

Overview of the Integrated Global Radiosonde Archive

IMKE DURRE, RUSSELL S. VOSE, AND DAVID B. WUERTZ

National Climatic Data Center, Asheville, North Carolina

(Manuscript received 15 October 2004, in final form 23 June 2005)

ABSTRACT

This paper provides a general description of the Integrated Global Radiosonde Archive (IGRA), a new radiosonde dataset from the National Climatic Data Center (NCDC). IGRA consists of radiosonde and pilot balloon observations at more than 1500 globally distributed stations with varying periods of record, many of which extend from the 1960s to present. Observations include pressure, temperature, geopotential height, dewpoint depression, wind direction, and wind speed at standard, surface, tropopause, and significant levels.

IGRA contains quality-assured data from 11 different sources. Rigorous procedures are employed to ensure proper station identification, eliminate duplicate levels within soundings, and select one sounding for every station, date, and time. The quality assurance algorithms check for format problems, physically implausible values, internal inconsistencies among variables, runs of values across soundings and levels, climatological outliers, and temporal and vertical inconsistencies in temperature. The performance of the various checks was evaluated by careful inspection of selected soundings and time series.

In its final form, IGRA is the largest and most comprehensive dataset of quality-assured radiosonde observations freely available. Its temporal and spatial coverage is most complete over the United States, western Europe, Russia, and Australia. The vertical resolution and extent of soundings improve significantly over time, with nearly three-quarters of all soundings reaching up to at least 100 hPa by 2003. IGRA data are updated on a daily basis and are available online from NCDC as both individual soundings and monthly means.

1. Introduction

Radiosondes have been launched on a daily or twice-daily basis at stations around the globe since the 1940s. During its 1- or 2-h ascent from the surface into the stratosphere, a radiosonde transmits its measurements to ground receiving stations where they are processed into pressure, temperature, dewpoint depression, and geopotential height. Wind direction and speed are obtained by tracking the position of the balloon during its ascent. Thermodynamic and wind observations may be provided at mandatory pressure levels, additional required levels, significant levels, and certain fixed-height increments. Mandatory pressure levels include those specified by the World Meteorological Organization (WMO 1996: 1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, and 10 hPa) as well as those additional levels suggested by the U.S. National

Weather Service (FCM-H3 2004: 7, 5, 3, 2, and 1 hPa). Surface observations taken at or near the launch site are included in the sounding as a “surface level.” Conforming to standards set forth by the WMO, the radiosonde, wind, and surface measurements are compiled into a report that is transmitted as a binary-coded message over the Global Telecommunications System (GTS) to various regional and national meteorological centers around the world, where they are processed, archived, and redistributed to other locations (WMO 1996).

Although radiosonde observations have traditionally been taken primarily for the purpose of operational weather forecasting, they are critical to other applications, including model verification, climate research, and the verification of satellite measurements (Finger and Schmidlin 1991; NRC Panel on Reconciling Temperature Observations 2000; Free et al. 2002; Durre et al. 2005). Radiosonde measurements also constitute the only source of upper-air information prior to the 1970s and have historically provided a higher vertical resolution than satellite observations. Consequently, various

Corresponding author address: Dr. Imke Durre, National Climatic Data Center, 151 Patton Avenue, Asheville, NC 28801.
E-mail: Imke.Durre@noaa.gov

TABLE 1. Data sources for IGRA. Source names are listed by categories referenced in the text. For each source, IGRA contains the periods of record, number of stations, and percentage of soundings listed.

	Data source	Period of record	Area of coverage	No. of stations	Percent of IGRA soundings
Core	NCDC historical GTS	1963–70	Global	820	7.94
	NCAR/NCEP GTS	1970–72	Global	848	3.01
	NCEP GTS	1973–99	Global	1517	64.06
	NCDC real-time GTS	2000–present	Global	1093	7.13
Other large scale	Russian GTS	1998–2001	Global	923	1.59
	U.S. Air Force	1946–73	Global	292	4.49
	Australian GTS	1990–93	Southern Hemisphere	170	0.15
Country specific	United States	1946–2001	United States + U.S. military	150	9.81
	Australian	1938–89	Australia and its territories	17	1.63
	Argentine	1958–91	Argentina	8	0.18
	South Korean	1984–92	South Korea	4	0.01

efforts have been undertaken to compile historical collections of these observations. The Radiosonde Data of North America (Schwartz and Govett 1992), the Tropical Ocean Global Atmosphere Coupled Ocean-Atmosphere Response Experiment (TOGA COARE) Upper-Air Sounding Archive (Loehrer et al. 1996), and the Historical Arctic Radiosonde Archive (Kahl et al. 1992) are examples of sounding archives constructed for the purpose of analyzing the weather and climate of a particular region. Global-scale datasets of soundings have been assembled by the reanalysis projects (Kalnay et al. 1996; Uppala 2005) and in the National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center's (NCDC's) Comprehensive Aerological Reference Data Set (CARDS) (Eskridge et al. 1995). As foundation of two analyses of bias-adjusted monthly mean temperatures (Lanzante et al. 2003a,b; Thorne et al. 2005), CARDS is perhaps the most widely used of these datasets. However, unlike the regional datasets, CARDS and the reanalysis input data have neither been made available in an easy-to-use format nor have they necessarily been subjected to a high level of scrutiny for proper station identification (see the appendix and see Haimberger 2005).

The Integrated Global Radiosonde Archive (IGRA) project at NCDC constitutes an effort to produce a user-friendly, easily accessible dataset of quality-assured radiosonde observations from around the world. Specifically, the goals of the IGRA project are 1) to combine as many reliable data sources as possible into one radiosonde archive, 2) to develop and apply quality assurance algorithms that remove gross errors in the data, 3) to put into place an automatic system for updating the resulting archive on a daily basis, and 4) to provide unrestricted online access to the data. This paper provides an overview of the merging and quality

assurance methods used in IGRA as well as a general description of the dataset. Data sources and merging procedures are discussed in section 2. An overview of the quality assurance approach is given in section 3. Section 4 lists the types of data and auxiliary information available as part of IGRA. Sections 5 and 6 contain a description of the final dataset and a brief comparison with other global-scale sounding archives, respectively. A summary and future plans for IGRA are provided in section 7. A discussion of the motivation for replacing CARDS and a comparison between IGRA and CARDS are presented in the appendix.

2. Data integration

a. Data sources

IGRA constitutes a compilation of 11 source datasets (Table 1) selected based on the timely availability of the data, the existence of documentation for codes and conventions, and data quality. The core of IGRA consists of four GTS-based datasets that were preprocessed at one of three locations in the United States: NCDC (1963–70 and 2000–present); the National Center for Atmospheric Research (NCAR) (December 1970–72); and the National Centers for Environmental Prediction (NCEP) (1973–October 1999). Since these datasets have nearly consecutive periods of record, their records were concatenated into one “core” time series per station. Depending on data availability, the resulting time series may begin as early as September 1963 and continue until present. Many of the concatenated core records contain a 2.5-month break between the end of the NCEP–NCAR GTS in October 1999 and the beginning of the NCDC real-time GTS in January 2000. This gap is, in many cases, filled in with data from other sources.

Two additional GTS data sources originate from the Australian Bureau of Meteorology (1990–93) and the All-Russian Institute for Hydrometeorological Information (1998–2001). For a variety of reasons, including differences in decoding practices, some messages transmitted over the GTS are decoded only at certain receiving centers and not at others. Thus, even though extensive duplication generally exists among the core, Australian, and Russian GTS data, the latter two sources occasionally supply soundings that are either not present or are incomplete in the core data.

Five other datasets are also in IGRA. With a period of record of 1946–73, a dataset compiled by the U.S. Air Force extends the records of many stations back in time from the 1960s to the 1950s or 1940s. The temporal completeness and vertical resolution of data at stations in the United States, Australia, Argentina, and South Korea are further enhanced by four country-specific sets of data that were archived before their transmission over the GTS and thus contain levels not found in the GTS data. (Six additional sources archived at NCDC were excluded from IGRA owing to questionable data quality, undocumented quality assurance flags, or unusual and undocumented conventions for reporting pibal observations.)

