ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 28, NO. 5, 2011, 1011–1022

Impacts of Reference Time Series on the Homogenization of Radiosonde Temperature

GUO Yanjun*[1,2,3] (郭艳君) and DING Yihui[1] (丁一汇)

[1] *Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing* 100029

[2] *National Climate Center, China Meteorological Administration, Beijing* 100081

[3] *Graduate University of Chinese Academy of Sciences, Beijing* 100049

(Received 21 June 2010; revised 23 November 2010)

ABSTRACT

Using radiosonde temperatures of 92 selected stations in China, the uncertainties in homogenization processes caused by different reference series, including nighttime temperature, the NCEP (National Centers for Environmental Prediction) and ERA-40 (European Centre for Medium-Range Weather Forecasts) forecasting background, are examined via a two-phase regression approach. Although the results showed limited consistency in the temporal and spatial distribution of identified break points (BPs) in the context of metadata events of instrument model change and correction method, significant uncertainties still existed in BP identification, adjustment, and impact on the estimated trend. Reanalysis reference series generally led to more BP identification in homogenization. However, those differences were parts of global climatic shifts, which may have confused the BP calculations. Discontinuities also existed in the reanalysis series due to changes in the satellite input. The adjustment values deduced from the reanalysis series ranged widely and were larger than those from the nighttime series and, therefore, impacted the estimated temperature trend.

Key words: China, radiosonde temperature, homogenization, uncertainty, reference time series

Citation: Guo, Y. J., and Y. H. Ding, 2011: Impacts of reference time series on the homogenization of radiosonde temperature. *Adv. Atmos. Sci.*, **28**(5), 1011–1022, doi: 10.1007/s00376-010-9211-3.

1. Introduction

Within the satellite and radiosonde temperature analysis community, there is agreement that the uncertainties in long-term changes are substantial. Changes in instrumentation and protocol are present in both radiosonde and satellite records, obscuring the modest long-term trends (IPCC, 2007).

It is generally accepted that uncertainties are present in radiosonde temperature time series, even if they are homogenized (IPCC, 2007). Free and Seidel (2005) argued that there are three sources of uncertainty: the data source, the sampling method, and the homogenization technique. The World Meteorological Organization (WMO) Climate Program guidelines on climate metadata and homogenization (Llanso, 2003) show at least 14 data homogenization assessment techniques and suggest more. Most of the techniques use successive differences instead of actual data to calculate correlation coefficients. If one or both of the sites has inhomogeneities within the time period used to calculate the correlation, this reduces the risk that the estimated correlation between the candidate and reference sites will be poor.

Previous research showed that homogenization by different techniques on the same original data might lead to conflicting conclusions. Therefore, it is highly necessary to compare different techniques of homogenization. Although several studies have compared various techniques (Free and Seidel, 2005; Reeves et al., 2007; Guo et al., 2008; Guo and Ding, 2009), the impact of the reference time series on homogenization has not yet been assessed. Thus, this paper presents a preliminary and original study on this topic.

The method used to form the reference time series can be important and may need to be tailored specifi-

*Corresponding author: GUO Yanjun, gyj@cma.gov.cn

cally to the network and adjustment methodology (Peterson et al., 1998). The optimum reference series are homogeneous and highly correlated with the target series and improve detection of break points (BP) in homogenization. It may be impossible to build a completely homogeneous reference series with unknown inhomogeneous data, but several techniques have been used to minimize potential inhomogeneities. In previous studies, the most popular technique to build reference series involves using the data of neighbor stations. However, neighbor reference series are not suitable for cases in which the instruments or correction methods are modified simultaneously over a large spatial scale, as in China (Zhai and Eskridge, 1996). The nighttime series is another widely applied reference approach because bias of solar radiation correction is the major source of error in radiosonde observation and the correction improvements generally have a larger impact on daytime data. Sherwood et al. (2005) demonstrated the difference between day and night time series can potentially reveal undocumented BPs. Reanalysis data for free-atmosphere temperature, including both temperature and wind data reproduced from model dynamics, can also aid the design of reference series (Haimberger, 2007; Guo et al., 2008; Guo, 2008; Guo and Ding, 2009).

The two-phase regression (TPR) has been successfully used in the homogenization of radiosonde temperature data in China. For example, Zhai and Eskridge (1996) and Guo et al. (2008) conducted experiments on several stations with nighttime reference. Guo and Ding (2009) homogenized the temperature time series from most of the radiosonde stations in China with an NCEP reanalysis reference. The aim of this paper is to demonstrate the impact of reference on temperature homogenization in China. Section 2 describes the applied data and the considered approach. The adjustments with different reference series (i.e. nighttime, NCEP reanalysis, and ERA-40 reanalysis) are compared in sections 3–6, and the impact of different reference series on estimated temperature trends by radiosonde data are presented in section 7. Some conclusions are deduced in section 8.

2. Materials and method

Radiosonde station histories are known to be incomplete and, in many cases, inaccurate. Therefore, adjustment of radiosonde temperature series must also consider the potential for breakpoints existing within the dataset. The raw radiosonde records are temperatures at seven mandatory pressure levels (850, 700, 500, 400, 300, 200, and 100 hPa) from the Chinese National Meteorological Information Center (NMIC)/China Meteorological Administration (CMA). Quality control (QC, see more details for QC in Appendix A) was performed on twice-daily observations at 0000 UTC and 1200 UTC via the hydrostatic method (Collins, 2000). Data that were filtered out by quality control were treated as missing. The time series were removed if more than 30% were missing. The time series data were considered missing if either data at the 0000 UTC or the 1200 UTC was missing. After this procedure, the number of candidate stations was reduced from 116 to 92 (Guo and Ding, 2009). The 0000 UTC and 1200 UTC series were merged into a single series for the final homogenization.

