

A New Approach to Homogenize Global Subdaily Radiosonde Temperature Data from 1958 to 2018

CHUNLÜE ZHOU,^a JUNHONG WANG,^{a,b} AIGUO DAI,^a AND PETER W. THORNE^c

^a *Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York, Albany, New York*

^b *New York State Mesonet, University at Albany, State University of New York, Albany, New York*

^c *Irish Climate Analysis and Research Units, Department of Geography, Maynooth University, Maynooth, Ireland*

(Manuscript received 14 May 2020, in final form 8 October 2020)

ABSTRACT: This study develops an innovative approach to homogenize discontinuities in both *mean* and *variance* in global subdaily radiosonde temperature data from 1958 to 2018. First, temperature natural variations and changes are estimated using reanalyses and removed from the radiosonde data to construct monthly and daily difference series. A penalized maximal *F* test and an improved Kolmogorov–Smirnov test are then applied to the monthly and daily difference series to detect spurious shifts in the mean and variance, respectively. About 60% (40%) of the changepoints appear in the mean (variance), and ~56% of them are confirmed by available metadata. The changepoints display a country-dependent pattern likely due to changes in national radiosonde networks. Mean segment length is 7.2 (14.6) years for the mean (variance)-based detection. A mean (quantile)-matching method using up to 5 years of data from two adjacent mean (variance)-based segments is used to adjust the earlier segments relative to the latest segment. The homogenized series is obtained by adding the two homogenized difference series back to the subtracted reference series. The homogenized data exhibit more spatially coherent trends and temporally consistent variations than the raw data, and lack the spurious tropospheric cooling over North China and Mongolia seen in several reanalyses and raw datasets. The homogenized data clearly show a warming maximum around 300 hPa over 30°S–30°N, consistent with model simulations, in contrast to the raw data. The results suggest that spurious changes are numerous and significant in the radiosonde records and our method can greatly improve their homogeneity.

KEYWORDS: Climate records; Data quality control; Radiosonde observations


1. Introduction

Since the 1950s, under the auspices of the World Meteorological Organization (WMO), balloon-borne radiosonde measurements have provided the only long-term and high-vertical-resolution subdaily records (twice daily for most stations) of temperature, humidity, and winds over the continents and many islands (Durre et al. 2018). These radiosonde data have been used to constrain weather forecasts (Benjamin et al. 2004; Lee et al. 2019; Naakka et al. 2019) and historical atmospheric reanalyses (Kalnay et al. 1996; Dee et al. 2011; Kobayashi et al. 2015; Hersbach et al. 2020), calibrate satellite data (Sun et al. 2017; Carminati et al. 2019), and study regional extremes (DeRubertis 2006; Waugh and Schuur 2018). In particular, radiosonde temperature data are crucial in quantifying and attributing atmospheric warming trends (Gaffen et al. 2000; Santer et al. 2005; Thorne et al. 2005; Karl et al. 2006; Sherwood et al. 2008; Fu et al. 2011; Thorne et al. 2011a,b; Santer et al. 2017), quantifying atmospheric humidity and water vapor trends (Dai et al. 2011; Zhao et al. 2012; Wang et al. 2016), and studying atmospheric instability and buoyancy changes (Chen et al. 2020).

However, these applications are severely hampered by spurious discontinuities or changes in the *mean* and/or *variance* of the radiosonde data arising from changes in instruments,

observational practices, manufacturer processing methods, and so on (Gaffen 1993; Thorne et al. 2005; Sherwood et al. 2008; Wang and Zhang 2008; Dai et al. 2011; Haimberger et al. 2012). For example, these artificial changes may have been propagated into many atmospheric reanalysis products by data assimilation systems, leading to unreliable long-term changes in these widely used products (Dai et al. 2011, 2013; Zhou et al. 2018). Even for reanalysis products during the satellite era since 1979 when radiosonde data account for only a fraction of the assimilated data, they could still have the single largest impact, as shown in the Modern-Era Retrospective Analysis for Research and Applications (MERRA; see https://gmao.gsfc.nasa.gov/forecasts/systems/fp/obs_impact/) (Rienecker et al. 2011); therefore, the discontinuities in radiosonde data can still degrade the quality of the reanalysis products during the satellite era.

The Fifth Assessment Report of the United Nations Intergovernmental Panel on Climate Change (IPCC AR5) (Hartmann et al. 2013) has also pointed out a medium to low confidence level in the detected long-term changes in tropospheric and stratospheric temperatures and their vertical structure, partly due to significant nonclimatic changes in the radiosonde data. Furthermore, atmospheric water vapor and humidity trends estimated based on radiosonde data depend critically on the quality of the temperature data (Dai et al. 2011; Zhao et al. 2012; Wang et al. 2016). Thus, reducing the discontinuities and the associated spurious changes in radiosonde temperature data is important for increasing our confidence in the detection and attribution of tropospheric and lower-stratospheric temperature

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Chunlüe Zhou, chunluezhou@gmail.com

DOI: 10.1175/JCLI-D-20-0352.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

and water vapor changes and for improving the quality of atmospheric reanalysis products.

Considerable efforts have been devoted by many groups to identify and remove spurious nonclimatic shifts in radiosonde monthly temperature data (Parker et al. 1997; Peterson et al. 1998; Lanzante et al. 2003; Free et al. 2005; Thorne et al. 2005; Haimberger 2007; Guo and Ding 2009; Thorne et al. 2011b; Haimberger et al. 2012; Chen and Yang 2014). Lanzante et al. (2003) and Free et al. (2005) adjusted artificial discontinuities in radiosonde monthly temperature series at 87 stations over the globe using metadata to create the Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC) dataset. However, many spurious shifts in radiosonde and other climate records are not documented by the notoriously incomplete metadata (i.e., station history logs); hence statistical methods have been developed to detect and remove spurious changes often in monthly mean time series for temperature and other variables (Reeves et al. 2007; Thorne et al. 2011b).

Most existing automated homogenization methods are for identifying and adjusting spurious shifts in the *mean* of a candidate time series (often for monthly data) by comparison with a reference series from different sources. For example, Parker et al. (1997) corrected discontinuities of monthly temperature by comparing with the collocated satellite-based Microwave Sounding Unit (MSU) temperature series since 1979 for constructing the Hadley Centre Radiosonde Temperature (HadRT) dataset. Thorne et al. (2005) used a composite of neighboring soundings as a reference series for developing the homogenized HadAT dataset (as an update to HadRT). Sherwood et al. (2008) utilized nighttime temperature as a reference series to correct the systematic errors in daytime temperature due to solar heating on instruments over regions where daytime and nighttime observations are available (only covering $\sim 1/3$ of the archived stations) for building the IUK (Iterative Universal Kriging) dataset. Haimberger (2007) applied a variant of the Standard Normal Homogeneity Test (SNHT) with ERA-40 [the European Centre for Medium-Range Weather Forecasts (ECMWF) 40-Year Re-Analysis] forecast data (Uppala et al. 2005) as a reference to establish the Radiosonde Observation Correction Using Reanalyses (RAOBCORE) dataset. Haimberger et al. (2012) further refined this approach by using the breakpoints identified to select apparently homogeneous segments from neighboring stations as a reference to build the Radiosonde Innovation Composite Homogenization (RICH) dataset, which has been assimilated into various reanalyses, including the ECMWF interim reanalysis (ERA-Interim; Dee et al. 2011), MERRA (Rienecker et al. 2011), and the 55-year Japanese Reanalysis (JRA-55; Kobayashi et al. 2015). Additionally, Guo and Ding (2009) and Chen and Yang (2014) homogenized radiosonde monthly temperature data over China using the reanalysis temperature series from NCEP-R1 (the National Centers for Environmental Prediction Reanalysis 1) (Kalnay et al. 1996) and ERA-40 (Uppala et al. 2005) as a reference series.

These methods based on comparison with a reference series have been questioned, especially due to the issues related to the selection of a reference series, including too short or no data overlap, sparse nearby stations, similar shifts in adjacent

stations, inhomogeneity in reanalysis data, and so on (Della-Marta and Wanner 2006; Dai et al. 2011; Zhou et al. 2018). For example, the ERA-40 initial and thus forecast data may still contain the systematic biases in radiosonde data, and their use as the reference series will thus potentially propagate these biases into the homogenized product. This appears to be the case for the spurious tropospheric cooling trend centered over North China and Mongolia that occurred mainly from 1950s to the early 1970s (Zhou and Zhang 2009) that is seen in the raw radiosonde data and many reanalysis products (Dai et al. 2013). Using the average of neighboring stations as the reference series is also problematic because radiosonde stations are often sparse and tend to have similar spurious shifts due to simultaneous changes in national networks, leading to difficulties in generating or discontinuities in such a reference series, particularly in large countries where most neighboring stations may suffer from similar data artifacts.

Homogenization efforts to date have only adjusted the discontinuities in the *mean* of monthly upper air temperature [i.e., the first-order moment of a probability density function (PDF)], with no attempts to adjust the discontinuities in the *variance* of the subdaily temperature data (i.e., the high-order moment), mainly due to the difficulties in detecting and adjusting spurious shifts in daily data that have large synoptic and local variability but low spatial correlations. Individual sounding reports are influenced by not only synoptic-scale fluctuations but also local processes that are complex and nonlinear, resulting in a relatively small decorrelation distance, especially for lower levels over topographically complex regions. Because of this, how to separate and remove the large natural variations from the artificial changes in subdaily temperature series is difficult but crucial for homogenizing the subdaily radiosonde data.

On the other hand, reliable daily or subdaily temperature data are needed for studying weather and climate extremes (Zhou and Wang 2016b; Zhou et al. 2019; Sippel et al. 2020) that greatly impact natural and social systems. Homogenized subdaily upper air temperature data are also needed for input into reanalysis assimilation systems and for calculating water vapor variables (Dai et al. 2011). The lack of reliable tropospheric temperature data also prevents us from effectively validating the enhanced warming over the tropical mid- to upper troposphere projected by climate models under increased greenhouse gases (GHGs) (Santer et al. 2005; Karl et al. 2006; Fu et al. 2011; Mitchell et al. 2013; Santer et al. 2017). Therefore, it has become increasingly urgent to develop an automated homogenization approach for building a homogenized subdaily upper air temperature dataset that contains minimal discontinuities in both the mean and variance, as noted previously in chapter 2.4 of IPCC AR5 (Hartmann et al. 2013).

Dai et al. (2011) made the first attempt to homogenize subdaily radiosonde humidity [i.e., dewpoint depression (DPD)] data from the 1950s to 2009 over the globe, by detecting changepoints in DPD's occurrence frequency and PDFs to remove spurious shifts in the subdaily time series. Unlike DPD, temperature is nonstationary and more variable, which makes its homogenization more difficult, although its sampling is likely more homogeneous than DPD (Dai et al.

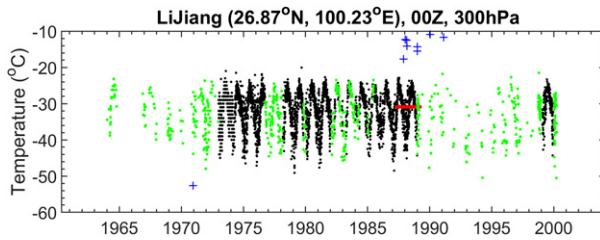


FIG. 1. Individual radiosonde temperature reports at 0000 UTC and 300 hPa at station Lijiang, China, from the merged dataset. Outliers (outside the ± 5 standard deviation range; blue plus signs) and duplicates (consecutive red dots) were removed. Some data points (green dots) were also excluded in our analysis due to insufficient monthly sampling (see the text for details). Black dots represent subdaily raw temperatures retained in our subsequent analysis.

2011). Building on Dai et al. (2011), we developed a new approach to create a homogenized global radiosonde dataset with subdaily temperatures by separately detecting and removing the major discontinuities in the mean and variance, which is referred to as the University at Albany Homogenized Radiosonde Subdaily dataset (UA-HRD).

To reach this goal, we first compiled all the available radiosonde temperature data and conducted initial quality controls, as described in section 2. In section 3, we described a four-step approach to detect and adjust discontinuities in the mean and variance of subdaily temperature records. We examined the correlations and homogeneities of two types of reanalysis products to estimate natural temperature variations and changes and then remove them from the raw radiosonde data to construct monthly and daily temperature difference series, as described in section 3a. In section 3b, the two types of changepoints in the mean and variance were analyzed and compared with available metadata. In section 3c, we described the adjustment methods to remove the discontinuities. The long-term trend and variance of the homogenized temperature data were analyzed and compared with those from the raw data in section 4. A summary is given in section 5.

2. Radiosonde data and preprocessing

Three radiosonde temperature datasets were first compiled in this study, including the Integrated Global Radiosonde Archive version 2 (IGRA2) built from 33 different data sources (Durre et al. 2018) and two distinct ERA-assimilated radiosonde datasets (Pralungo et al. 2014) used in ERA-40 and ERA-Interim, respectively.

The outliers and duplicates in each data source were detected and removed. One example is shown in Fig. 1 using the 0000 UTC reports at 300 hPa from the Lijiang station, China. The outliers are defined as data points outside the climatological mean plus and minus five standard deviations (SD) range (blue pluses in Fig. 1) calculated using the whole anomaly data series for each pressure level and observation time. The duplicates are five or more consecutive identical values separated by larger than 30 elapsed days (red dots in

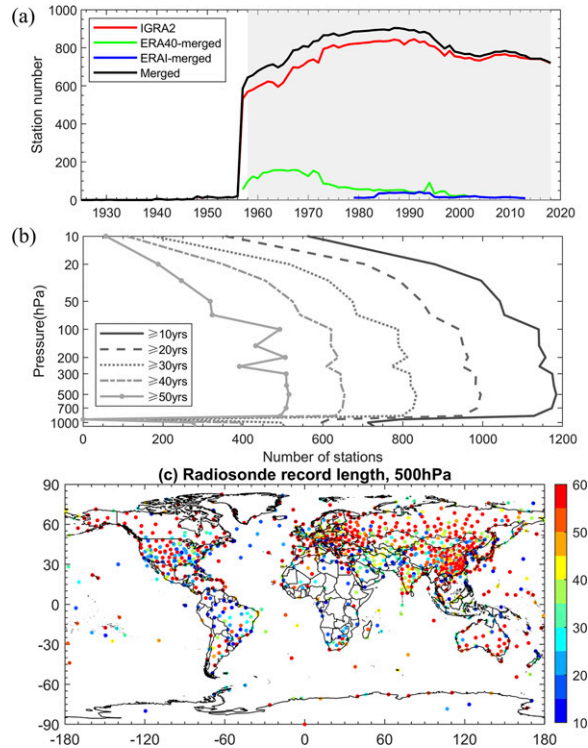


FIG. 2. (a) Time series of the number of radiosonde stations with 10 or more years of data from IGRA2 and ERA-40- and ERAI-assimilated radiosonde datasets, and from our merged dataset. The gray background shows the study period from 1958 to 2018. (b) Vertical profiles of the number of radiosonde stations with n or more years ($n = 10, 20, 30, 40, 50,$ and 60) of data during 1958–2018. (c) Radiosonde record length (in years) at 500 hPa, the level that has the most data. Note that station start and cessation dates are highly variable.