In most data sources, stations are identified only by their station number and location. Consequently, information such as the name and country of the station was obtained from external sources: GTS metadata from NCEP and NCDC; the station inventory of the Global Historical Climatology Network (GHCN) (Peterson and Vose 1997); WMO Publication 9, Volume A (WMO 2004); and a list of station moves affecting National Weather Service stations (Elliott et al. 2002). In those rare cases in which significant discrepancies exist in the information provided by the various lists, online searches were used to determine any necessary corrections.

b. Data comparisons

A set of intersource data comparisons was performed to check for any inconsistencies in station number assignments or widespread systematic discrepancies among data sources. Using all elements at the five mandatory levels between 850 and 300 hPa, the data for each station in any one source were compared with the data for all other stations in every other source. Taking into account differences in processing procedures among the various data sources, two overlapping station records are considered to closely match each other if a significant percentage of the differences between pairs of values fall within the similarity thresholds listed in Table 2 (i.e., the “percentage of similarity” exceeds a

TABLE 2. Similarity thresholds used when comparing data from different sources.

Variable	Threshold
Geopotential	
Height	10 m
Temperature	0.2°C
Dewpoint depression	0.5°C
Wind speed	2 m s ⁻¹
Wind direction	10°

specified value). One would expect to find such a match when data for the same station (e.g., 72210) are available from different sources, but not for two entirely different stations. Yet the latter situation does occur on occasion. For example, for station 72210 in the core GTS data and station 72211 in the U.S. data source, 99.7% of compared values are “identical” during the overlapping period of 1992–95. Based on an examination of station history information and the various sources of data, such cases were handled either by excluding one or both station records from further processing or, as in the aforementioned example, by reassigning one of the records to the station number of the other.

The comparison results further reveal a number of cases in which overlapping records for a particular station from different sources are less similar than might be anticipated or desirable. For example, for approximately one quarter of the stations compared, the percentage of similarity is less than 90% for at least one data element. Such relatively low similarities tend to be more common during the 1950s and 1960s than in later years. The disparities imply that the integration of different data sources can result in spurious shifts and additional noise in the resulting dataset. As a result, the construction of a single merged archive from multiple sources necessitates the development of merging procedures that minimize the risk of introducing such undesirable characteristics.

c. Station selection and data merging

The core IGRA station network consists of land-based stations with data in the NCDC real-time GTS since these are the stations with the most reliable location information. This network is supplemented with identifiable stations that no longer report observations but significantly enhance the spatial coverage during the historical record (Fig. 1). Given this combined network, the selection of data sources to be used takes place on a station-by-station basis. For any particular station, the core GTS data are used as the base record and supplemented with only those sources for which the percentage of similar values equals at least 90% for

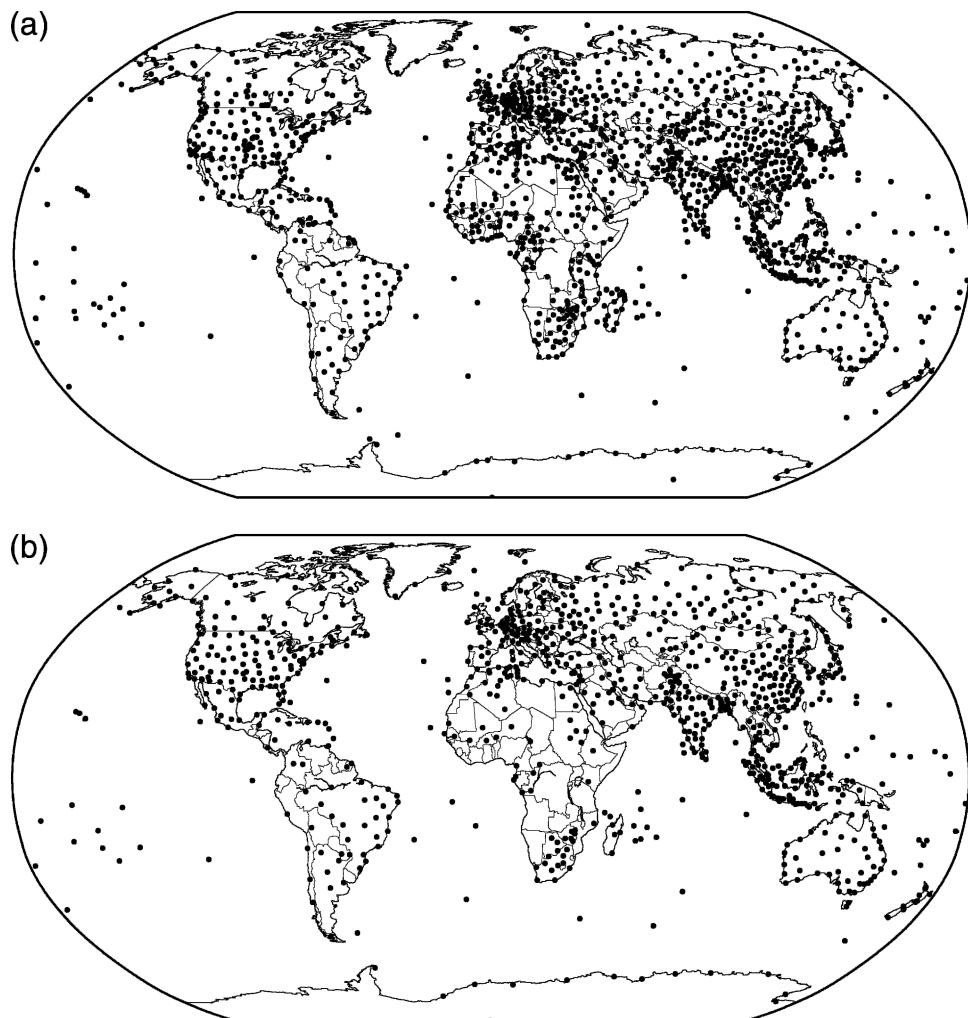


FIG. 1. Locations of (top) all IGRA stations and (bottom) IGRA stations that were active during 2003.

each data element in all possible comparisons. Any new source whose record does not provide a period of overlap for comparison with at least one other source is excluded from that particular station's record.

Once the sources to be used for a station have been selected, their data are merged on a sounding-by-sounding basis. When soundings with the same time stamp are available from multiple sources, the sounding with the largest number of values is chosen. The same procedure is also used to eliminate multiple occurrences of soundings for the same station and time within any one data source, which may arise from transmission or processing errors. Since some data sources report the nominal observation time (e.g., 0000 UTC) as the observation hour, while others report the hour closest to the launch time (e.g., 2300 UTC), the sounding with the largest number of values is also retained when identical soundings appear consecutively within

2 h of each other. Allowing for differences in data processing, two soundings from different sources are considered identical if at least 90% of the absolute differences between values at levels common to both soundings fall within the previously defined similarity thresholds (Table 2). Consecutive soundings that meet these criteria of similarity and whose time stamps are more than 2 h apart are discarded (i.e., the duplication of their data is considered erroneous).

Two additional procedures are then applied to the merged dataset. First, with the purpose of identifying cases in which identical soundings are reported simultaneously at more than one station, the mandatory-level 850-to-300-hPa data of concurrent soundings from all stations are compared. Approximately 60 000 soundings (0.2%) were identified as interstation duplicates and removed from the dataset. Second, composite records were created for a number of stations whose

radiosonde observations were reported under two or more station numbers over time. Many such changes in station number occurred without a discernible change in station location and were the result of changes in the numbering system used by the WMO (e.g., at Canadian stations in 1977). The compositing procedure merges the records of such stations into one record, which is then assigned the station number of the most recent station. In addition, at stations in the contiguous United States during the 1990s, radiosonde observations were moved from one site to another site close enough to reflect the same regional atmospheric conditions (Elliott et al. 2002). The records of such stations are also combined as long as they are located within 150 km of each other and their periods of record do not overlap. The 151 composite stations are identified in the IGRA station list, and the dates and times of the first and last soundings of each original station record and the corresponding composite record are listed in an auxiliary documentation file. Users engaged in climate change studies are advised to consider the potential impact of the compositing on their specific analysis, particularly when the emphasis is on the planetary boundary layer.

