $0.02 \text{ K} (10 \text{ yr})^{-1}$ ] in terms of 300–850-hPa trends in the Tropics. The strong warming trends in the tropical lower troposphere as suggested by the Remote Sensing

Systems (RSS) MSU lower-troposphere (LT) products (Santer et al. 2005) are not supported either by the radiosondes or by the ERA-40 bg.

The necessity of this bg adjustment prior to the homogenization and the sensitivity of the homogenization results on the bg adjustment (see section 9) limit the accuracy of the homogenization results. It is expected that the size of adjustments necessary for the bg from future reanalyses will be much less due to more advanced treatment of satellite biases.

## 7. Adjustment of the climatology of radiosonde time series

Even after successful homogenization, the climatology of a radiosonde record may have a large bias if the most recent part of the time series is biased. The reasons are typically as follows:

- Some countries still operate radiosondes with non-negligible temperature biases. Examples are the Chinese and most Russian radiosondes.
- Some radiosonde time series end well before the year 2000. Prominent examples are the weather ships in the Atlantic and Pacific Oceans. The most recent parts of these time series are likely biased.
- Some time series, for example, over Russia in the 1990s, have gaps that are wider than the averaging intervals used for calculating the break profiles. In this case the time series before the gap cannot be adjusted by the homogenization procedure.

Adjusting the climatology and the choice of “trusted” radiosonde types whose measurements are left untouched, as done here, is a delicate task since it implies preference to certain radiosonde types. However, this task cannot be avoided for climate data assimilation purposes. The involved biases are large enough to cause noticeable biases of the resulting analyses or even to trigger rejection of a substantial part of the radiosonde measurements, particularly at high altitudes in the satellite era.

The adjustment of the climatology is performed after the time series homogenization; it is the final adjustment step. Therefore the time series should already be without any detectable breaks. The challenge is now to calculate a reference that is less biased than the most recent part of the tested time series.

Many ascents launched in recent years have relatively small biases up to the 10-hPa level or the biases are well documented and can be corrected. A subset of 342 radiosonde stations has been identified where the climatology is considered unbiased after homogenization. A map of these can be found in Haimberger

(2006). These are stations that have been equipped with Vaisala RS80/90/92 radiosondes or temperature sensors, with MeiseiIII radiosondes and Sippican radiosondes. Their performance is well documented in radiosonde intercomparison studies, such as Nash et al. (2005).

At all other stations (e.g., with VIZ radiosondes, Chinese and Russian radiosondes) the most recent part (usually the most recent 8 yr, if available) of the records has been regarded as biased and therefore the records have been adjusted.

A reference time series has been calculated from a neighbor composite of obs – bg differences from the above 342 radiosonde stations. The size of the adjustment is calculated by comparing the obs – bg difference of the tested series with the composite obs – bg difference of the neighboring time series. The mean obs – bg difference of the tested time series should be very close to the mean obs – bg difference of the composite, since the spatial pattern of an 8-yr-averaged bg temperature field is smooth and the bg temperature gradients are considered realistic. Since the obs – bg differences are such small increments, their composites are less affected by data gaps than composites of absolute temperatures.

The composite obs – bg difference is a weighted mean of neighboring homogenized radiosondes using weights decreasing exponentially with distance from the radiosonde station to be adjusted. A decorrelation distance of 3000 km has been used. At least 20 stations have been required to be available for the composite.

Experience with this approach has been encouraging. Many biases of the most recent parts of time series, as evident for example over Russia or at the weather ships in the Atlantic, could be significantly reduced (see Figs. 17 and 18).

## 8. Adjustment results for selected individual stations and regions

A small subsample of radiosonde records is investigated to demonstrate the strengths and the limitations of RAOBCORE. For most countries similar examples can be found since practically all long radiosonde time series contain inhomogeneities. Even if the radiosonde records presented contain large breaks, they have high information content.

### a. Adjustments in the satellite era

Since homogenization works backward in time we begin with the more recent period 1979–2005. While the ERA-40 bg contains more information independent

of radiosondes than in the presatellite era, there is also the risk of spurious breaks due to insufficient bias adjustment between satellites (see section 6). The following examples should demonstrate that adjustment with obs – bg difference series is possible despite these problems.

One is Bethel, Alaska, which had a relatively homogeneous time series between 1961 and 1989 but then three marked breaks in 1989, 1992, and 1995. Test statistics for this station have been shown already in Figs. 5 and 6. Due to space constraints the analysis in this and other examples is restricted to the 50-hPa level in this study.

Figure 10 shows time series of unadjusted and adjusted 1200 – 0000 UTC temperature differences at Bethel. One can see how different the day–night differences are depending on the radiosonde system in use (VIZ before 1989, then Space Data radiosondes, then Vaisala RS80 radiosondes, according to CARDS). These jumps can only come from changes of the observing system and are independent of the ERA-40 bg.

The ERA-40 bg comes into play for the adjustments. The adjustments for the 0000 and 1200 UTC ascents are calculated independently from obs – bg time series at 0000 and 1200 UTC. It is worth emphasizing that the 1200 – 0000 UTC obs time series is never used for adjustment purposes in RAOBCORE; it is used for break detection purposes only.

As one can see from Fig. 10b the diagnosed breakpoints coincide well with metadata events in the 1980s–90s. The breaks in 1976 and in the 1960s are smaller, but one can see from visual examination of the unadjusted 1200 – 0000 UTC time series that there are breaks indeed. The adjusted 1200 – 0000 UTC time series is much more homogeneous and has an almost neutral trend, as is expected for this difference series. The homogeneity of the 1200 – 0000 UTC time series is a good consistency check for RAOBCORE since this time series has not been used for calculating the adjustments.

Bethel is not the only radiosonde station in Alaska affected by instrument changes. As can be seen from Fig. 11, the unadjusted radiosonde records over Alaska and Canada yield rather different temperature trends for the period 1989–2004 at 0000 and 1200 UTC and the trends are also spatially rather heterogeneous due to the different station histories. After adjustment with RAOBCORE, the temperature trends at 0000 and 1200 UTC are in good agreement and also the spatial homogeneity is much better. Note that there were substantial adjustments not only for the daytime (0000 UTC) ascents. This is consistent with findings from radiosonde intercomparisons (Nash and Schmidlin 1987) that

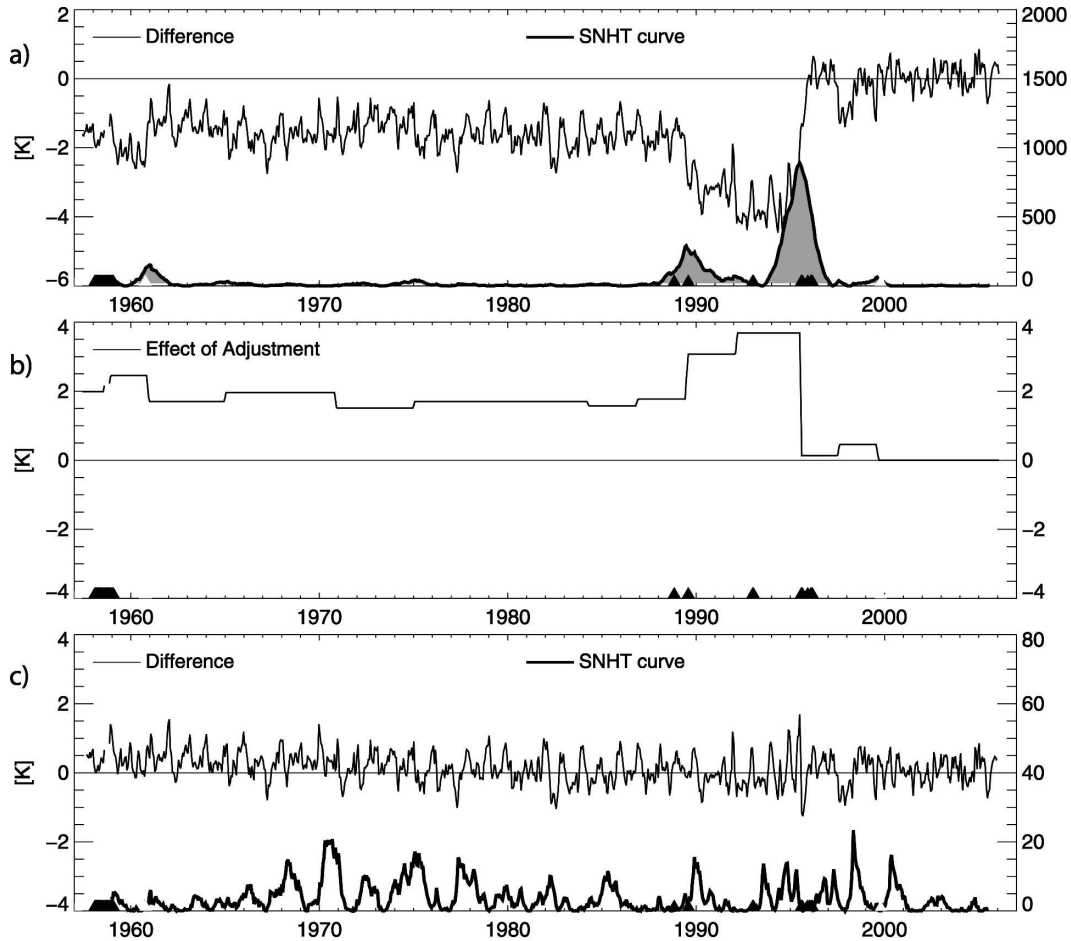


Figure 10:

FIG. 10. (a) Thin curve is unadjusted 1200 – 0000 UTC radiosonde temperature difference at 50 hPa for station Bethel (70219, Alaska). Thick curve is SNHT test statistic (right axis). Peaks in SNHT test statistic indicate abrupt changes in the mean difference. Triangles at the bottom indicate changes of radiosonde type and on-site radiation correction, as documented by Aguilar (2000). (b) Effect of adjustments applied by RAOBCORE on 1200 – 0000 UTC difference; (c) 1200 – 0000 UTC difference time series after adjustment. Note different scale of right axis compared to (a).

showed substantial nighttime deviations between VIZ and Vaisala RS80 radiosondes and contradicts the impression one may get from Sherwood et al. (2005), who stress the importance of daytime biases.