The TPR method is a technique to detect BPs which was initially applied by Solow (1987) and was developed by Easterling and Peterson (1995). The TPR tests the significance of a two-phase fit to each point in a series of differences between target and reference series using (1) a likelihood ratio statistic and (2) the difference in the means of the difference series before and after the potential discontinuity evaluated using a t-test (assessed at the 95% confidence interval). Each of the discontinuities that was thus identified was further tested using a multi-response permutation procedure. The adjustment was calculated as the difference in the means of the (station reference) difference series for all data before and after the discontinuity, and it was applied to all data points preceding the break point. Appendix B provides a brief overview [or see Guo et al. (2008) and Guo and Ding (2009) for more detail].

TPR was applied to three references which include the nighttime series (1200 UTC), the reanalysis series from the NCEP, and the series from the ERA-40. The NCEP reanalysis data were provided in a monthly averaged format and were available for 1948–2005 in grid intervals of 2.5° (Kistler et al., 2001). Data corresponding to the selected candidate station locations in China were extracted from the global reanalysis archive as a representative sample of the Global Climate Observing System (GCOS) Upper-air Network (Daan, 2002). Further information on this data set can be found at www.cpc.ncep.noaa.gov/products/wesley/reanalysis.html.

The observations used in ERA-40 were gathered from many sources. The observing systems evolved considerably over the reanalysis period, with data provided by a succession of satellite-borne instruments from the 1970s onward, supplemented by increasing numbers of observations from aircraft, ocean buoys, and other surface platforms, but with a declining number of radiosonde ascents after the late 1980s. The data available to the public were from 1 September 1957 to 1 August 2000. The horizontal resolution was

Table 1. Major changes of instrumentation model and correction method for Chinese radiosonde observation during last 50 years.

	Change content	Date
Instrumentation model change	Change RZ 049 to GZZ-2	Around 1966 (1963–1969)
	Change GZZ-2 to GTS1-LBand Radar	Around 2002
Correction method change	Radiation correction at levels upper than 300 hPa	1966
	Acceleration of gravity change	1999
	Radiation correction for all levels	2001

Table 2. BP numbers detected at each level by the three references.

Level (hPa)	Nighttime	NCEP	ERA-40
100	163	308	278
200	110	437	269
300	166	298	270
400	110	208	241
500	105	206	178
700	62	113	110
850	50	60	82
Total	766	1630	1428

$2.5° \times 2.5°$ with 23 vertical levels (7, 5, 3, 2, and 1 hPa, in addition to those used in the NCEP reanalysis). Details of this data set can be found at www.ecmwf.int/research/era/era40survey/.

During the last 50 years, radiosonde observation in China experienced several changes of instrument and method which were recorded by metadata for each station. Based on the metadata of radiosonde stations in China, major changes in both instrumentation model and correction method are summarized in Table 1. For example, during the last 50 years, three radiosonde modes (RZ049, GZZ-2 and GTS1-LBand Radar) were used in China. RZ049 is Russia mode which was used in 1950s and early 1960s, the other two modes are made by China. GZZ-2 is also be called as mode 59 which replaced RZ049 around 1966. The latest mode change is around 2002: GTS1-LBand Radar replaced GZZ-2. There are 3 correction method change in past 50 years. Although the accuracy and completeness of the metadata may not be perfect, these records are helpful in estimating the timing of changes in the error characteristics of the radiosonde temperature time series. The significance of metadata events was examined by comparison of BP detected by different references.

3. BP distributions as identified by different reference series

In the TPR approach, the homogenization was performed level by level. The t-test was applied to examine the significance of the initially identified BP, and, if it was significant above 95% confidence level, the BP was adjusted. Table 2 lists the BP numbers accounting for all stations detected at each level using the nighttime, NCEP, and ERA-40 reference time series. The numbers of BPs detected by the nighttime reference series were much smaller than the reanalysis, with a ratio less than 1:2. The BP numbers roughly increased as the altitude increased. For example, the minimum BP numbers were commonly at 850 hPa, which were only 14%–30% of the maxima. This finding is reasonable and consistent with our expectation, because instrument bias can be enhanced at higher elevations and lower air density due to influences from solar radiation.

Figure 1 exhibits the frequency distribution of BP numbers identified in individual radiosonde temperature time series in China for 1958–2005 by three references. Due to the TPR procedure, BPs were identified individually by mandatory levels, and the vertical coherence of BPs was discernable but not completely (Guo et al., 2008). The problem was not taken into account in this work, and the BP numbers shown in Fig. 1 are the total for all 7 levels. According to the three references, the distributions of identified BPs from the same temperature series are different. In certain ranges, station BPs from the two reanalysis references (NCEP and ERA-40) exhibit more similar correlation than those from nighttime and either reanalysis reference. For example, according to the NCEP reference, 4–7 BPs were detected at 75% of the stations, and 2–7 BPs were detected at 90% of the stations with the ERA-40 reference. The BP numbers detected by the nighttime reference had a nonsymmetrical distribution. The peak frequency in this case was 2 BP, and

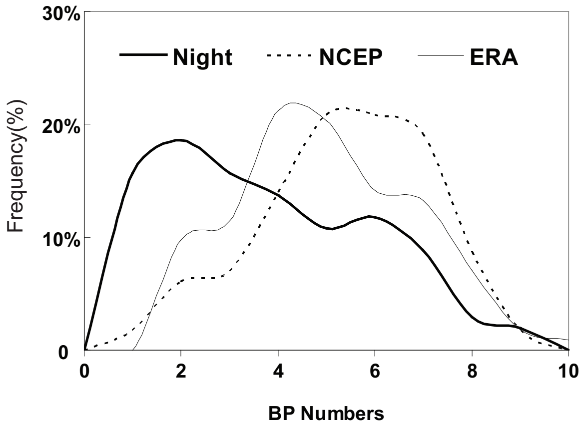

Fig. 1. Comparison of the frequency distributions of BP numbers produced by three different references, for individual radiosonde temperature time series in China for 1958–2005.