Fig. 1). Less than 0.1% of the data were removed using these two checks.

These quality-controlled data were then merged with preference given to IGRA2 to create a comprehensive, global 0000 and 1200 UTC radiosonde temperature dataset at the surface and 16 standard levels, namely 1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, and 10 hPa. Merging was achieved through (i) matching station identifiers or names, (ii) assessing data similarity (i.e., over 50% of the overlapping data having a temperature difference of $<0.2^{\circ}\text{C}$ at the matching stations within 40 km; Durre et al. 2006, 2018), and (iii) including the remaining unique stations in the final dataset.

To calculate monthly anomalies and make the homogenization robust, we removed any month with fewer than 10 days with valid data, and any year with fewer than 3 months with data (i.e., at least 10 days having reports within the month) from our analysis, as shown by green dots in Fig. 1. This removed about 2% of the merged data holdings. As a result, a total of 1184 stations with 10 or more years of data on one or more pressure levels are included in this study. ERA-assimilated radiosonde data sources contribute 59 new stations and about 9% of the merged data (Figs. 2a,b). There

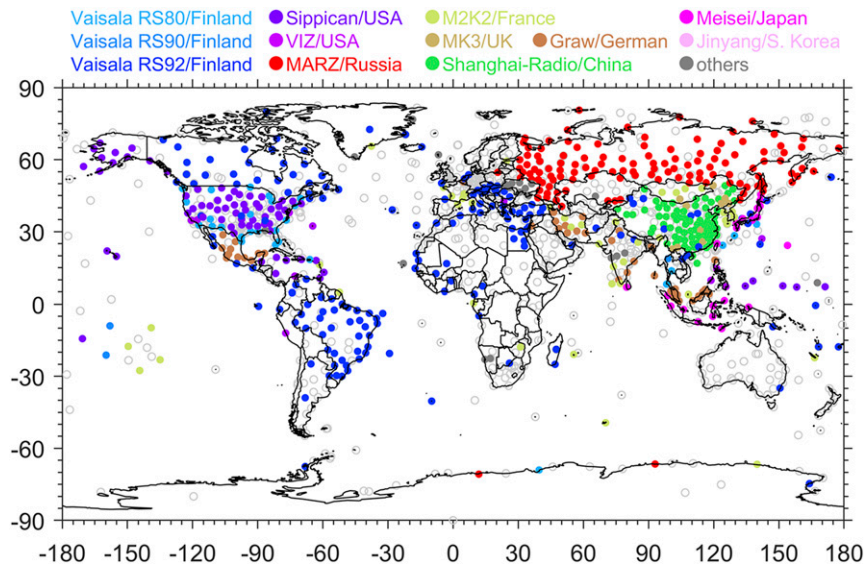


FIG. 3. Global distribution of the 1184 radiosonde stations included in the merged dataset colored by the 12 radiosonde types used by different countries during the period 2007–12. Gray circles are for radiosondes with unknown types during the period. Changes in sonde types through time have been ubiquitous at almost all radiosonde stations in different countries.

are very few (<50) stations before 1957; thereafter, the number of stations increased steadily from ~650 around 1958 to ~900 in the 1980s and then declined in the 1990s and remained around 750 in recent years (Fig. 2a). Thus, our analysis focuses on 1958–2018, during which 1184 stations have 10 or more years of data for at least one pressure level (Fig. 2a). Most of the stations have sufficient long-term data from 1958 to 2018 except at the upper levels (Fig. 2b) or over South America, Africa, and the Middle East (Fig. 2c) and the data at 925 hPa are available generally only from about 1992 to 2018 and only from low elevation stations. Figure 3 shows the radiosonde types obtained from the radiosonde metadata from the IGRA2 archive (<https://www1.ncdc.noaa.gov/pub/data/igra/history/>) for the period 2007–12. At least 12 different types of radiosondes were used at that time, often distinguished by countries (Fig. 3). However, it should be stressed that radiosonde model usage has varied enormously through time at all radiosonde stations (Thorne et al. 2011b).

The temperature data series at 0000 and 1200 UTC for each of the surface (often measured by surface station instruments, not by radiosonde sensors) and 16 standard pressure levels were analyzed separately. They were converted into daily anomalies by removing the 1958–2018 mean for each day; these anomalies were used in subsequent analyses described in section 3.

3. Homogenization method

Previous homogenization studies (Wang 2008; Dai et al. 2011; Haimberger et al. 2012; Zhou et al. 2018) have demonstrated the critical importance of a reliable reference series for detecting and adjusting discontinuities in a time series. This

is because a good reference series can remove most of the real climate changes and synoptic variations (the noise) in a time series, and thus enhance the signal (the spurious shifts) to noise ratio and make it possible for statistical detection and removal of the spurious shifts. We examined various datasets, and determined that the NOAA-20CRv3 (the Twentieth Century Reanalysis version 3 produced by National Oceanic and Atmospheric Administration) monthly temperature series (Slivinski et al. 2019) and the JRA-55 high-frequency variations of daily temperatures can be used as reliable reference series to remove natural monthly and daily variations from the radiosonde data over 1958–2018 (see section 3a for details). The decision was based on characteristics of available reanalysis products (spatial and temporal coverage and their performance documented by prior studies) and several tests applied to them, including their homogeneity and correlations with radiosonde data. These reference series were used to create the monthly and daily difference series for detecting and adjusting the spurious changes. They contain most of the natural variations and long-term changes in the raw radiosonde data, thus allowing the homogenized series to preserve many physical variations and changes, such as those associated with large-scale circulation/weather patterns.

We developed a four-step approach (Fig. 4) to detect and adjust spurious shifts in the mean and variance of subdaily radiosonde temperature data for each observation time and at each pressure level as follows. An example is given for the Lindenberg station in Germany in Fig. 5. More details on each step are provided below. Note that the same procedures were applied separately to the surface-level temperature series as it may contain different

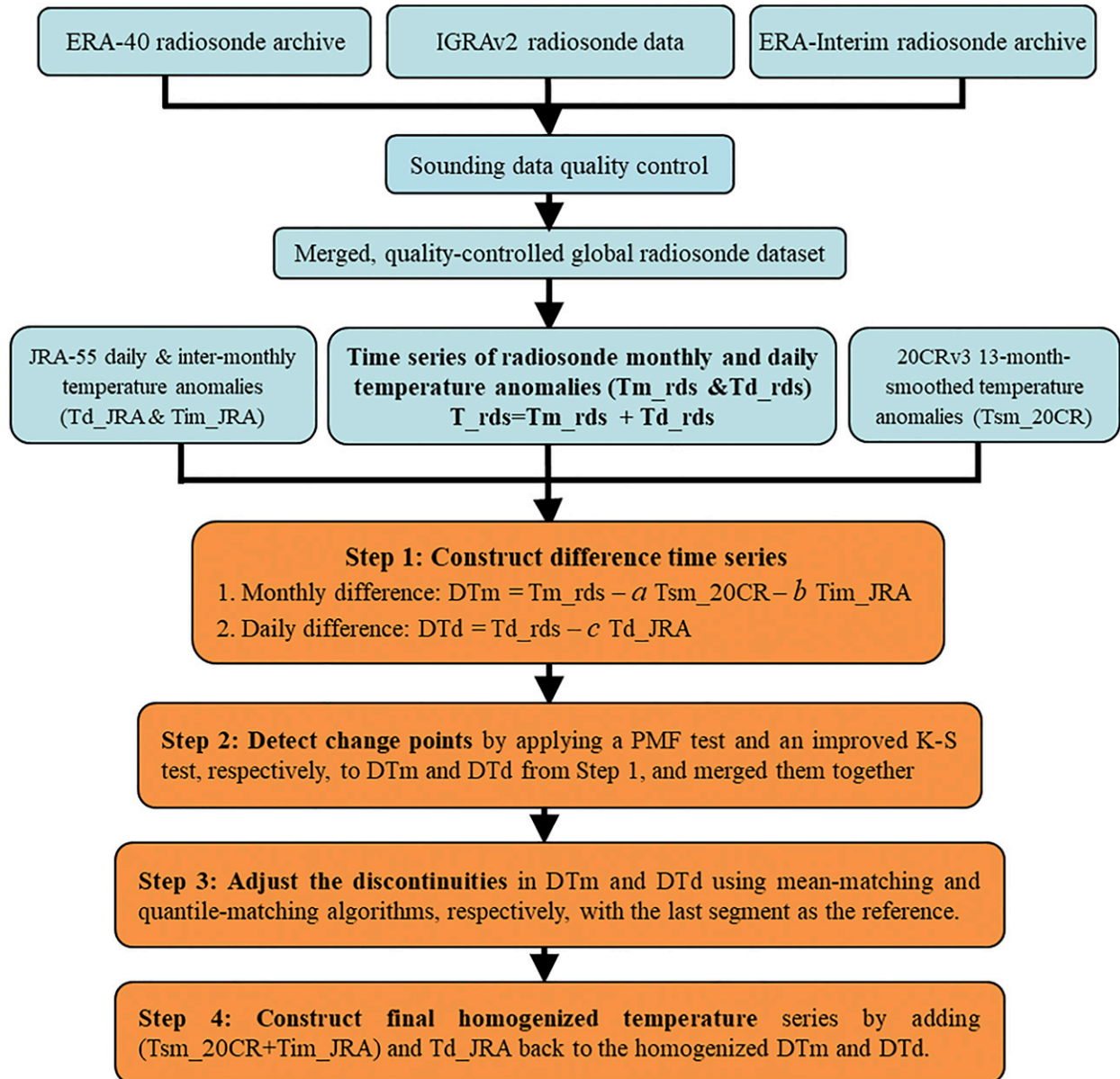


FIG. 4. Schematic diagram showing our radiosonde data processing and the four-step homogenization method. Note that the subdaily data for each level and observation time were converted into anomalies by removing their long-term (1958–2018) mean for each day before all subsequent analyses. “DT” represents temperature difference series, subscripts “m” and “d” denote, respectively, monthly mean and daily value; “_rds,” “_20CR,” and “_JRA” denote, respectively, radiosonde, NOAA-20CRv3, and JRA-55; and subscripts “sm” and “im” denote, respectively, 13-month moving-averaged data and the monthly data with the 13-month moving average being removed (referred to as intermonthly variations). Also, a is the linear regression coefficient between Tsm_20CR and Tsm_rds , b is between Tim_JRA and Tim_rds , and c is between Td_JRA and Td_rds .

change points, except that it only has one level and the reanalysis data at the lowest model level are used in Eqs. (1) and (2) for this case.

Step 1: Construction of two difference series

The first temperature difference series (DTm) is for detection and removal of the shifts in the *mean*; it is constructed as follows:

$$DTm = Tm_rds - a Tsm_20CR - b Tim_JRA, \quad (1)$$

where Tm_rds is the raw radiosonde monthly mean temperature anomaly relative to the 1958–2018 mean; Tsm_20CR is the 13-month moving-averaged NOAA-20CRv3 monthly temperature anomaly; and Tim_JRA is the JRA-55 monthly temperature anomaly with its 13-month moving average being removed (referred to as intermonthly variation). In Eq. (1), a

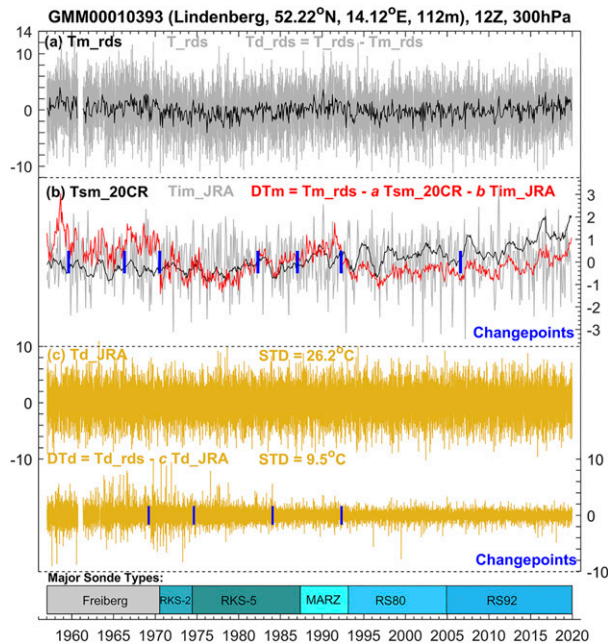


FIG. 5. An example to illustrate the steps to homogenize the monthly and daily radiosonde temperature anomalies (Tm_rds and Td_rds in $^{\circ}C$) (a) at 1200 UTC at 300 hPa at Lindenberg, Germany. (b),(c) Monthly and daily difference series (DTm and DTd), respectively. DTm is constructed by Tm_rds minus 13-month moving-averaged anomalies from NOAA-20CRv3 (Tsm_20CR) and intermonthly variations from JRA-55 (Tim_JRA). DTd is the difference between Td_rds and JRA-55 daily anomalies (Td_JRA). See the text for more details. Blue vertical lines indicate the detected changepoints. The radiosonde types are shown by colored rectangles at the bottom.

(b) is the linear regression coefficient between the similarly filtered radiosonde data and Tsm_20CR (Tim_JRA), that is, with the regression done separately as the variations are on different time scales. Tsm_20CR does not contain radiosonde-induced spurious changes since NOAA-20CRv3 did not assimilate radiosonde data. It should be able to describe long- and short-term climatic variations and changes, since NOAA-20CRv3 is forced with analyzed sea surface temperature (SST) and historical GHG and aerosol forcing. In section 3 we provided some assessments on the extent to which this is indeed the case. Tim_JRA should not include the spurious jumps caused by JRA-55's use of the radiosonde data because such jumps often occur on longer than 13-month time scales, as shown in section 3a below. We used Tim_JRA , instead of the intermonthly variations in NOAA-20CRv3, because it has stronger correlations with Tm_rds (i.e., the intermonthly variations in radiosonde data) than NOAA-20CRv3 does [see section 3a(2) below]. Thus, we used Tsm_20CR and Tim_JRA , respectively, to represent and set aside real long-term climatic changes and intermonthly natural variations in Tm_rds , and remove them from Tm_rds for constructing the difference series (Fig. 5). Both reference series are subject to homogeneity tests [see section 3a(1) below] and correlation evaluations [see

section 3a(2) below] before being used here. Note that since NOAA-20CRv3 temperature data end in 2015, we used JRA-55 temperature data to extend the NOAA-20CRv3 series to cover 2016–18 after adjusting the JRA-55 data over 2016–18 by the mean offset (JRA-55 minus NOAA-20CRv3) over 2011–15.

The second difference series (DTd) is for detection and removal of the shifts in the intramonthly variance of the daily time series. It is defined as the residual of the regression of Td_rds on Td_JRA for each pressure level and observation time:

$$DTd = Td_rds - c Td_JRA, \quad (2)$$

where Td_rds ($=T_rds - Tm_rds$) represents representing the intramonthly component of the raw radiosonde anomalies (T_rds); Td_JRA ($=T_JRA - Tm_JRA$) represents the intramonthly component of the JRA-55 data (T_JRA); and Tm_JRA is the JRA-55 monthly mean temperature anomaly. In Eq. (2), c is the linear regression coefficient between Td_rds and Td_JRA . Td_JRA was tested and confirmed to be homogeneous with strong correlations with Td_rds before being used here (see section 3a below).