The second example is Darwin, Australia, which switched from Philips MKIII radiosondes to Vaisala RS80 radiosondes in May 1987. This station has been analyzed more thoroughly for example by Free et al. (2002). Darwin is difficult since only 0000 UTC ascents have been available before 1987 and even those had some gaps at the 50-hPa level. Figure 12 shows the adjustments suggested by RAOBCORE for this station. The break size is estimated 2 K, which is in accord with Free et al. (2002). There is a clear indication of a

break in 1963 as well, which is in good agreement with breaks suggested by Lanzante et al. (2003a) for this station. Additional breaks in Fig. 12 seem evident in 1976 and 1981, but these have not been adjusted since the significance criteria have not been met.

The robustness of the RAOBCORE adjustment procedure may be seen from Fig. 13. The spurious 1200 – 0000 UTC differences are reduced at almost all stations that launch two radiosondes per day during the period 1988–90. The largest adjustments can be found over the United States, China, and Australia. The spatial consistency of the differences (as measured by the “cost”  $C$  defined in appendix A) has improved as well compared to Fig. 1, and the amplitude of the differences has been

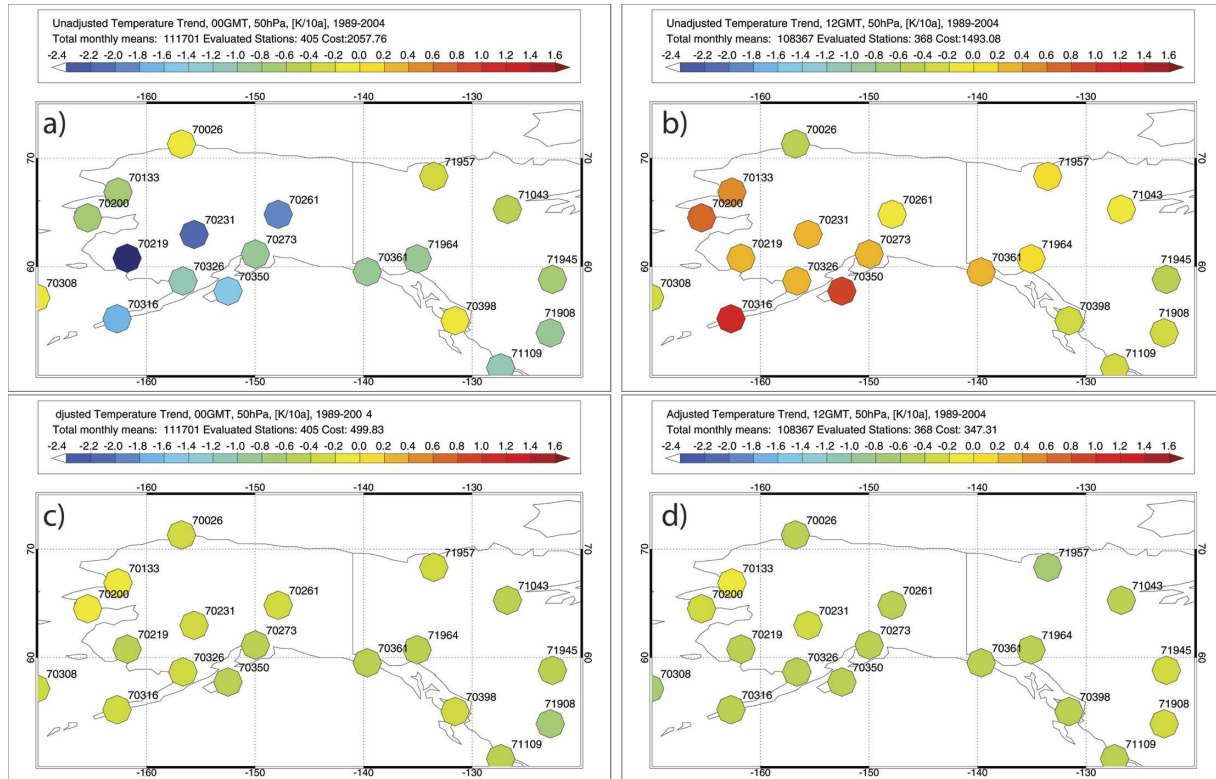


FIG. 11. Radiosonde temperature trends during 1989–2004 in  $\text{K decade}^{-1}$  over Alaska/northern Canada at 50 hPa. (a) Unadjusted obs(0000 UTC) time series, (b) from unadjusted obs(1200 UTC), (c) from adjusted obs(0000 UTC), and (d) from adjusted obs(1200 UTC). Numbers are WMO station IDs.

reduced to a few tenths of a kelvin except in the Tropics, in agreement with recent results from Free and Seidel (2005).

### b. Adjustments in the presatellite era

In the presatellite era, the ERA-40 bg does not contain much more upper-air information than is available from radiosondes. Therefore the issue of dependence of the bg on the radiosondes to be tested becomes particularly important. Further the radiosonde density in the Tropics and the Southern Hemisphere was very sparse, making neighbor intercomparisons rather difficult. Some examples are given indicating that adjustment of radiosondes using obs – bg time series is still possible.

The dependence of the bg on the radiosondes to be tested is expected to be particularly large for (i) remote stations and (ii) large countries using the same radiosonde equipment. The degree of dependence may be assessed qualitatively by looking, for example, at Mesural radiosondes in the late 1960s–early 1970s. Particularly the Mesural FM43 radiosondes had large radiation errors. This radiosonde type was in use in France and in former French dependencies (e.g., many

Pacific islands). This gives an opportunity to see whether the same breaks are diagnosed over data rich and extremely data sparse areas. Figure 14 shows obs – bg differences for local daytime ascents at the French radiosonde station Trappes, France (1200 UTC), and at the southeast Pacific station Rapa, Chile, before and after adjustment with RAOBCORE.

Both time series show very large biases around 1970 when Mesural-FMO-43B radiosondes were in use. These were replaced by Mesural-FMO-44C radiosondes in 1972 at Trappes and in 1976 at Rapa. The suggested corrections from obs – bg time series are about 20% weaker at Rapa, probably because the bg is influenced by Rapa itself. A correlogram from the stratospheric obs – bg time series at Rapa (not shown) also indicated some dependence of the bg on the observations at Rapa: the serial correlation of the obs – bg differences decreases from values of 0.3 for 1-day lags to insignificant levels for lags of 4 days. Nevertheless it is evident that the bg contains enough information to yield sensible break estimates even in this area. The obs – bg time series at Rapa after adjustment of the observations (Fig. 14d) looks much more homogeneous.