75% of the time series contained 1–4 BP.

The number of stations with the BPs in each year from the three homogenization procedures and the major metadata events are shown in Fig. 2. The white and shaded bars in Fig. 2 denote metadata events (mode and correction method changes, respectively) from the radiosonde network in China. The correction methods change occurred at most stations simultaneously in 1966, 1999, and 2001, whereas the instrument modes were renewed gradually around 1966 (1963–1969) and after 2002. The three homogenization procedures were robust to correction method or instrument changes in 1966 and to correction changes in 1999 and 2001 for the stations with a peak in BP around 1966 and 2000. However, the frequent change of instrument mode during 1963–1969 and 2002–2007 did not lead to corresponding increases in the number of stations with BPs and implies that correction method improvement is more likely to cause a BP than instrument change.

How to distinguish BPs from actual climatic shifts is one of the challenges in temperature time series homogenization. Previous studies have clarified that four climatic events that caused temperature shifts may be confused with BPs: the Agung eruption in March 1963, the El Chichón eruption in April 1982, the Pinatubo eruption in June 1991, and the strongest ever ENSO event in 1997–1998 (Free and Angell, 2002; Seidel et al., 2004). These events are shaded in Fig. 2 over two-year periods. The peak number of stations with BPs from reanalysis homogenization around 1978 and 1986 were most likely related to the introduction of satellite observations to the reanalysis models. Problems with the National Oceanic and Atmospheric Administration (NOAA) satellites NOAA-4 and NOAA-9 coincided with the end of NOAA-11 and the start and end of NOAA-14 (Christy and Norris, 2006). Many authors have noted that the discontinuities in the lower stratosphere and upper troposphere around 1979 were caused by biases from both the NCEP and ERA models (Santer et al., 1999; Kistler et al., 2001; Sturaro, 2003; Bengtsson et al., 2004; Wang et al., 2006).

4. Consistency of identified BPs with the metadata

Most homogenization approaches generally identify BPs with limited accuracy in the timing of BPs. The TPR approach treats two BPs as one if they are de-

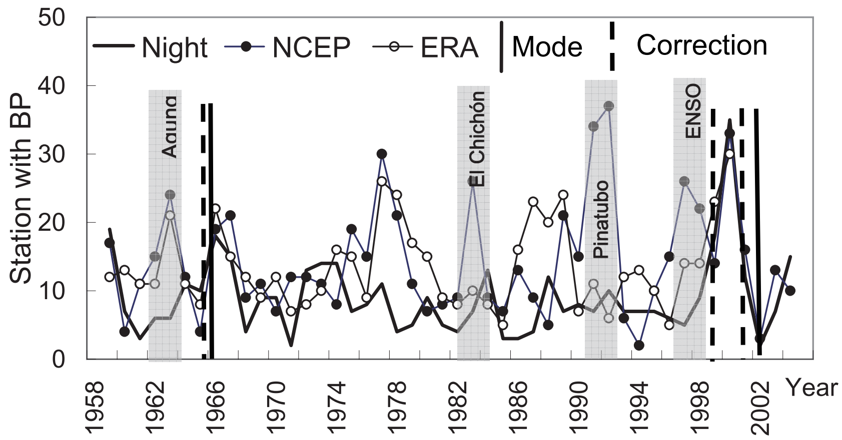

Fig. 2. Number of stations with BP for each year, and metadata events of instrument mode change (solid line) and correction method changes (dashed line) at 92 Chinese stations for 1958–2005, the shaded area indicate the period of volcano eruption and ENSO events.

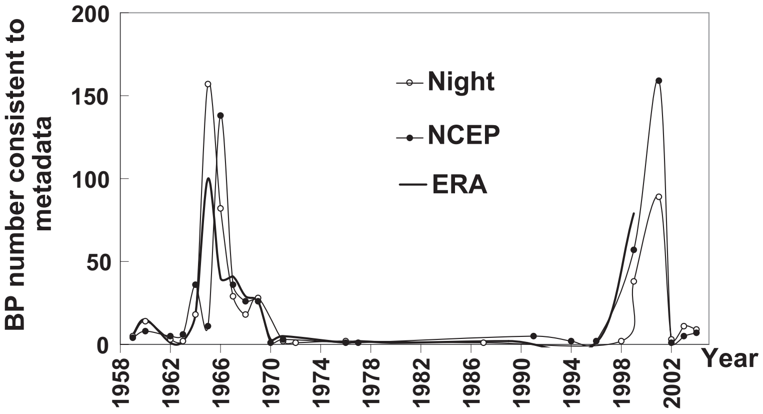

Fig. 3. Comparison of BP numbers consistent to metadata events for each year by three procedures at 92 Chinese stations for 1958–2005.

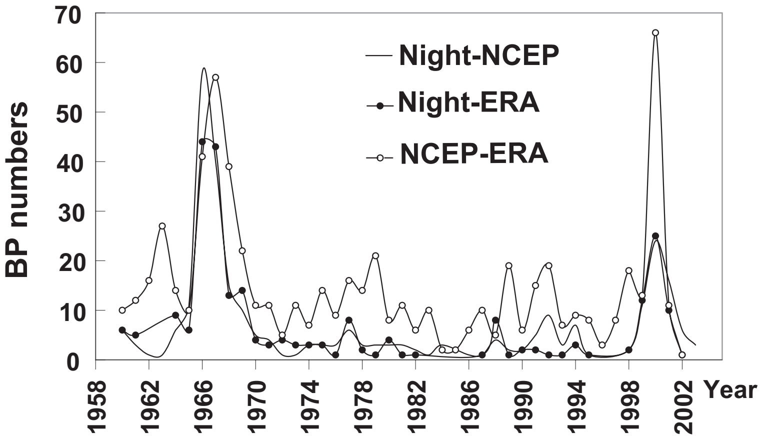

Fig. 4. Comparison of BP numbers consistent with each other for each year for 1958–2005.

tected within two adjacent years. Using this time interval, we investigated the temporal consistency of detected BPs with the metadata events (Fig. 3). The peak values in Fig. 3 are surprisingly consistent in both 1966 and 2000, when the correction method was changed in the Chinese radiosonde network (see Fig. 2).