3. Quality assurance

The quality of radiosonde data is compromised by a variety of observation, transmission, and processing problems (Schwartz and Doswell 1991; Gandin et al. 1993; Gaffen 1994). In general, quality assurance procedures for sounding data rely on principles of internal consistency, basic physical relationships, and/or statistical methods (Kahl et al. 1992; Eskridge et al. 1995; Loehrer et al. 1996; Collins 2001a,b). Some approaches employ a decision-making algorithm that takes into account the results of multiple tests, while others apply a sequence of independent checks. Since the performance and complexity of the decision-making approach are highly dependent on the number and types of checks applicable to any particular data point, the sequential approach is more straightforward to evaluate when working with a dataset with variable temporal and spatial resolution. Consequently, a sequential approach is employed in IGRA.

To account for the variety of errors that may be present, the IGRA quality assurance system consists of a series of specialized algorithms that are applied successively. Each successive check makes a binary decision on the quality of a value, level, or sounding; either the data item passes the check and remains available or it is identified as erroneous and thus set to missing. As discussed in Peterson and Vose (1997), this approach relieves the end user from the burden of determining the meaning of quality flags. However, for users inter-

ested in making their own binary decision based on our quality assessment results, record-keeping files listing erroneous values are provided by the authors upon request. For all checks, the thresholds used to identify erroneous values were selected based on a careful evaluation of both summary statistics and specific examples of the values identified as unrealistic.

The IGRA quality assurance procedures can be grouped into seven general categories: fundamental “sanity” checks, checks on the plausibility and temporal consistency of surface elevation, internal consistency checks, checks for the repetition of values, climatologically based checks, checks on the vertical and temporal consistency of temperature, and data completeness checks (Table 3). The first four categories eliminate gross errors that might compromise the performance of subsequent algorithms. The climatology and temperature consistency checks identify outliers based on station-specific climatological parameters and are applicable only when sufficient data are available for computing the required statistics. Although all variables are quality assured, temperature, pressure, and geopotential height receive somewhat greater scrutiny in order to facilitate operational climate monitoring activities at NCDC.

a. Fundamental sanity checks

Each data source undergoes two sanity checks, the first being a basic plausibility check to determine whether the date, observation hour, launch time, and data values in each sounding fall within certain gross plausibility limits (Table 4). The date and time limits identify instances of invalid days of the month (e.g., 31 April), invalid times of day, and soundings with a missing observation hour. Soundings with such invalid dates or times are excluded from further processing. The data limits are chosen so as to remove values that clearly exceed all known world extremes, such as temperatures less than -120°C or greater than 70°C . Overall, 0.25% of all date/time stamps as well as 0.025% of all data values were found to be implausible.

The second sanity check, which focuses on “duplicate” data, identifies cases in which two or more data levels within a sounding have identical pressure values or, if no pressure is reported, identical heights. Such cases of level duplication are addressed by removing any data values that differ among the duplicate levels and combining the remaining data into one level. For example, a sounding may contain two 500-hPa levels, one with geopotential height, temperature, and dew-point depression and one with geopotential height, wind direction, and wind speed. If the geopotential height values at the two levels are identical, the data

TABLE 3. IGRA quality assurance procedures and their impact.

Category	QA procedure	Items checked	Items deleted
Fundamental sanity checks	Date check	Year, month, day, hour	Sounding
	Release time check	Release time	Release time
	Observation value check	p, z, T, d, ws, wd	Individual values
	Duplicate level check	Pressure or height	Levels
Station elevation checks	Elevation inspection (manual)	Surface height	Surface height
Internal consistency checks	Hypsometric check	p, z	Individual levels
	Height sequence check	z	Levels
	Multiple surface levels check	Level-type indicator	Levels
	Surface inspection (manual)	p, T	Level
	Below-surface level check	p, z	Level
	Obs hour/release time check	Obs hour—release time	Sounding
	Zero-speed wind check	ws, wd	ws, wd
Checks for repetition of values	Temporal runs check (generic)	p, z, T	Levels or values
	Temporal runs check (by hour)	p, z, T	Levels or values
	Vertical runs check	T	Values
	Joint vertical runs check	T, d, ws, wd	Values
	Frequent erroneous values check	z, T	Values
	Fixed geopotential height	z (Russian GTS only)	Values
Climatological checks	Tier 1	p, z, T	Levels or values
	Tier 2	p, z, T	Levels or values
Additional checks on temperature	Crazy profile check	T	T soundings
	Generic vertical outlier check	T	Values
	Vertical sore-thumb check	T	Values
	Temporal sore-thumb check	T	Values
Data completeness checks	Lone dewpoint depression check	d, T	Values
	Lone wind value check	ws, wd	Values
	Incomplete level check	p, z, d, T, ws, wd	Levels
	Surface-only sounding check	Level-type indicator	Sounding
	Isolated sounding check	Date and time	Sounding

from the two levels are combined into one level containing all variables. If, however, the two geopotential height values do not agree, then the geopotential heights are removed from both levels, and the remaining values are combined into one level from which only geopotential height is missing. Of the more than 30 million soundings processed, approximately one-

quarter contained duplicate levels, with an average of three such levels per sounding. Discrepancies in data values, however, were found only at a few percent of these duplicate levels.

b. Checks on surface elevation

Surface observations are frequently included in a sounding as a surface level identified by a special level type indicator. The height of such levels generally originates either from the source of the sounding data or from various station lists used during initial processing at NCDC. The accuracy and temporal consistency of these heights can thus be compromised by errors in the original data sources or station lists, by processing problems, or by the integration of multiple sources reporting different elevations for the same station in time. Consequently, it was necessary to develop procedures for the removal of gross errors and unrealistic temporal variations in surface level heights.

The two surface elevation checks involved the computation of “monthly median elevations” as well as the

TABLE 4. Plausibility limits used in basic validity checks.

Item	Valid range
Year	1938–present
Month	1–12
Day	1–last day in month
Observation hour	0–23
Launch time (h)	0–23
Launch time (m)	0–59
Pressure	1100–0.1 hPa
Geopotential height	–1000 to 70 000 m
Temperature	–120° to 70°C
Dewpoint depression	0° to 70°C
Wind speed	0–150 m s ^{–1}
Wind direction	0°–360°

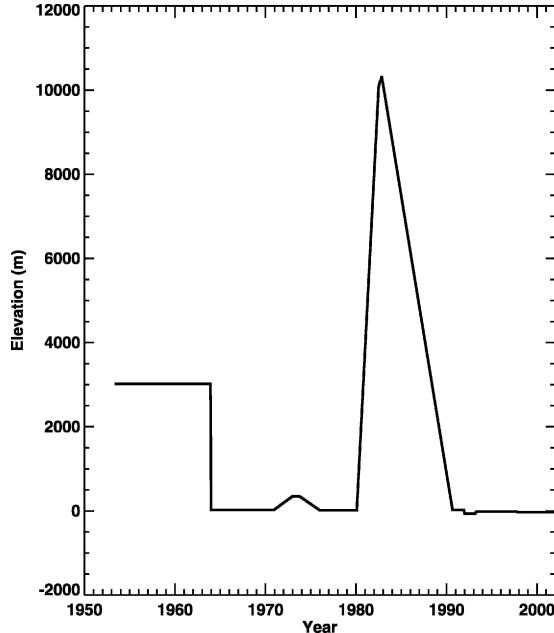


FIG. 2. Time series of monthly median surface elevation reported at Atyran, Kazakhstan.

inspection of elevation time series for unrealistic spikes or jumps. First, isolated errors were removed and inter-source discrepancies were reduced by replacing the surface level height in each sounding with the monthly median elevation generated from all available soundings for the corresponding station, year, and month. Next, each station's time series of monthly median elevations was examined for unrealistic features, periods with implausible elevations were identified, and the respective surface level heights were set to missing. In inspecting the elevation time series, features considered unrealistic included any combination of the following characteristics: significant (>50 m) discontinuities or spikes in the time series, inconsistencies with corresponding time series of surface pressure, and a large discrepancy with either the elevation reported in WMO (2004) or the elevation of the nearest grid point in the Global One-Kilometer Base Elevation (GLOBE) dataset (NGDC 2004).

An example of a station with implausible and temporally inconsistent elevations is shown in Fig. 2. This station, Atyran, Kazakhstan, has a WMO elevation of -28 m, a GLOBE elevation of -37 m, and a mean surface pressure of 1018 hPa. Thus, the monthly median elevations around 3000 m up to the early 1960s, around 500 m in the mid-1970s, and around $10\,000$ m in 1982 are grossly inconsistent with the remainder of the time series as well as with the other sources of station elevation. Consequently, Atyran's surface level heights during these months were set to missing in IGRA.