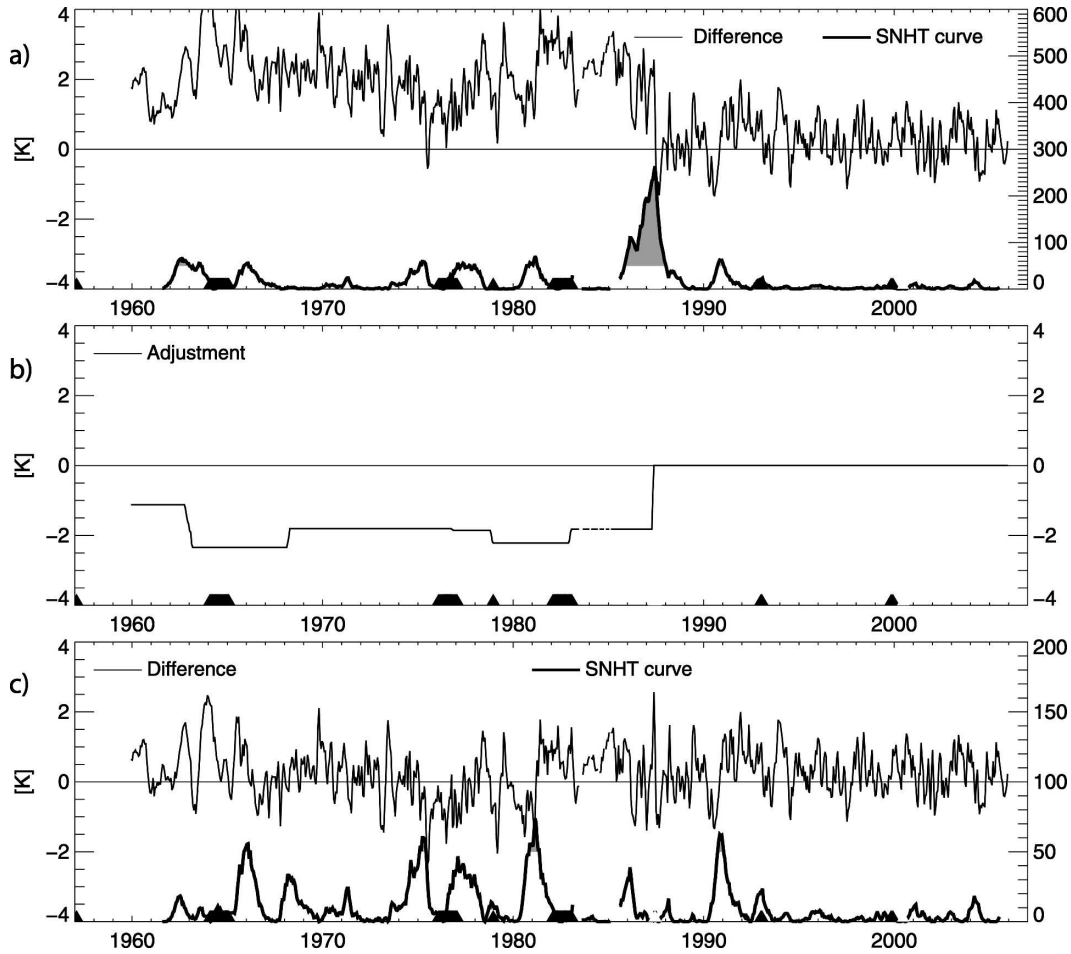


FIG. 12. The obs – bg difference at 50 hPa, 0000 UTC for the radiosonde station Darwin (12.43°S, 130.87°E). (a) Unadjusted time series, (b) adjustment applied, and (c) adjusted time series.

At Trappes the standard deviation of the obs – bg difference is much smaller since the bg forecast is more accurate in this region and also the serial correlation of the innovation time series is practically zero (except near a breakpoint). The adjusted obs – bg time series (Fig. 14b) is again more homogeneous although not all instationarities are removed from the strongly varying time series between 1969 and 1972.

Not only Trappes but all French radiosonde sites are affected by the Mesural problem with daytime stratospheric temperature biases in the order of 10 K. The bg over Europe seems almost unaffected by this problem, although most of the biased Mesural observations have been accepted by the ERA-40 quality control system. This can be concluded from the obs – bg time series at station Stuttgart, Germany (Fig. 15), which is situated just “downstream” of France and does not have documented changes around 1970. The obs – bg temperature time series of this station was homogeneous in 1969 and 1972 when France changed the radiosonde

instrumentation. If the bg were influenced by the breaks of the French temperature time series, a shift would be visible in the obs – bg time series at Stuttgart.

While over Europe this result may be plausible since there is much independent information available to make the bg apparently immune against even large breaks; the situation may be different when a large country changes its radiosonde equipment simultaneously. Figures 16 and 17 provide two examples that even in those situations the bg is independent enough to be used for adjustment. In the Japanese radiosonde time series (Fig. 16), major breaks are adjusted in 1968, 1981, and 1993–94. The Japanese composite of adjusted 1200 – 0000 UTC time series no longer shows spurious breaks (Fig. 16b).

A similar example is far eastern Russia (station IDs 31000–33000; Fig. 17). The most prominent shift in 1969 is caused by a change from MARS to RKZ5 radiosondes and is adjusted well by RAOBCORE. This composite also shows the difficulties in data availability

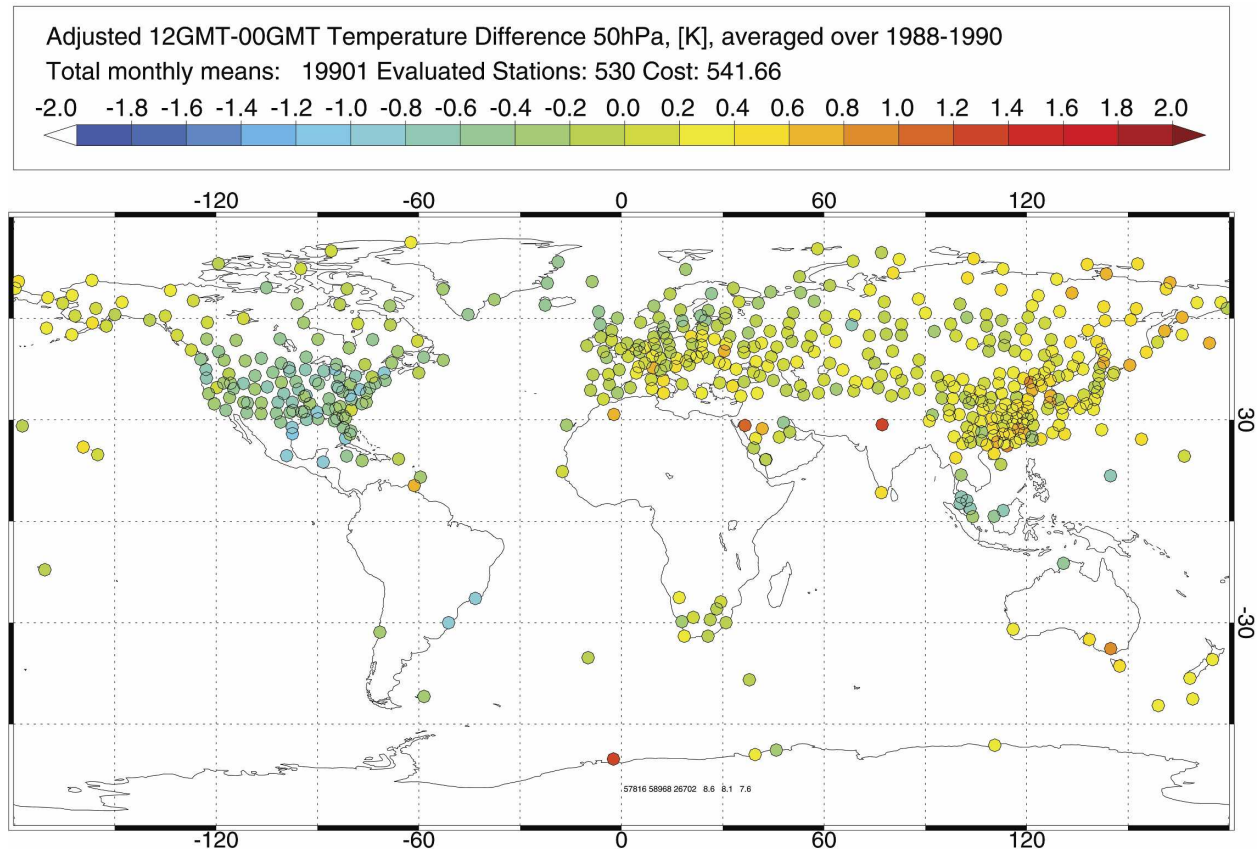


FIG. 13. Adjusted 1200 – 0000 UTC temperature differences in 50 hPa for the period 1988–90. Note improved spatial homogeneity compared to Fig. 1a.

over Russia in the 1990s. The most recent parts of most stations, before the data gaps had to be adjusted by about 1 K since their obs – bg time series, showed a distinct bias compared to a composite of mainly Japanese and Alaskan stations. The resulting adjusted 1200 – 0000 UTC time series is nevertheless relatively stationary (apart from a remaining annual cycle) as it should be. The daily mean adjustments in 1969 are about  $-0.7$  K at 50 hPa over eastern Russia (31000–33000). This compares well with daily mean adjustments of  $-0.6$  K over western Russia (south of  $60^{\circ}\text{N}$ , west of  $60^{\circ}\text{E}$ ), which have been taken at similar solar angles. This is remarkable since the bg over western Russia is potentially influenced by European radiosondes, whereas over eastern Russia it is more likely influenced by Alaskan and Japanese radiosondes (the latter have large breaks in 1968). Only the positive 1200 – 0000 UTC differences after adjustment in the 1960s is an indication of a slight overcorrection, due to a too-warm nighttime bg or a too-cool daytime bg caused by biases in the radiosondes 12 h earlier.

Figure 18 shows that the RAOBCORE adjustments substantially improve the spatial consistency of 1200 –

0000 UTC time series also in the early period 1964–66, although a few stations remain where the adjustments by RAOBCORE seem insufficient. Note also that many weather ships were operational in the 1960s until the early 1970s. For these ships, which operated up to the early 1970s, the adjustment of the most recent period removes the large biases visible in Fig. 18a.