Step 2: Detection of changepoints in the mean and variance

We applied the Penalized Maximal F (PMF) test developed by Wang (2008) at a significance level of 0.05 to the DTm series for each of surface, pressure levels, and observation times to detect spurious changepoints in the mean of the radiosonde temperature data. Similarly, we applied an improved variant of the Kolmogorov–Smirnov (K-S) test at a significance level of 0.001 from Dai et al. (2011) to the DTd series to detect spurious changepoints in the variance of the temperature data. The critical values of the K-S test statistic were estimated in appendix A. An example of the test results is shown in Figs. 5b and 5c. Using the significance level of 0.05 for the PMF test and the significance level of 0.001 for the K-S test can detect comparable and reasonable numbers of changepoints (see section 3b).

Noise in the time series combined with uncertainties of the tests mean that it is unlikely that a single changepoint would be uniquely identified on the same date by the two tests at different levels. To avoid identifying excessive changepoints, after several tests the detected changepoints at all the pressure levels from both the PMF and K-S tests were merged as follows: all changepoints within 180 days were grouped together and only the one in the middle of the group (if there are three or more changepoints) or the one with the larger test statistic (for two changepoints) was kept for each group. This means that our final changepoints will be 180 or more days apart.

Step 3: Adjustment of spurious changes

The mean-matching (MM) and quantile-matching (QM) algorithms from Wang et al. (2010) were applied to the DTm and DTd series to adjust their spurious discontinuities for each level and observation time. Up to five years of data from the segments before and after each merged changepoint (from step 2) were used to adjust the discontinuities in DTm and DTd with the last segment as the baseline (Fig. 4).

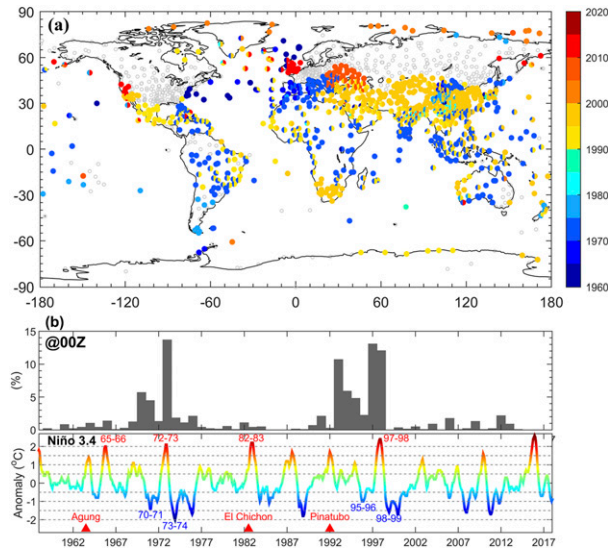


FIG. 6. (a) Map of the years with the detected change points in the NOAA-20CRv3 monthly temperature series at 0000 UTC at the grid boxes collocated with radiosonde stations based on the PMF test at all levels. Gray circles show no detectable change points and many stations have only one change point. (b) (top) Histograms (gray bars) of the years with the detected change points shown in (a) and (bottom) the time series of the Niño-3.4 index (colored curve) with the strong El Niño years labeled with red numbers and the strong La Niña years labeled with blue numbers. The three significant volcanic eruptions are also marked as Agung, El Chichón, and Pinatubo.

Step 4: Creation of the homogenized temperature series

The two homogenized difference series (DTm and DTd) from step 3 were added back onto the set-aside components ($aTsm_{20CR} + bTim_{JRA}$, and cTd_{JRA} from step 1) to obtain homogenized temperature anomaly series (Fig. 4).

a. Construction of monthly and daily temperature difference series

To construct the reference series used in Eqs. (1) and (2) to remove real physical variations and changes in the data, the reference series are required to have data back to 1958, be homogeneous with little or no impact from the radiosonde inhomogeneity, and perform well in depicting real climate and weather signals. After detailed analyses of several available atmospheric reanalyses, NOAA-20CRv3 and JRA-55 reanalysis products were selected for constructing the reference series based on their homogeneities [section 3a(1) below] and strong correlations with the radiosonde data [section 3a(2) below] at the monthly and daily time scales.

NOAA-20CRv3 was recently produced using an updated coupled atmosphere–land model of the NCEP Climate Forecast System (CFS v14.0.1) via an ensemble Kalman filter and four-dimensional incremental analysis (Slivinski et al. 2019). NOAA-20CRv3 has 3-hourly outputs on a horizontal T254 grid (equivalent to 60-km spacing at the equator) and 64 vertical levels up to 0.3 hPa and 80 ensemble members. The

model is constrained by several key climate forcings [historical time-varying CO_2 concentrations, volcanic aerosols, ozone concentrations, and solar variations as in phase 5 of the Coupled Model Intercomparison Project 5 (CMIP5)], observed SSTs, sea ice, and sea level pressure (SLP) but without assimilating the radiosonde or satellite observations (Slivinski et al. 2019). Thus, NOAA-20CRv3 temperatures should not be affected by nonclimatic shifts in radiosonde data and thus could be used to represent and set aside most natural variations and long-term changes that are associated with SSTs and SLP. Previous studies have shown that a predecessor (NOAA-20CRv2c) of NOAA-20CRv3 is capable of simulating long-term changes in near-surface monthly temperature (Parker 2011; Zhou et al. 2018), and NOAA-20CRv3 is greatly improved over its predecessor in many aspects including the radiative effects of ozone, volcanic aerosols, and solar variations (Slivinski et al. 2019).

In contrast, many other reanalysis products assimilate radiosonde temperature data (Zhou et al. 2018) and other upper-air observations, such as NCEP-R1 (Kalnay et al. 1996) and JRA-55 (Kobayashi et al. 2015), and thus they most likely inherit some of the discontinuities in the radiosonde data. For example, NCEP-R1 and JRA-55 show several apparent jumps in the late 1950s, 1970s, and 2000s at 300 hPa at Kuqa station (China) (Fig. B1), when the observation system was upgraded and/or the radiation correction method was changed in China (Guo et al. 2016). These spurious changes are evident in both the NCEP-R1 and JRA-55 but not in NOAA-20CRv3 (Fig. B1). As a result, JRA-55 displays spurious cooling trends over most of Asia at 300 hPa during the period 1958–2018, when NOAA-20CRv3 shows warming (Fig. B2). These spurious changes are reflected primarily in the monthly-mean temperature series, which can be removed from the JRA-55 daily temperature series, with the residual (i.e., Td-JRA) representing the high-frequency component (i.e., intramonthly variation) that can be used to remove synoptic variations in radiosonde temperature series.

1) ASSESSMENT OF HOMOGENEITY OF THE REANALYSIS REFERENCE SERIES

Homogeneity of the NOAA-20CRv3 monthly temperature series at each grid box collocated with radiosonde stations was first assessed via the PMF test at a significance level of 0.01. The significance level of 0.01 has been adopted to detect reasonable change points in the raw monthly series in many previous studies due to large variability in the raw series compared with a difference series (Wang 2008; Zhou et al. 2017, 2018). The years with a detected change point (after merging the detected change points from all the levels) are shown in Fig. 6a. It is found that most locations have one or two detectable change points (Fig. 6a) and these change points are concentrated around 1972, 1982, and 1998 around which either a strong El Niño or La Niña event occurred (Fig. 6b). Spatial patterns of the years with the detected change points (Fig. 6a) are very similar to ENSO's well-documented effect on near-surface temperature (Davey et al. 2014). Meanwhile, volcanic eruptions often cause sudden changes in tropospheric and stratospheric temperatures that last for several years (Santer

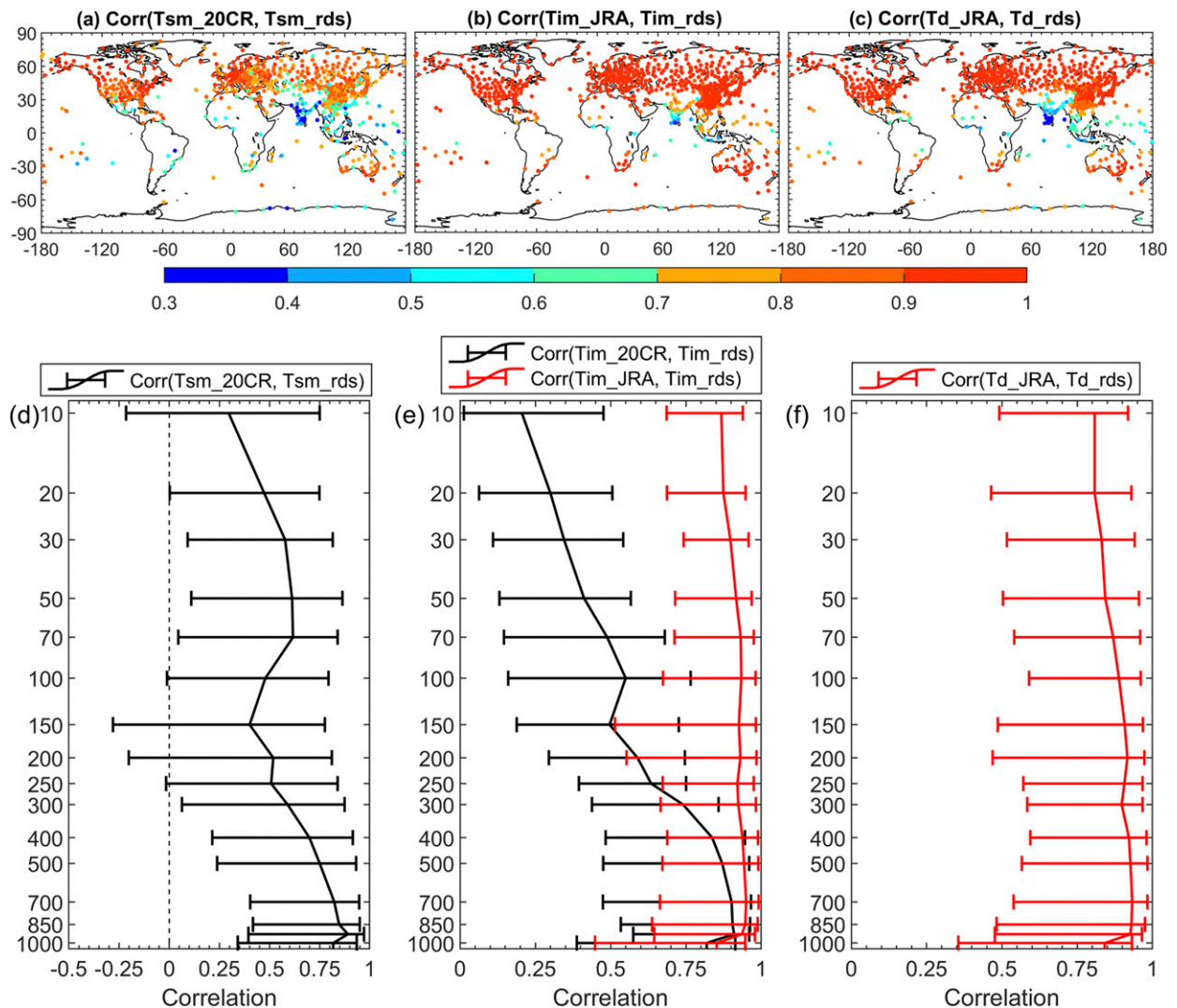


FIG. 7. (a) Correlation coefficients between the collocated 13-month moving-averaged temperature anomalies from NOAA-20CRv3 (Tsm_20CR) and radiosonde (Tsm_rds) datasets at 0000 UTC and 700 hPa from 1958 to 2018. (b) As in (a), but using the monthly data from JRA-55 (Tim_JRA) or NOAA 20CRv3 (Tim_20CR) and radiosonde (Tim_rds) datasets, with the 13-month moving average being removed. (c) As in (a), but using daily temperature anomalies from JRA-55 (Td_JRA) and radiosonde (Td_rds) datasets, with the monthly mean being removed from daily anomalies. (d)–(f) The corresponding vertical profiles of the globally averaged correlation coefficients, with the line representing the median and the error bar showing the 5%–95% spatial ranges.

et al. 2014). These changes are also detected, particularly for the Pinatubo eruption in 1991 (Fig. 6b). However, the detection of the Agung (in 1963) and El Chichón (in 1982) eruptions is less successful (Fig. 6b), mainly due to the simulated weak signal of these volcanic eruptions or substantial concurrent variability in upper air temperatures. When we looked at temperature series at some typical stations, we can obviously see the signals of ENSO events and volcanic eruptions. Therefore, these detected changepoints are mainly due to real abrupt climatic changes and NOAA-20CRv3 monthly temperature data since 1958 contain no obviously spurious sudden shifts attributable to artificial causes, although this may not be true for earlier years, other variables, or at other time scales. Note that

intermonthly variations of JRA-55 temperature (i.e., Tim_JRA) were also tested using the PMF test at a significance level of 0.01 and found to be homogeneous. As such, the Tsm-20CR and Tim_JRA in Eq. (1) can be used to remove these and other natural variations in radiosonde monthly temperature series. This will substantially reduce the probability of detecting the changepoints associated with real abrupt climatic changes.

Homogeneity of the Td_JRA series at each level and each collocated location was also examined using the improved K-S test at a significance level of 0.001 adopted from Dai et al. (2011). No significant changepoints could be detected. Note that JRA-55 daily temperature anomaly after the removal of its 13-month moving average was also tested to be homogeneous.

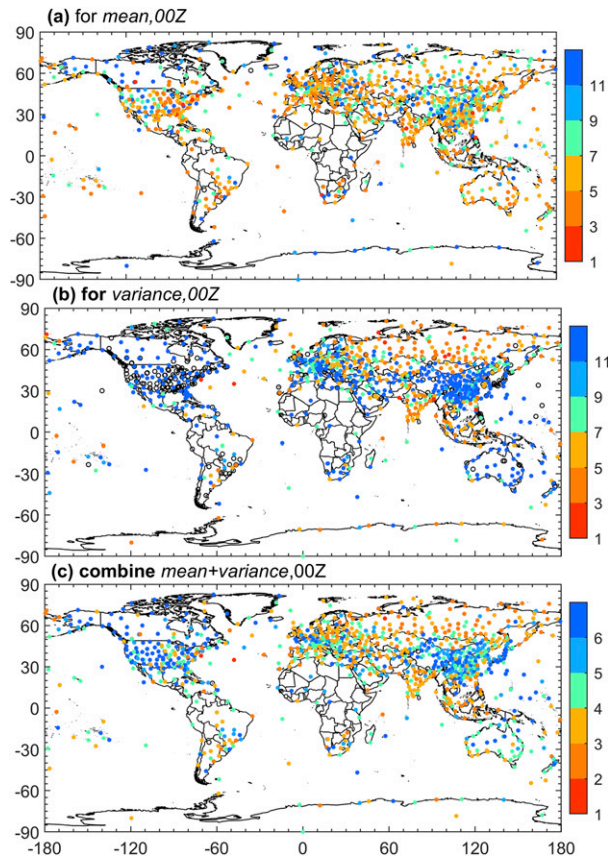


FIG. 8. Mean segment length (in years) separated by the detected changepoints for 0000 UTC at the 1184 stations based on the shifts in the (a) mean and (b) variance, and (c) their combination. Black circles show no detected changepoints at the stations.