As a result of the procedures described above, the insertion of the monthly median elevation resulted in a change from the original surface level height in approximately 3% of all soundings with a designated surface level. Based on the time series inspection, the surface level height was removed from an additional 1% of surface levels. Since these procedures require both manual inspection and the availability of data for an entire month, they are not part of the system that updates the archive on a daily basis. In update mode, the height of a surface level is set to the station's most recent known elevation, and internal consistency checks are used to remove any grossly erroneous elevations.

c. Internal consistency checks

The internal consistency checks developed for IGRA address cases of physical inconsistency among different variables or among values of one variable at different levels within a sounding. For instance, two algorithms evaluate the physical consistency of pressure and geopotential height. Another series of checks ensures that a sounding contains at most one valid surface level and no below-surface levels. Additional checks include one that compares the release time to the reported observation hour and one that evaluates wind direction when the wind speed is 0.

The first algorithm comparing pressure and geopotential height is similar to a hydrostatic check (Gandin 1988) but is independent of the temperature profile within the sounding examined. In this "hypsoemetric check," the range of plausible pressure values for any given height is determined from the hypsometric equation using the extreme values of the average temperature of the atmospheric layer between the surface and the level in question. The extremes of the layer-average temperature are computed using the lapse rates from the *U.S. Standard Atmosphere, 1976*, and assuming surface temperatures of -60°C for the cold extreme and 60°C for the warm extreme. Given these parameters, the hypsometric check removes gross inconsistencies, such as 30-hPa levels with geopotential heights of 0 and surface levels with geopotential heights of 3000 m (Fig. 3). Such inconsistencies were found at 0.09% of the approximately 800 million levels in the dataset.

Although the hypsometric check removes gross inconsistencies between pressure and height, it does not guarantee the monotonic increase of geopotential height with decreasing pressure. To ensure that this basic relationship holds true in all soundings, a second algorithm, the "height sequence check," compares the changes in pressure and height between all possible pairs of levels within a sounding. In this iterative mul-

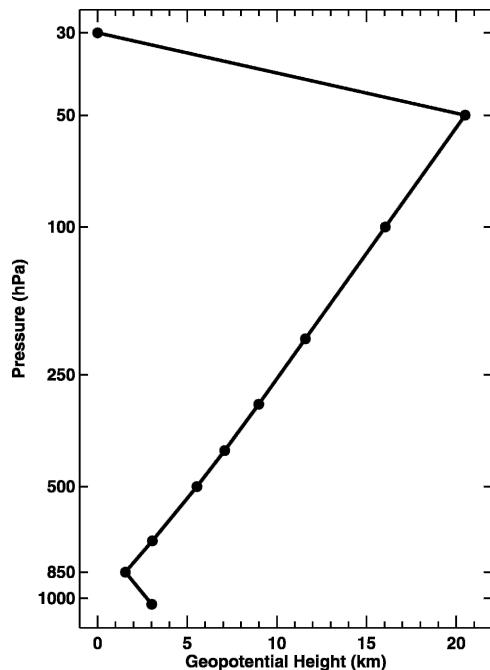


FIG. 3. Log pressure as a function of geopotential height reported at each level of a sounding taken at 1200 UTC 29 Jan 1963 in Atyran, Kazakhstan.

timestep procedure, the height of each pressure level k is compared with the height of every level j having a higher pressure. If the geopotential height of level k is found to be less than or equal to the geopotential height of level j , the numbers of violations for levels j and k are each incremented by 1. Once all possible pairs of levels within the sounding have been compared, the level with the largest number of violations is removed. This process is then repeated until no more violations are found. Based on the height sequence check, approximately 0.003% of the levels in the dataset were removed.

Following the hypsometric and height sequence checks, each sounding is inspected for the existence of multiple surface levels. In soundings in which more than one surface level remains, all such levels are deleted. When a level containing only height and wind values is located at the same elevation as the surface pressure level, the two levels are merged into one surface level. Of the 28 million soundings processed, approximately 55% contained a valid surface pressure level, 8.4% required the merging of surface pressure and wind levels, and 0.04% contained multiple surface levels. In addition, a one-time manual inspection of the historical records of surface pressure and temperature was aimed at identifying gross shifts or inconsistencies in the two variables. This analysis revealed unrealistic features that prompted the removal of surface levels for

1968–70 at former Soviet Union stations as well as for 1967–72 and 1992–97 at Chinese stations.

Several of the data sources contain levels whose pressure or geopotential height is below the surface pressure or elevation of the station. In general, these “below surface” levels consist of data that have been extrapolated from the surface down to any mandatory pressure that happens to fall below the surface. When extrapolated levels are flagged as such in the source dataset, they are automatically excluded from IGRA. However, because some extrapolated levels are not correctly labeled and because transmission errors can also produce below-surface levels, an additional check identifies all types of below-surface levels. Specifically, a pressure level is considered to fall below the surface if its pressure is higher than the pressure of the surface level or its geopotential height is less than the height of the surface level. In a sounding without a valid surface level, any pressure level whose geopotential height is at least 10 m below the median elevation of the current month is removed. Based on these thresholds, 0.05% of the levels processed were identified as below-surface levels.

An examination of the data revealed the necessity for two additional simple consistency checks. In the check comparing the observation hour of a sounding with the corresponding reported launch time, soundings are deleted if the launch time deviates by more than 3 h from the observation hour. Differences of such magnitude were identified in approximately 0.25% of all soundings. Another check removes wind direction and speed when the speed is equal to 0 and the direction is neither 0° nor 360° , a condition found at 0.16% of all levels.

d. Checks for the repetition of values

The next set of checks looks for runs of values in time and in the vertical. A run is defined as the repetition of a value over a certain number of consecutive soundings or levels, ending with a change to another nonmissing data value; the absence of a value in a sounding or level does not interrupt a run.

The following four checks are applied:

- 1) a check for runs in surface pressure, surface- and mandatory-level temperature, and mandatory-level geopotential height that extend over more than 15 consecutive soundings;
- 2) an hour-specific (e.g., 0000 UTC) runs-in-time check analogous to check 1;
- 3) a procedure that looks for temperatures of the same value extending across at least five consecutive surface/mandatory levels or across at least five significant levels in a sounding; and

- 4) a pairwise vertical run check that identifies the repetition of the same value in either temperature and dewpoint depression or wind direction and speed over at least five consecutive pressure or height-only levels. Among the more interesting runs identified are cases of 40 consecutive 1000-hPa surface levels, -7.5°C temperatures at nine consecutive mandatory levels between 850 and 30 hPa in a sounding, ten 24.4°C temperatures at significant levels between 937 and 429 hPa, and 0 wind speed and direction throughout an entire sounding.

The manual inspection of extremely long runs also revealed the existence of several peculiar data problems. These problems consist of excessively frequent occurrences of certain temperature or geopotential height values within specific geographical regions, periods, data sources, and atmospheric levels. In the most egregious case, mandatory levels at and above 100 hPa (as well as at 1000 hPa) contain an unusually high number of 7.1°C temperatures at many stations during November and December 1967. All such values were eliminated by specifically designed checks, as they might otherwise seriously impact the quality of IGRA data. All in all, the various procedures for identifying excessive repetition of values removed approximately 0.02% of all data values.

e. Climatological checks

A two-tiered set of climatological checks removes geopotential height, temperature, and pressure values that deviate by more than a certain number of standard deviations (STDs) from their respective long-term means. In the first tier, the climatological means and STDs are calculated for the entire period of record for each station and pressure level, whereas in the second phase, the climatological statistics are stratified by time of year and time of day. Owing to their less stringent data requirement, the tier-1 checks can be applied to a larger number of data values than the tier-2 checks. On the other hand, the tier-2 checks allow for the use of tighter thresholds in the identification of outliers because their STDs do not reflect the seasonal and diurnal variations included in the tier-1 statistics. Furthermore, the tier-2 statistics are not computed until after the tier-1 checks have been applied and thus are based on a cleaner set of data.