Figure 19 shows the time series of the adjustments per month applied to the global radiosonde network. The combined length of all time series divided by the total number of adjustments applied (6233) yields about one adjustment every 7 yr. While this number may seem large, it is in good agreement with Thorne et al. (2005a) and is also supported by the number of radiosonde-type changes derived from GTS (about 700 from 1990 onward) and documented in CARDS (about 6700). About 30% of the adjustments coincide with metadata events within  $\pm 14$  days of the detected breakpoint. Most peaks are related to changes in large countries (e.g., Russia, France in 1969). There are no peaks evident in 1973, 1975, 1976, and 1978–79 where major changes in the ERA-40 observing system occurred.

The adjustments for all 1184 stations are available

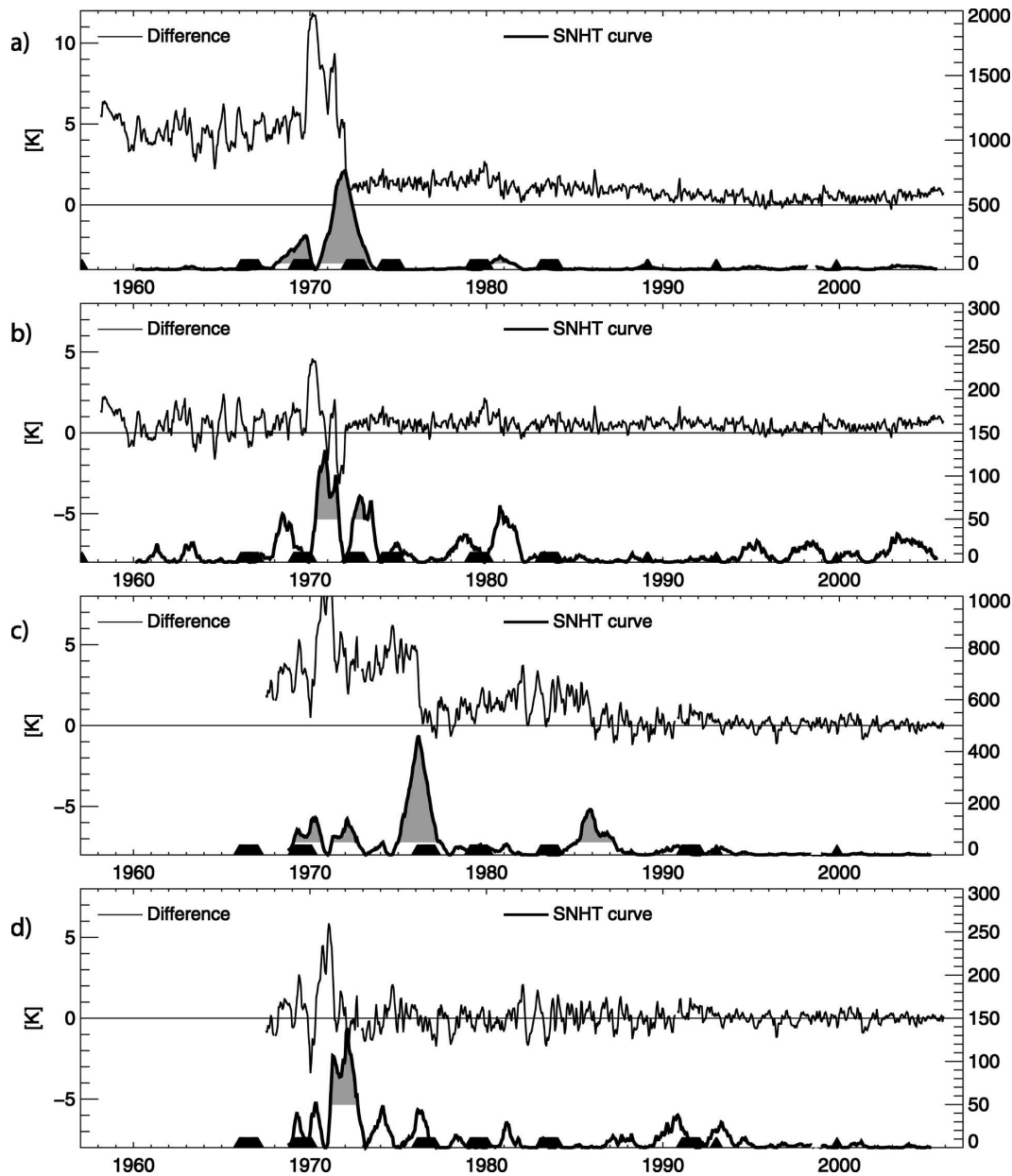


FIG. 14. Obs - bg difference at the 50-hPa level at 1200 UTC for radiosonde stations Trappes (7145, France, 48.77°N, 2.02°E, change from Mesural-FMO-1940B to Mesural-FMO-1943B in 1969 and to Mesural-FMO-1944C in 1972) and Rapa (91958, Chile, 27.61°S, 144.33°W, change from Mesural-FMO-1940B to Mesural-FMO-1943B in 1969 and to Mesural-FMO-1944C in 1976). Trappes (a) before and (b) after adjustment with RAOBCORE. Rapa (c) before and (d) after adjustment with RAOBCORE.

from <http://www.univie.ac.at/theoret-met/research/RAOBCORE/> as plots and as an ASCII-formatted file. The dataset will be updated annually, as soon as the input data from IGRA and ECMWF become available.

### 9. Sensitivity experiments

The homogenization process has several sources of uncertainties, as outlined, for example, by Thorne et al.

(2005a) and McCarthy et al. (2006, manuscript submitted to *J. Climate*). Since RAOBCORE is an automated system, it has been possible to test the sensitivity with respect to various parameters of the adjustment system.

The sensitivity of the following parameters has been tested: (i) 50–100 and 300–850-hPa layer global mean trends, which may be compared, for example, with the results collected in Fig. 3.4 of Karl et al. (2006); (ii)

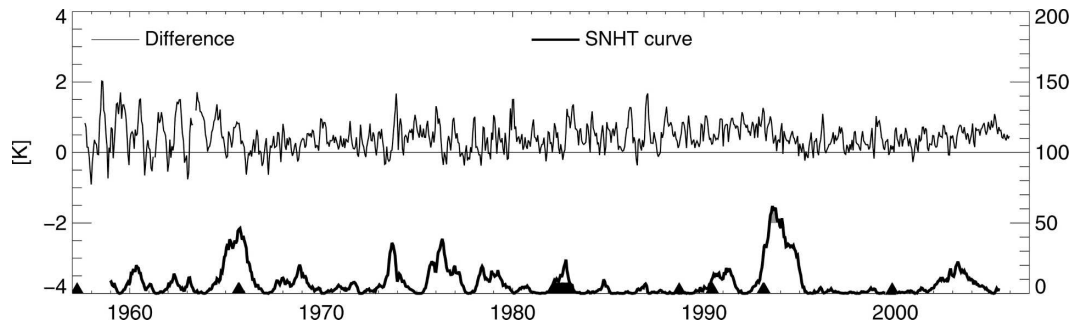


FIG. 15. Time series of obs - bg difference at 50 hPa at 0000 UTC for station Stuttgart (10739, Germany, 48.83°N, 9.2°E) downstream of the French radiosondes.

tropical (20°N–20°S) mean trends for the same period; (iii) spatial consistency of 50-hPa 1979–2004 trends, as defined in appendix B; (iv) spatial consistency of 1200 – 0000 UTC differences averaged over 1964–66; and (v) breakpoint count. The main results are summarized in Table 1.

Time series have been used for the trend comparison only if less than 24 months of data out of the 26-yr period 1979–2004 have been missing. If only 0000 or 1200 UTC trends have been available for a station, these have been regarded as daily mean trends. If both have been available, their average has been taken. The daily mean trends have been averaged to  $10^\circ \times 10^\circ$

latitude–longitude grid boxes and the mean trends from these grid boxes have been used to calculate the tropical (20°S–20°N) and global mean trends. The intermediate calculation of  $10^\circ \times 10^\circ$  trends helped alleviate the effects of the uneven spatial distribution of radiosondes.