If those BPs detected by a statistical approach can be verified by a given history record, the reliability of the homogenization approach is proved by the consistency of identified BPs and the metadata events. Table 3 compares the percentage of detected BP numbers consistent with the metadata at every level for the three procedures, which is defined as detected BPs corresponding temporally to metadata events within two years. The BPs identified by the nighttime reference series were more consistent with the metadata events than those identified by the reanalysis reference series. A total of 67% of the BPs can be explained by metadata records with a maximum of 100% at 700 hPa. Generally, the proportion of BPs consistent with the metadata decreases with height and implies that the reliability of all procedures decreases with the increasing height.

5. Consistency of BPs identified by different reference series

The consistency of the BP distributions identified by different procedures is an indirect measure of the uncertainty in the homogenization. Figure 4 compares the BP numbers that were consistent with each other for each year and reference series for the 92 Chinese stations for 1958–2005. As with the metadata consistency, the interconsistencies of the reference series BPs show two significant peaks around 1966 and 2000 in each case. This may imply that all three homogenization procedures are more reliably relationship-efficient than the metadata.

Figure 5 compares the BP percentages that were

Table 3. Percentage (%) of detected BP numbers consistent with the metadata at every level for the three references.

Level (hPa)	Nighttime	NCEP	ERA-40
100	49%	25%	24%
200	65%	28%	38%
300	56%	31%	31%
400	79%	32%	32%
500	72%	27%	40%
700	100%	33%	40%
850	88%	38%	37%
Total	67%	29%	33%

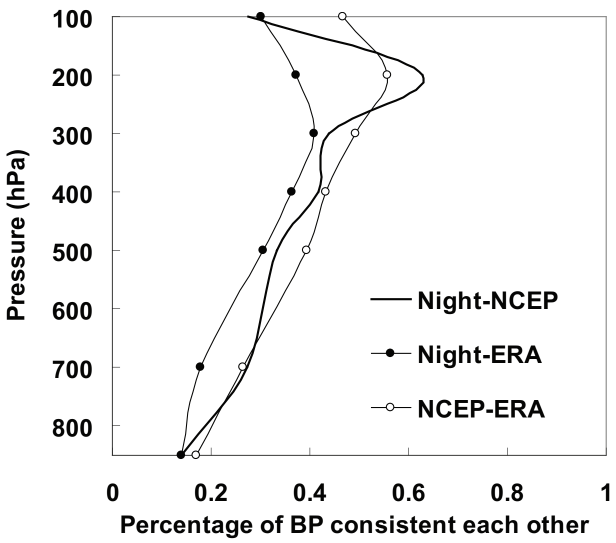

Fig. 5. Comparison of BP percentages that were consistently detected at each level by the three references.

consistently detected at each level by the three procedures. The proportion of BPs consistent with each other increased with increasing height, with a minimum of 14%–17% at 850 hPa. The three procedures commonly detected more significant BPs at higher levels and were more consistent with the metadata events at lower levels. Thus, the increased interconsistency of BPs detected by the three procedures may indirectly reveal greater uncertainties at the upper level.

6. Adjustment analysis

Figure 6 compares the distribution of adjustment values from the three homogenization procedures with different reference time series: nighttime, NCEP reanalysis, and ERA-40 reanalysis reference series. Adjustment of the nighttime reference series shows a single peak. The majority of adjustments (88%) were between −0.2 K and 0.4 K, and the frequency decreased dramatically outside this range. The frequency distribution of adjustments by NCEP and ERA-40 reanalysis reference series were similar with a near-symmetrical bimodal pattern, but the central values were different. The two dominant ranges of adjustment of NCEP were from −0.6 K to −0.2 K and from 0.4 K to 0.8 K, with both comprising a frequency of 30%. The adjustment of ERA-40 mainly ranged between −0.4 K and 0.0 K and between 0.6 K and 1.0 K, with frequencies of 37% and 34%.

The differences in adjustment for the three procedures can be seen from their statistical characteristic pressure profiles (Fig. 7). Only the mean adjustment from the ERA-40 reference series varied regularly with increasing height. The other two profiles showed significant vertical variability. The NCEP adjustment, in particular, exhibited contrast adjustment at 700 and 850 hPa. The homogenization using the nighttime reference series altered the original time series less than the other two procedures because the mean absolute adjustment from the nighttime reference series was clearly smaller than that from other two procedures at every level. Lanzante et al. (2003) noted that the bias of radiosonde temperature should increase with increasing height, with the profile showing a larger adjustment at higher levels. However, we found a minimum adjustment at 400 hPa. The variability of adjustment, as given by the mean standard deviation (MSD in Fig . 7), showed a similar pattern, with greater variability at upper and lower levels, but a modest variability in mid-levels.