The means and STDs of surface pressure and temperature as well as mandatory-level geopotential height and temperature are calculated using biweight statistics as described by Lanzante (1996). The biweight statistics tend to be more resistant to outliers that may be present in data that have not undergone advanced qual-

ity assurance. For the tier-1 checks, a mean and STD are produced as long as at least 120 values are available for a given station, level, and variable during the station's period of record. For the tier-2 checks, statistics are calculated for 45-day windows centered on each day of the year and in 3-h windows, provided that at least 150 values are available for any station, level, and variable in a given time interval. The means and STDs at other pressure levels (e.g., significant levels) are derived as needed by interpolating linearly with respect to the logarithm of pressure between the nearest adjacent mandatory levels. Recognizing that actual changes in temperature with height are not always linear, we compared the statistics derived by linear interpolation with those computed using all available levels (mandatory and significant) in 1-hPa slabs throughout the troposphere and stratosphere. Visual inspection of the two types of climatological profiles at a set of 87 globally distributed stations (Lanzante et al. 2003a) revealed few significant differences, suggesting the linearity assumption is viable from a quality assurance perspective.

To choose thresholds for labeling values as outliers, we visually compared, for all stations, the time series prior to the climatological checks to those following the application of the tier-1 and tier-2 checks, using various thresholds between three and seven STDs. We subjectively identified thresholds such that the algorithms neither remove a disproportionate number of values within the normal range of variability nor fail to remove a significant number of points that are clear outliers. In the tier-1 check, a threshold of six STDs was chosen for all three variables. For the tier-2 check, a threshold of five STDs was chosen for geopotential height, temperature, and below-normal surface pressure, and a threshold of four STDs was selected for above-normal pressure. (The asymmetric thresholds for above- and below-normal surface pressure were set in recognition of the fact that high-pressure anomalies tend to be smaller in magnitude than low-pressure anomalies.) These thresholds resulted in the removal of approximately 0.1% of all pressure, temperature, and geopotential height values by the tier-1 and tier-2 checks.

f. Additional checks on temperature

The inspection of various temperature time series and soundings revealed that the climatological check alone is incapable of satisfactorily removing all outliers without also removing realistic extremes. Figures 4 and 5 show examples of a time series and a sounding with outliers that are clearly erroneous when viewed in context with other temperatures within their temporal and vertical vicinity. However, to address outliers that pass

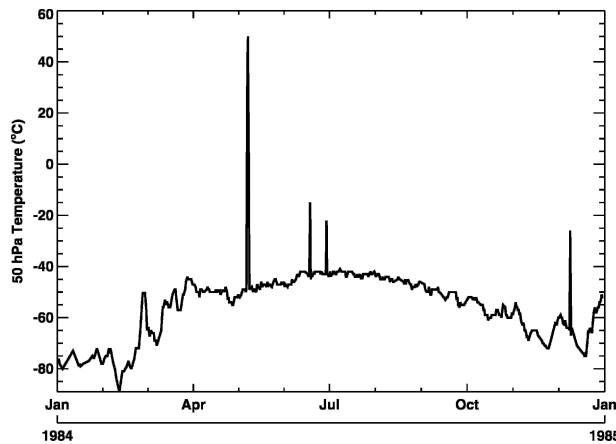


FIG. 4. Time series of temperature reported at the 50-hPa pressure level at Jan Mayen, Norway, during 1984.

the climatological checks but are vertically or temporally inconsistent, additional vertical and temporal consistency checks were developed specifically for temperature. These procedures are described briefly here and in more detail in a separate paper that is in preparation at the time of this writing.

The supplemental vertical consistency checks for temperature employ z-score profiles derived from the tier-2 climatological means and STDs. For instance, an entire temperature profile is eliminated if it is judged to

be grossly abnormal in terms of either its median z score or its median absolute level-to-level z-score difference. Additional checks remove one or more temperatures from a profile if the z scores are clearly inconsistent with either the entire profile or values at adjacent levels. When applied to IGRA, the procedures together identified 0.08% of all temperatures as vertically inconsistent.

Two temporal consistency checks are also applied to surface and mandatory-level temperatures. These checks are based on z scores derived using the overall mean and STD for any station and level, provided that at least 120 such values remain following the climatological and vertical consistency checks. The first identifies outliers that differ by more than two STDs from all other temperatures within ± 22.5 days, while the second variant uses a difference threshold of one STD and time window of 2.5 yr on either side of the potential outlier. Both variants examine only those temperatures whose absolute z score is greater than 2.5 and require that temperatures be available on at least half of the days in the time window. The temporal consistency checks together removed approximately 0.004% of the temperatures from IGRA.

g. Checks for data completeness

The IGRA quality assurance process also ensures that the dataset adheres to certain minimum requirements for completeness. For example, each station must have at least 100 soundings. An “isolated sounding check” eliminates groups of fewer than three soundings surrounded by at least 31 days without data, groups of fewer than 15 soundings surrounded by gaps of three months (92 days), and groups of fewer than 28 soundings flanked by gaps of half a year (182.5 days).

Within a sounding, wind speed and direction must always appear together, and a dewpoint depression may exist only if it is accompanied by a temperature at the same level. A pressure level is retained if it contains valid thermodynamic data and/or valid wind data. Levels with a height but no pressure are permitted to exist if they contain valid wind data. A sounding may consist of any combination of pressure levels and height–wind levels, as long as there is at least one nonsurface level.

4. Availability of data and metadata

IGRA is available at no charge from the NCDC Web site. In addition to the individual soundings, NCDC provides monthly means of geopotential height, temperature, as well as zonal and meridional wind components at the surface, tropopause, and mandatory levels for the nominal times of 0000 and 1200 UTC.

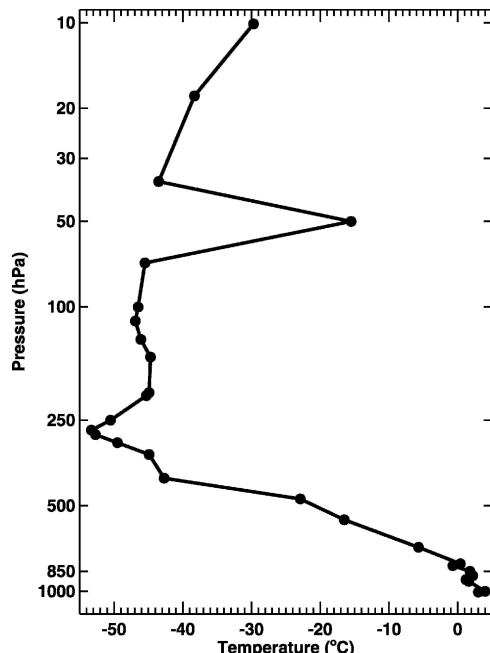


FIG. 5. Vertical profile of temperature reported at Jan Mayen, Norway, at 0000 UTC 18 Jun 1984, which is the time of an outlier shown in Fig. 4 at the 50-hPa pressure level.

IGRA is updated on a daily basis with GTS messages received as part of the NCDC real-time GTS data source (Table 1) on the previous day. Using the same procedures that were applied to the historical data, the update process ensures that soundings and levels are properly sorted, removes duplicate levels and soundings, and employs all applicable quality assurance procedures. Checks that require data for periods of time longer than a few days, such as the runs-in-time check and the check for temporal consistency in temperature, are not applied as part of the daily update process. These algorithms will instead be applied when revised versions of IGRA are created.

At present, IGRA metadata include the name, most recent location, and period of record of each station. Additional metadata, available from the authors upon request, include the station history information collected by Gaffen (1996) and recently updated through contacts with representatives from WMO member countries. Although many data sources contained additional metadata, such as the type of radiosonde used in a sounding, inconsistencies in the coding conventions used over time and across data sources complicate efforts to interpret this information and reconcile it with the other available station history information. Consequently, the processing of this information was left for future versions of IGRA.

5. Description of the dataset

IGRA consists of quality-assured soundings at over 1500 globally distributed stations with varying periods of record. Although the overall period of record is 1938 to present, the length and completeness of a record vary widely among stations, and the vertical resolution, vertical extent, and completeness of soundings improve considerably over time. Mandatory levels generally include geopotential height, temperature, wind direction, and wind speed. Beginning in 1969, dewpoint depression is usually also available in the lower and middle troposphere but becomes scarcer in the upper troposphere because of the general practice to discontinue humidity measurements at temperatures less than -40°C (Elliott and Gaffen 1991; Garand et al. 1992). Temperature and dewpoint depression are also available at significant thermodynamic levels (which usually do not include geopotential height). Wind observations are reported at significant thermodynamic levels or at separate levels whose elevation is defined by pressure and/or height.