The first two rows BG and BGADJ describe the unadjusted and adjusted versions of the ERA-40 + ECMWF background as they are used in the sensitivity experiment below. The original ERA-40 + ECMWF bg (BG) has almost neutral trends in the lower stratosphere, which are not supported by the University of Alabama in Huntsville (UAH) and RSS satellite prod-

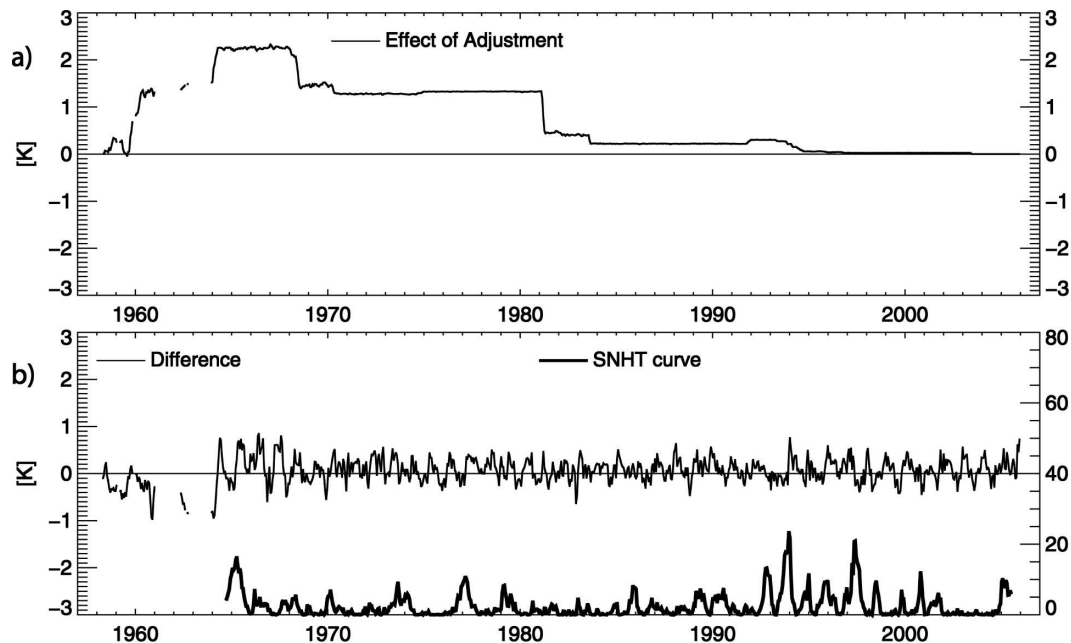


FIG. 16. (a) Effect of RAOBCORE adjustments on 1200 – 0000 UTC series of composite of northern Japanese radiosondes (IDs 47400–47700). (b) 1200 – 0000 UTC series after adjustment. Documented breaks in 1968, 1981, and 1993–94.

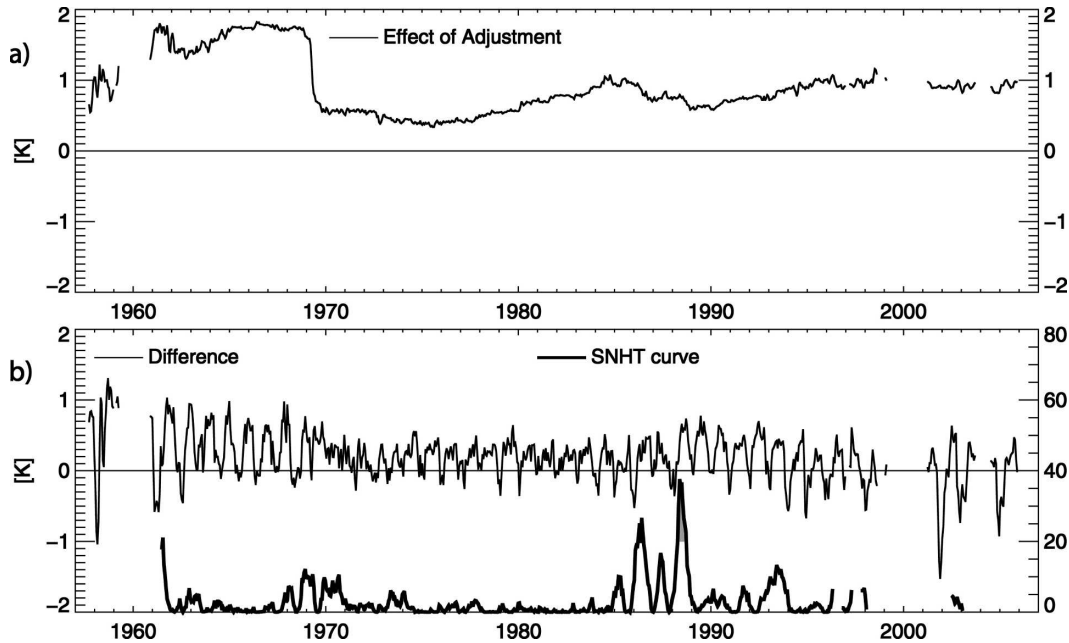


FIG. 17. (a) Effect of RAOBCORE adjustments on obs(1200 UTC) – obs(0000 UTC) time series for far eastern Russian radiosondes (IDs 31000–33000). (b) 1200 – 0000 UTC difference of adjusted time series. Note data gaps in recent periods. The strong annual cycle of the difference remains to be adjusted.

ucts and even less by the radiosondes. As one can see from row BGADJ, the adjustment described in section 6 introduces moderate stratospheric cooling (similar to that of MSU satellite products) into the bg.

Row UNADJ refers to the unadjusted radiosonde dataset used as input for RAOBCORE. Note the strong stratospheric cooling and large spatial heterogeneity compared to the bg.

Row RAOBCORE summarizes the results of RAOBCORE as used for the plots in this study, that is, with bg adjustment, with use of metadata, and with the thresholds described in section 4. The radiosonde trends adjusted with RAOBCORE show much less cooling than UNADJ and are very close to the existing homogenized radiosonde datasets [Hadley Centre Atmospheric Radiosonde Temperature Product version 2 (HadAT2), Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC); see Karl et al. 2006]. Although the RAOBCORE trends are not directly compared to satellite data in this paper, it is clear that they remain more negative than those derived from satellite products. The spatial consistency of both trends and 1200 – 0000 UTC difference after the adjustment is much better than for the original radiosonde dataset but is still not as high as the consistency of the ERA-40 bg trends and 1200 – 0000 UTC differences.

The other rows in Table 1 show results of sensitivity experiments. For NOMETA, RAOBCORE has been

applied with constant prior probability 0.02; that is, metadata information has been completely discarded. For ONLYMETA, the prior probability has been set to zero if no metadata were available and to 0.999 when metadata events (radiosonde-type changes and radiation correction changes) were documented. NOBGC refers to a RAOBCORE run where the bg was not adjusted before application of RAOBCORE. In the experiment NOBGC\_ONLYMETA the bg has not been adjusted and metadata have been treated as in ONLYMETA. STRICT refers to an experiment where the thresholds for break detection and adjustments have been set higher (40% higher  $T_k^s$  values required and 60% larger break size required than for the default).

The results suggest that the adjusted trends are relatively insensitive to changes in metadata treatment and to the increase of the breakpoint thresholds. However, there is sizeable sensitivity to the background adjustment applied before application of RAOBCORE. If the unadjusted bg is used as reference, the resulting RAOBCORE-adjusted trends shift toward more warming by more than  $0.3 \text{ K} (10 \text{ yr})^{-1}$  in the tropical stratosphere and by about  $0.04 \text{ K}$  in the tropical troposphere. In this case the adjusted radiosonde trends fit better to satellite-derived temperature trends but are still within the range of uncertainty of upper-air trends (see again Karl et al. 2006). The bg adjustment has relatively little impact on breakpoint count and on the