In summary, we found that the Tsm_{20CR} and Tim_{JRA} in Eq. (1), and Td_{JRA} in Eq. (2) are likely to be homogeneous and thus may be used as reference time series.

2) VALIDITY OF USING REANALYSIS AS REFERENCE SERIES

A good reference series should not only contain no spurious shifts, but also be correlated strongly with the data series to be homogenized. Figure 7 shows correlation coefficients between the collocated JRA-55 or NOAA-20CRv3 and radiosonde temperature variations on time scales of 13 months and longer, 1–13 months, and daily time scales at 700 hPa (Figs. 7a–c) and their averaged vertical profiles (Figs. 7d–f). Even though NOAA-20CRv3 does not assimilate any upper-air data, its 13-month smoothed temperature series [i.e., Tsm_{20CR} in Eq. (1)] still correlates significantly with similarly smoothed raw radiosonde data (Figs. 7a,d), and thus it can be used to reduce variations and long-term changes in the radiosonde series. Please note that such smoothed series from JRA-55 (Tsm_{JRA}) may contain spurious changes propagated from the radiosonde data (cf. Figs. B1 and B2) and thus should not be used as a reference in Eq. (1). For the intermonthly variations (Tim), Fig. 7e shows that JRA-55 has higher

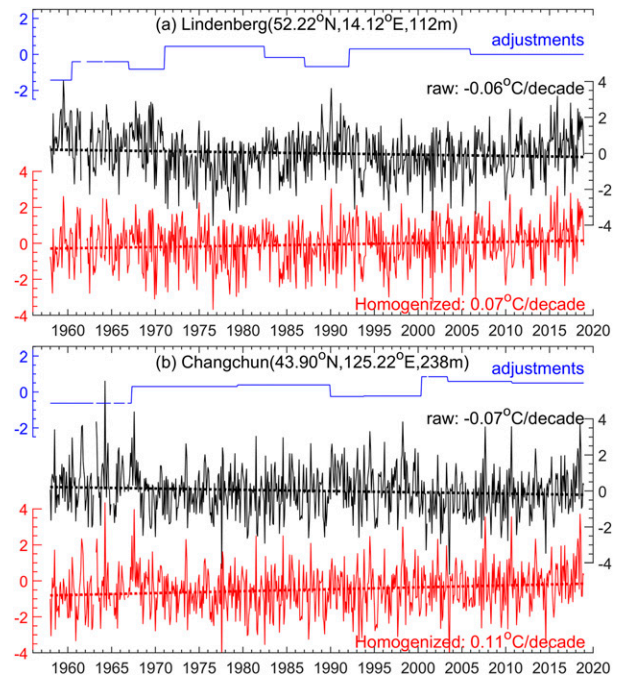


FIG. 9. Comparison of the time series of the raw (black) and homogenized (red) radiosonde monthly temperature anomalies at 1200 UTC and 300 hPa at (a) Lindenberg, Germany, and (b) Changchun, China. The blue line represents the mean adjustments added to the black line. The dashed line is the linear trend, with its slope shown in the same color.

correlations ($r \sim 0.95$) than NOAA-20CRv3 ($r = 0.2\text{--}0.9$) with the radiosonde data, especially for the upper levels. This implies that Tim_{JRA} , rather than Tim_{20CR} , should be used to substantially reduce the intermonthly variations in radiosonde temperatures. The collocated high-frequency anomalies Td_{rds} and Td_{JRA} have correlation coefficients ranging from 0.82 to 0.94 (Figs. 7c,f), and thus Td_{JRA} is able to substantially remove high-frequency variations in Td_{rds} . Note that relatively low correlations over India and South America (Figs. 6a–c) are likely due to the well-documented low quality of radiosonde data there (Raj et al. 1987; Lanzante et al. 2003; Thorne et al. 2005). Because of these strong correlations, the combination of the Tsm_{20CR} and Tim_{JRA} is capable of removing a large part of the natural variations and changes in the radiosonde monthly anomalies through Eq. (1), and the Td_{JRA} can be used to remove most of the high-frequency variations in radiosonde daily temperature through Eq. (2), as illustrated in Fig. 5.

The reanalysis data may contain systematic biases compared with the radiosonde data, and the use of the scaling factors (obtained through linear regression) in Eqs. (1) and (2) is designed to minimize the impact of such biases in our analysis. For some cases (mainly at 150–200-hPa levels) over India, the regression coefficients b in Eq. (1) are smaller than 0.2 or greater than 2 because of a few large jumps in those radiosonde temperature series. For those cases, the coefficient b is calculated from the first-order difference series; otherwise, it is set to 1. Furthermore, some reanalysis products including JRA-55 have a date-matching error (i.e., one day

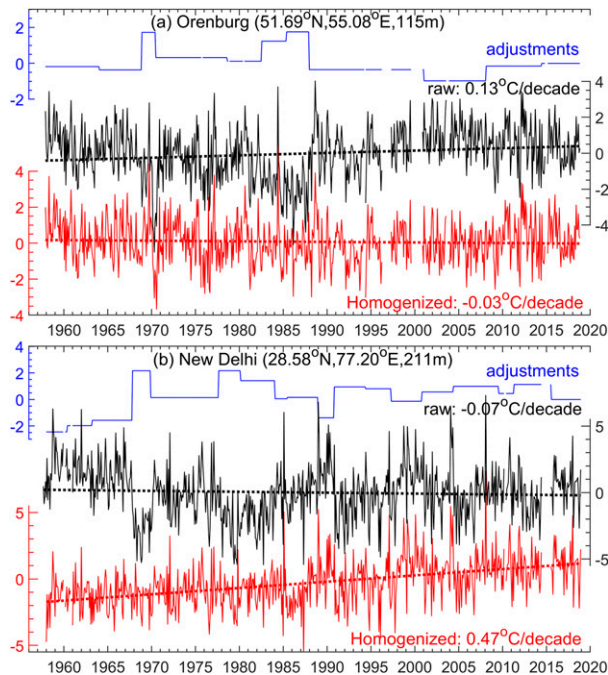


FIG. 10. As in Fig. 9, but at (a) Orenburg, Russia, and (b) New Delhi, India.

ahead of the radiosonde) for the 0000 UTC temperature data during 1970–73 and 1988–95 (Fig. C1). This error likely results from their assimilation of radiosonde subdaily temperature data from various sources including NCEP Global Telecommunication System (NCEP-GTS) messages from 1970 to 1999 that had the date error (Durre et al. 2018) (see appendix C for more details). This date error in the JRA-55 0000 UTC temperature series was corrected in our analysis.

b. Detection of changepoints in the mean and variance

In step 2 (Fig. 4), we employed a PMF test at a significance level of 0.05 to the monthly temperature difference series to detect spurious shifts in the *mean* and an improved K-S test at a significance level of 0.001 to the daily temperature difference series to detect artificial

discontinuities in the intramonthly *variance*. After several sensitivity tests, these confidence levels were chosen in order to have a reasonable and comparable number of changepoints at most of the stations. Sixty percent of the detected changepoints appeared in the mean, and 40% of them in the variance, with $\sim 11\%$ seen in both the mean and variance. This implies that it is important to consider spurious shifts in the variance besides the mean shifts for homogenization of the radiosonde daily temperature data.

The mean length of the segments separated by the detected changepoints shows distinct spatial patterns for the mean and variance, but both display a country-dependent pattern (Fig. 8), which is a result of the usages and changes of country-wide radiosonde instruments (Fig. 3). The globally averaged mean segment length is ~ 7.2 (14.6) years for the mean (variance)-based test and ~ 5.5 years after combining them. For surface-level temperature, global-averaged mean segment length is ~ 16.4 (22.7) years for the mean (variance)-based test and ~ 13.1 years when combined.

We collected all available metadata from the IGRA2 archive in an attempt to validate the detected changepoints to the extent these incomplete records permit. Overall, about 56% of our detected changepoints are confirmed by known metadata events (i.e., having one recorded event within one year of the detected changepoint) including instrument changes, radiation correction method changes, station relocations, and changes in observation practices. Conversely, about 53% of the available metadata events correspond to at least one changepoint (i.e., having at least one changepoint within one year of the metadata event). For example, the detected changepoints at Lindenberg station, Germany, are generally confirmed by its recorded instrument changes from Freiberg (1958–71), to RKS-2 (1971–74), then RKS-5 (1974–87), followed by MARZ (1987–92) and RS80 (1992–2004), and then finally to RS92 (2004–18) (Fig. 5). Lindenberg is highly unusual in a global context in the preservation of its metadata record and the care and attention applied to its time series.

c. Adjustments of detected discontinuities

1) MEAN-MATCHING ADJUSTMENT FOR SPURIOUS MEAN SHIFTS

In step 3 (Fig. 4), to remove the detected shifts in the mean of the monthly temperature difference series, we adopted a

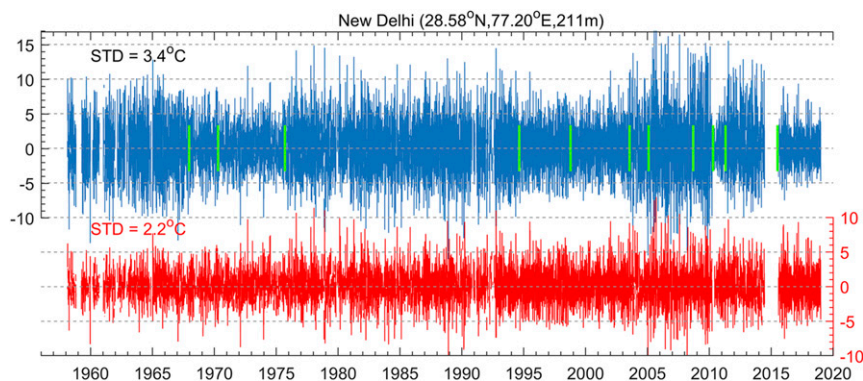


FIG. 11. Comparisons of the time series of the raw (light blue) and homogenized (red) radiosonde daily temperature anomalies (Td_{rds} ; $^{\circ}\text{C}$) at 1200 UTC and 300 hPa at New Delhi, India.

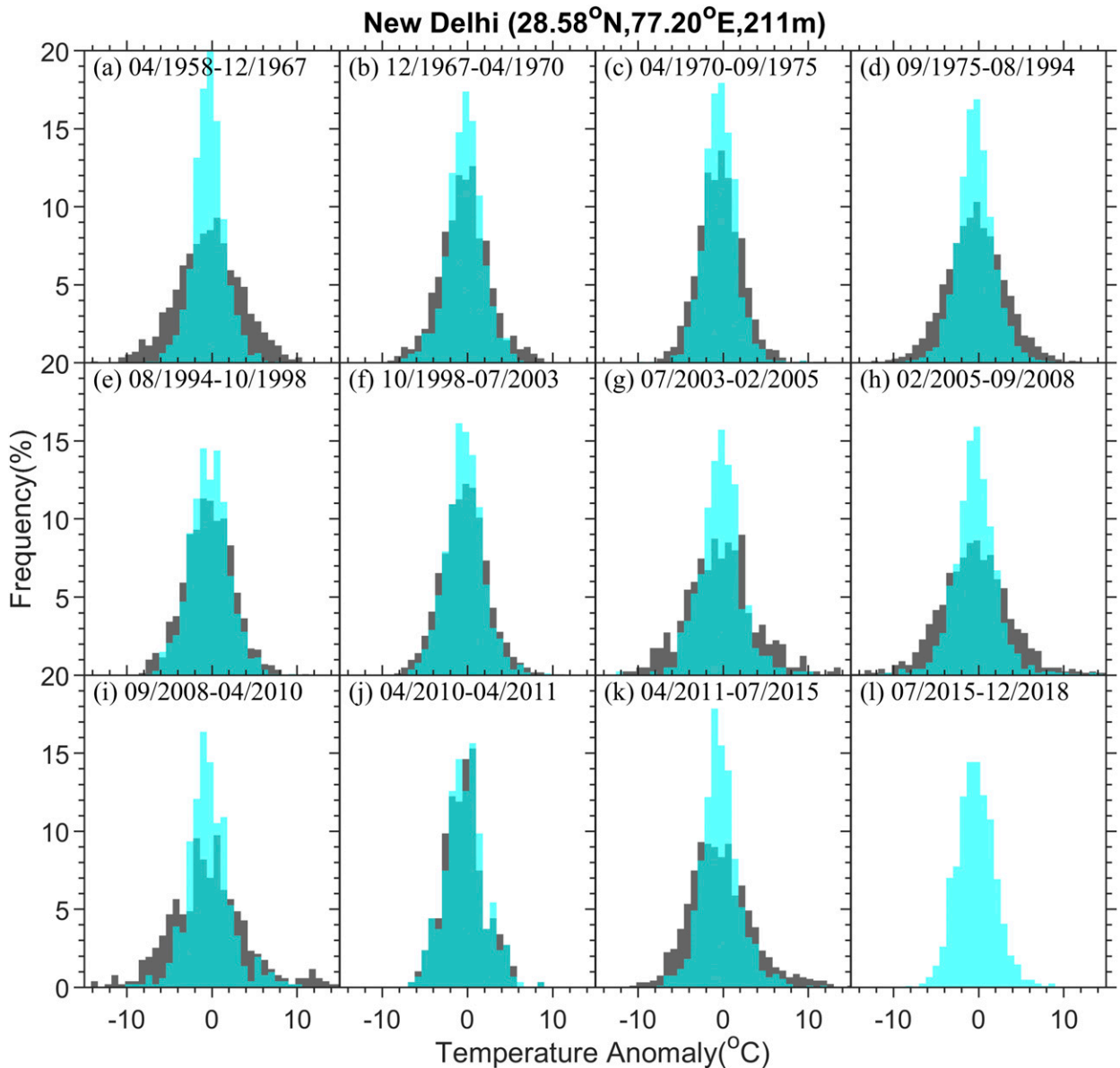


FIG. 12. (a)–(k) Histograms of the high-frequency component of daily temperature anomalies at 1200 UTC and 100 hPa at New Delhi, India. Each panel is from a different segment. Black (cyan) bars are for the raw (homogenized) data. (l) The latest reference segment used to adjust the histograms of all prior segments.