The dataset contains slightly more than 28 million soundings with a total of 800 million levels. Approximately 20 million of the soundings from roughly 1250 stations contain temperature measurements, with the

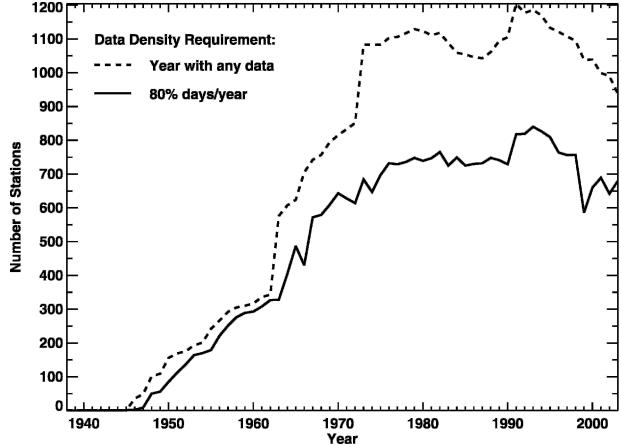


FIG. 6. Number of stations in IGRA for each year where data are available on at least one day of the year (dashed line) and at least 80% of the days in the year (solid line).

remainder consisting of only pibal observations. As shown in Table 1, 82% of the soundings originate from the GTS-based core data sources, while the other large-scale sources and country-specific datasets contribute 6% and 12%, respectively. The most frequent observation times available are 0000 and 1200 UTC beginning in 1958 and 0300 and 1500 UTC before that year. The majority of stations take observations twice daily at or near those observation times, and some provide observations even more frequently; however, a number of stations only have one sounding per day for extended periods due to a lack of equipment or observers.

As indicated by Fig. 1, IGRA contains stations in most areas of the globe. The spatial coverage is most complete in Europe and sparsest in northern Canada, interior Antarctica, and equatorial Africa. However, the total number, spatial distribution, and temporal completeness of stations vary considerably over time (Fig. 6). For each year between 1938 and 2003, Fig. 6 displays the total number of stations (dashed line) and the number of stations reporting one or more soundings on at least 80% of possible days (solid line). During the early part of the record, the number of stations increases from one station (in Tasmania) in 1938 to several hundred in the early 1960s when most of the stations report data on more than 80% of the days. By the time the number of stations peaks in 1991, approximately 840 of the available 1180 stations report at least one sounding on at least 80% of the days. Station closings are responsible for the decline in the number of stations in recent years. Relative to the map of all stations (Fig. 1, top), the distribution of the 937 stations active in 2003 (Fig. 1, bottom) exhibits the most pronounced deficit in western equatorial Africa.

The jumps in the number of stations in 1946, 1963, and 1973 (Fig. 6) are related to changes in the number or type of data sources contributing to IGRA (Table 1). Before the beginning of the first GTS data source in September 1963, the U.S. Air Force and country-specific U.S. sources each account for nearly half of the soundings (approximately 48% and 44%, respectively), with the remainder provided by the country-specific sources for Australia (nearly 8%) and Argentina (<1%). Consequently, during this early period, IGRA stations are concentrated in the contiguous United States, Alaska, and the former Soviet Union, with additional stations in parts of the North Atlantic, Southeast Asia, Argentina, and coastal Australia. With the jump in 1963, coverage of western Europe, China, and Japan begins, while many stations in Africa, Brazil, central Asia, and India do not become available until the late 1960s or early 1970s.

The change in vertical resolution and extent over time is illustrated by time series of the average number of mandatory and total levels per sounding (Fig. 7) as well as time series of the percentage of soundings reaching up to at least 100 or 10 hPa (Fig. 8). Before the 1960s, soundings consist primarily of mandatory levels below 100 hPa. By the early 1960s, most of the soundings contain observations up to the 100-hPa level and include some significant levels. The addition of large numbers of stations with varying degrees of data completeness accounts for the overall drop in the percentage of soundings reaching into the stratosphere during the late 1960s and 1970s. Overall, however, the vertical resolution of soundings continues to improve, as indicated by the rather monotonic rise in the total number of levels per sounding. By 2003, the average sounding consists of 11 mandatory and 35 additional levels, and

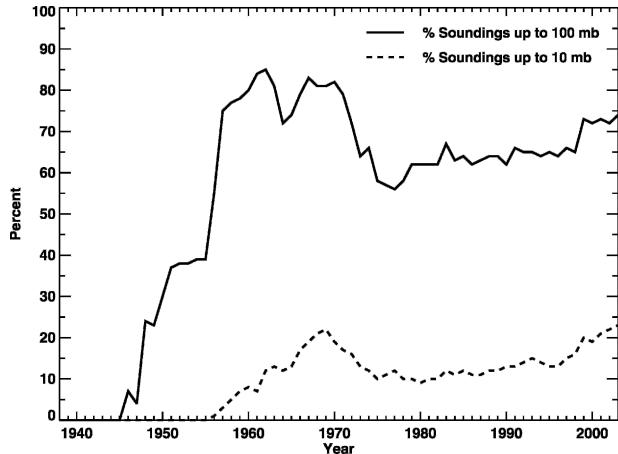


FIG. 8. Percentage of soundings reaching up to at least 100 (solid line) and 10 hPa (dash line).

74% (35%) of all soundings reach at least a 100-hPa (10-hPa) level.

6. Comparison with other global-scale datasets

As mentioned in the introduction, there have been other efforts to compile global-scale radiosonde datasets. The most recent of these efforts have been undertaken in support of reanalysis projects at NCEP (Kalnay et al. 1996) and the European Centre for Medium-Range Weather Forecasts (ECMWF) (Uppala 2005). Another relevant compilation is the CARDS dataset previously produced by NCDC (Eskridge et al. 1995). Like IGRA, these three data archives merged together numerous sources of upper-air data and applied some degree of quality assurance. An obvious question is, "How different are these compilations from IGRA?" The most straightforward answer lies in the relative accessibility of each archive as well as the quantity of data contained therein.

As discussed in greater detail in the appendix, user access to CARDS is hampered by inconsistencies and complexities in its data format and inadequacies in its quality assurance system, complications that are not present in IGRA. A comparison of the data holdings of IGRA and CARDS (see the appendix) indicates that the two datasets differ in terms of the number, length, completeness, and overall quality of station records. Even though CARDS contains 868 additional stations, the vast majority of their records are rather short and incomplete. While CARDS provides greater spatial coverage before 1970, IGRA exhibits somewhat greater post-1990 data coverage (Fig. A1). In this respect, the results are somewhat similar to those obtained from Haimberger's (2005) ERA-40-to-IGRA

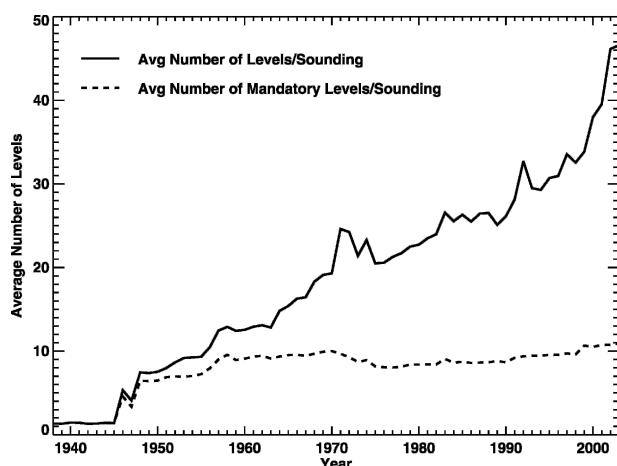


FIG. 7. Average number of mandatory (dash line) and total (solid line) levels per sounding for each year of data in IGRA.

comparison and can be attributed to the IGRA project's particular attention to record integrity. This interpretation is further supported by an analysis of selected time series of monthly mean temperature anomalies derived from the two datasets (Free et al. 2005), which reveals that the CARDS station records for the surface and 1000-hPa level are somewhat more likely to exhibit unrealistic shifts than the corresponding IGRA time series (e.g., Fig. A2).