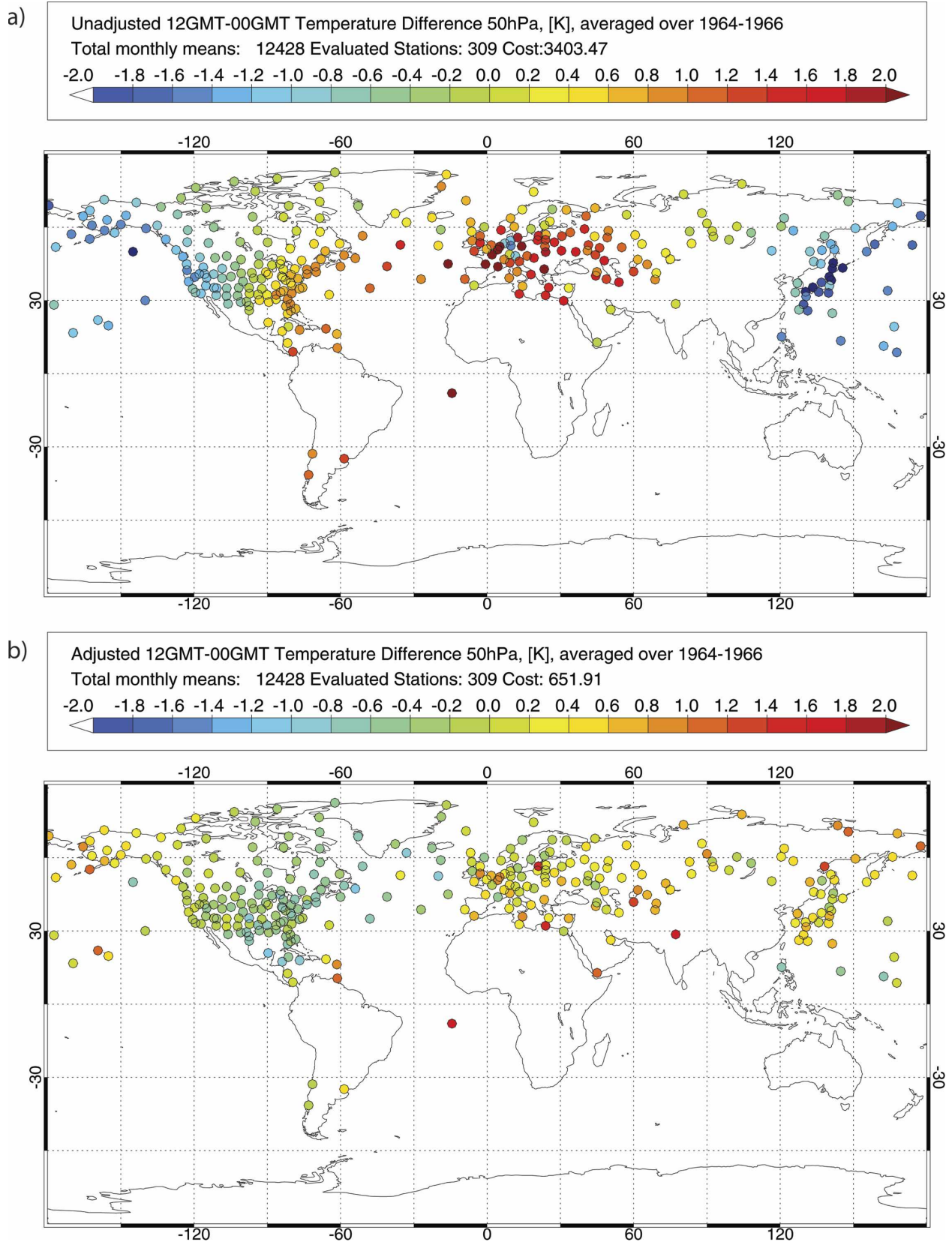


FIG. 18. The 1200 – 0000 UTC temperature differences in 50 hPa for the period 1964–66 (a) before and (b) after adjustment with RAOBCORE. Note weather ship stations over the Atlantic and Pacific whose time series end in the 1970s; their climatologies have been adjusted as described in section 7.

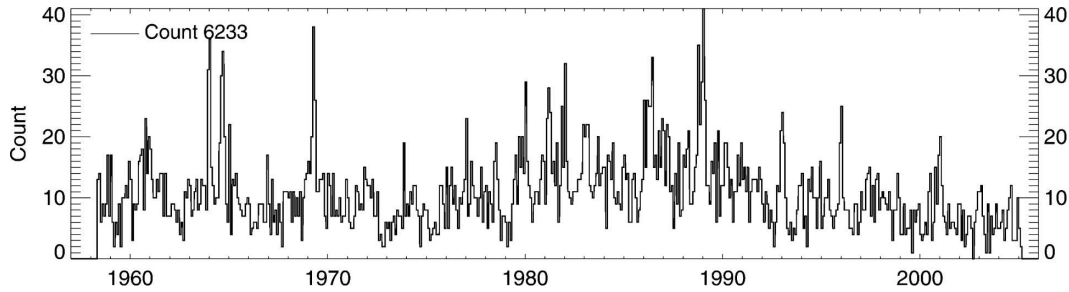


FIG. 19. Time series of number of adjustments per month. Peaks are related to metadata events in large countries, for example, radiation correction changes in 1982 and 1993. Break count includes climatology adjustments as explained in section 7.

spatial consistency of adjusted radiosonde time series. This shows that high spatial consistency of trends and 1200 – 0000 UTC differences, while a desirable property, is not sufficient to determine the quality of the diagnosed global mean upper-air trends.

While the unadjusted bg trends are most likely too weak in the stratosphere (see Fig. 3.4 in Karl et al. 2006) so that the experiment NOADJ is regarded as rather extreme, it shows the importance of a homogeneous reference. In view of mounting evidence for pervasive biases of the unadjusted radiosonde dataset particularly in the Tropics (Sherwood et al. 2005; Randel and Wu 2006), the bg adjustment applied in the best estimate of this paper (row RAOBCORE) may seem aggressive. There is indeed the possibility that the bg adjustment removes climate signals from the MSU and other satellite instruments. This is suggested by the good correspondence between ERA-40 MSU-4 equivalent and RSS (Mears et al. 2003) datasets, at least in the stratosphere, from circa 1989 onward (Santer et al. 2004). Therefore the difference in stratospheric trends be-

tween experiments RAOBCORE and NOADJ is probably larger than the true uncertainty of the radiosonde trends.

The sensitivity of the trends with respect to the bg adjustment is sizeable, even if adjustments are made only at documented breakpoints. Since the breakpoints are practically the same in the ONLYMETA and NOBGC\_ONLYMETA experiments, the trend differences must mainly come from break estimation, not from break detection. As is indicated in Fig. 7 the uncertainty in break estimation is at least 0.5 K in the stratosphere, even in recent years. Since almost every radiosonde time series contains at least one break between 1979 and 2004, trend differences in the order of 0.3 K, depending on the reference used, are not surprising. It has further been tried to use shorter intervals (4 yr instead of 8 yr) for break estimation, but this did not reduce the sensitivity to the bg adjustment either.

The high values of  $C$  and low break count values for the NOMETA and STRICT experiment suggest that many breaks remain undetected in these experiments.

TABLE 1. Results from sensitivity experiments with RAOBCORE. Trends in  $\text{K decade}^{-1}$  for globe and Tropics (in parentheses) valid for the period 1979–2004. Cost is value of cost function defined in Eq. (A1) at the 50-hPa level for 1200 – 0000 UTC temperature differences averaged over 1964–66 and for 1979–2004 temperature trends. Break count is total number of breaks detected, including about 760 adjustments of climatologies (see section 7).

| Acronym        | Description  | 50–100-hPa   | 300–850-hPa | Cost         | Cost 50-hPa     | Break count |
|----------------|--|--------------|-------------|--------------|-----------------|-------------|
|                |  | trend        | trend       | 50-hPa trend | 1200 – 0000 UTC |             |
| BG             | Unadjusted bg (except shift at 2001)<br>ERA-40 + ECMWF bg temperatures | –0.05(0.11)  | 0.15(0.02)  | 139          | 360             |             |
| BGADJ          | Adjusted bg  | –0.39(–0.34) | 0.11(–0.03) | 134          | 356             |             |
| UNADJ          | Unadjusted radiosondes   | –0.83(–0.94) | 0.09(–0.01) | 564          | 3401            |             |
| RAOBCORE       | RAOBCORE best estimate (control run)                                   | –0.66(–0.65) | 0.11(0.00)  | 192          | 652             | 6233        |
| NOMETA         | No metadata, constant prior probability 0.02                           | –0.65(–0.65) | 0.11(0.00)  | 193          | 757             | 5963        |
| ONLYMETA       | Adjustments only at documented changes                                 | –0.70(–0.73) | 0.12(0.03)  | 494          | 1652            | 2516        |
| NOBGC          | No adjustment of bg before RAOBCORE                                    | –0.44(–0.31) | 0.13(0.05)  | 197          | 638             | 6097        |
| NOBGC_ONLYMETA | No adjustment of bg before RAOBCORE,<br>only documented breaks         | –0.51(–0.46) | 0.15(0.06)  | 503          | 1635            | 2557        |
| STRICT         | $\Delta T$ of 0.5 K required at two places in break<br>profile         | –0.66(–0.64) | 0.11(0.02)  | 303          | 895             | 4093        |

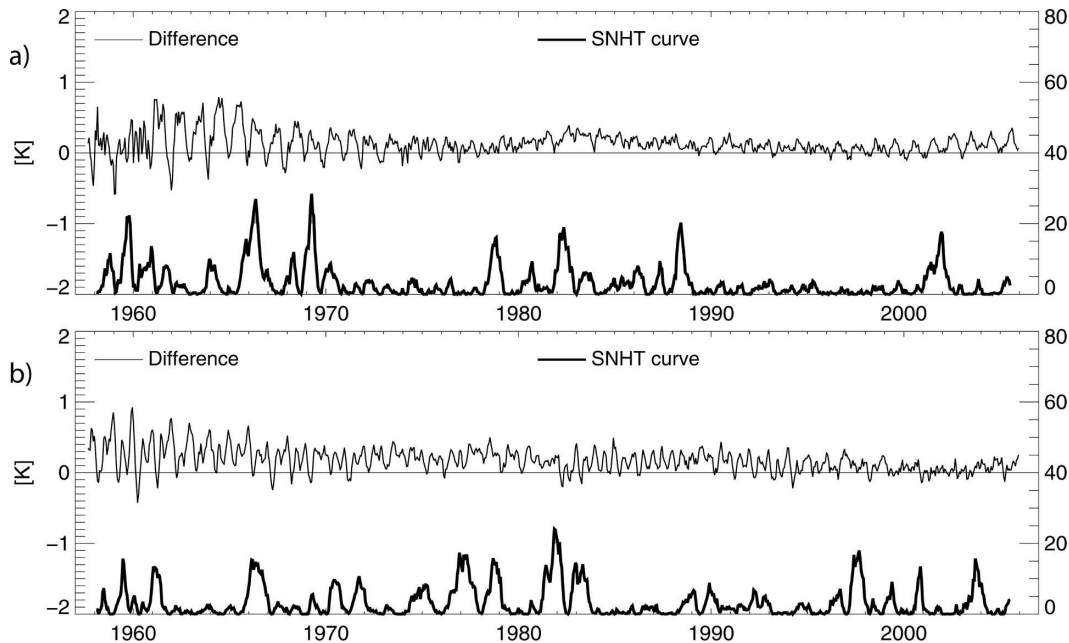


FIG. 20. Time series of composite mean 1200 – 0000 UTC differences from radiosondes (a) between 30°W and 40°E and (b) between 120°E and 120°W after adjustment with RAOBCORE. Note better homogeneity and almost neutral trends compared to Fig. 2.