mean-matching adjustment by using up to five years of data (if the segment length is 5 years or longer) before and after each detected changepoint (Thorne et al. 2005; Haimberger 2007). Starting from the last changepoint, the mean difference (over up to 5 years) around the changepoint is used to adjust the data points within the entire segment before the changepoint, so that after the adjustment the mean shift around the changepoint disappears. This process moves sequentially backward until all the mean shifts were removed (Figs. 9 and 10). The latest segment is used as the reference segment because it contains the most recent data collected by the most advanced instruments and thus is likely to be most reliable. Clearly, such a

mean-matching adjustment implicitly assumes that the mean shift estimated using the difference series data around a changepoint is due to nonclimatic changes. This may be invalid if the difference series (DTm; Fig. 5b) still contains substantial natural variations or long-term changes that may contribute to the estimated mean shift around a changepoint. This is a common issue in all mean-matching adjustments used in data homogenization (Peterson et al. 1998; Reeves et al. 2007); it further emphasizes the critical importance of minimizing the natural variations and changes in the difference series. The set-aside component (i.e., $aTsm_{20CR} + bTim_JRA$) preserves most natural temperature variations and changes, making the

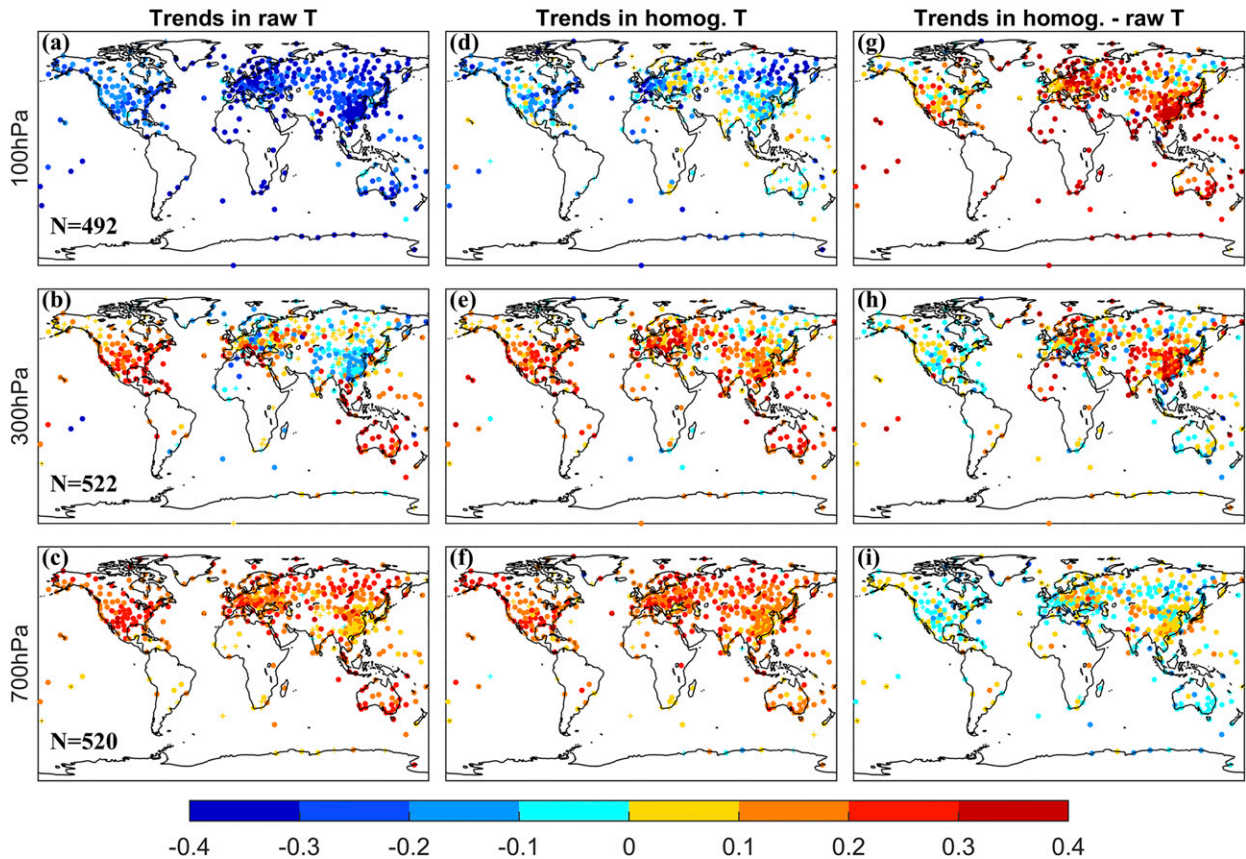


FIG. 13. Linear trends ($^{\circ}\text{C decade}^{-1}$) from 1958 to 2018 in (a)–(c) raw and (d)–(f) homogenized annual-mean temperature data, as well as (g)–(i) their difference at (top) 100, (middle) 300, and (bottom) 700 hPa. To cover a comparable time period, only stations with a data length greater than 30 years and with at least 1 year of data for each decade were included here. The trends significant at the significance level of 0.05 are shown as dots and the nonsignificant trends are shown as plus signs.

adjustment of DTm less affected by them. Despite this potential problem, based on our visual examination and the fact that the majority of our detected changepoints are confirmed by the known metadata, we concluded that the mean shifts estimated around our detected changepoints are likely mainly due to nonclimatic changes, and thus the adjustment should improve the homogeneity and quality of the radiosonde temperature data.

Four examples of the mean adjustment are presented in Figs. 9 and 10, which show clear improvements in long-term homogeneity of the data. Two apparent warm biases during 1958–71 and 1987–92 due to the slow response time of the Freiberg and MARZ sensors at Lindenberg are largely removed after the adjustment (Fig. 9a). The adjustments remove most of the apparent discontinuities and increase the linear trend during 1958–2018 from $-0.06^{\circ}\text{C decade}^{-1}$ at Lindenberg, Germany (Fig. 9a).

At Changchun, China, apparent systematic shifts in the raw data around 1967, 1990, and 2000 are largely removed by the mean adjustment (Fig. 9b). Zhai (1997) argued that relatively large biases in the 1960s in Chinese radiosonde temperature data are associated with frequent changes in instruments and radiation correction methods. The large decrease around 2000 (Fig. 9b) results from the known upgrades of the sounding systems around that time, which led to national use of L-band radar

and electronic radiosondes after 2002 at all Chinese stations (Guo and Ding 2009; Chen and Yang 2014; Guo et al. 2019).

Many of the detected changepoints at the Orenburg (Russia) and New Delhi (India) stations are readily apparent by visual inspection alone (Fig. 10). It is well documented that Indian and Russian radiosonde data contain large inhomogeneities because of their frequent instrument changes and other causes (Raj et al. 1987; Parker et al. 1997; Lanzante et al. 2003; Thorne et al. 2005; Schroeder 2009). The sudden cold biases during 1969–70 and 1983–88 at Orenburg (Fig. 10a) cause a downward trend from 1958 to 1988 but there remains an overall warming ($0.13^{\circ}\text{C decade}^{-1}$) from 1958 to 2018, and the adjustments improve the homogeneity and reduce the long-term trend at this station. Due to frequent instrumental changes, the radiosonde temperature series from New Delhi (Fig. 10b) shows large short-term fluctuations, which are especially marked during 1968–70, 1989–91, and 2011–15. Our adjustments significantly reduce these jumps and lead to a much larger warming trend (Fig. 10b); however, our confidence in the homogenized Indian radiosonde data is comparatively low because of the very poor quality of Indian radiosonde data.

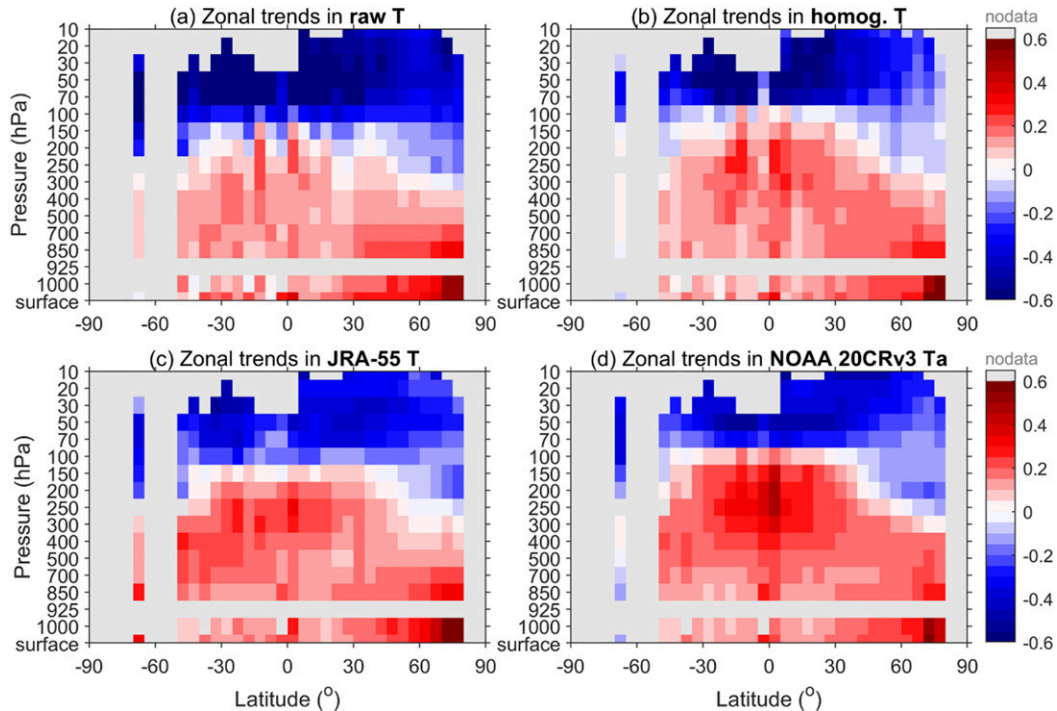


FIG. 14. Zonally averaged latitude–height distributions of the linear trend (in $^{\circ}\text{C decade}^{-1}$) from 1958 to 2018 in atmospheric temperatures from the (a) raw and (b) homogenized radiosonde datasets, and (c) JRA-55 and (d) NOAA-20CRv3 sampled at the radiosonde stations. The data from Fig. 13 were used, and a minimum of three $5^{\circ} \times 5^{\circ}$ grid boxes was required for estimating a zonal average for any given latitude band and time step.

2) QUANTILE-MATCHING ADJUSTMENT FOR SPURIOUS VARIANCE SHIFTS

To adjust spurious discontinuities in the *variance* of the daily temperature difference series, we employed a quantile-matching algorithm of Wang et al. (2010) using up to five years of data before and after each detected changepoint to obtain the adjustment amount for each quantile. Data in each segment were first grouped into 10 quantile categories. Sensitivity tests showed that using 8 to 12 quantile categories produced similar results. Starting from the last changepoint and moving sequentially backward, the category-mean differences between the two adjacent segments (using data up to 5 years) are calculated from quantiles 1 to 10 and then fitted with a natural spline to estimate the adjustment amount for adjusting the data within each quantile of the entire segment before the changepoint. Using the spline fitting to estimate the adjustment amount substantially reduces the impact from the division of the quantile categories. Again, the latest segment is used as the reference without any adjustment. The quantile-matching algorithm is a state-of-the-art approach to adjust the histograms of two samples to be similar, and it has been used to homogenize near-surface daily air temperature (Trewin 2013), daily precipitation (Wang et al. 2010), and subdaily radiosonde humidity data (Dai et al. 2011).

Figure 11 shows the comparison of the homogenized versus raw time series of the high-frequency component of the radiosonde daily temperature at the New Delhi station, whose discontinuities

are typical for other Indian stations. Spurious discontinuities in the *variance* are very apparent during 1958–66, 1975–95, and 2008–10 (Fig. 11), probably due to frequent changes in instruments and observation practices. The quantile-matching adjustment effectively removes these apparent discontinuities, resulting in more homogeneous variance from 1958 to 2018 (Fig. 11). Although we do not have a ground truth to validate the adjusted daily series, Fig. 11 clearly shows that the adjustments improve the homogeneity of the *variance* of the daily data, and this improvement is also seen outside India (see section 4c below).

Figure 12 shows the comparisons of the segment histograms of the homogenized versus raw data at New Delhi. As expected, the histograms of the homogenized data are more comparable to each other and to the reference segment (Fig. 12). The adjustment also works well on some short segments (Fig. 10). Thus, the quantile-matching adjustment greatly improves the homogeneity of the *variance* in daily temperature data.

In the final step, we obtained the homogenized temperature series by adding two homogenized DTm and DTd series back onto the set-aside components (i.e., $a\text{Tsm}_{20\text{CR}} + b\text{Tm}_{\text{JRA}}$, and $c\text{Td}_{\text{JRA}}$).

4. Impacts of homogenization

a. Impacts on long-term trends

The mean adjustment to the monthly anomaly series can significantly alter the long-term trend. Figure 13 shows that

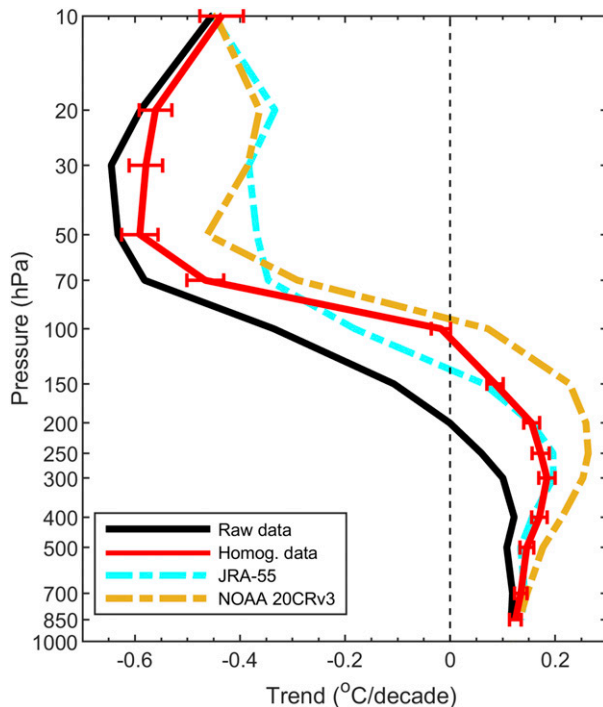


FIG. 15. Low-latitude (30°S – 30°N) mean vertical profiles of the linear trend (in $^{\circ}\text{C decade}^{-1}$) from 1958 to 2018 in the raw and homogenized radiosonde temperature data, and the JRA-55 and NOAA-20CRv3 sampled at the radiosonde stations. The 5%–95% confidence intervals of the homogenized trends are shown as error bars. The data from Fig. 13 were used here.

trends in the homogenized temperature series display more spatially coherent patterns than those in the raw data. In particular, at 300 hPa over China, the homogenized temperature data (Fig. 13e) show positive trends from 1958–2018 (consistent with other regions in Eurasia), in contrast to negative trends in the raw data (Fig. 13b) that have been propagated into several reanalysis products, including NCEP-R1 and JRA-55 (Figs. B1 and B2). The main reason for this change is the removal in the homogenized data of the spurious temperature decreases from the 1960s to 1970s and around 2000 (cf. Fig. 9b) due to the known sounding system changes and upgrades in China as mentioned previously in section 3c(1).