7. Impact on the estimated trend

To examine the impact of the different homogenization reference series on the upper-atmospheric temperature, we compared the difference of average trend between the original series and those adjusted by the three homogenization procedures (ADJ-ORI) over China for both 1958–2005 and 1979–2005 in Fig. 8. All of the homogenizations, employing different reference series, altered the estimated trend for 1958–2005 according to a similar pattern, but with different strengths. The negative (positive) values at the lower

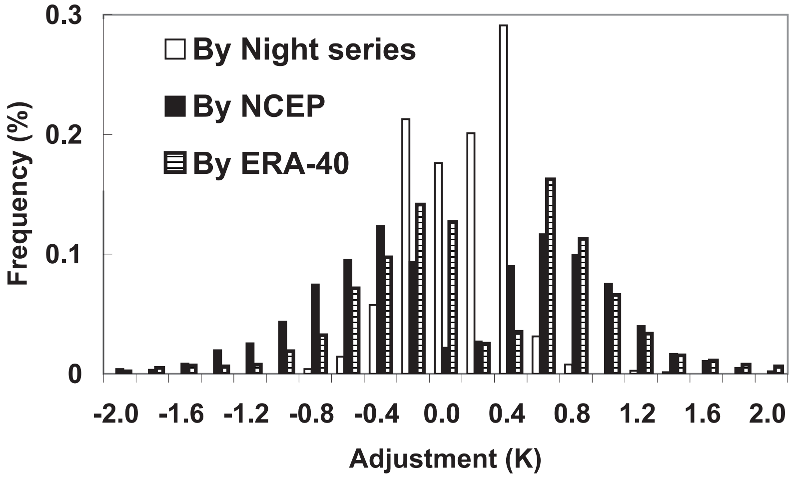

Fig. 6. Distribution of adjustment values for the three homogenizations procedures for 1958–2005.

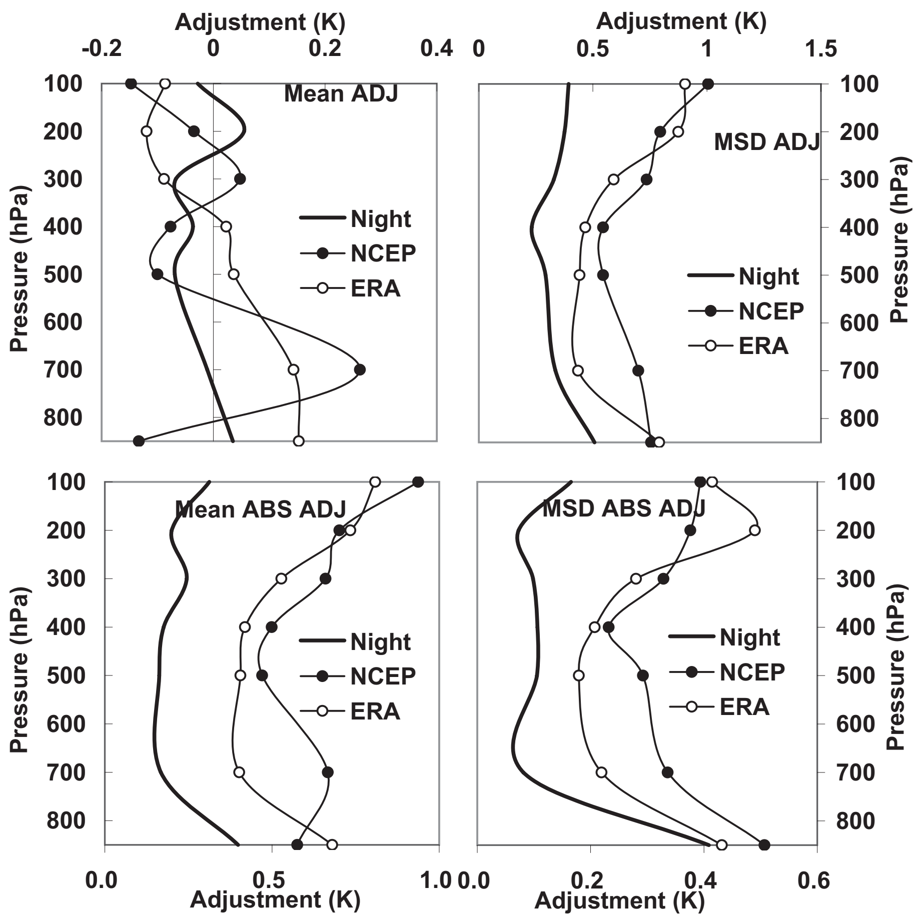

Fig. 7. Comparison of the statistical characteristics of adjustment with the increasing height for the three procedures for 1958–2005. The labels are ADJ: adjustment; MSD: Mean standard deviation; ABS: Absolute.

Table 4. The average trend over China for both 1958–2005 and 1979–2005 at the selected level of the original (ORI), reference series (REF), and difference between original and adjusted series for the three homogenization procedures (ADJ-ORI).

		Level (hPa)						
		850	700	500	400	300	200	100
1958–2005								
ORI		0.18	0.10	−0.02	−0.06	−0.24	−0.14	−0.20
REF	Nighttime	0.12	0.06	−0.05	−0.09	−0.23	−0.13	−0.15
	NCEP	0.08	−0.02	−0.10	−0.17	−0.27	−0.13	−0.01
	ERA(58–03)	0.02	−0.0	−0.11	−0.15	−0.22	0.01	−0.10
ADJ-ORI	Nighttime	−0.03	−0.02	−0.02	−0.01	0.02	0.01	0.04
	NCEP	0.03	−0.03	0.00	0.00	0.01	0.03	−0.16
	ERA(58–03)	−0.03	−0.01	0.00	0.03	0.08	0.18	0.13
1979–2005								
ORI		0.34	0.24	0.18	0.14	0.03	−0.23	−0.69
REF	Nighttime	0.33	0.24	0.19	0.16	−0.02	−0.26	−0.69
	NCEP	0.32	0.18	0.0	−0.05	−0.19	−0.43	−0.79
	ERA(58–03)	0.30	0.23	0.26	0.28	0.29	0.11	−0.24
ADJ-ORI	Nighttime	0.00	0.00	0.01	0.02	−0.04	−0.02	−0.01
	NCEP	0.04	−0.01	−0.08	−0.09	−0.14	−0.14	−0.04
	ERA(58–03)	0.04	0.05	0.10	0.18	0.27	0.32	0.31

(upper) level indicate that the homogenization weakened warming (cooling) in the lower (upper) atmosphere. However, the differences in ADJ-ORI corresponding to the nighttime reference series were rather smaller than those from the other two procedures.