One has to accept that there are thousands of breaks that need to be adjusted in order to get a spatially consistent dataset. The impact of these parameters on trends is small, which indicates that the largest breaks are found even without metadata and large thresholds.

The sensitivity experiments stress the importance of the reference series used for homogenization. The uncertainties in the bg adjustment lead to uncertainties that are similar to the differences between existing homogenized radiosonde and satellite datasets.

## 10. Conclusions and outlook

This paper has documented a method called RAOBCORE, which uses innovations (observations minus background forecasts) from a frozen data assimilation system, such as ERA-40 (Uppala et al. 2005), for automatic homogeneity adjustments of radiosonde temperature data. Innovations back to 1958 have been used. The method has been designed to make the adjusted radiosonde dataset as suitable for reanalyses as possible. This involves not only realistic trends but also completeness of the radiosonde dataset and adjustment of the absolute temperatures, not anomalies. It could be shown with several examples that RAOBCORE substantially reduces spurious trends in the 1200 – 0000 UTC differences (cf. Fig. 2 with Fig. 20) and improves the internal spatial consistency of the radiosonde measurements.

The global trend figures from RAOBCORE show good agreement with existing global homogenized radiosonde datasets, such as HadAT (Thorne et al. 2005a) and LKS (Lanzante et al. 2003b). More detailed comparisons with regional radiosonde datasets, for example, Comprehensive Alpine Radiosonde Dataset (CALRAS; Haerberli 2006) as well as MSU records (Mears et al. 2003; Christy et al. 2003), are in preparation.

Inhomogeneities introduced into the background forecast temperatures due to changes in the (satellite) observing system and the dependence of the background forecast error on the station to be tested and adjusted are regarded as the largest potential sources of uncertainty when using innovations. While the results from individual stations suggest that the latter problem is relatively small, the sensitivity of the RAOBCORE-adjusted trends with respect to the global mean bg is not negligible. The uncertainty (which is estimated  $0.3 \text{ K (10 yr)}^{-1}$  for the global mean 50–100-hPa layer and  $0.05 \text{ K}$  for the 300–850-hPa layer) of trends from the RAOBCORE-adjusted dataset can be reduced below that of existing upper-air datasets only if the inhomogeneities in the global mean obs – bg time series (as shown in Fig. 8 for the 50-hPa layer) can be attributed more clearly to either the bg or the obs.

The behavior of the ERA-40 data assimilation system in the satellite period is still investigated and in-



creasingly well understood (Uppala et al. 2006). With this building knowledge the uncertainty of the ERA-40 bg and thus of the resulting RAOBCORE trends can likely be reduced. It is quite possible that the bg adjustment used in this study was too aggressive and that therefore the RAOBCORE best estimate for global trends after homogenization still shows too much cooling/little warming. A more thorough investigation of this matter is under way but is beyond the scope of this article.

Some radiosonde temperature time series have a substantial bias even in their most recent parts. For these stations the biases have been reduced by adjusting with a composite of homogenized neighboring radiosondes in an extra step after the homogenization procedure. This step is not compulsory for climate trend analysis but adds substantial value for climate data assimilation applications, which are affected by absolute biases (Dee and Da Silva 1998; Uppala et al. 2006).

Apart from the correction algorithm itself this article has documented gaps in both the ERA-40 and the IGRA (Durre et al. 2006) radiosonde datasets. The union of both datasets is about 5%–10% larger than the individual datasets. Efforts to create a comprehensive global radiosonde dataset must therefore continue.

In view of the results gained so far, it seems worthwhile to perform a pilot climate data assimilation of the period 1939–57 in order to be able to adjust also these time series using the innovations gained from such an assimilation. Numerous radiosondes are available for this period (Durre et al. 2006; Bronnimann 2003). Innovations from a four-dimensional variational data assimilation (4DVAR) system would add valuable information to these data, which are otherwise very difficult to homogenize. Preliminary results with RAOBCORE also indicate that homogenization based on analysis of innovations is feasible also for radiosonde winds. Put into a more general context, the method of reanalysis (or climate data assimilation) together with proper archival of the innovations may become a key method for the improvement of historical observations. The recent success of assimilations using surface pressure only (Compo et al. 2006) together with potential application for homogenization makes the idea of a 100-yr reanalysis attractive, despite the sparseness of the available data.

The efforts described here are part of a long-term activity to provide a mature and complete homogenized radiosonde dataset suitable as input for the next major European reanalysis planned for the end of this decade.

*Acknowledgments.* The foundation for this work has been laid during the European Commission Contract

MEIF-CT-2003-503976, which enabled the author to work at ECMWF in 2004. I thank Sakari Uppala and Adrian Simmons in particular for their continuous support. After 2004 the work has been funded by project P18120-N10 of the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (FWF). Christina Tavolato and Stefan Sperka were helpful in detecting bugs in the adjustment procedure. The comments of the anonymous reviewers led to substantial improvements of the original manuscript.

## APPENDIX A

### A Spatial Consistency Measure for Trends and Day–Night Differences

To estimate the spatial consistency of trends and 1200 – 0000 UTC differences, a simple cost function is defined as

$$C(p) = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N \{ \Delta x_{ij}(p) \exp[-d_{ij}(1000 \text{ km})^{-1}] \}^2, \quad (\text{A1})$$

where  $p$  is the pressure level,  $i, j$  are station indices,  $N$  is the number of stations,  $\Delta x_{ij}$  is the difference quantity between stations, and  $d_{ij}$  is the spherical distance between stations in kilometers. Here  $\Delta x_{ij}$  is either the difference between trends  $(\Delta \partial T / \partial t)_{ij}$  or the difference between 1200 – 0000 UTC differences at stations  $i, j$ . Strongly deviant trends/differences at densely covered areas contribute most to the cost. The cost function  $C$  for radiosondes should reach similar values as the ERA-40 bg or satellite products after homogenization.

## APPENDIX B

### Adjustment of the Global Mean bg Time Series

In section 6 it has been shown that the global mean bg has spurious breaks due to insufficient satellite bias corrections. The signature of these breaks can be found throughout the globe. Although it has a horizontally smooth pattern compared to the distribution of breaks in radiosonde records, it is not constant. In general the influence of the satellite data and excessive latent heating on the bg is smaller in regions with dense radiosonde observation coverage and in the extratropics. Therefore the adjustment is scaled with the radiosonde density in the following way:

$$\Delta \text{bg}(\lambda, \varphi, p, t) = -\overline{\text{obs} - \text{bg}}(p, t) w(\lambda, \varphi, t), \quad (\text{B1})$$

where  $\Delta \text{bg}(\lambda, \varphi, p, t)$  is the global mean obs – bg difference and the weighting function  $w$  is defined as

$$W(\lambda, \varphi, t) = [0.5 + \cos(\varphi)][1.2 - \rho(\lambda, \varphi, t)/\rho_{\max}(t)]/1.5. \quad (\text{B2})$$

The stronger weight at low latitudes helps to reduce the strong heating of the bg in the Tropics, which is considered excessive. The radiosonde observation density  $\rho$  is defined as

$$\rho(\lambda, \varphi, t) = \sum_i^{N(t)} \exp[-d_i(\lambda, \varphi)/700\text{km}]f_i, \quad (\text{B3})$$

where  $d_i$  is the spherical distance between location  $(\lambda, \phi)$  and radiosonde  $i$ . The factor  $f_i$  takes into account whether the radiosonde reports once ( $f_i = 1$ ) or twice daily ( $f_i = 5$ ). It was set to 5, not 2, since sites with twice daily launches tend to be better maintained. Here  $N(t)$  is the number of active radiosondes and  $\rho_{\max}(t)$  is the maximum radiosonde density found at a particular time. Finally, the weights  $w$  are adjusted by a constant factor such that the global mean  $\Delta b_g$  is equal to  $-\text{obs} - \text{bg}$  when averaged over all radiosonde stations. Figure 18 in Haimberger (2005) shows the spatial pattern of the weight  $w$ , which shows little variation in time.