Compared with the raw data, the homogenized data show significantly enhanced warming trends at the surface and in the middle to lower troposphere, especially over Central and East Asia and northern Africa (Figs. 13e,f), which is consistent with near-surface air temperature trends (Zhou and Wang 2016a). The trend differences between the homogenized and raw data increase with height (Fig. 13), which is consistent with larger systematic biases in radiosonde temperatures at higher altitudes (Mears et al. 2006; Thorne et al. 2011a).

Zonally averaged temperature trends from the homogenized data show a warming maximum around 300 hPa over 30°S – 30°N that is evident also in NOAA-20CRv3 and JRA-55 but not in the raw data (Figs. 14 and 15). The vertical structures of the tropospheric warming trends are more comparable

between JRA-55 and the homogenized data than among the other datasets (Figs. 14 and 15), whereas the stratospheric cooling seems to be more consistent among the homogenized data, JRA-55, and NOAA-20CRv3 (Figs. 14 and 15). Due to the opposite influences from tropospheric warming and stratospheric cooling, it is difficult to accurately estimate temperature trends near the tropopause, especially around the Arctic (Fig. 14). The enhanced warming over the tropical upper troposphere has been a robust feature in climate models under increased GHGs (Santer et al. 2005, 2017), but it has been questioned because of the lack of such a warming maximum in the raw radiosonde data (Fig. 14a) (Thorne et al. 2011b; Mitchell et al. 2013). Our results reaffirm previous suggestions that such an inconsistency may be due to the inhomogeneities in the raw radiosonde data, and our homogenized data confirm such a tropospheric warming maximum (Figs. 14b and 15). More detailed comparisons with other homogenized radiosonde datasets, satellite observations and climate model simulations will be reported in a follow-on paper.

The 5th–95th-percentile uncertainty range of the homogenized trends increases with height (Fig. 15), which is due to higher temperature variability and lower data availability at the upper levels. Temperature trends in the raw data are outside the 5th–95th-percentile range of the homogenized trends (Fig. 15), indicating a significant adjustment for the raw data. Note that our homogenization has relatively minor impacts on surface-level temperature trends, with generally enhanced warming over Central Asia and Canada but reduced warming over southern Europe and southwestern North America (figure not shown).

b. Impacts on the quasi-biennial oscillation

The quasi-biennial oscillation (QBO) is a quasi-periodic oscillation (~ 28 months) in the equatorial zonal wind between the easterlies and westerlies in the tropical lower stratosphere (Fig. 16) (Trenberth 1980; Butchart et al. 2020). The QBO is a key feature for lower stratospheric temperature variability (Baldwin et al. 2001; Butchart et al. 2020). Since NOAA-20CRv3 has weak QBO signals (Fig. 16d), the QBO signal in lower stratospheric radiosonde temperatures is retained in the monthly temperature difference series [DTm in Eq. (1)]. Figure 16 shows similar 13–36-month signals between the homogenized and raw temperature data, suggesting that the homogenization preserves the QBO signal in radiosonde temperatures.

c. Impacts on the variance

Figure 17 compares the time-averaged standard deviation of the high-frequency component of the homogenized and raw daily temperature data. The homogenized data show more consistent latitude-dependent patterns for all the pressure levels that are likely associated with different weather patterns. In particular, Indian stations present abnormally large variances in the raw data (Figs. 17a–c) that is significantly reduced in the homogenized data (Figs. 17d–f). Figures 17g–i show that the homogenization generally reduces the variance in the daily data, especially over India, and the reduction in magnitude increases with heights, in accordance with larger biases in

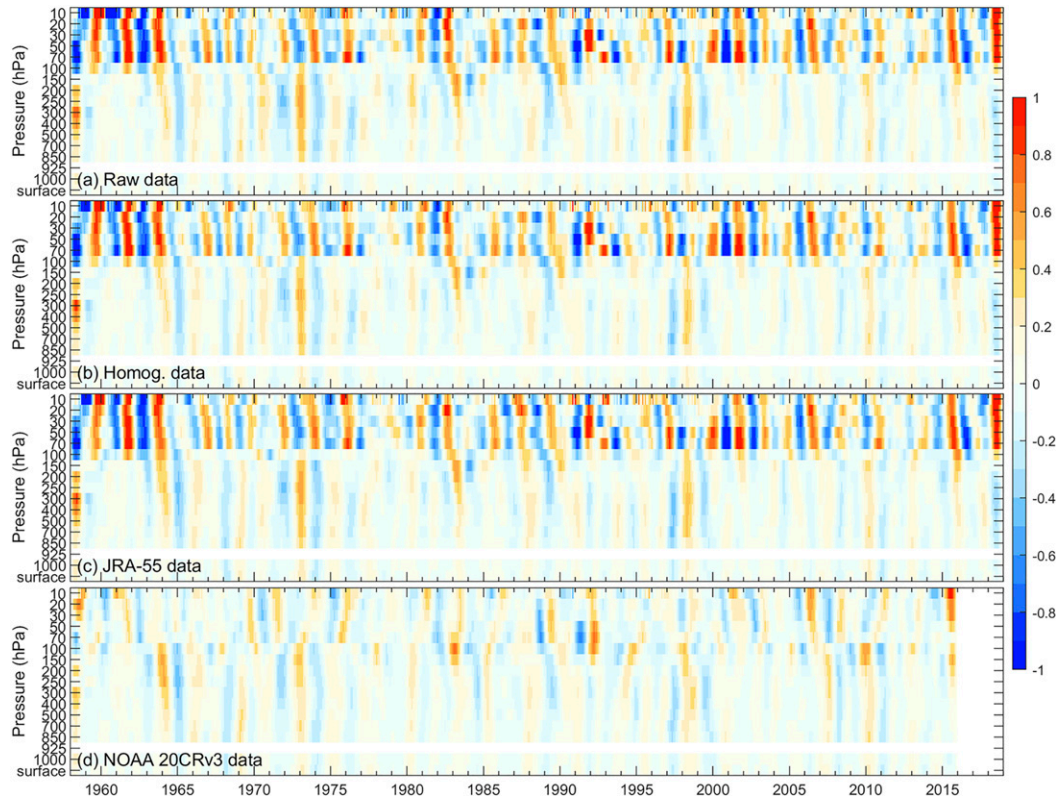


FIG. 16. Low-latitude (30°S – 30°N) mean time–height distributions of the 13–36-month bandpass filtered temperature anomalies ($^{\circ}\text{C}$) from the (a) raw and (b) homogenized radiosonde data, and (c) JRA-55 and (d) NOAA-20CRv3 sampled at the radiosonde stations. The data from Fig. 13 were used here.

radiosonde temperature data at higher altitudes (Mears et al. 2006; Thorne et al. 2011a). While we do not have a ground-truth to validate the variance of the homogenized data, Fig. 11 clearly shows that our quantile-matching based adjustments improve the homogeneity of the variance of the daily series, and Figs. 17g–i show that this improvement is seen not only for Indian stations, but also for stations in Southeast Asia, Europe, and other regions. The improved variability in daily temperature data is critically important for studying extremes and for improving future reanalysis products. The homogenization also makes the surface-level temperature variance more consistent over time than the raw data.

5. Conclusions and discussion

To improve the homogeneity of radiosonde temperature data, we have developed a four-step automated approach to effectively detect and adjust spurious shifts in both the mean and variance of the subdaily radiosonde temperature records from 1958 to 2018 at 1184 stations globally. We started with compiling a complete and quality-controlled global radiosonde data collection from 1958 to 2018 for observations near the surface and at 16 standard pressure levels. The final merged dataset mainly comes from IGRA2 with data gaps filled with data from two ERA-assimilated radiosonde datasets. The

four-step homogenization method is summarized in Fig. 4 and briefly described below.

After verifying the absence of detectable inhomogeneities of the 13-month averaged component in NOAA-20CRv3 and short-term components in JRA-55 temperature data (Tsm_20CR, Tim_JRA, and Td_JRA) and their correlations with the collocated radiosonde temperature data, we constructed the monthly and daily temperature difference series using Eqs. (1) and (2) for each observation time and each pressure level at each station. The strong correlations between the radiosonde and reanalysis data (Fig. 7) enable us to remove most of the natural variations and changes from the difference series using the reanalysis-based reference series [i.e., Tsm_20CR, Tim_JRA, and Td_JRA in Eqs. (1) and (2)]. This not only improves the detection and adjustment of spurious changes, but also verifiably preserves important natural temperature variations such as from QBO, ENSO, and volcanic eruption effects in the homogenized data.

In step 2, we employed the Penalized Maximal F (PMF) test of Wang (2008) to the monthly difference series (DTm) and an improved variant of Kolmogorov–Smirnov (K-S) test of Dai et al. (2011) to the daily difference series (DTd) to detect spurious changepoints in the mean and variance, respectively. Approximately 60% of the detected changepoints appear in the mean and 40% in the variance, with $\sim 11\%$ seen in both.

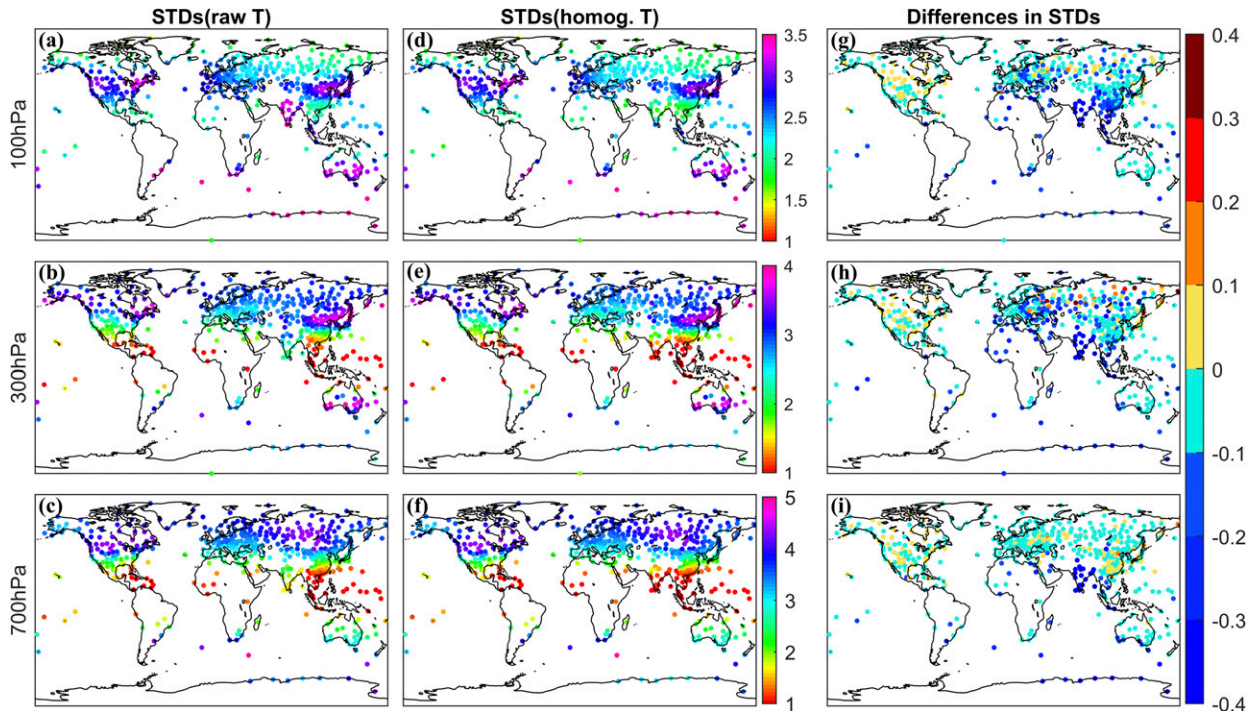


FIG. 17. The 1958–2018-averaged standard deviations (STDs; in $^{\circ}\text{C}$) of the (a)–(c) raw and (d)–(f) homogenized daily temperature anomalies, as well as (g)–(i) their difference at (top) 100, (middle) 300, and (bottom) 700 hPa from 1958 to 2018. The STDs were calculated for each year using the daily anomalies with the monthly mean being removed and then averaged over all the years. The daily data at the stations from Fig. 13 were used here.

The mean length of the segments separated by the detected changepoints display a country-dependent pattern, likely associated with the usages and changes of country-wide radiosonde instruments. Globally averaged mean segment length

is ~ 7.2 years for the mean shifts, ~ 14.6 years for the variance shifts, and ~ 5.5 years after combining them. About 56% of the detected changepoints are confirmed by available metadata and $\sim 53\%$ of the documented changes by available metadata

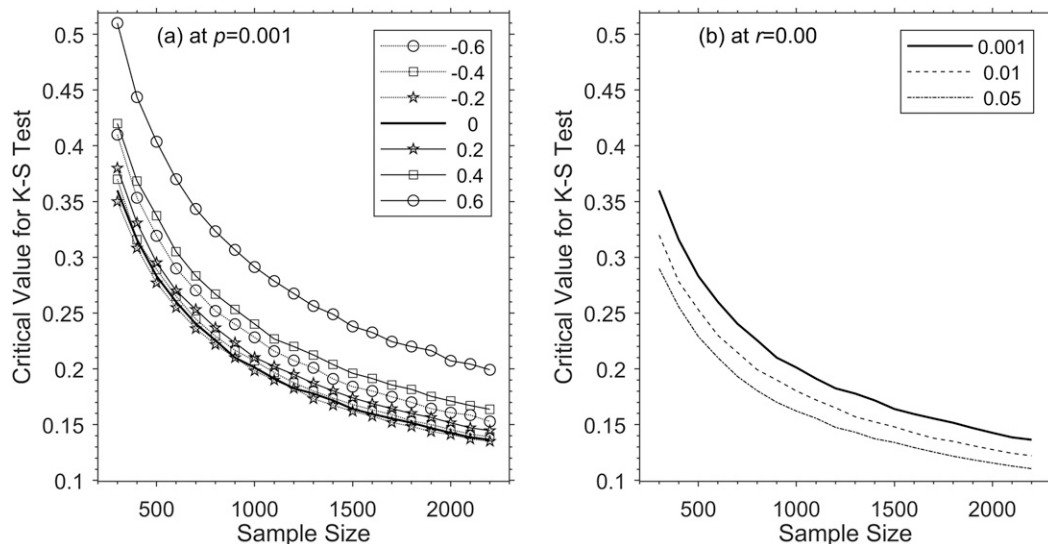


FIG. A1. Critical values estimated from 200 000 Monte Carlo simulations for each case as a function of the (a) sample size and lag-1 autocorrelation with a fixed significance level of 0.001 and (b) sample size and significance level with no autocorrelation for an improved variant of the Kolmogorov–Smirnov (K-S) test.