Quantitative analysis of the impact of the different homogenization reference series on the upper-atmospheric temperature trend is shown in Table 4. The change in the trend due to homogenization by the nighttime series was general smaller than that by other procedures. Both the NCEP and ERA-40 homogenization procedures altered the estimated trend over China for 1958–2005 by almost the same amplitude at 200–850 hPa. They were different only at 100 hPa, where the cooling trend was enhanced by the ERA-40 reference but weakened by the NCEP reference.

Lower stratospheric cooling and tropospheric warming were more intense in the last two decades, proceeding at roughly twice the rate as the changes before 1980. Meanwhile, the impacts of homogenization on the estimated trend changed in accordance with the identified BPs (see sections 3 and 4). The nighttime reference series still altered the estimated trend by smaller amplitude, but with a different vertical pattern. In particular, it changed the trend relatively significantly around 300 hPa. The NCEP reference series altered the estimated trend in a similar vertical pattern, but with less strength than for 1958–2005. The ERA-40 reference series affected the estimated trend very significantly at levels higher than 200 hPa, with the maximum ADJ-ORI of 0.18 and 0.32 K $(10\ \mathrm{yr})^{-1}$ for 1958–2003 and 1979–2003, respectively.

8. Discussion and conclusion

The necessity of reference series in homogenization should indicate the importance of their reliability. However, the fact that the metadata in the Chinese radiosonde network occurred almost simultaneously meant that we could not use the series of neighboring stations. In this paper, we examined the suitability of three reference series for homogenization at 92 stations selected from the radiosonde network in China. The references were: nighttime, NCEP reanalysis, and ERA-40 reanalysis. Our results show limited consistency in the temporal and spatial distributions of identified BPs. The BPs detected using the three homogenization procedures were consistent with two correction method changes but less consistent with instrument mode changes. Furthermore, the BP ratios tended to vary similarly with the increasing height.

The most important findings of this study are the uncertainties in homogenization attributed to different reference time series. The reanalysis reference series generally lead to higher BP identification in homogenization. However, those differences were common to climatic shifts (i.e. volcanic eruptions and a strong ENSO event) that may have led to a temperature shift, confusing the BP. The NCEP reference series appeared to be less able to distinguish between BP and climatic shift because it identified more stations containing BPs in comparison with the ERA-40 series. This may be elucidated by different assimilations between the two reanalysis models. The satellite radiances used in the NCEP series were reanalyses of retrievals, but those in

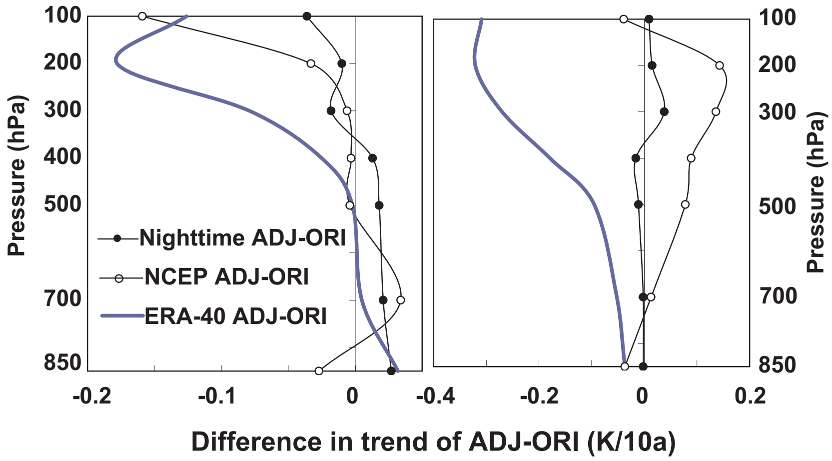

Fig. 8. Difference of the average trend between original trend (ORI) and their adjusted (ADJ) by the three homogenization procedures over China for 1958–2005 (left) and 1979–2005 (right).

the ERA-40 were raw. Retrieval data were used to estimate the vertical temperature and humidity profiles through a series of empirical and statistical relationships, whereas raw radiances are direct measurements of atmospheric radiation acquired by satellite sensors. Incorporating raw radiances requires more computational time and higher power, but it eliminates the errors associated with the retrieval process (Bromwich et al., 2007).

It is essential to be aware of the peak numbers of stations with BPs from the reanalysis homogenization. The introduction of satellite observations to the reanalysis model was most likely to cause peak BPs as shown in Fig. 2. Although the reanalysis were produced using a fixed model and analysis system, the input observations were not homogeneous because of historical changes in the observing system. Several studies have indicated the limitations of reanalysis products for global climate change studies. Kistler et al. (2001) showed that the strongest effect of the introduction of satellite observations in 1979 on the NCEP reanalysis occurred above 200 hPa. The discontinuities in the lower stratosphere and upper troposphere around 1979 were caused by biases from both the NCEP model and the National Environmental Satellite Data and Information Service (NESDIS) retrievals (Santer et al., 1999). Sturaro (2003) found that the spurious shifts in the temperature fields in 1979 were broader than earlier estimates. For ERA-40, Bengtsson et al. (2004) argued that the changes in the observing system had huge impacts on the global mean linear trends of temperature, integrated water vapor, and kinetic energy. Trenberth and Smith (2006) diagnosed a spurious break in the ERA-40 temperature analyses related to the assimilation of MSU-3 satellite radiances at the end of the NOAA-9 period.