With this choice of parameters the maximum adjustment for an individual radiosonde site is about 1.6 times the global mean adjustment (e.g., in the South Pacific), and the minimum adjustment is below 0.2 times the global mean adjustment (over central Europe, China). It has been tuned to reduce as much as possible the conspicuous jumps due to the erroneous bias correction of the *NOAA-4* radiances between January 1975 and September 1976. Figure 19 in Haimberger (2005) shows the  $\text{obs} - \text{bg}$  series averaged over the radiosondes south of  $25^\circ\text{N}$  before and after adjustment with the weighted global mean bg. Although this simple adjustment of the bg described here can by no means remove all biases, at least the breaks in January 1975 and September 1976 are substantially reduced in the average over this data-sparse region. For other breaks, for example, due to problems with *NOAA-9*, other horizontal weighting functions may be optimal, but so far this has not been tested.

#### REFERENCES

- Aguilar, E., 2000: The upper air station history: A brief description. 7 pp. [Available from National Climatic Data Center, 151 Patton Ave., Asheville, NC 28801-5001.]
- Alexandersson, H., and A. Moberg, 1997: Homogenization of Swedish temperature data. Part I: Homogeneity test for linear trends. *Int. J. Climatol.*, **17**, 25–34.
- Andrae, U., N. Sokka, and K. Onogi, 2004: The radiosonde temperature bias correction in ERA-40. ECMWF ERA-40 Project Rep. Series 15, 34 pp.
- Bronnimann, S., 2003: A historical upper air-data set for the 1939–44 period. *Int. J. Climatol.*, **23**, 769–791.
- Christy, J. R., and W. B. Norris, 2006: Satellite and VIZ–radiosonde intercomparisons for diagnosis of nonclimatic influences. *J. Atmos. Oceanic Technol.*, **23**, 1181–1194.
- , —, 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.
- Compo, G. P., J. S. Whitacker, and P. D. Sardeshmukh, 2006: Feasibility of a 100-year reanalysis using only surface pressure data. *Bull. Amer. Meteor. Soc.*, **87**, 175–190.
- Courtier, P., and Coauthors, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quart. J. Roy. Meteor. Soc.*, **124**, 1783–1807.
- Daan, H., 2002: Guide to the GCOS surface and upper air networks: GSN and Guan. Tech. Doc. WMO-TD 1106, GCOS-73, 37 pp.
- Dee, D. P., 2004: Detection and correction of model bias during data assimilation. *Proc. ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean*, Reading, United Kingdom, ECMWF, 65–74.
- , and A. Da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.*, **124**, 269–295.
- DeGroot, M. H., 1986: *Probability and Statistics*. Addison Wesley, 723 pp.
- Ducre-Robetaille, J., L. Vincent, and D. Boulet, 2003: Comparison of techniques for detection of discontinuities in temperature series. *Int. J. Climatol.*, **23**, 1087–1101.
- Durre, I., R. Vose, and D. B. Wuertz, 2006: Overview of the Integrated Global Radiosonde Archive. *J. Climate*, **19**, 53–68.
- ECMWF, cited 2000: IFS documentation (CY23r4). [Available online at <http://www.ecmwf.int/research/ifsdocs/CY23r4/index.html>.]
- Eskridge, R. E., J. K. Luers, and C. R. Redder, 2003: Unexplained discontinuity in the U.S. radiosonde temperature data. Part I: Troposphere. *J. Climate*, **16**, 2385–2395.
- Free, M., and D. J. Seidel, 2005: Causes of differing temperature trends in radiosonde upper air data sets. *J. Geophys. Res.*, **110**, D07101, doi:10.1029/2004JD005481.
- , and Coauthors, 2002: Creating climate reference datasets: CARDS workshop on adjusting radiosonde temperature data for climate monitoring. *Bull. Amer. Meteor. Soc.*, **83**, 891–899.
- , D. J. Seidel, J. K. Angell, J. Lanzante, I. Durre, and T. C. Peterson, 2005: Radiosonde atmospheric temperature products for assessing climate (RATPAC): A new data set of large-area anomaly time series. *J. Geophys. Res.*, **110**, D22101, doi:10.1029/2005JD006169.
- Gaffen, D. J., 1996: A digitized metadata set of global upper-air station histories. NOAA Tech. Memo. ERL ARL-211, 3 pp.
- Haerberli, C., 2006: *The Comprehensive Alpine Radiosonde Dataset (CALRAS)*. Vol. 4. *Wiener Meteorologische Schriften*, Facultas, 297 pp.
- Haimberger, L., 2005: Homogenization of radiosonde temperature time series using ERA-40 analysis feedback information. ECMWF ERA-40 Project Rep. Series 23, 75 pp.
- , 2006: Bias correction of conventional observations. *Proc. ECMWF–NAF Workshop on Bias Estimation and Correction in Data Assimilation*, Reading, United Kingdom, ECMWF, 93–106.
- Harris, B. A., and G. A. Kelly, 2001: A satellite radiance bias correction scheme for data assimilation. *Quart. J. Roy. Meteor. Soc.*, **127**, 1453–1468.
- Hernandez, A., G. Kelly, and S. Uppala, 2004: The TOVS/

- ATOVS observing system in ERA-40. ECMWF ERA-40 Project Rep. Series 16, 45 pp.
- Hollingsworth, A., D. B. Shaw, P. Lönnberg, L. Illari, and A. J. Simmons, 1986: Monitoring of observation and analysis quality by a data assimilation system. *Mon. Wea. Rev.*, **114**, 861–879.
- Karl, T. R., S. J. Hassol, C. D. Miller, and W. L. Murray, Eds., 2006: Temperature trends in the lower atmosphere: Steps for understanding and reconciling differences. Climate Change Science Program and the Subcommittee on Global Change Research Rep., 180 pp.
- Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247–267.
- Lanzante, J. R., 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.
- Lewis, J. M., S. Lakshminarayanan, and S. Dhall, 2006: *Dynamic Data Assimilation: A Least Squares Approach*. Cambridge University Press, 654 pp.
- Li, X., G. Kelly, S. Uppala, R. Saunders, and J. Gibson, 2006: The use of VTPR raw radiances in ERA-40. ECMWF ERA-40 Project Rep. Series 21, 24 pp.
- Luers, J. K., and R. E. Eskridge, 1995: Temperature corrections for the VIZ and Vaisala radiosondes. *J. Appl. Meteor.*, **34**, 1241–1253.
- 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.
- Nash, J., and F. J. Schmidlin, 1987: WMO international radiosonde intercomparison: Final report. Tech. Doc. WMO/TD Rep. 195, 123 pp.
- , R. Smout, T. Oakley, and S. Kurnosenko, 2005: WMO intercomparison of high quality radiosonde systems: Final report. WMO Rep., 118 pp.
- Onogi, K., 2000: The long term performance of the radiosonde observing system to be used in ERA-40. ECMWF ERA-40 Project Rep. Series 2, 77 pp.
- Randel, W. J., and F. Wu, 2006: Biases in stratospheric and tropospheric temperature trends derived from historical radiosonde data. *J. Climate*, **19**, 2094–2104.
- , and Coauthors, 2004: The SPARC intercomparison of middle-atmosphere climatologies. *J. Climate*, **17**, 986–1003.
- Redder, C. R., J. K. Luers, and R. E. Eskridge, 2004: Unexplained discontinuity in the U.S. radiosonde temperature data. Part II: Stratosphere. *J. Atmos. Oceanic Technol.*, **21**, 1133–1144.
- Santer, B. D., and Coauthors, 2004: Identification of anthropogenic climate change using a second-generation reanalysis. *J. Geophys. Res.*, **109**, D21104, doi:10.1029/2004JD005075.
- , and Coauthors, 2005: Amplification of surface temperature trends and variability in the tropical atmosphere. *Science*, **309**, 1551–1556.
- 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.
- Sherwood, S., J. Lanzante, and C. Meyer, 2005: Radiosonde daytime biases and late 20th century warming. *Science*, **309**, 1556–1559.
- Thorne, P. W., D. E. Parker, S. F. B. Tett, P. D. Jones, M. McCarthy, H. Coleman, and P. Brohan, 2005a: Revisiting radiosonde upper air temperatures from 1958 to 2002. *J. Geophys. Res.*, **110**, D18105, doi:10.1029/2004JD005753.
- , and Coauthors, 2005b: Vertical profiles of temperature trends. *Bull. Amer. Meteor. Soc.*, **86**, 1471–1476.
- Trenberth, K. E., and L. Smith, 2006: The vertical structure of temperature in the Tropics: Different flavors of El Niño. *J. Climate*, **19**, 4956–4970.
- Uppala, S., and Coauthors, 2005: The ERA-40 re-analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012.
- , G. Kelly, B. K. Park, P. Kallberg, and A. Untch, 2006: Experience in estimation of biases in ECMWF reanalyses. *Proc. ECMWF/NWP-SAF Workshop on Bias Estimation and Correction in Data Assimilation*, Reading, United Kingdom, ECMWF, 207–224.
- von Storch, H., and F. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.