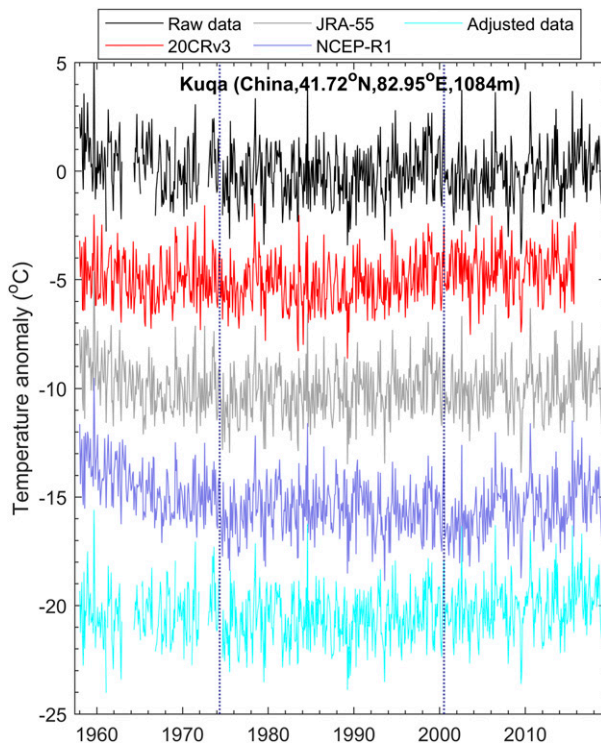


FIG. B1. Time series comparisons of the monthly temperature anomalies for 0000 UTC at 300 hPa at one example station (Kuqa, China) among the raw (black) and homogenized (cyan) data and reanalysis products including JRA-55 (gray), NOAA-20CRv3 (red), and NCEP-R1 (blue). The shifts, e.g., around 1974 and 2000 (vertical dashed lines), still exist in the JRA-55 and NCEP-R1 series because of their assimilation of the raw or underadjusted radiosonde temperature data.

correspond to at least one detected changepoint. These results suggest that many spurious discontinuities are embedded in the radiosonde temperature data, and it is important to consider spurious shifts in both the mean and variance when homogenizing them.

In step 3, we adopted a mean-matching and a quantile-matching algorithm (Wang et al. 2010) to adjust the discontinuities in the DTm and DTd, respectively, with the latest segment as the reference. Finally, the homogenized data series is obtained by adding the two homogenized difference series (DTm and DTd) back to the set-aside components [$aT_{sm_20CR} + bT_{im_JRA}$, and cT_{d_JRA} in Eqs. (1) and (2)].

The impact of the homogenization on long-term trends, QBO signals, and the variance were assessed by comparing them in the raw and homogenized data. Long-term (1958–2018) trends in the homogenized temperature data show more coherent spatial patterns than the raw data. The homogenized data show enhanced warming trends in the middle-to-lower troposphere over central and East Asia and northern Africa, but do not show the spurious cooling around 300 hPa over northern China and Mongolia seen in the raw data and many reanalysis products. A tropospheric warming maximum around 300 hPa over 30°S–30°N is absent in the raw data, but is

present in the homogenized data. Thus, the lack of such a tropospheric warming maximum in previous analyses of radiosonde data (Thorne et al. 2011b; Mitchell et al. 2013) is likely due to the impact of the inhomogeneities in these data. Our homogenized data confirm the existence of such a tropospheric warming maximum present in some homogenized datasets, reanalysis products, and climate models with increased GHGs (Santer et al. 2005; Trenberth and Smith 2006; Thorne et al. 2011b; Haimberger et al. 2012; Mitchell et al. 2013; Santer et al. 2017). The homogenization generally reduces the variance and leads to more consistent latitudinal variations of the variance in daily temperatures, especially for Indian stations.

Our results show that spurious shifts in the mean and variance in the raw subdaily radiosonde temperature data are numerous and nonnegligible. The four-step homogenization approach developed here is effective for detecting and removing these shifts. The improved spatial coherence of the trends and the improved temporal coherence of the variance in the homogenized data suggest that the homogenization improves the quality of the data, although a lack of the truth prevents a thorough validation. We believe that our homogenized dataset (referred to as the University at Albany Homogenized Radiosonde Subdaily dataset or UA-HRD) at 16 standard pressure levels from 1958 to 2018, which will be made available to the community at <ftp://aspen.atmos.albany.edu/data/UA-HRD/>, has many potential applications, including studying tropospheric warming trends, constraining atmospheric reanalyses, evaluating climate simulations, helping homogenize radiosonde humidity data, and studying climate extremes.

Acknowledgments. This study was supported by the U.S. National Oceanic and Atmospheric Administration (Grant NA18OAR4310425). IGRA2 and ERA-assimilated radiosonde daily data (v1.5) were obtained, respectively, from <https://www1.ncdc.noaa.gov/pub/data/igra/> and <https://www.univie.ac.at/theoret-met/research/raobcore/>. NOAA-20CRv3 and JRA-55 data were downloaded from <https://www.esrl.noaa.gov/psd/data> and <http://jra.kishou.go.jp/>, respectively. The Niño-3.4 index was obtained from https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php. Prof. Leo Haimberger and two anonymous reviewers are thanked for providing constructive reviews which served to improve the analysis presented herein.

APPENDIX A

The K-S Test

The conventional Kolmogorov–Smirnov (K-S) test has been widely used to test whether two given PDFs are statistically different (Press et al. 1992). Here, an improved variant of the K-S test as described in Dai et al. (2011) was adopted to test whether two samples from two moving segments have similar or different distributions. For detecting unknown changepoints, we need to search and test each data point within a data window to see if the two segments separated by this data point have different PDFs. As such, the critical value for a given

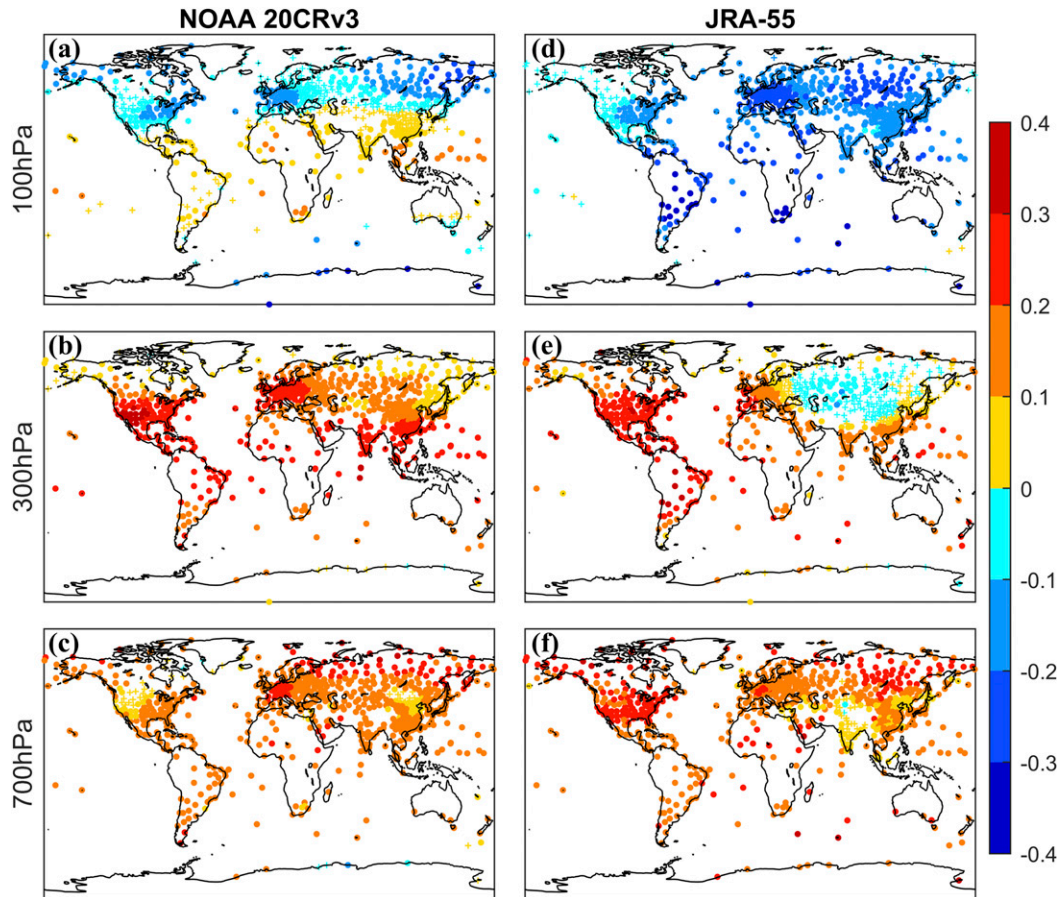


FIG. B2. Linear trends from 1958 to 2018 in the (a)–(c) NOAA-20CRv3 and (d)–(f) JRA-55 annual-mean temperature data at (top) 100, (middle) 300, and (bottom) 700 hPa.

significance level for the K-S test will need to consider the impacts of the lag-1 autocorrelation (r_1) and sample size (N), often through empirical simulations as done in Dai et al. (2011).

Following the procedures of Dai et al. (2011), we used 200 000 Monte Carlo simulations for each case of the r_1 and N values to estimate the critical values (Fig. A1) for use in detecting unknown changepoints. Dai et al. (2011) generated the critical values for the positive r_1 values shown in Fig. A1, here we extended the critical values to cover cases with negative r_1 values, which occurred at a few stations. The critical values nonlinearly increase with r_1 , but decreases with the sample size; and they are smaller for negative r_1 than for positive r_1 (Fig. A1).

APPENDIX B

Time Series Comparisons

Figure B1 shows time series comparisons at Kuqa station in China among the raw and homogenized data, and three reanalysis products, namely, JRA-55, NOAA-20CRv3, and NCEP-R1. JRA-55 and NCEP-R1 assimilated most of the raw

radiosonde data, but NOAA-20CRv3 did not. Even though JRA-55 assimilated the homogenized radiosonde temperature data from RICH (Haimberger et al. 2012), it still shows several large shifts (e.g., around 1974 and 2000) at the Kuqa station (Fig. B1). Partly due to the inadequate adjustments for the radiosonde data, JRA-55 displays a spurious cooling pattern at 300 hPa over Asia from 1958 to 2018 (Fig. B2), which is also seen in the raw data over China and Mongolia (Fig. 13h). This seems to suggest that the adjustments made in RICH may be insufficient.

APPENDIX C

Correction of Date Errors in JRA-55

Reanalysis products including JRA-55 assimilate radiosonde subdaily temperature data from various sources including Global Telecommunication System (GTS) messages from 1970 to 1999 maintained by the National Centers for Environmental Prediction (NCEP). The NCEP-GTS dataset was one of the main sources in IGRA version 1 (IGRA1) but was replaced by a comparable dataset from 1973 to 1999 built by the U.S. Air Force (USAF) 14th Weather Squadron in

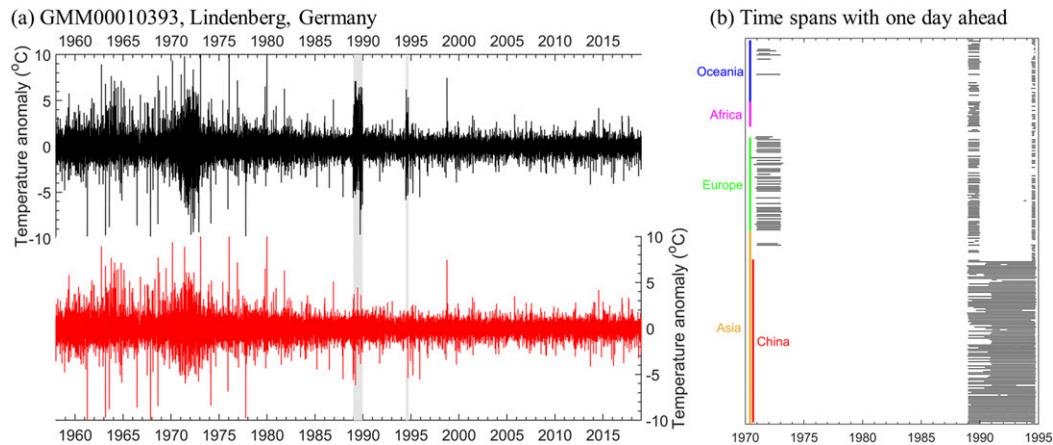


FIG. C1. (a) Time series of 0000 UTC daily temperature difference series between the radiosonde and JRA-55 data of the same date (black) and between the radiosonde data and yesterday's JRA-55 data (red) at the Lindenberg station in Germany. The gray shading indicates the time periods when JRA-55 daily data are shifted backward by 1 day. (b) These periods with the date error in JRA-55 are represented by a black line for each station, with their typical regions labeled on the y axis.

IGRA2. Compared with the NCEP–GTS dataset, the USAF dataset has more complete data from the 1970s to 1980s, particularly for Europe and China, and exhibits similar spatial completeness thereafter. More importantly, the dates of observations reported for 0000 UTC are correct in the USAF dataset, but incorrect in the NCEP–GTS dataset because it is recorded as one day ahead (Durre et al. 2018), which would lead to the same date error for 0000 UTC in reanalysis products that assimilated the NCEP–GTS dataset, including JRA-55.

To correct this date error in JRA-55 daily temperature data for 0000 UTC, in theory it is possible to detect the date error by directly comparing the dates of the identical observations between IGRA2 and IGRA1. However, this is not entirely feasible in practice because of the different data coverages between IGRA2 and IGRA1, especially from the 1970s to 1980s. Therefore, we used another method: (i) perform the K-S test on the 0000 UTC daily temperature difference series (after removing monthly means) between IGRA2 and JRA-55 (i.e., Td_{rds} and Td_{JRA}) at the same date (black lines in Fig. C1) to obtain the segments with different variances that may arise from the date error, instrument changes and so on; and (ii) determine the segment with the date error if the standard deviation of the segment is 25% larger than that of the same segment from another daily temperature difference series between IGRA2 and yesterday's JRA-55 data (red line in Fig. C1a). The 25% value was obtained by multiple tests.

Results show that the 0000 UTC daily data over the segment from 1988 to 1995 at many stations in JRA-55 are one day ahead of IGRA2 (Fig. C1), and these periods with the date error roughly overlap those derived based on direct comparisons of the dates with identical values in IGRA2 and IGRA1. It is worth noting that the segment from 1970 to 1973 was also detected to have the date error at many European stations (Fig. C1), despite there being insufficient data in IGRA1 during this period to confirm it independently. These date errors in JRA-55 were corrected before any subsequent analysis.