Further differences in the impact of reference series on homogenization can be seen from their adjustments. The magnitude of adjustment values deduced from reanalysis references ranged widely, and the vertical variability of the adjustments was larger than that of the nighttime series. The larger adjustment logically led to a greater impact on the estimated trend, as discussed in section 7. Reanalysis reference series, generally, caused a larger difference between the estimated trends of the original and adjusted temperature series. However, the relationship between reference series and the estimated trend depends on the study period. For 1958–2005, the NCEP and ERA-40 reference series altered the trend in similar patterns and amplitudes. The ERA-40 data modified the trend for 1979–2003 very efficiently, with a maximum of 0.32 K $(10\ \mathrm{yr})^{-1}$. This was certainly caused by the switch from ERA-40 to the operational ECMWF data assimilation system, and some changes in the ERA-40 satellite observing system did, indeed, lead to breaks in the background forecasts (Christy and Norris, 2006; Haimberger, 2007). These problems were most likely the reason for the rather weak stratospheric cooling and strong upper-troposphere heating trends seen in the ERA-40 background forecasts compared to those seen in the radiosonde and NCEP data, which caused a larger difference in the trend for 1979–2003 (ADJ-ORI of ERA in Table 4).

Homogenization using the nighttime series as a reference requires calculation of the solar elevation angle difference between the two observation times (0000 UTC and 1200 UTC). This allows the effect of the radiosonde solar radiation correction to be reflected in the day–night difference series, even if the series for most of China at 1200 UTC corresponds to 1700–2000 LST. Because the solar angle varies with the season, it is likely that for at least part of the year, day–night difference series are useful at almost all locations. Generally, daytime observations are more likely than nighttime observations to have problems because of the effects of solar radiation and historical changes in methods to reduce the effects of solar radiation (i.e. shielding of the radiosonde, ventilation, and software corrections). Because the magnitude of solar radiation during the day is usually much greater than the magnitude of infrared radiation at night, the nighttime observations should have less error. In addition, solar radiation error corrections to the radiosonde are much larger (Sherwood et al., 2005). Many previous analyses have found that the day–night difference is the most useful tool for finding discontinuities in radiosonde temperature series (Gaffen, 1994; Luers and Eskridge, 1998; Lanzante et al., 2003). The results of this paper prove that homogenization could provide BP identification that is able to distinguish between a BP and a climatic shift. The impact of climatic shift to homogenization must be artificially eliminated when the NCEP series is applied as a reference (e.g. Guo and Ding, 2009). The effect of the nighttime reference series on the estimated trends appears too small to prove the necessity of homogenization.

Acknowledgements. We are grateful to Prof. HUANG Bingxun for helpful comments and Prof. Thomas C. Peterson for providing the FORTRAN code used in the two-phase regression method. This work was jointly supported by the National Basic Research Program of China (973 Program) under Grant No. 2010CB428606, special finance from the China Meteorological Administration (GYHY200906017), the National Natural Science Foundation of China under Grant No 40775045, and the Key Technologies R&D Program of China under Grant No. 2007BAC29B04.

APPENDIX A

Quality Control

For Quality Control (QC), we used the hydrostatic method, as first described by Collins and Gandin (1990). The basic form of the hydrostatic residual (S) for a layer between the two mandatory levels $l1$ and $l2$, each containing a height and temperature, is give as (Collins, 2000)

$$S^m_{l1,l2} = Z_{l2} - Z_{l1} - A_{l1,l2} - B_{l1,l2}(T_{l1} + T_{l2}) \,, \quad \text{(A1)}$$

where T is the virtual temperature in °C, and Z is the geopotential height. The coefficients A and B are given by

$$A_{l1,l2} = \frac{RT_0}{g}\ln\left(\frac{p_{l1}}{p_{l2}}\right) \quad \text{and} \quad B_{l1,l2} = \frac{R}{2g}\ln\left(\frac{p_{l1}}{p_{l2}}\right) \quad \text{(A2)}$$

where $T_0 = 273.15$ K, R is the gas constant for dry air, and g is the acceleration of gravity. For S, Zhai (1997) deduced a critical value for temperature time series.

APPENDIX B

Methods of Detecting Break Points and Adjustment

The two-phase regression (TPR) method, a technique initially described by Solow (1987), was used to detect break points. Easterling and Peterson (1995) developed a variation of the TPR in which the regression lines were not constrained to meet, and a linear regression was fitted to the part of the difference series before the point being tested and another part after the year being tested:

$$T(\text{i}) = T_{\text{CAN}} - T_{\text{REF}} \quad \text{and} \quad T(i) = \mu + \alpha i \,. \quad \text{(B1)}$$

T_{CAN} and T_{REF} represent candidate and reference series, respectively. The residual sum of squares from a single regression through the entire time series was also calculated with a critical value $U(i)$:

$$U(i) = \frac{(\text{RSS}_1 - \text{RSS}_2)/3}{\text{RSS}_2/(n-4)} \,, \quad \text{(B2)}$$

for which RSS_1 is the residual sum of squares for $i = 1, \ldots .c$, and RSS_2 is the residual sum for $i = c+1, \ldots .n$. The procedure tests the significance of a two-phase fit to each point in a series of differences from a reference series using Eq. (A1) a likelihood ratio statistic with the two residual sums of squares and Eq. (A2) the difference in the means of the difference series before and after the potential discontinuity as evaluated using a t-test (Mielke, 1991). If the break point was significant at the 5% level (probability: $P = 0.05$), it was considered a true discontinuity. The time series was subdivided into two at that year, and a break point was identified.