While IGRA is available online as station-by-station ASCII files, the reanalysis input data are not as readily accessible, thus complicating efforts at direct data comparisons. Since the input data to the NCEP–NCAR reanalysis are distributed only in the form of individual data sources rather than as a single comprehensive dataset (J. Woolen 2005, personal communication), a direct comparison between IGRA and the NCEP–NCAR dataset is not possible. The ECMWF's archive of 40-yr ECMWF Re-Analysis (ERA-40) input data, on the other hand, can be requested from NCAR, albeit not in a format readily suitable for the analysis of station time series. Consequently, a comparative analysis of the amounts of data in the ECMWF and IGRA archives is only feasible after considerable data processing. Nevertheless, such a comparison has been performed by Haimberger (2005), who found that the ERA-40 input dataset contains more data before the 1990s, while IGRA contains a larger number of soundings in the 1990s. This finding is consistent with the results from the IGRA/CARDS comparison (Fig. A1).

7. Conclusions

IGRA consists of historical records of quality-assured soundings from 1500 globally distributed stations. The historical data and real-time updates are available online from NCDC along with relevant inventories and station history information. The archive is the result of integrating data from 11 different sources and applying a sequence of specialized quality assurance algorithms. Even though IGRA provides fewer stations and less pre-1970 spatial coverage than other global-scale datasets, its records tend to exhibit a higher level of completeness and integrity. In general, the highest-quality data in IGRA are temperature, geopotential height, and surface pressure at stations with relatively complete records in these variables.

Given these characteristics, IGRA is suitable for a wide range of applications, including, for example, comparisons between measurements from radiosondes and other observing systems, the verification of output from model simulations, and studies of boundary layer structure. Since IGRA data have not yet been adjusted

for inhomogeneities resulting from changes in instrumentation or observing practices, users interested in utilizing these data for climate change analyses are advised to refer to one of several IGRA-derived products. Currently, these include the Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC; Free et al. 2005) and Haimberger's (2005) station time series adjusted with the Radiosonde Observation Bias Correction Using Reanalyses (RAOB-CORE).

In constructing the next version of IGRA, the primary goal will be to acquire and integrate data for periods and regions for which an improvement in data coverage is most needed. With the availability of additional data sources such as the ERA-40 input dataset (Uppala 2005) and a collection of World War II era observations (Brönnimann 2003), it may be possible to augment the early records in IGRA without compromising our requirements for record integrity. An enhancement in temporal coverage should also be feasible at a number of Chinese stations where the reliable data sources at our disposal during the construction of version 1 of IGRA lacked data between 1973 and 1990. In addition, recently digitized data from certain African countries may help to enhance the overall data coverage in a part of the world where data are particularly sparse.

Finally, during future revisions of IGRA, improvements to the quality assurance system will also be explored. Potential enhancements include climatologically based temporal and vertical consistency checks on geopotential height, an algorithm for the identification of invalid tropopause levels, procedures for detecting unrealistically large wind speeds, and additional checks on dewpoint depression. Furthermore, the utility of both spatial consistency checks and comparisons with first-guess fields from the NCEP–NCAR and ERA-40 reanalyses in the quality assurance process will be investigated.

Acknowledgments. We thank Byron Gleason and Claude Williams for providing helpful and constructive comments on an initial version of the manuscript. Suggestions from two anonymous reviewers contributed to the further improvement of this paper.

APPENDIX

Comparison between IGRA and CARDS

During the 1990s, the CARDS project acquired radiosonde data from over 20 different sources and placed them into one common format in which they are

stored in the NCDC archive. The various sources were then combined into one dataset and passed through a quality assurance system based on Gandin's (1988) concept of complex quality control (CQC). Despite efforts to overcome difficulties with incomplete documentation, unreadable sections of tape, and limitations in storage capacities during the reformatting process, the resulting archived source datasets contain residual inconsistencies, some of which made it through into the final CARDS dataset. These include inconsistencies in station numbering, undocumented or unreadable variations in the data format, and duplicate records of various types. In addition, miscommunication led to misidentification of some station numbers or observation times for a portion of soundings in one of the principal sources of CARDS data for the late 1950s and early 1960s [the Massachusetts Institute of Technology dataset obtained from NCAR].

Feedback from users highlighted other areas requiring attention. Examples include the presence of impossible surface levels (e.g., at 70 hPa), an extensive amount of rounding of highly precise wind measurements, and the removal of some near-surface temperature inversions in regions where strong inversions are common. These issues can be traced to the fact that the CQC procedures do not fully address the most egregious problems and ignore fundamental properties of the atmosphere. For example, the systematic removal of strong but realistic inversions may be the result of the system's reliance on a combination of the hydrostatic balance and synoptic-scale relationships (Eskridge et al. 1995) when evaluating instantaneous observations whose variability is likely to contain a significant subsynoptic component.

Even though the IGRA project made use of the same reformatted datasets, its carefully tested procedures avoid the inconsistencies and deficiencies present in CARDS by focusing on the identification of reliable data records, the detection of the most significant and most common types of errors, and the preservation of local phenomena. The development of these techniques was aided not only by an awareness of the types of problems encountered by the CARDS project, but also by more advanced computing capabilities and experience with successful surface datasets such as GHCN (Peterson and Vose 1997).

As a result of the differing processing approaches, the two datasets differ in terms of the number, length, completeness, and, in some cases, the overall quality of station records. When counted by station number, 1491 stations are common to both datasets, 45 are found only in IGRA, and 1021 are found only in CARDS. Of the IGRA stations not found in CARDS, the majority (33)

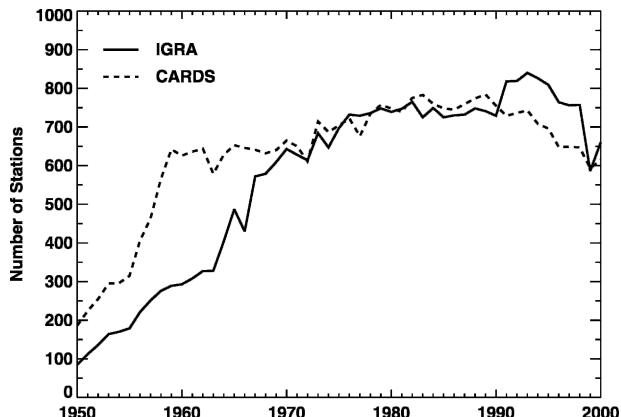


FIG. A1. Number of stations per year with at least one sounding on at least 80% of the days in the year for IGRA (solid line) and CARDS (dashed line).

began reporting data during or after 1990, and many report only wind observations. Data for 154 of the additional CARDS station numbers are contained in IGRA as part of the composite records of more recent stations. Many of the remaining 868 stations found only in CARDS have rather short or incomplete records and, thus, do not augment the volume of data nearly as much as the sheer number of stations may suggest.

A year-by-year comparison of the two datasets is provided by a plot of the number of 80% complete stations per year in CARDS and IGRA for the common period of record from 1948 to 2000 (Fig. A1). Here, the line plotted for IGRA is analogous to the corresponding line in Fig. 6, and a station is counted for a particular year and dataset if its record in the respective dataset contains at least one sounding on at least 80% of the days in that year. The figure indicates that CARDS contains a considerably larger number of stations until the early 1960s, when the number of stations available in IGRA begins to increase rapidly. During the 1970s and 1980s, the two datasets contain approximately the same number of 80% complete stations, although some year-to-year variation is apparent. For much of the 1990s, the number of such stations is approximately 100 larger in IGRA than in CARDS. The differences before the 1970s are the result of the stricter requirements for the inclusion of stations and sources that are employed in the construction of IGRA. These requirements lead to the exclusion of a larger fraction of stations from the first half of the record than from the second half since both the confidence in the identification of stations and the level of agreement among data sources are lower in earlier years.