REFERENCES

- Baldwin, M., and Coauthors, 2001: The quasi-biennial oscillation. *Rev. Geophys.*, **39**, 179–229, <https://doi.org/10.1029/1999RG000073>.
- Benjamin, S. G., B. E. Schwartz, E. J. Szoke, and S. E. Koch, 2004: The value of wind profiler data in U.S. weather forecasting. *Bull. Amer. Meteor. Soc.*, **85**, 1871–1886, <https://doi.org/10.1175/BAMS-85-12-1871>.
- Butchart, N., J. A. Anstey, Y. Kawatani, S. M. Osprey, J. H. Richter, and T. Wu, 2020: QBO changes in CMIP6 climate projections. *Geophys. Res. Lett.*, **47**, e2019GL086903, <https://doi.org/10.1029/2019GL086903>.
- Carminati, F., S. Migliorini, B. Ingleby, W. Bell, H. Lawrence, S. Newman, J. Hocking, and A. Smith, 2019: Using reference radiosondes to characterise NWP model uncertainty for improved satellite calibration and validation. *Atmos. Meas. Tech.*, **12**, 83–106, <https://doi.org/10.5194/amt-12-83-2019>.
- Chen, J., A. Dai, Y. Zhang, and K. L. Rasmussen, 2020: Changes in convective available potential energy and convective inhibition under global warming. *J. Climate*, **33**, 2025–2050, <https://doi.org/10.1175/JCLI-D-19-0461.1>.
- Chen, Z., and S. Yang, 2014: Homogenization and analysis of China radiosonde temperature data from 1979 to 2012. *Acta Meteor. Sin.*, **72**, 794–804, <https://doi.org/10.11676/qxb2014.046>.
- Dai, A., J. Wang, P. W. Thorne, D. E. Parker, L. Haimberger, and X. L. Wang, 2011: A new approach to homogenize daily radiosonde humidity data. *J. Climate*, **24**, 965–991, <https://doi.org/10.1175/2010JCLI3816.1>.
- , H. Li, Y. Sun, L.-C. Hong, L. Ho, C. Chou, and T. Zhou, 2013: The relative roles of upper and lower tropospheric thermal contrasts and tropical influences in driving Asian summer monsoons. *J. Geophys. Res. Atmos.*, **118**, 7024–7045, <https://doi.org/10.1002/jgrd.50565>.
- Davey, M., A. Brookshaw, and S. Ineson, 2014: The probability of the impact of ENSO on precipitation and near-surface temperature. *Climate Risk Manage.*, **1**, 5–24, <https://doi.org/10.1016/j.crm.2013.12.002>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.

- Della-Marta, P. M., and H. Wanner, 2006: A method of homogenizing the extremes and mean of daily temperature measurements. *J. Climate*, **19**, 4179–4197, <https://doi.org/10.1175/JCLI3855.1>.
- DeRubertis, D., 2006: Recent trends in four common stability indices derived from U.S. radiosonde observations. *J. Climate*, **19**, 309–323, <https://doi.org/10.1175/JCLI3626.1>.
- Durre, I., R. S. Vose, and D. B. Wuertz, 2006: Overview of the integrated global radiosonde archive. *J. Climate*, **19**, 53–68, <https://doi.org/10.1175/JCLI3594.1>.
- , X. Yin, R. S. Vose, S. Appleyquist, and J. Arnfield, 2018: Enhancing the data coverage in the integrated global radiosonde archive. *J. Atmos. Oceanic Technol.*, **35**, 1753–1770, <https://doi.org/10.1175/JTECH-D-17-0223.1>.
- Free, M., 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, <https://doi.org/10.1029/2005JD006169>.
- Fu, Q., S. Manabe, and C. M. Johanson, 2011: On the warming in the tropical upper troposphere: Models versus observations. *Geophys. Res. Lett.*, **38**, L15704, <https://doi.org/10.1029/2011GL048101>.
- Gaffen, D. J., 1993: Historical changes in radiosonde instruments and practices. WMO Instruments and Observing Methods Rep. 50, 123 pp.
- , M. A. Sargent, R. Habermann, and J. R. Lanzante, 2000: Sensitivity of tropospheric and stratospheric temperature trends to radiosonde data quality. *J. Climate*, **13**, 1776–1796, [https://doi.org/10.1175/1520-0442\(2000\)013<1776:SOTAST>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1776:SOTAST>2.0.CO;2).
- 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**, 1037–1051, <https://doi.org/10.1175/2008JCLI2480.1>.
- , S. Zhang, J. Yan, Z. Chen, and X. Ruan, 2016: A comparison of atmospheric temperature over China between radiosonde observations and multiple reanalysis datasets. *J. Meteor. Res.*, **30**, 242–257, <https://doi.org/10.1007/s13351-016-5169-0>.
- , C. Zou, P. Zhai, and G. Wang, 2019: An analysis of the discontinuity in Chinese radiosonde temperature data using satellite observation as a reference. *J. Meteor. Res.*, **33**, 289–306, <https://doi.org/10.1007/s13351-019-8130-1>.
- Haimberger, L., 2007: Homogenization of radiosonde temperature time series using innovation statistics. *J. Climate*, **20**, 1377–1403, <https://doi.org/10.1175/JCLI4050.1>.
- , C. Tavolato, and S. Sperka, 2012: Homogenization of the global radiosonde temperature dataset through combined comparison with reanalysis background series and neighboring stations. *J. Climate*, **25**, 8108–8131, <https://doi.org/10.1175/JCLI-D-11-00668.1>.
- Hartmann, D. L., and Coauthors, 2013: Observations: Atmosphere and surface. *Climate Change 2013: The Physical Science Basis*, Cambridge University Press, 159–254.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Karl, T., S. Hassol, C. D. Miller, and W. Murray, 2006: Temperature trends in the lower atmosphere: Steps for understanding and reconciling differences. A Report by the Climate Change Science Program and Subcommittee on Global Change Research, 180 pp.
- Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, <https://doi.org/10.2151/jmsj.2015-001>.
- 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, [https://doi.org/10.1175/1520-0442\(2003\)016<0224:THOMRT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<0224:THOMRT>2.0.CO;2).
- Lee, M.-H., J.-H. Kim, H.-J. Song, J. Inoue, K. Sato, and A. Yamazaki, 2019: Potential benefit of extra radiosonde observations around the Chukchi Sea for the Alaskan short-range weather forecast. *Polar Sci.*, **21**, 124–135, <https://doi.org/10.1016/j.polar.2018.12.005>.
- Mears, C., C. Forest, R. Spencer, R. Vose, and R. Reynolds, 2006: What is our understanding of the contribution made by observational or methodological uncertainties to the previously reported vertical differences in temperature trends? *Temperature Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences*, Climate Change Science Program and the Subcommittee on Global Change Research Rep., 71–88.
- Mitchell, D., P. Thorne, P. Stott, and L. Gray, 2013: Revisiting the controversial issue of tropical tropospheric temperature trends. *Geophys. Res. Lett.*, **40**, 2801–2806, <https://doi.org/10.1002/grl.50465>.
- Naakka, T., T. Nygård, M. Tjernström, T. Vihma, R. Pirazzini, and I. Brooks, 2019: The impact of radiosounding observations on numerical weather prediction analyses in the Arctic. *Geophys. Res. Lett.*, **46**, 8527–8535, <https://doi.org/10.1029/2019GL083332>.
- Parker, D. E., 2011: Recent land surface air temperature trends assessed using the 20th Century Reanalysis. *J. Geophys. Res.*, **116**, D20125, <https://doi.org/10.1029/2011JD016438>.
- , M. Gordon, D. Cullum, D. Sexton, C. Folland, and N. Rayner, 1997: A new global gridded radiosonde temperature data base and recent temperature trends. *Geophys. Res. Lett.*, **24**, 1499–1502, <https://doi.org/10.1029/97GL01186>.
- Peterson, T. C., and Coauthors, 1998: Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. Climatol.*, **18**, 1493–1517, [https://doi.org/10.1002/\(SICI\)1097-0088\(19981115\)18:13<1493::AID-JOC329>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T).
- Pralungo, R. L., L. Haimberger, A. N. Stickler, and S. Brönnimann, 2014: A global radiosonde and tracked balloon archive on 16 pressure levels (GRASP) back to 1905—Part 1: Merging and interpolation to 00:00 and 12:00 GMT. *Earth Syst. Sci. Data*, **6**, 185–200, <https://doi.org/10.5194/essd-6-185-2014>.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes in Fortran 77: The Art of Scientific Computing*. Vol. 2, Cambridge University Press, 933 pp.
- Raj, Y. E. A., V. Mathew, and J. C. Natu, 1987: Discontinuities in the temperature and contour heights resulting from change of instruments at Indian radiosonde stations. *Mausam*, **38**, 407–410.
- Reeves, J., J. Chen, X. L. Wang, R. Lund, and Q. Q. Lu, 2007: A review and comparison of changepoint detection techniques for climate data. *J. Appl. Meteor. Climatol.*, **46**, 900–915, <https://doi.org/10.1175/JAM2493.1>.
- Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, <https://doi.org/10.1175/JCLI-D-11-00015.1>.

- Santer, B. D., and Coauthors, 2005: Amplification of surface temperature trends and variability in the tropical atmosphere. *Science*, **309**, 1551–1556, <https://doi.org/10.1126/science.1114867>.
- , and Coauthors, 2014: Volcanic contribution to decadal changes in tropospheric temperature. *Nat. Geosci.*, **7**, 185–189, <https://doi.org/10.1038/ngeo2098>.
- , and Coauthors, 2017: Comparing tropospheric warming in climate models and satellite data. *J. Climate*, **30**, 373–392, <https://doi.org/10.1175/JCLI-D-16-0333.1>.
- Schroeder, S. R., 2009: Homogenizing the Russian Federation upper air climate record by adjusting radiosonde temperatures and dew points for instrument changes. *21st Conf. on Climate Variability and Change/89th AMS Annual Meeting*, Phoenix, AZ, Amer. Meteor. Soc., 11–15.
- Sherwood, S. C., C. L. Meyer, R. J. Allen, and H. A. Titchner, 2008: Robust tropospheric warming revealed by iteratively homogenized radiosonde data. *J. Climate*, **21**, 5336–5352, <https://doi.org/10.1175/2008JCLI2320.1>.
- Sippel, S., N. Meinshausen, E. M. Fischer, E. Székely, and R. Knutti, 2020: Climate change now detectable from any single day of weather at global scale. *Nat. Climate Change*, **10**, 35–41, <https://doi.org/10.1038/s41558-019-0666-7>.
- Slivinski, L. C., and Coauthors, 2019: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quart. J. Roy. Meteor. Soc.*, **145**, 2876–2908, <https://doi.org/10.1002/qj.3598>.
- Sun, B., A. Reale, F. H. Tilley, M. E. Pettey, N. R. Nalli, and C. D. Barnett, 2017: Assessment of NUCAPS S-NPP CrIS/ATMS sounding products using reference and conventional radiosonde observations. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **10**, 2499–2509, <https://doi.org/10.1109/JSTARS.2017.2670504>.
- Thorne, P. W., D. E. Parker, S. F. Tett, P. D. Jones, M. McCarthy, H. Coleman, and P. Brohan, 2005: Revisiting radiosonde upper air temperatures from 1958 to 2002. *J. Geophys. Res.*, **110**, D18105, <https://doi.org/10.1029/2004JD005753>.
- , and Coauthors, 2011a: A quantification of uncertainties in historical tropical tropospheric temperature trends from radiosondes. *J. Geophys. Res.*, **116**, D12116, <https://doi.org/10.1029/2010JD015487>.
- , J. R. Lanzante, T. C. Peterson, D. J. Seidel, and K. P. Shine, 2011b: Tropospheric temperature trends: History of an ongoing controversy. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 66–88, <https://doi.org/10.1002/wcc.80>.
- Trenberth, K. E., 1980: Atmospheric quasi-biennial oscillations. *Mon. Wea. Rev.*, **108**, 1370–1377, [https://doi.org/10.1175/1520-0493\(1980\)108<1370:AQBO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1980)108<1370:AQBO>2.0.CO;2).
- , and L. Smith, 2006: The vertical structure of temperature in the tropics: Different flavors of El Niño. *J. Climate*, **19**, 4956–4973, <https://doi.org/10.1175/JCLI3891.1>.
- Trewin, B., 2013: A daily homogenized temperature data set for Australia. *Int. J. Climatol.*, **33**, 1510–1529, <https://doi.org/10.1002/joc.3530>.
- Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, <https://doi.org/10.1256/qj.04.176>.
- Wang, J., and L. Zhang, 2008: Systematic errors in global radiosonde precipitable water data from comparisons with ground-based GPS measurements. *J. Climate*, **21**, 2218–2238, <https://doi.org/10.1175/2007JCLI1944.1>.
- , A. Dai, and C. Mears, 2016: Global water vapor trend from 1988 to 2011 and its diurnal asymmetry based on GPS, radiosonde, and microwave satellite measurements. *J. Climate*, **29**, 5205–5222, <https://doi.org/10.1175/JCLI-D-15-0485.1>.
- Wang, X. L., 2008: Penalized maximal F test for detecting undocumented mean shift without trend change. *J. Atmos. Oceanic Technol.*, **25**, 368–384, <https://doi.org/10.1175/2007JTECHA982.1>.
- , H. Chen, Y. Wu, Y. Feng, and Q. Pu, 2010: New techniques for the detection and adjustment of shifts in daily precipitation data series. *J. Appl. Meteor. Climatol.*, **49**, 2416–2436, <https://doi.org/10.1175/2010JAMC2376.1>.
- Waugh, S., and T. J. Schuur, 2018: On the use of radiosondes in freezing precipitation. *J. Atmos. Oceanic Technol.*, **35**, 459–472, <https://doi.org/10.1175/JTECH-D-17-0074.1>.
- Zhai, P., 1997: Some gross errors and biases in China's historical radiosonde data (in Chinese). *Acta Meteor. Sin.*, **55**, 563–572, <https://doi.org/10.11676/qxxb1997.055>.
- Zhao, T., A. Dai, and J. Wang, 2012: Trends in tropospheric humidity from 1970 to 2008 over China from a homogenized radiosonde dataset. *J. Climate*, **25**, 4549–4567, <https://doi.org/10.1175/JCLI-D-11-00557.1>.
- Zhou, C., and K. Wang, 2016a: Land surface temperature over global deserts: Means, variability and trends. *J. Geophys. Res. Atmos.*, **121**, 14 344–14 357, <https://doi.org/10.1002/2016JD025410>.
- , and —, 2016b: Coldest temperature extreme monotonically increased and hottest extreme oscillated over Northern Hemisphere land during last 114 years. *Sci. Rep.*, **6**, 25721, <https://doi.org/10.1038/srep25721>.
- , —, and Q. Ma, 2017: Evaluation of eight current reanalyses in simulating land surface temperature from 1979 to 2003 in China. *J. Climate*, **30**, 7379–7398, <https://doi.org/10.1175/JCLI-D-16-0903.1>.
- , Y. He, and K. Wang, 2018: On the suitability of current atmospheric reanalyses for regional warming studies over China. *Atmos. Chem. Phys.*, **18**, 8113–8136, <https://doi.org/10.5194/acp-18-8113-2018>.
- , K. Wang, D. Qi, and J. Tan, 2019: Attribution of a record-breaking heatwave event in summer 2017 over the Yangtze River Delta. *Bull. Amer. Meteor. Soc.*, **100** (1), S97–S103, <https://doi.org/10.1175/BAMS-D-18-0134.1>.
- Zhou, T., and J. Zhang, 2009: Harmonious inter-decadal changes of July–August upper tropospheric temperature across the North Atlantic, Eurasian continent, and North Pacific. *Adv. Atmos. Sci.*, **26**, 656–665, <https://doi.org/10.1007/s00376-009-9020-8>.