A homogenized array was created by

$$T(i) = \begin{cases} \mu_1 + \alpha_1 i \,, & 1 \leqslant i \leqslant c \\ \mu_2 + \alpha_2 i \,, & c < i \leqslant n \end{cases} \,, \quad \text{(B3)}$$

with α_1 and α_2 calculated by

$$\begin{aligned} \alpha_1 &= \frac{\sum\limits_{i=1}^{c}(i-\bar{i})(T_i-\overline{T})}{\sum\limits_{i=1}^{c}(i-\bar{i})^2} \,, \quad \mu_1 = (\overline{T} - \alpha_1\bar{i}) \quad \text{and} \\ \alpha_2 &= \frac{\sum\limits_{i=c+1}^{n}(i-\bar{i})(T_i-\overline{T})}{\sum\limits_{i=c+1}^{c}(i-\bar{i})^2} \,, \quad \mu_2 = (\overline{T} - \alpha_2\bar{i}) \,. \end{aligned} \quad \text{(B4)}$$

The adjustment that was applied to all data points prior to the discontinuity was the difference in the means of the two windows of the difference series. This test was repeated for all years of the time series up until the discontinuity became insignificant; then adjusted time series were gained. Identification was performed level-by-level at 7 levels of candidate stations. Several studies have proved the TPR method is suitable for use with Chinese radiosonde network (Zhai and Eskridge, 1996; Guo et al., 2008)

REFERENCES

Bengtsson, L., S. Hagemann, and K. I. Hodges, 2004: Can climate trends be calculated from reanalysis data? *J. Geophys. Res.*, **109**, D11111, doi: 10.1029/2004JD004536.

Bromwich, D. H., R. L. Fogt, K. I. Hodges, and J. E. Walsh, 2007: A tropospheric assessment of the ERA-40, NCEP, and JRA-25 global reanalyses in the polar regions. *J. Geophys. Res.*, **112**, D10111, doi: 10.1029/2006JD007859.

Christy, J. R., and W. B. Norris, 2006: Satellite and VIZ–radiosonde inter-comparisons for diagnosis of nonclimatic influences. *J. Atmos. Oceanic Technol.*, **23**, 1181–1194.

Collins, W. G., 2000: The operational complex quality control of radiosonde heights and temperatures at the national centers for environmental prediction. Part I: Description of the method. *J. Appl. Meteor. Climatol.*, **40**, 137–151.

Collins, W. G., and L. S. Gandin, 1990: Comprehensive hydrostatic quality control at the National Meteorological Center. *Mon. Wea. Rev.*, **118**, 752–2767.

Daan, H., 2002: GCOS-73: Guide to the GCOS surface and upper air networks: GSN and GUAN. World Meteorological Organization Tech. Doc. WMOTD-1106, Geneva, Switzerland, 37pp.

Easterling, D. R., and T. C. Peterson 1995: A new method for detecting undocumented discontinuities in climatological time series. *International Journal of Climatology*, **15**, 369–377.

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

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.

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

Guo, Y., 2008: Advances in uncertainties in upper-air temperature trends. *Advances in Earth Science*, **23**, 24–29. (in Chinese)

Guo, Y., P. W. Thorne, M. P. McCarthy, H. A. Titchner, B. Huang, P. Zhai, and Y. Ding, 2008: Radiosonde temperature trends and their uncertainties over eastern China. *International Journal of Climatology*, **28**, 1269–1281.

Guo, Y., and Y. Ding, 2009: Long-term free-atmosphere temperature trends in China derived from homogenized in situ radiosonde temperature series. *J. Climate*, **22**(4), 1037–1051.

Haimberger, L., 2007: Homogenization of radiosonde temperature time series using innovation statistics. *J. Climate*, **20**, 1377–1403.

IPCC, 2007: *Climate Change*, 2007: *The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC*, Solomon et al., Eds., 996pp. [Available at http://ipcc-wg1.ucar.edu/]

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, 2003: Temporal homogenization of monthly radiosonde temperature data. Part I: Methodology. *J. Climate*, **16**, 224–240.

Llanso, P., Ed., 2003: Guidelines on climate metadata and homogenization. WMO Tech. Doc. 1186, 51pp.

Luers, J. K., and R. E. Eskridge, 1998: Use of radiosonde temperature data in climate studies. *J. Climate*, **11**, 1002–1019.

Mielke, P. W., 1991: The appliacation of multivariate permutation methods based on distance functions in the earth sciences. *Earth-Science Reviews*, **31**, 51–71.

Peterson, T. C., and Coauthors, 1998: Homogeneity adjustments of in situ atmospheric climate data: A review. *International Journal of Climatology*, **18**, 1493–1517.

Reeves, J., J. Chen, X. L. Wang, R. Lund, and Q. Lu, 2007: A review and comparison of changepoint detection techniques for climate data. *J. Appl. Meteor.*, **46**, 900–915.

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

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. R. Lanzante, and C. Meyer, 2005: Radiosonde daytime biases and late 20th century warming. *Science*, **309**, 1556–1559, doi: 10.1126/science.1115640309.

Solow, A. R., 1987: Testing for climate change: An application of the two-phase regression model. *J. Appl. Meteor.*, **26**, 1401–1405.

Sturaro, G., 2003: A closer look at the climatological discontinuities present in the NCEP/NCAR reanalysis temperature due to the introduction of satellite data. *Climate Dyn.*, **21**, 309–316.

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.

Wang, X. L., V. R. Swail, and F. W. Zwiers, 2006: Climatology and changes of extratropical cyclone activity: Comparison of ERA-40 with NCEP–NCAR reanalysis for 1958–2001. *J. Climate*, **19**(13), 3145–3166, doi: 10.1175/JCLI3781.1.

Zhai, P. M., 1997: Some gross errors and biases in China's historical radiosonde data. *Acta Meteorologica Sinica*, **55**, 563–572. (in Chinese)

Zhai, P. M., and R. E. Eskridge, 1996: Analysis of inhomogeneities in Radiosonde temperature and humidity time series. *J. Climate*, **9**, 884–894.