While a detailed comparison of the IGRA and CARDS quality assurance systems is beyond the scope

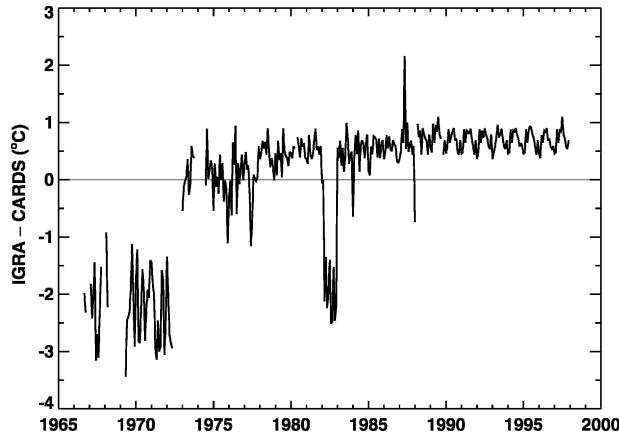


FIG. A2. IGRA–CARDs difference in monthly 1000-hPa temperature anomalies for 0000 UTC at Darwin, Australia. For each month of the year, anomalies are computed relative to all years available in both datasets before the resulting CARDs anomalies are subtracted from the corresponding IGRA anomalies.

of this paper, some insight can be gained from a comparison of time series of monthly temperature anomalies for the surface and mandatory levels at the 87 Lanzante et al. (2003a) stations (Free et al. 2005). While the time series from the two datasets tend to capture the same variability, some significant differences exist at the surface and 1000-hPa levels where data have previously been found to be particularly problematic (Gandin et al. 1993; Lanzante et al. 2003a). As exemplified by the IGRA–CARDs differences in 0000 UTC 1000-hPa temperature anomalies for Darwin, Australia (Fig. A2), a number of these cases are the result of an unrealistically large shift in the CARDs time series that is either reduced or not present in the corresponding IGRA data. In the Darwin case, the greater homogeneity and reduced length of the IGRA time series is likely to be related to the exclusion of the Australian country-specific source data due to poor agreement with the GTS data. The above discussion reveals that, even though CARDs contains a larger number of stations and some longer records than IGRA, the IGRA records tend to be more homogeneous and robust as a result of the specific merging and quality assurance procedures employed. As discussed in the main text, the acquisition of additional data sources, particularly for the early portion of the record, is likely to help further improve the spatial and temporal coverage in future versions of IGRA without compromising on the quality of the records.

REFERENCES

- Brönnimann, S., 2003: A historical upper-air data set for the 1939–44 period. *Int. J. Climatol.*, **23**, 769–791.
- Collins, W. G., 2001a: 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.*, **40**, 137–151.
- , 2001b: The operational complex quality control of radiosonde heights and temperatures at the National Centers for Environmental Prediction. Part II: Examples of error diagnosis and correction from operational use. *J. Appl. Meteor.*, **40**, 152–168.
- Durre, I., T. Reale, D. Carlson, J. Christy, M. Uddstrom, M. Gelman, and P. Thorne, 2005: Improving the usefulness of operational radiosonde data. *Bull. Amer. Meteor. Soc.*, **86**, 411–418.
- Elliott, W. P., and D. J. Gaffen, 1991: On the utility of radiosonde humidity archives for climate studies. *Bull. Amer. Meteor. Soc.*, **72**, 1507–1520.
- , R. J. Ross, and W. Blackmore, 2002: Recent changes in NWS upper-air observations with emphasis on changes from VIZ to Vaisala radiosondes. *Bull. Amer. Meteor. Soc.*, **83**, 1003–1017.
- Eskridge, R. E., O. A. Alduchov, I. V. Chernykh, P. Zhai, A. C. Polansky, and S. R. Doty, 1995: A Comprehensive Aerological Reference Data Set (CARDs): Rough and systematic errors. *Bull. Amer. Meteor. Soc.*, **76**, 1759–1775.
- FCM-H3, cited 2004: Federal Meteorological Handbook No. 3: Rawinsonde and pibal observations. [Available online at <http://www.ofcm.gov/fmh3/text/default.htm>.]
- Finger, F. G., and F. J. Schmidlin, 1991: Meeting review: Upper-Air Measurements and Instrumentation Workshop. *Bull. Amer. Meteor. Soc.*, **72**, 51–56.
- Free, M., and Coauthors, 2002: CARDs workshop on adjusting radiosonde temperature data for climate monitoring: Meeting summary. *Bull. Amer. Meteor. Soc.*, **83**, 891–899.
- , J. K. Angell, I. Durre, J. R. Lanzante, T. C. Peterson, and D. J. Seidel, 2004: Using first differences to reduce inhomogeneity in radiosonde temperature datasets. *J. Climate*, **17**, 4171–4179.
- , 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 dataset of large-area anomaly time series. *J. Geophys. Res.*, **110**, D22101, doi:10.1029/2005JD006169.
- Gaffen, D. J., 1994: Temporal inhomogeneities in radiosonde temperature records. *J. Geophys. Res.*, **99**, 3667–3676.
- , 1996: A digitized metadata set of global upper-air station histories. NOAA Tech. Memo. ERL ARL-211, 37 pp. [NTIS PB 96147491; available from Dian J. Seidel, NOAA/Air Resources Laboratory (R/ARL), 1315 East-West Highway, Silver Spring, MD 20910.]
- Gandin, L. S., 1988: Complex quality control of meteorological observations. *Mon. Wea. Rev.*, **116**, 1137–1156.
- , L. L. Morone, and W. G. Collins, 1993: Two years of operational comprehensive hydrostatic quality control at the National Meteorological Center. *Wea. Forecasting*, **8**, 57–72.
- Garand, L., C. Grassotti, J. Hall, and G. L. Klein, 1992: On differences in radiosonde humidity-reporting practices and their implications for numerical weather prediction and remote sensing. *Bull. Amer. Meteor. Soc.*, **73**, 1417–1423.
- Haimberger, L., 2005: Homogenization of radiosonde temperature time series using ERA-40 analysis feedback information. ERA-40 Project Report Series, No. 23, ECMWF, Reading, United Kingdom, 73 pp. [Available online at <http://www.ecmwf.int/research/reports/era40/report23.pdf>.]

- [ecmwf.int/publications/library/ecpublications/_pdf/era40/ERA40_PRS23.pdf.\]](http://ecmwf.int/publications/library/ecpublications/_pdf/era40/ERA40_PRS23.pdf)
- Kahl, J. D., M. C. Serreze, S. Shiotani, S. M. Skony, and R. C. Schnell, 1992: In situ meteorological sounding archives for Arctic studies. *Bull. Amer. Meteor. Soc.*, **73**, 1824–1830.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Lanzante, J. R., 1996: Resistant, robust and nonparametric techniques for analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.*, **16**, 1197–1226.
- , —, and —, 2003a: Temporal homogenization of monthly radiosonde temperature data. Part I: Methodology. *J. Climate*, **16**, 224–240.
- , —, and —, 2003b: Temporal homogenization of monthly radiosonde temperature data. Part II: Trends, sensitivities, and MSU comparison. *J. Climate*, **16**, 241–262.
- Loehrer, S. M., T. A. Edmonds, and J. A. Moore, 1996: TOGA COARE upper-air sounding data archive: Development and quality control procedures. *Bull. Amer. Meteor. Soc.*, **77**, 2651–2672.
- NGDC, cited 2004: Global Land One-kilometer Base Elevation (GLOBE). [Available online at <http://www.ngdc.noaa.gov/seg/topo/globe.shtml>.]
- NRC Panel on Reconciling Temperature Observations, 2000: *Reconciling Observations of Global Temperature Change*. National Academy Press, 85 pp.
- Peterson, T. C., and R. S. Vose, 1997: An overview of the Global Historical Climatology Network Temperature Database. *Bull. Amer. Meteor. Soc.*, **78**, 2837–2849.
- Schwartz, B. E., and C. A. Doswell III, 1991: North American rawinsonde observations: Problems, concerns, and a call to action. *Bull. Amer. Meteor. Soc.*, **72**, 1885–1896.
- , and M. G. Govett, 1992: A hydrostatically consistent North American radiosonde data base at the Forecast Systems Laboratory, 1946–present. NOAA Tech. Memo. ERL FSL-4, 81 pp. [NTIS PB 112225.]
- Thorne, P. W., D. E. Parker, S. F. B. Tett, P. D. Jones, M. McCarthy, H. Coleman, P. Brohan, and J. R. Knight, 2005: Revisiting radiosonde upper-air temperatures from 1958 to 2002. *J. Geophys. Res.*, **110**, D18105, doi:10.1029/2004JD005753.
- Uppala, S. M., and Coauthors, 2005: The ERA-40 ReAnalysis. *Quart. J. Roy. Meteor. Soc.*, in press.
- WMO, 1996: *Guide to Meteorological Instruments and Methods of Observation*. 6th ed. WMO Rep. 8, World Meteorological Organization, Geneva, Switzerland.
- , cited 2004: *Observing Stations*. Vol. A, *Weather Reporting*, Publication 9. [Available online at <http://www.wmo.ch/www/ois/volume-a/vola-home.htm>.]